

Computational Modeling of Toxicity and Ecotoxicity of Engineered Nanoparticles

Thesis submitted
by
JOYITA ROY

Doctor of Philosophy (Pharmacy)

Department of Pharmaceutical Technology
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India
2024

Dedicated to science

**JADAVPUR UNIVERSITY
KOLKATA – 700 032, INDIA**

1. Title of the thesis:

Computational Modeling of Toxicity and Ecotoxicity of Engineered Nanoparticles

2. Name, Designation & Institution of the Supervisors:

Dr. Kunal Roy

Professor

Drug Theoretics and Cheminformatics Laboratory

Department of Pharmaceutical Technology Jadavpur University

Kolkata 700 032

West Bengal, India

3. List of publications

Research Articles

Related to thesis

1. **Roy, J.** and Roy, K., 2024. Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors. Beilstein J. Nanotechnol. 2024, 15, 297–309. <https://doi.org/10.3762/bjnano.15.27>.
2. **Roy, J.**, Pore, S. and Roy, K., 2023. Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches. Beilstein Journal of Nanotechnology, 14(1), pp.939-950.
3. **Roy, J.** and Roy, K., 2023. Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: a nano-QSAR approach. SAR and QSAR in Environmental Research, pp.1-16.
4. **Roy, J.** and Roy, K., 2022. Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells. Nanotoxicology, 16(5), pp. 629-644. <https://doi.org/10.1080/17435390.2022.2132887>.
5. **Roy, J.** and Roy, K., 2022. Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to Escherichia coli: Categorization and data gap filling for untested metal oxides. Nanotoxicology, 16(2), pp.152-164. <https://doi.org/10.1080/17435390.2022.2038299>.
6. **Roy, J.** and Roy, K., 2022. Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (Me Ox NPs) used in nanomedicines: a mechanistic QSAR approach. Environmental Science: Nano, 9(9), pp.3456-3470. DOI: [10.1039/D2EN00303A](https://doi.org/10.1039/D2EN00303A).
7. **Roy, J.** and Roy, K., 2021. Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach. Environmental Science: Nano, 8(11), pp.3395-3407. DOI: [10.1039/D1EN00733E](https://doi.org/10.1039/D1EN00733E).

Publications not related to thesis

1. Paul, R., **Roy, J.** and Roy, K., 2023. Prediction of soil ecotoxicity against *Folsomia candida* using acute and chronic endpoints. SAR and QSAR in Environmental Research, 34(4), pp.321-340.
2. De, P., **Roy, J.**, Bhattacharyya, D. and Roy, K., 2020. Chemometric modeling of PET imaging agents for diagnosis of Parkinson's disease: a QSAR approach. Structural Chemistry, 31, pp.1969-1981.
3. **Roy, J.**, Ojha, P.K., Carnesecchi, E., Lombardo, A., Roy, K. and Benfenati, E., 2020. First report on a classification-based QSAR model for chemical toxicity to earthworm. Journal of hazardous materials, 386, p.121660. <https://doi.org/10.1016/j.jhazmat.2019.121660>.
4. **Roy, J.**, Ghosh, S., Ojha, P.K. and Roy, K., 2019. Predictive quantitative structure–property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs). Environmental Science: Nano, 6(1), pp.224-247. DOI: [10.1039/C8EN01059E](https://doi.org/10.1039/C8EN01059E)
5. **Roy, J.**, Ojha, P.K. and Roy, K., 2019. Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table based descriptors. Nanotoxicology, 13(5), pp.701-716. <https://doi.org/10.1080/17435390.2019.1593543>.

Review

6. **Roy, J.** and Roy, K., 2023. Evaluating metal oxide nanoparticle (MeOx NP) toxicity with different types of nano descriptors mainly focusing on simple periodic table-based descriptors: A mini-review. Environmental Science: Nano..
7. Ojha, P.K., Kumar, V., **Roy, J.** and Roy, K., 2021. Recent advances in quantitative structure–activity relationship models of antimalarial drugs. Expert Opinion on Drug Discovery, 16(6), pp.659-695. <https://doi.org/10.1080/17460441.2021.1866535>.

4. List of patents: **Nil**

5. List of presentation /(s) in National /International/Conferences/ Workshops:

International conference/ Workshops

- **Joyita Roy**, Kunal Roy, (2022). **Poster presentation- In-person:** Exploring toxicity of metal oxide engineered nanoparticles (MeOx ENPs) used in nanopesticides to BEAS-2B and RAW 264.7 cells with QSTR, Nano QSTTR and Read-Across Approaches. **23rd Euro QSAR 2022**, 23rd Symposium on Quantitative Structure-Activity Relationship, **September 26-30, 2022**, Heidelberg, Germany.
- **Joyita Roy**, Kunal Roy (2022). Assessing cytotoxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines using a QSTR approach. **ACS Fall 2022**, Chicago-IL and Hybrid, **August 21 - 25, 2022**.

- **Joyita Roy**, Kunal Roy (2021) Assessment of toxicity of metal oxide and hydroxide nanoparticles using QSAR modeling approach. International Symposium on Drug Design and Development Research (**DDDR- 2021**), **December 17-18, 2021**, GIPER, Kashipur, India.
- **Joyita Roy**, Kunal Roy (2021). Nano-QSTR modeling of metal oxide nanoparticles (MeOxNPs) to E. coli: Application of the Small Dataset Modeler tool **OpenTox Euro 2021**, OpenTox Virtual Conference on **September 20-24, 2021**.
- **Joyita Roy**, Kunal Roy (2020). Chemometric modeling of PET imaging agents for diagnosis of Parkinson's disease: a QSAR approach. International Conference on Currents Trends in Pharmaceutical and Medical Sciences, **February 26-29, 2020**, GIPER, Kashipur, India.

National conference

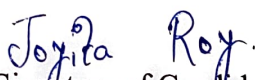
- **Joyita Roy**. Attended webinar on “Young Scientist Colloquium 2023” organized by Material Research Society of India (MRSI), Kolkata in association with School of Material Science and Nanotechnology, Jadavpur University in collaboration with Indian Institute of engineering and Science and technology, Shibpur on **1st December 2023**.
- **Joyita Roy**. Attended webinar on 8th Indo-US Workshop on Mathematical Chemistry 2022 organised by Sharda University held on **September 13-17, 2022**.
- **Joyita Roy**. Attended webinar on Health Informatics Summit organized by the Department of Computational Biology, Indraprastha Institute of Information Technology Delhi (IIIT-D) and APbians (Bioinformatics Society) held on **October 16-19, 2021**.
- **Joyita Roy**. Attended webinar on “Biodiversity and Environmental Protection during Pandemic Outbreak of COVID 19” organized by International Society of Waste Management Air and Water (ISWMAW) jointly by Chairman Prof. Dr. Sadhan Kumar Ghosh of India and Co-Chairman Dr. Abas Basir of Sri Lanka on **5th June 2020**.

Workshops, Training

- Drug Discovery Hackathon 2020.

“Statement of Originality”

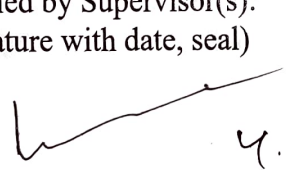
I, **Joyita Roy** registered-on **27th July, 2020** do hereby declare that this thesis entitled **“Computational Modeling of Toxicity and Ecotoxicity of Engineered Nanoparticles”** contains a literature survey and original research work done by the undersigned candidate as part of Doctoral studies. All information in this thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work. I also declare that I have checked this thesis as per the “Policy on Anti Plagiarism, Jadavpur University, 2019”, and the level of similarity as checked by iThenticate software is **9%***.


Signature of Candidate:

Date: **4/4/2022**

Certified by Supervisor(s):
(Signature with date, seal)

1.


4. 9. 2022
DR. KUNAL ROY
PROFESSOR & EX-HEAD
DEPT. PHARMACEUTICAL TECHNOLOGY
JADAVPUR UNIVERSITY
KOLKATA - 700 032, INDIA

*Ignoring smaller match of up to 14 words (as per the UGC) and ignoring own publications as the source.

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled “*Computational Modeling of Toxicity and Ecotoxicity of Engineered Nanoparticles*” submitted by **Smt. Joyita Roy** who got her name registered on **27th July, 2020** for the award of Ph.D. (Pharmacy) degree of Jadavpur university is absolutely based upon her own work under the supervision of **Dr. Kunal Roy** that neither her thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

1. _____

(Signature with Date and Seal)

DR. KUNAL ROY
PROFESSOR & EX-HEAD
DEPT. PHARMACEUTICAL TECHNOLOGY
JADAVPUR UNIVERSITY
KOLKATA - 700 032, INDIA

Acknowledgements

I deem it a pleasure and privilege to work under the guidance of **Dr. Kunal Roy**, Professor, Drug Theoretics & Cheminformatics Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata. I express my deep gratitude and regard to my revered mentor for suggesting the subject of this thesis and rendering me his thoughtful suggestions and rational approaches to this thesis work. I am also greatly indebted to him for his valuable guidance throughout my PhD tenure that enabled me to complete the work.

I am thankful to the Research Advisory Committee members **Prof. Tarun Jha**, and **Prof. Tapan Kr. Maity** for timely evaluating my progress and also for their valuable feedback on my work. I also express my gratitude to the Head, Professors, Registrar, Dean, Faculty of Engineering and Technology, non-teaching staff of the Department of Pharmaceutical Technology, Jadavpur University and the entire University Authority for facilities provided to carry out this work. My sincere thanks to the also thank the **Department of Atomic Energy-Board of Research in Nuclear Sciences (DAE-BRNS)** and **Indian council of medical research (ICMR), New Delhi** for providing financial assistance in the form of a **Senior Research Fellowship (SRF)** in the past four years.

With a deep sense of thankfulness and sincerity, I acknowledge the continuous encouragement, perpetual assistance, and cooperation from my senior Dr. Probir Kr. Ojha, Dr. Vinay Kumar and Mainak Chatterjee. Their constant support and helpful suggestions have helped me to accomplish this work in time. I would like to express my special thanks to my friends and my juniors Arkaprava Banerjee, Sapna Kumari Pandey, Souvik Pore and Shilpayan Ghosh, Ankur Kumar, Subham Kr. Pandey, Doelima Bera, and MD. Mobarak Hossain who all have extended their helping hands and friendly cooperation all through my work.

A word of thanks to all those people associated with this work directly or indirectly whose names I have been unable to mention here. Finally, I would like to thank my parents **Mr. Hiralal Roy** and **Mrs. Debika Roy** for their unconditional love and affection; their support keep me calm and strong during many undulations in the path of life. I would like to specially acknowledge my husband **Mr. Kuntal Saha Chowdhury**, for his constant support and encouragement without which this work would have been incomplete. I would also like to acknowledge by mother-in-law **Mrs. Swapna Saha Chowdhury** and father-in-law **Mr. Chandra shekhar Saha Chowdhury** for their support and patience through by dissertation work.

Date: 4/4/2024
Place: Kolkata

Joyita Roy
JOYITA ROY

Preface

The research work presented in this dissertation is the result of constant effort over the last four years. This dissertation is mainly focused on different computational tools and techniques that have been employed for the prediction of toxicity of nanoparticles against different relevant toxicity endpoints. Nowadays, different *in silico* techniques are being routinely applied for predicting toxicities and properties of different nanoparticles as a rational alternative technique to animal testing. The prime objective of this dissertation was to develop *in silico* modeling of nanoparticles, especially metal oxide nanoparticles, and use those models, further for risk assessment. Nanotechnology shows beneficial application starting from cosmetic industry, food industry, material science or in communication and environmental sector for removing pollution. The rapid increase in engineered nanoparticles in recent years is because of their unique magnetic, electronic, and optical properties. Along with the advantages of nanotechnology, there is some problems or risk related to it. It has been observed that nanoparticles may endanger ecosystem and other species through generation of different cytotoxic responses (such as cytogenecity, mutagenicity or ecotoxicity). These nanoparticles are mostly present in the environment as multicomponent but we have very little nanoparticle toxicity knowledge and safety data. Experimental toxicity determination is an unfeasible approach to fill the existing toxicity data gap. Animal testing cannot alone suffice the need for the entire assessment of those nanoparticles. So, an alternative method should come in. Hence, assessment of properties and toxicity of the existing as well as new nanoparticle is necessary. The *in silico* modeling techniques have been proven advantageous over experimentation; these are fast, cheap, and environmentally safe alternatives that can provide reliable predictions. These computational models can be developed with limited toxicity data in hand and using those developed models, one can get predictions of toxicity of new or untested nanoparticles.

Over the past few decades, *in silico* modeling has evolved a lot and grabbed sufficient attention of the scientific community. Quantitative structure-activity relationships (QSARs), Read-across predictions, and Machine learning (ML) models are becoming more popular now in predictive toxicology. These methods are similarity-based approaches where the prediction of endpoints can be obtained from structurally similar chemical analogs. The majority of work in this dissertation deals with the development of predictive and statistically sound nano-QSAR models; although different other algorithms like read-across predictions, and ML have also been used. A standard QSAR methodology comprises the steps i.e., data collection and curation, descriptor calculation and dataset division, feature selection and model development, validation of the developed model, and mechanistic interpretation of the response in light of the identified structural features. Various statistical tools are used for developing the model whereas, the developed model must be validated using strict internal and external validation protocols. Organization for Economic Cooperation and Development (OECD) prescribes the guidelines of QSAR modeling that have been followed in all presented works. During this study, many software tools have been employed that are freely available at http://teqip.jdvu.ac.in/QSAR_Tools/. The 3R strategy (Reduction, Refinement, and Replacement) of animal experimentation has been followed throughout this study; therefore, no animal has been harmed.

The works presented in this dissertation are as follows:

Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using QSAR modeling approach

Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides.

Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.

Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells.

Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach.

Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches.

Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.

Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across, and ML algorithms: QSAR models for *Daphnia magna* and zebrafish.

The entire work has been presented under the following chapters in this dissertation:

Chapter 1: Introduction

Chapter 2: Present Work

Chapter 3: Materials and Methods

Chapter 4: Results and Discussions

Chapter 5: Conclusion

References

Appendix: Reprints

The overall background knowledge of the nanoparticle toxicity assessment, the computational approaches, and the previous works done in this area of research have been discussed under the heading “*Introduction*”. Additionally, information on a brief history of QSAR, the basic principles of in-silico approaches that have been employed, and the validation protocol of in-silico models have also been provided. Chapter “*Present Work*” briefly discusses the performed research works along with the collected datasets. In the “*Materials and Methods*” section, the methodology of the entire work done has been thoroughly discussed. Schematic representations of all studies and the important numerical formula used for calculations have been provided in this section to avoid ambiguities. The results of all studies performed have been discussed in the “*Results and Discussions*” in detail. An overall “*Conclusion*” has been provided in this dissertation followed by the “*References*”. Most of the studies from this dissertation work have been published in *reputed international journals* and also presented at

different national and international conferences. Those published works are thus also included in the ultimate section of this dissertation under the section “*Reprints*”. However, the research work performed and presented in the limited scope of this dissertation constitutes only a small portion of the broad spectrum of work in this area of research. Various interesting aspects of research that have arisen in this dissertation, can be further investigated more appropriately in the future study.

Abbreviation

1D QSAR	One dimensional QSAR	OECD	Organization for Economic Cooperation and Development
2D QSAR	Two dimensional QSAR	TEM	Transmission electron microscopy
3D QSAR	Three dimensional QSAR	MD	Molecular dynamics
4D QSAR	Four dimensional QSAR	PLS	Partial least squares
5D QSAR	Five dimensional QSAR	LUMO	Lowest unoccupied molecular orbital
6D QSAR	Six dimensional QSAR		
7D QSAR	Seven dimensional QSAR	PT	Periodic table
AD	Applicability domain	PRI	Prediction reliability indicator
MeOx NPs	Metal oxides nanoparticles	QTTR	Quantitative toxicity-toxicity relationship
BSS	Best subset selection		
CCC	Concordance correlation coefficient	QSAR	Quantitative structure-activity relationship
NMs	Nanomaterials		
DNA	Deoxyribonucleic acid	RA	Read-across
EC	Effective concentration	AI	Artificial Intelligence
ECHA	European Chemical Agency	LDM	Liquid drop model
ED	Euclidean distance	REACH	Registration, evaluation, authorisation and restriction of chemicals
TUNEL	TdT-mediated dUTP-biotin nick end labeling		
ETA	Extended topochemical atom	RF	Random forest
EU	European Union	RMSEC	Root mean square error of calibration
GA	Genetic algorithm		
GB	Gradient boost	RR	Ridge regression
GK	Gaussian kernel	U.S. EPA	United States Environmental Protection Agency
IA	Independent action	SMILES	Simplified Molecular Input Line Entry Systems
EPR	Electron paramagnetic resonance	S-MLR	Stepwise multiple linear regression
IC	Inhibitory concentration	SVM	Support vector machine
S-MLR	Stepwise multiple linear regression	US EPA	United States Environment Protection Agency
SVM	Support vector machine		
LD	Lethal dose	WAP	Weighted average prediction
LDA	Linear discriminant analysis	WFD	Water Framework Directive
LK	Laplacian Kernel	WHO	World Health Organization
MAE	Mean absolute error	XGB	Extreme gradient boost
MCC	Matthews correlation coefficient	LDH	Lactate dehydrogenase
ML	Machine learning	QSPR	Quantitative structure-property relationship
kNN	k-nearest neighbors	QC	Quantum chemicals
TEMP	2,2,6,6-tetramethylpiperidine	ANN	Artificial neural networks
EDL	Electrical Double Layer		

Contents

Acknowledgements	<i>i</i>
Preface.....	<i>ii</i>
Abbreviations.....	<i>v</i>
 Chapter 1: Introduction.....	 1
1.1 Advancement in human uses of nanomaterials	1
1.2 Nanoparticle threat to the living organisms and the environment	2
1.3 The role of regulatory agencies in nanoparticles (NPs) risk assessment	4
1.4 Toxicity assessment of nanoparticles	6
1.4.1 Conventional techniques for nanoparticle toxicity assessment	6
1.4.1.1 In vitro assessment methods	6
1.4.1.1.1 Proliferation assay	6
1.4.1.1.2 Apoptosis assay	6
1.4.1.1.3 Necrosis assay	7
1.4.1.1.4 Oxidative stress assay	7
1.4.1.2 In vivo toxicity assessment methods	7
1.4.2 Computational techniques in nanoparticle toxicity assessment	8
1.5 In silico predictive modeling approaches	8
1.5.1 Quantitative structure-activity relationship (QSAR)	8
1.5.1.1 Background, its brief history, and evolution of QSAR	9
1.5.1.2 Objectives of QSAR	11
1.5.1.3 Classification of QSAR	12
1.5.1.3.1 Based on dimensionality	12
1.5.1.3.2 Based on chemometric methods	12
1.5.1.4 OECD Guidelines and QSAR modelling	13
1.5.1.5 QSAR methodology	14

1.5.1.5.1 Data preparation and descriptor calculation	14
1.5.1.5.2 Currently available nano-toxicity descriptors	15
1.5.1.5.2.1 Software supporting the calculation of PT descriptors	18
1.5.1.5.3 Dataset division	20
1.5.1.5.4 Feature selection	21
1.5.1.5.5 Model development	21
1.5.1.5.6 Validation protocols	22
1.5.1.5.7 Model interpretation	22
1.5.2. Read -across (RA) predictions	27
1.5.3. Machine learning (ML)	27
1.5.3.1 RF (random forest)	27
1.5.3.2 AdaBoost (Adaptive Boosting)	27
1.5.3.3 Gradient Boost (GB)	28
1.5.3.4 Extreme Gradient Boost (XGBoost)	28
1.6 Application of in silico modeling approaches for nanoparticle toxicity assessment	29
Chapter 2: Present Work.....	31
2.1 Objectives of the dissertations	33
2.2 Datasets employed for the development of nano-QSAR / ML models.	33
2.1.1 Dataset I (Study 1)	34
2.1.2 Dataset II (Study 2)	35
2.1.3 Dataset III (Study 3)	35
2.1.4 Dataset IV (Study 4)	36
2.1.5 Dataset V (Study 5)	36
2.1.6 Dataset VI (Study 6)	37
2.1.7 Dataset VII (Study 7)	37
2.1.8 Dataset VIII (Study 8)	38

Chapter 3: Materials and Methods.....	39
3.1. Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach.	39
3.1.1 Toxicity data	39
3.1.2 Descriptor calculation	40
3.1.3 Data set division	40
3.1.4 Model development	40
3.1.5 Model validation	43
3.2. Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to Escherichia coli: categorization and data gap filling for untested metal oxides.	44
3.2.1 The dataset	44
3.2.2 Molecular descriptors	45
3.2.3 Model development: Application of the Small Dataset Modeler	45
3.2.4 Reliability of predictions of untested NPs	47
3.2.5 Statistical validation metrics	50
3.3. Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.	50
3.3.1 Dataset	50
3.3.2 Molecular Descriptors	53
3.3.3 Data set splitting	54
3.3.4 Chemometric tools	54
3.3.5 Validation metrics	55
3.3.6 Applicability domain (AD)	55
3.4. Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7cells.	55
3.4.1 The dataset	55
3.4.2 Molecular Descriptors	57
3.4.3 Small Dataset Modeler for feature selection	57

3.4.4 Read-Across based predictions	58
3.4.5 Applicability domain	58
3.4.6 Nano-QTTR model development	59
3.4.7 Statistical validation metrics	60
3.5. Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach.	61
3.5.1 Dataset	61
3.5.2 Periodic table descriptor calculations	64
3.5.3 Model development	64
3.5.4 Validation metrics	64
3.6. Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO ₂ with periodic table descriptors using machine learning approaches.	66
3.6.1 Dataset	66
3.6.2 Descriptor calculation	69
3.6.3 Splitting of data set and hyper-parameter tuning	69
3.6.4 Feature Selection with RF	69
3.6.5 Model development	70
3.6.6 SHAP analysis	70
3.6.7 Model validation	71
3.7. Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.	72
3.7.1 Dataset	72
3.7.2 Descriptor calculation	75
3.7.3 Splitting of the data sets	75
3.7.4 Model development	76
3.7.5 Model validation	76
3.8. Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across and ML algorithms: QSAR models for <i>Daphnia magna</i> and zebrafish.	80

3.8.1 Theoretical descriptors	81
3.8.2 Dataset division	81
3.8.3 Important feature selection	82
3.8.4 Regression based QSAR model development	82
3.8.5 Read-across (RA)	82
3.8.6 Development of various ML (machine learning) models	83
Chapter 4: Results and Discussions.....	84
4.1. Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach.	84
4.1.1 Loading plot of the PLS model	86
4.1.2 Model randomization	86
4.1.3 Mechanism of toxicity	90
4.1.4 Comparison with previously published models	91
4.2. Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to Escherichia coli: categorization and data gap filling for untested metal oxides	92
4.2.1 Interpretation of the obtained descriptors	93
4.2.2 Proposed mechanism of toxicity to E. coli cell.	95
4.2.3 Comparison with the previous work	96
4.3. Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.	97
4.3.1 Interpretations of different plots	99
4.3.2 Descriptor description	103
4.3.3 Mechanism of the toxicity	109
4.3.4 Comparison with the previous work	111
4.4. Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells.	112
4.4.1 Prediction by the Small Dataset Modeler	115
4.4.2 Statistical aspects of the nano-QTTR models	117

4.4.3 Chemical Read-Across Analysis	119
4.4.4 Interpretation of the descriptors for the BEAS-2B cell toxicity model	121
4.4.5 Mechanism of toxicity	124
4.5. Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach	125
4.5.1 MLR models for inflammatory potential	127
4.5.2 Evaluation of the predictive ability of the developed MLR model for fold change	130
4.5.3 Features responsible for the inflammatory effect to THP-1Cell	130
4.5.4 Comparison with the previous study	132
4.6. Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO ₂ with periodic table descriptors using machine learning approaches.	132
4.6.1 Diagnosis based on SHAP value	133
4.6.2 Results of Model Validation for all ML Methods	135
4.6.3 General Mechanism of Toxicity	136
4.6.4 Comparison with the previous work	137
4.7. Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.	138
4.7.1 QSPR model for Zeta potential	140
4.7.2 QSPR model for cell damage	142
4.7.3 Chemical Read-Across Analysis	142
4.7.4 Interpretation of the descriptors	143
4.7.5 Importance of Zeta potential as a descriptor	146
4.7.6. Utilization of the metal oxide cell damage knowledge for cancer treatment	146
4.7.7 Comparison with the previously published literature	147
4.8. Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across, and ML algorithms: QSAR models for <i>Daphnia magna</i> and Zebra fish.	148
4.8.1 QSAR model with <i>Daphnia magna</i> (Model 8.1)	149
4.8.2 QSAR model with Zebra fish (Model 8.2)	151

4.8.3 Discussion similarity and read across (RA) predictions	152
4.8.4 Discussion on ML (Machine Learning) models	153
4.8.5 The NP effects on morphological parameters of D. magna and hatching of zebra fish	154
4.8.6 Comparison of ML models with QSAR models and the previous work	156
4.8.7 Future research and considerations	157
Chapter 5: Conclusion.....	158
References.....	164
Appendix: Reprints	

CHAPTER - 1

INTRODUCTION

Chapter 1: Introduction

1.1 Advancement in human uses of nanomaterials

The use of nanomaterials has drastically changed the field of science and technology worldwide. These particles, which range from 1 to 100 nm in size, have unique physical, chemical, and biological properties that have a wide range of applications in various fields (Bayda et al., 2019). The concept of nanoparticles and nanotechnology has been around since the 20th century. During the discovery of quantum theory it was estimated that managing the nano-sized physiochemical properties would be a challenging task (Roy and Roy, 2023). Nanomaterial products have been used since ancient times when metallic nanoparticles were used for glass-making in Egypt and Mesopotamia in the fourteenth and thirteenth centuries BCE (Schaming and Remita, 2015). It was discovered that some glasses have a red color because they contain copper nanoparticles or cuprous oxide nanoparticles. Only recently has a proper understanding and use of nanomaterials come to light. The concept of nanomaterials was first applied by Richard Feynman (Feynman et al., 2022) in the late 1950s. It was after the discovery of fullerenes in 1985, that the application of nanomaterials gained momentum. Nano-sized materials have been in commercial use for a very long time. Nano compounds with their remarkable properties are ranked highly in the research domain. Nanoparticles (NPs) are a wide class of materials that include substances ranging from less than 100 nm at least (Laurent, 2010). Based on their overall shape, these materials can be classified as 0D, 1D, 2D, or 3D (Tiwari, 2012) (**Figure 1.1**). There are two major classes of nanoparticles: a) carbon nanoparticles and b) metal oxide nanoparticles (Gajewicz et al., 2012; Artiles et al., 2011; Puzyn et al., 2010). The term “ENMs” includes engineered nanoparticles, nanofibers, nanoplates, quantum dots, and other nanostructured surfaces (Domingues et al., 2022). Metal oxide nanoparticles (NPs) have unique properties due to their size and high density of edge or corner sites. Nanoparticles are being increasingly used in different areas like pharmacy, space technology, environmental engineering, stain-resistant clothing and environmental monitoring. Currently, ENMs with their innumerable properties are also used in water treatment, food, agriculture, textiles, medicines, pharmaceuticals, cosmetics, and electronics as well as in other fields. Nanomaterials are used in many everyday products. A report by Markets and Markets predicts that the global nanomaterials market will grow to \$75.64 billion by 2025 and with a compound annual growth rate of 13.2% from 2020. The nano industry includes 10,150 nano products marketed by 3246 companies worldwide. The engineered nanoparticles (ENPs) is estimated to rise at a rate of 25% per year (Abbas et al., 2020).

Engineered nanoparticles (NPs) are currently utilized in over 5,000 commercial products, ranging from medical equipment and fuel additives to plastics (<https://nanodb.dk/>). Nanomedicines (Roy et al., 2022) are one of the applications of nanotechnology that have sparked hope in the field of medical sciences. The incorporation of nanostructures in this field has not only created advanced opportunities but also generated perspectives on issues such as targeting tissues or cells, controlled release of drugs on a specific target, and early detection of diseases. In recent years, more attention has been given to nanostructures as adsorbents and catalysts to remove harmful and toxic contaminants from the environment (Yang et al., 2006; Liu et al., 2004). In theory, nanoparticles can be engineered from substances, such as semiconductor nanocrystals, organic dendrimers, and carbon fullerenes, and possess highly desirable properties like electrical, thermal, and mechanical that are useful in commercial, medical, and environmental sectors.

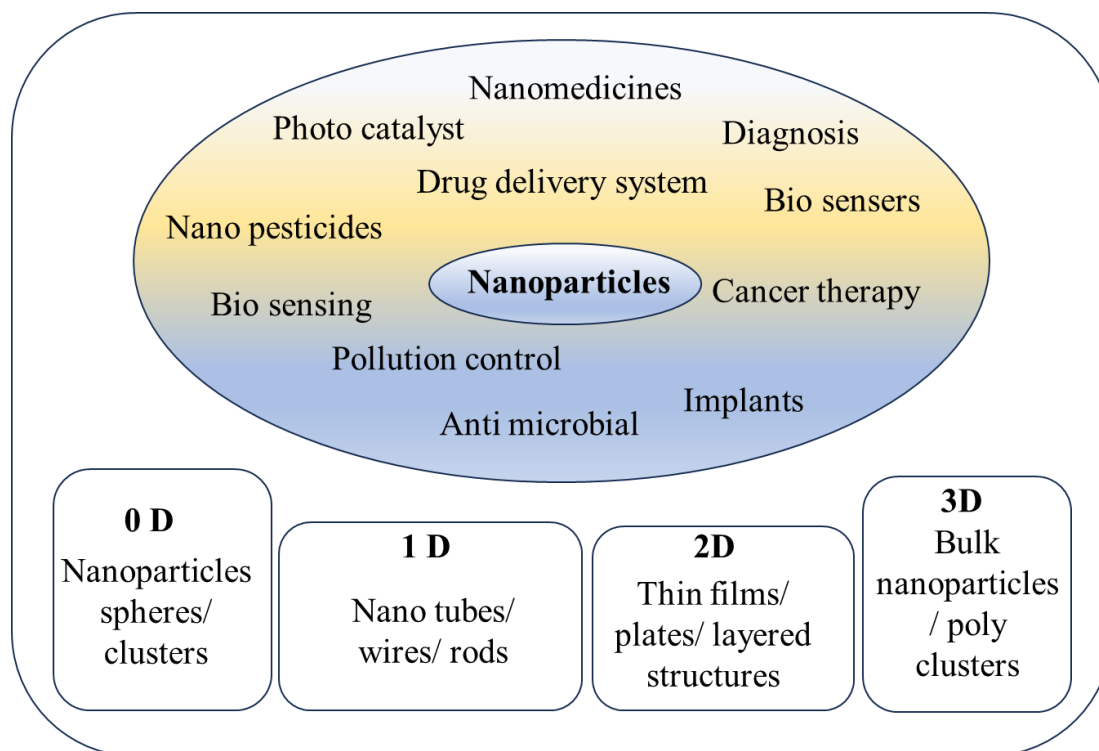


Figure 1.1 Types and applications of nanoparticles

1.2 Nanoparticles threat to the living organisms and the environment

Engineered nanoparticles (ENPs) can be released into the environment at various stages such as during production, manufacture, use, and disposal. These are considered novel entities that fall under planetary boundaries, which aim to identify a "safe operating space" for human societies to thrive (Steffen et al., 2015). MeOx NPs have gained popularity in recent years for their potential to enhance the quality of human life. However, there are growing concerns regarding the risk associated with these nano-sized particles to the environment and human beings. People are unknowingly exposed to different NPs through products used in daily life, such as skin care products (skin absorption, or medical use), through disposal, during production, inhalation or oral uptake. Upon uptake they may cross the various biological barriers present in the species and lead to accumulation in different tissues and organs (Yan et al., 2023) (see **Figure 1.2**). Nowadays, many diseases are caused by exposure to harmful chemicals or materials whose behavior is not fully understood, and these harmful substances may be hidden behind the perceived benefits of the products they are used in. Engineered nanoparticle (ENP) toxicity is a global environmental threat, and the analysis of nanoparticles in the terrestrial and aquatic environment and the evaluation of risks have gained significant attention due to their increasing use worldwide. These particles are likely to pose a risk to human and cause different abnormalities such as bronchitis, Parkinson's, Alzheimer's, asthma, and many more (Ahamed et al., 2014). Recently, a report found submicron-sized plastic in a healthy donor (Roy and Roy, 2023). This finding was not surprising to researchers, as an experimental study in 2002 showed the translocation of particulates in the bloodstream in that size range. Since then, many more studies have come up with the translocation of nano-sized particles in blood circulation that may reach different tissues and even cross the blood-brain barrier to reach the brain during the fetal stage (Nemmar et al., 2002). Nanotechnology and nanomedicines have brought many

benefits to the world, but it is crucial to address their potential negative effects. Concerns about the health and safety of NPs have led to the development of a advanced research field called nanotoxicology. Nanotoxicology research has recently concentrated on reducing the negative effects of NPs with the assistance of various agencies in order to have a safe and viable NPs application. The world is witnessing a significant transformation due to the impact of nanotechnology and nanoparticles. These tiny substances have become central players in our economies. However, it has been known since ancient times that exposure to such substances can be harmful to human health, and miners (an example of occupational exposure to materials) which leads to health problems. Today, there is a growing concern about the potential exposure and adverse outcomes for the final users of products containing nanotechnology and nanoparticles. For instance, MeOx NPs used in various consumer products are suspected to end up in aquatic or terrestrial environments, where their toxicity mechanism is yet to be fully explored. Exposure to these substances has been shown to cause growth inhibition of algae (Ribeiro et al., 2014), immobilization and genotoxic effects on *Daphnia magna* (*D. magna*) (Park et al., 2010), and cytotoxicity in fish cells. It is, therefore, complex to assess the potential risk (Raftis and Miller, 2019) and safety of MeOx NPs through different routes and sources. A complete toxicity profile, in the form of cell viability and cytotoxicity studies, should be conducted before the application of nano-sized materials to ensure safety.

The increased global demand for nanoparticles comes with a significant risk of causing toxicity to human health and the environment. It is crucial to develop fast and reliable methods for predicting the properties, toxic behavior, and environmental impact of these nanoparticles to avoid any potential harm. Efficient nano-toxicological testing is necessary to ensure a safer and healthier future. However, with the rapidly expanding market for nanoparticles, it is challenging to regulate their safety through traditional in vivo and in vitro tests. Additionally, assessing nanoparticles using traditional methods is becoming increasingly difficult due to the limited data. To address these issues, various strategies have been implemented to evaluate the toxicity of nanoparticles. Thus, assessing the potential risks, that workers, final users, and the environment face, is of utmost importance (Lebre et al., 2022).

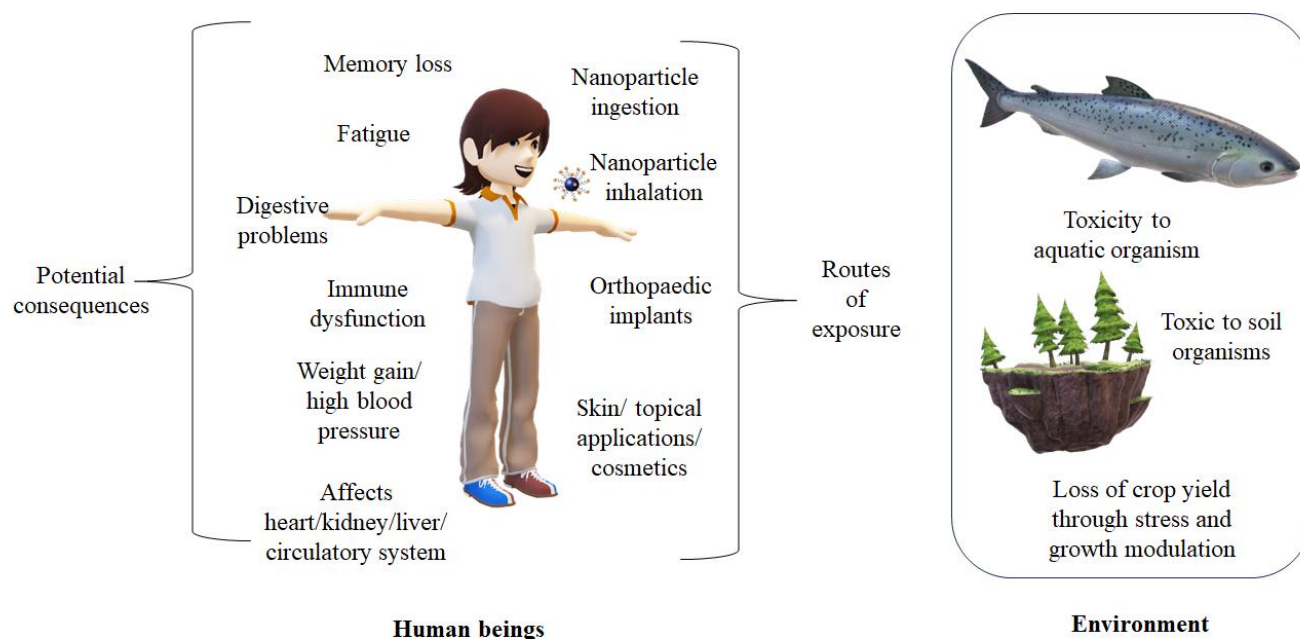


Figure 1.2 Exposure to materials leading to health problems for both humans and the environment

1.3 The role of regulatory agencies in nanoparticles (NPs) risk assessment

The creation and use of novel nanoparticles (NP) may lead to a wider gap in our understanding of their exposure and potential hazards to the ecosystem. Laboratories and industries must have information on human exposure to nanoparticles to define reliable safety regulations for people working with them. Although attempts were made to develop corresponding instructions, they are far from the ideal. The identification, risk analysis, risk impact, and economic aspect of the relevant legislation documents are the four components that make up a real risk assessment (<https://cloudogre.com/risk-assessment/>). For instance, the European Food Safety Authority has carried out an environmental risk assessment (<https://www.efsa.europa.eu/en/topics/topic/environmental-risk-assessment>) Regulation No. 1107/2009 has been used for this purpose. Different approaches have been used to account for engineered nanoparticle hetero-aggregation in large-scale models for the fate and distribution of nanoparticles in ecological systems (Dale et al., 2015). Chemical safety assessment in Europe is regulated by a set of legal acts (Mech et al., 2019), among which the most valuable is REACH regulation (EC 1906/2006) that mention the term “chemical substances”. Though the term “substance” is not suitable for ENPs. Another, important information with ENPs is its link with the individual components. In this context, “composition” (as also used in REACH) cannot explain ENPs components and it should not be considered under the term “chemical composition”, which is generally used with elemental analysis (Puzyn et al., 2010). Thus, to address such misexplained definition, amendment was performed by REACH regulation (EC No 1907/2006)22, which was enforced on 1 January 2020. These amendment introduced the term “nanoforms” for ENPs that can distinguish nano structures of particle substances that can also be present as larger particles or structure i.e., ion. A nanoform is a natural or artificial material that contains particles in an unbound state, as an aggregate or an agglomerate, and where one or more of the particles' external dimensions fall between 1 and 100 nm for 50% or more of the particles in the number size distribution. This includes, fullerenes, graphene flakes, and single-wall carbon nanotubes with one or more external dimensions less than 1 nm. (ECHA, 2019).

According to the Toxic Substances and Cancer Act (TSCA), to be classified as a specific kind of "chemical substance," ENPs need additional information about their chemical identity, processing techniques, use cases, exposure conditions, and other items. In addition to being beneficial to currently expanding regulatory frameworks, the data model for ENP representation should be manageable. The descriptors generated based on the data model might also be appropriate for TSCA and REACH compliance. With the introduction of REACH, *in silico* models, such as the Quantitative Structure-Activity Relationships (Q)SAR approach, were used for chemical safety assessments in order to reduce the need for animal testing and to close knowledge gaps for a wide range of chemicals (EC 2008). When the *in silico* method and information related to the regulatory aspect are carefully examined, it can be seen that approximately 30% of REACH case studies used read-across and/or (Q)SAR models in lieu of experimental evidence (ECHA 2014). In its “Guidance Document on the Validation” of (Q)SAR models, the OECD offered a toolbox with five principles required to validate (Q)SARs for chemicals should fulfill (OECD, 2007). A (Q)SAR model must have a defined endpoint, or response value, a clear algorithmic form, a defined AD (domain of applicability), suitable goodness-of-fit, robustness, and predictive ability measurements, as well as a mechanistic interpretation, in order to be used for regulatory purposes. Compared with traditional chemicals, hazard data, and effects sets for nanoparticles that support computational approaches are still limited. Because there are insufficient data to explain the mode-of-action

(MOA) or adverse outcome pathway (AOP) mechanisms for nanoparticles, it is difficult to quantitatively correlate physicochemical descriptors to specific biological effects and fate characteristics of nanoparticles (Klaire et al., 2008). Potential pathways through which nanoparticles could have detrimental effects on health or the environment include oxidative stress and the production of ROS (reactive oxygen species), inflammation, protein denaturation, damage to membranes and DNA, disruption of energy transduction, loss of immunity, upregulation of granuloma formation, damage to mitochondria, and release of toxic constituents (**Figure 1.3**), which ultimately causes cell death.

Based on their reactivity and solubility, several chemistry-based predictive models have been developed for nanoparticles (NPs). These models show promise in predicting toxicity. However, in the field of NP toxicology, more robust advancements are required to overcome the gap of knowledge and data. In this regard, computational methods can assist in focusing on in vivo and in vitro fate testing and assessment schemes. Computational methods can also help in analyzing and predicting the outcome of functional test and assays that are designed to indicate specific MOA or extrinsic properties of NPs.

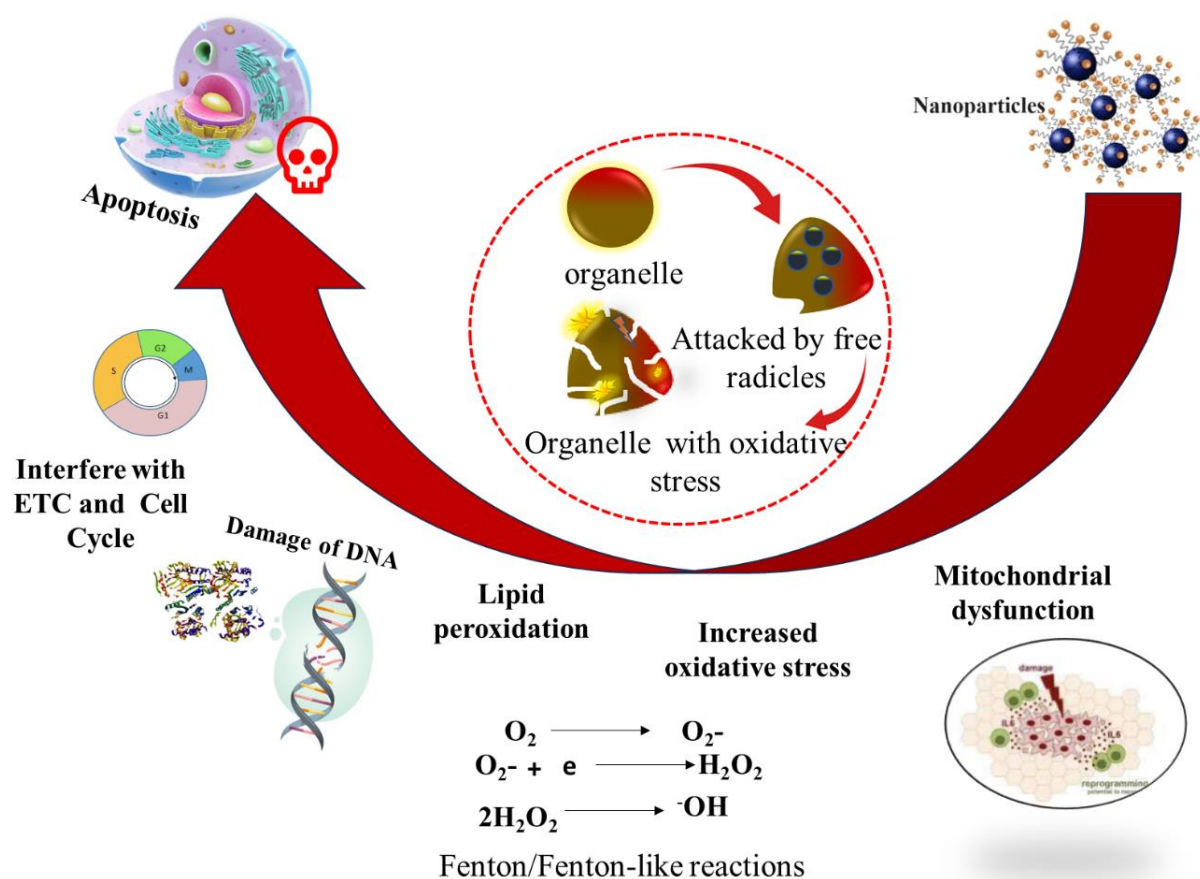


Figure 1.3 Mechanism of toxicity of nanoparticles

1.4 Toxicity assessment of nanoparticles

Nanotoxicology is the study of the harmful effects of nanoparticles on living organisms. It involves assessing the toxicity of these particles *in vivo* and *in vitro*, which is similar to the methods used for other chemicals and drugs. Different models, including environmental (Suhendraet et al., 2020), animal (Valic et al., 2020), *in vitro* (Kathawala et al., 2013), and *in silico* (Oberlintner et al., 2021), are used to evaluate the potential effects of nanoparticles. However, adapting conventional methods to assess nanotoxicity has led to accuracy and confidence issues, as engineered nanoparticles can interfere with traditional methods. Therefore, it is essential to have a good knowledge of their characteristics (Oberdörster et al., 2010), such as composition, size, shape, dispersity, surface charge, surface functionality, and protein corona formation. These characteristics affect a particle's ability to cross biological barriers and bring additional challenges, such as agglomeration or aggregation (Gatoo et al., 2014). *In vitro*, nanoparticle toxicity assessment is an important method that is cost-effective, faster and has minimal ethical concerns. It involves assessing different types of assays, including proliferation, apoptosis, necrosis, oxidative stress, and DNA damage assays.

1.4.1. Conventional techniques for nanoparticle toxicity assessment

1.4.1.1 *In vitro* assessment methods

1.4.1.1.1 Proliferation assay

This assay used cellular metabolism through the assessment of metabolically active cells. The tetrazolium salt, 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT), was used for *in vitro* toxicity assessment of nanoparticles (Sayes et al., 2007). This technique yields quick and reproducible results requiring minimum manipulation of the model cells (Marshall et al., 1995). The MTT assay measures the tetrazolium salt, and can lead to alteration as per the change of culture media additives (Molinari et al., 2003), media pH (Jabbar et al., 1989), ascorbate (Natarajan, 2000), and cholesterol (Abe et al., 1999). Thus, XTT or WST-1 assays that produce soluble dyes are selected over the MTT assay. Further, the incorporation of [³H] thymidine is mostly used in cases of cell proliferation though it was avoided due to high-cost (Hussain et al., 2002). The clonogenic assay was used for counting the proliferating cells through visual inspection on nanoparticle exposure.

1.4.1.1.2 Apoptosis assay

Apoptosis is a significant indicator of nanoparticle toxicity *in vitro*. Apoptosis and DNA damage (Ryter et al., 2007) are thought to be the result of excessive free radical generation. Several studies reported the induction of apoptosis through nanoparticles. In study performed by Ahamed et al., 2008 using silver nanoparticles showed apoptosis in the mouse embryonic stem cells. In another study, the treatment of larval tissues of *Drosophila melanogaster* was investigated with silver nanoparticles with concentration (50 and 100 µg/ml) for 24 hrs to 48 hrs. which resulted in the caspase-3 and caspase-9 (apoptosis marker), which indicate the involvement apoptotic pathway due to Ag⁺ nanoparticle (Ahamed et al., 2010) in *D. melanogaster*. The up-regulation of p38 protein expression and genes in *D. melanogaster* species demonstrated that silver nanoparticles are dose and time-dependent. The cascade of apoptosis pathway and cell death is a result of extensive DNA damage. The apoptosis can be

detected using several assays including Annexin-V assay (Lee et al., 2007), Comet assay (Jin et al., 2007), TdT-mediated dUTP-biotin nick end labeling (TUNEL) assay (Mo et al., 2005), and morphological changes. The decrease of size of cells and DNA fragmentation are included as indications of apoptosis. The use of gel electrophoresis can easily differentiate between apoptotic and necrotic modes of cell damage. Toxicological assessment frequently uses markers like propidium iodide (PI) and Annexin-V. The intensity of fluorescence is the measurement used in this type of assay. The fluorescence indicates the leakage of the plasma membrane induced by a caspase-dependent pathway that is caused by the binding of Annexin-V binds to phosphatidylserine. The damage of the DNA is detected using Comet assay, here the fragmented DNA migrates out of the cell upon the applied electric current. On the other hand, undamaged DNA remains within the cell. The tail in the in the fluorescence detector represents the damaged DNA whereas the head is the intact DNA. The extended tail is correlated with the size and shape of the DNA damage and its distribution within the comet.

1.4.1.1.3 Necrosis assay

Necrosis is a method of measuring the viability of cells by assessing the integrity of the membrane. It is done by observing the amount of dye (Neutral Red (2-amino-3 methyl-7-dimethylaminophenazoniumchloride) (Monteiro-Riviere, 2005) and Trypan Blue (Huang et al., 2004)) intake. Neutral Red is a weakly cationic supravital dye that produces a deep red color at slightly acidic pH levels that can easily diffuse through the plasma membrane. This dye can accumulate in the lysosomes where it binds with the matrix through electrostatic or hydrophobic bonds. The changes (uptake or binding) in the Neutral Red indicate the exposure to xenobiotics or nanoparticles (Nemes et al., 1979; Lüllmann-Rauch, 1979) that decreases cell viability. By assessing the change via binding with and uptaking the Neutral Red dye, distinguishing between viable and dead cells is possible.

1.4.1.1.4 Oxidative stress assay

Exposure of nanoparticles to cells can cause ROS and RNS production. To detect this, the conventional method uses the reaction of 2,2,6,6-tetramethylpiperidine (TEMP) with O_2^- (stable radical), which can be identified using electron paramagnetic resonance (EPR) via X-band. However, this method is often hindered by its high cost. An alternative approach that is also cost-effective has come up in the form of fluorescent probe molecules (Gomes et al., 2005). However, these probes can react with a variety of reactive species, leading to misleading results sometimes (Kumar et al., 2017).

1.4.1.2 In vivo toxicity assessment methods

Animal models (mice and rats) are typically used to perform in vivo toxicity assessments. The methods for assessing in vivo toxicity include examining biodistribution, clearance, haematology, and histopathology. Various biodistribution studies are conducted to determine the route of localization of nanoparticles to various tissues or organs. Radiolabels are used to detect nanoparticles in live or killed animals (Kim et al., 2001). The clearance of nanoparticles is assessed by examining their excretion and metabolism at different time points after exposure. Investigating the similarity of nanoparticles with the cell activity of alive systems is an attractive but difficult task. As it is not practical to experimentally test each nanomaterial's size, type, and shape, experimentalists and *in silico* scientists must collaborate in search of common terminology to pave the way for more predictive nanotoxicology.

1.4.2. Computational techniques in nanoparticle toxicity assessment

Computational approaches are being adopted to bridge the gap between experimental and nanotoxicology studies. With the huge amount of experimental data generated by modern nanotechnology research, *in silico* approaches are becoming a preferred alternative to *in vivo* and *in vitro* studies. This is because examining thousands of MeOx NP-based commercial products through classical experiments would be expensive and risk animal sacrifice (Fourches et al., 2010). Furthermore, the creation of strong computational experiments is more affordable and often approved by international regulatory bodies, such as the US Environmental Protection Agency (EPA), for the purpose of toxicity data gap mapping of MeOx nanoparticles (NPs) lacking experimental data. However, the rapid increase in the quantity and heterogeneity of literature data on nanoparticles is making it more difficult to efficiently screen data based on their quality and completeness. As a result, creating trustworthy models that account for the experimental setup is difficult. Another major issue is choosing the list of experimental conditions that should be available in order to construct a model. Researchers frequently use *in silico* approaches such as machine learning (ML) (Gajewicz et al., 2018; Fjodorova et al., 2017), quantitative structure-activity relationship (QSAR) (Kar et al., 2014; Ojha et al., 2019; Puzyn et al., 2011; Winkler et al., 2013), and read-across (Gajewicz et al., 2017; Lamon et al., 2019) are among the tools that researchers commonly use to model the various toxicity endpoint related to nanoparticles. Without a thorough understanding of the interaction mechanisms, QSAR is anticipated to capture the complex relationships between the macroscopic characteristics or activities of nanoparticles and their microscopic molecular structures. However, since nanoparticles have special properties, properly characterizing materials when studying their properties is difficult. Therefore, it is essential to choose a proper approach to build a QSAR model related to nanoparticles because they often provide unique useful properties.

1.5. *In silico* predictive modeling approaches

Since the introduction of computational technology in nanotoxicity assessment, various *in silico* algorithms have been used to generate predictive models. These algorithms correlate observed responses with structural and biological aspects of nanoparticles. Among these approaches, classical Quantitative Structure-Activity Relationship (QSAR), read-across (RA) predictions, and machine learning (ML) are the most common. The development of *in silico* models involves data gathering, descriptor generation, modeling, model evaluation, and model interpretation. These are general steps followed by all algorithms.

1.5.1. Quantitative structure-activity relationship (QSAR)

Quantitative Structure-Activity Relationship (QSAR) is a statistical method used to assess nanoparticle toxicity. It involves creating a mathematical model that correlates molecular properties (such as biological, physicochemical, or toxicological properties) with specific aspects of nanoparticle chemistry. This is done using various quantitative parameters known as descriptors (Roy et al., 2015), which can be obtained through computational or experimental means. To create a QSAR model, a series of compounds with a known response are used. If experimental data is not available, QSAR models can predict a compound's characteristics by establishing a relationship between its structural properties and activity. In other words,

molecular responses are determined by a set of structural features, which are represented by molecular descriptors. This relationship is typically expressed using equation 1.1.

$$\text{Molecular response} = (\text{activity/toxicity/property}) = f(\text{quantitative structural information or properties}) = f(\text{Descriptors}) \quad (1.1)$$

1.5.1.1. Background, its brief history, and evolution of QSAR

It can be challenging to figure out the historical beginning of Quantitative Structure-Activity Relationship (QSAR) since the fundamental idea has gone through significant changes from 1860 to the time of Hansch (Purcell et al., 1973; Rekker, 1992; Tute, 1990; van De Waterbeemd, 1992). In the past, researchers lack code to represent the chemical structure using quantitative terms. However, the mathematical expression of the biological property was suggested to be of great importance, even though they were unable to confirm this relationship using other examples. During the 19th and 20th centuries, many theories related to the effects of narcotic chemicals and narcosis emerged. Later, Meyer (Meyer, 1899) and Overton (Overton, 1899) suggested that the depressant action of a group of organic compounds is correlated with their oil/water partition coefficients (Lifnick, 1989; Lipnick, 1991).

L.P. Hammett in 1935 made a revolutionary contribution to the theoretical organic chemistry domain by establishing a relationship between the chemical reactivity and structural features of benzene derivatives using rate constant and electronic constant terminologies. The rate constants in this case are k_X and k_H , and the equilibrium constants of substituted and unsubstituted benzenes are K_X and K_H . The constant ρ is a constant, and σ_X is the Hammett constant (Hammett, 1970), which provides electronic information as shown in the following equations 1.2 and 1.3.

$$\log K_{R-X} - \log K_{R-H} = \rho\sigma \quad (1.2)$$

$$\log k_{R-X} - \log k_{R-H} = \rho\sigma \quad (1.3)$$

Hammett ‘ σ ’ was defined by ionization constant terms, and ρ represents the reaction constant which is specific for the reaction type. The “Hammett electronic constant” represented as “ σ ” was attempted by many reserchers in consecutive years to correlate the biological activity (Hansen, 1962) with the Hammett electronic constant (σ) as demonstrating a separate equation (equation 1.4) (Hansen, 1962; Zahradník, 1962; Zahradník, 1962). By employing the LFER technique, Taft introduced steric feature E_s and allowed a separate assessment of polar, steric, and resonance effects by performing acid- and base-catalyzed hydrolysis of aliphatic esters. Corwin Hansch is considered the ‘Father of QSAR’ (Martin & Stouch, 2011) for his notable contribution to the QSAR paradigm.

$$\log \tau_i - \log \tau_{Et} = \alpha\beta_i \quad (1.4)$$

Here, τ_i and τ_{Et} resemble the biological property, “ α ” is a constant related to the biological property, whereas “ β_i ” resembles the contribution of the substituent group X_i . Equation 1.4 is the generalized representation of the unspecified biological activity terms.

The modern QSAR is related to linear free energy relationship or LFER in general, and the Hammett equation in specific. Hansch and Fujita (1964) published a research paper named “ ρ - σ - π analysis. A method for correlating biological activity and chemical structure”, which was an initial step of the modern QSAR that is still used (Hansch and Fujita, 1964). Hansch paid attention to the effect of hydrophobicity and to its related substituent species towards the response (y). Hansch developed an equation (similar to a Hammett equation) to show the hydrophobicity of aromatic hydrocarbons by representing hydrophobicity as an additive property of a molecule as shown in equation 1.5:

$$\log (P_{R-X} / P_{R-H}) = \pi_X \quad (1.5)$$

where “ π_X ” denotes the partition coefficient of the substituent “X” of a substituted aromatic hydrocarbon. This was followed by combining Hammett’s electronic constant (σ) and the hydrophobicity measure (π) into a single equation by Fujita and Hansch. Fujita identified that the free energy variables, e.g., π , $\log P$, σ , etc. can be combined into a single equation to represent biological activity by implementing a logarithmic transformation of the concentration term to keep the LFER formalism. The proposed multiparametric model has been given in the following equation 1.16 –

$$\log 1/C = a \pi + b \sigma + \dots + \text{constant} \quad (1.6)$$

The famous Free-Wilson approach (Free & Wilson, 1964) mathematically correlates the biological activity of a congeneric series of chemicals with the common contribution of the parent moiety plus the contribution of each structural substituent. Equation 1.7 presents the Free-Wilson model where BA is the biological activity assumed to comprise the average parent moiety contribution μ and the contribution of each substituent a_i . Here, x_i is a Boolean variable denoting the presence (=1) or absence (=0) of a specific structural fragment.

$$BA = \sum a_i x_i + \mu \quad (1.7)$$

Fujita and Ban (1971) considered a log-transformed response and provided a modified version of this equation to overcome its shortcomings (Equation. 1.8).

$$\log BA = \sum G_i x_i + \mu \quad (1.8)$$

Here, “ μ ” denotes the contribution of the “parent moiety”, “ G_i ” represents the contribution of “chemical fragments” and “ x_i ” represents the “presence/ absence of G_i ”.

In 1973, C. Hansch and S. Unger defined good practice in QSAR, comprising five criteria that must be considered before selecting a “best equation”. The five criteria were as follows, i) selection of independent variables considering the widest possible number of variables; ii) justification of the choice of independent variables by statistical procedures; iii) the principle of parsimony, i.e., all things being equal, one accepts the simplest model; iv) at least five to six data points per variable should be present to avoid chance correlations; v) try to find a qualitative model of physicochemical or biochemical significance, etc.

1.5.1.2. Objectives of QSAR

The principle objectives of QSAR are:

1. Prediction of new compounds with better activity.
2. Better understanding of the modes of action.
3. Lead compound optimization with decreased toxicity and better properties (Tong et al., 2005).
4. Reduction of the expensive experimental cost, time, and manpower by introducing effective compounds through reduced exhaustive methods.

To achieve the objectives mentioned earlier, a detailed knowledge of the following aspects is necessary:

1. The mode of action of the molecules must be understood.
2. Various factors that control the experimental conditions of the molecules must be considered.
3. A thorough examination of molecular structures and their properties is required.

Quantitative structure-activity relationship (QSAR) is an interdisciplinary study of chemistry, biology, and statistics. By predicting the essential structural requirements necessary to obtain a molecule with optimized activity, toxicity, or property, QSAR analysis provides a good platform for synthesizing a relatively smaller number of chemicals with improved activity/toxicity/property of interest.

The key objective of QSAR analysis is the development of a mathematical relationship between the response and chemical features of a series of compounds. Hence, mathematically for a series of compounds, this is about developing a correlation between a Y and several X variables where Y is the dependent variable remaining on the left side of the equation and the X variables are independent entities and a QSAR relationship can be simply stated as follows in Equation 1.9

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n \quad (1.9)$$

The response or the Y variable is also termed an 'endpoint' while the X features are predictor variables, and a QSAR relationship can be simply stated as follows in Equation 1.9. 'Y' is the response variable (independent) being modeled for a set of n number of compounds with the predictor variables $x_1, x_2, \dots, x_n$, etc. possessing coefficient values $a_1, a_2, \dots, a_n$, respectively, and a_0 is the constant term. With the values of Y and x variables known, the coefficients ($a_1, a_2, \dots, a_n$) along with the constant (a_0) can be easily solved by giving an explicit mathematical equation correlating biological activity or toxicity or physicochemical properties of compounds with their chemical features.

1.5.1.3. Classification of QSAR

1.5.1.3.1 Based on dimensionality

QSAR models can be classified into different classes based on the descriptor complexity. They are (i) 0D-QSAR, (ii) 1D-QSAR, (iii) 2D-QSAR, (iv) 3D-QSAR, (v) 4D-QSAR, (vi) 5D QSAR, and (vii) 6D-QSAR and (viii) 7D-QSAR. **Table 1.1** gives examples of descriptors based on their dimensions.

Table 1.1. Classification of descriptors based on their dimensionality.

Dimensionality	Interpretation	References
1D	Sub-structural fragments	(Todeschini and Consonni, 2009)
2D	Structural patterns, i.e., the topology of the molecules (without 3D representation)	(Todeschini and Consonni, 2009)
3D	These representations are produced by first providing the molecules' static three-dimensional structures as input. Numerous geometrical and spatial characteristics are considered.	(Guner, 2005; Todeschini and Consonni, 2009)
4D	Ligands are represented as an ensemble of configurations.	(Hopfinger et al., 1997)
5D	As 4D-QSAR combined with explicit models representing various induced-fit scenarios.	(Vedani and Dobler, 2002)
6D	As 5D-QSAR combined with the simultaneous analysis of several solvation models.	(Vedani et al., 2005)
7D	Such analysis comprises of real receptor or target-based receptor model data.	(Polanski, 2009)

1.5.1.3.2 Based on chemometric methods

QSAR methods are divided into two categories: linear and non-linear, based on the type of chemometric techniques used. Linear techniques include stepwise multiple linear regression (S-MLR), principal component analysis (PCA), partial least-squares (PLS), and genetic function approximation (GFA). However, PLS and GFA can also be used in developing non-linear models. Furthermore, the chemometric field has introduced various predictive modeling techniques, such as non-linear regression and algorithmic techniques like k-nearest neighbors (kNN), support vector machine (SVM), artificial neural networks (ANN), and Bayesian neural nets (BNN) (Roy and Mitra, 2011), among others.

1.5.1.4 OECD Guidelines and QSAR modelling

QSAR is a highly interdisciplinary field that involves the use of chemistry, biology, and statistics. It is a effective method for establishing mathematical relationships among the structural and chemical features of a molecule with its biological properties. The field provides a great platform for predicting the essential structural requirements necessary for optimizing the activity, toxicity, and other properties of interest. This helps in the synthesis of a smaller number of chemicals with vastly improved properties. To achieve these objectives, certain principles are adopted by the Organization for Economic Co-operation and Development (OECD) http://www.oecd.org/data_oecd/33/37/37849783.pdf.

The guidelines are as follows:

OECD principle 1 (*A defined endpoint*)- This principle commands transparency to be maintained during the selection of endpoint data for modeling. The QSAR models are anticipated to be created with homogeneous datasets made up of response data generated by a single protocol or assay.

OECD principle 2 (*An unambiguous algorithm*)- When creating predictive QSAR models, this principle lays out a clear methodology to follow. This includes the features selection process, dataset partitioning, and pre-treatment methodology. Therefore, the goal of this guideline is to increase transparency in the model-building process, making it both repeatable for other people and explicative for reaching endpoint estimates.

OECD principle 3 (*A defined domain of applicability*)- This third OECD principle emphasizes the significance of the chemical/reaction domain of applicability. Every QSAR model created with a given combination of chemicals has a unique theoretical domain that yields a trustworthy prediction within that domain.

OECD principle 4 (*Appropriate measure of goodness-of-fit, robustness, and predictivity*)- Knowing the statistical judgment of a model's predictability and stability is possible thanks to the fourth OECD principle. Using a training set, the internal model performance is measured by fitness and robustness measures, and external predictivity is measured using test sets. This offers an appropriate equilibrium between the extreme scenarios, specifically overfitting and underfitting of the prediction based on the model. Division of the dataset into training and test sets is one of the principal strategies to determine the internal stability and external predictivity of a developed QSAR model. In simpler terms, all such validation exercise is meant to verify the closeness of a prediction/ estimation by a model with respect to its experimental observation.

OECD principle 5 (*A mechanistic interpretation, if possible*)- The goal of the fifth OECD principle is to support a sound mechanistic foundation for the response that is being modeled. Only the intended analogs can be designed and developed with the help of precise knowledge about the mechanisms of action of chemicals toward a process. When proving mechanistic information toward a modeled endpoint, molecular descriptors are essential.

A graphical workflow and an explanation of some of the main points of a standard QSAR modeling approach that adheres to the OECD protocol are provided in the following **Figure 1.4**.

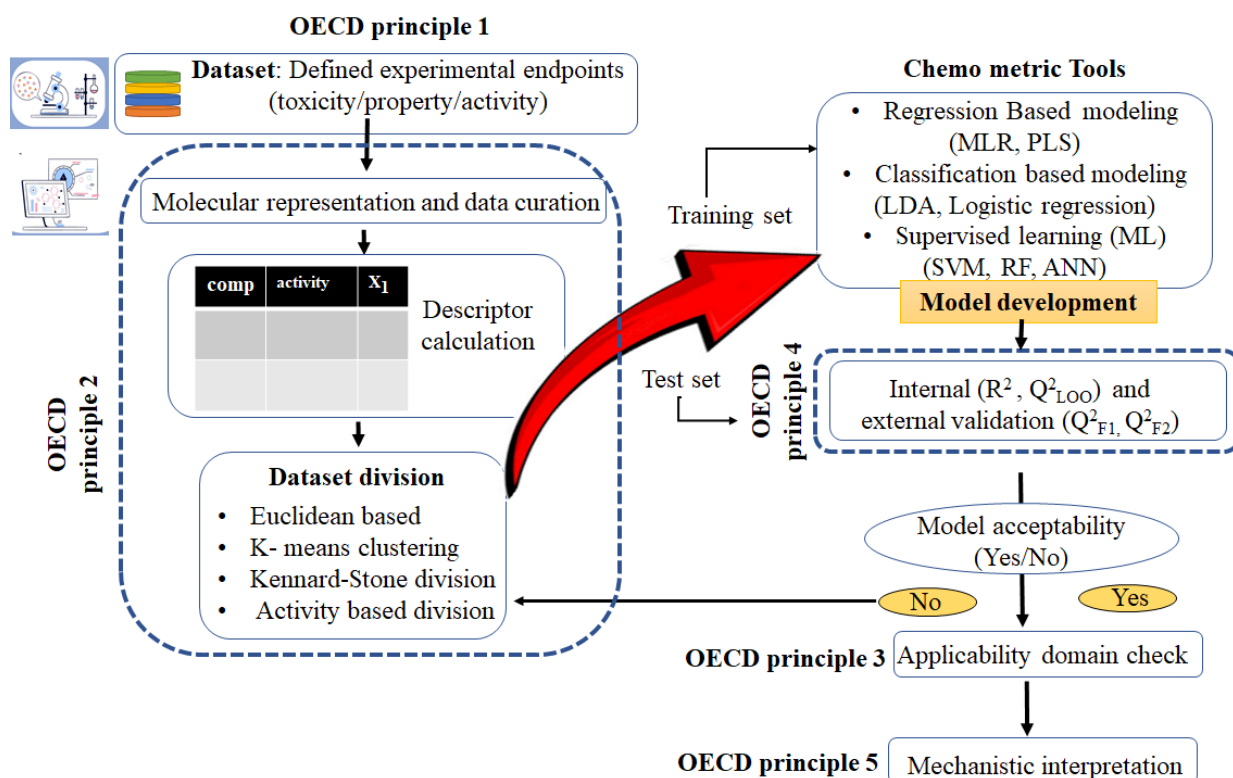


Fig 1.4 QSAR modeling approach according to OECD guidelines / principles

1.5.1.5 QSAR methodology

Developing a Predictive QSAR model involves a series of crucial and well-defined steps. These steps include data preparation, analysis, validation, and interpretation. It is important to note that 'data' refers to both the response and predictor variables. Careful handling of the data is crucial for the success of QSAR modeling. Mathematics provides an abstract platform to establish a quantitative correlation between these variables. Therefore, it is our responsibility to ensure that the preparation and analysis of chemical data are meticulous to avoid any loss of essential chemical attributes. With our expertise and attention to detail, we can confidently deliver accurate and reliable QSAR models. To provide further clarity, a brief discussion of each step of QSAR modeling is presented below.

1.5.1.5.1 Data preparation and descriptor calculation

This component comprises the preparation of the QSAR data matrix composed of response and predictor variables.

1. In order to develop QSAR models, the inputs required are chemical structures and corresponding activity/property data which can be confidently obtained from databases or scientific literature. The biological/ toxicological/ physicochemical response is then subjected to conversion into a molar unit followed by logarithmic transformation maintaining data uniformity. The mistake in experimentation itself cannot be avoided by a QSAR modelers.

2. With an accurate characterization of the molecular structure, a relation can be established between the molecular structure information and the property with a significant correlation. Hence, molecular descriptors are the core element of any QSAR or *in silico* models.

The current descriptors for nanoparticles are limited in their ability to accurately model the complex structure and interactions of nanoparticles. While experimental descriptors extracted from transmission electron microscopy (TEM) and spectral analysis have been used for model development, they are subject to variations in laboratory conditions and require standard references (Roy and Roy, 2023). Additionally, empirical descriptors such as quantum or molecular dynamics (MD) simulation descriptors are limited in their applicability due to high computational resources and commercial software expertise.

1.5.1.5.2 Currently available nano-toxicity descriptors

- **Nano descriptors derived from using quantum-chemical calculations**

Several QSAR models were developed by different groups of authors using quantum-chemical calculations. For example, QSAR models were reported to describe the cytotoxicity of 17 metal oxide NPs to the bacteria *E. coli*. (Puzyn et al., 2011) Zhou et al. (2017) modeled metal oxide NPs to *E. coli* employing DFT-derived (B3LYP method) quantum-chemical descriptors to reveal the significance of LUMO (lowest unoccupied molecular orbital) energy and Cp (molar heat capacity) properties for the developed models. Numerous software tools, both commercial and open-source, like Gaussian and VASP (Vienna Ab initio Simulation Package) offer semiempirical and/or ab initio methods that offer insights into the geometric and electronic properties of NMs. Most of the studies used quantum chemical-based descriptors to explore the toxic features of the nanostructures for which a certain quantum-chemical background is essential to perform the calculation and they are quite computationally demanding. The ab initio-based descriptor calculation is desired to explore the influence of the electronic properties of NMs.

- **Nano descriptors derived from experimental properties**

Nanostructure properties can also be obtained from the whole nanoparticle information including the diameter, surface area, constitutional properties, number of atoms, number of metal atoms, number of surface atoms, etc., or through energy-related properties such as potential energy of surface atoms and atoms in the metal, descriptors for any surface coatings, zeta potential, aqueous solubility, etc. Therefore, to explore new ways that can represent nanomaterial data, image-based descriptors (TEM and structural representation) were used that can be easily applied in neural network modeling. Further, image-based descriptors derived from images of nanoparticles, such as sphericity, allow testing of the established image processing methods and applying them to experimental micrographs. Very recently, Varsou et al. (2020) developed the NanoXtract tool (automated online tool) for the extraction of NP image descriptors derived from TEM images and described how these image descriptors could be used to predict the NP zeta potential. A similar image-processing approach was performed by Odziomek et al. (2017) to calculate geometric descriptors. Further, Mac Fhionnlaoich and Guldin (Mac Fhionnlaoich and S.Guldin, 2020) used entropy information for the characterization of NPs (nanoparticles), thus providing a better estimation of NP properties' distribution. Experimental results are also used to derive thermodynamic descriptors that

characterize the enthalpy of the formation of metal oxides and metal atoms, the enthalpy and Gibbs energy difference between the neutral/cation and neutral/anion, and the heat capacities.

- **Nano descriptors derived from 1D text representation**

Selecting suitable descriptors to explain the toxicity of nanomaterials remains challenging. Simplified molecular input-line entry system (SMILES) based descriptors have also gained momentum in nano-QSAR modeling but after some critical analysis, SMILES-based descriptors were found to bear certain drawbacks for describing endpoints of limited NPs (Blekos, 2023). SMILES-based optimal descriptors were used by Toropov et al. (2012) to develop QSAR models employing web-based CORAL software (<http://www.insilico.eu/coral>). In 2010, (Fourches et al., 2010) SMILES-based nano descriptors were calculated for 109 fluorescent magnetic NPs. Further, a novel sequence of symbols was also generated based on the SMILES descriptors in the extended form. They include information encoding molecular structure, experimental conditions, and physiochemical properties. These string characters are then merged to generate quasi-SMILES descriptors through Monte Carlo optimization in the CORAL (CORrelations and Logic), software package (Leone et al., 2018)

- **Nano descriptors derived from liquid drop model (LDM), and metal-ligand binding (MLB) theory**

The liquid drop model (LDM)-based descriptor calculation method is an innovative approach that incorporates physics to understand the interaction of NPs with biological systems. Sizochenko et al. (2014) computed different fragmental simplex descriptors, along with various ligand–metal binding characteristics of NPs, to reflect the molecular structure. With this representation, any molecule can be viewed as a system consisting of various fixed compositions and topology fragments. The characteristics of a single metal ion were described through the theory of metal-ligand binding (MLB). The LDM model portrays nanoparticles as spherical clusters or aggregates in a solution, with densely packed constituent particles, resulting in a density equal to the mass density. LDM depicts important features such as surface area and surface-to-volume ratio in a simplified manner.

- **Nano descriptors from full nanostructures**

Several nanoparticle descriptor calculation methods have been developed but most of them only consider particles that are of a single size, uncoated, and spherical in shape. This limitation has prompted the development of more advanced methods that can calculate virtual nanostructures by taking into account core materials, shape, size, chemical structure, and the number of ligands (Yan et al., 2020). One such method is the PubVINAS database which offers a Python-based program for calculating virtual nanostructures. However, this calculation method requires extensive resources and may involve commercial platforms or software for calculations.

- **Nano descriptors from MD (molecular dynamics) simulation**

Descriptors derived from MD (molecular dynamics) simulation are based on complete nanostructures. These nano descriptors have the potential to derive interaction between NMs with water or with other biological media or the so-called “nano–bio” interface (Nel et al., 2009). MD simulation does not specifically consider electrons and therefore chemical reactivity or excited states cannot be identified. They can easily investigate non-covalent

interactions like electrostatic and van der Waals forces. These descriptors primarily yield the following features: ligand-water hydrogen bonds, solvent-accessible surface area, NM–solvent interaction energy, and NM–protein interactions.

- **Nano descriptors from the periodic table (PT)**

Simple periodic table-based descriptors have been successfully used as alternatives to quantum-chemical descriptors to model various endpoints of NPs (Kar et al., 2014). The introduction of periodic table descriptors in the study of NP toxicity has made the modeling challenges easier (Fig. 3). These can be easily calculated from the periodic table without expert knowledge and are effective enough to produce similar or better results than quantum chemical descriptor-based models. Periodic table-based descriptors (simple molecular descriptors) have been a promising tool for modeling the toxicity of metal oxide nanoparticles (MeOx NPs) with high statistical quality as well as interpretability. Basic properties such as electronegativity and the metal cation's charge associated with a particular oxide (χ_{ox}) can be quickly and easily calculated using the periodic table. It is noteworthy that these fundamental characteristics remain constant across the entire size range of NPs, negating a common issue whereby NPs' physical attributes change with size range. Modeling studies of the properties and toxicity of nanomaterials have been reported (Li et al., 2022; Basant et al., 2017; Wyrzykowska et al., 2016) but not in the context of periodic table-based descriptors. **Table 1.2.** and **Figure 1.6** shows the description of the descriptors generation-wise. **Figure 1.5** shows the various applications of PT descriptor.

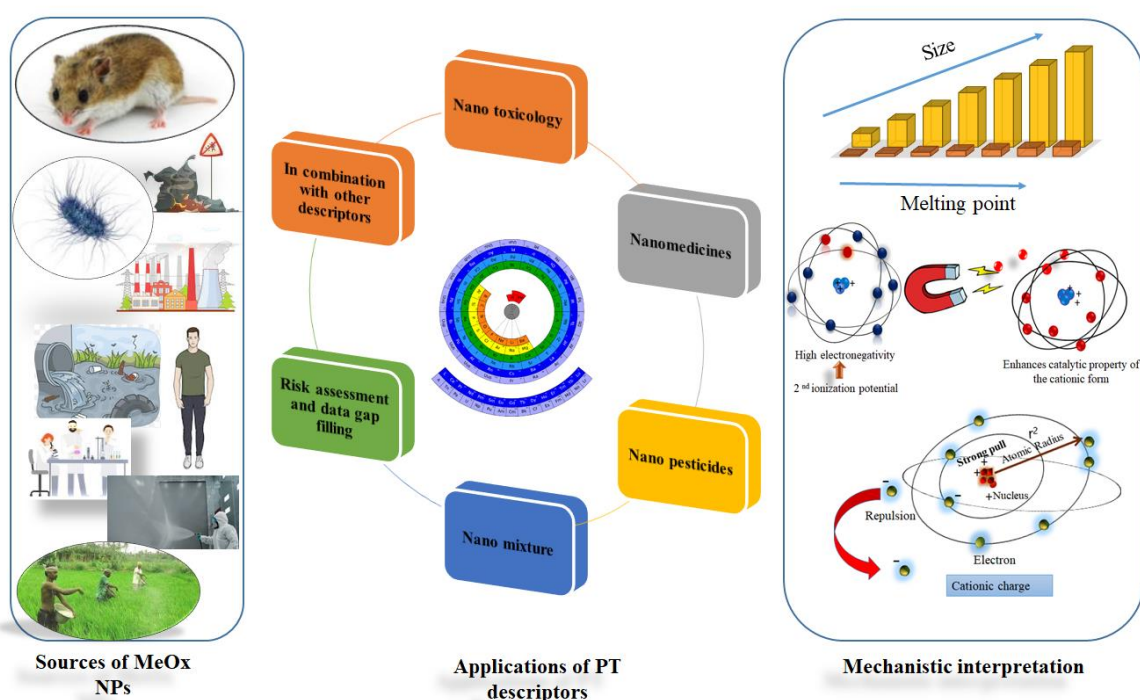


Figure 1.5 Application of PT descriptors

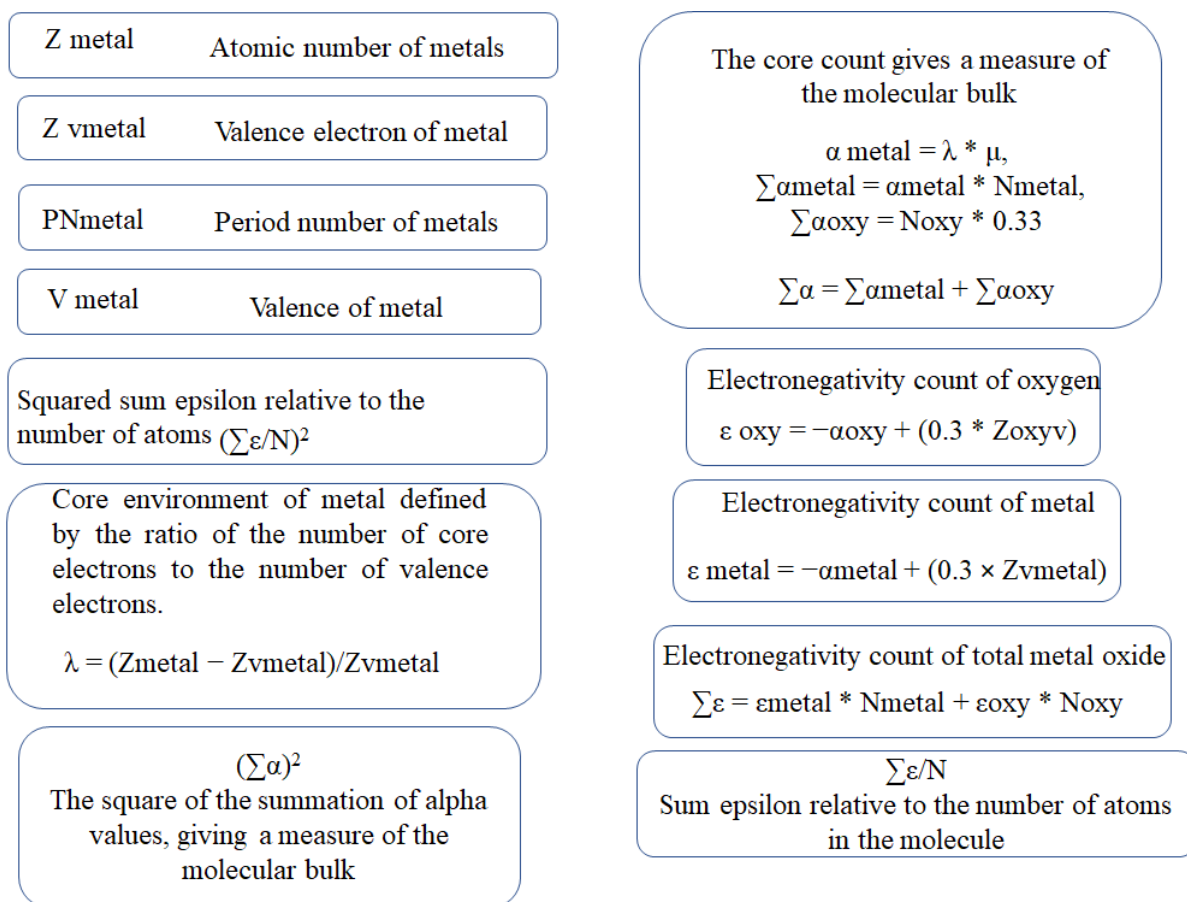
1.5.1.5.2.1 Software supporting the calculation of PT descriptors

Most of the descriptors mentioned above can be calculated with simple formulas without the need for any software. Simple descriptors for metals and semimetals can be obtained from the Elemental Descriptor Calculator software present in the DTC lab software tool website (http://teqip.jdvu.ac.in/QSAR_Tools). This software calculates a total of 31 descriptors connected to the metal, semimetal, and nonmetals obtained from the periodic table. It computes descriptors for elements having atomic numbers less than 103. Elements with the least ionization potential energy of the corresponding class (metal/nonmetal/semimetal) are considered for respective descriptor calculation. The calculated periodic table-based descriptors could be applied alone or in combination with other simple descriptors for mixture modeling.

Table 1.2. Description of the descriptors generation-wise.

<i>No.</i>	<i>Mathematical expression</i>	<i>Description</i>
<i>1st generation descriptors</i>		
1	D1_Metals	Total number of metals
2	D2_SemiMetal	Total number of semi-metals
3	D3_HeteroNonMetals	Total number of heteroatoms that are not metallic (N, O, F, P, S, Cl, Se, Br, I, At), omitting hydrogen.
4	Metals_SumIP	Ionization potential energy of all the metals added together (kJ/mol).
5	SemiMetals_SumIP	Ionization potential energy of semi-metals added together (kJ/mol).
6	Electrons_ActiveM	Number of electrons in an active metal.
7	AtWt_ActiveM	Atomic wt. of active metal.
8	MassNo_ActiveM	Mass quantity of the metal in action.
9	Neutons_ActiveM	Number of neutrons of active metal.
10	IP_ActivM	Ionization potential energy (kJ/mol) of active metal.
11	X_ActivM	The active metal's electronegativity.
12	VWR_ActivM	van der Waal radius of the metal in use.
13	Electrons_Activ_SM	No. of electrons of active metalloid/semi-metal.
14	AtWt_Activ_SM	Atomic wt. of active metalloid/semi-metal.
15	MassNo_Activ_SM	Mass no. of active metalloid/semi-metal.
16	Neutons_Activ_SM	The quantity of neutrons in an active metalloid or semi-metal
17	IP_Activ_SM	Ionization potential energy of active metalloid/semi-metal
18	X_Activ_SM	Electronegativity of active metalloid/semi-metal
19	VWR_Activ_SM	van der Waals radius of active metalloid/semi-metal

20	Electrons_Activ_NM	The quantity of electrons in an active non-metal.
21	AtWt_Activ_NM	Atomic wt. of active non-metal
22	MassNo_Activ_NM	Number of mass of active non-metal
23	Neutrons_Activ_NM	Total neutrons in an active non-metallic.
24	IP_Activ_NM	Ionization potential energy of active non-metal
25	X_Activ_NM	Electronegativity of active non-metal
26	VWR_Activ_NM	van der Waals radius of active non-metal
27	SuM_Active_M_SM	Summation of the number of metals and semi-metal
28	SuMElectrons_Active_M_SM	Summation of electrons of active metal and semi-metal
29	SuMAAtWt_Active_M_SM	Summation of At. wt of active metal and semi-meta
30	SuMMassNo_Active_M_SM	Summation of mass no. of active metal and semi-metal
31	SuMNeutrons_Active_M_SM	Summation of neutrons of active metal and semi-metal
32	χ_{Ox}	Oxidation number
33	N oxygen	Several oxygen atoms are present in the metal oxide.
34	Molar volume	
35	Atomic rad	Atomic radius
36	electronegativity	-
37	Thermal conductivity	-
38	Melting point	-
39	Electron affinity	-
40	Density	-
41	Mol wt	Molecular weight
42	Cationic charge	-
43	$\sum \chi$	Total electronegativity of the metal in particular Me_{Ox} .
44	$\sum \chi/\text{no}$	Total electronegativity of metal in specific Me_{Ox} relative to several oxygen.

2nd generation descriptorsFigure 1.6 2nd generation descriptors

1.5.1.5.3 Dataset division

The dataset is split into two sections, the training set and the test set, in order to assess the predictive model's performance. This division is based on particle similarity. The model development is undertaken using the training set, while the test set is kept separate for external predictivity assessment. The internal predictivity of the model, i.e., its predictive performance on the training set, is also evaluated. Generally, the number of compounds assigned to the training set is higher than that of the test set. As part of the dataset partitioning process, it is necessary to ensure that the compounds in the test set and the training set share the same chemical space. By doing this, the training set's representativeness over the test set is guaranteed. Various methods may be employed for dataset division, including: a) treatment of the predictor variables, b) treatment of the response variable, and c) random selection. There is no fixed rule for determining the ratio of training and test set compounds in a dataset. Nonetheless, it is usually advised to retain between 70 and 80 percent of the data in the training set and 20 to 30 percent in the test set for any dataset that contains more than 50 data points (Roy et al., 2008). The ratio of training and test sets should be selected based on the "Bias-variance dilemma," as discussed by Almaliki (2018) and Kononenko and Kukar (2007).

To divide compounds based on their predictor variables, the first approach is to group them by chemical similarity determined through descriptor matrix operations. From each group, a user-defined fraction of compounds is assigned to the training and test sets. Several techniques can be employed to divide chemicals into groups, such as k-means clustering, the Kennard Stone algorithm, the sphere exclusion principle, principal component analysis (PCA) based selection, Kohonen's self-organizing map, statistical molecular design, Extrapolation-oriented test set selection, and more. In the response variable-based division approach, compounds are categorized as diverse based on their biological, physicochemical, or toxicological response values. This method involves sorting the entire data matrix using the response column and selecting a predetermined fraction of compounds for the training and test sets from different zones, with a specific pattern such as every fourth compound. On the other hand, in the random division approach, the compounds are randomly assigned to the training and test sets based on a user-defined fraction.

1.5.1.5.4 Feature selection

Feature selection is the process of identifying important predictor variables that are strongly associated with the response variable. Software programs frequently produce hundreds or thousands of molecular descriptors; however, only a small number of these have a strong correlation with the activity. Hence, it is imperative to use appropriate feature selection tools to identify these critical variables. It is noteworthy that a large number of the descriptors exhibit intercorrelation, a factor that adversely impacts multiple facets of QSAR analysis. A higher number of compounds than descriptors is needed for some statistical method. Various chemometric tools are employed to select potential variables that correlate with endpoint data from the entire descriptor matrix. Then, the selected variables undergo appropriate statistical operations to develop the final model. Some feature selection tools used in chemometric modeling studies include stepwise variable selection, genetic algorithm, best subset selection, variable subset selection, and factor analysis.

1.5.1.5.5 Model development

Regression-based or classification-based modeling approaches can be used to develop QSAR models. These days, the most popular method is machine learning algorithms based on artificial intelligence. While models based on classification use discrete data with binary responses, models based on regression use continuous data. Common regression-based modeling approaches include MLR (multiple linear regression), PLS (partial least squares) PCR (principle component regression), and Ridge Regression, while common classification-based modeling approaches include LDA (linear discriminant analysis) and logistic regression. During the model development step, the best-selected features are combined in a single equation using an explicit formalism. Stepwise-MLR (stepwise multiple linear regression), GFA-MLR (genetic function approximation- multiple linear regression), G/PLS (genetic PLS), PLS-DA (PLS followed by discriminant analysis), and other similar techniques are all preceded by a feature selection step.

1.5.1.5.6 Validation protocols

After developing a predictive model, it is important to determine its statistical reliability. In addition to developing models, the goal of QSAR analysis is to use them to forecast the behavior of novel or untested chemicals. Therefore, it is necessary to ensure the stability and predictivity of the model. Various statistical metrics such as R^2 , R_{adj}^2 , are computed to evaluate

the model's fitness, Q_{LOO}^2 , $r_{m(LOO)}^2$ internal stability, and Q_{F1}^2 , $r_{m(test)}^2$ external predictivity. If the computed values are above the threshold limits, then the model can be considered acceptable, while external predictivity refers to the judgment on test set prediction. Validation of discriminant model parameters, such as sensitivity, specificity, accuracy, precision, F-value, and receiver operating characteristic (ROC) analysis, can be done. Major regression and classification validation metrics are listed in **Tables 1.3 and 1.4** respectively. Each QSAR model must meet the requirements of the OECD principle 4 in terms of robustness, predictivity, and goodness-of-fit in order to be considered adequate (OECD, 2014).

1.5.1.5.7 Model interpretation

Once a QSAR (Quantitative Structure-Activity Relationship) model is developed and its acceptability is verified through various metrics, the final step is to ensure that the modeled features have mechanistic interpretability. This means establishing a strong connection between the chemistry of the compounds and their biological/toxicological action or physicochemical properties. This understanding is essential for comprehending the mechanism of action involved. By combining the experimental results and observations from the model, each step of the process of behavioral manifestation of chemicals can be explicitly explained. Such knowledge is valuable in designing and developing potent analogs.

Table 1.3. Regression-based validation metrics

Metrics	Formula	Description
R^2	$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{cals})^2}{\sum(Y_{obs} - \bar{Y}_{training})^2}$	A regression model's goodness-of-fit is verified by the determination coefficient. It calculates the difference between the observed and predicted data. One is the maximum value of R^2 , signifying a perfect correlation. Here, Y_{obs} represents observed response, Y_{calc} is fitted response, $\bar{Y}_{training}$ denotes mean of the observed responses of the training set.
R^2_{Adj}	$R^2_{Adj} = \frac{\{(n - 1) \times R^2\} - p}{n - p - 1}$	The adjusted determination coefficient or explained variance or adjusted R^2 (R^2_{Adj}) is a measure of goodness of fit. It is the modified version of the determination coefficient. It incorporates the information of number of samples and independent variables. “n” stands number of data points and p is the number of descriptor. Whereas R^2 is the determination coefficient.
$Q^2(Q^2_{Loo})$	$Q^2 = 1 - \frac{\sum(Y_{obs(Train)} - Y_{pred(Train)})^2}{\sum(Y_{obs(Train)} - \bar{Y}_{training})^2}$	One indicator of the robustness of internal validation is the cross-validated correlation coefficient. Typically, leave-one-out (LOO) cross-validation is used to predict a single data point by removing it from the training set and using the model that was created using the remaining molecules. Here, $Y_{obs(train)}$ stands for training set observed response, $Y_{pred(train)}$ denotes training set predicted response and $\bar{Y}_{training}$ mean of the observed responses of the training set.
$RMSE_c$	$RMSE_c = \sqrt{\frac{\sum(Y_{obs(train)} - Y_{calc(train)})^2}{n_{train}}}$	The internal validation metric for the model's prediction error in the training set is represented by the root mean square error of calibration. The computed value of should ideally be less. In equation, $Y_{obs(train)}$ represents observed response of training set, $Y_{calc(train)}$ stands for fitted response and N_{train} is the number of training data points.

Q_{F1}^2

$$Q_{F1}^2 = 1 - \frac{\sum(Y_{obs(test)} - Y_{pred(test)})^2}{\sum(Y_{obs(test)} - \bar{Y}_{training})^2}$$

Q_{F1}^2 judge the external correlation coefficient of the predictability of the model. It shows the correlation between the observed and predicted data of test set. The $Y_{obs(test)}$ denotes observed response of the test set, $Y_{pred(test)}$ is predicted response of the test set and $\bar{Y}_{training}$ denotes mean of the observed responses of the training set.

 Q_{F2}^2

$$Q_{F2}^2 = 1 - \frac{\sum(Y_{obs(test)} - Y_{pred(test)})^2}{\sum(Y_{obs(test)} - \bar{Y}_{test})^2}$$

This metric employed to predict or judge the external correlation coefficient which also evaluates the external predictability of the model. The number of training set compounds determines the value. Here, $Y_{obs(test)}$ represents test set observed response, $Y_{pred(test)}$ is the predicted response and $\bar{Y}_{test}$ stands for mean of the observed responses of the test set.

 CCC

$$CCC = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$$

CCC is the concordance correlation coefficient. CCC measures the precision and accuracy. x_i shows observed response of i^{th} test compound, y_i represents the fitted response of i^{th} test compound, whereas, $\bar{x}$ gives information of the mean of observed response value of test compounds and $\bar{y}$ is the mean of predicted response data of test compounds.

 $RMSE_p$

$$RMSE_p = \sqrt{\frac{\sum(Y_{obs(test)} - Y_{pred(test)})^2}{n_{test}}}$$

Root mean square error of prediction shows the external validation and measures the prediction error in the validation set. Here, $Y_{obs(test)}$ represents the test set observed response, $Y_{pred(test)}$ represents fitted response, and n_{test} denotes the number of test data points.

Table 1.4 classification-based validation metrics

Sl.	Metrics	Formula
1.	Sensitivity (%)	$Sensitivity = \frac{TP}{TP + FN}$
2.	Specificity (%)	$Specificity = \frac{TN}{TN + FP}$
3.	Precision (%)	$Precision = \frac{TP}{TP + FP}$
4.	Accuracy (%)	$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$
5.	F-measure (%) (harmonic mean of recall)	$F - measure(\%) = \frac{2}{\frac{1}{Precision} + \frac{1}{Sensitivity}}$
6.	G-means (geometric mean)	$G - means = \sqrt{Specificity \times Sensitivity}$ $P_r(a) = \frac{(TP + TN)}{(TP + FP + TN + FN)}$

7. **Cohen's kappa (k)**

$$P_r(e) = \frac{\{(TP + FP) \times (TP + FN)\} + \{(TN + FP) \times (TN + FN)\}}{(TP + FN + FP + TN)^2}$$

8. **Matthews correlation coefficient (MCC)**

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

1.5.2. Read -across (RA) predictions

The European Chemicals Agency (ECHA) has defined read-across (RA) as a technique to predict information about a substance's properties. This is done by using data from similar substances. RA is based on the principle that similar substances should have similar properties. It works like a local QSAR model, where chemically or biologically similar substances are used to predict the properties of a target substance. The read-across data gap filling techniques include four different approaches: one-to-one, one-to-many, many-to-one, and many-to-many (Oomen et al., 2015). However, two different analogues may have different toxicity patterns, increasing the chance of false predictions. The category approach is a more robust prediction technique that uses a large number of analogs to identify trends within categories and forecast the toxicity of unidentified analogs. Both qualitative and quantitative data can be used with read-across.

1.5.3. Machine learning (ML)

This section introduces all ensemble learning models. ML is a subset of AI (Artificial Intelligence), where the machine learns from data, improves the performance from past experiences, and makes a prediction based on it (Marcílio et al., 2020). In the supervised learning approach, a model is trained on labeled datasets. Regression analysis algorithms are trained and learned from both input features and output labels. Regression analysis seeks a mapping function from the input features for a continuous output function. Machine Learning (ML) approaches can be classified into unsupervised learning and supervised learning. In the supervised learning approach, a model is trained on labeled datasets. Regression analysis algorithms are trained and learned from both input features and output labels. Regression analysis seeks a mapping function from the input features for a continuous output function. A set of QSAR descriptors is used to train the algorithm and extract information in unsupervised learning. When using descriptive modeling techniques, it helps to identify significant data features without requiring the use of dependent variables (Y variables).

1.5.3.1 RF (random forest)

Ensemble learner, RF (random forest) is often used for its flexibility. Whether it is regression or classification, RF is a versatile learning method that can handle both. It works by building several “decision trees” in the training phase and generates average forecasts of various decision trees involved. In other words, it combines the results of different decision trees to make the best possible decision. Though the goal variable in classification-based issues is categorical, numerical values are present in regression. One advantage of RF is its capacity to analyze large datasets with greater efficiency. It can be regarded as a dimensionality reduction method since it analyzes large input data and finds all the important variables. While handling RF datasets, the model emphasizes the importance of parameters, which are considered to be highly helpful aspects (El Naqa et al., 2015).

1.5.3.2 AdaBoost (Adaptive Boosting)

Adaboost is one of the best-boosting algorithms. It uses an ensemble learning method. This approach of machine learning is based on the idea of creating accurate prediction rules by combining many relatively weaker and inaccurate rules and assisting in alleviating overfitting issues. It is possible to make a smarter learner by altering the training data intelligently and constructing many sub-models. During the training phase, an infinite number of decision trees are included. In the process of building the first decision tree, inaccurate data are indicated within the main model.

The same data are used as input for different models. Until a predetermined number of base learners are produced, this process is repeated. It uses a weighted average relying on the subsets to determine whether it should be included in the finalized model. In reality, some data may include something linear, and others may not. Therefore, utilizing the ensemble AdaBoost allows us to capture the nonlinear predictions and make a precision prediction for such data (Russell, 2010).

1.5.3.3 Gradient Boost (GB)

In 2002, Friedman (2002) suggested an ensemble learning algorithm for both regression and classification. GB method is associated with each repeat of the randomly chosen training data set with the fundamental model. The overfitting is exhibited by randomly subsampling the training set data; by doing so, the execution time and model accuracy are also improved. Since every repetition of the model must include small data (as a training set) the regression becomes quicker. Other parameters that must be modified or changed in the GB approach are the shrinkage aspect, also known as the learning rate, which should not be kept too high, and n-trees, which should not be retained too little (Breiman, 2001).

1.5.3.4 Extreme Gradient Boost (XGBoost)

In a similar manner as described (Freund, 1997), for effective and scalable implementation performance, XGBoost, an ensemble machine learning algorithm based on tree boosting, employs a gradient-boosting framework. With the use of multiple models for gradient enhancement, ensemble learning employs multiple predictions and produces results with good adjustments for continuous variables and outliers. It is an efficient tool for dealing with both regression and classification problems. The basic idea is to build “N” regression trees to train each subsequent tree using the residual from the previous tree. Models are built recursively until there is no improvement in the results obtained. The new models predict the residuals of the prior model and then collectively provide the final predictions (Saadat et al., 2022). The gradient descent algorithm is used to minimize the loss while adding new models. Then, these individual predictors or classifications are ensemble to give more strong and more precise predictions. Tuning can be done using the grid search method. The workflow of the ML algorithm is represented in **Figure 1.7**.

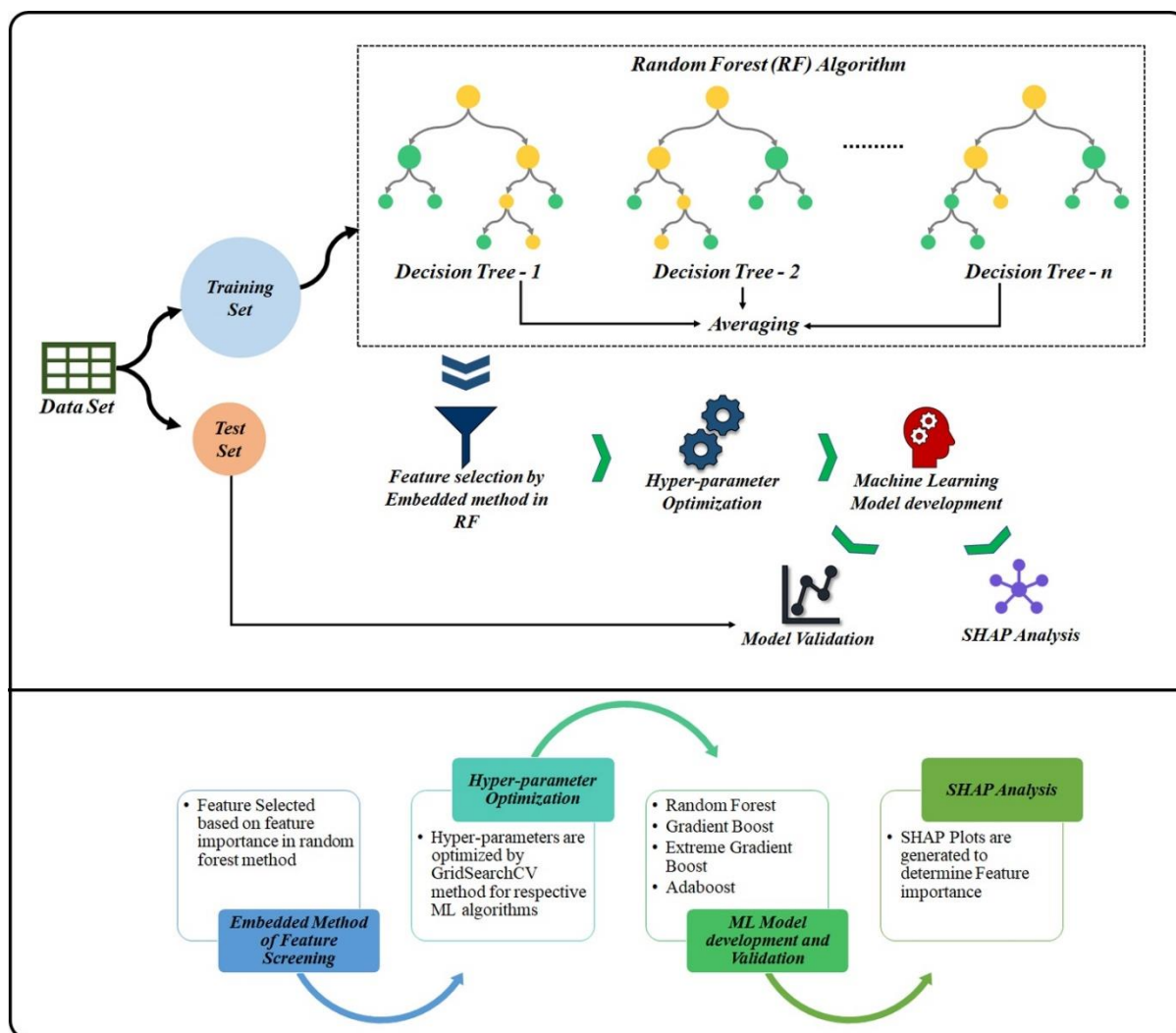


Figure 1.7 Workflow of the ML algorithm.

1.6. Application of in silico modeling approaches for nanoparticle toxicity assessment

Scientists worldwide are researching nanoparticle risk assessment using *in silico* techniques such as QSAR and machine learning. They aim to create accurate prediction models and reliable guidelines. Selected case studies are mentioned here. In their study, Puzyn et al. (2011) used all available quantum-chemical descriptors to characterize the cytotoxicity of 17 metal oxide NPs (MONPs) to the bacterium *E. coli*. They found that the enthalpy of formation of a gaseous cation, one of the features, was able to predict 86% of the correlation with the cytotoxicity. By employing periodic table-based descriptors to conduct the same work, Kar et al. (2014) postulated that the primary cause of oxidative stress in *E. coli* is still the release of electrons, which starts the formation of ROS. The enthalpy of formation of a gaseous cation (ΔH_{me+}) and polarization force (Z/r) are good candidates to explain the cytotoxicity to *E. coli*, according to Mu et al. (2016), who also performed QSAR modeling using the same dataset. Consequently, Basant and Gupta (2017) employed a random forest (RF) classification model to describe the cytotoxicity using enthalpy of formation of a gaseous cation and oxygen in weight percentage as descriptors. Additionally, utilizing molar refractivity, polar surface area,

and oxygen percent, Singh and Gupta (2014) created an ensemble learning model. The importance of LUMO (lowest unoccupied molecular orbital) energy and Cp (molar heat capacity) properties for the developed models was revealed by Zhou et al. (2017), who modeled metal oxide NPs to *E. coli* using DFT-derived (B3LYP method) quantum-chemical descriptors. The same dataset was used by Sizochenko et al. (2014), who used descriptors based on the liquid drop model (LDM). The toxicity of MONPs was found to be correlated with the number of metal atoms in oxide, the charge of a metal cation, the number of oxygen atoms in oxide, and metal electronegativity, according to Fjodorova et al. (2017), who employed artificial neural network (ANN) models. Zhnag et al., (2023) performed ML techniques, specially support vector machine (SVM) and neural network (NN) to determine the cytotoxicity of mixtures of nanoparticles using different mixing ratios (22 binary combinations) on *Escherichia coli*.

Following that, these mathematical models were applied to determine the unknown activity/toxicity of NPs using available NPs with known toxicities from experiments or computed descriptors pertaining to the structural and electronic characteristics of the nanoparticles under study. **Figure 1.8** shows the general work flow of nano-QSAR modelling.

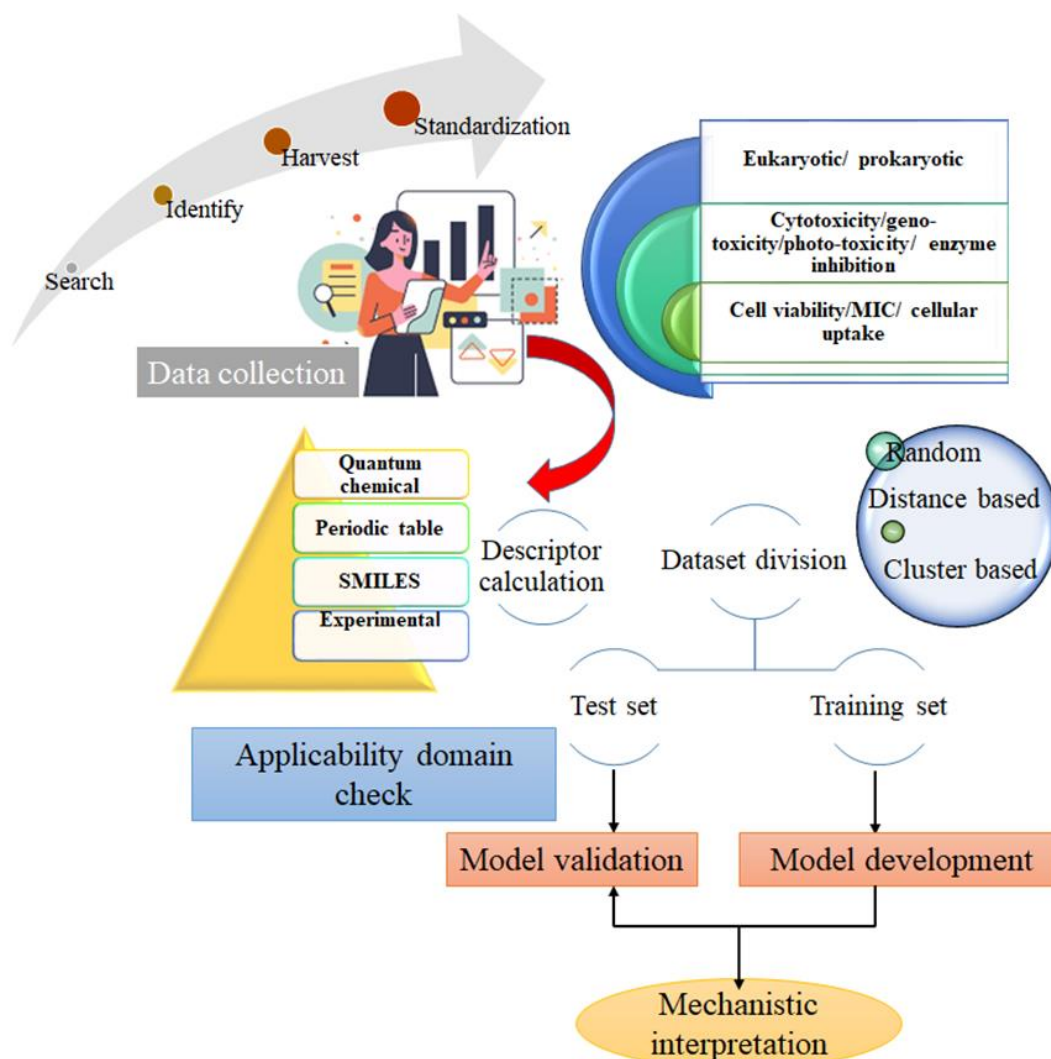


Figure 1.8 General Workflow of nano-QSAR modeling.

CHAPTER - 2

PRESENT WORK

Chapter 2: Present Work

The concept of nanoparticles and nanotechnology emerged in the 20th century. From the beginning of the quantum theory, it was clear that it would be difficult to handle the physiochemical properties of nano-sized compounds. Nano-sized materials have been in commercial use for a long time. Currently, metallic oxide nanoparticles (MeOx NPs) with their numerous properties are used in different fields like water treatment, food, agriculture, textiles, medicines, pharmaceuticals, cosmetics, and electronics (Tripathi et al., 2019; Wu et al., 2020). Thus, MeOx NPs have become an essential part of human life in recent years. However, with the increasing use of nano-sized MeOx NPs, concerns have also arisen regarding their potential risks to the environment and human health. These concerns are growing as nanoparticles are released intentionally or unintentionally into the environment (Subhan et al., 2021). For example, MeOx NPs used in various consumer products may end up in aquatic or terrestrial environments, where the toxicity mechanism is yet to be completely explored (Nemmar et al., 2002). The toxicity of the single NPs has been tested intensively for recent decades but the combination of a single NP with other NPs or metals may cause co-exposure effects on living organisms. Therefore, it is crucial to consider this issue and take necessary measures to prevent any potential harm. Given the rapid development of nanotechnology, it is impractical to test all NPs on a case-by-case basis (Pan et al., 2016). Additionally, the dynamic properties of MeOx NPs, such as size, shape, charge, chemical composition, crystal structure, and surface modification, make it challenging to determine their toxic effects experimentally. To tackle these issues, different strategies have been implemented to evaluate the toxicity of nanoparticles. This is the initial step in assessing the toxicity and potential risks of nanoparticles to workers, users, and the environment. The field of nanotoxicology relies mostly on traditional methods used to assess the toxicity of other chemicals, including drugs, in vivo and in vitro. Toxicity is measured based on three parameters: dose, dimension, and durability. When it comes to nanoparticles, the properties that induce specific toxicity effects and the mechanisms that mediate the adverse effects are still not well understood. As a result, many questions remain open in terms of NP risk assessment. Chemical legislative frameworks such as Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) have considered such non-experimental methods reliable. Another, Regulation (EC) No 396/2005 covers matters relating to allowed limits of residues in food and the official controls. Regarding the safety aspect of human and animal health, the European Regulations (EU) No 283/2013 and (EU) No, 284/2013 were set up for plant production products i.e., PPP, and for data requirement of the active products used in nano pesticides. Nanotechnology, nanomedicines, and nanotoxicology also go hand in hand. The increased use of engineered nanoparticles (ENPs) in pharmaceutical sciences has imposed major concerns about their possible nanotoxicity and has emerged as a major sub-discipline of nanotechnology. Determination of the long-term toxic effect of NPs may consume a lot of time in clinical trials involving animal models and human volunteers. Toxicity and biokinetics information of nano-products along with the information on unintended human exposure or intended delivery for medical purposes is necessary to determine the perceived risks of nanoparticles. Considering these facts, the determination of toxicity for nanomedicines is very important. Hence, regulatory bodies are searching for ways to assess their risk before application as pharmaceuticals. Nanoinformatics in combination with nanoparticles (NPs) has attracted the attention of many scientists to create comprehensive databases comprising physical and chemical features of the materials based on which Quantitative Structure-Activity Relationship (QSAR) models are generated. *In silico* toxicological methods, such as QSARs play an efficient role in complementing and reducing

animal testing. QSAR can also be used as an excellent tool for determining the physicochemical properties of NPs computationally involving various types of descriptors. These studies can be utilized to forecast the properties or toxicity of those NPs that are yet to be explored or synthesized. In the field of nanotoxicity, only a few studies have been done using QSAR models and only a few reviews concerning the risk of NPs have been reported to date (Puzyn et al., 2011; Burello et al., 2011; Puzyn et al., 2009). Furthermore, to date, all QSAR models have predicted the toxicity of NPs without giving proper information regarding important factors like dose, hydrodynamic size, etc. which can be a determining factor in influencing toxicity. Over the last years, many cheminformatics predictive models (Nano-QSAR) have become a notion to address the limitation of knowledge of nanoparticle behavior. In water bodies, aquatic organisms are exposed to mixtures of nanoparticles with varying compositions over time. These nanoparticles can accumulate in organisms and transfer through the food chain, ultimately reaching the top consumers. The accidental intake of nanoparticles can lead to adverse consequences beyond conventional contamination (Handy et al., 2008). The large surface area-to-volume ratio and surface chemistry of nanoparticles make them potentially more toxic. While many studies have focused on the impact of nanoparticles on human health and the environment, there is still limited knowledge about their environmental fate and their potential to accumulate in the food chain (Nelms et al., 2018). Although the toxicity of nanoparticles on aquatic organisms has been studied in the context of risk assessment, there is still insufficient knowledge about the toxic impact of ENPs in the aquatic ecosystem.

Computational modeling is a useful technique for predicting toxicity and has been recognized by the United States Environmental Protection Agency (U.S. EPA) as an efficient method. It helps in reducing the need for animal testing and provides more accurate results, making it an essential resource for researchers. Many scientists have developed nano-QSAR models to predict the toxicity endpoints of biological systems. These models establish a relationship between the independent and dependent variables, i.e., the physicochemical properties of MeOx NPs and the toxic effects. Both prokaryotic and eukaryotic organisms have been extensively studied to determine MeOx NP toxicity. However, predictive models do not address the problem of missing data in nanoparticle datasets. Hence, there is a lack of models that can fill the data gaps. The nano-read across approach is a technique that fills data gaps by predicting the information of endpoints for one nanoparticle using an experiment or calculated data from another nanoparticle. In the absence of appropriate validated QSAR models, the nano-read across approach can be an attractive technique to fill the existing data gap for toxicity analysis from a regulatory perspective (OECD, 2007). The read-across method has been successfully applied to predict the toxicity of many compounds (Sizochenko et al, 2018) (Gajewicz et al, 2014). This is particularly much needed in the field of nano-safety with limited high-quality toxicological data availability. The general idea of read-across is to extrapolate the endpoint information for one or more compounds by using data from other compounds having the same endpoint information which is considered similar in one or another way. The fundamental concept of read-across is based on the similarity among chemicals and it follows a hypothesis that similar compounds will have similar biological activities. Regardless of whether the endpoint is quantitative or not, read-across follows some basic steps. First, structural features common to the two compounds under consideration should be identified, and secondly, it should be assumed that the toxicological effect or environmental fate property (i.e., the endpoint) could be used to assess the missing value of the property or activity of the targeted compound. QSAR models are usually constructed using a linear regression, but ENP toxicity can evoke non-linear responses. Therefore, non-linear methods are explored to develop reliable and predictive models. In recent years, the use

of non-linear techniques has increased to deal with complex non-linear multidimensional interactions. However, most *in silico* modeling methods deal with small datasets with restricted diversity, which can lead to inconsistency between eco-toxicological experiments in the literature (Forest et al., 2019; Gajewicz et al., 2018).

2.1. Objectives of the dissertations

Therefore, in the present dissertation work, several *in silico* techniques were employed to study computational modeling to predict the toxicity endpoints of biological systems. The main purpose of the dissertation was to utilize various *in silico* tools to identify the key descriptors or features responsible for modulating toxicity of NPs towards living organisms and to understand how QSAR, read-across-based predictions, and supervised machine learning (ML) models predict toxicity upon exposure to metallic ENPs. Therefore, the current study focuses on the application of an efficient approach that is capable of predicting different properties of unknown nanoparticles based on information derived from limited data sources. One of the criterion to enhance a nanotoxicity model is the selection of the appropriate structural descriptors of NPs that are in charge of enhancing NP toxicity. Periodic table-based descriptors have been a promising tool in predicting toxicity profile and risk assessment of MeOx NPs with high predictivity and interpretability. This type of descriptor can indicate the features and mend the mechanism interpretation. Some properties (size, zeta potential, molecular weight, mass percentage of metal elements, and cation charge) are investigated to have a better understanding of NP structure and its influence on toxicity characterization. The major part of the work deals with the development of predictive and statistically robust QSAR models.

2.2. Datasets employed for the development of nano-QSAR / ML models.

In this dissertation, eight different *in silico* prediction modeling studies have been carried out and presented. For performing the requisite *in silico* studies, several datasets were collected from various reliable sources as mentioned in **Table 2.1**.

Table 2.1. Collected nanotoxicity datasets employed in this dissertation.

Datasets	Cell	Target/ endpoints	No. compounds	references
I	Macrophage RAW 264.7 cells	LDH release (ldh _R)	25	Forest et al., 2019
II	<i>E. coli</i>	-log LC ₅₀ (dark and photo induced condition)	17	Pathakoti et al., 2014
III	<i>HaCat cell</i>	log(1/LC ₅₀)	18	Gajewicz et al., 2015
		% Cell death	15 MeO _x NPs (total 117 data points),	Choi et al., 2019
IV	BEAS-2B cell (ATP assay method).	log (1/EC ₅₀)	9	Zhang et al., 2015
	BEAS-2B cell (MTS assay method)		10	

	RAW 264.7 cells (MTS assay method).		10	
	RAW 264.7 cells (ATP assay method).		12	
V	THP-1 cell line	fold change (FC _{IL-1β}) proinflammatory cytokine interleukin (IL)-1 β (IL-1b) release	32	Huang et al., 2020
		$-\ln(p/(1-p))$	54	Mitteroecker and Bookstein, 2011
VI	HK-2 cells	Cell viability	72	Sang et al., 2022
VII	BEAS-2B cells	Zeta potential	18	Cao et al., (2020)
		Cell damage	131	Toropova et al. (2015)
VIII	<i>D. magna</i>	$\log(1/EC_{50})$	28	NanoToxdb database
	Zebrafish	Enzymatic activity of metalloprotease zebrafish hatching enzyme	24	Sizochenko et al., 2017

2.1.1. Dataset I (Study 1)

In the current study, data on the toxic potential of 25 NPs (nanoparticles) from 6 metal (hydro) oxides families (CeO_2 , SiO_2 , TiO_2 , $AlOOH$, ZnO , and $Ni(OH)_2$) were obtained from laboratory experiments from a literature (Forest et al., 2019) that were expressed in terms of LDH release (ldh_R) from macrophage RAW 264.7 cells. QSAR models were developed to computationally evaluate the cytotoxicity of MeOx NPs towards RAW 264.7 cells. Both MLR and PLS models were developed with 25 NPs (nanoparticles) from 6 metal (hydro) oxide families (SiO_2 , TiO_2 , CeO_2 , $AlOOH$, ZnO , and $Ni(OH)_2$) using the release of lactate dehydrogenase (ldh_R) from the cell as the endpoint. Multiple linear regression (MLR) models were developed using the best subset selection method followed by enhancement of model derived predictions by the application of the intelligent consensus predictor (ICP) tool. On the other hand, a partial least squares (PLS) model was developed using the Small Dataset Modeler software which can handle a limited number of data points. Periodic table descriptors and some physicochemical descriptors have been used to characterize the nanostructure of MeOx NPs. All models were validated meticulously by employing internal and external validation metrics followed by the randomization test and applicability domain (AD) study. The statistical metric values suggested the robustness of the models. The insights from both models suggested that electronegativity, solubility, formation of a metal cation increase the cytotoxicity whereas non-metal NPs and an increase of the oxidation degree contribute to low toxicity towards RAW 264.7 cells. A short proposed mechanism has also been explained to determine the cause of cellular toxicity by the MeOx NPs.

2.1.2. Dataset II (Study 2)

In the present study, the cytotoxicity data of 17 MeO_xNPs to *E. coli* (Pathakoti et al., 2014) was expressed in terms of negative logarithm of lethal concentration causing 50% growth inhibition of *E. coli* (-log LC₅₀). The dataset was considered for 2D QSAR modeling under two experimental conditions, i.e., in the dark and in the presence of light. Multiple linear regression (MLR) models were developed using 17 MeO_xNPs on *Escherichia coli* (*E. coli*) bacteria cells under both light and dark conditions. The models were developed applying Small Dataset Modeler software, version 1.0.0 (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab/) which generates models with a limited number of data points. Periodic table-based descriptors (both 1st and 2nd generation) were used for the modeling purpose. Two statistically significant MLR models based on photo-induced toxicity and dark-based toxicity were developed. From the developed models, we interpreted that an increase in valency and oxidation state of the metal will decrease the cytotoxicity whereas the atomic radius of the metal and electronegativity of MeO_xNPs influences the toxicity towards *E. coli* cells. The MLR models were validated using different internal validation metrics. Additionally, we have collected 42 MeO_xNPs as an external set to observe the predictive power of the two developed MLR models and categorized them into toxic and non-toxic classes. The chemical features selected in the developed models are important for understanding the mechanisms of nanotoxicity. Thus, the developed models can be a scientific basis for designing safer NPs.

2.1.3. Dataset III (Study 3)

In the present study, two different cytotoxicity datasets of metal oxides were employed for performing nano-QSAR modeling toward the HaCat cell line. The first dataset (3.1) comprises 18 nano-metal oxides taken from a previously published paper (Gajewicz et al., 2015). Their cytotoxicity is expressed as logarithms of the inverse molar lethal concentration i.e., log (1/LC₅₀). The second dataset (3.2) consists of 15 MeO_x NPs (total of 117 data points), taken from the Safe and Sustainable Nanotechnology (S₂NANO) database built by a research group in North Korea (Choi et al., 2019). The cytotoxicity was expressed in the percentage (%) of cell death. Multiple Linear Regression (MLR) modeling with 16 MeO_x NPs dataset and Partial Least Squares (PLS) regression modeling with 15 MeO_x NPs dataset (117 data points with different physical features and conditions) were performed to determine the cytotoxicity towards HaCat cells. Easily interpretable periodic table descriptors (1st and 2nd generation) were calculated from the primary descriptors (De et al., 2018). Additionally, a few periodic table descriptors were calculated with the help of Elemental Descriptor Calculator software available from (<https://dtclab.webs.com/software-tools>) and some previously reported physical descriptors were employed for modeling. The Genetic Algorithm (GA) followed by the Best Subset Selection (BSS) technique was applied before the model development. The final MLR model (4 descriptors) and PLS (2 LVs) model for the two datasets were rigorously validated with various stringent validation criteria following the strict rules of OECD guidelines. The statistical data of the MLR model were robust and reliable. Features like electronegativity, formation of metal cations, an increase of solubility, and dose were important for the mechanism of toxicity. Therefore, these models can be efficient tools in designing modern materials with optimal properties and minimal potential toxicity.

2.1.4. Dataset IV (Study 4)

In the present work, we have applied computational modeling approaches to explore the influence of metal oxide nanoparticles on the toxicity of bronchial epithelial (BEAS-2B) and murine myeloid (RAW 264.7) cells (Zhang et al., 2015). The cytotoxicity data were obtained from two different assay methods (i.e., MTS and ATP) for a single cell line. A total of four datasets were used for modeling purposes with a half-maximal effective concentration in logarithmic ($\log(1/EC_{50})$) scale as the endpoint. Here, four (4) datasets containing MeOx ENPs of 9, 10, 10, and 12 respectively were taken for modeling purposes. Combining compounds from two datasets (ATP assay data) out of the four datasets from RAW 264.7 and BEAS-2B cell lines, common eight MeOx ENPs were used to develop the nano-quantitative toxicity-toxicity relationship (nano-QTTR) model. The same four datasets were further utilized to predict the similarity-based approach. Initially, partial least squares (PLS) regression models were developed applying the Small Dataset Modeler software version 1.0.0 (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab/) using four datasets having effective concentration ($EC_{50}\%$) as the endpoints and employing only periodic table descriptors. To further explore the predictions, we applied a read-across approach using the descriptors selected in the QSAR models. Also, the inter-endpoint cytotoxicity relationship modeling (quantitative toxicity-toxicity relationship or QTTR) was conducted to evaluate the toxicological characteristics of MeOx nanoparticles. The results obtained from the PLS models and the QTTR model were found to be statistically significant. It was found that the result obtained by nano-read-across provided a similar level of accuracy as provided by QSAR. The models show that the formation of metal cations and electronegativity play a role in the manifestation of cytotoxicity. Thus, it is feasible to design safe nanopesticides that could be more effective than conventional analogs.

2.1.5. Dataset V (Study 5)

The present work attempts to develop nano-QSAR models (both regression as well as classification-based) to determine the inflammatory potential of MeOxNPs based on the THP-1 cell line. A comprehensive dataset comprising 32 MeOxNPs (Dataset 5.1) was used to develop a regression model with fold change (FC) of pro-inflammatory cytokine interleukin (IL)-1 β release in the THP-1 cell line as the endpoint. Further, the same number of MeOx NPs with varying doses was modelled for the cell viability $[-\ln(p/(1-p))]$ endpoint (Dataset 5.2) (Mitteroecker and Bookstein, 2011). The results obtained from regression models were statistically significant. The descriptors obtained from the developed predictive models inferred that dose, electronegativity and the presence of metal ions and oxygen can be responsible for IL-1 β leakage from the THP-1 cell line. Based on our results, we can conclude that periodic table-based descriptors, incorporated into the QSAR models, are reliable for modelling pro-inflammatory potential. Researchers can use these comprehensive results to design metal oxide nanoparticles with lower toxicity and determine the cause of pro-inflammatory conditions induced by MeOxNPs

2.1.6. Dataset VI (Study 6)

The present study reports the ensemble learning approach-based nano-quantitative structure-activity relationship (QSAR) models for predicting the cytotoxicity of heavy metals adsorbed on nano-TiO₂ to human renal cortex proximal tubule epithelial (HK-2) cells. A mixture of nano-TiO₂ powders was added to HK-2 cells in Hyclone DMEM medium supplemented with 10% fetal bovine serum (FBS) and 100 mg penicillin/streptomycin and maintained at 37 °C in the presence of 5% carbon dioxide. Nine concentrations of heavy metal salts were added to a constant amount of nano-TiO₂ (25 µmol/L) (Sang et al., 2022). Ensemble learning-based nano-QSAR implementing gradient boosting and bagging algorithms, i.e., random forest (RF), Ada boost, Gradient boost, and Extreme Gradient boost were constructed and utilized to establish statistically significant relationships between the structural properties of NPs and the cause of cytotoxicity. To demonstrate the predictive ability of the developed nano-QSAR models, the simple periodic table descriptors requiring low computational resources were utilized. Thus, the work manifests that ML in conjunction with periodic table descriptors can be used to explore the features and predict unknown compounds with similar properties

2.1.7. Dataset VII (Study 7)

In the present study, Dataset 7.1 consists of 18 metal oxide nanoparticles (MeOx NPs) that were explored on five types of metal oxides. This data was obtained from Cao et al. (2020) work, where the zeta potential of MeOx NPs was measured in a cell culture of 20% foetal bovine complete medium. Dataset 7.2 was taken from Toropova et al. (2015) where the cell damage measurement was performed based on the uptake of propidium iodide (PI). The dataset is related to four doses (50, 100, 150, and 200 µg/mL) and seven exposure times (from 1 to 7 hrs) that result in 132 MeOx NP data points. The goal of the study is to construct fully validated property-based models to extract relevant features for estimating and influencing the zeta potential and obtaining the toxicity profile regarding cell damage in the treatment of cancer cells. To achieve this, QSPR modeling was first performed with 18 metal oxide (MeOx) NMs to measure their materials properties using periodic table-based descriptors. The features obtained were later applied for zeta potential calculation (imputation for sparse data) for MeOx NMs that lack such information. To further clarify the influence of zeta potential in cell damage, a QSPR model was developed with 132 MeOx NPs to understand the possible mechanism of cell damage. The results showed that zeta potential, along with seven other descriptors, had the potential to influence oxidative damage through free radical accumulation, which could lead to changes in the survival rate of cancerous cells. This study also guides safer designing and toxicity assessment of MeOx NPs using the developed QSPR and QSAR models

2.1.8. Dataset VIII (Study 8)

For this study on the toxicity testing of metal oxide nanoparticles on aquatic organisms, two datasets were used. The first dataset (Dataset 8.1) was sourced from the NanoTox database (<http://iitrindia.org/envis/Default.aspx>) against *D. magna* and included 28 different MeOx NPs. It provided data on their physiochemical properties such as size and the duration of exposure during toxicity testing. The data for *D. magna* was expressed in mg/L, but it was later converted to a negative logarithm scale after conversion to a molar scale. The second dataset (Dataset 8.2) was obtained from the literature (Sizochenko et al., 2017) and was used to predict the toxic outcomes in zebrafish. It analyzed the enzymatic activity of metalloprotease zebrafish hatching enzyme, ZHE1 for 24 metal oxides. Regression, read-

across approach (RA), and machine learning (ML) algorithms in accordance with the principles of OECD (Organization for Economic Cooperation and Development) were performed for the creation of QSAR modeling. The aim of this work was to investigate which features are important drivers of in vivo nanotoxicity in *D. magna* and *Danio rerio* using simple periodic table-derived descriptors. Further, we have examined the effectiveness of read-across-derived similarity measures in comparison to traditional QSAR models. The study emphasized the importance of developing transferable, reproducible, and easily interpretable models for the early identification of nanoparticle features contributing to aquatic toxicity.

CHAPTER - 3

MATERIALS AND METHODS

Chapter 3: Materials and Methods

The present dissertation aims to implement a transparent methodological framework for the development of predictive QSAR models for toxicity assessment of the NPs targeted against various toxicity endpoints. The development of model-friendly databases based on the FAIR (findability, accessibility, interoperability, reuse) principles has significantly contributed to improving scholarly data quality. FAIR datasets play a crucial role in enhancing computational modeling in nanotoxicology by promoting machine-actionability and improving the reusability of computational systems for searching and using data without human intervention. Although some data may not be suitable for proposed repositories, they still hold great value for research, reproducibility, and reuse.

3.1. Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach.

3.1.1. Toxicity data

3.1.1.1 Series 1 NPs (nanoparticles) compounds

In this work, data on the toxic potential of 25 NPs (nanoparticles) from 6 metal (hydro) oxides families (CeO_2 , SiO_2 , TiO_2 , AlOOH , ZnO , and Ni(OH)_2) (**Table 3.1**) were obtained from laboratory experiments from a literature (Forest et al., 2019) that were expressed in terms of LDH release (ldh_R) from macrophage RAW 264.7 cells. A total of forty descriptors were calculated covering both physicochemical and periodic table-based descriptors (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>).

Table 3.1 Toxic potential data of 25 NPs (nanoparticles) from 6 metal (hydro) oxides in terms of ldh_R .

CompNo	Metal Oxides(hydroxides)	LDH
1	O[Al]=O	1.1
2	O[Al]=O	1.08
3	O[Al]=O	1.05
4	O[Al]=O	1.07
5	O[Al]=O	1.07
6	$[\text{Ni}+2].[\text{OH}^-].[\text{OH}^-]$	2.76
7	$[\text{Ni}+2].[\text{OH}^-].[\text{OH}^-]$	2.57
8	$[\text{Ni}+2].[\text{OH}^-].[\text{OH}^-]$	2.99
9	$[\text{Ni}+2].[\text{OH}^-].[\text{OH}^-]$	2.2
10	$[\text{Ni}+2].[\text{OH}^-].[\text{OH}^-]$	2.07
11	O=[Ni]	0.88
12	O=[Zn]	2.62
13	O=[Zn]	2.81
14	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	2.02
15	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	2.58
16	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	2.08
17	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	0.98
18	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	2.08
19	$[\text{O}-2]=[\text{Ce}+4]=[\text{O}-2]$	0.82
20	O=[Ti]=O	0.88

21	O=[Ti]=O	1.01
22	O=[Ti]=O	1.29
23	O=[Ti]=O	1
24	O=[Si]=O	1.2
25	O=[Si]=O	1.09

3.1.2. Descriptor calculation

To develop nano-QSAR models, we have gathered information related to MeO_x NPs (hydroxides) to encrypt their toxicity towards macrophage RAW 264.7 cells. The complexity and versatile behavior of NPs complicate the nano descriptor calculation. The nano-QSAR models were developed using descriptors derived from the periodic table and some physicochemical descriptors derived from the experimental procedure as already reported in a previous study (Forest et al., 2019). Periodic table descriptors were calculated from the Elemental Descriptor Calculator Software developed in our laboratory (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>), while some additional information about the metals was added directly from the periodic table as descriptors. The selection of periodic table descriptors was advantageous because of their ease of calculation.

3.1.3. Data set division

Division of a dataset is crucial for the development of statistically significant and predictive models. For the development of MLR models, the toxicity dataset was divided into a training set (containing 18 compounds) and a test set (containing 7 compounds) with the help of a modified *k*-medoids clustering technique keeping the highest and lowest toxic data of NPs in the training set. The training set was utilized for model development whereas the test set was used to assess the predictivity of the MLR models. Note that the partial least squares (PLS) model was also obtained with the Small Dataset Modeler considering the whole dataset without any dataset division (Ambure et al., 2019).

3.1.4. Model development

Managing small datasets due to the non-availability of experimental data for a particular endpoint is a pivotal work for the QSAR researchers. As the existing dataset contains limited data, we have applied Small Dataset Modeler *version 1.0.0* to the whole dataset. This software performs dataset curation, “exhaustive” double cross-validation, and performs consensus predictions as optimal model selection techniques. It also involves the DCV (Double cross-validation) (as shown in **Figure 3.1**) method for model development for the small dataset without the need to divide the data into training and test sets. The inner loop used as the modeling set is constantly spit into calibration and validation sets containing *n* and *n-r* compounds respectively for deriving all possible “*k*” combinations. Based on the defined calibration and validation sets, the tool allows the management of validation set (*r*) compounds. The calibration set uses the GA-MLR (genetic algorithm-MLR) technique to develop MLR models whereas the quality of the developed MLR models is predicted with the validation set. All validation metrics (both internal and external metrics) were calculated with the help of an exhaustive double cross-validation technique for the respective models.

Further, we tried dividing the dataset using the modified *k*-medoids clustering technique Subset Selection v2.1 software was used for model development. The training set (with forty

descriptors) was used for the development of models from which the four best MLR models were selected based on the mean absolute error (MAE) criteria (Roy et al., 2016). Further, ICP (Intelligent Consensus Prediction) was performed with the four MLR models to enhance the predictability of the models (Roy et al., 2018) through intelligent selection of multiple models (**Figure 3.2**).

3.1.4.1 Double Cross-validation (DCV)

Double Cross-validation (DCV) consists of two nested loops named internal and external loops. The dataset in the external cross-validation loop is split into a training set and a test set. The training set in the internal loop is used for model selection and development; on the other hand, the test set is used for the assessment of the developed models. The training set (in the internal loop) is repeatedly divided into calibration and validation sets. The calibration data set is utilized to generate various models whereas the validation set is used to determine the error of the models. At last, the model having the least prediction error (in the internal loop) is chosen. The test dataset in the outer loop is then employed to assess the predictability of the selected model in the inner loop. This method of splitting the training set into calibration and validation obviates the bias caused by the usage of a single training set of selected composition.

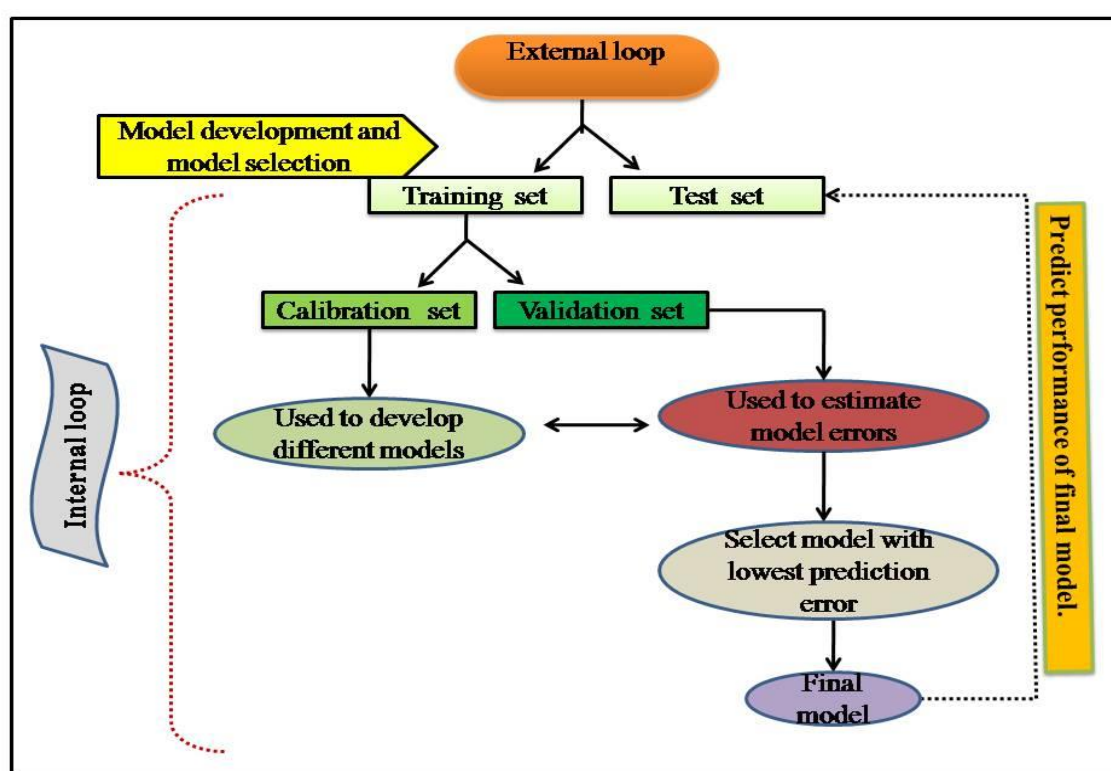


Figure 3.1 Double cross-validation methodology

3.1.4.2 ICP (Intelligent Consensus Predictions)

This software judges the performance of the consensus models and relates them with the individual MLR model-derived prediction quality based on the MAE criteria (95%). It is assumed that an individual model might not predict all individual test set compounds with the same accuracy; in other words, one QSAR model may be best for the prediction of one test set compound while another QSAR model may be best for the prediction of some other test set compound. So, here in our current study, we have selected four models (M1.1-M1.4) and performed ICP to investigate the external prediction quality of test set compounds has been improved through an “intelligent” selection of multiple MLR models. Four different ways of consensus prediction are employed they are:

- i) **Consensus model 0 (CM0):** Ordinary consensus.
- ii) **Consensus model 1 (CM1):** Average of predictions from all qualified individual models.
- iii) **Consensus model 2 (CM2):** Weighted average predictions (WAP) from all qualified individual models.
- iv) **Consensus model 3 (CM3):** Best selection of predictions (compound-wise) from individual models.

Consensus models are more reliable due to the following reasons:

1. Consensus predictions are precise and reliable as they do not rely on a single model but on multiple models and the predictions are not biased.
2. Consensus predictions may consider the WAPs from the individual models along with compound-wise prediction of the individual models thus selecting the model that is best for a particular test compound.
3. It also checks prediction by taking AD criteria, Dixon Q-test, and Euclidean distance into consideration. Checking AD before prediction makes the prediction value more acceptable.
4. For limited data points, we can rely more on consensus models than individual models as it enhances the applicability domain.

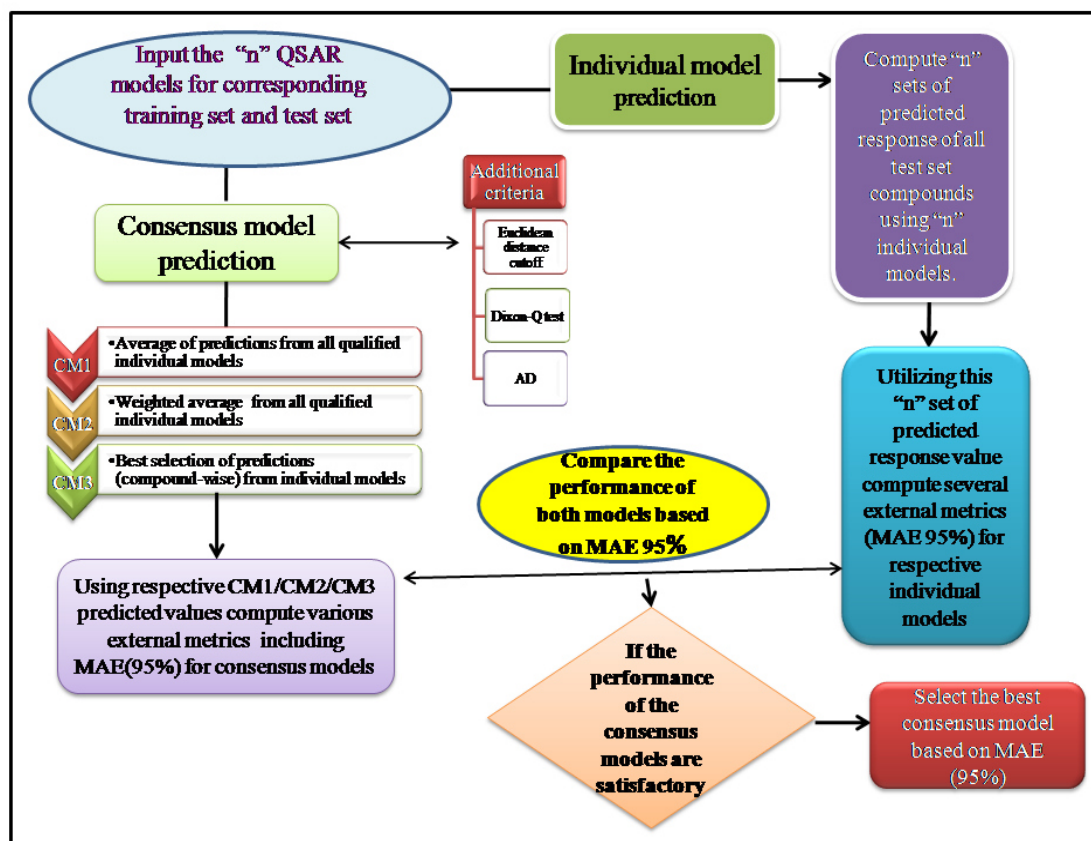


Figure 3.2 Algorithm of the ICP tool

3.1.5. Model validation

Validation of a QSAR model is crucial for nano-QSAR models. A rigorous assessment of the model quality for the measurement of stability, fitness, robustness, and predictivity of the developed models was carried out. Here, all the models were evaluated based on the OECD proposed criteria. In the present work, we have calculated R^2 , R^2_{adj} , variance ratio (F), and standard error of estimate (s) and leave-one-out squared correlation coefficient (Q^2_{LOO}) for internal model quality and validation (Chirico et al., 2011). We have also computed the leave-many-out squared correlation coefficient ($Q^2_{LMO (20\%)}$) for the final PLS model (Roy et al., 2015). Additionally, we have also calculated the r^2_m metrics and mean absolute error (MAE) (Roy et al., 2016). Further, other external validation metrics (R^2_{pred} and Q^2_{F2}) were also calculated for the test set in the case of the MLR models. The whole workflow is well depicted in Figure 3.3.

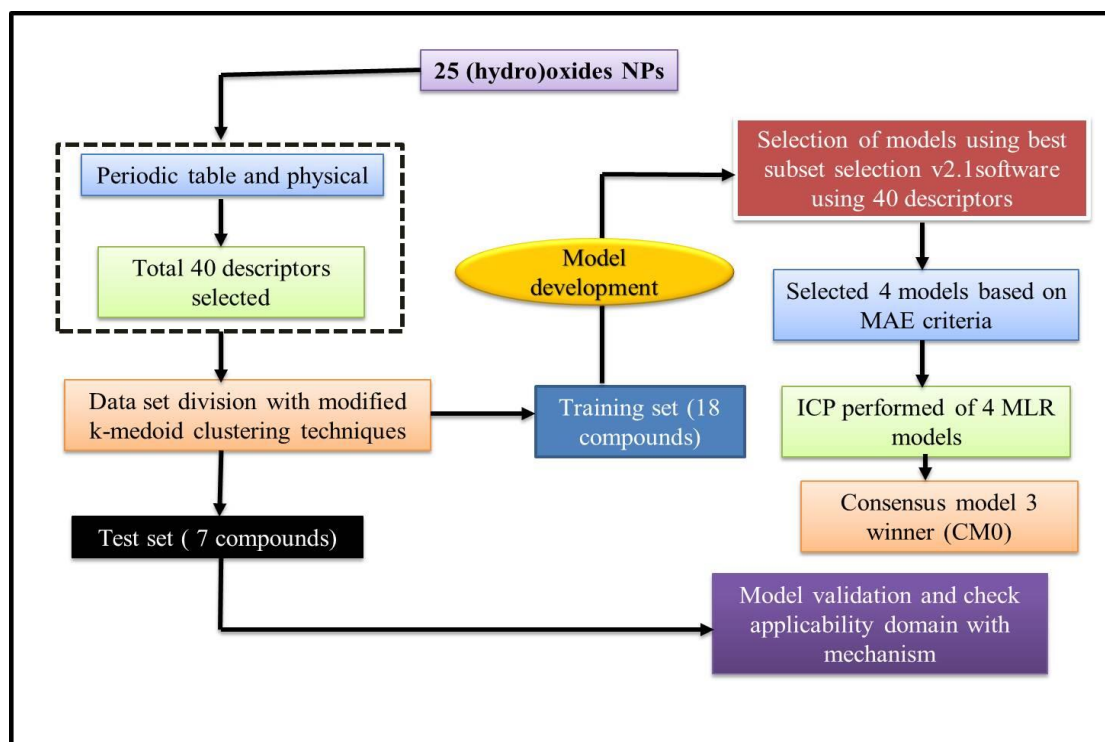


Figure 3.3 Flowchart of the work (Study 1)

3.2. Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides.

3.2.1 The dataset

The cytotoxicity data of 17 MeOxNPs to *E. coli* was obtained from the previously published research paper by Pathakoti et al. 2014, and was expressed in terms of negative logarithm of lethal concentration causing 50% growth inhibition of *E. coli* ($-\log LC_{50}$) (Table 3.2 and Table 3.3). The dataset was considered for 2D-QSAR modeling under two experimental conditions, i.e., in the dark and in the presence of light.

Table 3.2. Toxicity data of MeOxNPs to *E. coli* under dark conditions.

Sl. No	MeOxNPs	$-\log LC_{50}$
1	ZnO	5.8
2	CuO	4.24
3	Y ₂ O ₃	5.79
4	In ₂ O ₃	2.83
5	Sb ₂ O ₃	3.12
6	Al ₂ O ₃	2.42
7	Fe ₂ O ₃	2.4
8	SiO ₂	2.54
9	ZrO ₂	2.58
10	TiO ₂	2.14
11	CoO	3.13
12	Cr ₂ O ₃	2.06

13	La ₂ O ₃	4.96
14	V ₂ O ₃	3.48
15	Bi ₂ O ₃	3.55
16	SnO ₂	2.53
17	NiO	3.79

Table 3.3. Toxicity data of MeOxNPs to *E. coli* under Photo-induced conditions

Sl. No	MeOxNPs	-log LC ₅₀
1	ZnO	6.23
2	CuO	5.71
3	Y ₂ O ₃	5.84
4	In ₂ O ₃	3.48
5	Sb ₂ O ₃	3.66
6	Al ₂ O ₃	2.75
7	Fe ₂ O ₃	2.54
8	SiO ₂	2.92
9	ZrO ₂	3.04
10	TiO ₂	4.68
11	CoO	3.33
12	Cr ₂ O ₃	2.06
13	La ₂ O ₃	5.56
14	V ₂ O ₃	3.78
15	Bi ₂ O ₃	4.02
16	SnO ₂	3.24
17	NiO	3.87

3.2.2 Molecular descriptors

The dynamic behavior of MeOxNPs complicates the descriptor calculation. Therefore, descriptor calculation for metal oxide nanoparticles is a crucial job for researchers. Hence, in the present work, we have calculated fifteen 2nd generation periodic table-based descriptors (De and Roy et al. 2021) along with forty-four descriptors using elemental descriptor software freely available from our laboratory (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>). The rest periodic table descriptors were taken directly from the periodic table based on the nature of the metals. A total sixty-one periodic table descriptors were calculated for modeling of the MeOxNPs.

3.2.3 Model development: Application of the Small Dataset Modeler

Before performing QSAR model development, usually division of a dataset into training and test sets is performed. However, as the present dataset contains limited number of data points; division of the dataset is not desirable. Here, the model development has been done using the Small Dataset Modeler, version 1.0.0 software (Ambure et al. 2019) without any data set division. Small dataset modeler works on the double cross-validation (Ambure et al. 2019) strategy as illustrated in **Figure 3.4**. The double cross-validation consists of two nested loops: internal (inner) and external (outer) cross-validation loops. The external loop divides the dataset into calibration and validation sets while in the internal loop, the calibration set is

used for model development and model selection. The predictivity assessment of the models is entirely done by the validation set. The Small Dataset Modeler software (Ambure et al. 2019) avoids the dataset splitting step in the inner loop; instead, the whole dataset is used as the ‘modeling set’. The modeling set is constantly divided into calibration and validation sets and all possible combinations (‘k’) of both sets are derived. The number of compounds in the validation set is user-defined and based on it, different combinations of calibration sets and validation sets are derived. The generated calibration sets are used to develop multiple linear regression (MLR) models applying the genetic algorithm-MLR (GA-MLR) technique (Devillers, 1996; Venkatasubramanian and Sundaram 2002). On the other hand, the validation set is used to obtain the prediction quality of the respective developed MLR models. The whole workflow is illustrated in **Figure 3.5**.

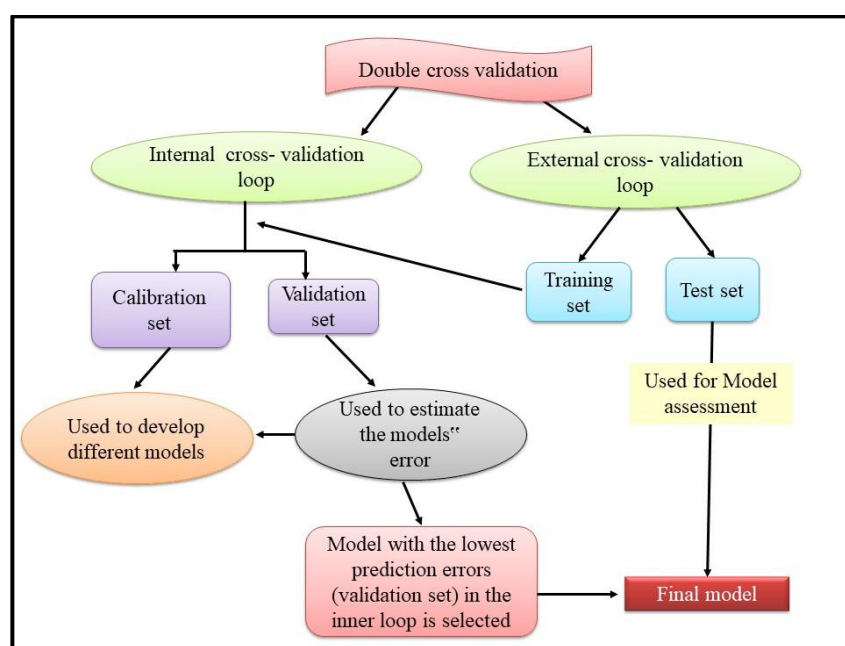


Figure 3.4 The Double Cross Validation (DCV) methodology

Additionally, internal and external validation metrics are calculated for all the selected models. The final model could be selected out of any of the five recommended methods.

1. MLR/PLS model with the least absolute MAE (95%) value in the validation set is selected.
2. MLR/PLS model with the least MAE (95%) value in the modeling set.
3. MLR/PLS model with highest $Q^2_{\text{Leave-many-out}}$ in the modeling set.
4. In the case of consensus predictions, the top-ranking model should be selected based on the least MAE (95%) values in the respective validation set. It works on two variants namely
 - a) Simple arithmetic average of prediction containing all the selected top models
 - b) Weighted average of prediction which works by assigning appropriate weight to the top selected models based on the mean absolute error gathered from leave-one-cross validation MAE_{cv} (95%).

5. A pool of unique descriptors is selected from the top three models with the least MAE (95%) of the validation set. Further, these descriptors are used for model development purposes. The best subset selection (BSS) method is used in the case of the MLR model to find the best descriptor combination, whereas in the case of PLS models, the models are formed by the descriptors selected from the top models.

The small dataset modeler software also combines data curation, exhaustive double cross-validation, along with consensus predictions in case of model development which is essential for small datasets.

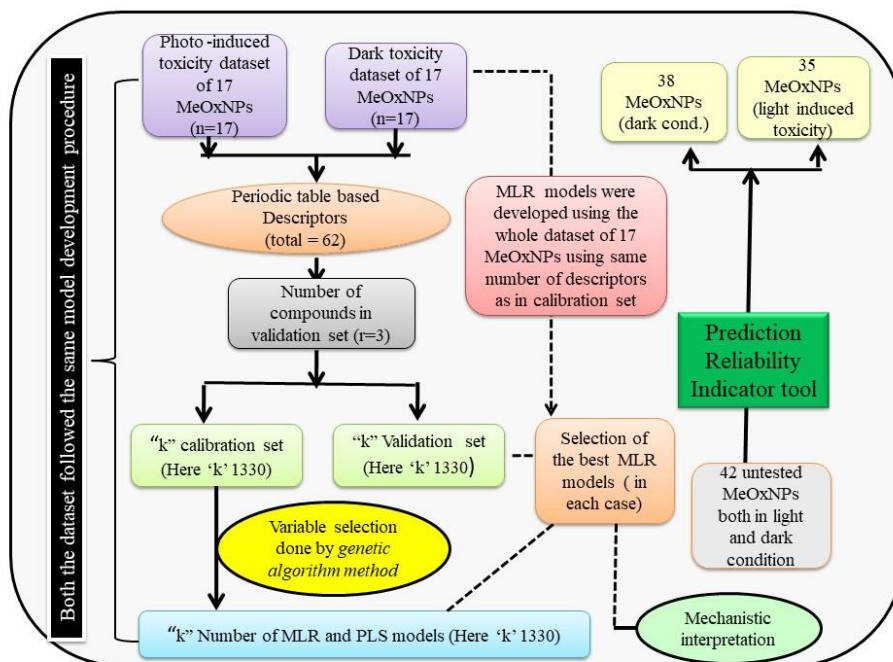


Figure 3.5. Flowchart of the approach used by the Small Dataset Modeler (Study 2)

3.2.4 Reliability of predictions of untested NPs

The estimation of checking reliability of predictions of the generated model for an external set is an interesting job when the data points are within the AD. A robust QSAR model can help in predicting the external dataset depending on the molecular features. Several tools are available for estimating the confidence of a QSAR model. Among them, one of the tools which can be relied on for the prediction of the response data of the external set is 'Prediction Reliability Indicator' tool available in <http://dtclab.webs.com/software-tools>. This tool categorizes the quality of predictions for the test set or a true external set as good, moderate and bad based on the three rules of scoring. Rule 1: scoring depending on the quality of leave-one-out prediction of the 10 closest training set compounds to the test/ external compounds, Rule 2: scoring based on the AD in terms of similarity using the standardization method (Roy et al. 2015) and the Rule 3: scoring based on the proximity of predicted value of the enquiry compound to the experimental mean of the training set responses. Thus, in the present study we have collected 42 MeOxNPs as external set (Table 3.4) to evaluate the predictive power of the developed MLR models and found that out of 42 MeOxNPs, 38 MeOxNPs showed good predictions with the dark condition toxicity data model and 35

MeOxNPs showed good predictions with the photo-induced toxicity data model. We were able to classify the untested metal oxides into toxic and nontoxic categories based on the mean toxicity value of the training set as the threshold **Figure 3.6**.

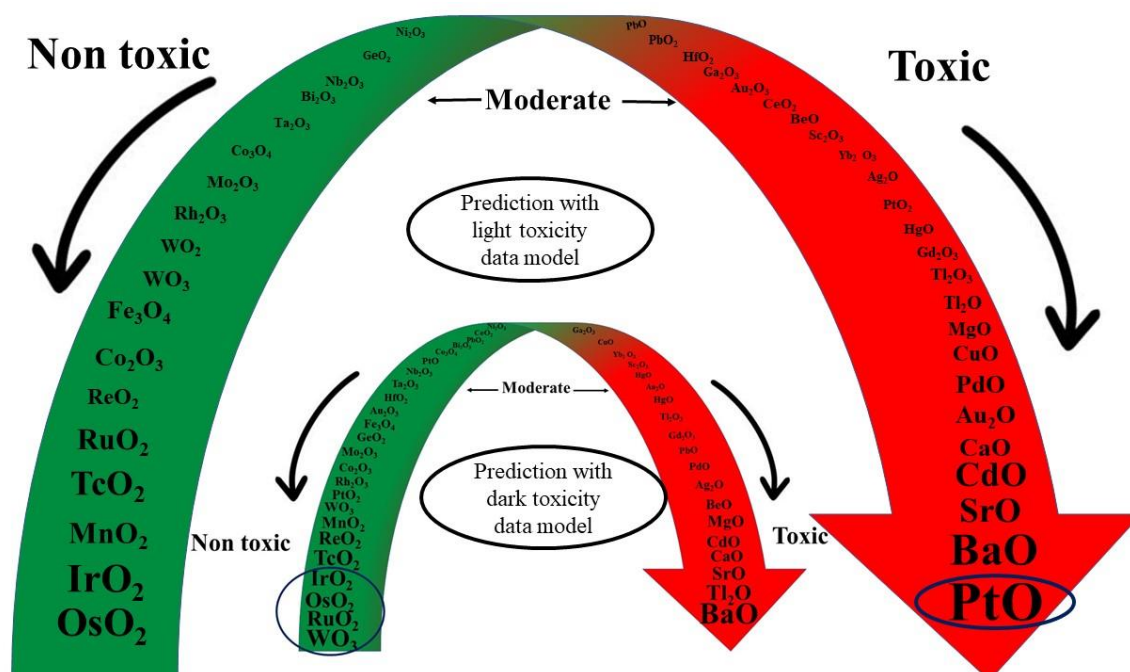


Figure 3.6. Prediction of the untested NPs with the developed MLR models shown in the increasing order of toxicity. The NPs within the circle represent the NPs which were not confidently predicted by the MLR models as per the “PRI tool”.

Table 3.4: Predicted toxicity values for 42 External dataset using light and dark toxicity MLR models.

Sl. No	Metal oxide	Valence	χ_{ox}	sq_sum_epsilon/N	atomic rad	light tox (1/log LC ₅₀) (predicted)	Dark tox (1/log LC ₅₀) (predicted)
1	Ag ₂ O	4	1	10.51921	125	5.070574	4.917
2	Au ₂ O	5	1	21.2521	135	5.933781	4.45
3	Au ₂ O ₃	5	3	7.650756	135	4.111201	2.774
4	BaO	2	2	2.772225	215	6.832478	6.553
5	BeO	2	2	0.286225	105	4.849354	5.013
6	Bi ₂ O ₃	5	3	0.0009	160	3.461121	3.124
7	CaO	2	2	0.216225	180	5.964974	6.063
8	CdO	2	2	3.3856	155	6.01467	5.713
9	CeO ₂	4	4	1.384544	185	4.746529	3.243
10	Co ₂ O ₃	6	3	0.17584	135	2.384563	2.167
11	Co ₃ O ₄	5	2.666667	0.089715	135	3.098022	3.053333
12	CuO	4	2	14.301	135	5.727334	4.219
13	Fe ₃ O ₄	6	2.666667	0.076018	140	2.446186	2.516333
14	Ga ₂ O ₃	3	3	0.02197	130	4.463944	3.918
15	Gd ₂ O ₃	3	3	0.682276	180	5.302425	4.618
16	GeO ₂	4	4	0.012594	125	3.662688	2.403
17	HfO ₂	4	4	2.700544	155	4.472873	2.823
18	HgO	4	2	8.208225	150	5.135902	4.429
19	IrO ₂	8	4	3.2761	135	1.349997	0.115
20	MgO	2	2	0.046225	150	5.492194	5.643
21	MnO ₂	7	4	0.345483	140	1.757295	0.792
22	Mo ₂ O ₃	6	3	2.876416	145	2.89644	2.307
23	Nb ₂ O ₃	5	3	2.709316	145	3.599048	2.914
24	Ni ₂ O ₃	4	3	0.204907	135	3.838458	3.381
25	OsO ₂	8	4	3.156544	130	1.258977	0.045
26	PbO	4	2	0.378225	180	4.536682	4.849
27	PbO ₂	4	4	0.1681	180	4.508525	3.173
28	PdO	4	2	10.5012	180	5.893161	4.849
29	PtO	6	2	46.44423	135	8.584526	3.005
30	PtO ₂	6	4	20.64188	135	5.127012	1.329
31	ReO ₂	7	4	3.039211	135	2.043254	0.722
32	Rh ₂ O ₃	6	3	3.407716	135	2.817634	2.167
33	RuO ₂	8	4	8.960044	130	2.036646	0.045
34	Sc ₂ O ₃	3	3	0.048107	160	4.917446	4.338
35	SrO	2	2	1.476225	200	6.433814	6.343
36	Ta ₂ O ₃	5	3	1.012036	145	3.371613	2.914
37	TcO ₂	7	4	1.037003	135	1.774958	0.722
38	Tl ₂ O	3	1	0.889878	190	5.480244	6.434
39	Tl ₂ O ₃	3	3	0.320356	190	5.403928	4.758
40	WO ₂	6	4	2.9241	135	2.752829	1.329
41	WO ₃	6	6	1.644806	135	2.581404	-0.347
42	Yb ₂ O ₃	3	3	0.894916	155	4.955919	4.268

3.2.5 Statistical validation metrics

Validation of the developed models is an important part of any QSAR study. Multiple approaches can be applied to the developed MLR models to rigorously measure and validate their fitness, stability, robustness, and predictivity. In the present study, we have computed statistical quality metrics like determination coefficient (R^2) and variance ratio (F) and validation metrics like leave-one-out determination coefficient (Q^2_{LOO}) and predicted residual sum of squares (PRESS). Additionally, r_m^2 metrics (Roy et al., 2012), root mean square error (RMSE), and mean absolute error (MAE) were also calculated (Roy et al., 2016).

3.3. Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.

3.3.1 Dataset

In the present work, two different cytotoxicity datasets of metal oxides were employed for performing nano-QSAR modeling towards the *HaCat* cell line presented as **Case Study 3.1** and **Case Study 3.2**.

Case study 3. 1: The first dataset comprises 18 nano-metal oxides taken from a previously published paper (Gajewicz et al., 2015). Their cytotoxicity is expressed as logarithms of the inverse molar lethal concentration i.e., $\log(1/LC_{50})$. Here, we have removed two MeOx -NPs because of their outlier behavior in the initial analysis **Table 3.5**. Further, another set of data was used for modeling in Case Study 2 to see the influence of dose on the cytotoxicity endpoint.

Case study 3. 2: The second dataset consists of 15 MeOx NPs (total 117 data points), taken from the Safe and Sustainable Nanotechnology (S2NANO) database built by a research group of North Korea (Choi et al., 2019). This database deals with information of various nanomaterials toxicity that is gathered from different experimental reports or literature. The cytotoxicity was expressed in the percentage (%) of cell death in **Table 3.6**. The *HaCat cell* was exposed to NPs with various physicochemical properties such as core size, hydrodynamic size, surface charge, and dose ranging from 0-1500 ppm.

Table 3.5. Cytotoxicity data of 18 nano-metal oxides (**Case study 3. 1**).

CompNo	Metal oxides (MeOx) NPs	$\log(1/LC_{50})$
*1	ZnO	3.32
2	TiO ₂	1.76
3	CoO	2.83
4	In ₂ O ₃	2.92
*5	NiO	2.49
6	V ₂ O ₃	2.24
*7	Y ₂ O ₃	2.21
8	Al ₂ O ₃	1.85
9	Bi ₂ O ₃	2.5
*10	La ₂ O ₃	2.87
11	Sb ₂ O ₃	2.31
12	SiO ₂	2.12
13	SnO ₂	2.67
14	ZrO ₂	2.02
15	Cr ₂ O ₃	2.3
*16	Fe ₂ O ₃	2.05

17	WO ₃	2.56
18	Mn ₂ O ₃	2.64

Table 3.6. Cell viability data of 15 Me_{ox} NPs taken from the Safe and Sustainable Nanotechnology (S2NANO) database (**Case study 3. 2**).

No.	Material	Core Size	Hydrodynamic Size	Surface Charge	Dose	% Cell death
1.	Al ₂ O ₃	44	372.3	-20.2	50	5.1
2.	Al ₂ O ₃	44	372.3	-20.2	100	18.4
3.	Al ₂ O ₃	44	372.3	-20.2	150	20.6
4.	Al ₂ O ₃	44	372.3	-20.2	200	26.5
5.	Al ₂ O ₃	44	372.3	-20.2	250	28
6.	Al ₂ O ₃	44	372.3	-20.2	500	30.5
7.	Al ₂ O ₃	44	372.3	-20.2	750	35.5
8.	Al ₂ O ₃	44	372.3	-20.2	1000	47.6
9.	Al ₂ O ₃	44	372.3	-20.2	1500	50.1
10. *	Bi ₂ O ₃	90	2029	-2.3	50	8.8
11.	Bi ₂ O ₃	90	2029	-2.3	100	10.7
12.	Bi ₂ O ₃	90	2029	-2.3	150	14.7
13.	Bi ₂ O ₃	90	2029	-2.3	200	22.8
14. *	Bi ₂ O ₃	90	2029	-2.3	250	30.8
15.	Bi ₂ O ₃	90	2029	-2.3	500	37
16.	Bi ₂ O ₃	90	2029	-2.3	750	36.7
17.	Bi ₂ O ₃	90	2029	-2.3	1000	41.7
18. *	Bi ₂ O ₃	90	2029	-2.3	1500	50.1
19.	Cr ₂ O ₃	60	616.8	5.7	50	4.8
20.	Cr ₂ O ₃	60	616.8	5.7	100	6.6
21.	Cr ₂ O ₃	60	616.8	5.7	150	10.4
22. *	Cr ₂ O ₃	60	616.8	5.7	200	13.2
23.	Cr ₂ O ₃	60	616.8	5.7	250	27.1
24. *	Cr ₂ O ₃	60	616.8	5.7	500	29.6
25.	Cr ₂ O ₃	60	616.8	5.7	750	42
26.	Cr ₂ O ₃	60	616.8	5.7	1000	71.2
27.	Cr ₂ O ₃	60	616.8	5.7	1500	76.8
28.	Fe ₂ O ₃	32	297.6	-18.1	50	7.6
29.	Fe ₂ O ₃	32	297.6	-18.1	100	11.6
30.	Fe ₂ O ₃	32	297.6	-18.1	150	13.2
31.	Fe ₂ O ₃	32	297.6	-18.1	200	16.9
32. *	Fe ₂ O ₃	32	297.6	-18.1	250	19
33.	Fe ₂ O ₃	32	297.6	-18.1	500	21.8
34.	Fe ₂ O ₃	32	297.6	-18.1	750	23.4
35. *	Fe ₂ O ₃	32	297.6	-18.1	1000	26.8
36.	Fe ₂ O ₃	32	297.6	-18.1	1500	50.7
37.	In ₂ O ₃	29.8	224.3	-9.6	50	-2
38.	In ₂ O ₃	29.8	224.3	-9.6	100	6.6

39.	In ₂ O ₃	29.8	224.3	-9.6	250	41.4
40.	In ₂ O ₃	29.8	224.3	-9.6	500	63.4
41. *	In ₂ O ₃	29.8	224.3	-9.6	750	68.7
42. *	In ₂ O ₃	29.8	224.3	-9.6	1000	71.2
43.	In ₂ O ₃	29.8	224.3	-9.6	1500	72.7
44.	La ₂ O ₃	45.6	672.9	-12.8	50	2
45.	La ₂ O ₃	45.6	672.9	-12.8	100	0.7
46.	La ₂ O ₃	45.6	672.9	-12.8	250	48.2
47. *	La ₂ O ₃	45.6	672.9	-12.8	500	72.4
48.	La ₂ O ₃	45.6	672.9	-12.8	750	65
49. *	La ₂ O ₃	45.6	672.9	-12.8	1000	65
50. *	La ₂ O ₃	45.6	672.9	-12.8	1500	80.8
51.	Mn ₂ O ₃	29.8	291.1	-3	50	9.1
52.	Mn ₂ O ₃	29.8	291.1	-3	100	17.5
53. *	Mn ₂ O ₃	29.8	291.1	-3	150	29.6
54.	Mn ₂ O ₃	29.8	291.1	-3	200	41.1
55.	Mn ₂ O ₃	29.8	291.1	-3	250	44.5
56. *	Mn ₂ O ₃	29.8	291.1	-3	500	49.1
57.	Mn ₂ O ₃	29.8	291.1	-3	750	60.9
58.	Mn ₂ O ₃	29.8	291.1	-3	1000	69.6
59.	Mn ₂ O ₃	29.8	291.1	-3	1500	77
60.	NiO	20	223.5	-12.1	50	-1
61.	NiO	20	223.5	-12.1	100	-1
62.	NiO	20	223.5	-12.1	250	46.3
63. *	NiO	20	223.5	-12.1	500	81.4
64.	NiO	20	223.5	-12.1	750	83.9
65. *	NiO	20	223.5	-12.1	1000	84.8
66.	NiO	20	223.5	-12.1	1500	86.4
67. *	Sb ₂ O ₃	150	640.3	-13.3	50	6.9
68.	Sb ₂ O ₃	150	640.3	-13.3	100	11.6
69.	Sb ₂ O ₃	150	640.3	-13.3	150	15.3
70.	Sb ₂ O ₃	150	640.3	-13.3	200	23.1
71. *	Sb ₂ O ₃	150	640.3	-13.3	250	26.2
72.	Sb ₂ O ₃	150	640.3	-13.3	500	32.1
73.	Sb ₂ O ₃	150	640.3	-13.3	750	34.6
74.	Sb ₂ O ₃	150	640.3	-13.3	1000	42.3
75.	Sb ₂ O ₃	150	640.3	-13.3	1500	53.5
76.	SnO ₂	46.1	264.9	-10.5	50	8.2
77.	SnO ₂	46.1	264.9	-10.5	100	13.2
78. *	SnO ₂	46.1	264.9	-10.5	250	50.1
79.	SnO ₂	46.1	264.9	-10.5	500	56
80.	SnO ₂	46.1	264.9	-10.5	750	65.6
81. *	SnO ₂	46.1	264.9	-10.5	1000	70.8
82.	SnO ₂	46.1	264.9	-10.5	1500	82
83.	TiO ₂	42.3	1307	-9.6	50	20.3
84.	TiO ₂	42.3	1307	-9.6	100	17.5
85.	TiO ₂	42.3	1307	-9.6	250	33.9
86.	TiO ₂	42.3	1307	-9.6	500	34.2
87.	TiO ₂	42.3	1307	-9.6	750	38.6
88. *	TiO ₂	42.3	1307	-9.6	1000	39.2

89.	TiO ₂	42.3	1307	-9.6	1500	59.7
90.	WO ₃	50	179.6	-9.1	50	-6
91.	WO ₃	50	179.6	-9.1	100	-4
92.	WO ₃	50	179.6	-9.1	250	33.9
93.*	WO ₃	50	179.6	-9.1	500	47.9
94.	WO ₃	50	179.6	-9.1	750	48.5
95.*	WO ₃	50	179.6	-9.1	1000	50.7
96.	WO ₃	50	179.6	-9.1	1500	61.5
97.	Y ₂ O ₃	38	1222.9	-10.7	50	-1
98.	Y ₂ O ₃	38	1222.9	-10.7	100	3.5
99.	Y ₂ O ₃	38	1222.9	-10.7	250	31.8
100.	Y ₂ O ₃	38	1222.9	-10.7	500	37.7
101.	Y ₂ O ₃	38	1222.9	-10.7	750	42
102.	Y ₂ O ₃	38	1222.9	-10.7	1000	42.6
103.	Y ₂ O ₃	38	1222.9	-10.7	1500	54.4
104.	ZnO	71	188.9	-10.8	50	66.2
105.	ZnO	71	188.9	-10.8	100	73.3
106.	ZnO	71	188.9	-10.8	250	87.9
107.	ZnO	71	188.9	-10.8	500	92.6
108.	ZnO	71	188.9	-10.8	750	98.1
109.	ZnO	71	188.9	-10.8	1000	99.07
110.	ZnO	71	188.9	-10.8	1500	100
111.	ZrO ₂	46.7	661.4	-8.5	50	2.9
112.	ZrO ₂	46.7	661.4	-8.5	100	6.9
113.	ZrO ₂	46.7	661.4	-8.5	250	26.5
114.	ZrO ₂	46.7	661.4	-8.5	500	31.8
115.	ZrO ₂	46.7	661.4	-8.5	750	33.9
116.	ZrO ₂	46.7	661.4	-8.5	1000	51.6
117.	ZrO ₂	46.7	661.4	-8.5	1500	57.5

*Test set

3.3.2 Molecular Descriptors

Nano-QSAR modeling was performed using the basic information of the metal oxides taken from the periodic table to encode the cytotoxicity and percentage cell death against the *HaCat* cell line. For **Case Study 3. 1**, a total of 69 descriptors were calculated using the basic information of the metal oxides from the periodic table directly, and other descriptors (2nd generation periodic table descriptors) were calculated from the primary descriptors (De et al., 2018). Additionally, a few periodic table descriptors were calculated with the help of Elemental Descriptor Calculator software available from (<https://dtclab.webs.com/software-tools>). For **Case Study 3. 2**, a total of 73 descriptors were calculated considering periodic table-based descriptors as well as some physicochemical descriptors taken from the experimental conditions (Choi et al., 2019). The advantage of using periodic table-based descriptors over other descriptors is that they can be easily calculated without using any sophisticated software.

3.3.3 Data set splitting

Selecting a proper training set and a test set is vital for the development of a robust QSAR model. The combination of metal oxides is different in each data point (Case Study 2). So, in the present study, after pretreatment and removal of noisy descriptors from both datasets, the dataset for each case was divided with dataset division software freely available from the DTC Lab site (<https://dtclab.webs.com/software-tools>). In **Case Study 3. 1**: Eleven (11) metal oxides were kept in the training set and five (5) metal oxides were present in the test set. On the other hand, in **Case Study 3. 2**, one hundred and six (94) metal oxide data points were present in the training set and twenty-six (23) metal oxide data points were present in the test set for model development and validation respectively.

3.3.4 Chemometric tools

Model development is a crucial job for the QSAR modeler. Therefore, after data pretreatment, the datasets were subjected to features selection by genetic algorithm (GA) (Venkatasubramanian, and Sundaram, 2002) followed by model development by Best Subset Selection (BSS) software (<https://dtclab.webs.com/software-tools>) based on the MAE criteria (Roy et al., 2016). The dataset containing sixteen metal oxides having cytotoxicity data towards *HaCat* cell line used periodic table descriptors and the final model with the best four descriptors combination was selected. In the second case, along with the periodic table-based descriptors, some physicochemical descriptors were also included. Partial Least Squares (PLS) regression was performed using four descriptors obtained in the MLR model. The PLS model was selected as the final model with two latent variables. The whole workflow is represented in **Figure 3.7**.

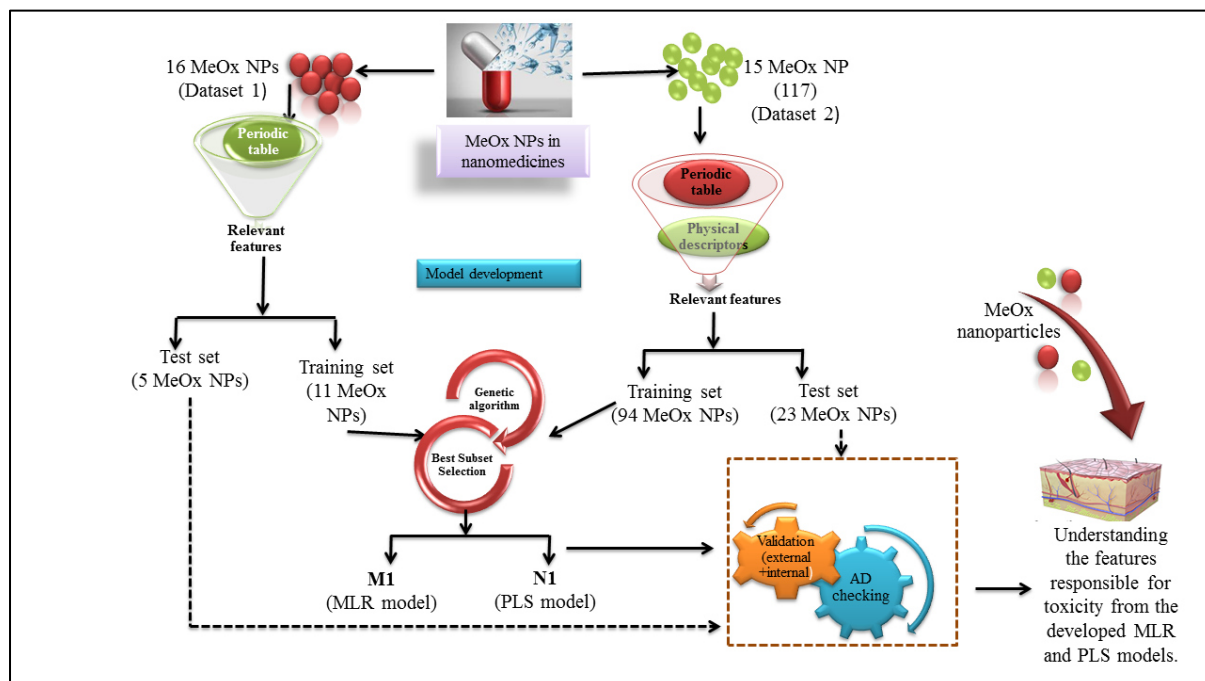


Figure 3.7. Flowchart of the work (Study 3).

3.3.5 Validation metrics

The validation of the developed models was done by both internal and external validation metrics. The internal quality metric R^2 determines the fitting potential of the models whereas Q_{Loo}^2 metric (Roy et al., 2012) measures the predictive ability of the model based on the training set compounds. Although, Q_{Loo}^2 metric measures the robustness of the model, it may not be sufficient enough to determine the performance of the model on a new set of compounds. The external validation metrics (Q_{F1}^2 , Q_{F2}^2) deal with the quality of the predictions for the test set compounds. Additionally, r_m^2 metrics and root mean squared error (RMSE) were also computed. The accepted threshold value for both internal and external metrics is 0.5.

3.3.6. Applicability domain (AD)

The applicability domain (AD) in QSAR stands for physicochemical, biological, and chemical space knowledge of training set based on which the model has been developed. The model is capable of predicting new compounds when they belong to the same chemical domain (Roy et al., 2015). The purpose of AD is to confirm the model assumptions. To investigate AD of the training set, one can analyze the features of the multivariate descriptor information of the training set compounds or can analyze similarity-based metrics. When a compound is highly dissimilar to the training set compounds, the reliability of the particular compound is unlikely. The concept of AD was used in the final model to avoid the unjustified extrapolation of prediction. In this study, we have checked the applicability domain for the MLR model with the leverage approach (Gramatica, 2007 ; Eriksson et al., 2003) using STATISTICA software (<http://www.statsoft.com/>), and for the PLS model, we have performed DModX (distance to model) approach (Wold, 2001) using SIMCA-P software (info@umetrics.com; www.umetrics.com) at 99% confidence level. Through the leverage approach, it is possible to determine whether a compound is within the chemical space or outside. If the leverage values (h) of the compounds are less than the critical value (h^*), then the prediction can be considered reliable and vice versa. The critical value (h^*), calculated as $h^* = 3p'/n$, where p' stands for the modeled variables plus one and 'n' stands for the number of compounds used in model development. We can easily detect the outlier from plotting leverages (Hat diagonal) against cross-validated standardized residuals (R), i.e (William's plot). Compounds with cross-validation standardized residual greater than three standard deviation units, i.e., $>2.5\sigma$ can be considered as outlier

3.4. Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells.

3.4.1. The dataset

The cytotoxicity data on RAW 264.7 and BEAS-2B cell lines were taken from previously published literature (Zhang et al., 2012). The cytotoxicity data were obtained from two different assay methods (i.e., MTS and ATP) for a single cell line. A total of four datasets were used for modeling purposes with the half-maximal effective concentration in the logarithmic ($\log(1/EC_{50})$) scale as the endpoint. Here, four datasets containing 9 (**Dataset 4. 1**) (**Table 3.7**), 10 (**Dataset 4. 2**) (**Table 3.8**), 10 (**Dataset 4. 3**) (**Table 3.9**), and 12 (**Dataset 4. 4**) (**Table 3.10**) MeOx ENPs respectively were taken for modeling purposes. Merging compounds from two datasets (ATP assay data) out of the four datasets from RAW 264.7 and BEAS-2B cell lines, common eight MeOx ENPs were used to develop the nano-quantitative

toxicity-toxicity relationship (nano-QTTR) model. The same four datasets were also further utilized for predictions using the similarity-based approach.

Table 3.7. Dataset 4. 1: Log (1/EC₅₀) in BEAS-2B cell (ATP assay method).

ID	Metal oxide	log (1/EC ₅₀)
1	Co ₃ O ₄	-1.37
2	Cr ₂ O ₃	-1.55
3	Ni ₂ O ₃	-1.8
4	CuO	-1.38
5	Mn ₂ O ₃	-1.12
6	CoO	-1.64
7	ZnO	-0.56
8	La ₂ O ₃	-19.26
9	Sb ₂ O ₃	-18.12

Table 3.8. Dataset 4. 2: Log (1/EC₅₀) in BEAS-2B cell (MTS assay method).

ID	Metal oxide	log (1/EC ₅₀)
1	Co ₃ O ₄	-3.71
2	Cr ₂ O ₃	-2.53
3	Ni ₂ O ₃	-2.52
4	CuO	-1.38
5	Mn ₂ O ₃	-1.05
6	CoO	-1.71
7	La ₂ O ₃	-56.34
8	Sb ₂ O ₃	-12.12
9	In ₂ O ₃	-5.62
10	Y ₂ O ₃	-15.42

Table 3.9. Dataset 4.3: Log (1/EC₅₀) in RAW 264.7 cell (MTS assay method).

ID	Metal oxide	log (1/EC ₅₀)
1	Co ₃ O ₄	-2.59
2	Cr ₂ O ₃	-2.67
3	Ni ₂ O ₃	-2.5
4	CuO	-1.12
5	Mn ₂ O ₃	-1.01
6	CoO	-1.8
7	ZnO	-1.24
8	HfO ₂	-16.75
9	NiO	-2.18
10	Yb ₂ O ₃	-20.72

Table 3.10. Dataset 4.4: Log (1/EC₅₀) in RAW 264.7 cell (ATP assay method).

ID	Metal oxide	log (1/EC ₅₀)
1	Co ₃ O ₄	-1.35
2	Cr ₂ O ₃	-1.07
3	Ni ₂ O ₃	-1.94
4	CuO	-1.16
5	Mn ₂ O ₃	-1.1
6	CoO	-1.68
7	ZnO	-0.91
8	Gd ₂ O ₃	-10.02
9	HfO ₂	-9.45
10	SnO ₂	-5
11	Y ₂ O ₃	-2.14
12	La ₂ O ₃	-8.52

3.4.2. Molecular Descriptors

Molecular descriptors are used to encode the properties of MeOx ENPs. The complex behavior of metal oxides complicates their descriptor calculations making it a difficult job for the QSAR researchers. Therefore, in the present study, fifty-eight periodic table descriptors were calculated which include twenty-three 2nd generation periodic table descriptors calculated from the primary descriptors (De et al., 2018), and thirty descriptors were taken from the Elemental Descriptor Calculator software available from (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>). Additionally, some basic information about the metals was taken from the periodic table to encode the cytotoxicity of RAW 264.7 and BEAS-2B cell lines.

3.4.3. Small Dataset Modeler for feature selection

The choice of appropriate structural characteristics that can quantitatively describe the variation of the MeOx ENPs' structures is essential for computational modeling. This is because the general concept of the read-across method is based on the assumption that similar compounds will bear similar responses to toxicity. Hence, to identify the structural features that can relate to the toxic features of MeOx NPs to RAW 264.7 and BEAS-2B cell lines, the Small Dataset Modeler software version 1.0.0 (Ambure et al., 2019) was utilized. The existing dataset contains limited data for modeling, and handling datasets with limited data points (less than 20) is problematic for statistical model development as lower degrees of freedom lead to statistically less interesting and unreliable models. The small dataset modeler software (http://teqip.jdvu.ac.in/QSAR_Tools/) can handle the situation as it works with the whole data set without training-test splitting and cross-validates the models using the double cross-validation technique. This software performs dataset curation, "exhaustive" double cross-validation, and performs consensus predictions as optimal model selection techniques. In the Double Cross-Validation (DCV) strategy, the internal loop (training set) is divided into a calibration set and a validation set where the calibration set is used for model development and model selection, and predictivity is entirely checked by the validation set. In the case of Small Dataset Modeler software (Ambure et al., 2019), the whole dataset is used as modeling set in the inner loop (avoiding the data splitting step in the inner loop). The modeling set is

constantly divided into all possible ‘k’ combinations of calibration and validation sets. The developed calibration set is used to develop multiple linear regressions (MLR) or partial least squares models (PLS) using the genetic algorithm-MLR (GA-MLR) technique (Devillers, 1996; Venkatasubramanian and Sundaram, 2002). Again, the validation set is used to derive the prediction quality of the respective developed models. The results obtained from the PLS models of BEAS-2B cell and RAW 264.7 cell datasets were found to be statistically significant. The descriptors obtained from the four PLS models with significant outcomes were used in the read-across technique.

3.4.4. Read-Across based predictions

After selecting appropriate descriptors from the four PLS models, a machine learning approach based on similarity measures was performed. This approach uses the Quantitative Read-Across v4.0 tool available from our laboratory website (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home>). The Read-across approach uses Euclidean distance, Gaussian kernel function, and Laplacian kernel function-based similarity techniques. This method requires optimization of hyper-parameters, i.e., sigma and gamma values; distance, and similarity thresholds. The best result obtained with the specific hyperparameter was selected as the final set to get the desired outcome. Each of all four datasets was divided into a training set (source compounds) and a test set (target compounds). Since the dataset was small (less than 20), conventional dataset division, i.e. 70% data points in the training set and 30% in the test set does not appear to be satisfactory. In such a case, the number of test set compounds would be very limited to check the quality of predictions. For read-across predictions, “many” source compounds are not usually required as the limited number of close source compounds would be sufficient to make efficient predictions. Therefore, 50% division was opted (Gajewicz, 2017), and the division was done randomly many times to further avoid any bias during modeling. The read-across method is well illustrated in **Figure 3.8**.

3.4.5. Applicability domain

The applicability domain (AD) in QSAR (Gadaleta et al., 2016) is the physicochemical, structural, and biochemical space knowledge of the training set compounds based on which the model has been developed. Prediction of the new compounds is possible when the new entity falls in the same chemical domain of the modeled descriptors. AD confirms the model assumption for which type of compounds the model is reliably applicable. Therefore, in this study, the AD assessment for the developed PLS models was done by the DModX approach, i.e., distance to model in the X-space with SIMCA 16.0.2 software. The AD was checked at a 99% confidence level ($D\text{-crit} = 0.009998$).

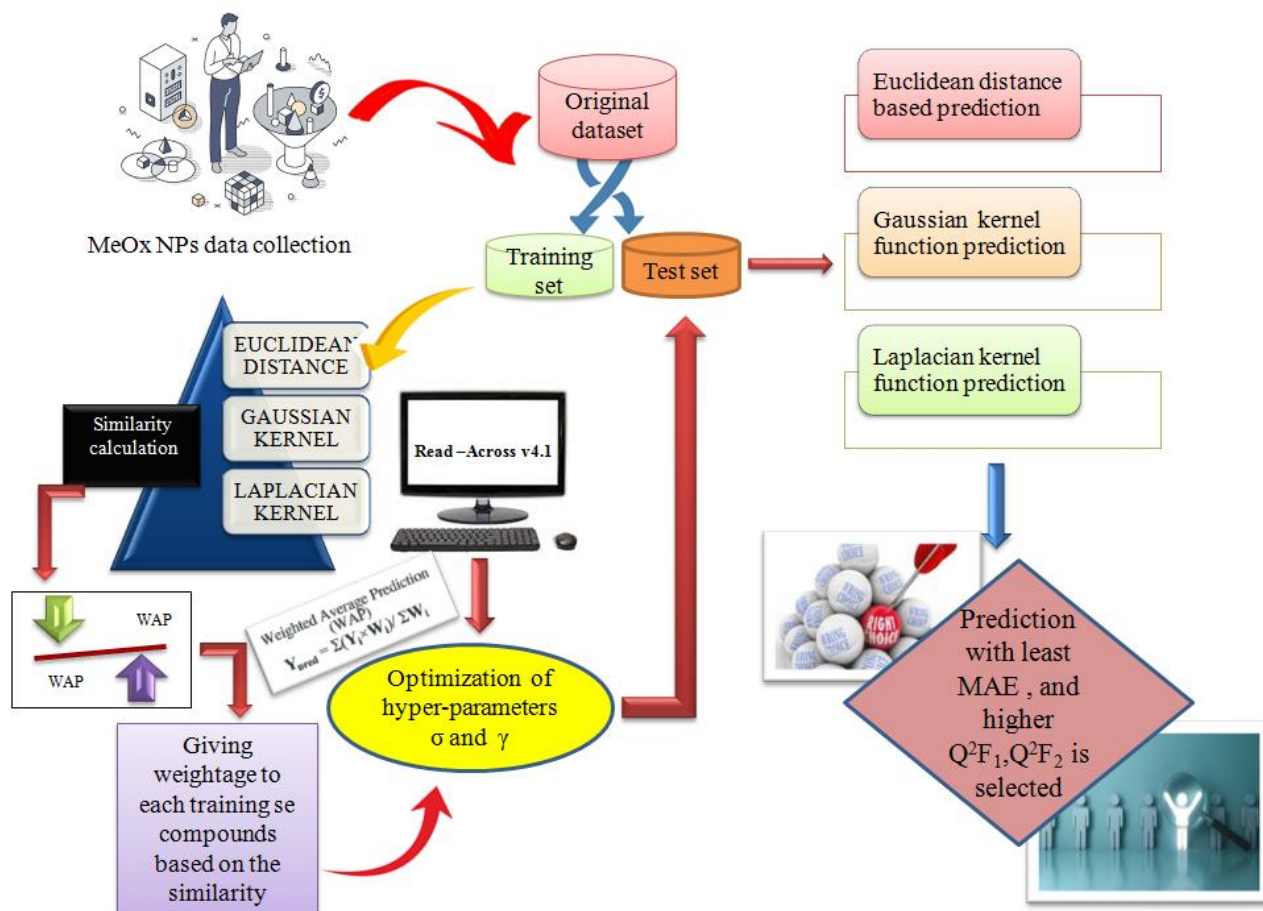


Figure 3.8. The Read-across flow diagram (Study 4).

3.4.6. Nano-QTTR model development

The common eight compounds were taken from the ATP assay of both BEAS-2B cell and RAW 264.7 cell datasets with cytotoxicity values ($EC_{50}\%$), for model development. It is interesting to note that initially without any chemical descriptors, just simple linear regression was performed between the cytotoxicity data of both BEAS-2B cells and RAW 264.7 cells. The encouraging results from the MLR model inspired us to develop the nano-QTTR model. Subsequently, identification of the most relevant descriptors from the pool of fifty-eight descriptors was done employing the stepwise regression technique (stepwise-MLR) (Darlington, 1990). Along with the selected descriptors, the cytotoxicity of one of the cells was used as the predictor variable to predict the cytotoxicity of the other cell using the partial least squares (PLS) method (Wold, 1995) (**Figure 3.9**). As we know that finding a large number of data points relevant to MeOx ENPs is difficult, therefore QTTR models may give us the insight to predict MeOx ENPs' toxicity to various species despite a limited number of available data points and help in data gap filling. The experimental values of the eight metals were available for the ATP assay methods (Dataset 4.1 and Dataset 4.4) and not for the MTS assay methods (Dataset 4.2 and Dataset 4.3). Therefore, these were not applied in QTTR modeling.

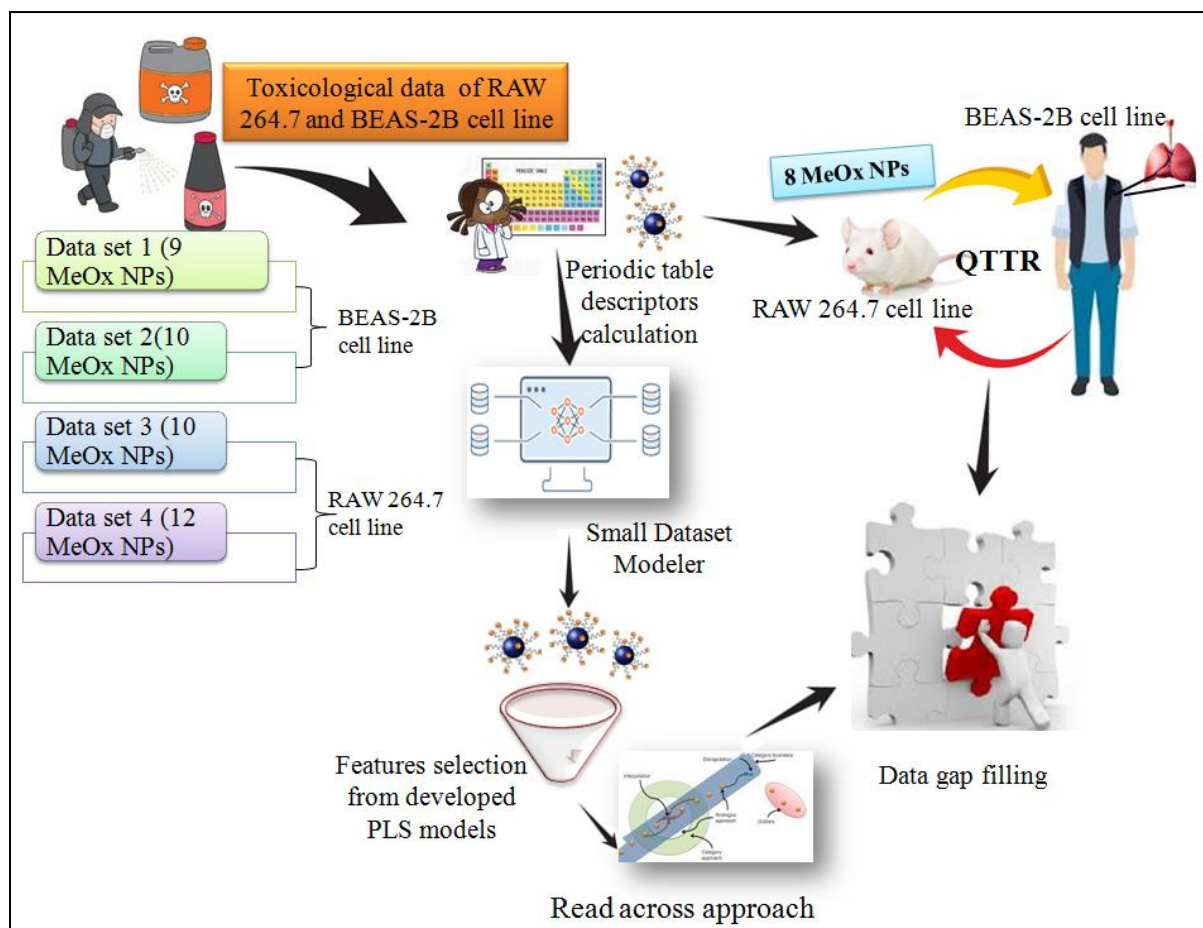


Figure 3.9. Flowchart of the present work (Study 4).

3.4.7. Statistical validation metrics

The validation of the developed model is crucial for the QSAR study. Multiple rigorous assessments were performed to determine the predictability, stability, and robustness of the developed models. All the models were evaluated as per the OECD guidelines. Here, we have computed the determination coefficient (R^2), leave-one-out determination coefficient (Q^2_{LOO}), and mean absolute error (MAE) (Roy et al., 2016). Additionally, r_m^2 metrics (Roy et al., 2012) for the final PLS models (Wold, 1995) were also calculated. Furthermore, Q^2_{F1} and Q^2_{F2} , RMSEP, and MAE were also calculated for read-across prediction.

For compliance of the developed models with the OECD guidelines for applicability in regulatory purposes, the following principles were applied:

Principle 1 (a defined endpoint): The responses being modeled in the present work for the different datasets were measured under identical conditions. Thus, the QSAR models were developed following the 1st OECD principle.

Principle 2 (an unambiguous algorithm): Various chemometric tools based on specific algorithms were employed for the calculation of the different categories of descriptors and subsequent QSAR model development using specific software packages. Thus, the model development pathway employed for the present studies follows a definite algorithm.

Principle 3 (a defined domain of applicability): The domain of applicability of all the statistically significant PLS QSAR models was analyzed using the DModX method. Thus, the selection of the best QSAR model was done in corroboration with this principle.

Principle 4 (appropriate measures of goodness-of-fit, robustness, and predictivity): All the developed models were rigorously validated using internal, external, and overall validation approaches. The quality of fitness and the predictive potential of the developed models were assessed based on the different validation metrics while the robustness of the models was judged using the randomization approach. The selection of the most significant models based on the acceptable values of the various validation metrics accounts for the compliance of the models with the 4th guideline.

Principle 5 (a mechanistic interpretation): All the descriptors appearing in the developed QSAR models could aptly define the essential structural attributes of the molecules imparting optimum endpoint values thus signifying suitable mechanistic interpretation of the developed models.

3.5. Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach.

3.5.1 Dataset

The toxicity data used for model development were obtained from the literature (Huang et al., 2020). 32 MeOxNPs were analyzed for fold change ($FC_{IL-1\beta}$), and cell viability. The doses for the cell viability dataset were 100-200 $\mu\text{g/mL}$, which resulted in 54 data points. Here, two datasets were used that is **dataset 5. 1** with 32 MeOxNPs with fold change ($FC_{IL-1\beta}$) as endpoint (i.e., **Table 3.11**) and **dataset 5. 2** with 54 MeOxNPs that have $-\ln(p/1-p)$ as its endpoint and considered 100-200 $\mu\text{g/mL}$ doses (**Table 3.12**).

We have changed the cell viability dataset endpoint to $(-\ln(p/1-p))$ for a better understanding of the toxicity (Mitteroecker and Bookstein, 2011). The release of cytokine interleukin (IL)-1beta (IL-1b) from the THP-1 cells and $(-\ln(p/1-p))$ were used as an indicator for the toxicity assessment and thus was used for further modeling purposes. The THP-1 cell line can be used in immune system disorder research, immunology research, and toxicology research.

The Fold Changes of IL-1b production (FC_{IL-1b}) were expressed as the level of MeOxNPs induced IL-1b release in THP-1 cells. It was calculated using the following equation 3.1:

$$FC_{IL-1\beta} = \frac{\text{avg}(\text{Sample})}{\text{avg}(\text{Control})} \quad (3.1)$$

Here, avg (Sample) stands for the average level of IL-1b ($\mu\text{g/mL}$) induced by specific MeOx NPs at each dose, and avg (Control) stands for the average level of IL-1b ($\mu\text{g/mL}$) for untreated THP- 1 cell lines.

Table 3.11. Dataset 5.1 MeOxNPs were analyzed for fold change ($FC_{IL-1\beta}$).

Sl. No	Metal oxide	Exp. Fold-change (FC)
1	Al ₂ O ₃	3.15
2	Co ₃ O ₄	0.44
3	Cr ₂ O ₃	0.5
4	Fe ₂ O ₃	2.83
5	TiO ₂	1.77
6	Y ₂ O ₃	26.63
7	ZrO ₂	2.34
8	Sm ₂ O ₃	46.44
9	In ₂ O ₃	3.38
10	La ₂ O ₃	38.63
11	Dy ₂ O ₃	29.69
12	Er ₂ O ₃	24.24
13	Bi ₂ O ₃	1.35
14	Ni ₂ O ₃	1.25
15	Yb ₂ O ₃	36.05
16	Gd ₂ O ₃	31.37
17	HfO ₂	1.98
18	ZnO	0.29
19	CuO	0.86
20	MnO ₂	25.04
21	Nd ₂ O ₃	25.61
22	Eu ₂ O ₃	42.96
23	Fe ₃ O ₄	3.61
24	WO ₃	0.73
25	CoO	0.31
26	SnO ₂	0.39
27	Pr ₆ O ₁₁	95.67
28	MoO ₃	1.82
29	Ta ₂ O ₅	9.96
30	Nb ₂ O ₅	2.71
31	Mn ₃ O ₄	41.88
32	MoO ₂	6.28

Table 3.12. Dataset 5.2 MeOxNPs were analyzed for $\ln(1-p)/p$.

Sl. No	Metal oxide	$\ln(1-p)/p$
1	Al ₂ O ₃ -6	-6.72423
2	CoO-6	-2.62733
3	CoO-7	-5.59572
4	Co ₃ O ₄ -7	-5.2933
5	CuO-7	-3.37769
6	Dy ₂ O ₃ -7	2.813481
7	Er ₂ O ₃ -6	3.3903
8	Er ₂ O ₃ -7	0.0016

9	Eu ₂ O ₃ -7	1.228181
10	Fe ₂ O ₃ -6	-0.65484
11	Fe ₂ O ₃ -7	0.145858
12	Fe ₃ O ₄ -7	1.394439
13	Gd ₂ O ₃ -6	6.906755
14	Gd ₂ O ₃ -7	-2.95714
15	Co ₃ O ₄ -6	-2.16433
16	MnO ₂ -6	-2.96569
17	MnO ₂ -7	-2.16866
18	Mn ₂ O ₃ -6	-3.12222
19	Mn ₂ O ₃ -7	-1.86003
20	Nd ₂ O ₃ -6	-3.61153
21	Nd ₂ O ₃ -7	-3.69691
22	NiO-6	-4.62588
23	NiO-7	-2.81348
24	Ni ₂ O ₃ -6	-1.89302
25	Ni ₂ O ₃ -7	-2.94444
26	Sb ₂ O ₃ -7	-2.79123
27	Cr ₂ O ₃ -6	-1.98865
28	Sm ₂ O ₃ -7	-2.54149
29	SnO ₂ -7	-1.57152
30	CeO ₂ -6	1.263337
31	TiO ₂ -6	4.027899
32	Eu ₂ O ₃ -6	2.078542
33	TiO ₂ -7	2.915349
34	HfO ₂ -7	-3.01394
35	Y ₂ O ₃ -6	-2.18947
36	Y ₂ O ₃ -7	-6.07254
37	Yb ₂ O ₃ -6	-1.86605
38	Al ₂ O ₃ -6	-1.96897
39	CoO-6	0.187346
40	CoO-7	-5.7414
41	Co ₃ O ₄ -7	-4.4364
42	CuO-7	-2.75687
43	Dy ₂ O ₃ -7	-1.89831
44	Er ₂ O ₃ -6	-4.34543
45	Er ₂ O ₃ -7	-2.97648
46	Eu ₂ O ₃ -7	-4.91255
47	Fe ₂ O ₃ -6	-4.85881
48	Fe ₂ O ₃ -7	-4.03373
49	Fe ₃ O ₄ -7	-2.56086
50	Gd ₂ O ₃ -6	-1.87729
51	Gd ₂ O ₃ -7	-5.17937
52	Co ₃ O ₄ -6	-2.49513
53	MnO ₂ -6	0.09447
54	MnO ₂ -7	2.624154

3.5.2 Periodic table descriptor calculations

We have taken some of the basic pieces of information from the periodic table to calculate descriptors for MeOx NPs. The metal's information was taken directly from the periodic table and calculated with the Elemental Descriptor Calculator software available from (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>), are known as the first-generation periodic table descriptors and those calculated using some relevant formulas are known as the second-generation periodic table descriptors (De et al., 2018). A total of forty-four descriptors were calculated out of which twenty-eight were 1st generation and sixteen were 2nd generation periodic table-based descriptors. These descriptors were calculated in no time which is convenient as well as cost-effective. It is worth mentioning here that these descriptors are independent of nanoparticle size, and these descriptors can be utilized in the mixture calculation of MeOxNPs which is problematic with quantum chemical descriptors.

3.5.3 Model development

We have developed MLR (multiple linear regression) models with FCIL-1 β (Dataset 5.1) and PLS model with $-\ln(p/(1-p))$ (Dataset 5.2) datasets. For regression modelling, at first, linear regression was performed with the first set of data comprising 32 MeOxNPs (Dataset 5.1). After randomly dividing the training and test sets (70:30), feature selection was carried out using the Genetic algorithm (GA). With the selected features, multiple linear regression was performed using the MINITAB Software (MINITAB is a Statistical Software of Minitab Inc., USA, <http://www.minitab.com>) for model development. A schematic diagram of the whole process is illustrated in **Figure 3.10**. For PLS regression modeling with $-\ln(p/(1-p))$ endpoint (Dataset 5.2) we have performed stepwise regression in MINITAB software. Further, partial linear regression (PLS) analysis was performed with the features obtained from stepwise regression procedure followed by best subset selection (BSS) (Mitteroecker and Bookstein, 2011).

3.5.4 Validation metrics

All the requisite internal and external validation metrics were computed using the training and test sets respectively for the MLR models. The fitting potential of regression-based models was determined based on some statistical metrics like the determination coefficient (R^2), the correlation coefficient of the model (Q^2), and root-mean-square error (RMSE) for internal prediction. External validation was checked by the validation metric Q^2_{F1} . The r_m^2 metrics namely r_m^2 and Δr_m^2 developed by Roy (Roy et al. 2012) were also employed for the present work. Further, the reliability of the developed MLR model developed using 32 data points was checked on a true external set using the PRI (Prediction Reliability Indicator) tool from the DTC website (http://teqip.jdvu.ac.in/QSAR_Tools/). The reliability of predictions was checked on 37 MeOx NPs (**Table 3.13**) that lacked experiment data. This tool categorizes the quality of prediction of true external set into three groups i) good (with composite score 3), ii) moderate (with composite score 2), and iii) bad (with composite score 1) (Roy et al. 2018).

Table 3.13: External dataset

Sl. no.	Metal oxides
1	Ag ₂ O
2	Au ₂ O
3	Au ₂ O ₃
4	AgO
5	GeO ₂
6	Rh ₂ O ₃
7	PtO ₂
8	PtO
9	PdO
10	PbO
11	PbO ₂
12	RuO ₂
13	Ir ₂ O ₃
14	OsO ₂
15	HgO
16	BeO
17	CdO
18	TcO ₂
19	ReO ₂
20	Tl ₂ O
21	Tl ₂ O ₃
22	Ga ₂ O ₃
23	V ₂ O ₃
24	CaO
25	SrO
26	Sc ₂ O ₃
27	BaO
28	Tm ₂ O ₃
29	Tb ₂ O ₃
30	CeO ₂
31	IrO ₂
32	MgO
33	MnO ₂
34	Mo ₂ O ₃
35	Nb ₂ O ₃
36	Ta ₂ O ₃
37	WO ₂

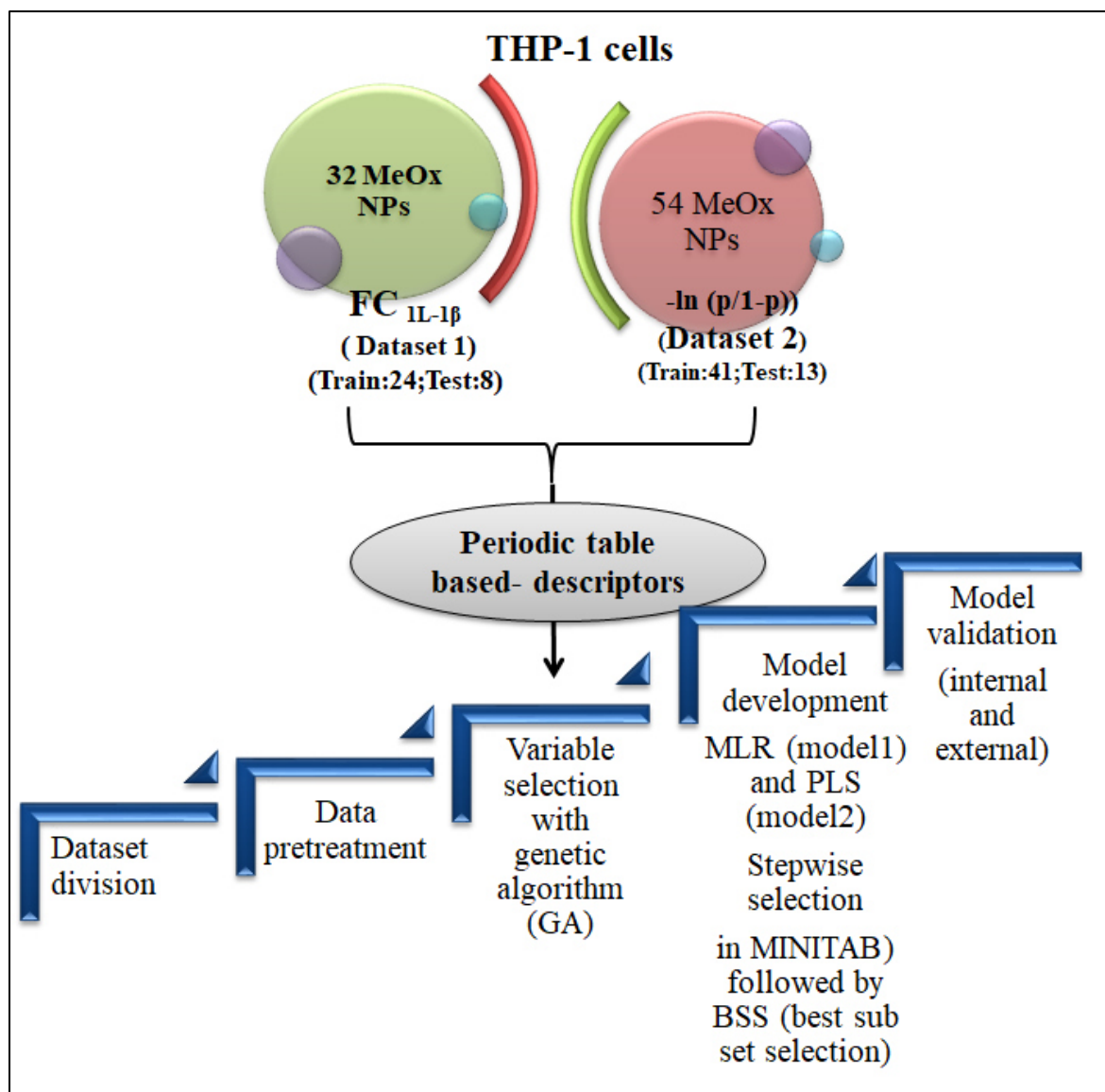


Figure 3.10. Workflow of the present work (Study 5).

3.6. Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches.

3.6.1 Dataset

The dataset was collected from the previously published literature (Sang et al., 2022). A mixture of nano-TiO₂ powders was added to HK-2 cells in Hyclone DMEM medium supplemented with 10% fetal bovine serum (FBS) and 100 mg penicillin/streptomycin and maintained at 37 °C in the presence of 5% carbon dioxide. Nine concentrations of heavy metal salts were added to a constant amount of nano-TiO₂ (25 μmol/L). The details of heavy metal concentrations are given in **Table 3.14**.

Table 3.14: Various concentrations of heavy metals in $\mu\text{mol/L}$.

Sl. no.	CdCl ₂	ZnCl ₂	CuSO ₄	NiCl ₂	Pb (NO ₃) ₂	MnCl ₂	SbCl ₃	CoCl ₂
1	10	60	30	100	100	100	5	10
2	20	90	60	200	200	200	10	20
3	30	120	90	300	300	300	15	30
4	40	150	120	400	400	400	20	40
5	50	180	150	500	500	500	25	50
6	60	210	180	600	600	600	30	60
7	70	240	210	700	700	700	35	70
8	80	270	240	800	800	800	40	80
9	90	300	270	900	900	900	45	90

Table 3.15. Cytotoxicity of heavy metals adsorbed on nano-TiO₂

Compound Id	MeOx NPs	Observed value
1	CdCl ₂	0.9238
2	CdCl ₂	0.7484
3	CdCl ₂	0.6705
4	CdCl ₂	0.5967
5	CdCl ₂	0.5490
6	CdCl ₂	0.4974
7	CdCl ₂	0.4838
8	CdCl ₂	0.4513
9	CdCl ₂	0.4049
10	ZnCl ₂	0.9631
11	ZnCl ₂	0.9467
12	ZnCl ₂	0.9285
13	ZnCl ₂	0.9339
14	ZnCl ₂	0.9057
15	ZnCl ₂	0.8973
16	ZnCl ₂	0.7987
17	ZnCl ₂	0.2918
18	ZnCl ₂	0.0863
19	MnCl ₂	0.9428
20	MnCl ₂	0.8859
21	MnCl ₂	0.7757
22	MnCl ₂	0.7296
23	MnCl ₂	0.6843
24	MnCl ₂	0.5563
25	MnCl ₂	0.5289
26	MnCl ₂	0.4538
27	MnCl ₂	0.4569
28	CoCl ₂	0.8723
29	CoCl ₂	0.6303
30	CoCl ₂	0.5937
31	CoCl ₂	0.5641
32	CoCl ₂	0.4969

33	CoCl ₂	0.4107
34	CoCl ₂	0.3362
35	CoCl ₂	0.3023
36	CoCl ₂	0.2776
37	CuSO ₄	0.9263
38	CuSO ₄	0.7974
39	CuSO ₄	0.7505
40	CuSO ₄	0.6778
41	CuSO ₄	0.4540
42	CuSO ₄	0.2084
43	CuSO ₄	0.0881
44	CuSO ₄	0.0327
45	CuSO ₄	0.0010
46	NiCl ₂	0.8850
47	NiCl ₂	0.7578
48	NiCl ₂	0.6436
49	NiCl ₂	0.5879
50	NiCl ₂	0.5257
51	NiCl ₂	0.4989
52	NiCl ₂	0.4900
53	NiCl ₂	0.4862
54	NiCl ₂	0.4582
55	Pb (NO ₃) ₂	0.6684
56	Pb (NO ₃) ₂	0.5450
57	Pb (NO ₃) ₂	0.4578
58	Pb (NO ₃) ₂	0.4426
59	Pb (NO ₃) ₂	0.4463
60	Pb (NO ₃) ₂	0.3873
61	Pb (NO ₃) ₂	0.3665
62	Pb (NO ₃) ₂	0.3706
63	Pb (NO ₃) ₂	0.3501
64	SbCl ₃	0.6694
65	SbCl ₃	0.5401
66	SbCl ₃	0.2747
67	SbCl ₃	0.0831
68	SbCl ₃	0.0177
69	SbCl ₃	0.0146
70	SbCl ₃	0.0123
71	SbCl ₃	0.0123
72	SbCl ₃	0.0058

HK-2 cells were utilized to determine the toxicity in this study using cell viability as the endpoint. HK-2 cells are a sensitive model for examining renal cytotoxicity. They grow in monolayers and are suitable for studying the proximal tubular toxicity of a variety of compounds (Menze et al., 2009). The main advantage of HK-2 cells is that they retain the basic morphological and functional properties of proximal tubular epithelial cells (Marcílio and Eler, 2020). Cell viability was measured by using equation 3.2.

$$S = \frac{A_{exp} - A_{blank}}{A_{control} - A_{blank}} \quad (3.2)$$

Here, S stands for cell survival rate, A_{exp} is the absorbance value of the experimental group, $A_{control}$ is the absorbance value of the control group, and A_{blank} is the absorbance value of the blank control group. **Table 3.15** contains the cytotoxicity of heavy metals adsorbed on nano-TiO₂ with varying concentration.

3.6.2 Descriptor calculation

Based on the characteristics of metals, we used periodic table-based descriptors, which are easily calculated from the periodic table. Simple molecular information is generated in no time, which is time-effective as well as cost-effective. The previously used descriptors by Kar et al.(2014) are metal electronegativity (χ), the sum of metal electronegativity for individual metal oxide ($\sum\chi$), the sum of metal electronegativity for individual metal oxide divided by the number of oxygen atoms present in a particular metal oxide ($\sum\chi/NO$), number of metal atoms (N_{Metal}), number of oxygen atoms (N_{Oxygen}), the charge of the metal cation corresponding to a given oxide (χ_{ox}) and molecular weight (MW), which are termed as 1st generation periodic table-based descriptors. The newly introduced sixteen descriptors are denoted as second-generation periodic table descriptors (De et al., 2018). In addition to being computationally less demanding, periodic table-based descriptors are size-independent.

3.6.3 Splitting of data set and hyper-parameter tuning

Splitting of the dataset into training and test sets was done before building the model. The training set was mainly used to fit the model, and the test set was used to measure the generalization ability of the developed model. Theoretically speaking, the dataset was divided based on a sorted response-based approach using the in-house dataset division tool (<https://dtclab.webs.com/software-tools>). In this study, the size ratio was set at 3:1 (training set: test set) for dataset division.

In any ML model development exercise, different models are trained for a dataset and the best-performing model is selected. However, there may be room for improvement in the model, and hyperparameter tuning can significantly improve the model. Therefore, here the optimal values of the hyperparameters of the models were obtained with the GridSearchCV algorithm using the Hyperparameter optimizer tool (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home/machine-learning-model-development-guis?pli=1>). GridSearchCV tests all combinations of values in the dictionary and evaluates the model using the Cross-Validation method for each combination. Therefore, we choose the hyperparameter combination with the best average MAE results from the validation sets.

3.6.4 Feature Selection with RF

The goal of feature selection techniques is to find the best set of features that allows one to build optimized models. Feature selection using random forest (RF) comes under the embedded method. The embedded method combines the benefits of the filter and wrapper techniques. This method encompasses the interaction of features while maintaining reasonable computational costs. In the embedded methods, each iteration of the model

training process is taken care of, and a few features are carefully extracted that contribute the most to the training process. More precisely, it measures how much impurity is reduced by averaging (weighted average) through each tree node with the selected features. Here, each node is equivalent to the number of training samples associated with it. Through the RF algorithm, we have selected the best contributing eight periodic table-based descriptors i.e., Conc, $\sum\chi$, atomic radius, IP_ActivM, Mol_Wt, χ of metal, D3_HeteroNonMetal, and the total atom in a molecule from the pool of 43 periodic table-based descriptors through sorting the most contributing features having highest Gini importance (Menze et al., 2009). The selected first eight descriptors (most contributing features) were further used for modeling using RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost algorithms.

3.6.5 Model development

This section introduces four classification models; all are ensemble learning models. ML is a subset of AI (Artificial Intelligence) where the machine learns from data improves performance from past experiences and makes a prediction based on it (Marcílio and Eler, 2020). In this study, along with RF, Gradient Boost, Extreme Gradient Boost, and AdaBoost were also performed. In the supervised learning approach, a model is trained on labeled datasets. Regression analysis algorithms are trained and learned from both input features and output labels. Regression analysis seeks a mapping function from the input features for a continuous output function. In this study, there was no intention to categorize but to predict quantitatively the responses. So, the supervised regression method was selected to map the function of heavy metals and predict the cytotoxicity of these metals on HK-2 cells using periodic table-based descriptors. For the model development, after the dataset division and feature selection, different ML algorithms were applied. The overall workflow is illustrated in **Figure 3.11**.

3.6.6 SHAP analysis

The feature importance in the model was performed using the Shapley Additive exPlanation (SHAP) method using SHAP version 0.41.0. SHAP framework takes into account the calculation of Shapley value. These values are calculated from the average marginal contribution of each feature from all conceivable coalitions. Firstly, the dataset is incorporated into the model, then the SHAP framework assigns the Shapley value to each feature that contributes to the corresponding output of the model. Therefore, SHAP helps to select the features based on the ranking algorithm (Criminisi et al., 2012). We have selected the best features with the one having the highest Shapley values for the training set. Since the standards method tends to overestimate the continuous variable.

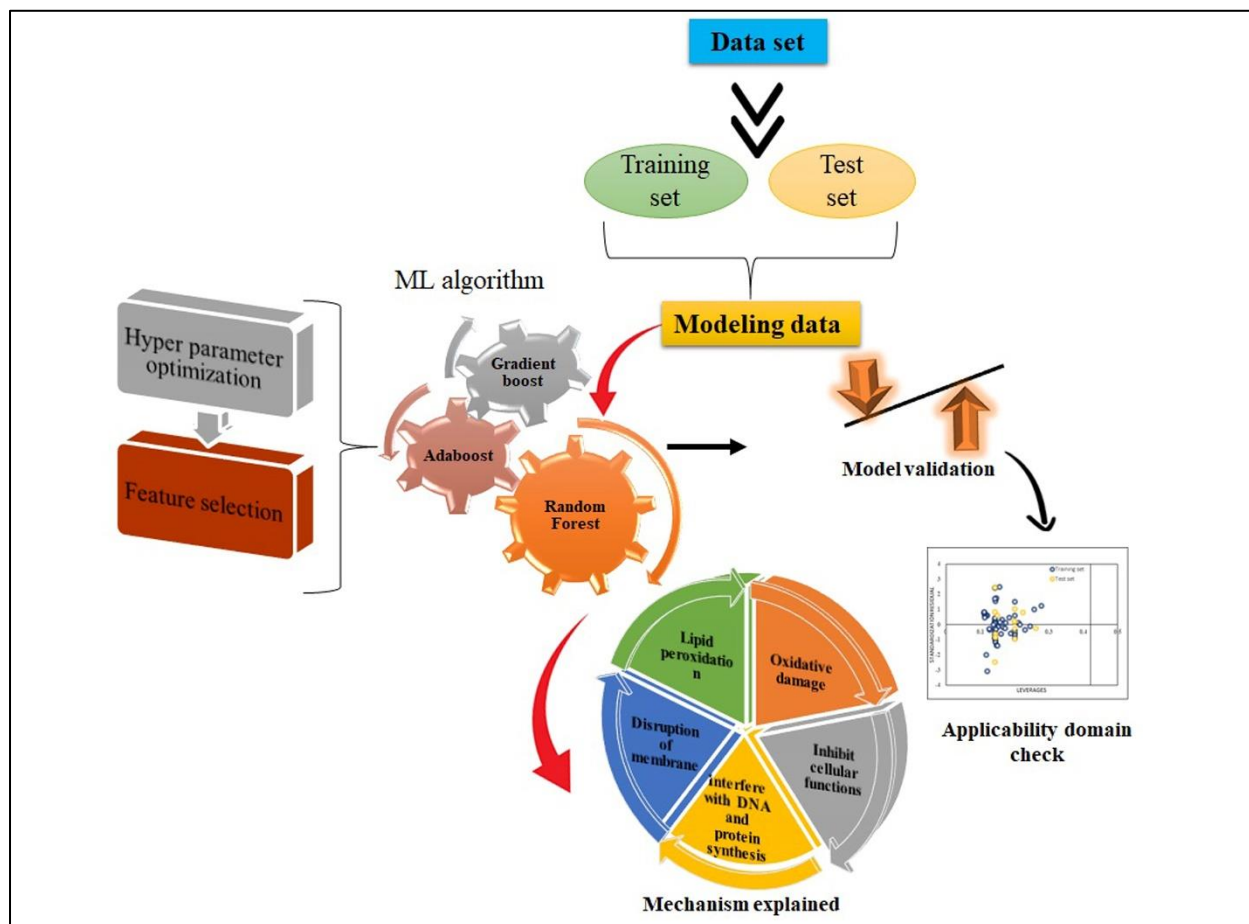


Figure 3.11. Flowchart of the present work (Study 6).

3.6.7 Model validation

A reliable model should pass the threshold values for different internal and external validation metrics. Internal validation generates the generalization ability and robustness of the model. On the other hand, the external prediction is used to validate the model. The most common metrics to measure internal quality are the coefficient of determination (R^2 and Q^2_{LOO}). Besides these, we have also calculated the root mean square error (RMSE) of the training set. $MAE_{(test)}$ (mean absolute error), RMSEP (root mean square error of prediction), Q^2F_1 , and Q^2F_2 were used for external validation or for the test set. The evaluation criteria included are as follows: Q^2_{LOO} and Q^2F_1 , greater than 0.5, RMSE should be between 0.2 and 0.5 whereas the closer the value of the MAE is to 0, the better (Pesantez-Narvaez et al., 2019).

3.7. Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.

3.7.1 Dataset

The study is based on two datasets, i.e., **dataset 7. 1** (Zeta potential) (**Table 3.16**) and **dataset 7. 2** (cell membrane damage). **Dataset 7. 1** consists of 18 metal oxide nanoparticles (MeOx NPs) that were investigated on 5 types of MeOx NPs (MO, MO₂, MO₃, M₂O₃, and M₃O₄). This data was obtained from Cao et al., (2020) where the Zeta potential of MeOx NPs was measured in a cell culture of 20% foetal bovine complete medium. **Dataset 7. 2** was taken from Toropova et al. (2015) where the cell damage measurement was performed based on the uptake of propidium iodide (PI). The dataset is related to the variation of four doses (50, 100, 150, and 200 µg/mL) and exposure time ranging from 1 to 7 hrs that result in 132 MeOx NP data points (**Table 3.17**).

Table 3.16. Dataset 7.1: Zeta potential of MeOx NPs

Compound Id	MeOx NPs	Zeta potential
1	O=[Bi]O[Bi]=O	17.1
2	O=[Ce]=O	17.4
3	O=[Co]O[Co]O[Co]=O	22.6
4	O=[Hf]=O	-17.8
5	O=[Mn]O[Mn]=O	-15.9
6	O=[Ni]	34.4
7	O=[Si]=O	-18.7
8	O=[Ti]=O	-12.2
9	O=[W](=O)=O	-23
10	O=[Zn]	13.9
11	O=[Zr]=O	-2
12	O=[Fe]O[Fe]=O	4.2
13	O=[Gd]O[Gd]=O	19.3
14	O=[La]O[La]=O	25.4
15	O=[Sb]O[Sb]=O	5.4
16	O=[Sn]=O	-20.5
17	O=[Yb]O[Yb]=O	46
18	O=[Cr]O[Cr]=O	2.2

Table 3.17. Dataset 7.2: Cell damage data

Compound Id	MeOx NPs	Cell damage
1	ZrO ₂	-3
2	YY.O ₃	-2.75
3	YY.O ₃	-3
4	YY.O ₃	-3
5	YY.O ₃	-3.125
6	YY.O ₃	-3
7	TiO ₂	-2.875
8	SnO ₂	-3
9	SnO ₂	-3

10	SnO ₂	-3.125
11	SiO ₂	-3.25
12	SiO ₂	-3.125
13	SiO ₂	-3.125
14	SiO ₂	-3.125
15	SiO ₂	-2.875
16	SbSbO ₃	-3.125
17	SbSbO ₃	-2.875
18	SbSbO ₃	-3.125
19	LaLa.O ₃	-2.625
20	HfO ₂	-2.875
21	GdGd.O ₃	-3
22	GdGd.O ₃	-3.05
23	GdGd.O ₃	-3.125
24	GdGd.O ₃	-3.125
25	GdGd.O ₃	-3
26	GdGd.O ₃	-2.75
27	FeFe.Fe.O ₄	-2.875
28	FeFe.Fe.O ₄	-3.5
29	CuO	-3.125
30	CrCr.O ₃	-2.875
31	CrCr.O ₃	-3
32	CoO	-4.375
33	CoO	-5
34	CoCo.Co.O ₄	-3
35	CeO ₂	-4.5
36	YbYb.O ₃	-2.875
37	YbYb.O ₃	-2.875
38	YY.O ₃	-3
39	YY.O ₃	-2.75
40	YY.O ₄	-3.125
41	WO ₃	-3.125
42	WO ₃	-3.5
43	WO ₃	-4.575
44	WO ₃	-3
45	SnO ₂	-3.25
46	SnO ₂	-3.125
47	SnO ₂	-3.125
48	SnO ₂	-3.125
49	SiO ₂	-3
50	SiO ₂	-3.125
51	SbSbO ₃	-4.625
52	SbSb.O ₃	-4.375
53	NiO	-3.575
54	GdGd.O ₃	-3.55
55	FeFe.Fe.O ₄	-3.125
56	FeFe.O ₃	-4.75
57	CuO	-2.875
58	CuO	-3
59	CuO	-1.5

60	CrCr.O ₃	-3.875
61	ZrO ₂	-3.125
62	ZrO ₂	-3.125
63	ZrO ₂	-3
64	ZrO ₂	-3
65	YbYb.O ₃	-2.875
66	YY.O ₃	-2.875
67	WO ₃	-4.5
68	WO ₃	-3
69	TiO ₂	-3
70	TiO ₂	-2.875
71	TiO ₂	-2.875
72	TiO ₂	-2.875
73	TiO ₂	-2.5
74	SnO ₂	-3
75	SiO ₂	-3.125
76	NiO	-3.55
77	NiO	-3.8
78	NiO	-4.5
79	NiO	-4.75
80	NiO	-4.875
81	InIn.O ₃	-3
82	InIn.O ₃	-3.125
83	InIn.O ₃	-2.875
84	InIn.O ₃	-3
85	InIn.O ₃	-3
86	HfO ₂	-2.8
87	HfO ₂	-2.9
88	GdGd.O ₃	-3
89	FeFe.O ₃	-3.625
90	FeFe.O ₃	-3.375
91	FeFe.O ₃	-4.125
92	CuO	-3
93	CrCr.O ₃	-4
94	CoCo.Co.O ₄	-4
95	CoCo.Co.O ₄	-4
96	CeO ₂	-2.75
97	CeO ₂	-2.875
98	CeO ₂	-3
99	ZrO ₂	-3
100	YbYb.OO ₃	-2.875
101	YbYb.O ₃	-3
102	YbYb.OO ₄	-2.875
103	YY.O ₃	-3
104	YY.O ₃	-2.75
105	YY.O ₃	-3
106	YY.O ₃	-2.875
107	WO ₃	-2.75
108	SnO ₂	-3.125
109	SnO ₂	-3

110	SnO ₂	-3
111	SnO ₂	-3.125
112	SiO ₂	-3.5
113	SiO ₂	-3.25
114	SbSb.O ₃	-3.5
115	MnO ₃	-3
116	LaLa.O ₃	-4.15
117	LaLa.O ₃	-3.675
118	LaLa.O ₃	-3.125
119	LaLa.O ₃	-2.875
120	InIn.O ₃	-2.875
121	InIn.O ₃	-3
122	InIn.O ₃	-2.875
123	InIn.O ₃	-3.125
124	InIn.O ₃	-3.175
125	HfO ₂	-2.75
126	GdGd.O ₃	-3.525
127	FeFe.Fe.O ₄	-2.625
128	CoCo.Co..O ₄	-3.875
129	CeO ₂	-2.875
130	CeO ₂	-2.75
131	CeO ₂	-3
132	AlAlO ₃	-2.875

3.7.2 Descriptor calculation

Selecting the appropriate descriptors is crucial for property and toxicity modeling. Quantitative values of chemical features (descriptors) play a significant role in determining the target endpoint. Therefore, in this study, we have calculated periodic table-based descriptors (PT descriptors) for calculating the relevant features contributing to respective property and toxicity endpoint. The physicochemical features encoding the information of MeOx NPs into PT descriptors were used to build prediction models for Zeta potential and cytotoxicity (cell damage). The basic information of MeOx NPs was directly taken from the periodic table and some were calculated with the Elemental Descriptor Calculator software available from (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>), termed as 1st generation periodic table descriptors. On the other hand, the 2nd generation PT descriptors were calculated using some relevant formulas (De et al., 2018). These descriptors are calculated without any expert intervention, and these are independent of size variations.

3.7.3 Splitting of the data sets

Dataset division or splitting of the dataset into a training set and a test set is essential for developing statistically robust nano-QSPR models. Each of the datasets, i.e., the Zeta potential dataset and cell damage dataset, was divided into training and test sets with a ratio of 7:3 using dataset division software present in the DTC lab software site (http://teqip.jdvu.ac.in/QSAR_Tools/). Accordingly, thirteen compounds were in the training set and five compounds in the test set for the Zeta potential dataset whereas for the cell damage dataset, 111 compounds were present in the training set, and the rest 21 compounds were in the test set. The training set compounds were used for feature selection and model

development; on the contrary, the test set was utilized for assessing the predictivity of the developed model.

3.7. 4 Model development

3.7.4.1 Zeta potential (ζ) QSPR model

To develop the property-based QSPR model, the training set was utilized for model development. The training set of 13 compounds was processed through feature selection via stepwise regression and genetic algorithm (GA) (Sivanandam et al., 2008). After feature selection, the training set was utilized for model development through stepwise regression using the MINITAB software (MINITAB is a Statistical Software of Minitab Inc., USA, <http://www.minitab.com>). The multiple linear regression (MLR) model was obtained with three descriptors keeping the F values to enter and remove 4 and 3.9, respectively. Finally, a PLS (partial least squares) model was developed with the selected features from the MLR model. The developed PLS model consisted of 1 LV (latent variable) which was also developed in the same software as the MLR model, i.e., MINITAB software.

3.7.4.2 Cell damage QSPR model

Previously developed QSPR model (Zeta potential dataset, i.e., Dataset 7.1) was utilized to calculate the Zeta potential of the MeOx NPs in the cell damage dataset (Dataset 7.2) that lacks the Zeta potential information (imputation of sparse data) and was used as a descriptor in case of model development along with the PT descriptors. Although the solvents used for testing metal oxides in both datasets differ, the work involves correlating the zeta potential data (experimental or computed) with the cell damage model as a descriptor. Cao et al. (2020) also used zeta potential as one of the determinants for the endpoint modeled. The zeta potential of all data points was determined in the same solvent, and this does not contribute to the variations in zeta potential values due to solvents. This work is similar to imputation in QSAR modeling, where a missing value is replaced by a predicted value from another model (Walter et al., 2022). The training set with 111 compounds (MeOx NPs) after feature selection through GA was further used for model development. The model development was performed with stepwise regression using the MINITAB software followed by the best subset selection (BSS) method. Further, to enhance the quality of predictions for the test set, we have performed a chemical read-across (RA) approach for the developed MLR model with eight descriptors.

3.7. 5 Model validation

The validation procedure is the prerequisite for the application of nano-QSPR models. Rigorous validation of the developed models was performed following the OECD (Organization for Economic Cooperation and Development) principles (OECD, 2007). Validation of the model includes both internal and external validation. Internal validation indicates the robustness and fitting of the developed model applying the training set, whereas the test set indicates the predictivity of the developed model for the new NMs. Common internal validation methods include the Leave-One-Out cross-validation (Q^2_{LOO}) algorithm and the Y-randomization test (Gramatica, 2007; Gütlein et al., 2013). Model fitting ability is expressed by the squared correlation coefficient (R^2) and mean absolute error (MAE). For judging the external predictivity for the test set, we chose the Q^2_{F1} and Q^2_{F2} metrics. According to the Golbraikh and Tropsha (Golbraikh and Tropsha, 2002) criteria, R^2 should be greater than 0.6 and Q^2 should be greater than 0.5 to meet the standard requirements of

external validation. The true external set was also used to evaluate the predicting power of the model. This was done using the Prediction reliability indicator (PRI) tool available from the DTC lab software tools (http://teqip.jdvu.ac.in/QSAR_Tools/). To further validate model 7.2 for the similarity-based prediction, we have performed chemical read-across analysis (RA).

3.7.5.1 Prediction Reliability Indicator (PRI) tool

Ensuring the reliability of predictions for a new set of data is a vital task. By making robust predictions based on molecular features, we can estimate the external set accurately. In this study, we used the 'Prediction Reliability Indicator tool (Roy et al., 2018) (http://teqip.jdvu.ac.in/QSAR_Tools/) to predict the response of a true external set comprising 49 MeOx NPs (Table 3.18). The tool categorizes the prediction quality as good, moderate, or bad based on certain scoring rules. To assess the predictive power of the developed QSPR models, we used the QSPR model (Model 7.1 with Zeta potential endpoint) to predict the response of the external set. Figure 3.12. shows the overall workflow of the present work, highlighting our confident approach to the study.

Table 3.18: External dataset (Study 7)

Sl No	Metal Oxides	χ_{ox}	tot_metal_alpha	Valence Electron Potential	zeta potential (predicted)
1	Ag ₂ O	1	23	11.4	71.9356
2	Au ₂ O	1	31.2	51	81.64
3	Au ₂ O ₃	3	31.2	51	66.64
4	BaO	2	5.4	21.3	25.0072
5	BeO	2	1	82	3.738
6	CaO	2	3	29	18.386
7	CdO	2	5.75	30	24.0475
8	Co ₂ O ₃	3	8.333333333	38.7	20.3448
9	CoO	2	4.166666667	38.7	18.9698
10	CuO	2	9.333333333	34	30.896
11	Dy ₂ O ₃	3	7	47.4	15.7996
12	Er ₂ O ₃	3	7.222222222	49	15.95933333
13	Eu ₂ O ₃	3	10.16666667	45.6	22.8974
14	Fe ₃ O ₄	2.666667	12	67	25.108
15	Ga ₂ O ₃	3	6.222222222	69.7	9.772133333
16	GeO ₂	4	2.333333333	109	-13.714
17	HgO	2	7.8	28.2	28.7668
18	Ho ₂ O ₃	3	13	47.9	28.4816
19	IrO ₂	4	8.2	92.2	2.0748
20	MgO	2	2.5	40	15.165
21	Mn ₃ O ₄	7	11.5	220	-38.445
22	MnO ₂	4	3.833333333	220	-32.275
23	Mo ₂ O ₃	3	20.5	88.6	36.4794
24	MoO ₂	6	10.25	88.6	-7.8531
25	MoO ₃	6	10.25	88.6	-7.8531
26	Nb ₂ O ₃	3	20	104	32.396
27	Nb ₂ O ₅	5	20	104	17.396

28	Nd ₂ O ₃	3	20	43.4	44.2736
29	Ni ₂ O ₃	3	8.666666667	42	20.408
30	OsO ₂	4	7.4	91.4	0.5276
31	PbO	2	3.9	24.2	21.2438
32	PbO ₂	4	3.9	24.2	6.2438
33	PdO	2	0.388888889	24.2	13.76513333
34	Pr ₆ O ₁₁	4	34.2	42.64	67.16856
35	PtO	2	15.4	92.2	32.4108
36	PtO ₂	4	15.4	92.2	17.4108
37	ReO ₂	4	7.3	180	-17.051
38	Rh ₂ O ₃	3	22	64	44.496
39	RuO ₂	4	10.75	64	13.0335
40	Sc ₂ O ₃	3	6.333333333	58	12.302
41	Sm ₂ O ₃	3	6.555555556	44.8	15.36253333
42	SrO	2	4.5	25.7	22.2278
43	Ta ₂ O ₃	3	14.2	110	18.866
44	Ta ₂ O ₅	5	7.1	110	-11.257
45	TcO ₂	4	5.125	180	-21.68375
46	Tl ₂ O	1	10.4	9.6	45.4504
47	Tl ₂ O ₃	3	10.4	9.6	30.4504
48	V ₂ O ₃	3	7	120	1.57
49	WO ₂	4	7.2	140	-9.424

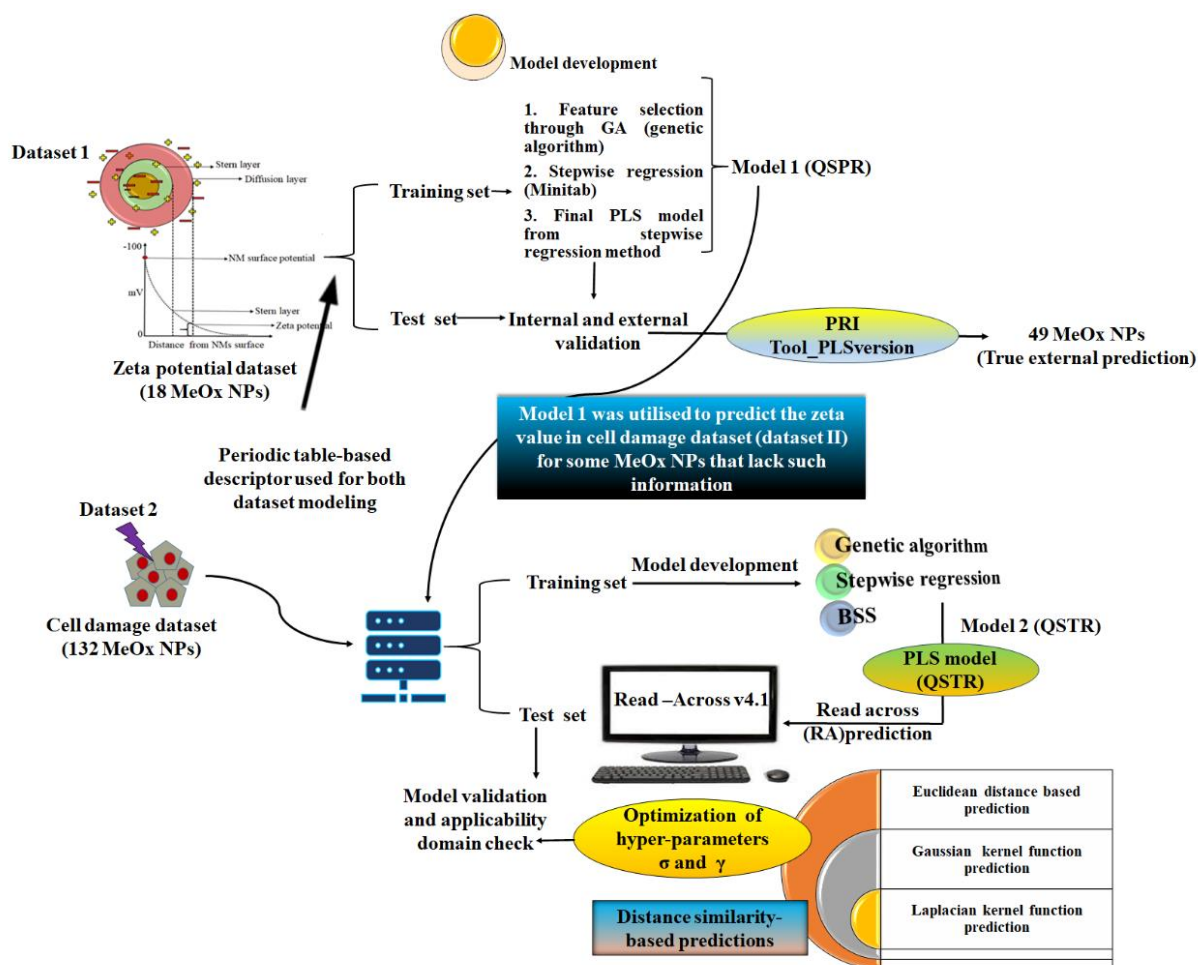


Figure 3.12. Workflow of the developed QSPR (model 7.1) and QSAR (model 7.2) models (Study 7).

3.7.5.2 Read-across (RA) analysis

The Reach-Across (RA) technique is a reliable and scientifically proven method used to predict the endpoint of a new compound, also known as the target compound. This technique involves utilizing data from similar substances that have a regular pattern resulting in structural similarity analogues and bearing similar physicochemical, toxicokinetic, toxicodynamic, and ecotoxicological properties (Chatterjee et al., 2022). Therefore, after selecting the appropriate descriptors from the PLS model (Model 2), we have applied the Quantitative Read-Across v4.0 tool available from our laboratory website (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home>). This tool uses a similarity-based approach based on Euclidean distance, Gaussian kernel function, and Laplacian kernel function. The method requires optimization of the hyper-parameters (sigma and gamma values; distance, and similarity thresholds). To ensure the best results, we used Dataset 7.2, which we divided into a 70% training set and a 30% test set. We further divided the training set into a sub-training and sub-test set to fine-tune the hyper-parameters by changing the default setting. Finally, we used the best hyper-parameters to predict the external set and achieved the best possible results through a rigorous process.

3.8. Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across and ML algorithms: QSAR models for *Daphnia magna* and zebrafish.

For this study on the toxicity testing of metal oxide nanoparticles on aquatic organisms, two datasets were used. The first dataset (Dataset 8.1) was sourced from the NanoTox database (<http://iitrindia.org/envis/Default.aspx>) against *D. magna* and included 28 different MeOx NPs represented in **Table 3.19**. It provided data on their physiochemical properties such as size and the duration of exposure during toxicity testing. The data for *D. magna* was expressed in mg/L, but it was later converted to a negative logarithm scale after converting into a molar scale. The second dataset (Dataset 8.2) as shown in **Table 3.20** was obtained from the literature (Sizochenko et al., 2017) and was used to predict the toxic outcomes in zebrafish. It analyzed the enzymatic activity of metalloprotease zebrafish hatching enzyme, ZHE1 for 24 metal oxides.

Table 3.19. Dataset 8.1: Toxicity data for *D. magna*

Sl. No	Metal oxides	log (1/EC ₅₀)
1	Ag ₂ O	4.39794
2	CeO ₂	1.931814
3	CeO ₂	1.718967
4	CeO ₂	1.9914
5	CeO ₂	1.595166
6	CeO ₂	1.369572
7	CeO ₂	1.688246
8	CeO ₂	4.920819
9	CuO	2.49485
10	CuO	3.337242
11	CuO	1.91364
12	Fe ₃ O ₄	6.638272
13	MgO	1.20852
14	TiO ₂	4.79588
15	TiO ₂	1.452151
16	TiO ₂	2.259637
17	TiO ₂	0.995679
18	TiO ₂	0.521434
19	TiO ₂	1.29243
20	TiO ₂	2.790485
21	ZnO	3.20621
22	ZnO	2.9914
23	ZnO	2.850781
24	ZnO	2.716699
25	ZnO	2.769551
26	ZnO	2.138107
27	ZnO	3.119186
28	ZnO	3

Table 3.20. Dataset 8.2: enzyme inhibitory activity of *zebrafish hatching enzyme, ZHE1*

Sl. No	Metal oxides	Enzyme Activity (A), mmol/mg·min
1	Al ₂ O ₃	1.17
2	CeO ₂	1.1
3	Co ₃ O ₄	1.25
4	CoO	1.17
5	Cr ₂ O ₃	0.68
6	CuO	0.62
7	Fe ₂ O ₃	1.15
8	Fe ₃ O ₄	1.02
9	Gd ₂ O ₃	1.1
10	HfO ₂	1.1
11	In ₂ O ₃	1.17
12	La ₂ O ₃	1.15
13	Mn ₂ O ₃	1.08
14	Ni ₂ O ₃	1.08
15	NiO	0.81
16	Sb ₂ O ₃	1.15
17	SiO ₂	1.1
18	SnO ₂	1.17
19	TiO ₂	1.1
20	WO ₃	1.15
21	Y ₂ O ₃	1.1
22	Yb ₂ O ₃	1.1
23	ZnO	0.7
24	ZrO ₂	1.13

3.8.1 Theoretical descriptors

A simplified theoretical descriptor, based on the periodic table, has been recently introduced (Roy and Roy, 2023). The descriptor set in the present study includes 41 descriptors, which include both 1st and 2nd generation PT descriptors (De et al., 2018). Among these descriptors, some were directly calculated from the periodic table, while the rest are derived from the primary descriptors (1st generation). One can find the complete list of these descriptors in the Introduction Section **Table 1.2**. Additionally, some related descriptors were calculated using Elemental Descriptor Calculator software, which is available from <https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>. It is worth noting that the descriptors can be computed easily and promptly without any expert knowledge.

3.8.2 Dataset division

The process of dividing a dataset is crucial when developing a model, and we have achieved this using the DTC lab software tool site (http://teqip.jdvu.ac.in/QSAR_Tools/). We used Euclidean distance (ED) based division for both datasets. The dataset was divided into 80:20 ratios. For the *Daphnia magna* dataset (Dataset 8.1), we used a training set of 23 NPs and a test set of 5 NPs. Similarly, for the zebrafish dataset (Dataset 8.2), we used 19 NPs for the

training set and 5 NPs for the test set. Our approach involved using the training set to develop the model and the test set to evaluate the model's performance.

3.8.3 Important feature selection

In a QSAR study, selecting the most relevant features that contribute to the toxicity endpoint is crucial. To extract these important features, several selection algorithms are available, including stepwise selection, genetic algorithm, all possible features selection from best subsets factor analysis, swarm optimization, and more. In our study, we used a genetic algorithm (GA) (Sivanandam et al., 2008), stepwise feature selection methods (Zhou et al., 2006), and the best selection software tool (http://teqip.jdvu.ac.in/QSAR_Tools/) to obtain the most suitable features for model development. The selected features were then used to build a regression-based model based on their corresponding response value or endpoint.

3.8.4 Regression based QSAR model development

Both dataset 8.1 and dataset 8.2 were trained using stepwise regression using the MINITAB software (MINITAB is a Statistical Software of Minitab Inc., USA, <http://www.minitab.com>) followed by running best selection (BSS) software tool to obtain the best possible combination of descriptors. The multiple linear regression (MLR) model was developed using four descriptors for dataset 8.1 and dataset 8.2. The final descriptors were selected based on their robustness (based on Q_{Loo}^2). Additionally, PLS (partial least squares) was performed for dataset 8.1 with three LVs (latent variables). The PLS regression strategy is a more generalized version of the “multiple linear regression (MLR)” model. The PLS model allows for the examination of highly correlated, noisy, and collinear data that contain significant ‘X’ variables (Wold, 2001).

3.8.5 Read-across (RA)

Read-across (RA) is a chemometric approach that uses data from a source compound to predict a property or response for a target compound, based on similarity (Schultz et al., 2015). It is an efficient method for predicting the properties of a query compound. To perform the RA predictions, the selected features from the QSAR model were used with the Read-Across-v4.1 tool (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home>), which employs three similarity-based approaches: Euclidean distance-based, Gaussian kernel similarity-based, and Laplacian kernel similarity-based. RA is also proposed as an alternative technique for data gap-filling and experimental hazard assessment of NPs. In this study, the training set was divided into sub-train and sub-test sets, followed by hyperparameter optimization (sigma and gamma values, distance, similarity thresholds, and close source compounds) (Feurer et al., 2019), to find the best similarity-based prediction. The best result was then applied to the entire training and test sets. Based on the RA hypothesis, the similarity function that gave the best prediction for the query compound was selected in terms of Q_{F1}^2 or MAE (Roy et al., 2016). The Read-Across-v4.0 software (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home>) was utilized for this work.

3.8.6 Development of various ML (machine learning) models

The study also explored machine learning algorithms (ML) to check the predictive ability of the created QSTR models. Various approaches such as support vector machine (SVM), linear SVM, random forests (RF) (Rigatti, 2017), partial least squares (PLS) (Wold, 2001), and ridge regression (RR) (McDonald, 2009) were tested. A machine learning regression model development software available from the DTC lab (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home/machine-learning-model-development-guis>) was used to perform the task. The grid search was performed to optimize the associated hyperparameters with 5-fold cross validation mean square error as an objective function in each ML approach (Lerman, 1980). The selected features from the respective MLR models were used for performing the ML algorithm. The best model was judged based on the external Q_{F1}^2 performance. **Figure 3.13** shows the workflow of the present work.

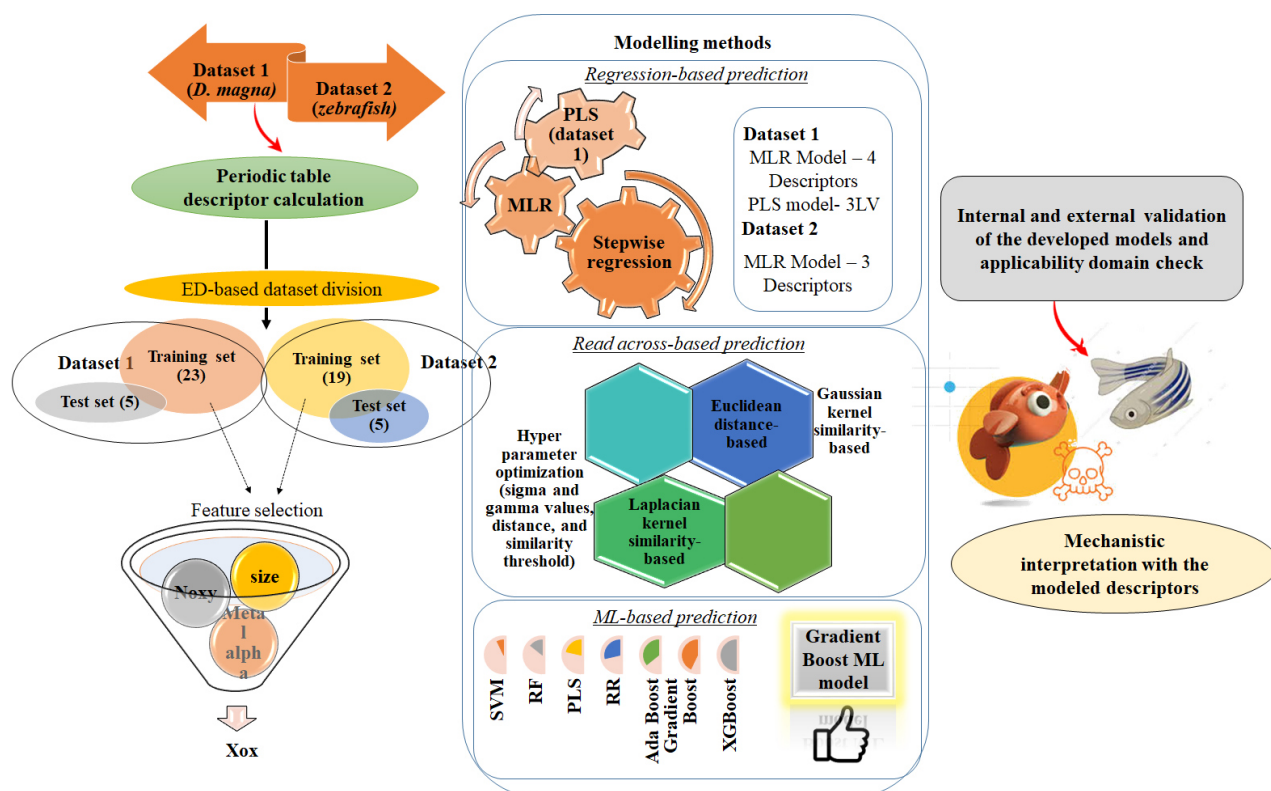


Figure 3.13. Workflow of the present work (Study 8).

CHAPTER - 4

RESULTS AND DISCUSSIONS

Chapter 4: Results and Discussion

4.1. Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach.

Linear regression models have been developed for the cytotoxicity data of 25 nano oxides (hydroxides) towards macrophage RAW 267 cells using the release of LDH (ldh_R) as the endpoint. The present work was divided into two sections; at first, the whole set was subjected to the small data set modeller due to a limited number of experimental data being available, and secondly, the same dataset was further divided into a training set and a test set to check the predictivity of the developed models (**Table 4.1 and Table 4.2**). The statistical values ($R^2=0.75$, $Q^2=0.664$) of the PLS model (N1.1) developed with the small dataset modeller were satisfactory and it was also noted that a few descriptors obtained from the MLR models (M1.1, M 1.2, M1.3, M1.4) were common to the PLS (N1.1) model thus justifying the importance of such descriptors. The MLR models (M1.1, M 1.2, M1.3, M1.4) were obtained through the Best Subset Selection method and the models were subjected to the Intelligent Consensus Prediction technique. The descriptors obtained from both the PLS model (N1.1) and MLR models (M1.1, M 1.2, M1.3, M1.4) were elaborately discussed (**Table 4.3**) below with their possible mechanisms. The scatter plots of MLR models are also shown in **Figure 4.1**.

The AD was also checked using the standardization approach for MLR models and DModX, i.e., distance to model in the X-space with SIMCA 16.0.2 software for the PLS model to calculate the AD. At 95% confidence level ($D\text{-crit} = 0.009998$) the AD of the PLS model was checked (**Figure 4.2**).

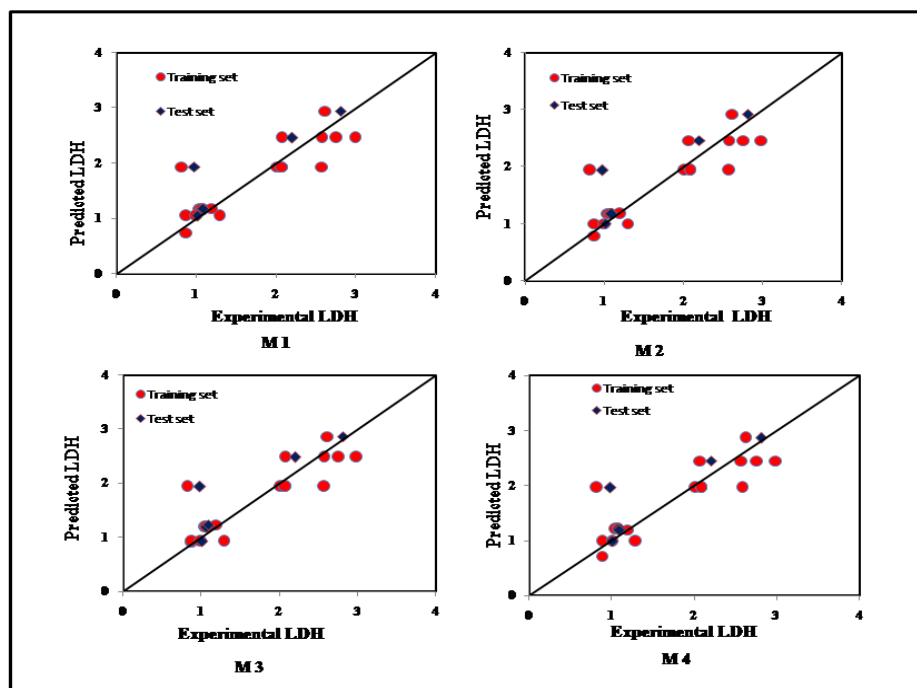


Figure 4.1 Scatter plots of MLR models (study 1)

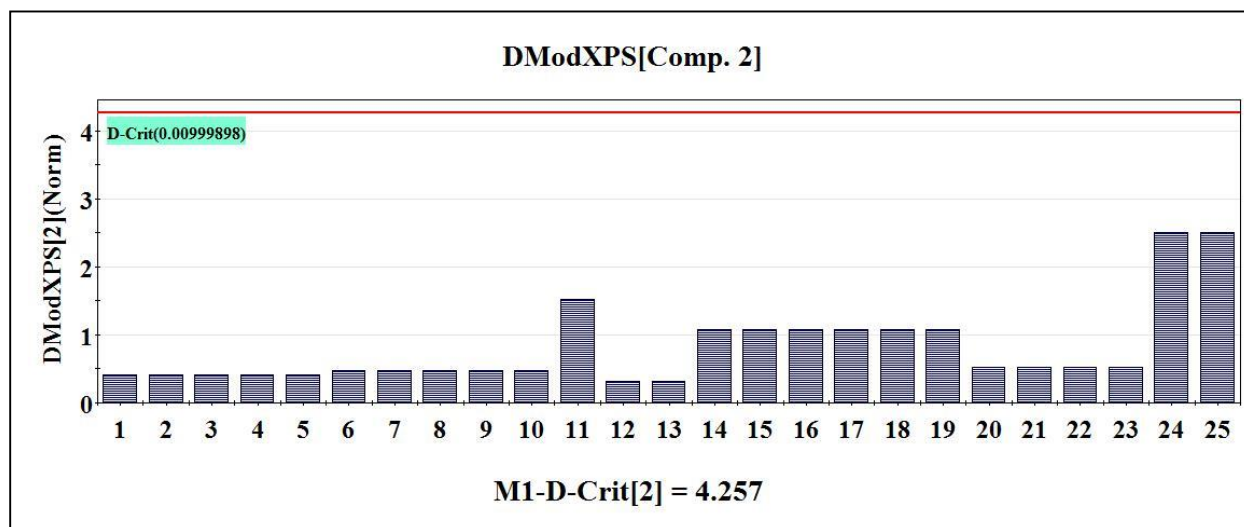


Figure 4.2 Applicability domain of model N1.1 (PLS model)

The Variable importance plot of the PLS model highlights the influential descriptors contributing to the cytotoxicity of MeO_xNPs to RAW267.2 cells. The descriptors of the N1.1 model appearing as per the VIP scores are solubility, Metals_SumIP, SumElectrons_ Active_M_SM. The descriptors from higher to lower contributions are shown in **Figure 4.3**

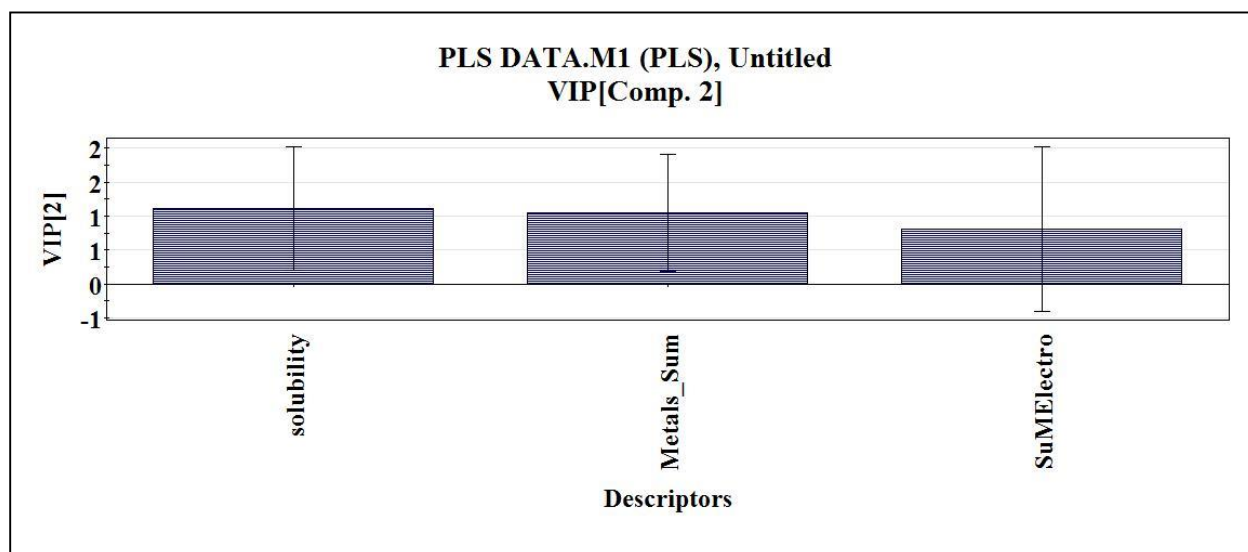


Figure 4.3 Variable importance plot (VIP) of N1.1 model

4.1.1 Loading plot of the PLS model

The loading plot gives the relationship between the X-variables (descriptors) and the Y-variable (LDH) of the PLS model as shown in **Figure 4.4**. The development of the loading plot was done using two PLS components (first and second components). This plot gives insights into how different variables influence the model. During the evaluation, the distance from the origin was taken into consideration in the loading plot. The descriptors located far away from the origin in the plot are considered to have a greater influence on the model and the Y response. The solubility descriptor is situated furthest from the origin and hence it has more impact with respect to the other two descriptors. It can also be confirmed from the VIP plot also.

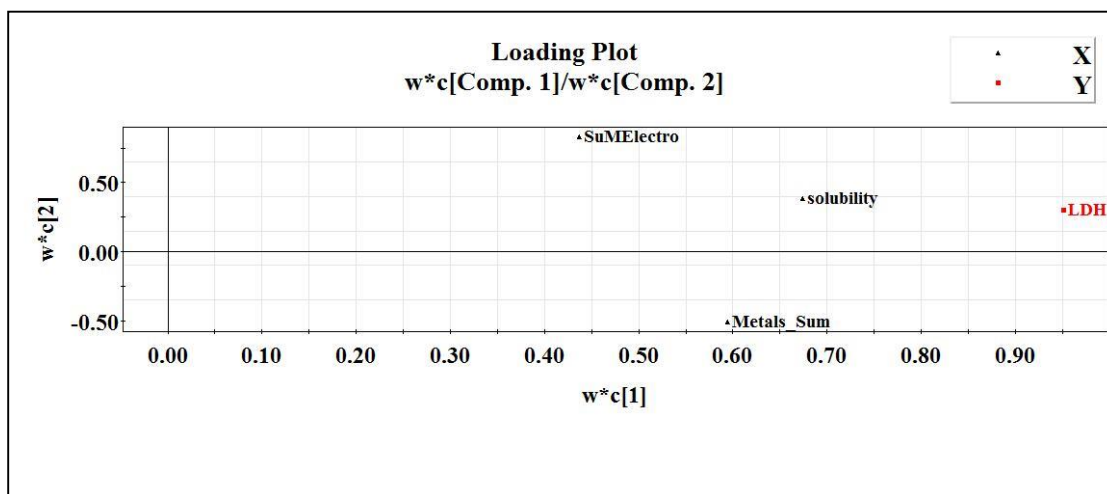


Figure 4.4 Loading Plot of the PLS model (N1.1)

4.1.2 Model randomization

Randomization of a model authenticates the robustness of the model. It signifies that the developed model is not obtained by chance correlation. The randomized model is developed by shuffling different combinations of X or Y variables leading to multiple models generated. Here, we have used 100 permutations, which can be randomized depending on the user's choice. The random models should have poor statistics to prove that the non-random model is not formed by chance correlation. We have provided the correlation between the original Y-vector and permuted Y-vector versus the cumulative R_y^2 and cumulative Q_y^2 plots as represented in **Figure 4.5**. The plot shows that the model is non-random as well as robust since $R_y^2=0.025$ and $Q_y^2 = -0.258$ and one can easily predict the cytotoxicity of the Me_{Ox}NPs from the developed PLS model.

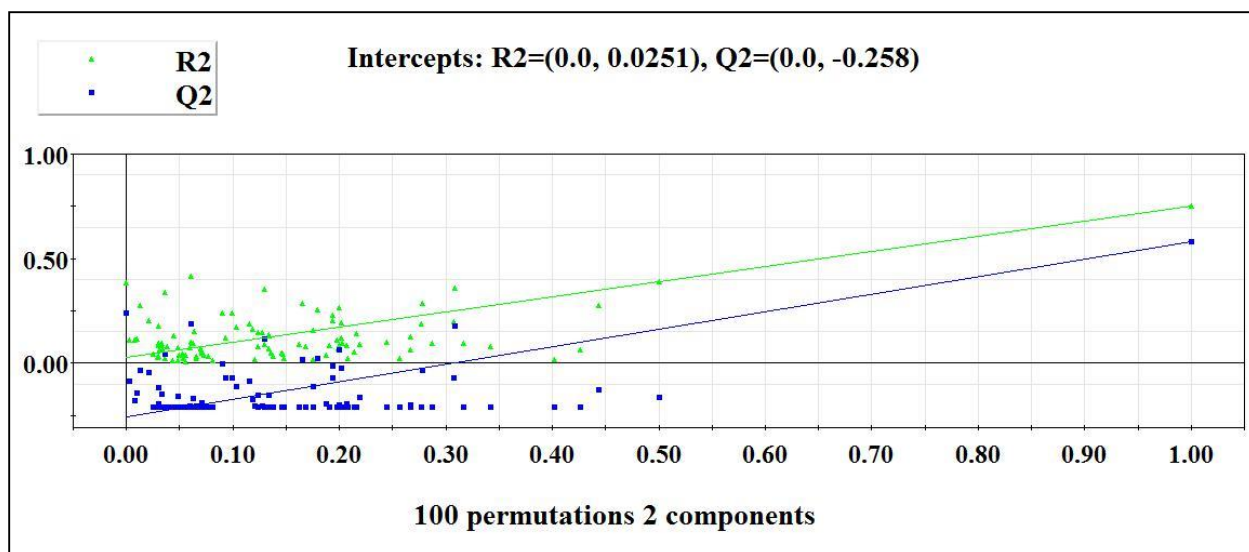


Figure 4.5 Randomization plot of the PLS model (N1.1)

Table 4.1. Statistical results of MLR models

Statistical parameters	Training set			Test set		MAE based criteria
	R^2	R^2_{adj}	Q^2	Q^2_{F1}	F value	
M1.1	0.749	0.695	0.611	0.722	13.92	Good
M1.2	0.750	0.696	0.612	0.720	13.98	Good
M1.3	0.750	0.697	0.624	0.712	14.03	Good
M1.4	0.747	0.693	0.620	0.707	13.77	Good

Table 4.2. PLS model result

Model	R^2	Q^2	$Q^2_{LMO(20\%)}$	$\bar{r}_m^{-2}$	Δr_m^2
N1.1	0.75	0.662	0.662	0.579	0.077

Study 1 result

MLR models

M1.1

$ldh_R = 1.2782(\pm 0.4431) + 0.00134(\pm 0.00055) \text{ Metals_SumIP} + 0.0322(\pm 0.0066) \text{ SumElectrons_Active_M_SM} + 0.20123(\pm 0.03801) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.749, R^2_{adj} = 0.695, Q^2 = 0.611, S = 0.421658, F = 13.92, \text{PRESS} = 3.85635, \text{MAE based criteria} = \text{Moderate}$
 $n_{\text{test}} = 7, Q^2_{F1} = 0.722, Q^2_{F2} = 0.682, \text{MAE based criteria} = \text{Good}$

M1.2

$ldh_R = -0.488(\pm 0.6306) - 0.00367(\pm 0.00058) \text{ Metals_SumIP} + 0.0406(\pm 0.0083) \text{ Cat_rad} + 0.197(\pm 0.0375) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.750, R^2_{adj} = 0.696, Q^2 = 0.612, S = 0.421001, F = 13.98, \text{PRESS} = 3.84633, \text{MAE based criteria} = \text{Moderate}$

$n_{\text{test}} = 7, Q_{F1}^2 = 0.720, Q_{F2}^2 = 0.679, \text{MAE based criteria} = \text{Good}$	
M1.3	$ldh_R = -0.150(\pm 0.6904) + 0.041596(\pm 0.0079) \text{Cat_rad} - 0.091(\pm 0.0137) \text{Oxidation degree of metal(N_oxy_M)}$ $+ 0.1834(\pm 0.045) \text{Solubility}$
	$n_{\text{training}} = 18, R^2 = 0.750, R_{\text{adj}}^2 = 0.697, Q^2 = 0.624, S = 0.420392, F = 14.03, \text{PRESS} = 3.72620, \text{MAE based criteria}$ $= \text{Moderate}$
	$n_{\text{test}} = 7, Q_{F1}^2 = 0.712, Q_{F2}^2 = 0.670, \text{MAE based criteria} = \text{Good}$
M1.4	$ldh_R = -0.436(\pm 0.6208) - 0.246(\pm 0.5132) \text{D1_Metals} + 0.0405(\pm 0.0088) \text{Cat_rad} + 0.0204(\pm 0.0377) \text{Solubility}$
	$n_{\text{training}} = 18, R^2 = 0.747, R_{\text{adj}}^2 = 0.693, Q^2 = 0.620, S = 0.423426, F = 13.77, \text{MAE based criteria} = \text{Moderate}$
	$n_{\text{test}} = 7, Q_{F1}^2 = 0.707, Q_{F2}^2 = 0.665, \text{MAE based criteria} = \text{Good}$
PLS model (N1.1)	$ldh_R = 1.44 + 0.19 \times \text{Solubility} + 0.01 \times \text{Metals_SumIP} + 0.029 \times \text{SumElectrons_Active_M_SM}$ $R^2 = 0.75, Q^2 = 0.664, \overline{rm}^2 = 0.579, \Delta rm^2 = 0.077$

Table 4.3. Description and Contribution of the descriptors obtained from MLR and PLS models.

Sl.no.	Descriptor	Description	Contribution
1	Metals_SumIP	The sum of ionization potential energy (kJ/mol) of metals.	+ve
2	D1_Metals	Total number of metals	+ve
3	Cat_rad	Distance between the nucleus and the electron in the outermost shell of an ion	+ve
4	SuMElectrons_Active_M_SM	Summation of electrons of active metal and semi-metal.	+ve
5	N_oxy_M	Oxidation degree or state of metal	-ve
6	Solubility	The ability of a solid, liquid, or gaseous chemical substance (referred to as the solute) to get dissolved in a solvent (usually a liquid) and form a solution.	-ve

The **Metals_SumIP** descriptor stands for the sum of ionization potential energy (kJ/mol) of metals. The ionization potential energy is the quantity of energy that an isolated gaseous atom in the ground electronic state must absorb to discharge an electron resulting in a cation. The **positive** contribution of the descriptor suggests that with an increase in the ionization potential, the toxicity also increases as found in the case of **nanoparticles 12, 8, and 6** whereas the reverse happens for **nanoparticles 24, 2, and 3** where with a decrease in ionization potential the toxicity also decreases. Ionization potential is related to electronegativity and electron affinity. If a molecule has more ionization potential that means that it is difficult to remove the valence electrons easily to form a cation which suggests that the molecule is more electronegative. The increase of electronegativity of the chemicals may increase the production of reactive oxygen species (ROS).

D1_Metals stands for the total number of metals. The presence of metals in the nanoparticles tends to increase the release of LDH from the cells as indicated by **nanoparticles 7, 8, 12** while **nanoparticle 24** shows a low LDH release from the RAW 264.7 cell line for being non-metallic in nature. This is correctly indicated by the **positive** contribution of the descriptor. Metal-based NPs can show tremendous toxicity and can pose detrimental biological interactions to mammalian cells (Jeng et al., 2006). Metal oxide nanoparticles with a larger surface area can penetrate deeply into the body due to their smaller size and possess surface charge which causes greater inflammatory response (Nemmar et al., 2002). Studies revealed that different nanoparticles modulate the inflammatory responses of macrophages in different ways.

The periodic table descriptor, **Cat_rad**, is the distance between the nucleus and the electron present in the outermost shell of an ion. The **positive** contribution of the descriptor suggests that with an increase in cationic radius, the LDH released from the cell increases as shown in **nanoparticles 14 and 88**. On the contrary, a decrease in cationic radius leads to less LDH release as indicated by **nanoparticles 3 and 24**. The addition of electrons increases the size of an atom through the repelling of the electrons within the shells. When the cationic radius increases, it becomes easier for the metal to lose electrons to form a cation. This happens because as the size of the atom increases, the outermost electrons feel a decrease of shielding by the inner-shell electron, and as a result electrons present in the outer shell do not “feel” the entire positive charge of the nucleus. Consequently, metal cations released from the metal oxides (Me_{ox}) may catalyze the formation of hydroxyl radicals to produce ROS via the so-called Haber-Weiss-Fenton cycle (Roy et al., 2019; Koppenol, 2001) leading to apoptosis to the cell through LDH leakage.

SuMElectrons_Active_M_SM stands for the summation of electrons of active metals and semi-metals. This descriptor **positively** contributes to cell toxicity. The presence of electrons can also be correlated to the higher electronegativity of that atom. The increase in electronegativity is responsible for the catalytic property of the metal cationic form as well as for the cytotoxicity of RAW 264.7 cells. This descriptor positively contributes as seen in **nanoparticles 15, 8** where an increase of electronegativity increases the release of LDH from the cell, but on the other hand, a decrease in electronegativity decreases the cytotoxicity and release of LDH from the cell as evidenced by **nanoparticles 23 and 3**. The presence of electronegativity is responsible for initiating cation formation in metal oxide (Ozben, 2007).

N_oxy_M stands for the oxidation degree or state of metal. This descriptor is **negatively** correlated with the cytotoxicity of the cell which means that an increase in oxidation degree decreases the cytotoxicity of the cell as evidenced by **nanoparticles 19 and 23**. On the other hand, the lower degree of oxidation enhances the cytotoxicity effect on the cell and promotes the release of LDH as seen in **nanoparticles 8 and 12**. A lower degree of oxidation makes the metal unstable due to an unfilled valence shell and therefore the metal tries to gain a higher oxidation state or a stable one. Transition metal has a different oxidation state due to partially filled -d and -f orbital shells (Walker et al., 2003). When transition metal tends to have a lower oxidation state, it means that the metal easily donates its valence electrons or can gain the electrons to form metal cations or enrich itself with electrons respectively but this is not so easily done if the oxidation state of the metal is increased; thus lower oxidation state of the metal is detrimental to the cell.

Another descriptor, **Solubility** stands for the ability of a solid, liquid, or gaseous chemical substance (referred to as the solute) to get dissolved in a solvent (usually a liquid) and form a solution. The solubility depends on three factors such as temperature, pressure as well as the solvent used. Solubility is one of the pivotal parameters that are involved in nanomaterial-induced toxicity (Schrand et al., 2010). The **positive** contribution of the descriptor suggests that an increase in solubility increases the cytotoxicity as in **nanoparticles 12 and 8**, and the opposite happens for **nanoparticles 11 and 19** with decreased solubility. The NP toxicity is related to its solubility, the dissolution of NPs under acidic conditions and as well as the presence of lipids and proteins (amino acids) which can induce toxicity by binding with them and interfering with the cell functions. A few studies by Xia et al. (2008) and Li et.al (2009) confirmed that the solubility of NPs has a strong relation with cellular toxicity.

4.1.3 Mechanism of toxicity

The wide applications of metallic NPs with their unique properties and structural alteration have led to the modification of biological functions along with potential ROS production as well as change of cellular signaling linked with cell death, proliferation, and differentiation. Reactive oxygen species (ROS) are small, short lived and reactive molecules that are continuously produced and destroyed by a wide variety of antioxidant defense mechanisms in the human body. The generation of ROS by the MeO_x NPs has been considered the major contributor and originator of cytotoxicity. Excessive production of ROS by MeO_x NPs causes oxidative damage, by destroying the cell function by acting on the biomolecules (lipids, proteins, and DNA) hampering their functionality, ultimately causing the death of the cell. The biological toxicity of MeO_x NPs occurs when they interact with solubilized metallic ions. The derived information from the various descriptors gave us insights that among the various features of the nanoparticles, electronegativity is the most prevailing feature responsible for cytotoxicity. Due to the higher electronegativity, the dipole charge distribution takes place between the atoms in the MeO_x NPs. The surface-charged NPs then cause electrostatic interaction with the cellular membrane (CM) of the cell which enhances the membrane-wrapping phenomena. It may also increase the surface tension and facilitate the entry of NPs through the generation of pores. Hydrophobic interaction is also possible with oxides of NPs. The cationic form of MeO_x NPs has a more disruptive influence on the membrane bilayer, and through membrane phase change they enter the cell (Behzadi et al., 2017). MeO_x NPs can interfere with the electron transport chain and depolarize with the mitochondrial membrane through activation of NADPH-related enzymes (nicotinamide adenine dinucleotide phosphate). The interaction of NPs blocks the electron transfer chain, thus elevating the cellular O₂·(oxygen radical) level through electron transfer from the respiratory carriers to O₂ (Soenen et al., 2011). The oxidative stress generated further affects the DNA by the formation of 8-hydroxyl-20-deoxyguanosine (8-OHdG) that ultimately leads to DNA damage, arrest of the cell cycle causing alteration of apoptosis mechanism, and damage to the cellular membrane which promotes the leakage of LDH from the membrane (Xia et al., 2006) **Figure 4.6**. Inflammatory responses by ROS are also enhanced by metallic NPs like TiO₂ NPs and SiO₂ NPs (Park et al., 2009). ZnO NPs cause cytotoxicity by triggering oxidative injury by generating ROS and releasing inflammatory mediators that finally lead to cell death in phagocytic RAW 264.7 cells.

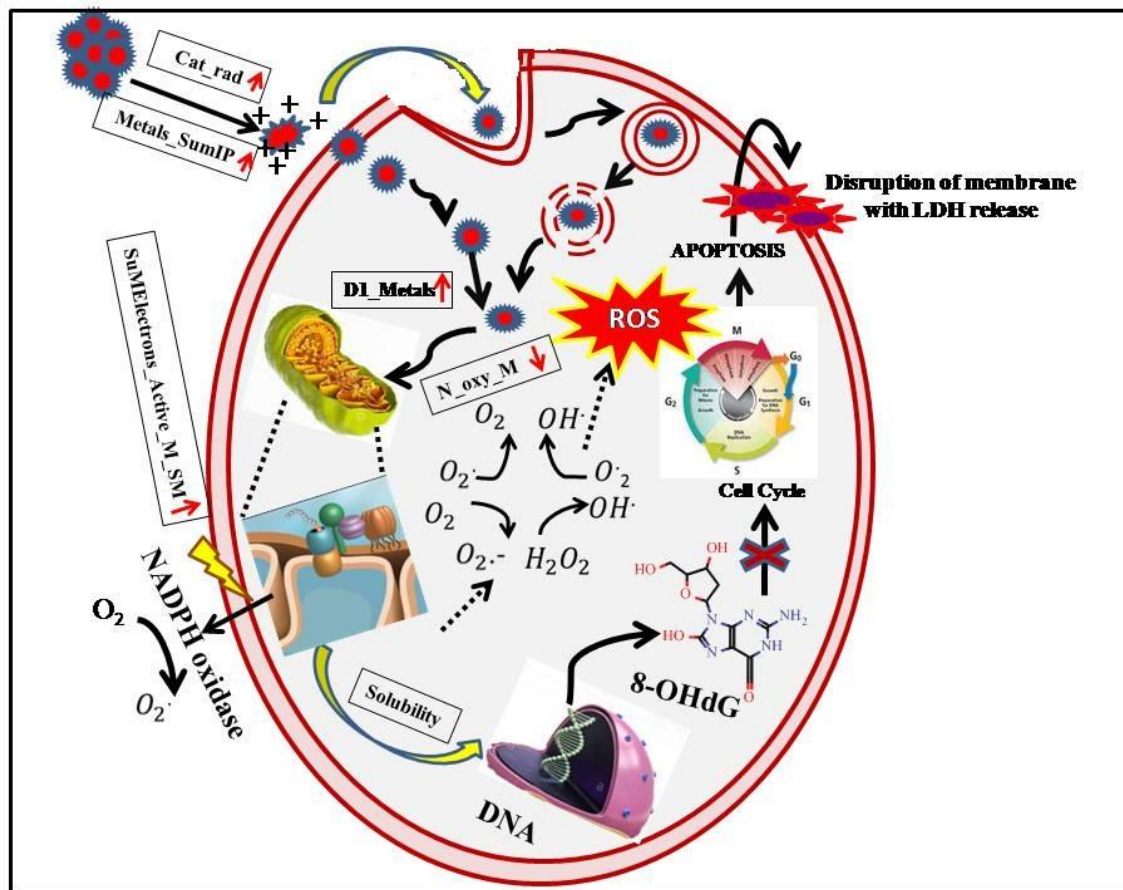


Figure 4.6 Mechanism of toxicity (study 1)

4.1.4 Comparison with previously published models

We have compared the presently developed models (both MLR and PLS) with the previously developed model (Forest et al., 2019) for cytotoxicity of the MeO_x NPs towards RAW 264.7 cells. However, a critical analysis was not possible due to the difference in the method of model development. In the previous study, the authors were unable to develop a sound QSAR model due to the availability of a small dataset. But, on the contrary, we were able to develop meaningful QSAR models with the help of the Small Dataset Modeler tool. In the previous study, the authors performed PLS regression and decision trees, whereas in the current study, PLS and MLR models were developed with simple freely available tools. The advantage of our study is that we have been successful in developing QSAR models with the same small dataset for which the previous authors failed to develop meaningful models with good statistical results. We have also proposed a mechanism of toxicity for better understanding.

4.2. Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides

Periodic table-based descriptors were used for the modeling of 17 MeOxNPs using the small dataset modeler software (Ambure et al., 2019). The statistical quality and validation metrics of both the models have shown acceptable values for R^2 , Q^2_{LOO} , RMSE, MAE, rm^2 metrics as per **Table 4.4**. The descriptors obtained from the MLR models of the two datasets dark (Dataset 2.1) and light-induced toxicity (Dataset 2.2) are explained in **Table 4.5**. The scatter plots (**Figure 4.7**) of the MLR models are also shown in this section.

Table 4.4. Details of statistical metrics obtained from both MLR models.

Statistical parameters	R^2	R^2_{adj}	Q^2_{LOO}	$\bar{r}_m^2$	Δr_m^2	MAE (95%)	Press value	F
Model 2.1	0.726	0.663	0.612	0.497	0.166	0.587	9.747	11.48
Model 2.2	0.77	0.717	0.627	0.543	0.054	0.482	8.494	14.48

Table 4.5. Detailed description of the obtained descriptors from two developed MLR models (A and B).

Serial No.	Descriptors	Description
1.	Atomic rad (pm)	This descriptor stands for the measure of the size of the atom. It typically measures the distance from the center of the nucleus to the outer most shell containing electrons.
2.	Valence	The valence of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules
3.	$(\sum \epsilon / N)^2$	It stands for the total electronegativity of the atoms of the MeOx by the number of atoms present that can be computed based on the electronegativity count ($\sum \epsilon$) of the oxides.
4.	χ_{ox}	It represents the oxidation number of the metal atom.

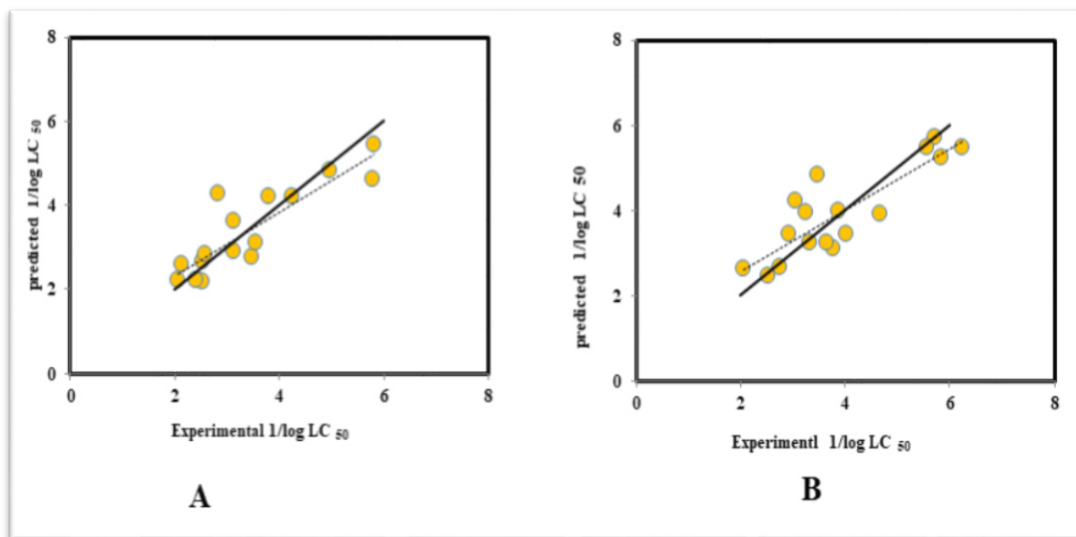


Figure 4.7 Scatter plots for the dark toxicity dataset (model 2.1) (A) and the photo induced dataset (model 2.2) (B).

4.2.1. Interpretation of the obtained descriptors

Model 2.1: Photo-induced toxicity

$$\log \frac{1}{LC_{50}} = 4.686 + 0.015 \text{atomic rad (pm)} - 0.725 \text{Valence} + 0.134 \text{sq_sum_epsilon/N} \quad (4.1)$$

$$Q_{Loo}^2 = 0.612, R^2 = 0.726, R_{adj}^2 = 0.663, \bar{r}_m^2 = 0.497, \Delta r_m^2 = 0.166$$

The first descriptor influencing the toxicity is **Atomic rad (pm)**. This descriptor stands for the measure of the size of an atom. It typically measures the distance from the center of the nucleus to the outer most shell containing electrons. This descriptor is **positively** correlated with the response which means that with an increase of the descriptor value, the toxicity data also increases as in **nanoparticles 11 and 8** as shown in equation 4.1; in contrast, a decrease of descriptor values as represented in **nanoparticles 13 and 9** decreases the toxicity value. The reason for such a phenomenon is that due to the increase of the atomic radii, the pull of the electrons by the neutron reduces since the number of the shells in the atom increases which is also known as the shielding effect. This makes the removal of the electrons from the atom easier and leads to the formation of metal cations.

The next descriptor, **Valence** or valency of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules. This also indicates the number of valence electrons in the outer most shells. The **negative** contribution of the descriptor suggests that with an increase of valence of the metal, the toxicity decreases as in **nanoparticles 16 and 17** whereas with a decrease of valence, the toxicity increases as observed in **nanoparticles 9 and 2**. The presence of few electrons in the outer most shells makes it easier for the metal to exchange

the valence electron with its neighboring element and form a covalent or electrovalent bond with it. On the other hand, it is also possible to lose its electron easily to get converted into its cationic form and accelerate the toxic reaction.

This second-generation descriptor $(\sum \epsilon/N)^2$ stands for the total electronegativity of the atoms of the MeOx by the number of atoms present that can be computed based on the electronegativity count ($\sum \epsilon$) of the oxides.
(equation 4.2).

$$\sum \epsilon = \epsilon_{metal} * N_{metal} + \epsilon_{oxy} * N_{oxy} \quad (4.2)$$

Here, ϵ_{metal} and ϵ_{oxy} are the electronegativity count of metal and oxygen atoms, respectively and N_{metal} and N_{oxy} are the number of metal and oxygen atoms. The **positive** coefficient of the descriptor in equation 4.2 infers that an increase in electronegativity favors the rise in toxicity as in **nanoparticles 1 and 2** whereas the nanoparticles with the presence of low electronegativity have less toxicity as observed in **nanoparticles 14 and 10**. It has been observed that due to the presence of high electronegativity, the atoms pull electrons from their neighboring elements or molecules leading to the development of an electrostatic bond with them.

Model 2.2: Dark-induced toxicity

$$\log \frac{1}{LC_{50}} = 6.433 - 0.838 \chi_{ox} + 0.014 \text{ atomic rad (pm)} - 0.607 \text{ Valence} \quad (4.3)$$

$$Q_{Loo}^2 = 0.627, R^2 = 0.77, R_{adj}^2 = 0.717, \bar{r}_m^2 = 0.543, \Delta r_m^2 = 0.054$$

The first descriptor in equation 4.3 contributing to the toxicity is χ_{ox} , which stands for the oxidation number of the metal atom. This descriptor contributes **negatively** towards the toxicity which is well observed in **nanoparticles 3 and 14** where an increased oxidation number leads to a decreased toxicity value and vice versa as exemplified in **nanoparticles 1 and 2**. The oxidation number reflects the ability of an atom to acquire, donate, or share electrons, and it is quite like the valence electrons. If the oxidation number is positive, then it means that the element has lost electrons and has fewer electrons in the outer most shell. On the other hand, a negative oxidation number suggests that the element has gained electrons and has more electronegativity. As described earlier in the case of the valence descriptor, here also, the loss of electrons can accelerate the formation of a cationic form of the metal that leads to the toxic reaction.

Other descriptors having an influencing effect on toxicity to *E. coli* are discussed previously in the case of Model 2.1.

The atomic rad and valence descriptors appearing in both model 2.1 and model 2.2 play an important role in influencing the toxicity of *E. coli*. Comparing both models, we can infer that the presence of light triggers the impact of toxicity of the metal oxides. It was noted that with an increase in atomic rad and valence descriptors values the respective toxicity value ($\log 1/LC_{50}$) was found to be more in the presence of light than in the dark. The positively contributing descriptor, atomic rad, led to $\log 1/LC_{50}$ values of 4.96 and 5.79 for **nanoparticles 11 and 8**

respectively in dark conditions whereas in the presence of light, the toxicity value increases to 5.56 and 5.84 for the same respective nanoparticles with the same descriptor values. The same observation was seen in the case of negatively contributing descriptor valence; in this case, too, with the same descriptor value the $\log 1/LC_{50}$ values for nanoparticles 9 and 2 in dark conditions were 2.42 and 5.8 respectively which became 2.75 and 6.23 in the presence of light.

4.2.2. Proposed mechanism of toxicity to *E. coli* cell.

The potential mechanism of cytotoxicity of MeOxNPs to *E. coli* cells was deduced with the help of information extracted from the descriptors of the nano-QSTR models. The insights obtained from the descriptors pointed out that atomic radius, electronegativity of the metal in metal oxides, and the number of electrons in the valence shell have a substantial effect on the mortality of *E. coli*. The cluster of MeOxNPs after forming metal cations due to the loss of electrons can easily interact with the proteins present on the surface of the membrane and can enter into the *E. coli* cells. An increase in electronegativity raises the catalytic properties of the metal oxides and enhances their intervention process. After the entry of the metal oxides in the plasma region through physiochemical interaction, it breaks the Me-O bond due to the presence of acidic conditions within the cells. The detached metal cation can lead to macro molecular complexation and destroy the molecular structure (Mu et al., 2016). On the other hand, the metal cation instigates the formation of free radicals ($\cdot OH$) while interacting with the ribosomes present in the *E. coli* cell than the metals naturally do (Roy et al., 2019). The elevated levels of radicals interact with the molecules in the cells and produce more reactive oxygen species. The ROS production causes nucleoid damage that causes DNA fragmentation and death of the cell due to the blockade of the synthesis of proper protein structure. The whole process of the toxicity mechanism is well depicted in Figure 4.8

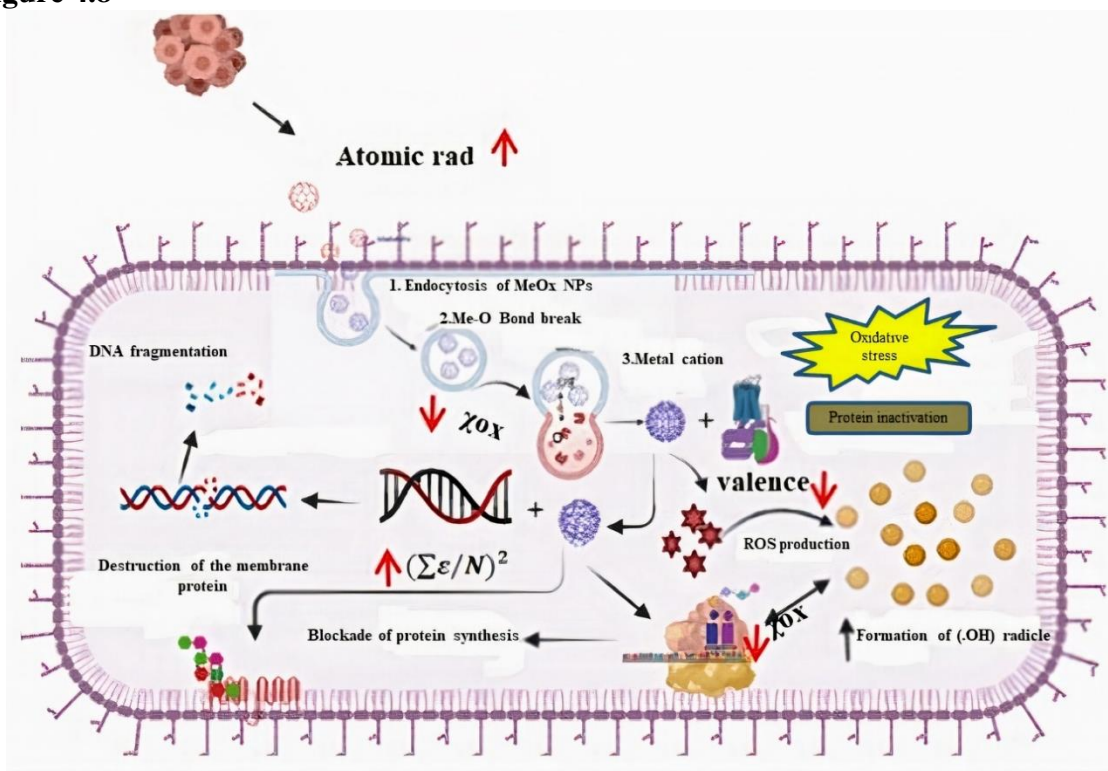


Figure 4.8 Toxicity mechanism of the MeOx NPs to *E. coli* cell (study 2).**4.2.3. Comparison with the previous work**

A comprehensive comparison is not possible with the previous work since the models were developed using different compositions of modeling sets. Pathakoti et al., 2014 used DFT (density functional theory) descriptors which require computationally intensive quantum chemical calculations whereas the work performed in the current paper used periodic table-based descriptors which are simple, easy, and do not require any expert knowledge. It is also noteworthy that in the previous paper, the authors reported only a few statistical quality metrics (SD, F, and R^2). They have kept aside a few data points for predictions and developed models using a lower number of data points. But in the present paper, we have used all available data points to avoid information loss and reported more extensively the quality metrics such as R^2 , F, Press value, R^2_{adj} , Q^2 , $\bar{r}_m^2$, Δr_m^2 metrics (**Tables 4.6 and 4.7**). In addition, the quality of predictions of the previous work is found to be poor as evidenced by high predicted residuals and negative R^2_{pred} values (not reported by the previous authors). We believe that handling a small dataset is a crucial job for a nano-QSAR modeler, and information may be lost while modeling a small dataset due to hold-out samples. This problem has been overcome in this work by using the small dataset modeler tool which uses the DCV method. By using the DCV method, we also omit the bias in description selection with a constant composition training set. Furthermore, the developed models were used to predict the toxicity of 42 untested metal oxides which do not have any reported toxicity data. The developed models successfully predicted most of the query NPs that were found to be within the AD, and with the help of the Prediction Reliability Indicator tool (PRI) their prediction quality was also tested. Additionally, we have also tried to provide a conceptual toxicity mechanism diagram for a better understanding of toxicity to *E. coli* cells (**Figure 4.8**).

Table 4.6: Comparison table for the photo-induced dataset

Statistical parameters	Number of data points (n)	R^2	F	Press value	R^2_{adj}	Q^2	$\bar{r}_m^2$	Δr_m^2	MAE (95%)	SD
Previous* work (Pathakoti et al., 2014)	Training set=13 Test set=4	0.804	20.51	-	-	-	-	-	-	0.63
Current work (model 2.1)	Whole dataset (n = 17)	0.726	11.48	9.747	0.663	0.612	0.497	0.166	0.587	-

* R^2_{pred} calculated to be a negative value.

Table 4.7: Comparison table for the dark dataset

Statistical parameters	Number of data points	R^2	F	Press value	R^2_{adj}	Q^2	$\bar{r}_m^2$	Δr_m^2	MAE (95%)	SD
Previous* work	Training set=13 Test set=4	0.87	33.83	-	-	-	-	-	-	0.48
Current work (model 2.2)	Whole dataset (n = 17)	0.77	14.48	8.494	0.717	0.627	0.543	0.054	0.482	-

* R^2_{pred} calculated to be a negative value.

4.3. Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.

Based on the toxicity data and calculated descriptors of the two datasets, we have tried to develop statistically significant QSAR models based on the regression approach to successfully predict the cytotoxicity of the *HaCat* cell line. The MLR model generated from the first dataset (**Model 3.1**) consists of only periodic table descriptors that were able to describe the cytotoxicity successfully without the need to calculate quantum chemical descriptors. On the other hand, in the case of the second dataset (**Model 3.2**), the PLS model having two latent variables was obtained using periodic table-based descriptors and physicochemical descriptors.

Model 3.1 (MLR)

$$\log\left(\frac{1}{LC_{50}}\right) = 1.46 + 0.0117\text{Amt rad (pm)} + 0.00520 \Delta \text{cond} \left(\frac{w}{mk}\right) - 0.000123 \text{M.P (k)} - 0.301 \chi_{ox}$$

$$n_{\text{training}} = 11, R^2 = 0.95, R^2_{adj} = 0.92, Q^2 = 0.83, S = 0.0953388, F = 29.03, \text{PRESS} = 0.186201$$

$$n_{\text{test}} = 5, Q^2_{F1} = 0.73, Q^2_{F2} = 0.62, \text{MAE based criteria} = \text{Good} \quad (4.4)$$

Model 3.2 (PLS)

$$\% \text{Cell death} = 113.295 + 0.039 \text{Dose} - 21.297 \text{SuM_Active_M_SM} - 15.400 \text{ cationic chrg} \\ - 1.194 \text{Heat of Fusion (kJ/mol)}$$

$$n_{\text{training}} = 94, R^2 = 0.76, Q^2 = 0.73, \text{LV} = 2, n_{\text{test}} = 23, Q^2_{F1} = 0.69 \quad (4.5)$$

Taking the *HaCat* cell cytotoxicity as the response variable, the regression models (MLR and PLS) yielded determination coefficients (R^2) of 0.95 and 0.76 respectively. The whole work was divided into two parts wherein the first part, we tried to develop regression-based models (MLR) using only periodic table descriptors and find out the features responsible for the toxicity of the particular metal oxides, and in the second part, another set of 15 metal oxides (total 117 data points) was utilized to develop a PLS model using periodic table based descriptors along with physicochemical descriptors to find out the influence of these descriptors on the toxicity to living

cells. The PLS model was constructed with all four descriptors from the obtained MLR model (**Model 3.1**). A PLS model can handle data that are strongly correlated or noisy. It finds new variables (i.e., latent variables) which are a linear combination of the original variable. **Figure 4.9** represents the scatter plot.

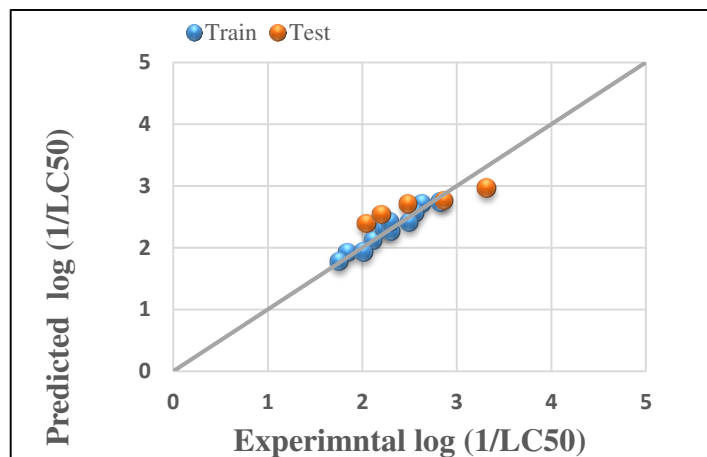


Figure 4.9 Scatter plot of Model 3.1

The models were validated according to the OECD principles; as properly validated models can be used for mechanistic interpretation (OECD, 2004) The values of all classical metrics were within the acceptable limit for both cases as given in **Table 4.8**. The descriptors obtained from both models are listed in **Table 4.9** and explained elaborately below. The AD for both the models (MLR and PLS) was checked using the Leverage approach (Gramatica, 2007) and DModX, i.e., distance to model in the X-space with SIMCA 16.0.2 software respectively. No compound was outside the AD as per William's plot **Figure 4.10** and DModX approach for the training and test set **Figure 4.11** and **4.12**.

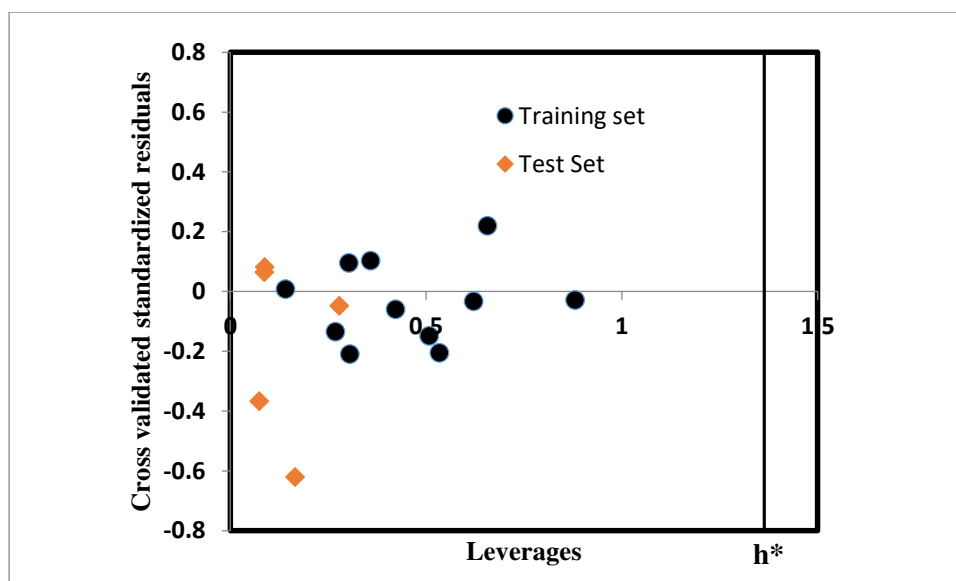
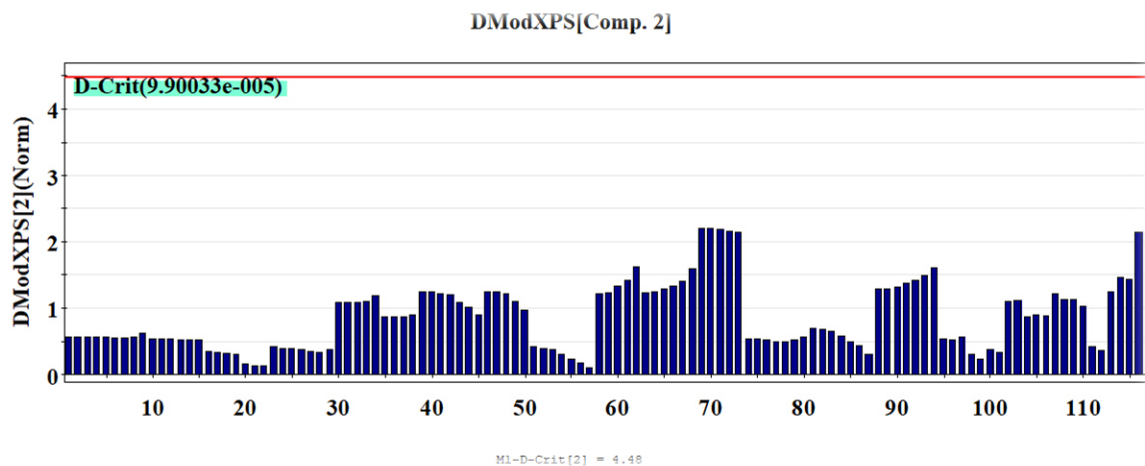
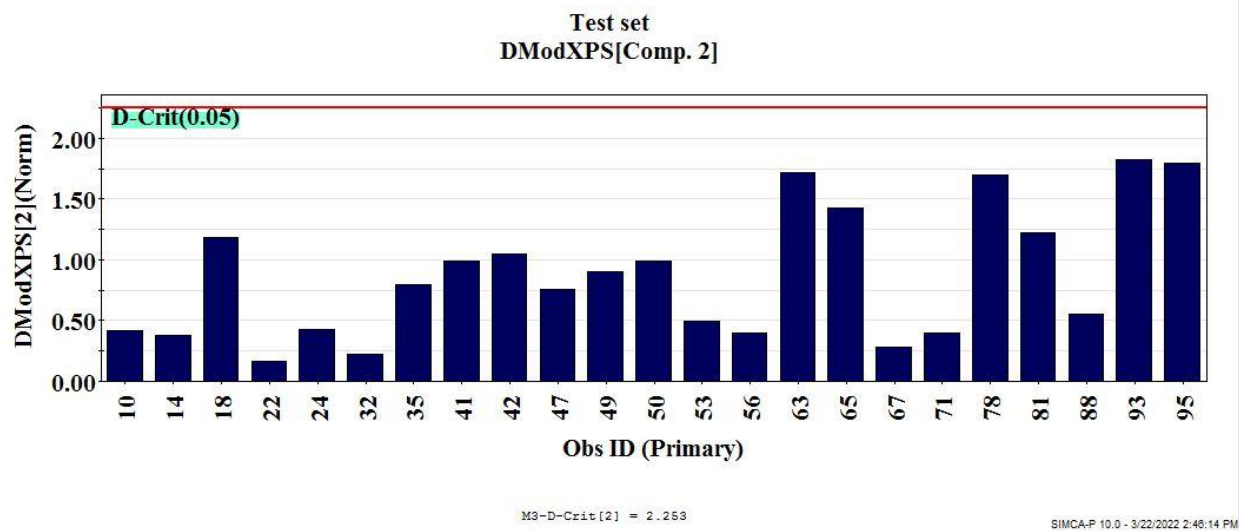


Figure 4.10 William's Plot of the MLR model (Model 3.1).**Figure 4.11** Applicability domain of model 2 for the training set (PLS model) (Model 3.2)**Figure 4.12** Applicability domain of model 3.2 for the test set (PLS model)

4.3.1 Interpretations of different plots

The variable importance plot (VIP) of the PLS model (Model 3.2) highlights the most influential descriptors of the MeO_x contributing to the cytotoxicity of the *HaCat* cell. In the case of the PLS model, the most contributing descriptors as per the plot were dose, cationic charge, SuM_Active_M_SM followed by descriptor heat of fusion **Figure 4.13**

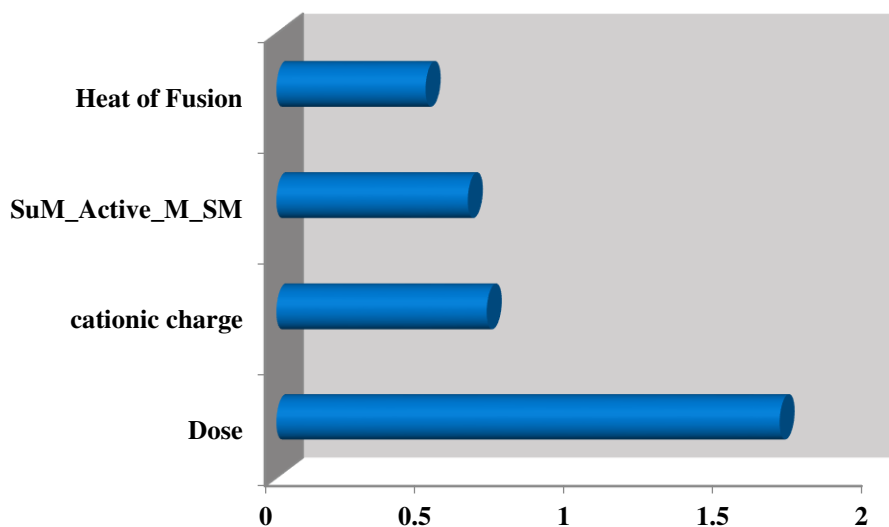


Figure 4.13 Variable importance plot of model 3.2 (PLS model)

The loading plot gives detailed information on the X-variable and Y-variables of the PLS plot as represented in **Figure 4.14**. The loading plot was developed with two PLS components. From **Figure 4.14** we can infer that the dose of the MeO_x plays a dominant role in determining the toxicity to *HaCat* cells. While evaluating the plot, it was observed that the descriptors located far away from the origin have more impact on the Y response of the model.

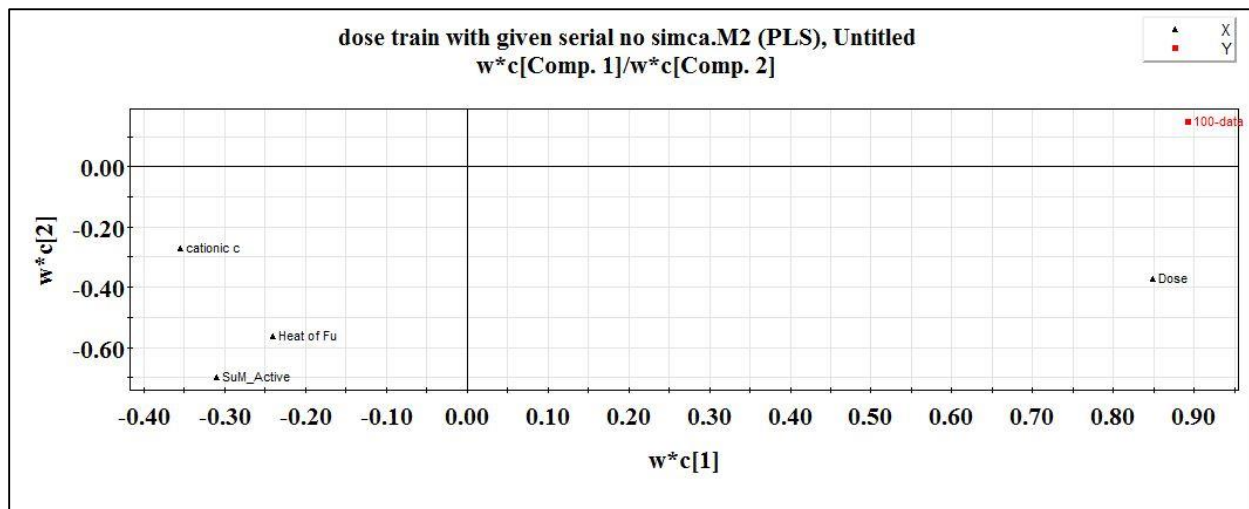


Figure 4.14 Loading plot of the PLS model (Model 3.2)

The coefficient plot shows the contribution of the modeled descriptors. Except for the dose descriptors, all other descriptors (heat of fusion, SuM_Active_M_SM, Cationic charge) were found to contribute negatively to the PLS model of the 15 MeO_x NPs dataset (**Figure 4.15**).

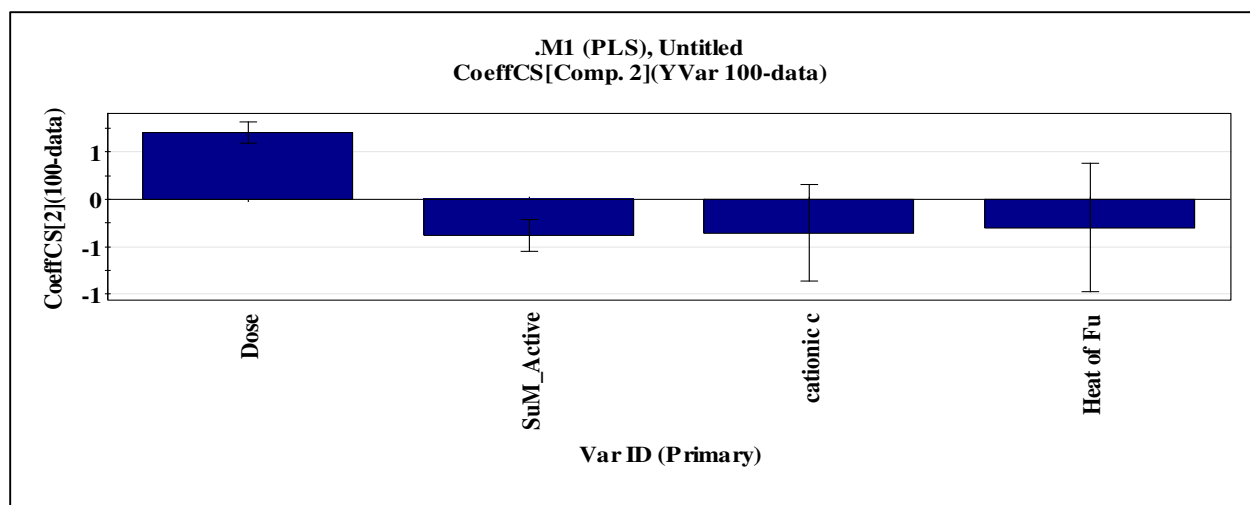


Figure 4.15 Coefficient plot of the PLS model (Model 3.2)

Randomization of the model authenticates the robustness of the developed model (Model 3.2) and ensures that it is not developed by chance correlation. The randomized model is developed by multiple shuffling of X and Y variables. Here, 100 permutations were done to generate the model. The developed random should have a poor statistic to show it is a nonrandom model. The correlation between original Y-vector and permuted Y-vector versus cumulative R_y^2 and cumulative Q_y^2 plots are represented in **Figure 4.16**. The plot shows that the model is nonrandom as well as robust since $R_y^2 \text{ intercept} = -0.0211$ and $Q_y^2 \text{ intercept} = -0.179$, and one can easily predict the cytotoxicity of the MeO_xNPs from the developed PLS model.

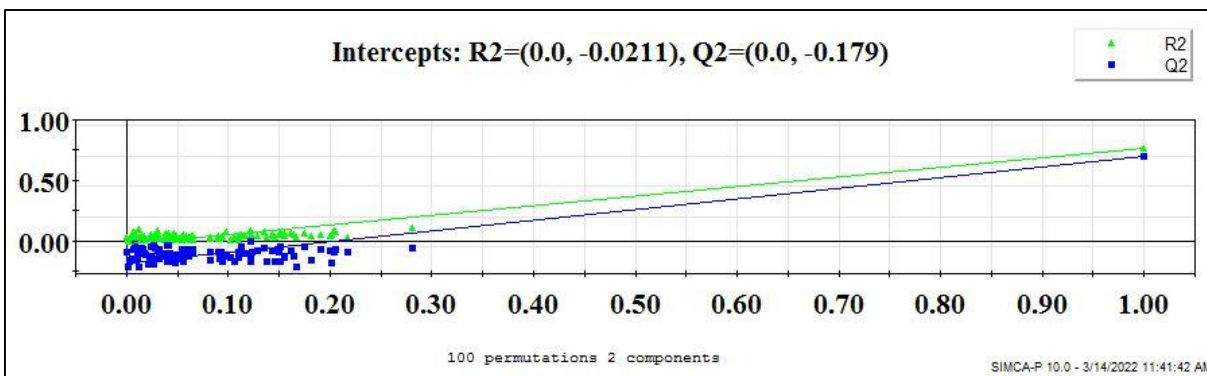


Figure 4.16 Randomization plot of the PLS model (Model 3.2)

Table 4.8. Statistical quality and validation parameters obtained from the developed MLR and PLS models (3.1 and 3.2)

Types of models		Descriptors	Division	Training set statistics					Test set statistics				
				Model R^2	Model $Q^2_{(LOO)}$	$\overline{r^2_{m(LOO)}}$	$\Delta r^2_{m(LOO)}$	F value	R^2_{pred} or Q^2F_1	Q^2F_2	CCC	$\overline{r^2_{m(test)}}$	$\Delta r^2_{m(test)}$
MLR (Model 3.1)	Case 1 (16 MeOx NPs)	4 (four)	Train: 11 test: 5	0.95	0.83	0.773	0.037	29.03	0.73	0.62	0.79	0.271	-0.505
	Case 2 (15 MeOx NPs)	4 (four)	Train:94 test: 23	0.76	0.73	0.63	0.038	71.69	0.69	0.63	0.81	0.61	0.123
PLS (Model 3.2)	P1 (Case 2)	2 (two)	Train:94 test: 23	0.76	0.73	0.63	0.038	71.69	0.69	0.63	0.81	0.61	0.123

Table 4.9. A detailed description of the obtained descriptors from two developed MLR models (Model 3.1 and Model 3.2).

Serial No.	Descriptors	Description
Model 3.1:		
1.	Atomic rad (pm)	This descriptor stands for the measure of the size of the atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons.
2.	Thermal cond(w/mk)	The transfer of energy is due to the random motion of molecules across a temperature gradient.
3.	Melting point	The temperature at which a matter changes its state from solid to liquid.
4.	χ_{ox}	It represents the oxidation number of the metal atom.
Model 3.2:		
Serial No.	Descriptors	Description
1.	Heat of fusion	The heat of enthalpy is the change of enthalpy of a fixed amount of metal resulting from the supply of energy typically in the form of heat
2.	Dose	A quantity of a substance taken or recommended to be taken at a particular time.
3.	SuM_Active_M_SM	Summation of several metals and semi-metals in the nanoparticles.
4.	Cationic charge	Ion which bears a positive charge formed when an atom loses its electron

4.3.2 Descriptor description

1st case study (Model 3.1):

The 1st generation periodic table-based descriptor **atomic rad (pm)** stands for the measurement of the distance between the nucleus and the valence electron. The **positive** coefficient of the descriptor suggests that with an increased atomic radius, the cytotoxicity increases in the *HaCat* cell line as observed in nanoparticles **WO₃** (where atom rad is 210 (pm) and $\log(1/LC_{50})$ is 2.56) and vice versa in the case of the nanoparticle **Al₂O₃** (atom rad 125 (pm) and $\log(1/LC_{50})$ is 1.85). This happens because as the atomic radius increases, the valence electron in the outermost shell keeps moving away from the nucleus which means that more atomic shell is being added. The addition of the number of shells gives a shield to the valence electron in the outermost shell. The electron in the outermost shell does not feel the counter pull of the nucleus and the presence of the electron in the inner shell repels the similarly charged electron which makes it easier to isolate and move out from its shell (**Figure 4.17**). The removal of the electron converts the metal into its cationic form. It is this cationic form of the metal that initiates the production of oxidative radicals in the cell which is detrimental to cell viability. Studies (Sun et al., 2019; Roy et al., 2019) on ROS- mediated mechanisms showed that cellular uptake of cationic NPs accelerates depolarization of the cell

membrane and facilitates calcium channel opening that stimulates mitochondria to generate more oxidative stress.

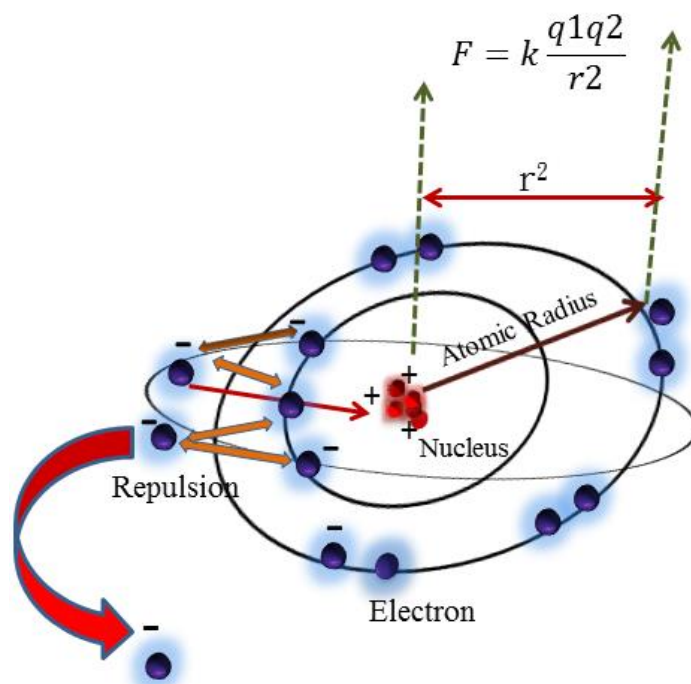


Figure 4.17 Electron removal as the atomic radius increases.

Another intrinsic property of metals is **thermal cond (w/mk)**, defined as the transfer of energy due to random motion of molecules across a temperature gradient. It is the measurement of the amount of heat that passes across per unit area of the material per unit time. As per the band theory, the atoms in the metal crystal are close to each other which leads to the overlapping of molecular orbitals on each other. This phenomenon in **Figure 4.18** suggests the presence of a large number of electrons in a small piece of metal and the closeness prefers them as a band (Walker et al., 2007). As per the band theory, there are two types of bands, one with partially filled electrons known as the conduction band, and the other with filled electrons known as the valence band. The ability of electricity to pass through the metal depends on the proximity of the valence and conduction bands. The **positive** regression coefficient of the descriptor suggests that with an increase in thermal conductivity, the cytotoxicity also increases as in nanoparticles **CoO** (thermal cond. is 100 (w/mk) and $\log(1/LC_{50})$ is 2.85), the reverse happens in nanoparticle **Al₂O₃** (thermal cond. is 2.37 (w/mk) and $\log(1/LC_{50})$ is 1.85). This happens because with the increase of thermal conductivity, the distance between the two bands decreases, and it becomes easier for the electron to move faster from the valence band to the conduction band. This redox potential occurring naturally in the cell may be the reason for the toxicity of nanoparticles.

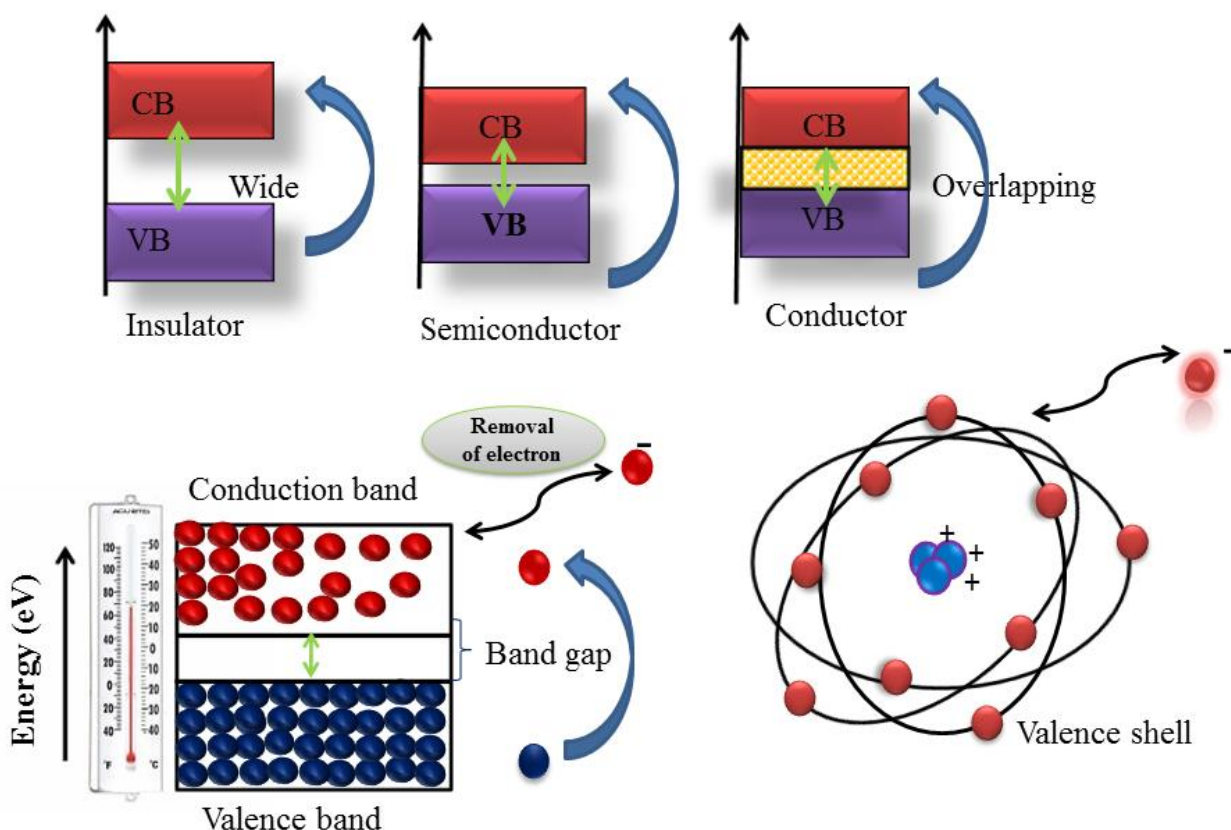


Figure 4.18 Electron removal through Band theory.

Another property of metal is **Melting point (MP)**, defined as the temperature at which a matter changes its state from solid to liquid. The melting point of any metal is primarily related to the bond strength. Materials having weaker bonds between atoms tend to have a low melting point. The **negative** coefficient of the descriptor suggests that with a decrease in the melting point, the toxicity of the nanoparticle increases as observed in **TiO₂** nanoparticles where the M.P is 1941 (k) and $\log(1/LC_{50})$ is 1.76. On the contrary, **V₂O₃** nanoparticles with less MP (30.7 (k)) have a $\log(1/LC_{50})$ value of 2.24. This is because the melting point is associated with the solubility of the metal. With the decrease of MP, the solubility increases. The solubility determines its transport to the active site with the cell thus causing toxicity. The low melting **MeO_x** easily dissociates when coming in contact with the body temperature **Figure 4.19**. The presence of temperature makes the molecules vibrate faster and break their intra-molecular bonds. Since nanoparticles have fewer bonds compared to their bulk counterparts, it becomes easier to break them.

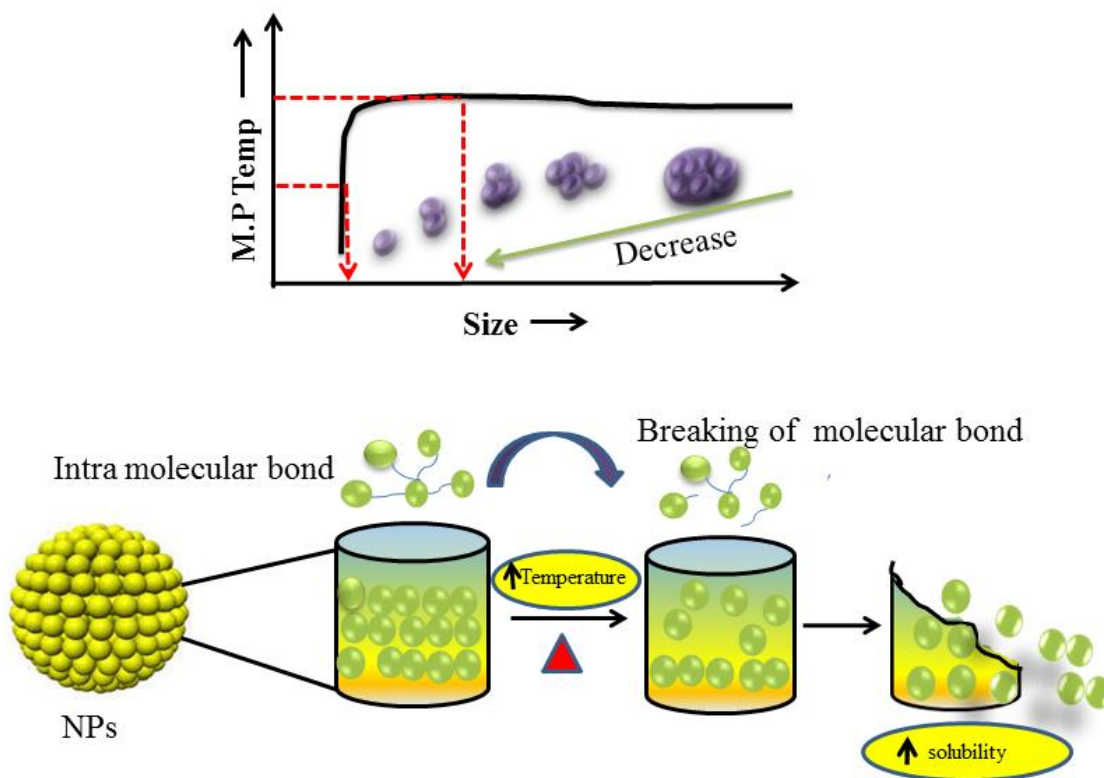


Figure 4.19 Increase of solubility with of melting point.

The periodic table descriptor, χ_{ox} stands for the oxidation number of the metal. The oxidation number reflects the hypothetical charges in an atom. Electronegativity indicates the distribution of electrons in a molecule, it can also determine the oxidation number of the atom. The oxidation number indicates the oxidation power of the metal in a particular environment. The transition metals may exist in different oxidation states as they have partially filled d- and f- shells (**Figure 4.20**). The atom with fewer valence electrons will have less oxidation state and atoms with less oxidation state are more harmful than the higher or stable oxidation state due to their tendency to become oxidized (Sengul and Asmatulu, 2020). The atom bearing a negative oxidation number consists of higher electronegativity whereas an atom with a positive oxidative number will have low electronegativity. The electronegativity is responsible for that catalytic property of the cationic form of the metal. The **negative** coefficient of the descriptor indicates that with a decrease of the descriptor value, the electronegativity increases and hence it increases the cytotoxicity to the *HaCat* cell as in nanoparticle TiO_2 (χ_{ox} is 4 and $\log(1/\text{LC}_{50})$ is 1.76) and vice versa in case of nanoparticle CoO (where χ_{ox} is 2 and $\log(1/\text{LC}_{50})$ is .83). The low oxidation metals are more toxic due to their tendency to get oxidized to a higher oxidation state.

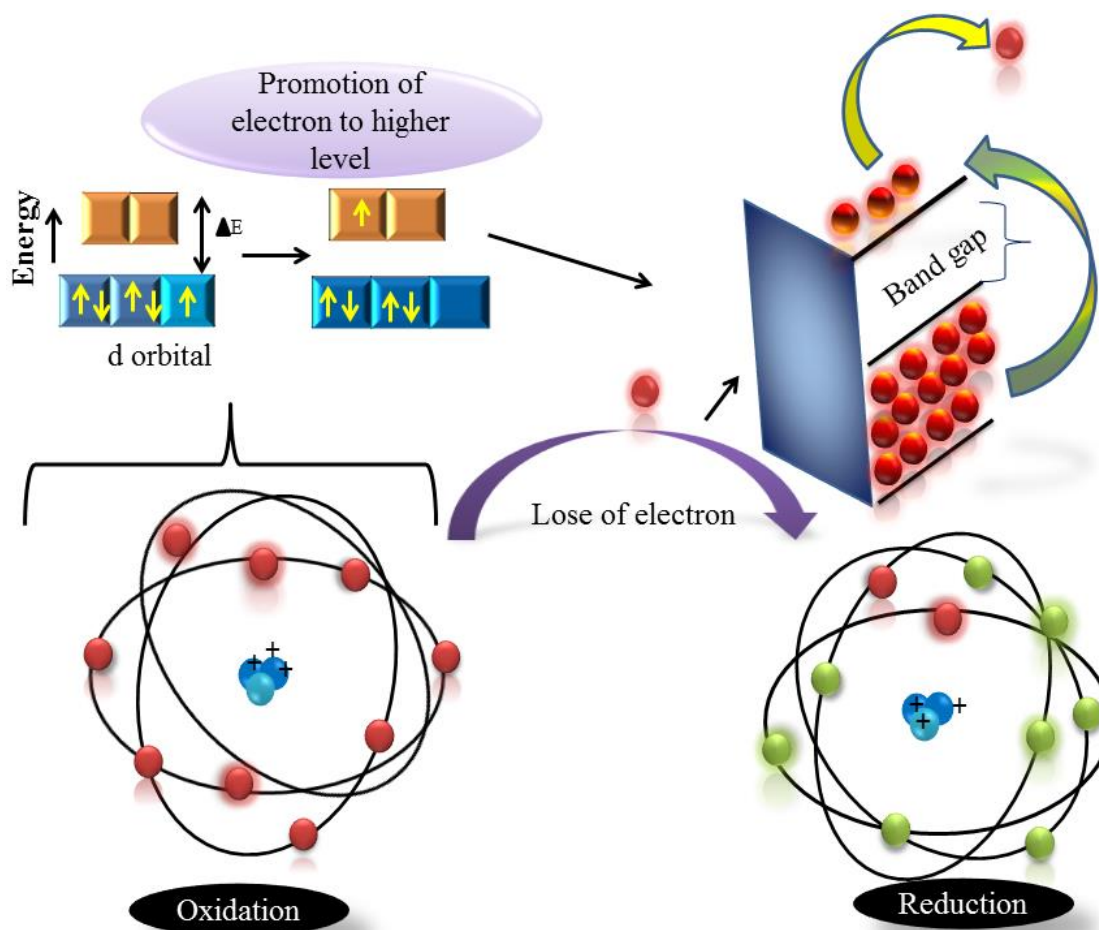


Figure 4.20 Electron removal with electronegativity.

2nd Case Study (Model 3.2):

The periodic table descriptor, the **heat of fusion** (ΔH_f) also known as the heat of enthalpy is the change of enthalpy of a fixed amount of metal resulting from the supply of energy typically in the form of heat. Latent heat is the amount of heat added or removed from the system to cause a phase change in the system **Figure 4.21**. The latter breaks the intermolecular forces and helps the system to expand by providing the necessary energy. The temperature at which it takes place is called the melting point. Based on the negative coefficient of the descriptor, a decrease in the heat of fusion leads to an increase in cytotoxicity and the percentage of cell death. In the nanoparticle **Sb₂O₃**, the heat of fusion was 19.87 (KJ/mol) whereas the % of cell death was 15.3%. On the other hand, the opposite was observed in nanoparticle **In₂O₃** where the heat of fusion was 3.26 (KJ/mol) and %cell death was 72.7 %. The heat of the fusion descriptor is inversely proportional to the solubility which means with the decrease of the heat of fusion the solubility increases thus increasing the toxicity.

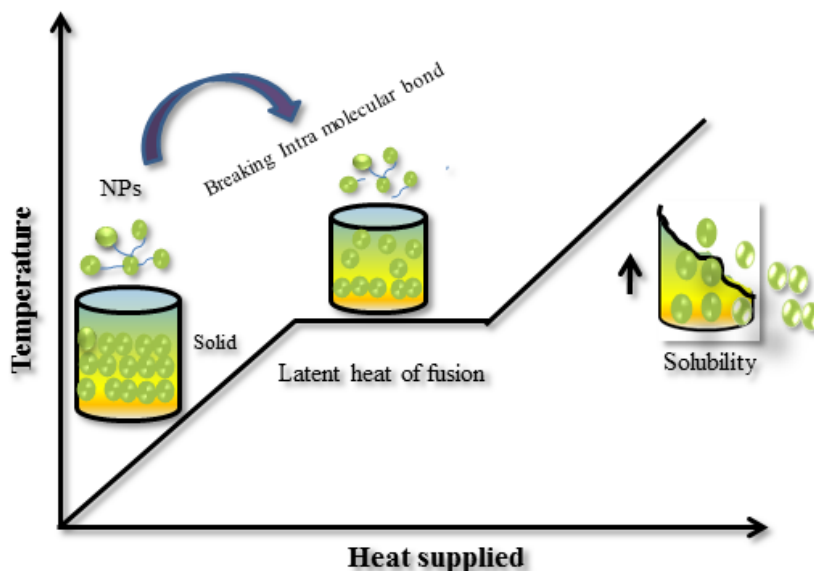


Figure 4.21 Mechanism of the heat of fusion

Another descriptor, influencing the cytotoxicity is the **Dose**. The knowledge of dose is very critical. Drugs that are beneficial at a low dose can be detrimental at a high dose. The **positive** contribution of the descriptors indicates that the increase of the dose of nanoparticles with the same metal core can decrease the cell viability as observed in the **nanoparticle SnO₂** (Dose 1500 and % Cell death is 82). On the contrary, with a decrease of the dose with the same metal core, the cell viability increases as seen in the **nanoparticle Al₂O₃** (Dose 50 and % Cell death is 5.1). Rizk et al. (2017) demonstrated that oxidative stress and liver functions were greatly dependent on the dose and time exposure of the cell to the nanoparticles. He also mentioned that genetic disturbance occurs at a higher dose and for a long duration. Studies have also supported the hypothesis that NPs increase free radicals inside the cell in a dose-dependent manner. The catalytic process can initiate the free radical generation until they are removed or degraded from the body.

SuM_Active_M_SM is the summation of several metals and semi-metals in the nanoparticles. The toxicity phenomenon of metallic NPs is complicated and depends on various physicochemical parameters. These NPs with their unique nature may have a different mechanism and may alter the amount of toxicity. It was also observed that different nanomaterials stimulate certain mechanisms. The presence of metal does not clearly state the increase or decrease of toxicity. It depends on solubility, oxidation number, etc. This descriptor contributes **negatively** to the toxicity which can be interpreted as the increase of several metals or semimetal in the metal oxides leading to an increase of the molecular size. The monovalent metal oxides are smaller in size than the divalent or trivalent metal oxides. Therefore, with the increase of metals in metal oxides, the cytotoxicity increases as in nanoparticle Sb₂O₃, the descriptor value is 2 and % cell death is 11.6% whereas, on the other hand, nanoparticle NiO having descriptor value 1 have %cell death equal to 87.4 %. It can be assumed that an increase in molecular size increases the size of the MeOx NPs and hence interferes with the penetration of the NPs through the pores of the HaCat cell leading to a less toxic effect.

The **cationic charge (Cat charge)** descriptor corresponds to ions that bear a positive charge formed when an atom loses its electron. The loss of electrons results in an overall positive charge. Cationic charges are related to cytotoxicity due to their electrostatic interaction with the negatively charged membrane. The **negative** coefficient of the descriptor validated that an increase of the cationic charge decreases the cytotoxicity as in nanoparticle **ZrO₂** (where the cationic charge is 4 and % Cell death is 2.9). On the other hand, the decrease in cationic charge to a certain level increases the toxicity towards the *HaCat* cell line as observed in nanoparticle **ZnO** (cationic charge is 2 and % Cell death is 99.07). Electronegativity is the tendency to attract electrons towards itself. Weak electronegative atoms tend to lose electrons and form cations whereas strong electronegative atoms acquire electrons and become an anion. Electronegativity decreases as the electrons in the valence cell decrease. The decreased electronegativity makes the atom electropositive. The electropositive atoms have high ionization potential, i.e., they require more energy to remove the electron from the valence shell. As the cationic charge increases, the ionization energy moves from 1st ionization potential to a higher form. And we know that the energy required by the 2nd and 3rd ionization potential is much higher than the 1st ionization potential. Electronegativity is the function of nuclear charge, i.e., the effective positive charge from the nucleus that an electron experiences. A large effective nuclear charge corresponds to a strong attraction between electrons and nuclei. The effective nuclear pull increases as the number of shells decreases as witnessed in the atomic radius descriptor in case 1. So, it becomes tougher for the electron to leave its valence shell against the strong nuclear pull **Figure 4.22**.

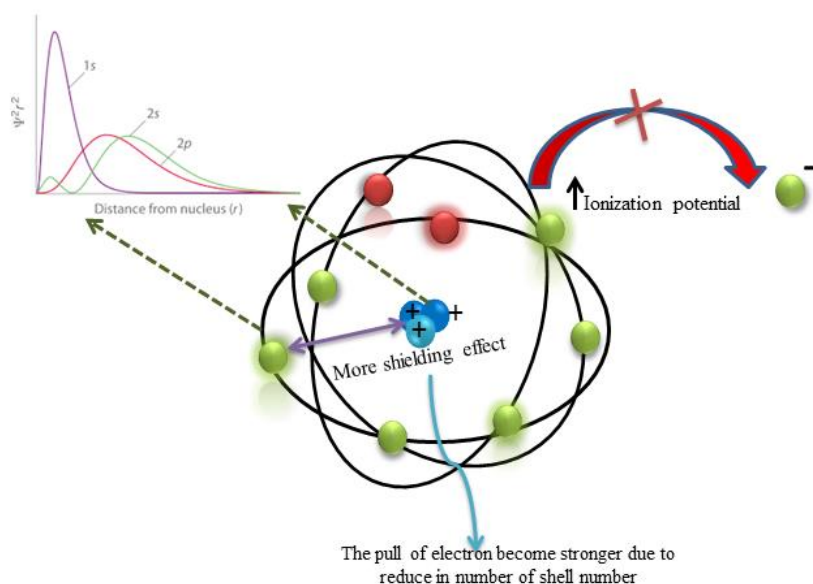


Figure 4.22 Mechanism of reduced toxicity due to increase the cationic charge.

4.3.3 Mechanism of the toxicity

The increase in production and use of NPs enhances their exposure to humans and the environment. Humans are more prone to NP exposure through inhalation, ingestion, dermal, or injection. When NPs come in contact with a human, they cause toxicity via various mechanisms. Features obtained from the modeled descriptors can well explain the mechanism of toxicity. Periodic table-based descriptors such as ionic charge affect the interaction of NPs with cell membranes having a greater

impact on the induced toxicity mechanism. The charged surface of NPs can be described by electronegativity, thermal conductivity, and atomic radius of MeO_x NPs. The presence of these descriptors promotes the formation of the cationic form of the metal. The physical interaction of NPs with the cell membrane is mainly governed by the cationic charge of the NPs. Due to the presence of a cationic charge, the NPs can easily enter the cell through endocytosis. The electrostatic interaction of the NPs with the cellular membrane produces cellular responses that are identified as a paradigm for cytotoxicity. The presence of a high electronegative ion also increases the catalytic property of the metal cation. Thus, electronegativity increases the toxicity of nanoparticles through the Haber–Weiss–Fenton cycle (Koppenol, 2002). Oxidative stress depends on the concentration of the NPs or in other words on the dose of the NPs. The VIP plot for the PLS model also pointed out the importance of dose descriptors in influencing cytotoxicity. The intracellular generation of oxidative stress in the targeted cells may form a malondialdehyde molecule that triggers the deterioration of the lipid membrane (lipid peroxidation) (Wang et al., 2017). Excessive production of ROS by NPs can damage RNA and DNA (nuclear components) and interfere with cellular signaling pathways, which may alter the gene expression that ultimately stops the cell growth mechanism **Figure 4.23**. The presence of metallic NPs also damages cells through cell membrane permeabilization in a spatial manner. The presence of metallic NPs binds with the cell wall and leads to cell death. Co NPs (Cobalt nanoparticles) produce DNA damage that leads to ataxia-telangiectasia mutated (ATM) protein activation that further increases the phosphorylation of p53 and Rad 51 protein expression. The heat of the fusion and melting point descriptors increase the solubility of the MeO_x NPs. The NPs release metallic ions to the cytoplasm that is directly proportional to the dissolution rate of NPs in a particular size range.

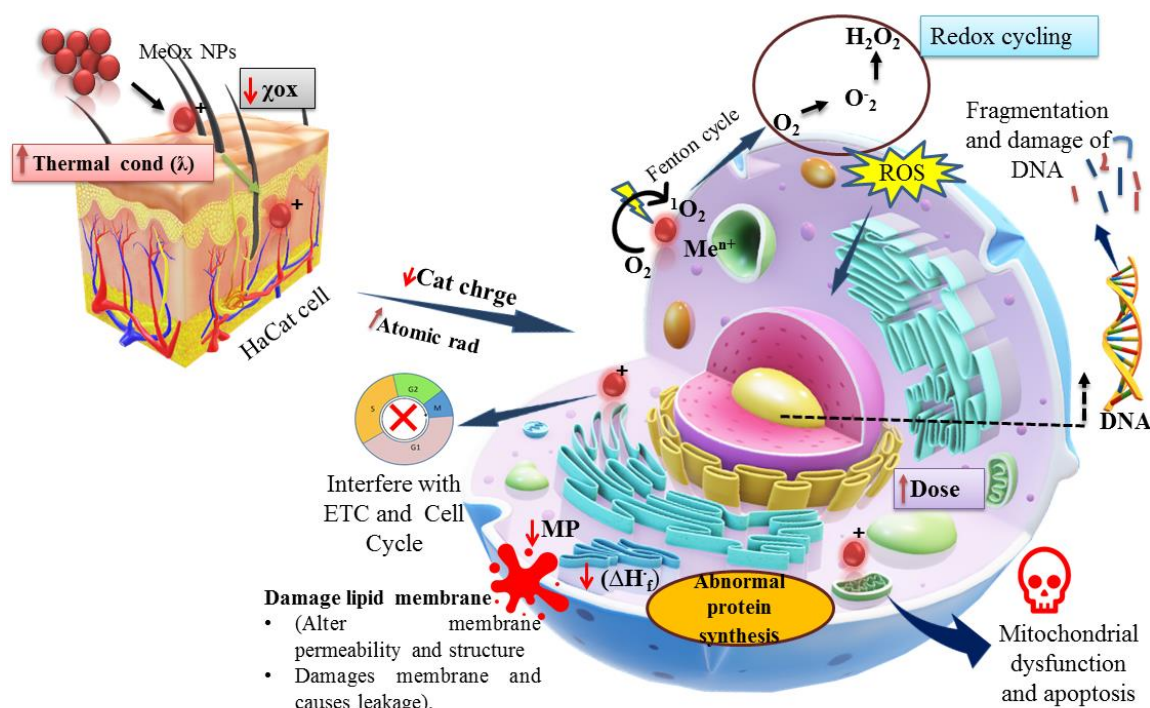


Figure 4.23 Mechanism of MeO_x NPs Toxicity via various pathways (study 3).

*ETC- Electron Transport Chain

4.3.4 Comparison with the previous work

This study focused on identifying the variables that govern the toxicity of MeO_x NPs to *HaCat* cells. Here, two datasets of MeO_x NPs were used having cytotoxicity to *HaCat* cells. In the study by Puzyn et al. (2011), quantum chemical descriptors (**Table 4.10**) were used that require a computational chemistry background to perform, and these are also computationally challenging. Generally, these are difficult to use due to their mathematical and computational complexities. In every case where a different metal oxide is used, energy minimization and changing of the orientation of the MeO_x are required before calculating the quantum descriptor. The use of quantum descriptors also has some drawbacks. For instance, determining the quantum descriptors for mixture nanoparticles can be challenging, but it can easily be accomplished with periodic table-based descriptors. The quantum chemical descriptor has failed to address directly the bulk effect of nanoparticles (Karelson et al., 1996). The periodic table descriptor obtained in our model corroborates with the findings of Puzyn et al. (2011) obtained from quantum chemical descriptors. It is worth mentioning here that fundamental properties like cation charge formation of metal oxide, LUMO, and HOMO energies can be easily explained from the information acquired from periodic table descriptors in no time. Additionally, such information is independent of size range and does not require minimization of energy for descriptor calculation. The other work performed by Choi et al. (2019) used two cell lines in the same model. According to the OECD guideline, to perform any QSAR study the experimental condition must be the same. So, here we have selected the data corresponding to the *HaCat* cells (only one cell line) with a specific assay method (CytoTox-Glo). The developed models were found to be statistically significant. We have also changed the endpoint to % cell death for a better understanding of metal oxide toxicity. A strict comparison was not possible in the case of 2nd dataset due to differences in model development procedure, the use of different cell lines and different assay methods.

Table 4.10. Comparison with the previously published work (Dataset 3.1)

Statistical parameters	Number of data points (n)	R^2	Q_{LOO}^2	F value	$\bar{r}_m^2$ Train	Δr_m^2 Train	R_{adj}^2	Q_{F1}^2	Q_{F2}^2	$\bar{r}_m^2$ Test	Δr_m^2 Test
Previous* work (Quantum descriptors)	Training set=10 Test set=8	0.93	0.86	44.6	-	-	-	0.83	-	-	-
Current work (model 3.1) (Periodic table descriptors)	Training set=11 Test set=5	0.95	0.83	29.03	0.773	0.037	0.92	0.73	0.62	0.27	-0.505

4.4. Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPs) used in nanopesticides to BEAS-2B and RAW 264.7 cells.

Periodic table descriptors were used for modeling four datasets (M 4.1-M 4.4) comprising 10, 9, 10, and 12 MeOx ENPs respectively. First, the small dataset modeler software (Ambure et al. 2019) was employed to find out whether the datasets were modelable or not. After developing statistically significant models with the small dataset modeler software, we have selected the modeled descriptors to perform read-across to fill the data gap. Furthermore, the inter-endpoint cytotoxicity comparison between BEAS-2B and RAW 264.7 cells was performed to evaluate the toxicological characteristics of MeOx ENPs. The models were validated according to the OECD principles (OECD, 2007). The statistical results are given in **Table 4.11**, and the values are within the acceptable criteria. The AD of the four PLS models was checked using DModX, i.e., distance to model in the X-space with SIMCA 16.0.2 software. At a 99% confidence level, the AD test was performed and all the compounds were found to be within the critical value. No compound was outside the AD as per the AD plot in **Figure 4.24**.

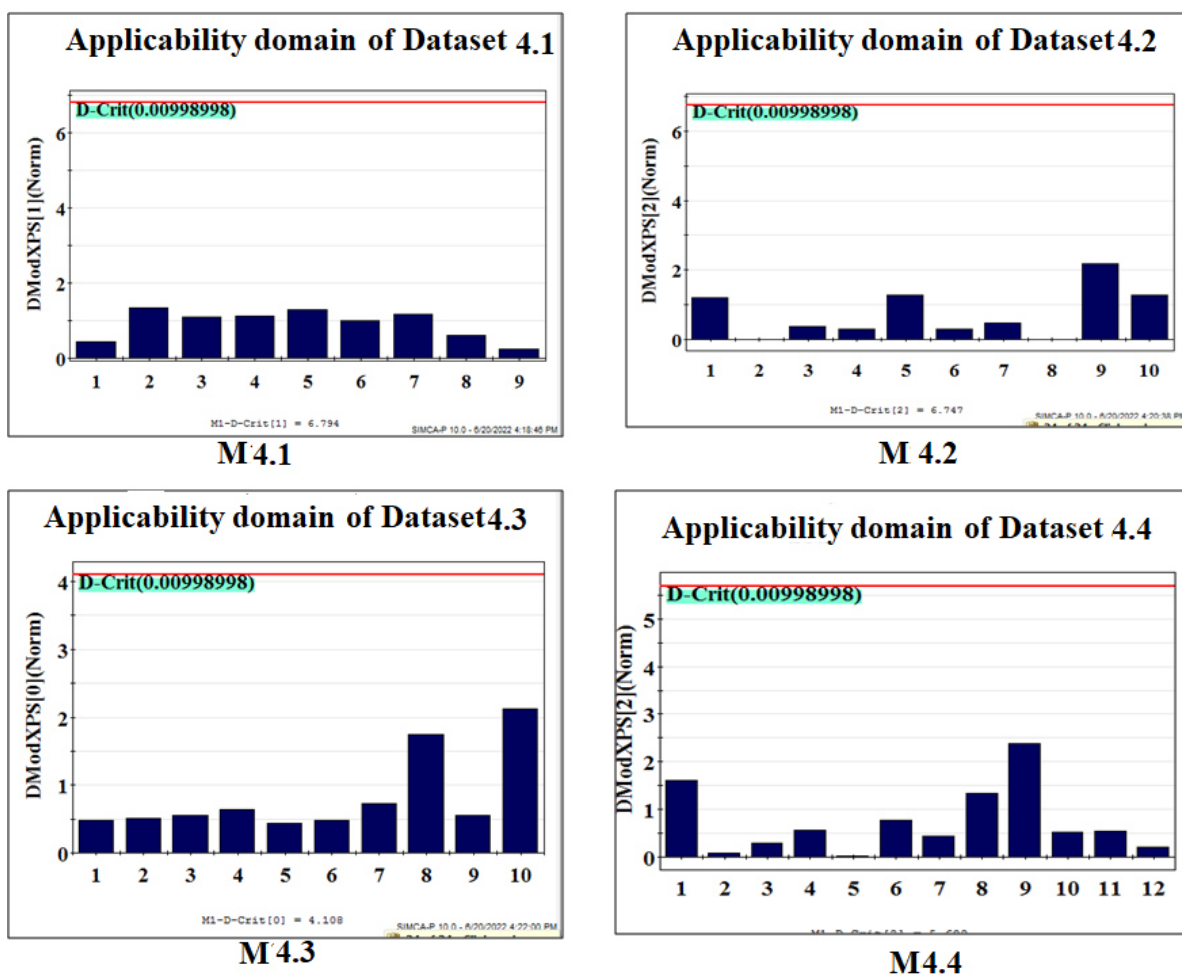


Figure 4.24 Applicability domain of the PLS models (M 4.1-M 4.4)

The order of importance of the descriptors of four PLS models towards the cellular uptake was determined from the Variable Importance Projections (VIPs) (**Figure 4.25**). The descriptors are ranked in decreasing order according to the VIP values. The developed PLS models were also validated using randomization tests through randomly shuffling (100 permutations) response data using SIMCA 10.0. software. Each random shuffle subsequently generates a new set of Q^2 values which are plotted against the correlation coefficient with the original and permuted Y values. The Q^2 and R^2 values should yield an intercept that is below zero, indicating the robustness and applicability of the model **Figure 4.26**. The obtained result showed that the models were not obtained by chance.

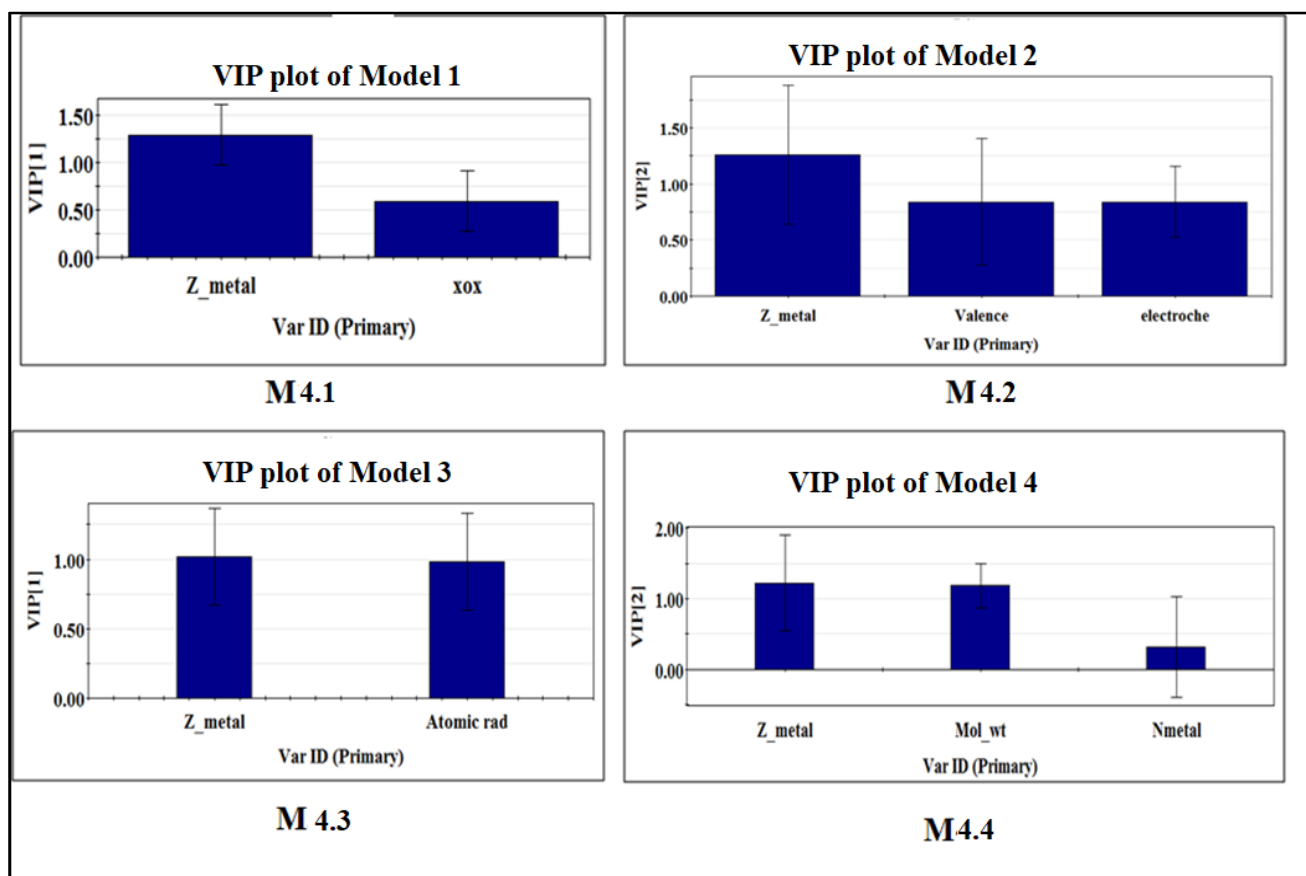


Figure 4.25. Variable Importance Projections (VIPs) of the PLS models (M 4.1-M 4.4)

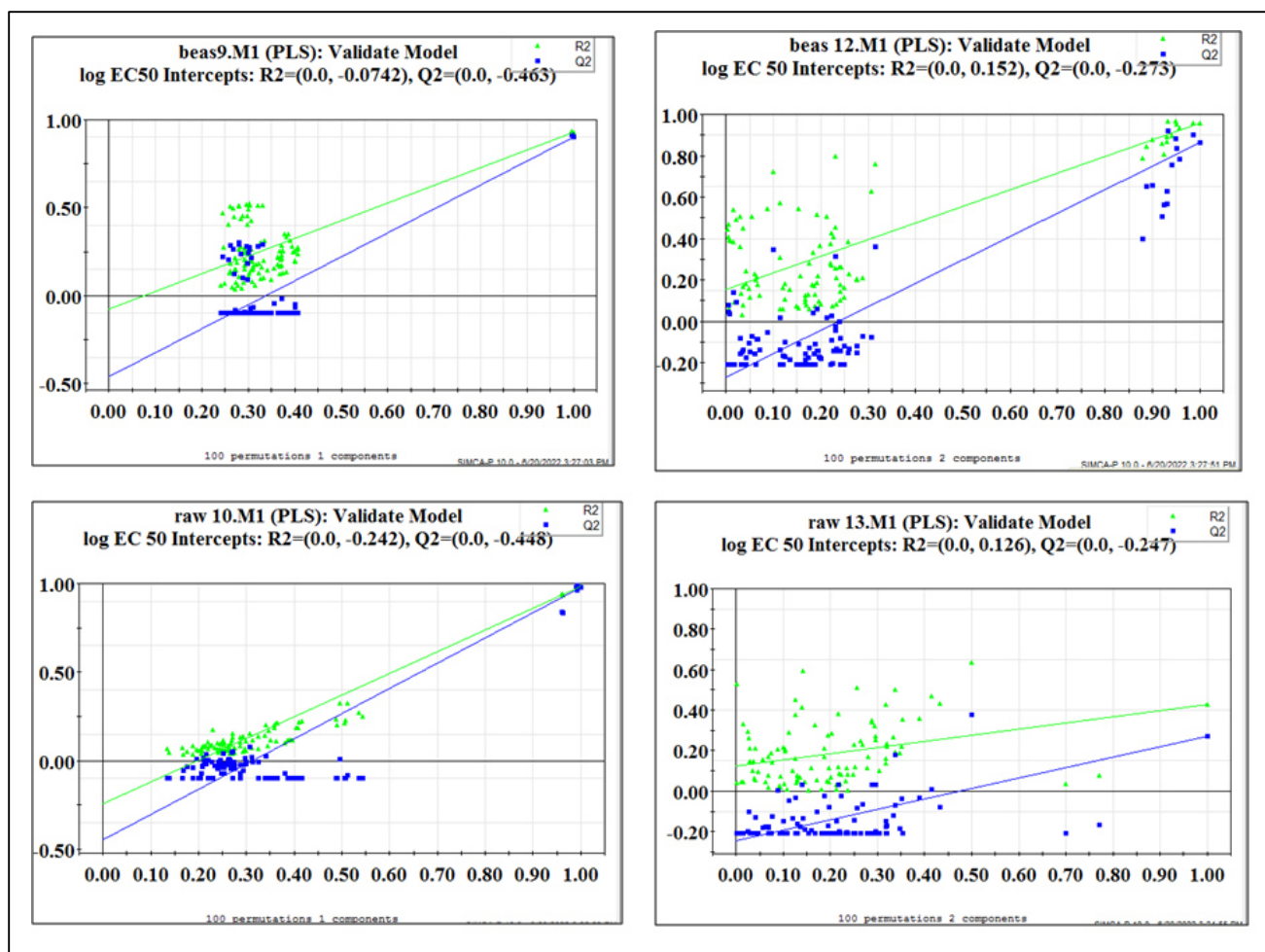


Figure 4.26 Y randomization plots of the PLS models (M 4.1-M 4.4)

The loading plots were checked for the developed PLS models. The loading describes the relationship between the descriptors, i.e., X variable with that of the response i.e., Y. Through the loading plot (**Figure 4.27**), the impact of descriptors on the response can be easily understood. The descriptors with a higher impact on the developed model will be situated far away from the origin.

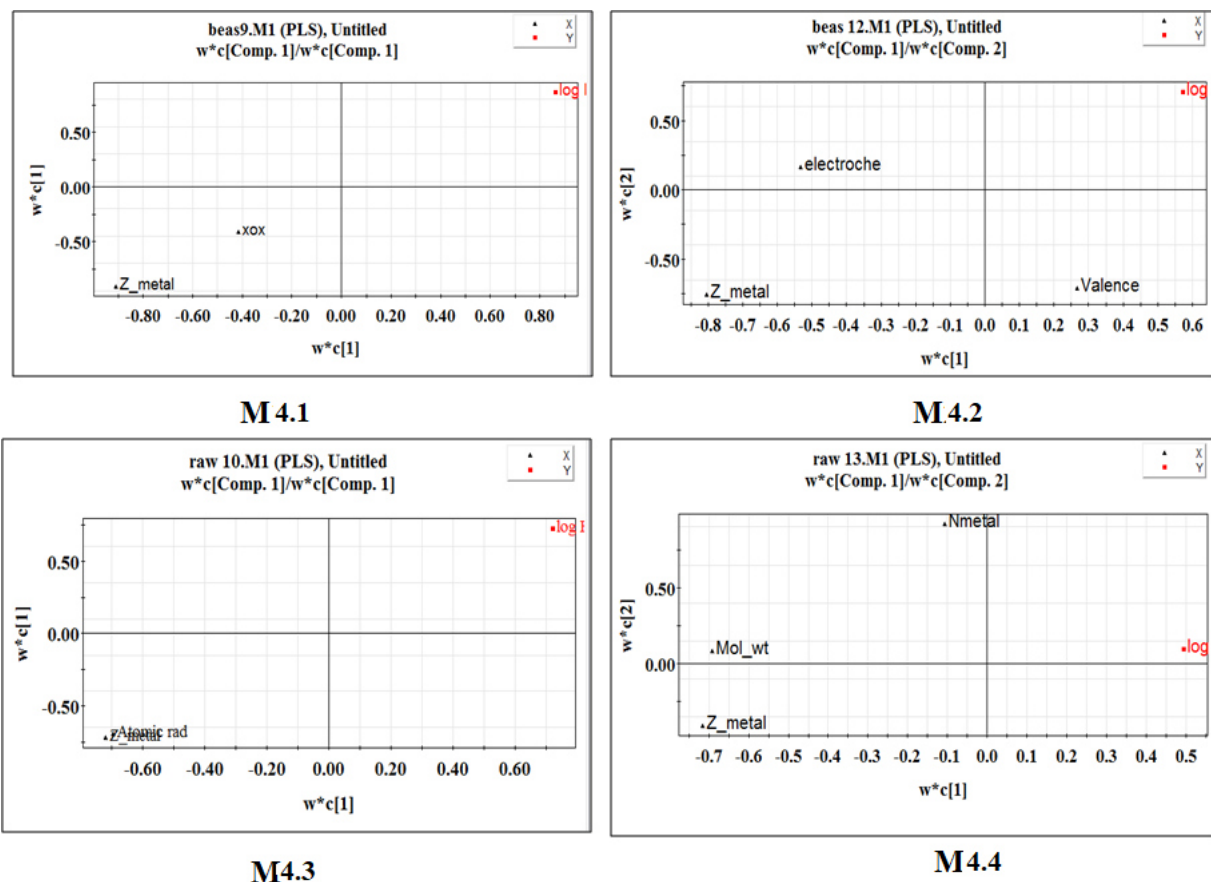
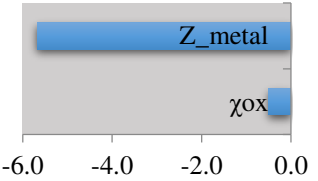
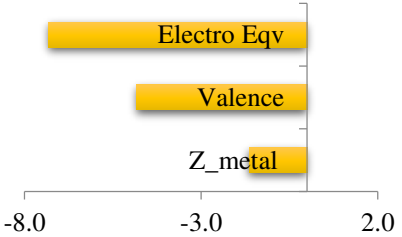
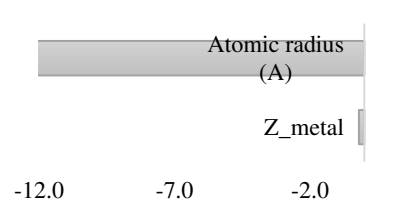
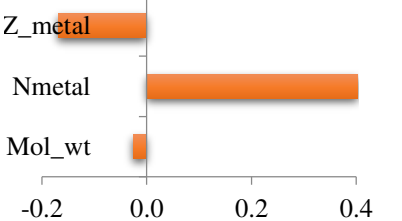


Figure 4.27. Loading plots of the PLS models (M 4.1-M 4.4)

4.4.1. Prediction by the Small Dataset Modeler

PLS regression models were developed using the four cytotoxicity datasets of BEAS-2B and RAW 264.7 cells performed using the two assay methods (MTS and ATP). The small dataset modeler version 1.0.0 software was employed to check the modelability of all the datasets. The availability of limited MeOx NP data restricted data division. The whole dataset was used for modeling by the software. The statistical results and modeled descriptors along with their contribution are given in **Table 4.11**. The results obtained from the PLS models of BEAS-2B cell ($Q^2_{\text{LOO}}=0.655-0.871$, $R^2=0.931-0.956$) and RAW 264.7 cell ($Q^2_{\text{LOO}}=0.946-0.969$, $R^2=0.964-0.982$) were statistically significant. The definitions of the descriptors obtained from the PLS models are given in **Table 4.12**.

Table 4.11. Statistical parameters of the developed PLS models with Small Dataset Modeler (M 4.1- M 4.4).

Statistical parameters				Modeled descriptors	R^2	Q_{Loo}^2	$\bar{r}_m^2$	Δr_m^2	MAE (95%)	Standardized contribution of descriptors
CELL	ASSAY	MODEL	MeO x NPs							
BEAS-2B cell	ATP	Model 4.1 (M 4.1) (LV1)	9	• Z_metal	0.931	0.871	0.849	0.027	2.154	
				• χ_{ox}						
	MTS	Model 4.2 (M 4.2) (LV2)	10	• Z_metal	0.945	0.651	0.277	0.465	4.187	
				• Electrochemical equivalent						
				• Valence						
RAW 264.7 cell	MTS	Model 4.3 (M 4.3) (LV1)	10	○ Atomic radius (A)	0.982	0.969	0.961	0.006	0.861	
				○ Z_metal						
	ATP	Model 4.4 (M 4.4) (LV2)	12	○ Mol_wt	0.964	0.949	0.933	0.022	0.458	
				○ Z_metal						
				○ Nmetal						

*LV- Latent variable

Table 4.12. Details of the modeled descriptors from developed PLS models (M 4.1 -M 4.4).

SL. no.	Descriptors	Description	Contribution
1	Z_metal	The atomic number of metals	-ve
2	χ_{ox}	Oxidation number	-ve
3	Electrochemical equivalent (Eq)	Mass of that element (in grams) transported by 1 coulomb of electric charge.	-ve
4	Valence	The valence of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules.	-ve
5	Mol_wt	Relative molecular mass	-ve
6	Nmetal	Number of metals	+ve

4.4.2. Statistical aspects of the nano-QTTR models

The linear regression performed between cytotoxicity data for the common eight compounds from BEAS-2B cells and RAW 264.7 cells yielded a squared correlation coefficient (R^2) of 0.5 without using any kind of calculated descriptors. To get a more robust model, along with the response value of one of the cells, some periodic table descriptors were also included. Using the RAW 264.7 cell response variable, the BEAS-2B cell cytotoxicity was used as a predictor variable and vice versa. The values of the leave-one-outcross-validation metrics (Q^2_{LOO}) were 0.37 and 0.50 respectively. Therefore, to get an interpretable and statistically significant model, along with some periodic table descriptors, the cytotoxicity of one cell type can be used as the predictor variable for determining the cytotoxicity of other cells.

RAW 264.7 cells and BEAS-2B cell toxicity correlation

Considering RAW 264.7 cell toxicity as the response and BEAS-2B cell cytotoxicity as a predictor variable along with some calculated periodic table descriptors, we have constructed a PLS model as below:

$$\log_{EC\ 50}Rawcell = -8.69557 - 0.22261 \log_{EC50}Beas\ cell + 4.4442X_ActivM$$

$$n = 8, R^2 = 0.90, R^2_{adj} = 0.90, Q^2_{LOO} = 0.86, LV = 1, PRESS = 6.60254 \quad (4.6)$$

In **equation 4.6**, X_ActivM is a 2nd generation periodic descriptor, which stands for the electronegativity of active metalloid/semi-metal.

BEAS-2B cell and RAW 264.7 cell toxicity correlation

Similarly, in this case, BEAS-2B cell toxicity was taken as the response, and RAW 264.7 cell toxicity was considered as a predictor variable along with a structural descriptor to develop a PLS model represented as **equation 4.7**

$$\log_{EC50} \text{Beas cell} = 13.6745 - 1.2606 \log_{EC50} \text{Rawcell} - 8.0142 \text{ Atomic radius}$$

$$n = 8, R^2 = 0.98, R_{adj}^2 = 0.99, Q_{LOO}^2 = 0.66, LV = 1, PRESS = 100.7 \quad (4.7)$$

In **equation 4.7**, the atomic radius is a periodic table descriptor that signifies the measure of the size of the atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons (Roy and Roy, 2021). Thus, it is evident that the toxicity values of the cell lines are interrelated to each other. The MeOx NPs responsible for murine cell line toxicity are also toxic to human cells. The nano-QTTR models can be employed for extrapolating cytotoxicity to BEAS-2B and murine myeloid RAW 264.7 cells for metal nanoparticles when the data for the other species are available.

In the scarcity of experimental nanotoxicity data, these results would indicate a new approach and contribute to more possibilities for comprehensive risk assessment of nanoparticles when cytotoxicity data is missing for a specific cell line. The predicted values are given in **Table 4.13**.

Table 4.13: Observed and predicted values for individual compounds for BEAS-2B and murine myeloid RAW 264.7 cells lines from QTTR models

ID	Metal oxide	Observed log (1/EC ₅₀)BEAS-2B cell	Predicted log (1/EC ₅₀)BEAS-2B cell	Observed log (1/EC ₅₀)RAW 264.7 cell	Predicted log (1/EC ₅₀)RAW 264.7 cell
1	Co ₃ O ₄	-1.37	-1.41	-1.35	-1.00
2	Cr ₂ O ₃	-1.55	-2.50	-1.07	-1.93
3	Ni ₂ O ₃	-1.8	-1.75	-1.94	-1.10
4	CuO	-1.38	-0.37	-1.16	-0.56
5	Mn ₂ O ₃	-1.12	-2.06	-1.1	-2.28
6	CoO	-1.64	-1.83	-1.68	-1.06
7	ZnO	-0.56	0.27	-0.91	-1.71
8	La ₂ O ₃	-19.26	-19.02	-8.52	-8.09

4.4.3. Chemical Read-Across Analysis

Read-across (RA) is a similarity-based approach for the prediction of a new compound referred to as ‘target compound’ based on the endpoint of other similar compounds known as the ‘source’ compounds. The similarity-based prediction method has gained enormous attention owing to its success in data gap-filling issues. In the present work, the quantitative-read across approach based on similarity for the four datasets was implemented using the descriptors obtained from the four PLS models for the four datasets. The datasets were divided into training and test sets for performing the read-across technique. In this work, we tried to maintain an equal number of compounds in training and test sets. Three similarities-based measures were involved in this work namely Euclidean Distance (ED) –based Gaussian Kernel (GK) Similarity-based, and Laplacian Kernel (LK) Similarity-based prediction. After optimizing the hyperparameters, i.e., σ (sigma value) and γ (gamma value) for all the datasets separately, it was observed that the read-across prediction was within acceptable criteria as given in **Table 4.14**. The work was performed using Read-Across-v4.0 (Chatterjee et al. 2022; Banerjee and Roy, 2022; Banerjee et al., 2022). The threshold values for distance can range from 1 to 0, similarity threshold values from 0 to 1, and the numbers of most similar training compounds from 2 to 10. The sigma (σ) and gamma value (γ) range from 0-2.

Table 4.14. Prediction quality of the features obtained from the PLS (M 4.1-M 4.4) models using the read-across approach.

Feature combinations		Hypothesis	Hyperparameters				Statistical parameters			
			σ	γ	Distance threshold	Similarity threshold	Q_{F1}^2	Q_{F2}^2	MAE	RMSEP
BEAS-2B cell	M 4.1 (9 MeOx NPs)	RA-ED	0.5	0.5	1	0	0.779	0.777	1.859	3.428
		RA-GK					0.993	0.993	0.506	0.622
		RA-LK					0.583	0.580	3.485	4.707
	M 4.2 (10 MeOx NPs)	RA-ED	1	1	1	0	0.912	0.531	3.169	3.733
		RA-GK					0.713	-0.525	5.962	6.734
		RA-LK					0.703	-0.575	6.317	6.844
RAW 264.7 cell	M 4.3 (10 MeOx NPs)	RA-ED	1.5	1.5	1	0	0.953	0.895	0.889	1.067
		RA-GK					0.784	0.519	1.432	2.284
		RA-LK					-0.062	-1.365	2.572	5.064
	M 4.4 (12 MeOx NPs)	RA-ED	1	1	1	0	0.833	0.832	1.875	3.113
		RA-GK					0.937	0.936	1.307	1.916
		RA-LK					0.914	0.913	1.432	2.236

4.4.4. Interpretation of the descriptors for the BEAS-2B cell toxicity model

	Model	Descriptors	Interpretation	Example
BEAS-2B cell toxicity	Model 4.1	Z _{metal}	The atomic number of an atom is the charge number of the nucleus. The increased electropositivity of the nucleus has a stronger pull on the valence electrons preventing them from leaving the outermost shell. As a result, the valence electrons cannot form a bond or influence ROS initiation since the electrons cannot move out due to the strong electropositive attraction of the nucleus.	La₂O₃ and Sb₂O₃ MeOx NPs with an increase in atomic number show a decrease in the toxic effect on the cell. On the other hand, the reverse phenomenon happens in Cr₂O₃ and Mn₂O₃ MeOx NPs, where a decrease in atomic number elevates the toxicity of the cell.
		χ_{ox}	The oxidation number or state reflects the electronic configuration. A reducible oxide contains metal ions with a maximum oxidation state, yet can be reduced to a lower oxidation state within certain limits. Such metals contain vacant oxygen vacancies after the loss of oxygen in the crystal lattice that may exist in the bulk or on or subsurface of nanomaterials (Baek et al, 2011). These vacancies result in the presence of reduced cations on the surface of nanoparticles with Lewis base character. The surface of Oxidizable oxides prevalently exhibits oxygen vacancies associated with the oxidized metals that behave like a Lewis acid, i.e., they accept electrons. Concerning their Lewis acid-base behavior, this nanoparticle vacancy acts as a binding site, particularly for biomolecules that may compose bio-corona responsible for adverse pathways	The increase in the oxidation number suggests that with the increase of the oxidation number, more electrons are removed from the metal oxide as in MeOx NPs CoO and CuO , and the metal results in the formation of metal cation. Again, the opposite happens in MeOx NPs Sb₂O₃ and Ni₂O₃ .

	Model 4.2	Electro Eqv	The electrochemical equivalent (Electro Eqv) of a chemical element indicates the mass of that element (in grams) transported by 1 coulomb of electric charge. It indicates the kinetics of corrosion rate and estimates the oxidizing power of metal in a specific environment. The number of valence electrons plays an important role here. This means that with an increase in the number of valence electrons, the descriptor value decreases. The lower oxidation state elements are found to be more harmful than the compounds with higher or stable oxidation states (Walker, et al., 2003). This metal ion can then radically combine with the macromolecules present inside the body and enhance toxicity through the Haber–Weiss–Fenton cycle (Koppenol, 2001).	The Electro Eqv descriptor negatively influences cytotoxicity in BEAS-2B, and as the descriptor value decreases, cytotoxicity increases just as it does in $\text{MeOxCr}_2\text{O}_3$. On the contrary, an increase of the descriptor value decreases the cytotoxicity as the wearing of the metal reduces the toxicity from its oxide as in $\text{MeOxNP Fe}_2\text{O}_3$.
		Valence	The first-generation periodic table descriptor Valence measures the combining capacity of the metal with other atoms when it forms chemical compounds or molecules. The different valence states of metal oxide exhibit different characteristics and patterns of toxicity.	The negative contribution of this descriptor suggests that an increase of valence as in CoO MeOx NP decreases the toxicity whereas the decrease of the descriptor value as observed in CuO increases the cytotoxicity of the cell.

RAW 264.7 cell toxicity	Model 4.3	Atm rad (A)	The size of the atom is associated with the distribution of electrons around the nucleus. When electrons are closer to the nucleus, they are more tightly bound, thus creating a higher electronegativity in the metal. Therefore, when the atomic radius is increased, the electrons are not bound tightly, which results in a reduction in the oxide's electronegativity, thus reducing the metal's toxic effect.	This descriptor contributes negatively to the cytotoxicity which means that the latter explanation is justified as observed in MeOx NPs Yb ₂ O ₃ and HfO ₂ where an increase of the radius causes a less toxic effect than the MeOx NPs CuO and ZnO having a larger atomic radius.
	Model 4.4	Mol_wt	Mol_wt describes the molecular weight of the metal. This descriptor contributes negatively to RAW 264.7 cell cytotoxicity. This happens because as the weight increases, the size of the metal oxide will also increase, and the penetration of the metal oxide through the cell will become limited. Reducing the size of the metal oxides increases the surface area, making the active size of the metal more accessible for chemical, electrical, magnetic, structural, and morphological interaction.	MeOx NPs Gd ₂ O ₃ and La ₂ O ₃ , bearing high molecular weight have low cytotoxicity whereas MeOx NPs CuO and CoO with low molecular weight possess higher toxicity to RAW 264.7 cells.
		N _{metal}	N _{metal} stands for the presence of metals in the metal oxide nanoparticles. The higher number of metals in the NPs can trigger more ROS production through the formation of metal cations. The higher cation formation may further accelerate bond formation with the protein in the biological system which may increase the chance of metals accumulating in the body making excretion difficult. The persistent nature of metals can cause carcinogenic, mutagenic, and teratogenic effects on the living cell that may consequently cause disease and abnormal conditions in the body (Briffa et al., 2020).	As this descriptor contributes positively toward RAW 264.7 cell cytotoxicity, an increase of the descriptor value increases the toxicity of the cell as observed in MeOx NPs Co ₃ O ₄ , whereas the opposite is observed in MeOx NPs HfO ₂ .

*The common descriptors from all the models are discussed.

4.4.5. Mechanism of toxicity.

The assessment of the detrimental effects of nanopesticides on living organisms demands a comprehensive understanding of their potential impact on the environment. Despite the benefits and the requirement of an agro-food sector, the enormous use of agrochemicals has led to the emission of a huge runway of pesticides into the environment. Further challenges in toxicity assessment of nanopesticides arise from other sectors through contamination of terrestrial and aquatic ecosystems and incorporation of toxic ingredients in food for animal and human consumption. Each of the individual sectors has its complexity and should be dealt with specific consideration, especially when the pesticides are present in low concentrations and dispersed form. The biotransformation of nanomaterials in the ecosphere may interact with various bio-films or biological organisms that can determine the fate, behavior, and persistence in the environment (Mu et al. 2014). It is noticeable that some terrestrial species come in direct contact with pesticides and hence different interactions can occur with or within the organism. The metal cation released from the metal oxide exhibits higher binding and cellular uptake to the cell than the negatively charged NPs. The charged particles induce different perturbations to the cellular membrane potential. The lower valence suggests the release of metal cations through corrosion would depolarize the cellular membrane and form holes. These holes lead to the initialization of metal cations inside the cell as well as diffusion of the cytosolic content out of the cell. The manifestation of the NPs is supposed to be dependent on the size and morphology of the NPs. MeOx NPs can enter the cell through various paths which in return depends on the size and possibility of interaction with various surface proteins and lipids, but the initial contact between NPs and cell (Djurišić et al. 2015) is usually through electrostatics as shown by the Electro Eqv descriptor. The decrease in size as indicated by the lower Mol_wt descriptor is thought to be correlated with higher uptake into the cell. The toxicity of the NPs inside the cell depends on their stability in the cellular environment (Zhu et al. 2013). The acidic surrounding of lysosomes can cause a redox reaction as demonstrated by the decrease of χ_{ox} descriptor. The binding properties of the NPs are explained by the Z_{metal} , as the decrease of this descriptor increases the cellular interaction leading to damage to the membrane and the production of ROS (reactive oxygen species). NPs can enter the mitochondria (the powerhouse of the cell) and affect their function. The nuclear envelope inhibits the entry of large NPs into the nucleus. The small-sized NPs enter the nucleus and exhibit strong affinity with the histones or other cell organelles through electronegativity. The NPs may cause decondensation of nuclear chromatin in mice which is driven by the interaction of NPs with the nucleus. The nano-specific behavior also intrinsically depends on the metallic nature of NPs used in pesticides as the descriptor N_{metal} . Their toxicity is associated with internalization and bioaccumulation by the non-targeted organism (Firdaus et al, 2018) (Fojtova, et al, 2019). Such interaction and bioaccumulation may alter oxidative stress-dependent and -independent pathways that affect proliferation and differentiation. Other signaling pathways such as MAPK, NF- κ B, and DNA damage are responsible for nanoparticle-induced oxidative stress. All these events as in **Figure 4.28** eventually cause the death of the respective cell (Fraceto et al. 2020).

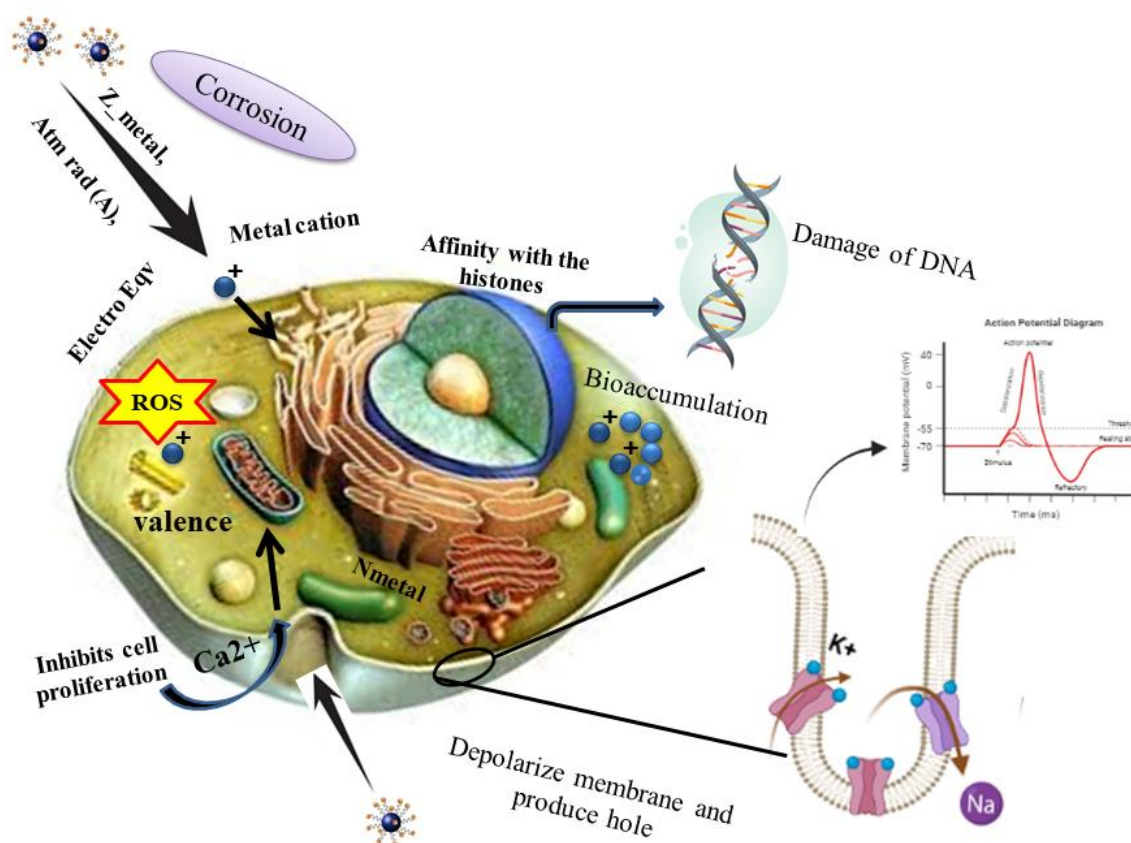


Figure 4.28 Mechanism of toxicity to BEAS-2B cell and RAW 264.7 cells (study 4).

4.5. Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach.

Successful development of the QSAR model depends on the quality of biological data. Although the available MeOxNPs data presented here are limited, but are sufficient for developing predictive models. Many authors have developed and published QSAR models with a small number of data points (Puzyn et al. 2011; Kar et al. 2014; Roy and Roy 2022a). The criterion of a useful model is the difference between R^2 and Q^2 being less than 0.3. With this, the MLR model ($FC_{IL-1\beta}$ dataset; Data set 5.1) satisfies this criterion ($R^2 = 0.89$, $Q^2 = 0.86$, $Q^2_{FI} = 0.86$). **Equation (4.8)** suggested that the developed MLR model could be applied to evaluate the inflammatory potential of new, untested MeOxNPs. The results obtained from the cell viability $[-\ln(p/(1-p))]$ data set (5.2) ($R^2 = 0.72$, $Q^2_{loo} = 0.53$, $Q^2_{FI} = 0.66$) with 54 data points and latent variable 4 **equation (4.9)** could explain 72 percent of variance while it could predict 53 percent of leaving-one-out (LOO) predicted variance. The scatter plots (**Figure 4.29**) suggest that the models have good predictions. **Table 4.16** gives the results of MLR models.

The AD checked with the standardization approach showed that for model 5.1, compound no **28** was an outlier in the training compounds, whereas there was no outlier compound in case of model 5.2. No compound was found to be outside the applicability domain for the test sets for all the models. As per the William plot (**Figure 4.30**).

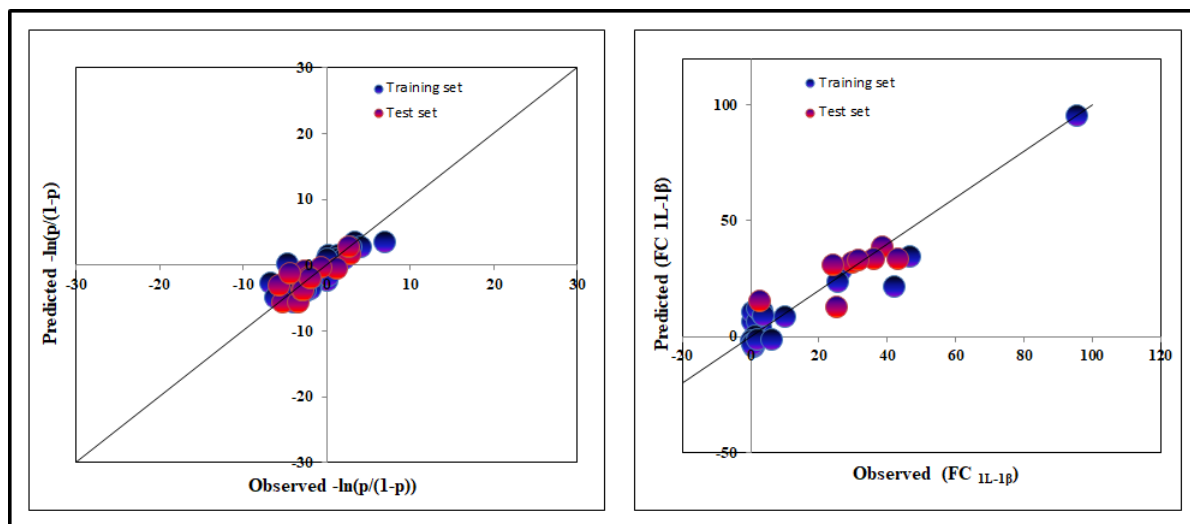


Figure 4.29 Scatter plots for the MLR models (FC_{IL-1β}, -ln (p/ (1-p))) (Model 5.1 and 5.2).

Table 4.16. Results of MLR models with different endpoints.

Model	Descriptors	Number of data points	Endpoints	R ²	Q ²	Q ² _{F1}
MLR 5.1	Nmetal, $\Sigma\chi$, atomic radius	32	FC _{IL-1β}	0.893	0.86	0.86
MLR 5.2	$\frac{\Sigma\chi}{nO}$, χ , Vpφ, atomic radius, dose, $\Sigma\chi$	54	-ln (p/(1-p))	0.72	0.55	0.66

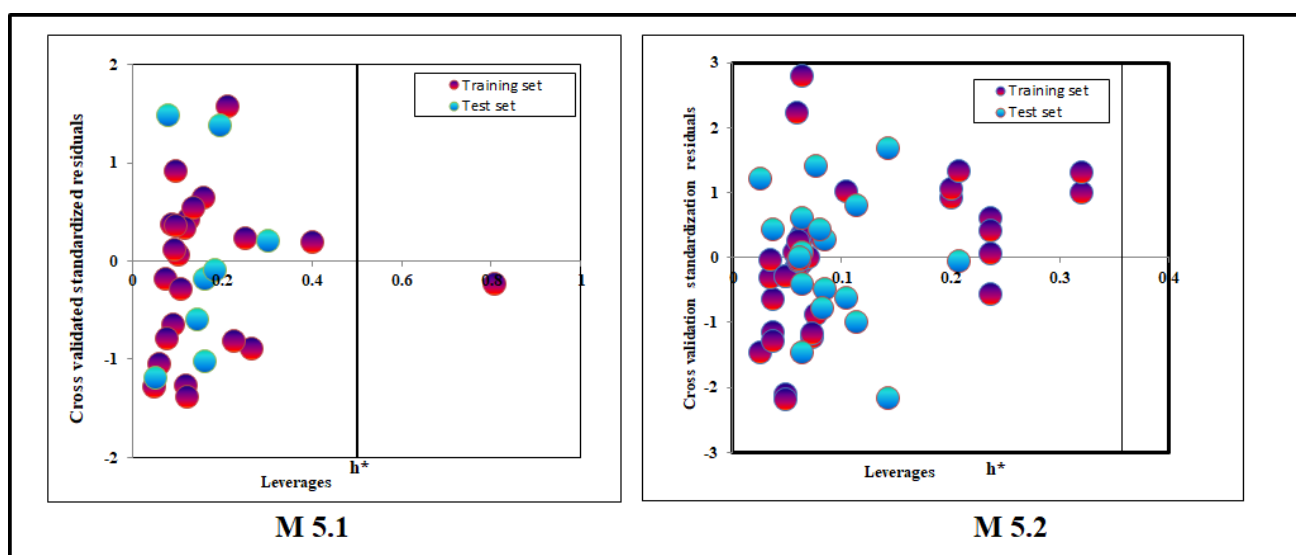


Figure 4.30 William's plot of the three MLR models.

4.5.1. MLR models for inflammatory potential

Model 5.1

$$FC_{IL-1\beta} = -32.2 + 30 N_{metal} - 13.4 \Sigma\chi + 14.9 \text{ atomic radius} \quad (1)$$

$$n_{train} = 24, R^2 = 0.893, Q_{LOO}^2 = 0.856, PRESS = 1637.56, F = 55.60, s = 7.80424$$

$$n_{test} = 8, Q_{F1}^2 = 0.86 \quad (4.8)$$

N_{metal} is 1st generation periodic table descriptor. This descriptor indicates the number of metals present in the MeOx NP. For example, if the MeOx NP contains one metal in the MeOx NP, then the descriptor value would be one, and so on. This descriptor contributes positively towards the Fold Change (FC_{IL-1 β}) endpoint which suggests that with an increase in metal number, the FC_{IL-1 β} increases. The aforesaid statement can be well understood from the Pr₆O₁₁ MeOx NP where the descriptor **N_{metal}** value is 6 and the FC_{IL-1 β} value is 95. On the contrary, ZnO MeOx NP bearing **N_{metal}** value 1 represents an FC_{IL-1 β} value of 0.29. Thus, it can be clearly depicted that the **N_{metal}** descriptor has a significant effect on toxicity. Certain metals have been reported to seriously affect the immune system potentially causing ill effects on health. The presence of metal, depending on its concentration, route of entry, and duration of exposure may lead to immunosuppressant or stimulation of immune cell activity. The metal oxides alter the expression level of adhesion molecules along with chemokine receptors, i.e., the key factor associated with immune response. This alteration affects the T-cell proliferative response to mitogens and chemotaxis (Lozano-Fernández et al., 2014). The induction of metals can form metal–protein aggregates through metal co-ordinate bond formation eventually causing pro-inflammatory function by the assembly of NLR3 (NACHT, LRR, and PYD domain-containing protein, and inflammasome initiating IL-1 β release from the cell (Pettersson et al., 2022).

$\Sigma\chi$ stands for the summation of electronegativity of the metals present within the MeOx NP. This descriptor contributes negatively towards the inflammatory effects which suggests that with the increase of the descriptor value, the toxicity decreases whereas the opposite takes place with the decrease of the descriptor value. This observation also was in agreement with the result that Co₃O₄ MeOx NP with low electronegativity tends to induce more inflammatory potential than Sm₂O₃ MeOx NPs with higher electronegativity. The MeOx NPs with less electronegativity will have fewer valence shell electrons, which in turn means that their oxidation number is less. From the previous studies, we know that metals with less oxidation number are more harmful than higher oxidation numbers (Roy and Roy, 2022b). The tendency of a metal to gain its higher oxidation state disrupts cellular processing through the generation of reactive oxygen species (ROS) which reduces the antioxidants of the cell and increases stress within the cell (Kang et al., 2007). The MeOx NPs with less electronegativity can easily enter the cell through phagocytosis. The increased cell stress can further modify the immune system and down-regulate the cortisol receptors. This down regulation, in turn, enhances the cytokine-mediated inflammatory processes by reducing the response of a cell to anti-inflammatory signals or stimuli (Segerstrom and Miller 2004).

The third period descriptor of the model is **Atomic radius**. This descriptor measures the size of the element, usually, it is the typical measurement of the distance from the nucleus to the outermost isolated electron. With an increase of the atomic radius, the formation of metal ions

becomes easier since with the increase in the number of shells, there is an increase of shielding effect and the outermost electrons in the valence shell does not experience the counter positive pull from the nucleus. The positive contribution of the descriptor suggests that MeOx NPs with a high descriptor value will have a high potential to cause inflammatory effect as in La₂O₃ MeOx NPs with descriptor value 2.74 has an FC_{IL-1 β} value 38.63 whereas CoO MeOx NPs with less Atomic radius value 1.53 tends to cause a minimum effect on the inflammatory potential with 0.31. The metal ion release through corrosion binds with the serum protein and forms hapten or hapten-like complexes which trigger the immune system. Some metal ions induce an increased level of superoxide anions that indicate neutrophil stimulation (Kumazawa et al. 2002; Mu et al. 2000) and may bind to phosphorylated cellular molecules which plays a major role in cell signaling. By binding to such messenger pathways, MeOx NPs may interfere with the signaling pathways and the immune function.

Model 5.2 (Cell viability)

$$-\ln(p/1-p) = -6.85457 + 0.01647 \text{ dose} - 4.51478 \chi + 8.48 \frac{\Sigma \chi}{nO} - 0.05770 \text{ atomic radius} + 0.02652 \text{ valence electron potential} - 0.42535 \Sigma \chi$$

$$n_{train} = 41, R^2 = 0.72, Q_{LOO}^2 = 0.55, PRESS = 138.486, F = 12.47, \quad s =$$

$$1.67892, \Delta r_m^2(LOO) = 0.0791, \bar{r}_{m(LOO)}^2 = 0.5056, LV = 4$$

$$n_{test} = 13, Q_{F1}^2 = 0.67, \Delta r_m^2(LOO) = 0.13174, \bar{r}_{m(LOO)}^2 = 0.553 \quad (4.9)$$

Dose of any metal oxide plays an important role in causing the inflammatory potential. This descriptor positivity contributes towards the release of cytokine from the cell. The higher the dose, the greater the release of cytokines from the respective cell as observed in La₂O₃-7 MeOx NP (where the dose is 200 $\mu\text{g/mL}$ and log₂ (FC_{IL-1 β}) is 5.27) whereas a low dose is not that effective as observed in Sm₂O₃-1 MeOx NP (where the dose is 3.1 $\mu\text{g/mL}$ and log₂ (FC_{IL-1 β}) is 0.07). The effect on T- cells depends on the route of administration and also on the dose of the NPs. An increased expression of cells and adverse effects have been observed at higher doses of MeOx NP (Shah and Dobrovolskaia 2018). This observation is supported by some literature reports such as that CeO₂, TiO₂ NPs, and Al₂O₃ activate ERK (1, 2); this activation is dose- dependent and does not activate at lower doses. ZnO NPs or Zn²⁺ salt was also dose-dependent, and phosphorylation of ERK (Extracellular signal-regulated kinases) is only detected at higher concentrations of Zn²⁺ ions (Simon-Vazquez et al. 2016). Chen et al. (2012) have also demonstrated that TiO₂ nanoparticles not only dose-dependently increased histamine secretion but also increased cytosolic Ca²⁺ concentration in rat mast cells.

The **valence electron potential (Vp ϕ)** descriptor provides a quantitative indication of elements reactivity and is based on the charge of the valence electrons and the ionic radius. This descriptor contributes positively towards the endpoint ($-\ln(p/(1-p))$). An increase in the descriptor value (220 eV) as observed in MnO₂ MeOxNP causes an increase in cell death (4.027) compared to the lower descriptor value (40.71 eV) in La₂O₃ MeOxNP where cell death is -1.57. The charge of the atom describes the detachment of the metal from the metal oxides. It is determined by the equation.

$$-ev = kn/r$$

Here, 'k' expresses the force exerted by the valence electron in volts which is equated to 14.33, 'n' is the valence, and 'r' is the ionic radius. A valence electron is furthest from the nucleus and is removed by spending less energy. Therefore, the electron relatively away has a smaller (1/r) value so the energy is larger. Decreased interaction of the nucleus is the principal reason why electrons away from the nucleus are higher in energy than the electrons near to the nucleus (Walker et al. 2003). The charge and radius ratio of the metal ion indicate the distance between the ion and another charge present, and the size of the electrostatic force created thus defines the index for ionic bond formation.

The $(\Sigma\chi)/nO$ descriptor stands for total metal electronegativity in a specific metal oxide relative to the number of oxygens. This ratio gives information on the electronegativity of the atom with respect to the number of oxygens present in the MeOx NP. With an increase of this ratio, there is a decrease of count cell deaths. The positive coefficient of this descriptor suggests that with an increase of the ratio as in CuO (1.9) there is an increase of cell inflammatory potential (6.9) whereas the reduction of $(\Sigma\chi)/nO$ ratio (0.56) decreases cell inflammatory potential (-3.37) as in CeO₂ MeOx NP. This descriptor also considers the presence of oxygen and indicates the effect of oxygen on the inflammatory potential. The increase in the number of oxygens relative to electronegativity enhances the change of reactive oxygen species which leads to stress in the cell. Inflammation and environment oxygen levels are linked with each other. A low level of oxygen count gives a high ratio; therefore, a low level of oxygen or hypoxia can lead to inflammatory conditions in the cell.

It is to be noted that electronegativity (χ) has a positive contribution but the summation of electronegativity ($\Sigma\chi$) has an opposite contribution. Therefore, it can be inferred that electronegativity describes the tendency to attract electrons. Metals with low χ tend to attract electrons or metals with low χ tend to lose electrons to an oxygen atom. Oxygen with more electrons traps protons (positively charged ions). Therefore, metals with lower χ have a higher affinity towards protons than MeOx NPs with higher χ . As per the proton sponge effect (Nel et al, 2006) **Figure 4. 31**, some nanoparticles enter the liposomal compartment and sequester protons supplied by the proton pump (v-ATPase). This retains the proton functions which leads to the retention of one Cl⁻ molecule and two molecules of water. Thus, it causes the rupture of lysosomal content, triggering excess IL-1b production and inflammation. Therefore, this hypothesis is supported by the developed MLR and classification models. Other common descriptors are described previously for other models.

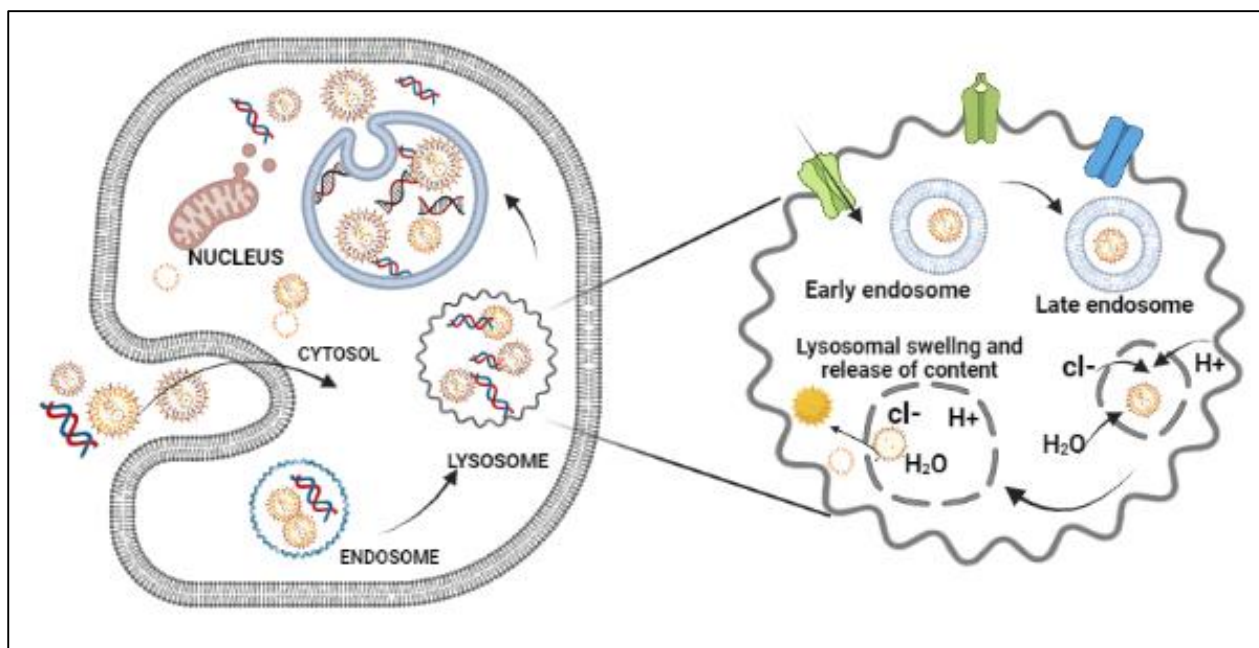


Figure 4. 31 Proton Sponge effect.

4.5.2. Evaluation of the predictive ability of the developed MLR model for fold change

To further evaluate the performance of the developed MLR model (**Equation 4.8**), new MeOxNPs were selected to verify the performance of the MLR model that lacked inflammatory potential response experimental information. These MeOxNPs were not included in the training set or test set. The “Prediction Reliability Indicator” (Roy et al. 2018a) tool was used to categorize the quality of predictions for the external set into three groups i) good (with composite score 3), ii) moderate (with composite score 2), and iii) bad (with composite score 1). Three criteria were used: 1) mean absolute error of leave-one-out predictions for 10 most close training compounds for each query molecule (Roy et al. 2018b); 2) applicability domain in terms of similarity based on the standardization approach (Roy et al. 2015); 3) proximity of the predicted value of the query compound to the mean training response (Roy et al. 2017) in different weightage schemes for making a composite score of predictions depending on the molecular features. Thus, we have collected 37 MeOx NPs as an external set (see in materials and methods **Table 3. 13**) to evaluate the predictive power of the developed MLR models and found that out of 37 MeOxNPs, 34 MeOxNPs showed good predictions except one (Tm_2O_3) which had moderate prediction while CeO_2 was found to be an outlier.

4.5.3. Features responsible for the inflammatory effect to THP-1Cells

To explore the hazardous effects of the MeOxNPs, the THP-1Cells were exploited to examine the toxic impact of MeOxNPs on cell viability and inflammatory responses. This was done since the behavior of differentiated THP-1 cells is similar to monocyte-derived macrophages (Auwerx, 1991) and phagocytic uptake is the prime pathway to remove non-soluble particulates (Geiser, 2007). We present a schematic diagram to collectively determine the inflammatory response of MeOxNPs. The regression models pinpointed that N_{metal} , $\Sigma\chi$, atomic radius, Valence electron potential, and $(\Sigma\chi)/nO$, Dose, $AtWt_Activ_SM$ are the most determining factors for inflammatory potential in THP-1 cells. The main feature derived from the developed models, i.e. presence of metals, has been found to manifest allergic reactions towards orthopedic implants (Foussereau and Laugier, 1966). Another indication is the

increased cytokines in the serum with orthopedic implants. The immune reactivity of metal ions is considered to be a type IV hypersensitive reaction (Büdinger et al. 2000). The presence of electronegative atoms triggers IL-1 β production and tends to rupture lysosomal content from the cell. IL-1 β is involved in a variety of cellular activities such as cell proliferation, differentiation, and cell death. The higher dose of the MeOxNPs also activates IL-1 β productions which activates the NALP-1 caspase-1 inflammasome cascade and participates in the process of cell pyroptosis, characterized by the release of massive cytokine levels (Bonaventura et al., 2018). Caspase-1 is generally present in inactive form as procaspase-1 but can get activated by the inflammasome. The activation of inflammasome further stimulates the cleavage of pro-IL-1 β by caspase-1, resulting in mature IL-1 β that is released from the cells. This cascade of reactions further activates the receptor in a paracrine manner producing proinflammatory gene products. **Figure 4.32** mechanism of inflammatory potential with sponge effect.

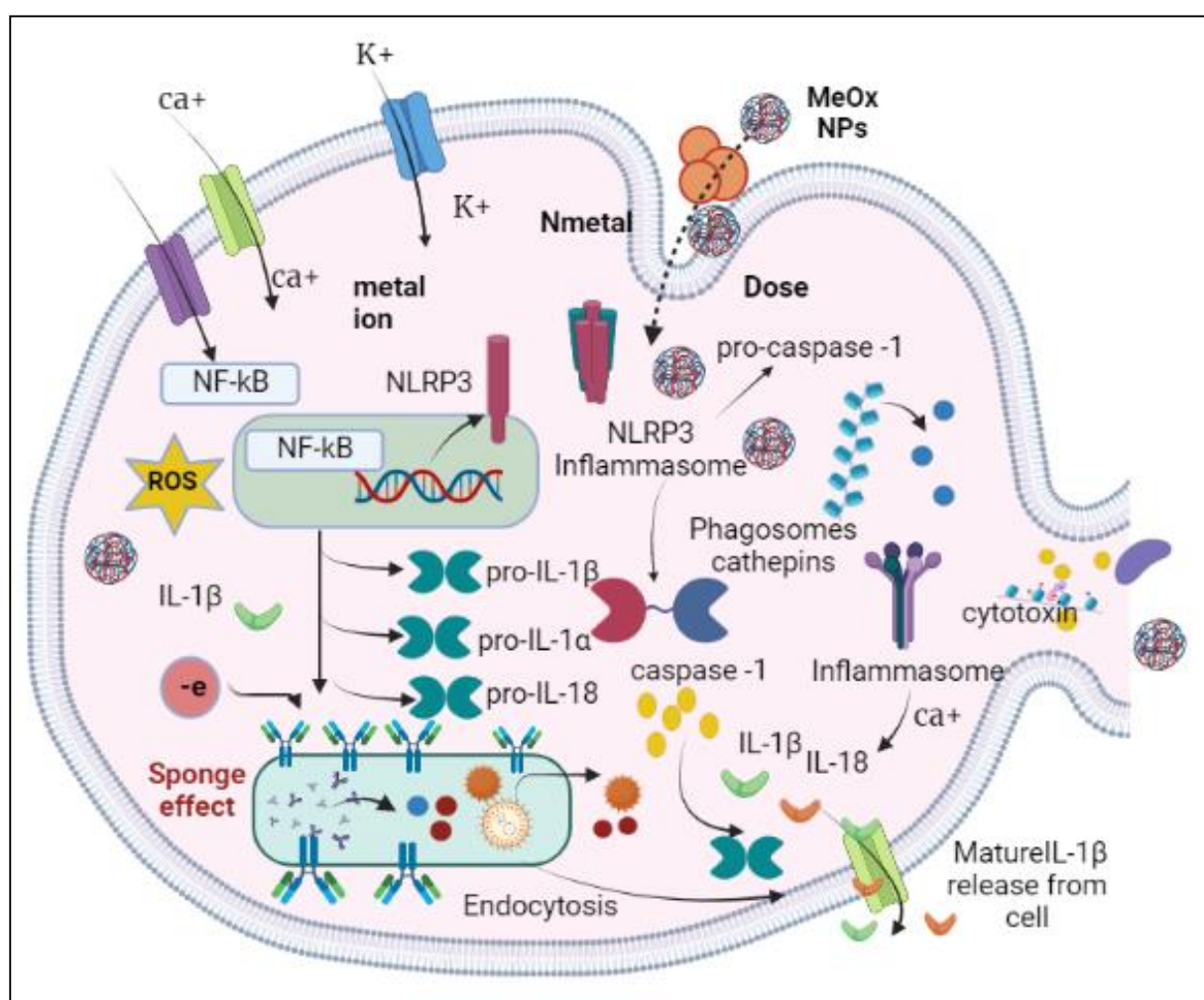


Figure 4.32 Mechanism of inflammatory potential with sponge effect, release of cytotoxins, and lysosomal leakage (study 5).

4.5.4. Comparison with the previous study

This study focuses on identifying the features governing the inflammatory potential of MeOx NPs to THP-1 cells. The previous study by Huang et al. (2020) performed the same study with 26 MeOx NPs ($FC_{IL-1\beta}$) applying machine learning techniques with quantum-mechanical computations, experimental characterization, and 5 periodic table descriptors to generate QSAR models (partial least squares). The developed model in the previous work showed a result of $R^2=0.89$, $Q^2=0.87$, and $Q^2_{F1}=0.80$. On the other hand, the present study used 32 MeOx NPs for modeling using the multi-linear regression method for generating a regression model using the in-house QSAR modeling tools. Again, the percent cell viability data was converted to logit form ($-\ln(p/(1-p))$) to evaluate the toxicity profile of the reported metal oxides. This endpoint was not modeled by the previous authors. The main novelty of the present work is that the whole work was performed using only periodic based-table descriptors that are computationally less demanding, independent of nanoparticle size, and do not require computational resources unlike quantum chemical descriptors. The result obtained in the present model with periodic table-based descriptor showed similar model quality ($R^2=0.89$, $Q^2_{loo}=0.86$, $Q^2_{F1}=0.86$) with the previous work (**Table 4.17**). The features obtained from the MLR model ($FC_{IL-1\beta}$) were also akin to that of the previous study involving machine learning techniques and quantum-mechanical descriptors which is more complicated.

Table 4.17 Comparison of the previous work with the present study.

Model	Descriptor s	Method	Endpoint	No. of compound s	R^2	Q^2	Q^2_{F1}
Previous model (Huang et al. (2020)	Mainly Quantum descriptors	Regression (PLS)	$FC_{IL-1\beta}$	26	0.89	0.87	0.80
Present work	Periodic table-based descriptors	Regression (MLR-M5.1)	$FC_{IL-1\beta}$	32	0.89	0.86	0.86
		and (PLS- M5.2)	$-\ln(p/(1-p))$	54	0.72	0.53	0.66

4.6. Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches.

In this research, we have used four ML models namely RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost to forecast the toxicity of nano-TiO₂ adsorbed heavy metals to HK-2 cells using periodic table-based descriptors (**Table 4.18**). The ML models were built using the features selected by the RF algorithm. Model specification and configuration were built by optimization of the hyperparameters. The AD was also diagnosed and all the compounds were found to be within the critical point (h^*) i.e., 0.42 as suggested by **Figure 4.33**. The applicability domain (AD) is the chemical space formed based on the descriptors of the training set compounds. The compounds in the space are considered reliable for predictions and those beyond the AD do not guarantee accurate predictions. AD plays an important role in determining the uncertainty of the predictions of specific molecules based on how close they

are to the training set compounds used to develop the model. AD is a valuable tool for the characterization of interpolation spaces based on the modeled descriptors and response functions.

The applicability domain developed here is based on the features of some specific heavy metal salts, that is, CdCl_2 , ZnCl_2 , MnCl_2 , CoCl_2 , CuSO_4 , NiCl_2 , $\text{Pb}(\text{NO}_3)_2$, and SbCl_3 . The developed model should be applicable to other closely related heavy metal salts.

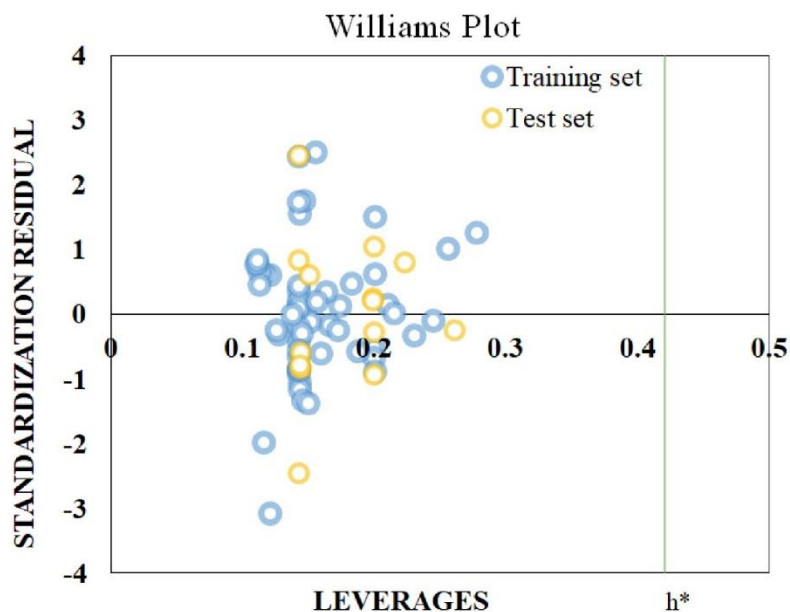


Figure 4.33 Williams plot.

4.6.1. Diagnosis based on SHAP value

The goal of SHAP is to explain the prediction of an instance by computing each feature to the prediction. Firstly, the SHAP value is used to calculate the magnitude of the contribution of each feature and then ranked to obtain the importance ranking of features. Features with large absolute Shapley values are important. Here, we have used the kernel method to calculate the SHAP value (Liu et al., 2022). The SHAP analysis and hyper parameter tuning (max_depth: "none" min_samples_leaf, min_samples_split, n_estimators) revealed that concentration, followed by atomic radius and IP_ActivM, ranked highest among the eight features (Conc, $\sum\chi$, atomic radius, IP_ActivM, Mol_Wt, χ of metal, D3_HeteroNonMetal, and atoms in the molecule) in the RF model. The hyper parameter setting n_estimators was kept 'at 80' for RF, while for 130, 70, and 170 for Gradient Boost, Extreme Gradient Boost, and AdaBoost respectively. The relative importance of each descriptor for all ML algorithms can be understood using the SHAP analysis (Figure 4.34). The SHapley Additive exPlanations (SHAP) methodology identifies the features contributing most to the model prediction. We can find that the conc (concentration of the heavy metal) descriptor is the most contributing one in all the EL algorithms. The Shapley values reflect the average marginal contribution of a feature value across all possible feature coalitions, both in terms of magnitude and direction.

Table 4.18: Statistical parameters and selected features from the developed ML models.

Method	R^2	$Q^2_{(LOO)}$	MAE train	RMSE C	Q^2_{F1}	Q^2_{F2}	MAE test	RMSE P	Optimized hyper parameters
Random Forest	0.96	0.72	0.13	0.2	0.94	0.94	0.14	0.19	'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 80, max_features:1.0, bootstrap: True random_state:0
Gradient Boost	0.99	0.77	0.06	0.08	0.82	0.81	0.2	0.34	'max_depth': 6, 'min_samples_leaf': 4, 'min_samples_split': 4, 'n_estimators': 130, max_features: None, random_state:0
Extreme Gradient Boost	0.94	0.46	0.16	0.26	0.83	0.83	0.25	0.32	'booster': 'gbtree', 'colsample_bytree': 0.3, 'max_depth': None, 'min_child_weight': 1, 'n_estimators': 70, 'subsample': 0.5
AdaBoo st	0.88	0.64	0.31	0.37	0.72	0.71	0.33	0.41	'loss': 'linear', 'n_estimators': 170, random_state:None

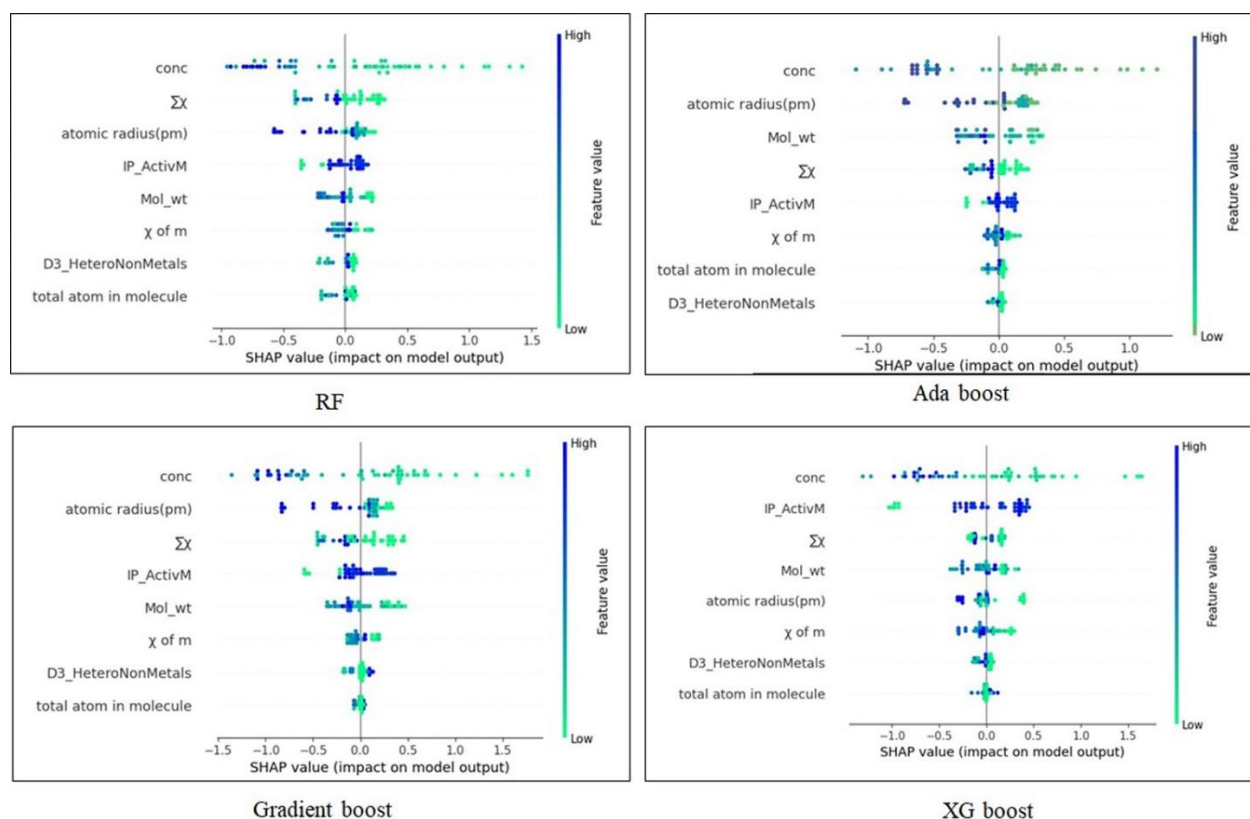


Figure 4.34 SHAP summary plot.

4.6.2. Results of Model Validation for all ML Methods

In order to determine if heavy metals and TiO₂ nanoparticles had any cytotoxic effects, eight important periodic table-based features were selected. The final models developed with RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost were evaluated using MAE_{train}, RMSE_{train}, R², and Q² for the training set and MAE_{test}, MSE, RMSE_{test}, Q²F₁, Q²F₂ metrics for the test set, and the results are shown in **Table 4.18**. According to the results, the MAE_{test} (0.14) was found to be the least for the test set in the RF method followed by AdaBoost, Gradient Boost, and XG Boost. The XG Boost gives the highest R² (0.99) for the training set while AdaBoost gives the lowest R² (0.88) with the highest MAE_{test} (0.33). Cross-validation statistics were obtained based on 20 times 5-fold repetitive CV along with 1000 times shuffle split CV (Mean±SEM) method. It is done to protect the model from overfitting when the data is limited. The results of the CV have a clear indication that the models do not memorize the correspondence between the descriptors since the outcome as R² is highest and the MAE value is lowest for the RF model after the repetitive CV method which suggests the superiority of the RF model to other models. **Figure 4.35** represents the cross-validation statistics based on 20 times 5-fold repetitive CV and 1000 times shuffle split CV method on R² and MAE for the developed ML model.

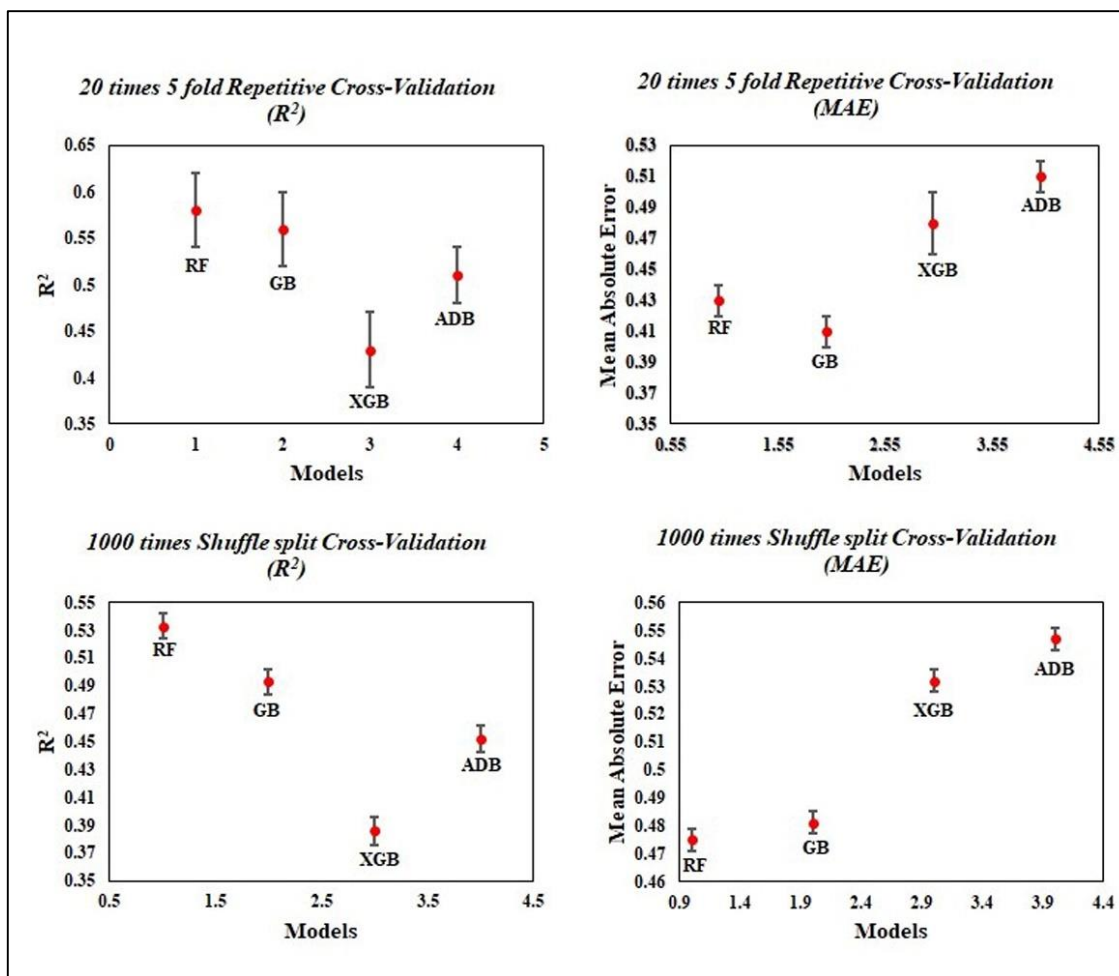


Figure 4.35 Cross-validation statistics based on 20 times 5-fold repetitive CV and 1000 times shuffle split CV method (Mean $\pm$ SEM)

4.6.3. General Mechanism of Toxicity

In the process of screening all descriptors from different ML methods, some common descriptors for heavy metals were discovered that are clear indicators of their importance to HK-2 cells. We found that the concentration of the heavy metal (conc), the atomic radius of the metal, electronegativity, and molecular weight of heavy metal influence HK-2 cells' survival. It was observed that conc, mol wt, atomic radius, and the total number of atoms in the molecule were of high importance in all the models. The increase of conc, mol_wt, and total atoms in a molecule is believed to increase toxicity. The severity of heavy metal is also time and dose-dependent. Among many other factors valence state plays an important role in metal toxicokinetics and toxicodynamics. Many studies have shown the increased level of concentration of heavy metals is correlated with the severity of hepatotoxicity and nephrotoxicity (Lund et al., 1991). Lead causes toxicity through an ionic mechanism followed by ROS (reactive oxygen species) generation. Another, biomarker for ROS is lipid peroxidation; (Niki, 2008). As a result, the free radical loses electrons and ultimately causes lipid peroxidation inside the cell membrane. The catalytic property of the metals is also responsible for increasing the toxicity of NMs causing a decrease in cell survival (Pujalté et al., 2011), and is well depicted in **Figure 4.36** The electronegativity and atomic radius are responsible for enhancing the catalytic property of the metal. The cationic form of the metal

thus generated also catalyzes the lipid peroxidation process (Roy and Roy, 2022) through the enhancement of endocytosis and the intrinsic property of the heavy metal. Their toxicity may be associated with internalization and bioaccumulation in the HK-2 cell. The increase in the concentration of heavy metal and its adsorption to nano TiO₂ induces toxicity by increasing the generation of ROS in the HK-2 cells.

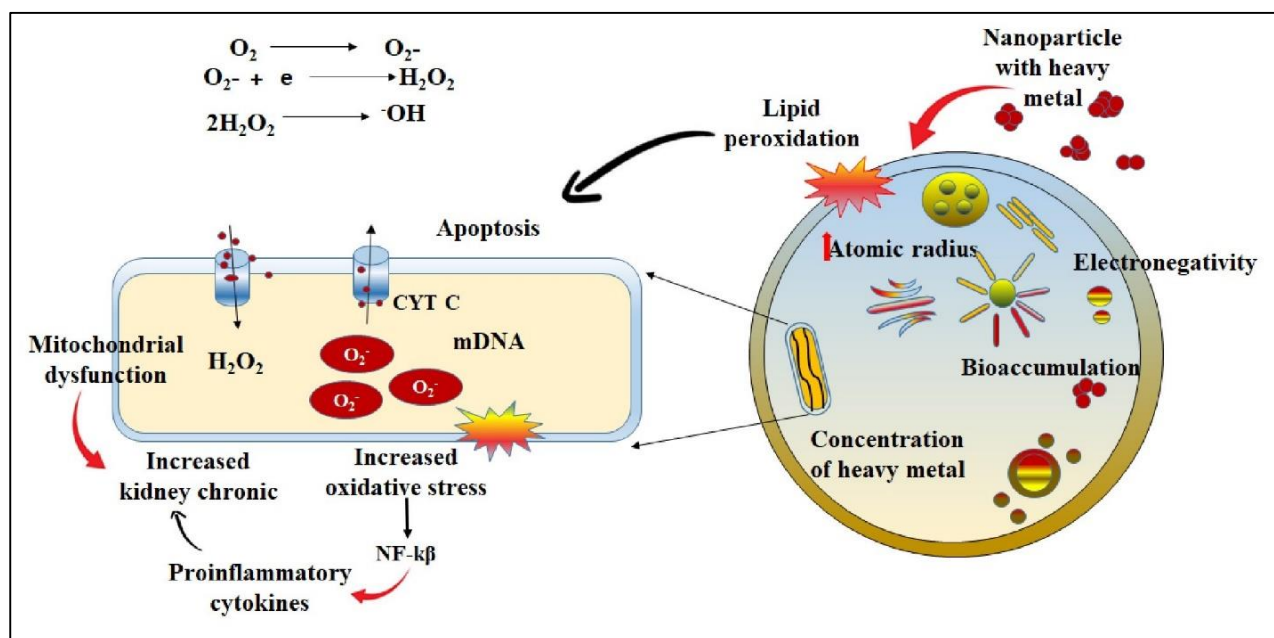


Figure 4.36 General Mechanism of toxicity to HK-2 cell by heavy metals (study 6).

4.6.4. Comparison with the previous work

The present work dealt with the model development of heavy metals of different concentrations with simple periodic table descriptors using various ML methods. The results obtained from the ML method suggest that the models have better predictivity than the models developed previously by Sang et al. (2022) as shown in **Table 4.19**. Sang et al. (2022) applied the random forest algorithm and AdaBoost algorithm for QSAR modeling using quantum mechanical descriptors while the present study involved the random forest algorithm, AdaBoost algorithm along with gradient boost and XGboost algorithm using simple periodic table-based descriptors that are easy to interpret and can be calculated in no time without the involvement of expert personnel. These descriptors simplify the nanostructure property calculation and determine the nano-level interaction without much computational intervention. The use of such descriptors saves time; they are also cost-effective and have a clear and straightforward physical meaning which facilitates the mechanical interpretation of the QSAR models developed with these descriptors. A direct comparison was not possible due to different dataset divisions and descriptors but the results obtained in the present work for the RF method were superior to that of the previous work.

Table 4.19: Comparison of the current work with the previous study.

Descriptors	Method	R^2	$Q^2_{(LOO)}$	MAE _{train}	RMSE _c	Q^2_{F1}	Q^2_{F2}	MAE _{test}	RMSE _p
Periodic table-based (Current study)	Random Forest	0.96	0.72	0.13	0.2	0.94	0.94	0.14	0.19
Quantum mechanical (Sang et al.)	Random Forest	0.85	0.70	-	0.06	0.86	0.85	-	0.10

4.7. Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.

To explore the physiochemical properties influencing the Zeta potential of the metal oxide nanoparticles (MeOx NPs), property-based modeling was performed considering Zeta potential as the Y-response (Model 7.1). Model 7.1 was developed with basic periodic table-based descriptors. The different validation metrics showed the models to be robust and of good predictivity. Furthermore, toxicity-based modeling (Model 7.2) was conducted to illustrate the impact of Zeta potential on the toxicity of BEAS-2B cell damage. The modeling aimed to create robust and predictive property and toxicity-based models capable of predicting new novel MeOx NPs with enhanced features. **Figure 4.37** shows the bubble plot for both dataset 7.1 and dataset 7.2. The green and red colors indicate the positive and negative coefficients of the respective descriptors. The size of the bubble represents the importance of the descriptors, where the smaller bubble size has less contribution to the respective endpoints than the larger bubble descriptor. The Williams plot in **Figure 4.38** shows that three compounds were outliers in the cell damage dataset. According to the PRI tool estimation on a true external set, out of 49 MeOx NPs, we confidently predicted 39 with good accuracy using this simple tool. This means that we were able to make predictions for untested metal oxides with greater confidence.

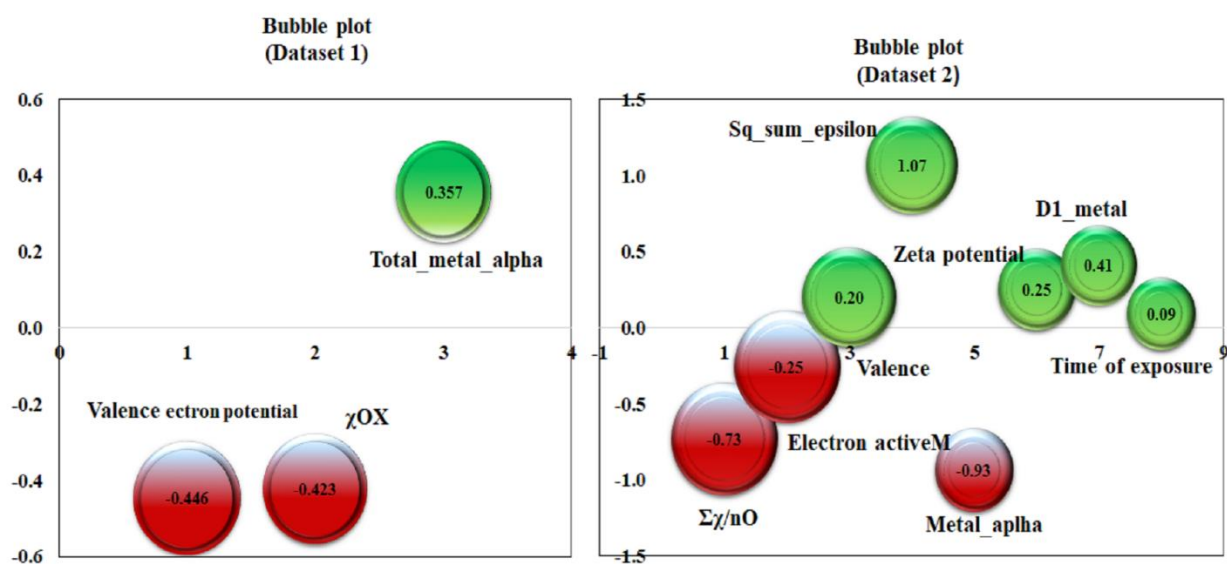


Figure 4.37 Bubble plot for dataset 7.1 (Model 7.1) and dataset 7.2 (Model 7.2).

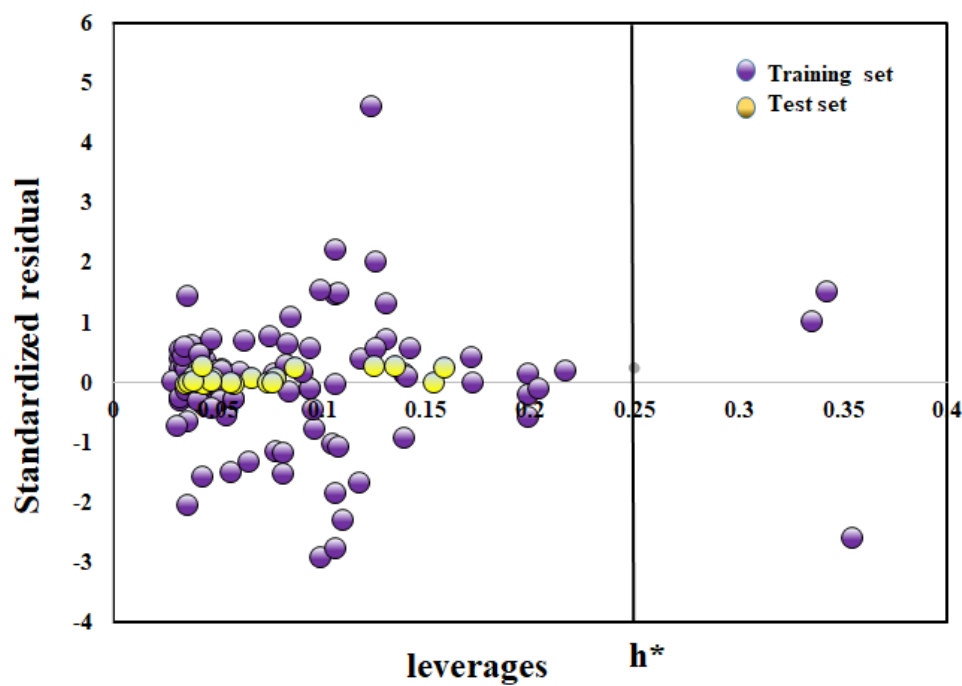


Figure 4.38 Williams plot for cell damage endpoint (model 7.2).

4.7. 1. QSPR model for Zeta potential

ζ -potential is the key parameter from the regulatory point and can directly affect the NPs' behaviour in the solution and their interaction with the biological organism (**Figure 4.39**). 18 MeOx NPs were modeled against the ζ -potential endpoint to obtain the partial least squares (PLS) model with one latent variable (LV).

Model 7.1 (PLS)

$$\zeta = 35.9157 - 8.4317 \chi_{ox} + 2.0002 \text{ tot_metal_alpha} - 0.1854 \text{ valence electron potential}$$

$$N_{train} = 13; R^2=0.80; Q_{(loo)}^2=0.67; N_{test} = 5; Q_{f1}^2=0.68; Q_{f2}^2 = 0.67; LV=1; MAE =; F=44.20; p=0.002 \quad (4.10)$$

Model 7.1 considered three descriptors to evaluate the influence of Zeta potential based on basic attributes. In equation 4.10, N_{train} and N_{test} stand for the number of training and test set compounds, respectively. R^2 is the determination coefficient; $Q_{(loo)}^2$ is LOO (leave-one-out) cross-validation R^2 . Again, Q_{f1}^2 and Q_{f2}^2 were calculated for external data predictions. The model parameters suggested the good predictive ability of the developed model as it passes the various statistical criteria (Golbraikh, and Tropsha, 2007). The descriptors depicted in the model also interpreted the influence of Zeta potential as discussed below.

The descriptor χ_{ox} pertains to the oxidation number of the metal, which represents the hypothetical charges within an atom. The Zeta potential decreases as the oxidation number increases, as indicated by the negative coefficient of the descriptor. A lower (negative) oxidation number indicates a higher electronegativity of the metal, which determines the electron distribution in a molecule. The metal's electronegativity also influences the catalytic property of the cationic form and the surface charge formed around the metallic oxide surface. The high electronegative surface of MeOx NPs (Kwon et al., 2001) affects their behavior and stability, thus determining the net charge of ions in a given medium. Certain MeOx are unstable and tend to become stable by increasing the size of NPs through agglomeration. To prevent this aggregation, NPs attract negative or positive ions from the medium to build a diffusion double layer. The electronegativity of the NPs is also dependent on the medium's pH (Toropov et al., 2018). In colloidal solutions, negatively charged metal oxides decrease the Zeta potential, which reflects stability based on the aggregation phenomenon. This is well observed in MeOx NPs, where an increase in the oxidation number (χ_{ox}) decreases the Zeta potential as in WO_3 NPs where the χ_{ox} value is 6 and Zeta potential value is -23 mV and vice versa occurs for NiO NPs where χ_{ox} value is 2 but the Zeta potential value is 34.4 mV.

Valence electron potential (valence eV ϕ) determines the elements' reactivity and is based on the charge of the valence electron and the ionic radius. It is denoted by equation 4.11

$$-eV = \frac{kn}{r} \quad (4.11)$$

Here, k expresses the energy (in electron volts) exerted by the valence electrons for an element with valence 'n', and the ionic radius 'r'.

This descriptor negatively contributes to the ζ -potential suggesting that with the increase of the valence electron potential of the metal, there will be a decrease in ζ -potential value as observed in Mn_2O_3 NP that has a valence electron potential value of 220 eV whereas the Zeta potential is -15.9 mV. On the other hand, Co_3O_3 NP shows an opposite result where the decrease in the valence electron potential value (38 eV) shows an increase in ζ -potential value (22.6 mV). MeOx NPs with large ionic radius tend to have low electron potential as it is inversely proportional to the ionic radius of the NPs. Thus, large-size NP particles tend to have low electron potential. NPs with lower valence potential are prone to leave the outer orbital easily than the one with higher potential thus allowing Electrostatic Double Layer (EDL) formation easier. If the solution with NPs shifts to lower ionic strength, then the ζ -potential increases as the EDL expands to balance the electrostatic force thus allowing the dispersion of NPs.

Tot_metal_alpha descriptor defines the core environment of the metal. It also defines the molecular bulk of the metal oxide. This descriptor has vital characteristics that are heavily influenced by the number of metals present in the metal oxide. Furthermore, the electronegative density of the metal is a crucial factor in determining the surface charge and stability of the NPs in the solution. Research indicates that an increase in the surface charge of the metal surrounding the NPs can significantly aid in its dispersion within the media and prevent flocculation. The positive regression coefficient suggests that an increase in the surface charge of the metal surrounding helps the NPs to be dispersed in the media and thus avoid flocculation. This phenomenon is observable in Yb_2O_3 NP with a high tot_metal_alpha value (13.6) having the highest ζ -potential (46 mV); on the other hand, SnO_2 NP with a descriptor value of 2.88 has a -20.5 mV ζ -potential value.

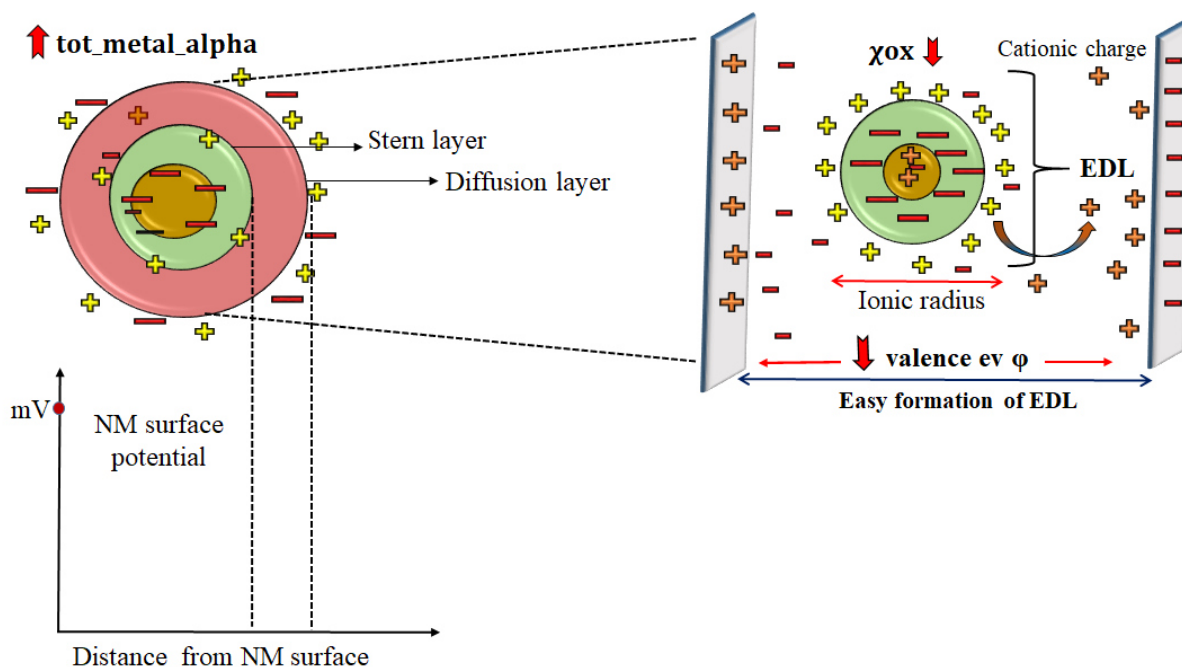


Figure 4.39 Zeta potential formation and influence phenomenon in respect to the modelled descriptors

4.7.2. QSPR model for cell damage

Model 7.2 (PLS)

$$\begin{aligned}
 \text{cell damage} = & -1.681 - 1.11 \frac{\Sigma \chi}{nO} + 0.295 \frac{\text{sq}_{\text{sumepsilon}}}{N} + 0.318 D1_{\text{metal}} - \\
 & 0.263 \text{Metal alpha} + 0.035 \text{time of exposure} + 0.00057 \text{zeta potential} + \\
 & 0.0057 \text{Electrons}_{\text{ActiveM}} - 0.079 \text{valence}
 \end{aligned}$$

$$\begin{aligned}
 N_{\text{train}} = 111; R^2 = 0.62; Q^2_{(loo)} = 0.54; \overline{rm^2_{loo}} = 0.389; \Delta rm^2_{loo} = 0.246; N_{\text{test}} = \\
 21; Q^2_{f1} = 0.653; Q^2_{f2} = 0.652; \overline{rm^2_{loo}} = 0.532; \Delta rm^2_{loo} = 0.183 \text{ LV}=7; MAE_{\text{test}} = 0.206; \\
 F = 19.87;
 \end{aligned}
 \tag{4.12}$$

Model 7.2 utilized eight descriptors to evaluate crucial attributes that can impact cell damage. Equation 4.12 shows the number of compounds used in the training and test sets represented by N_{train} and N_{test} , respectively. Additionally, R^2 and $Q^2_{(loo)}$, the determination coefficient and leave-one-out (LOO) cross-validation were employed. Furthermore, external data prediction calculations were made using Q^2_{f1} and Q^2_{f2} . The model parameters confidently demonstrate its exceptional predictive ability, meeting various statistical criteria (Golbraikh and Tropsha, 2007). The descriptors used in the model were well-interpreted and are comprehensively discussed in a later section. Note that zeta potential has appeared as a significant descriptor in defining the cell damage. On removal of Zeta potential as a descriptor, the model quality decreases. Using the predictions from one model as a descriptor for another model is done to fill the data gap or to determine the missing values. This approach is similar to the imputation methodology, which creates a model embedded within another model. Instead of using dummy variables for quantitative prediction, a useful imputation method can predict various types of inputs. It is worth noting that many existing works utilize imputation techniques for this purpose (Whitehead et al., 2023). In QSAR studies, it is not unusual to use a model-derived prediction as a descriptor for other model development or prediction when the endpoint has been tested under different experimental or varying conditions (as in the case of interspecies modeling). This approach is reliable and aims to establish a correlation between different conditions to fill the data gap.

4.7.3. Chemical Read-Across Analysis

The developed QSPR (PLS) model for dataset 7.2 provided 8 descriptors that were utilized for read-across predictions. Three similarity-based prediction methods, namely Euclidean Distance (ED)-based, Gaussian Kernel (GK) Similarity-based, and Laplacian Kernel (LK) were employed. Upon optimizing the dataset, it was concluded that the RA-ED (read across-based on the Euclidean Distance) function outperformed the others, as shown in **Table 4.20**. The Read-Across-v4.0 software (Banerjee et al., 2022) was utilized for this work. After performing RA, the resultant Q^2_{f1} increased from 0.65 to 0.766.

Table 4.20: Results for read across prediction using different similarity-based approaches.

Feature combinations	Hypothesis	Hyper parameter				Statistical parameters			
		σ	γ	Distance threshold	Similarity threshold	Q_{F1}^2	Q_{F2}^2	MAE	RMSEP
Model 7.2 (132 MeOx NP) 8 descriptors (7 LV)	RA-ED	1.75	1.75	1	0	0.766	0.765	0.177	0.252
	RA-GK					0.764	0.763	0.178	0.254
						0.724	0.724	0.195	0.274
	RA-LK								

4.7.4 Interpretation of the descriptors

The periodic table descriptor $\Sigma\chi/nO$ stands for the total metal electronegativity in a specific metal oxide relative to the number of oxygens. This descriptor takes into account the crucial role of oxygens in causing cell damage. Concerning the cell damage endpoint, this descriptor has a **negative** effect, indicating that an increase in the number of oxygens compared to the electronegativity sum results in a lower ratio of the descriptor. Thus, a high concentration of oxygens in the metal oxide can expedite the oxidative damage process, leading to the production of more reactive oxygen species (ROS) and causing more cell damage. The CoO NP shows that a high $\Sigma\chi/nO$ value (1.88) leads to less cell damage (-4.38) whereas a low value ($\Sigma\chi/nO=0.77$) has more cell damage (-2.50) as observed for the TiO₂ NP. The production of ROS (reactive oxygen species) can enhance the catalytic activity of Fenton/Fenton-like reactions, but can also result in cellular damage (Roy et al., 2019). ROS can break down the basic components of the cell, including DNA, proteins, and lipids. ROS can cause double-stranded breaks in DNA by converting guanine to 8-oxoguanine. This conversion can lead to mispairing with adenine, resulting in transversion mutations. Proteins can also be damaged when their amino acid side chains are oxidized by ROS. Exposure of lipids to ROS can result in lipid peroxidation, which can cause cell damage and generate reactive by-products that further damage the cell.

The second-generation periodic table-based descriptor $\text{sq_sum_epsilon}/N$ or $(\Sigma\epsilon/N)^2$ stands for the sum of the electronegativity of the atoms of the metal oxide scaled by the number of atoms present, which is calculated based on the electronegativity count ($\Sigma\epsilon$) of the oxides.

$$\Sigma\epsilon = \epsilon_{\text{metal}} * N_{\text{metal}} + \epsilon_{\text{oxy}} * N_{\text{oxy}} \quad (4.13)$$

In equation 4.13, ϵ_{metal} and ϵ_{oxy} are the electronegativity count of metal and oxygen atoms, respectively and N_{metal} and N_{oxy} are the number of metal and oxygen atoms. The **positive**

coefficient of the descriptor infers that an increase in electronegativity favors the rise in cell damage as in CuO nanoparticles where a high $(\sum \epsilon/N)^2$ value (9.93) causes more cell damage (-2.87), whereas the nanoparticles with the presence of low electronegativity ($(\sum \epsilon/N)^2=0.018$) causes less toxicity (-4.625) as observed in Sb₂O₃ nanoparticles. It has been observed that due to the presence of high electronegativity, the atoms pull electrons from their neighboring elements or molecules leading to the development of an electrostatic bond with the protein constituents present in the biological system. The high electronegativity also influences the formation of metal cations. The increase of catalytic property of the metal cation enhances the toxicity through the generation of reactive oxygen species, thus, causing damage to the membrane of the respective cell (Roy et al., 2019). The high electronegativity steals electrons from the molecules producing free radicals. Free radicals are unstable and highly reactive. These short-lived radicals are unable to leave the sub-cellular location where they are generated without being reduced leading to oxidative damage (Ge et al., 2009). The presence of high electronegativity metals in the cellular membrane can lead to the leakage of cellular content (Roy and Roy, 2023).

The **D1_{metal}** descriptor signifies the total number of metals in the MeOx NPs. An increase in the number of metals can have a detrimental effect to cells by impacting ROS generation. A higher presence of metals tends to cause more damage to the cell as compared to a low count of metal in the metal oxide. This is well observed by the **positive** coefficient of the descriptor. The positive coefficient of the D1 metal descriptor indicates that an increase in metal count in the MeOx NPs causes more cell damage (-2.63) as observed in Fe₃O₄ MeOx NP ($D1_{metal}=3$). On the other hand, the low metal count as in CoO ($D1_{metal}=1$) nanoparticles causes less cell damage (-4.375). Metal ions possess dual properties. They can generate reactive hydroxyl radicals, resulting in oxidative damage to proteins. Moreover, they can bind non-specifically to amino acid residues and replace existing metal ion enzyme active sites, leading to abnormal protein folding. Protein aggregation diseases are a type of neurodegenerative diseases that occur when proteins lose their structure and deposit in the brain, causing a loss of their function. These diseases are the most common type of neurodegenerative diseases. Many of these structures are highly toxic to the cell (Dobson, 2004). The folding of the protein also causes a drop in the immune system because certain structures do not induce the production of antibodies (Bossy-Wetzel et al., 2004).

Metal alpha defines the core environment of the metal. This descriptor represents the ratio of the number of core electrons to the number of valence electrons. Metal alpha descriptors describe the electron density of the metal. This descriptor is calculated using equation 4.14

$$\alpha_{metal} = \lambda * \mu \quad (4.14)$$

Further, λ is $(Z_{metal} - Z_{vmetal})/Z_{vmetal}$, and μ is the $1/(PN_{metal})$; here Z_{metal} is the atomic number, Z_{vmetal} stands for the valence electrons of the metal, PN_{metal} stands for the periodic number in the periodic table. The **negative** coefficient of the descriptor signifies the low electron affinity of the metal oxide to accept electrons. This means that the metal has a propensity of having a cationic charge, which leads to the catalytic power of metal cations. For example, in WO₃, the metal alpha value is 7.2 and cell damage is -4.57. In contrast, Al₂O₃ with a metal alpha value of 1.66 causes higher damage to cells (-2.8). Metal cations are more harmful than normal nanoparticles. This is because their electropositivity and inherent toxicity increase significantly with atomic weight. In addition, the formation of metal–ligand bonds has a direct impact on

the metal's toxicity. Furthermore, it is a well-established fact that each metal has an affinity constant for various ligands, which means that most metal cations can form stable complexes with a wide variety of ligands, further increasing their potential toxicity.

In the field of physical chemistry, the **Zeta potential** (ζ -potential) is a crucial parameter that measures the surface charge of particles relative to their size. In the colloidal system, ζ -potential is widely used as an indicator to reflect stability. It is important to note that NPs with higher positive charges can be more harmful than those with higher negative charges. Moreover, positively charged NPs with like charges interact more significantly with cells, leading to greater cell damage. Another crucial factor to consider is that NPs with a higher Zeta potential, regardless of their charge, are more easily absorbed by cells due to the electrostatic interaction between dispersed particles and the effective electric charge on the surface of the NP (Toropov et al., 2018). This feature is particularly relevant to their biological activity, especially their ability to bind to and be absorbed by cell membranes. For instance, Cr_2O_3 NP, which has a high ζ -potential (2130 mV) and a high cell damage propensity; whereas, the metal oxide Y_2O_3 , with a low ζ -potential (-23 mv), causes less damage to cells (-4.5). The increase in Zeta potential enhances the accumulation of nanoparticles on the surface of cells. The intensity of accumulation determines the toxicity parameter of the nanoparticles. The concept of ζ -potential plays a vital role in adhesion to the hydro-water interface and solid surfaces, providing an idea about the viability and permeability of the cell membrane under stress. As most of the cell surface carries a negative charge, metals with higher zeta potential can easily enter the cell, increase the production of ROS, and have a mechanical effect on the membrane, leading to depolarization of the membrane and cell damage.

Electron Active M is a representation of the number of electrons that an active metal possesses. Active metals are known for their quick and robust reactions owing to the electron arrangement in their structure. These metals contain free electrons in their outermost shell that can readily create a cation by interacting with other atoms and initiating a chemical reaction. The delocalized electrons can easily interact with macro proteins, leading to the acceleration of damage to the biological membrane. Unstable free radicals, which are electrons in the outermost shell, tend to react quickly to become stable and initiate oxidative stress in the cell. The Electron Active M descriptor indicates that a **positive** coefficient leads to more oxidative stress and more damage to the cell with an increase in free radicals. WO_3 has a high descriptor value of 74, resulting in high cell damage (-2.8), while Cr_2O_3 NP has a low descriptor value of 24, leading to low cellular damage (-4). Transition metals are capable of forming coordinate complexes with the imidazolyl group of histidine. These metal ions are redox-active and can play a crucial role in the production of Reactive Oxygen Species (ROS) within the cell. The reduced forms of these redox-active metal ions are involved in the Fenton reaction, which generates hydroxyl radicals from hydrogen peroxide. Similarly, the Haber-Weiss reaction involves the oxidized forms of redox-active metal ions and superoxide anion, which generates the reduced form of the metal ion. This reduced form can then be coupled to Fenton chemistry to produce hydroxyl radicals. ROS further accelerates the damage to the cell.

Valence, which is symbolized as V, is a factor that contributes to cell damage. It indicates the number of electrons in the outermost shell of an atom that is available for chemical bonding and is similar to other descriptors that provide information about free electrons. The insights obtained from the developed model 7.2 strongly suggest that an increase in valence (7) leads to a decrease in cell damage (-3), as observed in MnO_3 NPs. This is supported by the **negative**

regression coefficient of the descriptor. Conversely, a low valence (2) poses a greater risk of harm to cell damage (-2), as seen in ZnO MeOx NP. Atoms with fewer electrons in their outer shell tend to lose them and become metal cations, which can damage cells (Roy et al., 2019). Cations aid in the transportation of metal ions across the cell surface by interacting with its negatively charged surface. Unfortunately, this interaction can lead to DNA damage through processes such as delocalization, redox chemistry, and the generation of reactive oxygen species (ROS).

Our research aimed to examine how the **time of exposure** to metal oxide affects cell damage, regardless of other physiochemical properties of MeOx NPs. Our findings were remarkable, indicating that exposure time plays a crucial role in cell damage, where prolonged exposure times increase the damaging potential. For instance, exposing cells to WO₃ NP for 7 hours resulted in a cell damage score of -2.75. On the other hand, exposure to Yb₂O₃ for only 1 hour still resulted in a severe score of -3.5. These results demonstrate the significance of considering exposure time when evaluating the potential risks of metal oxide exposure. When nanoparticles (NPs) are exposed to living organisms for an extended period, they can cause inflammatory conditions that lead to physical, muscular, and neurological degeneration or increased intensity of oxidative stress. This happens because longer exposure times enhance the toxicity mechanism of NPs. On the other hand, short-term exposure does not show significant effects on the cells. NPs can induce oxidative stress by impairing antioxidant defenses in humans when they are chronically exposed to NPs.

4.7.5 Importance of Zeta potential as a descriptor

The developed QSPR model without the Zeta potential descriptor shows $R^2 = 0.47$; $Q^2_{(100)} = 0.34$, which is quite less than the desired acceptance criteria. The obtained results indicate that the fitting and robustness of the developed model without the presence of the zeta potential descriptor is unsatisfactory. Therefore, to achieve the fit and predictive power of the model, we included the Zeta potential descriptor along with the other seven descriptors. In the presence of Zeta potential, the statistical quality and internal validation metrics increased ($R^2 = 0.62$; $Q^2_{(100)} = 0.54$) showing the stability and predictive ability of the model. The observation also infers the importance of the Zeta potential descriptor in the development of a good model to predict the basic attributes of the toxicity profile for the cell damage dataset.

4.7.6. Utilization of the metal oxide cell damage knowledge for cancer treatment

Nanoparticles (NPs) have shown immense potential in treating various diseases owing to their small size and high surface-to-volume ratio, which makes them an effective drug delivery system. Metal NPs can lead to greater signal amplification, greater sensitivity, and higher detection. However, NPs with properties that generate reactive oxygen species (ROS) can increase cell damage. In cancer cells, rapid proliferation leads to an imbalance of oxygen, abnormal structure, and blood supply, making the tumor microenvironment (TME) prone to hypoxic conditions (Gilkes et al., 2014). Insufficient oxygen reduces ROS generation, which decreases the efficacy of oxygen-dependent therapies like photodynamic therapy (PDT), chemodynamic therapy (CDT), and radiation therapy. The information derived from the positive contribution of **D1_{metal}** descriptor (model 7.2) draws attention to the fact that metal oxides are good candidates for generating oxidative stress in the cell. The treatment based on PDT requires oxygen concentration, so a low oxygen level in TME lowers the efficacy of such

treatment. The $\Sigma\chi/nO$ descriptor suggested a higher oxygen requirement for damaging the cells, and it indicates the presence of oxygen in the metal oxide nanoparticle that can increase the sensitivity to PDT. Furthermore, transition metals can catalyze Fenton/Fenton-like reactions (Cao et al., 2021), generating highly oxidative species that can kill tumor cells. The presence of electronegativity of the metal oxides helps in the permeability of the NPs through the cell membrane. The formation of cationic forms of metal can also affect the pH of the cell, and increase the catalytic properties of metal oxides, thereby increasing ROS generation. Tumor cells have a mechanism for dealing with hypoxia, acidosis, and high glutathione (GSH) levels, which promote drug resistance, especially for ROS-dependent drugs (**Figure 4.40**). However, metal oxides can reverse the TME condition by supplying oxygen and suppressing hypoxia-inducible factor 1 and CD39/CD73 in T cells, which reduces the immunosuppression effect of tumors.

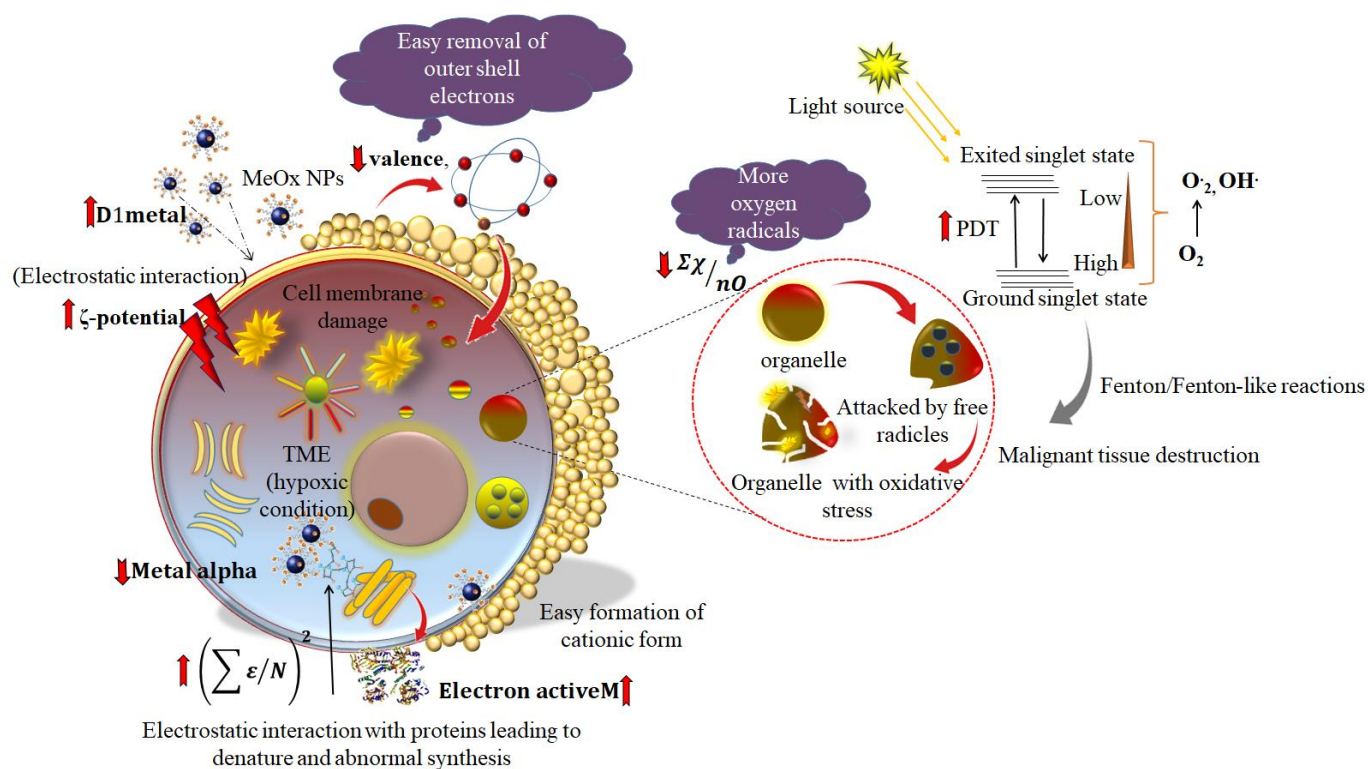


Figure 4.40 Interpretation of descriptors with respect to cell damage (endpoint) mechanism for cancer cells (study 7).

4.7.7. Comparison with the previously published literature

This study successfully develops a QSPR model with a cell damage endpoint that uses the Zeta potential value as a descriptor. The descriptor was calculated using another model that used Zeta potential as the endpoint (Y-response) for its QSPR model development. The QSPR models were developed with simple periodic table-based descriptors that do not depend on size

or any other experimental conditions. These descriptors are easy to calculate, less expensive, and can be calculated by anyone without the need for expert personnel.

The study provides in-depth knowledge about the properties and causes of toxicity of nanoparticles using simple regression-based models. Although it is important to note that a direct comparison with a previous study by Toropova (Toropova et al., 2015) is not possible due to the different data division methods (five random splits) and the use of different types of descriptors (Optimal nano-descriptors) and also due to the dissimilar modeling methods (Monte Carlo method), it is clear that the statistical metric values for the developed model in the present study are similar to those of the previous study (the best-split results only shown) as presented in **Table 4.21**. Furthermore, we have proposed an effective mechanism to treat cancerous cells with the cell-damaging property of MeOx NPs.

Table 4.21. Comparison with a previous study for the statistical parameters.

Sl. no.	Q^2	R^2_{Train}	S_{train}	Q^2F_1	R^2_{Test}	MAE_{test}	S_{test}
Current study (Model 7.2)	0.538	0.621	0.361	0.768	0.767	0.181	0.377
Previous study (best split) (Toropova et al., 2015)	0.486	0.512	0.387	-	0.822	-	0.318

4.8. Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across, and ML algorithms: QSAR models for *Daphnia magna* and *Zebra fish*.

One crucial aspect of nano-toxicological modeling is exploring the correlation between the physiological properties of nanoparticles and their toxicological behavior. To achieve this, we first need to develop and validate models, followed by understanding a clear definition and interpretation of the variables that are important for modeling the respective toxicity endpoint. By decoding the information in the descriptors, we can understand how nanoparticles affect organisms and their reproduction in aquatic environments. The statistical parameters suggest good predictivity of both datasets. **Figure 4.41** confirms and supports the developed models' predictability.

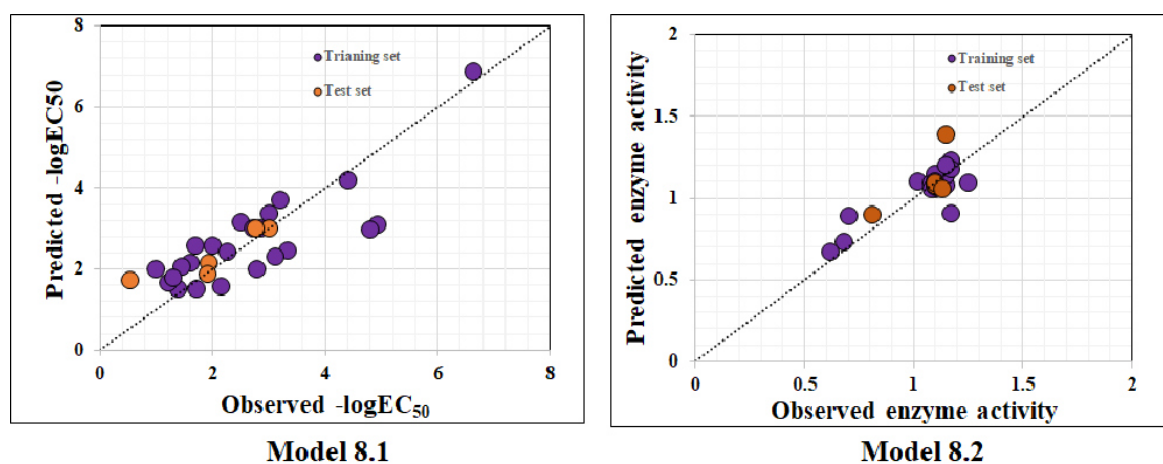


Figure 4.41 Scatter plots of model 8.1 (*D. magna*) and model 8.2 (zebrafish).

We have also checked the AD status of both datasets and found that for dataset 8.1, two compounds from the training set were outside the AD. whereas for dataset 8.2 no compound from the training and test set was outside the AD space (**Figure 4.42**).

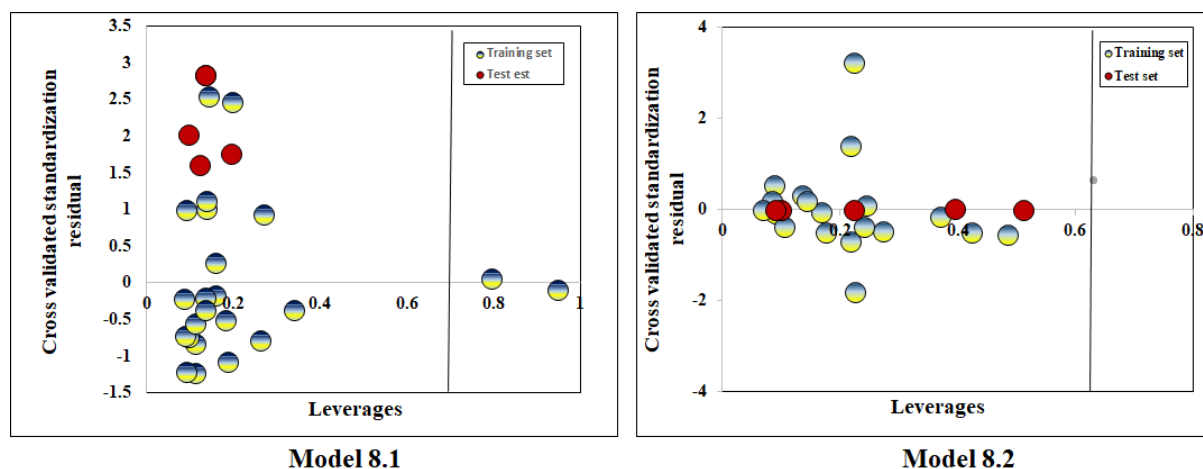


Figure 4.42 AD plots of model 8.1 (*D. magna*) and model 8.2 (zebrafish).

4.8.1. QSAR model with *Daphnia magna* (Model 8.1)

The structural effect analysis is one of the important tools in the field of ecotoxicity to predict hazards caused by the NPs to different organisms. Therefore, in the present study, the nano-QSAR toxicity modeling with 28 MeOx NPs for *Daphnia magna* (Dataset 8.1) was developed for the toxicity assessment. The developed model showed (Equation 4.15) good statistical performance concerning its endpoint ($-\log EC_{50}$). The PLS model obtained with three latent variables has metric values of $Q_{Loo}^2 = 0.57$, $R^2 = 0.69$, $Q_{F1}^2 = 0.74$. The basic molecular information like size, number of metals, oxidation number, and metal alpha were integrated into the developed model for *Daphnia magna* toxicity.

Model 8.1

$$-\log(EC)_{50} = 5.373 - 0.0712 * size + 1.220 * Nmetal - 0.806 * \chi ox + 0.0323 * sum alpha$$

Training set

$$n = 24, R^2 = 0.69, Q_{Loo}^2 = 0.57, s = 0.86, MAE =, F = 14.08, p < 0.05, LV = 3$$

Test set

$$n = 7, R^2 = 0.58, Q_{F1}^2 = 0.74, s = 0.86, MAE = 0.40, p < 0.05 \quad (4.15)$$

The **size** of nanoparticles (NPs) is crucial in determining their toxicity. Studies have found that there is an inverse relationship between the size of NPs and their accumulation in *Daphnia magna*, meaning that smaller NPs can easily pass through cell membranes compared to larger ones (Schultz et al., 2003). Consequently, the toxicity of *Daphnia magna* increases as the size of NPs decreases as in Fe_3O_4 with size 6 nm has a toxicity value of 6.64, since smaller particles can penetrate the organism's epithelial tissue, enter the bloodstream, and damage or alter cells in different organs. On the other hand, MgO with a larger size of 48 nm has a lower toxicity value (1.2). The size of NPs is also related to their surface reactivity, which greatly affects their dissolution behavior (Fischer et al., 2021). Studies (Rosenfeldt et al., 2015; Shin et al., 2021)

have suggested that NPs with sizes similar to those of the food consumed by *Daphnia magna* (1 μm up to 50 μm , and occasionally up to 70 μm) (Brun et al., 2017) are most harmful to the organism. Furthermore, smaller NPs are more toxic than larger ones, which has been demonstrated by several studies (Park et al., 2011). Therefore, the size of NPs is an important factor to consider when assessing their dispersion stability in a specific medium, and it is a driving parameter for their toxicity.

The **Nmetal** descriptor indicates the number of metals in the NPs. It has been observed that an increase in the number of metals can lead to higher toxicity to *Daphnia magna*. This is evident in NPs where an increase in metal content has resulted in a greater toxicity compared to NPs with a smaller number of metals. The presence of a higher number of metals as in **Ag₂O** ($N_{\text{metal}}=2$) may cause oxidative stress and genotoxicity with the reported toxicity value of 4.39. On the contrary, **CeO₂** with metal count 1 has a toxicity value of 1.99 signifying a less toxic effect than the higher metal count of metal oxides. Reactive oxygen species such as hydroxyl radical ($\text{OH}^\cdot$), superoxide anion ($\text{O}_2^\cdot$), and hydrogen peroxide (H_2O_2) are vital signaling molecules that can induce ROS (Reactive Oxygen Species) production. This ROS can also affect other signaling pathways such as MAPKs (mitogen-activated protein kinases), Nrf2, HIF-1 (hypoxia-inducible factor 1), and nuclear factor-kappa B (NF- κ B), which can result in severe apoptosis, DNA damage, and genotoxicity (Dominguez et al., 2015). However, the generation of ROS production is material-specific. The higher presence of metal count can cause excessive production of ROS which leads to DNA damage through its breakage, a phenomenon that has been shown in *Daphnia Magna* (Maselli et al., 2017; Park et al., 2010).

One of the descriptors that determines the toxicity of *Daphnia magna* is the oxidative number of the metals i.e., **χ_{ox}** . The **negative** coefficient of this descriptor indicates that as the value of the descriptor decreases, the toxicity increases as observed for **CeO₂**, where with an increase in the descriptor value ($\chi_{\text{ox}}=4$) there is a fall in the toxicity value (1.7), whereas the opposite is true as the oxidation number increases; for example, **Ag₂O** has a descriptor value of 1 while the corresponding toxicity value is 4.39. Nanoparticles with a lower oxidative number are more toxic than those with a higher value. When a metal has a lower oxidation number, it has fewer valence electrons in its outermost shell. This makes the metal behave as a reducing agent, with a greater tendency to lose electrons and act like a base. This can lead to the formation of cations, which can then participate in the formation of coordination complexes, synthesis of other compounds, or protein complexes (Remelli et al., 2016). Studies (Sheehan et al., 2001) have also reported the inactivation of GST (Glutathione-S-transferase) isoforms when organisms were exposed to metal ions. It is known that the GST enzyme has many functions, and different forms of the enzyme have different mechanisms of action. Thus, the inactivation of the GST enzyme could proceed through the non-specific binding of metal ions to the thiol group of the GST moiety, forming metal thiolate clusters whose conformation is different from the non-bound thiol groups and this change in the structure would affect the function.

The descriptor **Sum alpha** refers to information about the core environment of the metal oxide (Roy and Roy, 2022). It indicates the number of metal and oxygen atoms present in the NP and provides insight into the molecular bulk of the metal oxide. A **positive** correlation has been observed between molecular bulk and the toxic effect on *Daphnia magna*. The electronegativity of the metal determines the charge and stability of the metal in the NPs. In turn, the presence of surface charge acts as a mediator for the formation of co-ordinate bonds with proteins and other compounds. A higher value of the descriptor causes an increase in the toxicity as in **CeO₂**, where the descriptor value (Sum alpha = 6.26) leads to a higher toxicity

(4.9). The opposite is observed in **MgO**, with a low descriptor value (Sum alpha=2.83) leading to lower toxicity ($-\log EC_{50} = 1.2$).

4.8.2. QSAR model with Zebra fish (Model 8.2)

A QSAR model (equation 4.16) was developed for the zebrafish (*Danio rerio*) dataset using the zebrafish hatching enzyme ZHE1 inhibition activity results as the endpoint for the abiotic enzyme assay. The GA algorithm was used for feature selection and model development through BSS method.

Model 8.2

$$\text{Enzyme Activity} = 0.892 + 0.00115 * \text{Mol}_{wt} + 0.128 * \text{Noxygen} - 0.0472 * \text{tot}_{\text{metal}_{\alpha}} - 0.00207 * \text{AtWt}_{\text{Activ}_{SM}}$$

Training set

$$n = 19, R^2 = 0.72, R^2_{adj} = 0.64, Q^2_{LOO} = 0.52, s = 0.105, \overline{r^2_{m(loo)}} = 0.38, \Delta r^2_{m(loo)} = 0.187, \text{MAE} =, F=12.87, p<0.05, \text{PRESS}=0.161$$

Test set

$$n = 5, R^2 = 0.94, Q^2_{F1} = 0.83, \overline{r^2_{m(test)}} = 0.56, \Delta r^2_{m(test)} = 0.189. \text{MAE} = 0.0263 \quad (4.16)$$

The descriptors appearing in Model 8.2 can demonstrate their potential towards the respective enzyme inactivation and at the same time describe different mechanisms of action based on four descriptors. It has been observed that the delay or inhibition of NPs towards the hatching enzyme activity ZHE1 (zinc-based matrix metalloproteinase hatching enzyme) is highly correlated with the response with an R^2 of 0.72.

The molecular weight (**Mol wt**) of metal oxides is an important descriptor that represents the molecular mass of the metal nanoparticles. It has been observed that the less soluble NPs can easily accumulate in the organs such as the heart, gall bladder, lung, and kidney of the zebrafish. During development, the zebrafish embryo is surrounded by a chorion, which is an acellular envelope that helps protect the embryo from foreign interference. The pores of the chorion play a critical role in the transportation of nutrients and oxygen from the aquatic environment to the embryo. The inactivation of the ZHE1 enzyme was found to be influenced by the molecular weight along with other parameters (Chen et al., 2020) by the present research work. **CuO**, with a molecular weight of 79.54, showed a less effect with an enzyme inactivation activity of 0.62. However, **In₂O₃** and **Co₃O₄**, with molecular weights of 277.63 and 240.79 respectively, were found to be more toxic with enzyme inactivation activity of 1.17 and 1.25, respectively. It has been observed that the large molecular weight nano substances do not enter the embryo but adhere to the surface of the chorion preventing oxygen exchange with the embryo (Ghobadian et al., 2015).

The **Noxygen** descriptor refers to the number of oxygen atoms associated with the metal oxide present in the nanoparticles (NPs). The metal oxide is believed to absorb photons and produce electron-hole pairs that interact with water and oxygen to form reactive oxygen species (ROS) (Bar-Ilan et al., 2012). These ROS can be toxic to *zebrafish* and cause damage to their tissues. This effect is known as oxygen toxicity and is a result of oxidative stress, which is the major pathway in nanoparticle toxicity. Exposure to metal oxide NPs can produce the hydroxyl

radical ($-\text{OH}^\cdot$), which is capable of inducing DNA damage and protein denaturation, demonstrating mutagenic and carcinogenic properties of nanoparticles (Manke et al., 2013). This is due to the phototoxic effect of the metal oxides in *zebrafish*, which is mediated via oxidative damage. The higher the oxygen count in the compound, the more dramatic the effect on the hatching of the zebrafish enzyme. **CoO** with oxygen number 4 showed an enzyme inactivation activity of 1.2, while **ZnO** ($N_{\text{oxy}}=1$) with less oxygen number in it reduced the ZHE1 inhibition activity to 0.7.

The **tot_metal_alpha** descriptor provides information on the core environment of metals present in the NPs. It is associated with the number of metals involved and their nature in the environment. A negative coefficient indicates that there are fewer metals, which means that the size of the NP will be smaller compared to NPs with greater metal counts. Metals with different chemical compositions have different developmental toxicity. The stability of the chemical composition is an important factor for inhibiting enzyme activity (ZHE1). The physical and chemical properties of the metal also have toxic effects on the zebrafish hatching enzyme. The release of metal ions in the physiochemical fluid can influence toxic potential. The histidine-based binding site is highly reserved in any fish-hatching enzyme. The dissolving metals may also impact the hatching of the fish. **Cr₂O₃** with a descriptor value of 15.3 has less effect on the inhibitory activity of the enzyme, whereas **SnO₂** with a descriptor value of 2.87 has more inhibitory activity for the ZHE1 enzyme. Therefore, for proper enzyme activity, the number of metals as well as their electron density and charge should be lower.

The descriptor **AtWt_Activ_SM** refers to the atomic weight of the active metalloid or semi-metal, which can affect the toxicity of the ZHE1 enzyme in zebrafish. This descriptor is negatively correlated with the toxicity endpoint. It helps to interpret the influence of metalloids or semi-metals on toxicity. The negative correlation suggests that their influence is not as significant as that of metals. This is because metalloids or semi-metals do not form cations quickly, and their electron clouds are smaller than those of metals (Roy et al., 2023). Therefore, the toxicity of metal NPs is much higher than that of NPs with less metallic characteristics. For instance, **SiO₂** with a descriptor value of 28.08 has a more inhibitory activity for the ZHE1 enzyme than metallic NPs with a descriptor value of 0 and an inhibitory activity of 1.25 (**Co₃O₄**).

3.3 Discussion similarity and read across (RA) predictions

In order to reinforce our choice for grouping, we employed the read-across method using the selected descriptors from both PLS models (model 8.1 and model 8.2). After optimizing the hyperparameters for both models, there was a significant improvement in the quality of external prediction quality, especially for dataset 8.1. For the *Daphnia magna* dataset, the RA-ED (read-across based on the Euclidean Distance) approach yielded the best results, while for the zebrafish dataset, the RA-LK (read-across based on the Laplacian kernel) approach was the best similarity-based prediction for 2 and 3 close source compounds, respectively. **Table 4.22** displays the results of the RA analysis.

Table 4.22. Statistical analysis of the developed RA models for both datasets with their optimized hyper parameters.

Feature combinations	Hypothesis	Hyper parameters				Statistical validation metrics			
		σ	γ	Distance threshold	Similarity threshold	Q^2_{F1}	Q^2_{F2}	MAE	RMSEP
Model 8.1 (<i>Daphnia magna</i>)	RA-ED	1.75	1.75	1	0	0.75	0.61	0.40	0.545
Model 8.2 (zebrafish)	RA-LK	0.25	0.25	1	0	0.95	0.95	0.01	0.024

4.8. 4 Discussion on ML (Machine Learning) models

We evaluated the performance of various machine learning techniques using supervised learning algorithms, such as Random Forest, AdaBoost, Gradient Boost, XGboost, and SVM (Support Vector Machine). We used the selected features from both models (model 8.1 and model 8.2) to perform respective model development using supervised regression algorithms. Our findings indicate that Model 8.1 produced better prediction outcomes with the ML algorithm. Among the different algorithms tested, Gradient Boost worked best for dataset 8.1. However, when we applied the ML approach to the second dataset, which contained zebrafish data, we did not observe any significant improvements (**Table 4.23**). The results were similar to the PLS model with 3 LVs. For Dataset 8.1, with EC₅₀ as the endpoint, the ML models outperformed the other modeling algorithms. The models differed in the MAE values, with Gradient Boost showing a lower value than the other models.

Table 4.23. Statistical results of the developed models with various ML algorithms for dataset 8.1 and dataset 8.2.

Methods	Training set				Test set			
<i>Daphnia magna</i>								
	R ²	Q ² _{loo}	MAE _{train}	RMSE _c	Q ² _{F1}	Q ² _{F2}	MAE _{test}	RMSE _P
RF	0.91	0.56	0.29	0.40	0.76	0.63	0.40	0.52
Adaboost	0.91	0.64	0.29	0.39	0.57	0.31	0.50	0.71
Gradient Boost	0.95	0.57	0.12	0.27	0.78	0.65	0.37	0.51
XGBoost	0.79	0.48	0.43	0.62	0.58	0.33	0.44	0.70
SVM	0.33	-0.14	0.86	1.12	0.15	-0.35	0.83	1.01
Ridge regression	0.69	0.53	0.77	0.76	0.75	0.61	0.34	0.53
<i>Zebrafish (Danio rerio)</i>								
PLS (3 LV)	0.72	0.51	0.06	0.09	0.83	0.83	0.039	0.05
XGBoost	0.72	0.52	0.06	0.09	0.83	0.83	0.039	0.05

The best model was predicted based on the lowest MAE by running the algorithms in basic settings. Supervised nonlinear learning algorithms have also made a significant progress in predicting toxicity compared to linear regression. The boosting algorithm converts weaker learners into stronger learners, and this ensemble learning technique reduces both bias and

variance (Mienye and Sun, 2022) **Figure 4.43** represents SHAP summary plot of model that demonstrates the influence of each descriptor towards the model performance.

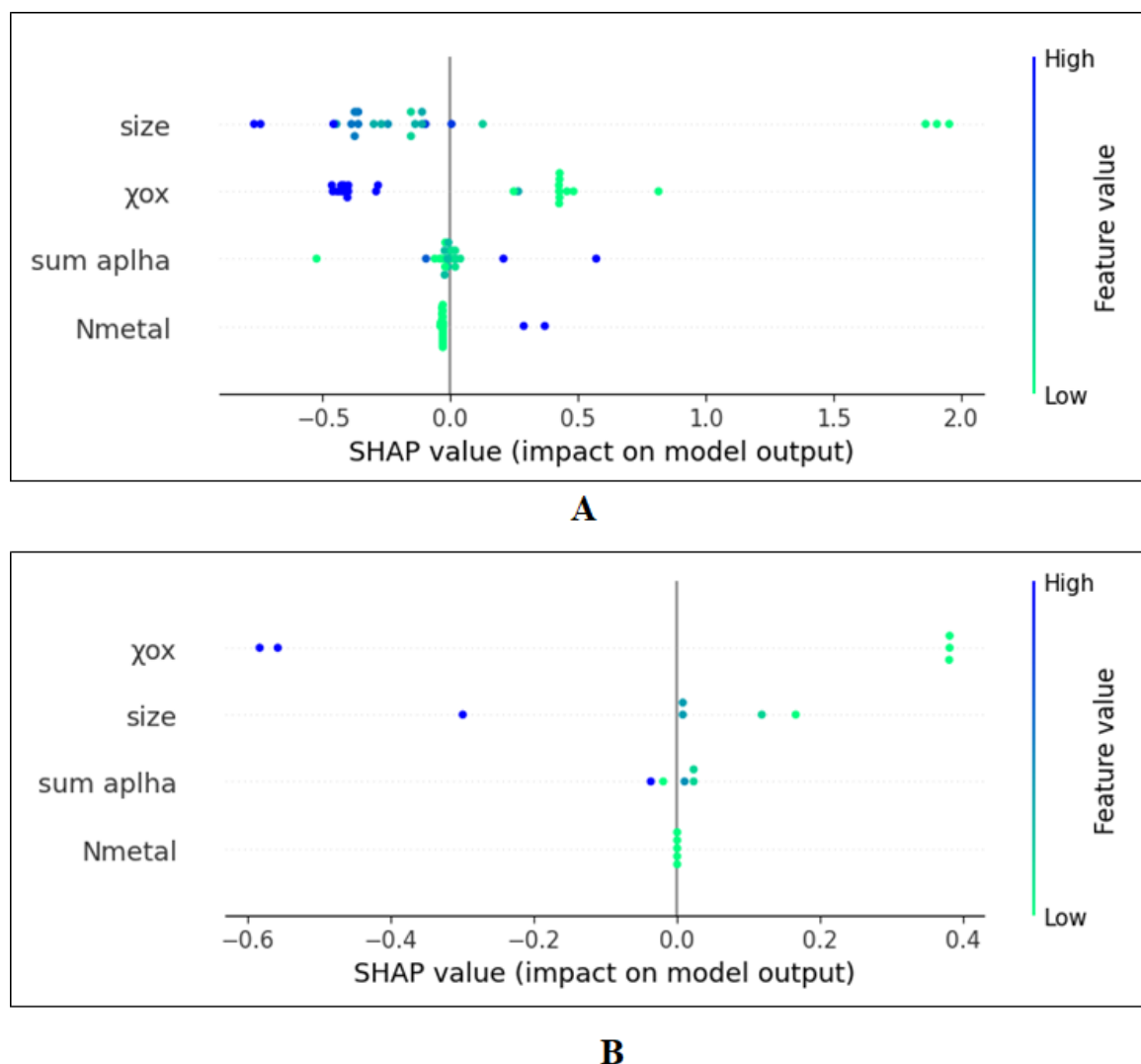
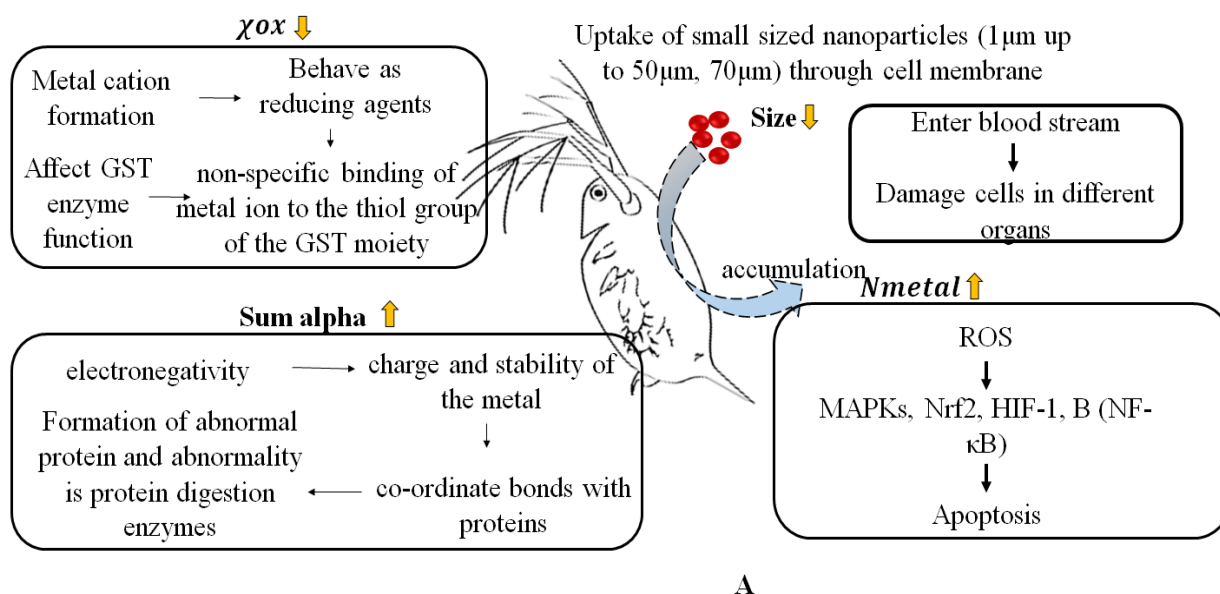


Figure 4.43 SHAP summary plots of model 8.1 (*D. magna*) for (A) the training set and (B) the test set.

4.8.5. The NP effects on morphological parameters of *D. magna* and hatching of *zebra fish*

The fifth principles of the OECD (The Organization for Economic Cooperation and Development) require a model with a mechanistic interpretation wherever possible. Such insights help to assign mechanisms towards toxicity. While machine learning (ML) descriptors are difficult to interpret, periodic table descriptors were able to interpret the reason for toxicity towards both *zebrafish* and *Daphnia magna* species. Although the physicochemical properties of ENPs are crucial for modulating the toxic effect of NPs, other physical and chemical properties of NPs also play an important role in determining the toxicity. The surface area of NPs is essential for their behavior; as high surface reactivity can result in severe toxicity on biota (Kovalishyn et al., 2018). The descriptors related to electrotopological indices and solubility are also related to toxicity through intermolecular interactions (Li et al., 2018; Roy

and Roy, 2023). Size has often been described as an important characteristic affecting NP toxicity. Intrinsic properties such as solubility and the ability of NP to aggregate were also considered to be important features for toxicity prediction. AgNPs (having antimicrobial properties) cause toxicity to aquatic organisms through oxidative stress and DNA damage (Tortella et al., 2020). The photocatalytic activity of TiO_2 may degrade other contaminants in the aquatic environment and favor the formation of toxic by-products. When NPs enter the aquatic environment, they inevitably interact with the aquatic organisms through electrostatic interactions. The presence of NPs led to an increase in bioaccumulation and toxicity in daphnia. Moreover, the Trojan horse effect could be influenced by exposure conditions. Reactive oxygen species (ROS) toxicity is common in nanoparticles towards daphnia. As important signaling molecules, ROS can affect other pathways, for instance, MAPKs (mitogen-activated protein kinases), Nrf2, HIF-1 (hypoxia-inducible factor 1), and nuclear factor-kappa B (NF- κ B), which can result in severe apoptosis, DNA damage, and genotoxicity. Similarly, for zebrafish toxicity, molecular weight and the presence of oxygen play an important role in enzyme inhibition. Molecules with high molecular weight tend to have poor solubility. Non-soluble particles could inactivate enzymes by blocking the outer surface, or in other words, the NPs adsorb the target enzyme. The core environment of the NPs also determines its binding with respective metal-sensitive active binding sites of the zebrafish enzyme (Lin et al., 2013). The core of the metal (charge and electronegativity) may facilitate the properties of metal ion release and interaction with the metal-binding enzyme center that leads to its inhibition. The surface property reflects the surface molecules and could be responsible for toxicity or inhibitory activity (Lee et al., 2007). **Figure 4.44** is the illustration of the toxicity mechanism.



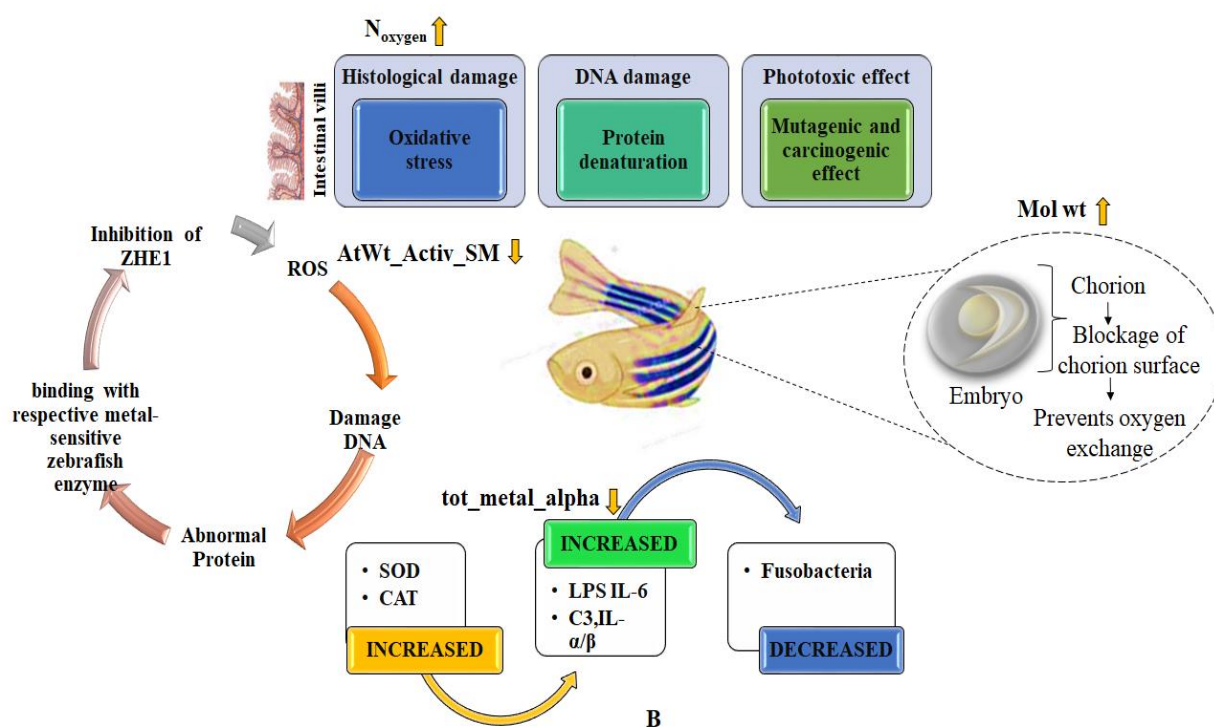


Figure 4.44 Illustration of the toxicity mechanism of *Daphnia magna* (A) and zebrafish (B) with simplified periodic table-based (PT) descriptors (study 8).

Note: GST- Glutathione-S-transferase, MAPKs (mitogen-activated protein kinases), Nrf2, HIF-1 (hypoxia-inducible factor 1), and nuclear factor-kappa B (NF- κ B), SOD-superoxide dismutase, CAT- catalase, LPS-lipopolysaccharide, IL- interleukin 6.

4.8.6. Comparison of ML models with QSAR models and the previous work

The development of ML models has improved the prediction accuracy of QSAR models for dataset 8.1, but there were hardly any changes observed for dataset 8.2. The developed models were compared with the predictive models developed by other authors too. In 2017, Sizochenko et al. (2017) used the non-linear regression tree M5P algorithm to predict the enzyme inactivation activity of 24 metal oxide NPs. Although the models showed higher values of prediction quality, they were not exactly comparable as the dataset was divided into training set, validation set, and external set. In contrast, the present work was performed by dividing the data set into training and test sets. The present work calculated simple periodic table descriptors that are easy to compute for new NPs without requiring an experimental setup, whereas the previous work used hydrodynamic radius, mass density, the Wigner–Seitz radius, and the covalent index. Moreover, a similarity-based approach (RA) was performed that showed an improvement in external prediction from 0.83 to 0.95, while minimizing the MAE value (0.028) represented in **Table 4.24**.

Table 4.24. Comparison between the present work with Sizochenko et al. (2017)

Model 8.1	Training set					Test set				Validation set		
	R^2	Q^2_{loo}	$\text{MAE}_{\text{train}}$	R^2_{adj}	RM SE_c	Q^2_{F1}	Q^2_{F2}	MAE_{test}	R^2_{adj}	RM SE_p	R^2_{adj}	RMS E_v
Sizochenko et al. (2017)												
Zebra fish (M5P algorithm)	-	-	-	0.85	0.06	-	-	-	0.80	0.11	0.7	0.18
Present work (model 8.2)												
PLS	0.72	0.52	0.06	0.64	0.09	0.83	0.83	0.03	0.93	0.16	-	-
RA	-	-	-	-	-	0.95	0.95	0.03	0.94	0.02	-	-
ML	0.724	0.515	0.06	0.64	0.09	0.83	0.83	0.04	0.93	0.05	-	-

The dataset 8.1 was not compared with any other studies since there was no previously reported QSAR modeling from the particular database i.e., the NanoToxdb database. Therefore, no such comparison was possible here.

4.8.7. Future research and considerations

Nanoparticles (NPs) are known to have an impact on the uptake, distribution, metabolism, and excretion of organisms, which can result in toxic effects. When NPs are present in aquatic environments, they often get exposed to other contaminants, which can further impact their toxicity. Therefore, it is important to understand the effects of NPs in aquatic environments and conduct additional research to quantify the toxicological effects on aquatic organisms. While there has been extensive research on the acute effects of NPs on organisms such as *Daphnia magna* and *D. rerio*, more research is needed on the long-term effects of low levels of NPs. A major gap in the research is the examination of the effects of NPs on different aquatic communities. The mechanism of NP toxicity may involve various processes mediated through the production of reactive oxygen species (ROS). However, ROS-mediated stress is not unique to NPs and may also be caused by other stressors in the environment. To further predict the potential toxicity of NPs, it is recommended to develop new *in-silico* models.

CHAPTER - 5

CONCLUSIONS

Chapter 5: Conclusion

Innovative scientific solutions play a key role in meeting the needs of new applications. The success of any research depends on the conclusions drawn from the results obtained through careful analysis. This analysis should be designed to address the given problem and explore both revealed and unrevealed scientific explanations to provide a comprehensive understanding of the issue at hand. To ensure that the research is both rigorous and accurate, it is important to use appropriate methodologies and techniques that are suited to the specific research question. This may involve using quantitative or qualitative methods, depending on the nature of the problem and the data available. It is also important to use reliable and valid measures, instruments, and tools to collect and analyze data. Furthermore, research should be conducted ethically and responsibly, taking into account the potential risks and benefits of the study for participants and society as a whole. This means ensuring the study is conducted in a way that minimizes harm and maximizes benefits. With access to new knowledge, we unlock the doors to deeper understanding and limitless possibilities within specific research domains.

In the present thesis, the development of *in silico* studies has been discussed, with special emphasis on nanoparticle toxicity prediction against several ecologically relevant endpoints. The toxicity of nanoparticles has been predicted in various fields, such as nanomedicines and nanopesticides. Then from terrestrial organisms to aquatic organisms. To ensure the safety and high performance of nanoparticles, nanotechnology has been utilized to focus on appropriate design and the use of relevant nanodescriptors that provide valuable nanotoxicological information. Predictive QSAR models, sometimes supplemented by ML algorithms, have been employed with the objective to aid in the creation of safe and effective nanoparticles. Read-across prediction methods have also been employed for the same purpose. Encoding nanostructure information in mathematical features or descriptors presents a challenge that has been addressed. This thesis delves into the efficacy of periodic table-based descriptors in nanotoxicology modeling, alongside other available nanodescriptors. It provides a comprehensive understanding of the fundamentals and mathematical derivations of second-generation periodic table descriptors, which possess clear physical and chemical meanings, rendering them easier to apply. These descriptors aid in understanding mechanisms and are computationally less demanding and more reproducible than other types of descriptors, such as QC (Quantum chemicals), SMILES, experimental, and MD (Molecular dynamics) simulation. The nano-QSPR/QSAR/QSTR models have been developed obeying the five OECD principles to avoid any ambiguity in the results obtained. The important structural features (specific to each dataset) have been identified that are responsible for nanoparticle toxicity. All the models have been strictly validated using conventional internal and external validation tests. To check the reliability of external validation, the domains of applicability of the developed models have been examined, which also provides knowledge about the chemical space of the respective models. An effort of data gap filling has been made in this thesis, where the toxicity of different hypothetical and/or real, untested nanoparticles has been predicted against corresponding ecological species.

The research conducted in this thesis not only aimed to predict toxicity against different organisms but also focused on improving the properties of nanoparticles to derive particles with better properties and less toxic effects. The mechanism of the effects on various cell lines was explained well, using the information obtained from the developed regression and supervised algorithms. The results have been fruitful and significant which lead to the conclusion that QSAR modeling can be applied for the property and toxicity determination of nanoparticles and the knowledge derived from the modeled descriptors can also be used to

design and synthesize new compounds with desired properties of interest. Finally, the individual studies conducted in this thesis work provided precise information and knowledge gained, which are described below:

5.1. Study 1: Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach.

Recent advances in NP engineering are believed to have a potential impact in diverse areas in the near future. Production of metal-based NPs for commercial use and other purposes has increased exponentially, and as a result, the assessment of their safety has also gained attention. Therefore, in this work, we have made efforts to develop interpretative nano-QSAR models that are reliable and capable of predicting not only the investigated NPs but also unexplored NPs if present within the applicability domain. As we know, working with small datasets is a critical job, here we have applied the Small Dataset Modeler version 1.0.0 to develop a PLS (N1.1) model without applying the conventional splitting approach. Further, MLR (M1.1-1.4) models were also generated through best subset selection, and then ICP was applied to obtain reliable predictions. After performing ICP, CM3 was found to be the winner model. The insights obtained from the various descriptors from both PLS and MLR models pointed out that electronegativity, solubility, formation of a metal cation, and an increase in the number of metal NPs increase the cytotoxicity whereas non-metal NPs and an increase of the oxidation degree contribute to low toxicity to RAW 267.4 cells. A short-proposed mechanism has also been explained to determine the cause of cellular toxicity by the MeOx NPs. We hope that the present work will be useful in establishing the cause of toxicity of unexplored NPs within the applicability domain, and the information obtained about the features of MeOx NPs would help in designing new NP entities with reduced toxicity to living cells. Considering the limited experimental data available for the toxicity of metal oxide and hydroxide nanoparticles, the developed models might be used for the prediction of toxicity of untested nanoparticles that may help in prioritizing further experiments.

5.2. Study 2: Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides

In the study, we have developed 2D MLR QSAR models to investigate the essential features in 17 MeOxNPs responsible for the cytotoxicity of *E. coli* cells with $-\log LC_{50}$ as the endpoint. This predominant nano-structure information has helped in identifying the cause of toxicity of MeOxNPs and has also suggested some mechanisms through which it can affect *E. coli* cells or other biological cells. The statistical results of the developed models are robust and can be used to extrapolate the endpoint if the experimental information of new MeOxNPs is unavailable. The periodic table descriptors used in this work can be easily obtained either using the Elemental Descriptor Calculator software (available from <https://dtclab.webs.com/software-tools>) or from the Periodic Table information. Now, knowing the regression coefficients of different descriptors in the reported model, one can easily compute the toxicity in an Excel sheet without requiring the use of any codes. The major findings of the present study are summarized below:

1. Only periodic table descriptors were used which can be calculated easily without any computational calculations. The reproducible descriptors can easily encode the cytotoxicity with good interpretability.
2. The application of the 'Small Dataset Modeler' software was emphasized in the current work. Division of the small dataset (17 compounds) into training and test sets is not desirable

here to generate and validate QSAR models to avoid the loss of information and introduce bias in descriptor selection.

3. The developed MLR models were conceptually simple, easy to apply, computationally inexpensive as well as can be used to reduce the data gap in untested NPs.

4. Here, we interpreted that an increase in valency and oxidation state of the metal in MeOxNPs will decrease the cytotoxicity whereas the atomic radius of the metal and electronegativity of MeOxNPs has a substantial influence on the toxicity toward *E. coli* cells.

5. This study can be used for predicting the cytotoxicity of untested MeOxNPs for which the experimental data are not known. The developed models can be used in designing eco-safe NPs with better properties as evidenced by the predictions of 42 MeOxNPs, out of which, 38 MeOxNPs showed good predictions with the dark condition toxicity data model and 35 MeOxNPs showed good predictions with the photo-induced toxicity data model. Therefore, these models could be used to predict the cytotoxicity of the untested NPs before their utilization in commercial products.

5.3. Study 3: Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOxNPs) used in nanomedicines: A mechanistic QSAR approach.

The ever-increasing use of NMs in the form of nanoparticles (NPs) in biomedical applications and consumer products has raised concerns about their safety for humans and the environment. Nanosafety has become an important part of nano education for future researchers and a highly demanding research area in recent times. Their potential to cause undesirable health effects to the environment and susceptible parts of the human population could be seen as an alarm and a call for preventive measures. The concept of nanotoxicology in medical sciences is needed for the safety evaluation of nanomedicines for their nanotechnology-based therapies and diagnostics to identify the potential hazards it may cause after application. Nanoparticle potential toxicity, as well as human toxicity, has prompted major scientific organizations to assess the risk that may be associated with new nanoparticles. Therefore, the present work was performed to identify the main features governing the cytotoxicity in human keratinous cells (*HaCaT*). Combining calculated periodic table descriptors with available experimental data allowed us to reveal and explain the features responsible for toxicity to the corresponding cell line through the development of a predictive and validated QSTR model. This work demonstrated the toxicity profile of *HaCaT cells* derived from two datasets. The insights obtained from the two datasets pointed out that periodic table descriptors can also well predict the relevant features responsible for the toxicity to the respective cell. The key factors influencing the toxicity of NPs were determined from the developed models (MLR and PLS). From the MLR model (Model 3.1), atomic radius and thermal conductivity were found to increase toxicity through metal cation formation. On the other hand, a decrease in melting point and increase of electronegativity of the atom were responsible for cell death through ROS production, and damaging DNA, mitochondria, and Lysosomes. The same observation was noticed in the case of the PLS model (Model 3.2) where the heat of fusion, an increase of cationic charge, and the dose have a substantial effect on the cytotoxicity. From the developed QSAR models, it was stated that the major factors for the toxicity were the formation of cations, an increase in solubility, and the dose of the NPs. This feature was responsible for controlling the physicochemical properties of metallic NPs to achieve safe and reliable MeOx NPs. This work opens a new horizon for awareness and knowledge for understanding the NPs' toxicity. The nanotoxicology information can help design modern material within a safe threshold with optimal properties and minimal potential toxicity.

5.4. Study 4: Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells.

Advancements in nanotechnology have paved innovations in the agro-food sector. Research work has focused on developing a nanoscale formulation of pesticides to enhance the efficacy and production of food. Nanotechnology has provided us with enormous opportunities while imposing a potential threat to human beings and living organisms. The small size particles have raised a question about their safety and their sustainability in the environment. It has been seen that the small-sized particulates with a large surface area led to unintended ecotoxicological effects on non-target organisms when applied to the agricultural field. Therefore, understanding the benefits and potential risks of MeOx ENPs used in food as pesticides is urgent and critical. Thus, in the present work, we have tried to identify the main features governing the toxicity of BEAS-2B and RAW 264.7 cell lines derived from four datasets. Periodic table-based descriptors were used to reveal the main features responsible for toxicity to the corresponding cells through the development of predictive QSTR models (M 4.1-M 4.4). Considering the knowledge and current ecotoxicological status of MeOx ENPs, PLS regression and the read-across approaches were used. Furthermore, the interspecies toxicity correlation study was performed between BEAS-2B and RAW 264.7 cells to fill the data gap that needs to be addressed for safer and sustainable use of pesticides. Particular emphasis was given to determining the fate and toxicological behavior of nano pesticides in the environmental compartments. The studies examined the role of metal cation formation, electro-positivity, binding capacity, and bioaccumulation of nanoparticles in contributing to toxicity. The periodic table descriptors that can be easily computed can successfully explain the cause of toxicity. The information derived from the PLS models of both the cell lines suggested that metal cation formation and bond-forming capacity influence the toxicity, whereas the presence of metal has an influential impact on the ecotoxicological effects. All the descriptors in the present study were found to have a negative impact on the toxicity except the number of metals which has a positive impact. It is also worth mentioning here that the toxicity depends on the organism type. Therefore, the knowledge derived from the work will not only help in data gap filling but also help in substantial enhancement in efficacy and stability of formulation of nano pesticides

5.5. Study 5: Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: A nano-QSAR approach.

Safe implementation of MeOx NPs requires an intimate knowledge of biological and nanomaterial interactions. Only a few nanotoxicological studies have investigated pro-inflammatory effects. The pro-inflammatory cytokine release of THP-1 cells can be a good indicator for MeOx NPs. The relationship of metal with the inflammation potential has been focused on in the present study. Models for fold change ($FC_{IL-1\beta}$) and cell viability [$-\ln(p/(1-p))$] using QSAR MLR (Model 5.1) and PLS (Model 5.2) methods respectively were developed that included the dose of MeOx NPs as a factor. The developed models were predicted and validated rigorously following the OECD (Organization for Economic Co-operation and Development) guidelines. The descriptors obtained from the developed models suggested that the number of metal ions or an increased number of metals influences inflammatory potential. Conversely, low electronegativity was found to be more detrimental than high electronegativity. We found that dose, low electronegativity, and the presence of metal ions and oxygen are connected to IL-1 β leakage from the cell. The metal ions can trigger inflammasome that leads to a cascade of reactions and activates IL-1 β productions. We suggest that MeOx NPs with lower electronegativity ($\chi_{me} \leq 1.55$) have more inflammatory potential. Summarizing our results, one can conclude that periodic table-based descriptors are efficient for modeling proinflammatory potential giving rise to statistically sound QSAR models. Our

results are in agreement with the literature sources. These comprehensive results and information can be helpful for further research in designing NPs with lower toxicity and deriving mechanistic information on the cause of proinflammatory conditions due to MeOx NP exposure.

5.6. Study 6: Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches.

We have performed the cytotoxicity modeling of eight heavy metal compounds adsorbed with nano TiO₂ on HK-2 cells and explored the features responsible for the toxicity mechanism. Many studies have shown co-exposure of metal, metalloid mixtures with heavy metals. The co-exposure may also be affected by the variation in dose at a biomarker level. Also, human co-exposure to heavy metals was found to have more profound renal damage than exposure to each of the elements alone. Hence, to elucidate the features responsible for such toxicity, in the present study, ML algorithms were applied along with periodic table-based descriptors for QSAR modeling. Experiment-independent periodic table-based descriptors were successful in producing better results than previous studies with Quantum chemical-based descriptors. PT descriptors, used in QSAR models, have strong theoretical guidance, which can help scientists design new entities with expected properties. As a part of the model development process, PT descriptors can be used in conjunction with other descriptors that are compatible with them. The application of periodic table-based descriptors is not only less computationally demanding but also size-independent. The ML algorithm along with periodic table descriptors has helped to evaluate the cell survival rate of HK-2 cells in less time and cost than expensive quantum chemical descriptors and experimental descriptors. Among all the algorithms, the random forest (RF) model shows the best prediction ability with $Q^2F_1 = 0.94$ and $MAE_{test} = 0.14$ for the test set. Hence, a good feature selection method reduces the computation time required to train a model. The SHAP analysis also emphasized the most significant features contributing to the model's performance. Here, we also have proposed a generalized mechanism for the most impactful features generated by the model. As a result, periodic table descriptors and machine learning can be used together to decipher features of unknown compounds and predict similar compounds.

5.7. Study 7: Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors.

The impact of nanoparticles on cell membranes has been thoroughly examined through nanotoxicological research and in-vitro modeling (Pathakoti et al., 2014; Qi et al., 2020). While toxicity endpoint is a well-explored topic, it is crucial to investigate non-fatal endpoints such as cell damage. The ζ -potential is a widely used parameter to characterize the properties of nanoparticles. However, cell membrane damage is influenced by various factors, including exposure time and dose. Thus, this study aimed to establish a relationship between the properties of nanoparticles and their toxicity, with a focus on cell membrane damage.

The study was divided into two parts. The first part involved modeling nanoparticles against ζ -potential to determine the features that can alter their properties. The second part focused on the causative elements that can influence toxicity and damage to the cell membrane. Both the QSPR model (Model 7.1) for ζ -potential and another model against cell damage (Model 7.2) were developed using periodic table-based descriptors. The QSPR model (Zeta potential) was able to predict ζ -potential for MeOx NPs without experimental values. The developed models showed good predictivity and robustness, confirming their effectiveness.

The features obtained from the models suggested that surface charge and electronegativity play a role in altering ζ -potential. Additionally, an increase in oxygen count, electronegativity, formation of cationic charge, and an increase in ζ -potential can influence membrane damage. Based on these findings, the authors propose that the damaging power of metal oxide nanoparticles can be harnessed in treating cancerous cells. This study not only identifies the features required to enhance the properties of nanoparticles but also provides knowledge for treating cancerous cells through cell damage techniques. The study can pave the way for researchers to use nanoparticles in clinical practices with confidence.

5.8. Study 8: Insights into nanoparticle toxicity against aquatic organisms using regression, Read-Across, and ML algorithms: QSAR models for *Daphnia magna* and zebrafish.

An understanding of the factors that are critical in determining the activity of nanoparticles is essential in establishing a relationship to the mechanism of toxicity. In this study, we investigated the effectiveness of *in silico* modeling to predict the nanotoxicity of metal oxides on *Daphnia magna* and zebrafish. We collected acute toxicity data from databases and literature reviews and used PLS modeling followed by read-across predictions and supervised machine learning (ML) on the respective datasets. The computational modeling was validated internally and externally. We found that several environmental parameters should be considered while addressing the toxicity of nanoparticles in the aquatic environment. The grouping hypothesis improved the external prediction for both datasets. Moreover, a nonlinear regression algorithm was believed to predict the toxicity of dataset 8.1 more confidently than dataset 8.2, in the latter case nonlinear models produced similar results as the regression-based linear QSAR model. The model for *Daphnia magna* hypothesized that nanoparticles' size, the number of metals, the core environment of the metal present in the metal oxide, and the oxidation number of the metal play a key role in the final expression of toxicity of nanoparticles to *Daphnia*. Similarly, oxidative stress is a major mechanism of toxicity for other models as well.

The nano-QSAR model for zebrafish reliably predicted the activity of nanoparticles towards zebrafish hatching enzyme. The interpretation of the descriptors from the developed model is related to the possible mechanism of nanoparticles. The presence of higher molecular weight, the core of the metal, and the presence of oxygen influence the enzyme inhibition activity. The enzyme inhibition is correlated with the ability of zebrafish embryos to hatch, and therefore, the inhibition of ZHE1 seems to be the factor driving hatch delay. Given that nanoparticles can inhibit several enzymes, other aquatic organisms depending on the hatching enzymes could be similarly affected when exposed to nanoparticles. The lack of hatching is sometimes overlooked as an endpoint for toxicology studies, but the results clearly show that some nanoparticles can cause death due to an inability to hatch. More research should be conducted to understand the interference of nanoparticles with hatching inhibition. Therefore, this study evaluates the predictive power of the proposed models, provides insights into the mechanism of toxicity towards aquatic organisms, and provides information to develop nanoparticles with less toxicity.

REFERENCES

References

- Abbas, Q., Yousaf, B., Ullah, H., Ali, M.U., Ok, Y.S. and Rinklebe, J., 2020. Environmental transformation and nano-toxicity of engineered nano-particles (ENPs) in aquatic and terrestrial organisms. *Critical Reviews in Environmental Science and Technology*, 50(23), pp.2523-2581.
- Abe, K. and Saito, H., 1999. Cholesterol does not affect the toxicity of amyloid β fragment but mimics its effect on MTT formazan exocytosis in cultured rat hippocampal neurons. *Neuroscience Research*, 35(3), pp.165-174.
- Ahamed, M., Karns, M., Goodson, M., Rowe, J., Hussain, S.M., Schlager, J.J. and Hong, Y., 2008. DNA damage response to different surface chemistry of silver nanoparticles in mammalian cells. *Toxicology and applied pharmacology*, 233(3), pp.404-410.
- Ahamed, M., Posgai, R., Gorey, T.J., Nielsen, M., Hussain, S.M. and Rowe, J.J., 2010. Silver nanoparticles induced heat shock protein 70, oxidative stress and apoptosis in *Drosophila melanogaster*. *Toxicology and applied pharmacology*, 242(3), pp.263-269.
- Ahamed, N., 2014. Ecotoxicity concert of nano zero-valent iron particles-a review. *J Crit Rev*, 1(1), pp.36-39.
- Almaliki, Z.A., 2018. Bias-variance dilemma?. Towards Data Science (<https://towardsdatascience.com/bias-variance>).
- Ambure, P., Gajewicz, A., Natália, M., Cordeiro, D.S. and Roy, K., 2019. A new workflow for qsar model development from small data sets: integration of data curation, exhaustive double cross-validation and a set of optimal model selection techniques. *J. Chem. Inf. Model*, 59(10), pp.4070-4076.
- Artiles, M.S., Rout, C.S. and Fisher, T.S., 2011. Graphene-based hybrid materials and devices for biosensing. *Advanced drug delivery reviews*, 63(14-15), pp.1352-1360.
- Auwerx, J., 1991. The human leukemia cell line, THP-1: a multifaceted model for the study of monocyte-macrophage differentiation. *Experientia*, 47(1), pp.22-31.
- Baek, S., Singh, R.K., Khanal, D., Patel, K.D., Lee, E.J., Leong, K.W., Chrzanowski, W. and Kim, H.W., 2015. Smart multifunctional drug delivery towards anticancer therapy harmonized in mesoporous nanoparticles. *Nanoscale*, 7(34), pp.14191-14216.
- Banerjee, A. and Roy, K., 2022. First report of q-RASAR modeling toward an approach of easy interpretability and efficient transferability. *Molecular Diversity*, 26(5), pp.2847-2862.
- Banerjee, A., Chatterjee, M., De, P. and Roy, K., 2022. Quantitative predictions from chemical read-across and their confidence measures. *Chemometrics and Intelligent Laboratory Systems*, 227, p.104613.
- Bar-Ilan, O., Louis, K.M., Yang, S.P., Pedersen, J.A., Hamers, R.J., Peterson, R.E. and Heideman, W., 2012. Titanium dioxide nanoparticles produce phototoxicity in the developing zebrafish. *Nanotoxicology*, 6(6), pp.670-679.

- Basant, N. and Gupta, S., 2017. Modeling uptake of nanoparticles in multiple human cells using structure–activity relationships and intercellular uptake correlations. *Nanotoxicology*, 11(1), pp.20-30.
- Basant, N. and Gupta, S., 2017. Multi-target QSTR modeling for simultaneous prediction of multiple toxicity endpoints of nano-metal oxides. *Nanotoxicology*, 11(3), pp.339-350.
- Bayda, S., Adeel, M., Tuccinardi, T., Cordani, M. and Rizzolio, F., 2019. The history of nanoscience and nanotechnology: from chemical–physical applications to nanomedicine. *Molecules*, 25(1), p.112.
- Behzadi, S., Serpooshan, V., Tao, W., Hamaly, M.A., Alkawareek, M.Y., Dreaden, E.C., Brown, D., Alkilany, A.M., Farokhzad, O.C. and Mahmoudi, M., 2017. Cellular uptake of nanoparticles: journey inside the cell. *Chemical society reviews*, 46(14), pp.4218-4244.
- Blekos, K., Chairetakis, K., Lynch, I. and Marcoulaki, E., 2023. Principles and requirements for nanomaterial representations to facilitate machine processing and cooperation with nanoinformatics tools. *Journal of Cheminformatics*, 15(1), p.44.
- Bonaventura, P., Lamboux, A., Albaredo, F. and Miossec, P., 2018. Differential effects of TNF- α and IL-1 β on the control of metal metabolism and cadmium-induced cell death in chronic inflammation. *PLoS One*, 13(5), p.e0196285.
- Bossy-Wetzel, E., Schwarzenbacher, R. and Lipton, S.A., 2004. Molecular pathways to neurodegeneration. *Nature medicine*, 10(Suppl 7), pp.S2-S9.
- Breiman, L., 2001. Random forests. *Machine learning*, 45, pp.5-32.
- Briffa, J., Sinagra, E. and Blundell, R., 2020. Heavy metal pollution in the environment and their toxicological effects on humans. *Heliyon*, 6(9).
- Brun, N.R., Beenakker, M.M., Hunting, E.R., Ebert, D. and Vijver, M.G., 2017. Brood pouch-mediated polystyrene nanoparticle uptake during *Daphnia magna* embryogenesis. *Nanotoxicology*, 11(8), pp.1059-1069.
- Büdinger, L., Hertl, M. and Büdinger, L., 2000. Immunologic mechanisms in hypersensitivity reactions to metal ions: an overview. *Allergy*, 55(2), pp.108-115.
- Burello, E. and Worth, A.P., 2011. QSAR modeling of nanomaterials. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 3(3), pp.298-306.
- Cao, J., Pan, Y., Jiang, Y., Qi, R., Yuan, B., Jia, Z., Jiang, J. and Wang, Q., 2020. Computer-aided nanotoxicology: risk assessment of metal oxide nanoparticles via nano-QSAR. *Green chemistry*, 22(11), pp.3512-3521.
- Cao, W., Jin, M., Yang, K., Chen, B., Xiong, M., Li, X. and Cao, G., 2021. Fenton/Fenton-like metal-based nanomaterials combine with oxidase for synergistic tumor therapy. *Journal of Nanobiotechnology*, 19, pp.1-35.
- Chatterjee, M., Banerjee, A., De, P., Gajewicz-Skretna, A. and Roy, K., 2022. A novel quantitative read-across tool designed purposefully to fill the existing gaps in nanosafety data. *Environmental Science: Nano*, 9(1), pp.189-203.

- Chen, E.Y., Garnica, M., Wang, Y.C., Mintz, A.J., Chen, C.S. and Chin, W.C., 2012. A mixture of anatase and rutile TiO₂ nanoparticles induces histamine secretion in mast cells. *Particle and fibre toxicology*, 9, pp.1-10.
- Chen, Z.Y., Li, N.J., Cheng, F.Y., Hsueh, J.F., Huang, C.C., Lu, F.I., Fu, T.F., Yan, S.J., Lee, Y.H. and Wang, Y.J., 2020. The effect of the chorion on size-dependent acute toxicity and underlying mechanisms of amine-modified silver nanoparticles in zebrafish embryos. *International journal of molecular sciences*, 21(8), p.2864.
- Chirico, N. and Gramatica, P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of chemical information and modeling*, 51(9), pp.2320-2335.
- Choi, J.S., Trinh, T.X., Yoon, T.H., Kim, J. and Byun, H.G., 2019. Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere*, 217, pp.243-249.
- Criminisi, A., Shotton, J. and Konukoglu, E., 2012. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and trends® in computer graphics and vision*, 7(2–3), pp.81-227.
- Dale, A.L., Lowry, G.V. and Casman, E.A., 2015. Much ado about α : reframing the debate over appropriate fate descriptors in nanoparticle environmental risk modeling. *Environmental Science: Nano*, 2(1), pp.27-32.
- Darlington, R.B., 1990. *Regression and Linear Models* McGrawHill. New York.
- De, P. and Roy, K., 2021. QSAR and QSAAR modeling of nitroimidazole sulfonamide radiosensitizers: application of small dataset modeling. *Structural Chemistry*, 32, pp.631-642.
- De, P., Kar, S., Roy, K. and Leszczynski, J., 2018. Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modeling for exploration of toxicity mechanisms. *Environmental Science: Nano*, 5(11), pp.2742-2760.
- Devillers, J., 1996. *Genetic algorithms in molecular modeling*. Academic Press.
- Djurišić, A.B., Leung, Y.H., Ng, A.M., Xu, X.Y., Lee, P.K., Degger, N. and Wu, R.S.S., 2015. Toxicity of metal oxide nanoparticles: mechanisms, characterization, and avoiding experimental artefacts. *Small*, 11(1), pp.26-44.
- Dobson, C.M., 2004. Experimental investigation of protein folding and misfolding. *Methods*, 34(1), pp.4-14.
- Domingues, C., Santos, A., Alvarez-Lorenzo, C., Concheiro, A., Jarak, I., Veiga, F., Barbosa, I., Dourado, M. and Figueiras, A., 2022. Where is nano today and where is it headed? A review of nanomedicine and the dilemma of nanotoxicology. *ACS nano*, 16(7), pp.9994-10041.
- Dominguez, G.A., Lohse, S.E., Torelli, M.D., Murphy, C.J., Hamers, R.J., Orr, G. and Klaper, R.D., 2015. Effects of charge and surface ligand properties of nanoparticles on oxidative

- stress and gene expression within the gut of *Daphnia magna*. *Aquatic toxicology*, 162, pp.1 - 9.
- ECHA (European Chemicals Agency), 2017. Guidance on Information Requirements and Chemical Safety Assessment: Appendix R7-1 for Nanomaterials Applicable to Chapter R7a Endpoint Specific Guidance. Version 2.0.
- El Naqa, I. and Murphy, M.J., 2015. What is machine learning? (pp. 3-11). Springer International Publishing, Cham, Switzerland, pp 3–11.
- Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T., McDowell, R.M. and Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives*, 111(10), pp.1361-1375.
- Feurer, M. and Hutter, F., 2019. Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, pp.3-33.
- Feynman R. P., Plenty of Room at the Bottom, in APS annual meeting, 1959, <http://www.its.caltech.edu/~feynman/plenty.html>, (accessed on 15 April 2022).
- Fischer, J., Gräf, T., Sakka, Y., Tessarek, C. and Köser, J., 2021. Ion compositions in artificial media control the impact of humic acid on colloidal behaviour, dissolution and speciation of CuO-NP. *Science of the Total Environment*, 785, p.147241.
- Fjodorova, N., Novic, M., Gajewicz, A. and Rasulev, B., 2017. The way to cover prediction for cytotoxicity for all existing nano-sized metal oxides by using neural network method. *Nanotoxicology*, 11(4), pp.475-483.
- Fojtová, D., Vašíčková, J., Grillo, R., Bílková, Z., Šimek, Z., Neuwirthová, N., Kah, M. and Hofman, J., 2019. Nanoformulations can significantly affect pesticide degradation and uptake by earthworms and plants. *Environmental Chemistry*, 16(6), pp.470-481.
- Forest, V., Hocheplied, J.F., Leclerc, L., Trouvé, A., Abdelkebir, K., Sarry, G., Augusto, V. and Pourchez, J., 2019. Towards an alternative to nano-QSAR for nanoparticle toxicity ranking in case of small datasets. *Journal of Nanoparticle Research*, 21, pp.1-14.
- Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S.Y., Mumper, R.J. and Tropsha, A., 2010. Quantitative nanostructure– activity relationship modeling. *ACS nano*, 4(10), pp.5703-5712.
- Foussereau, J., 1966. Allergic eczemas from metallic foreign bodies. *Transactions of the St. Johns Hospital Dermatological Society*, 52, pp.220-225.
- Fraceto, L.F., De Castro, V.L.S., Grillo, R., Ávila, D., Oliveira, H.C. and Lima, R., 2020. *Nanopesticides*. Springer International Publishing.
- Free, S.M. and Wilson, J.W., 1964. A mathematical contribution to structure-activity studies. *Journal of medicinal chemistry*, 7(4), pp.395-399.
- Freund, Y. and Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), pp.119-139.

- Friedman, J.H., 2002. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4), pp.367-378.
- Gadaleta, D., Mangiatordi, G.F., Catto, M., Carotti, A. and Nicolotti, O., 2016. Applicability domain for QSAR models: where theory meets reality. *International journal of quantitative structure-property relationships (IJQSPR)*, 1(1), pp.45-63.
- Gajewicz, A., 2017. Development of valuable predictive read-across models based on “real-life”(sparse) nanotoxicity data. *Environmental Science: Nano*, 4(6), pp.1389-1403.
- Gajewicz, A., Puzyn, T., Odziomek, K., Urbaszek, P., Haase, A., Riebeling, C., Luch, A., Irfan, M.A., Landsiedel, R., van der Zande, M. and Bouwmeester, H., 2018. Decision tree models to classify nanomaterials according to the DF4nanoGrouping scheme. *Nanotoxicology*, 12(1), pp.1-17.
- Gajewicz, A., Rasulev, B., Dinadayalane, T.C., Urbaszek, P., Puzyn, T., Leszczynska, D. and Leszczynski, J., 2012. Advancing risk assessment of engineered nanomaterials: application of computational approaches. *Advanced drug delivery reviews*, 64(15), pp.1663-1693.
- Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T. and Leszczynski, J., 2015. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), pp.313-325.
- Gatoo, M.A., Naseem, S., Arfat, M.Y., Mahmood Dar, A., Qasim, K. and Zubair, S., 2014. Physicochemical properties of nanomaterials: implication in associated toxic manifestations. *BioMed research international*, 2014.
- Ge, Y., Zhang, Y., Xia, J., Ma, M., He, S., Nie, F. and Gu, N., 2009. Effect of surface charge and agglomerate degree of magnetic iron oxide nanoparticles on KB cellular uptake in vitro. *Colloids and Surfaces B: Biointerfaces*, 73(2), pp.294-301.
- Geiser, M., 2010. Update on macrophage clearance of inhaled micro-and nanoparticles. *Journal of aerosol medicine and pulmonary drug delivery*, 23(4), pp.207-217.
- Ghobadian, M., Nabiani, M., Parivar, K., Fathi, M. and Pazooki, J., 2015. Toxic effects of magnesium oxide nanoparticles on early developmental and larval stages of zebrafish (*Danio rerio*). *Ecotoxicology and environmental safety*, 122, pp.260-267.
- Gilkes, D.M., Semenza, G.L. and Wirtz, D., 2014. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nature Reviews Cancer*, 14(6), pp.430-439.
- Golbraikh, A. and Tropsha, A., 2002. Beware of q²!. *Journal of molecular graphics and modelling*, 20(4), pp.269-276.
- Gomes, A., Fernandes, E. and Lima, J.L., 2005. Fluorescence probes used for detection of reactive oxygen species. *Journal of biochemical and biophysical methods*, 65(2-3), pp.45-80.
- Gramatica, P., 2007. Principles of QSAR models validation: internal and external. *QSAR & combinatorial science*, 26(5), pp.694-701.

- Guner, O.F., 2002. History and evolution of the pharmacophore concept in computer-aided drug design. *Current topics in medicinal chemistry*, 2(12), pp.1321-1332.
- Gütlein, M., Helma, C., Karwath, A. and Kramer, S., 2013. A large-scale empirical evaluation of cross-validation and external test set validation in (Q) SAR. *Molecular Informatics*, 32(5-6), pp.516-528.
- Hammett, L.P., 1970. *Physical organic chemistry: reaction rates, equilibria, and mechanisms*. 2nd ed. McGraw-Hill, New York.
- Handy, R.D., Henry, T.B., Scown, T.M., Johnston, B.D. and Tyler, C.R., 2008. Manufactured nanoparticles: their uptake and effects on fish—a mechanistic analysis. *Ecotoxicology*, 17, pp.396-409.
- Hansch, C. and Fujita, T., 1964. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, 86(8), pp.1616-1626.
- Hansch, C., Maloney, P.P., Fujita, T. and Muir, R.M., 1962. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, 194(4824), pp.178-180.
- Hansen, O.R., Hallberg, B., Ronquist, G., Briggs, T., Haslewood, G.A.D. and Flood, H., 1962. Hammett series with biological activity. *Acta Chem. Scand*, 16(7). 1593–1600.
- Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B., Albuquerque, M., Madhav, P.J. and Duraiswami, C., 1997. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *Journal of the American Chemical Society*, 119(43), pp.10509-10524.
- Huang, M., Khor, E. and Lim, L.Y., 2004. Uptake and cytotoxicity of chitosan molecules and nanoparticles: effects of molecular weight and degree of deacetylation. *Pharmaceutical research*, 21, pp.344-353.
- Huang, Y., Li, X., Xu, S., Zheng, H., Zhang, L., Chen, J., Hong, H., Kusko, R. and Li, R., 2020. Quantitative structure–activity relationship models for predicting inflammatory potential of metal oxide nanoparticles. *Environmental health perspectives*, 128(6), p.067010.
- Hussain, S.M. and Frazier, J.M., 2002. Cellular toxicity of hydrazine in primary rat hepatocytes. *Toxicological Sciences*, 69(2), pp.424-432.
- Jabbar, S.A.B., Twentyman, P.R. and Watson, J.V., 1989. The MTT assay underestimates the growth inhibitory effects of interferons. *British journal of cancer*, 60(4), pp.523-528.
- Jeng, H.A. and Swanson, J., 2006. Toxicity of metal oxide nanoparticles in mammalian cells. *Journal of Environmental Science and Health Part A*, 41(12), pp.2699-2711.
- Jin, Y., Kannan, S., Wu, M. and Zhao, J.X., 2007. Toxicity of luminescent silica nanoparticles to living cells. *Chemical research in toxicology*, 20(8), pp.1126-1133.
- Kang, S., Pinault, M., Pfefferle, L.D. and Elimelech, M., 2007. Single-walled carbon nanotubes exhibit strong antimicrobial activity. *Langmuir*, 23(17), pp.8670-8673.

- Kar, S., Gajewicz, A., Puzyn, T. and Roy, K., 2014. Nano-quantitative structure–activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. *Toxicology in Vitro*, 28(4), pp.600-606.
- Kar, S., Gajewicz, A., Puzyn, T., Roy, K. and Leszczynski, J., 2014. Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and environmental safety*, 107, pp.162-169.
- Karelson, M., Lobanov, V.S. and Katritzky, A.R., 1996. Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical reviews*, 96(3), pp.1027-1044.
- Kathawala, M.H., Xiong, S., Richards, M., Ng, K.W., George, S. and Loo, S.C.J., 2013. Emerging in vitro models for safety screening of high-volume production nanomaterials under environmentally relevant exposure conditions. *Small*, 9(9-10), pp.1504-1520.
- Kim, S.C., Kim, D.W., Shim, Y.H., Bang, J.S., Oh, H.S., Kim, S.W. and Seo, M.H., 2001. In vivo evaluation of polymeric micellar paclitaxel formulation: toxicity and efficacy. *Journal of controlled release*, 72(1-3), pp.191-202.
- Klaine, S.J., Alvarez, P.J., Batley, G.E., Fernandes, T.F., Handy, R.D., Lyon, D.Y., Mahendra, S., McLaughlin, M.J. and Lead, J.R., 2008. Nanomaterials in the environment: behavior, fate, bioavailability, and effects. *Environmental Toxicology and Chemistry: An International Journal*, 27(9), pp.1825-1851.
- Kononenko, I. and Kukar, M., 2007. Machine learning and data mining. Horwood Publishing, Chichester.
- Koppenol, W.H., 2001. The Haber-Weiss cycle—70 years later. *Redox Report*, 6(4), pp.229-234.
- Kovalishyn, V., Abramenko, N., Kopernyk, I., Charochkina, L., Metelytsia, L., Tetko, I.V., Peijnenburg, W. and Kustov, L., 2018. Modelling the toxicity of a large set of metal and metal oxide nanoparticles using the OCHEM platform. *Food and chemical toxicology*, 112, pp.507-517.
- Kumar, V., Sharma, N. and Maitra, S.S., 2017. In vitro and in vivo toxicity assessment of nanoparticles. *International Nano Letters*, 7(4), pp.243-256.
- Kwon, C.W., Poquet, A., Mornet, S., Campet, G., Delville, M.H., Treguer, M. and Portier, J., 2001. Electronegativity and chemical hardness: two helpful concepts for understanding oxide nanochemistry. *Materials Letters*, 51(5), pp.402-413.
- Lamon, L., Aschberger, K., Asturiol, D., Richarz, A. and Worth, A., 2019. Grouping of nanomaterials to read-across hazard endpoints: a review. *Nanotoxicology*, 13(1), pp.100-118.
- Laurent, S., Forge, D., Port, M., Roch, A., Robic, C., Vander Elst, L. and Muller, R.N., 2008. Magnetic iron oxide nanoparticles: synthesis, stabilization, vectorization, physicochemical characterizations, and biological applications. *Chemical reviews*, 108(6), pp.2064-2110.
- Lebre, F., Chatterjee, N., Costa, S., Fernández-de-Gortari, E., Lopes, C., Meneses, J., Ortiz, L., Ribeiro, A.R., Vilas-Boas, V. and Alfaro-Moreno, E., 2022. Nanosafety: an evolving

- concept to bring the safest possible nanomaterials to society and environment. *Nanomaterials*, 12(11), p.1810.
- Lee, K.J., Nallathamby, P.D., Browning, L.M., Osgood, C.J. and Xu, X.H.N., 2007. In vivo imaging of transport and biocompatibility of single silver nanoparticles in early development of zebrafish embryos. *ACS nano*, 1(2), pp.133-143.
- Leone, C., Bertuzzi, E.E., Toropova, A.P., Toropov, A.A. and Benfenati, E., 2018. CORAL: Predictive models for cytotoxicity of functionalized nanozeolites based on quasi-SMILES. *Chemosphere*, 210, pp.52-56.
- Lerman, P.M., 1980. Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 29(1), pp.77-84.
- Li, J., Wang, C., Yue, L., Chen, F., Cao, X. and Wang, Z., 2022. Nano-QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: A review. *Ecotoxicology and Environmental Safety*, 243, p.113955.
- Li, K.G., Chen, J.T., Bai, S.S., Wen, X., Song, S.Y., Yu, Q., Li, J. and Wang, Y.Q., 2009. Intracellular oxidative stress and cadmium ions release induce cytotoxicity of unmodified cadmium sulfide quantum dots. *Toxicology in Vitro*, 23(6), pp.1007-1013.
- Li, Z., Omidvar, N., Chin, W.S., Robb, E., Morris, A., Achenie, L. and Xin, H., 2018. Machine-learning energy gaps of porphyrins with molecular graph representations. *The Journal of Physical Chemistry A*, 122(18), pp.4571-4578.
- Lifnick, R.L., 1989. Hans Horst Meyer and the lipid theory of narcosis. *Trends in Pharmacological Sciences*, 10(7), pp.265-269.
- Lin, S., Zhao, Y., Ji, Z., Ear, J., Chang, C.H., Zhang, H., Low-Kam, C., Yamada, K., Meng, H., Wang, X. and Liu, R., 2013. Zebrafish high-throughput screening to study the impact of dissolvable metal oxide nanoparticles on the hatching enzyme, ZHE1. *Small*, 9(9-10), pp.1776-1785.
- Lipnick, R.L., 1986. Charles Ernest Overton: Narcosis studies and a contribution to general pharmacology. In *Studies of Narcosis: Charles Ernest Overton* (pp. 14-22). Dordrecht: Springer Netherlands.
- Liu, Phang, I.Y., Shen, L., Chow, S.Y. and Zhang, W.D., 2004. Morphology and mechanical properties of multiwalled carbon nanotubes reinforced nylon-6 composites. *Macromolecules*, 37(19), pp.7214-7222.
- Liu, Y., Liu, Z., Luo, X. and Zhao, H., 2022. Diagnosis of Parkinson's disease based on SHAP value feature selection. *Biocybernetics and Biomedical Engineering*, 42(3), pp.856-869.
- Lozano-Fernández, T., Ballester-Antxordoki, L., Pérez-Temprano, N., Rojas, E., Sanz, D., Iglesias-Gaspar, M., Moya, S., González-Fernández, Á. and Rey, M., 2014. Potential impact of metal oxide nanoparticles on the immune system: the role of integrins, L-selectin and the chemokine receptor CXCR4. *Nanomedicine: Nanotechnology, Biology and Medicine*, 10(6), pp.1301-1310.
- Lullmann-Rauch, R., 1979. Drug-induced lysosomal storage disorders. 49-130.

- Lund, B.O., Miller, D.M. and Woods, J.S., 1991. Mercury-induced H₂O₂ production and lipid peroxidation in vitro in rat kidney mitochondria. *Biochemical Pharmacology*, 42, pp.S181-S187.
- Mac Fhionnlaoich, N. and Guldin, S., 2020. Information entropy as a reliable measure of nanoparticle dispersity. *Chemistry of Materials*, 32(9), pp.3701-3706.
- Manke, A., Wang, L. and Rojanasakul, Y., 2013. Mechanisms of nanoparticle-induced oxidative stress and toxicity. *BioMed research international*, 2013.
- Marcílio, W.E. and Eler, D.M., 2020, November. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 340-347). Ieee.
- Marshall, N.J., Goodwin, C.J. and Holt, S.J., 1995. A critical assessment of the use of microculture tetrazolium assays to measure cell growth and function. *Growth regulation*, 5(2), pp.69-84.
- Martin, Y. and Stouch, T., 2011. In tribute to Corwin Hansch, father of QSAR. *Journal of computer-aided molecular design*, 25(6), pp.491-491.
- Maselli, V., Siciliano, A., Giorgio, A., Falanga, A., Galdiero, S., Guida, M., Fulgione, D. and Galdiero, E., 2017. Multigenerational effects and DNA alterations of QDs-Indolicidin on *Daphnia magna*. *Environmental pollution*, 224, pp.597-605.
- McDonald, G.C., 2009. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), pp.93-100.
- Mech, A., Rasmussen, K., Jantunen, P., Aicher, L., Alessandrelli, M., Bernauer, U., Bleeker, E.A.J., Bouillard, J., Di Prospero Fanghella, P., Draisci, R. and Dusinska, M., 2019. Insights into possibilities for grouping and read-across for nanomaterials in EU chemicals legislation. *Nanotoxicology*, 13(1), pp.119-141.
- Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W. and Hamprecht, F.A., 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10, pp.1-16.
- Mienye, I.D. and Sun, Y., 2022. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10, pp.99129-99149.
- Mitteroecker, P. and Bookstein, F., 2011. Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, 38, pp.100-114.
- Mo, Y. and Lim, L.Y., 2005. Paclitaxel-loaded PLGA nanoparticles: potentiation of anticancer activity by surface conjugation with wheat germ agglutinin. *Journal of controlled release*, 108(2-3), pp.244-262.
- Mohd Firdaus, M.A., Agatz, A., Hodson, M.E., Al-Khazrajy, O.S. and Boxall, A.B., 2018. Fate, uptake, and distribution of nanoencapsulated pesticides in soil–earthworm systems and implications for environmental risk assessment. *Environmental toxicology and chemistry*, 37(5), pp.1420-1429.

- Molinari, B.L., Tasat, D.R., Palmieri, M.A., O'Connor, S.E. and Cabrini, R.L., 2003. Cell-based quantitative evaluation of the MTT assay. *Analytical and quantitative cytology and histology*, 25(5), pp.254-262.
- Monteiro-Riviere, N.A., Nemanich, R.J., Inman, A.O., Wang, Y.Y. and Riviere, J.E., 2005. Multi-walled carbon nanotube interactions with human epidermal keratinocytes. *Toxicology letters*, 155(3), pp.377-384.
- Mu, Q., Jiang, G., Chen, L., Zhou, H., Fourches, D., Tropsha, A. and Yan, B., 2014. Chemical basis of interactions between engineered nanoparticles and biological systems. *Chemical reviews*, 114(15), pp.7740-7781.
- Mu, Y., Kobayashi, T., Sumita, M., Yamamoto, A. and Hanawa, T., 2000. Metal ion release from titanium with active oxygen species generated by rat macrophages in vitro. *Journal of Biomedical Materials Research: An Official Journal of The Society for Biomaterials and The Japanese Society for Biomaterials*, 49(2), pp.238-243.
- Mu, Y., Wu, F., Zhao, Q., Ji, R., Qie, Y., Zhou, Y., Hu, Y., Pang, C., Hristozov, D., Giesy, J.P. and Xing, B., 2016. Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. *Nanotoxicology*, 10(9), pp.1207-1214.
- Natarajan, M., Mohan, S., Martinez, B.R., Meltz, M.L. and Herman, T.S., 2000. Antioxidant compounds interfere with the 3. *Cancer detection and prevention*, 24(5), pp.405-414.
- Nel, A., Xia, T., Madler, L. and Li, N., 2006. Toxic potential of materials at the nanolevel. *science*, 311(5761), pp.622-627.
- Nel, A.E., Mädler, L., Velegol, D., Xia, T., Hoek, E.M., Somasundaran, P., Klaessig, F., Castranova, V. and Thompson, M., 2009. Understanding biophysicochemical interactions at the nano–bio interface. *Nature materials*, 8(7), pp.543-557.
- Nelms, S.E., Galloway, T.S., Godley, B.J., Jarvis, D.S. and Lindeque, P.K., 2018. Investigating microplastic trophic transfer in marine top predators. *Environmental pollution*, 238, pp.999-1007.
- Nemes, Z., Dietz, R., Lüth, J.B., Gomba, S., Hackenthal, E. and Gross, F., 1979. The pharmacological relevance of vital staining with neutral red. *Experientia*, 35, pp.1475-1476.
- Nemmar, A., Hoet, P.H., Vanquickenborne, B., Dinsdale, D., Thomeer, M., Hoylaerts, M.F., Vanbilloen, H., Mortelmans, L. and Nemery, B., 2002. Passage of inhaled particles into the blood circulation in humans. *Circulation*, 105(4), pp.411-414.
- Niki, E., 2008. Lipid peroxidation products as oxidative stress biomarkers. *Biofactors*, 34(2), pp.171-180.
- Oberdörster, G., 2010. Safety assessment for nanotechnology and nanomedicine: concepts of nanotoxicology. *Journal of internal medicine*, 267(1), pp.89-105.
- Oberlintner, A., Likozar, B. and Novak, U., 2021. Hydrophobic functionalization reactions of structured cellulose nanomaterials: Mechanisms, kinetics and in silico multi-scale models. *Carbohydrate Polymers*, 259, p.117742.

- Odziomek, K., Ushizima, D., Oberbek, P., Kurzydłowski, K.J., Puzyn, T. and Haranczyk, M., 2017. Scanning electron microscopy image representativeness: morphological data on nanoparticles. *Journal of microscopy*, 265(1), pp.34-50.
- OECD Principles for the validation, for regulatory purposes, of (Quantitative) Structure Activity Relationship models. 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides, and Biotechnology. 2004. Paris, France: Organisation for economic co-operation and development. Available at: <http://www.oecd.org/env/ehs/riskassessment/validationofqsarmodels.html>
- OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment. OECD. 2014. Doi:10.1787/9789264085442-en
- OECD, O.; Environment health and safety publications series on testing and assessment No. 69, Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models. 2007
- Ojha, P.K., Kar, S., Roy, K. and Leszczynski, J., 2019. Toward comprehension of multiple human cells uptake of engineered nano metal oxides: quantitative inter cell line uptake specificity (QICLUS) modeling. *Nanotoxicology*, 13(1), pp.14-34.
- Oomen, A.G., Bleeker, E.A., Bos, P.M., Van Broekhuizen, F., Gottardo, S., Groenewold, M., Hristozov, D., Hund-Rinke, K., Irfan, M.A., Marcomini, A. and Peijnenburg, W.J., 2015. Grouping and read-across approaches for risk assessment of nanomaterials. *International journal of environmental research and public health*, 12(10), pp.13415-13434.
- Ozben, T., 2007. Oxidative stress and apoptosis: impact on cancer therapy. *Journal of pharmaceutical sciences*, 96(9), pp.2181-2196.
- Pan, Y., Li, T., Cheng, J., Telesca, D., Zink, J.I. and Jiang, J., 2016. Nano-QSAR modeling for predicting the cytotoxicity of metal oxide nanoparticles using novel descriptors. *RSC advances*, 6(31), pp.25766-25775.
- Park, E.J., Yoon, J., Choi, K., Yi, J. and Park, K., 2009. Induction of chronic inflammation in mice treated with titanium dioxide nanoparticles by intratracheal instillation. *Toxicology*, 260(1-3), pp.37-46.
- Park, M.V., Neigh, A.M., Vermeulen, J.P., de la Fonteyne, L.J., Verharen, H.W., Briedé, J.J., van Loveren, H. and de Jong, W.H., 2011. The effect of particle size on the cytotoxicity, inflammation, developmental toxicity and genotoxicity of silver nanoparticles. *Biomaterials*, 32(36), pp.9810-9817.
- Park, S.Y. and Choi, J., 2010. Geno-and ecotoxicity evaluation of silver nanoparticles in freshwater crustacean *Daphnia magna*. *Environmental Engineering Research*, 15(1), pp.23-27.
- Pathakoti, K., Huang, M.J., Watts, J.D., He, X. and Hwang, H.M., 2014. Using experimental data of *Escherichia coli* to develop a QSAR model for predicting the photo-induced cytotoxicity of metal oxide nanoparticles. *Journal of Photochemistry and Photobiology B: Biology*, 130, pp.234-240.

- Pesantez-Narvaez, J., Guillen, M. and Alcañiz, M., 2019. Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks*, 7(2), p.70.
- Pettersson, M., Almlin, S., Romanos, G.E. and Johansson, A., 2022. Ti ions induce IL-1 β release by activation of the NLRP3 inflammasome in a human macrophage cell line. *Inflammation*, 45(5), pp.2027-2037.
- Polanski, J., 2009. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. *Current medicinal chemistry*, 16(25), pp.3243-3257.
- Pujalté, I., Passagne, I., Brouillaud, B., Tréguer, M., Durand, E., Ohayon-Courtès, C. and L'Azou, B., 2011. Cytotoxicity and oxidative stress induced by different metallic nanoparticles on human kidney cells. *Particle and fibre toxicology*, 8, pp.1-16.
- Purcell, W.P., Bass, G.F. and Clayton, J.M., 1973. Strategy of drug design: a guide to biological activity. A molecular guide to biological activity. Wiley and Sons, New York.
- Puzyn, T., Leszczynska, D. and Leszczynski, J., 2009. Toward the development of “nano-QSARs”: advances and challenges. *Small*, 5(22), pp.2494-2509.
- Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T.P., Michalkova, A., Hwang, H.M., Toropov, A., Leszczynska, D. and Leszczynski, J., 2011. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature nanotechnology*, 6(3), pp.175-178.
- Qi, R., Pan, Y., Cao, J., Jia, Z. and Jiang, J., 2020. The cytotoxicity of nanomaterials: Modeling multiple human cells uptake of functionalized magneto-fluorescent nanoparticles via nano-QSAR. *Chemosphere*, 249, p.126175.
- Raftis, J.B. and Miller, M.R., 2019. Nanoparticle translocation and multi-organ toxicity: a particularly small problem. *Nano Today*, 26, pp.8-12.
- Rekker, R.F., 1992. The history of drug research: from Overton to Hansch. *Quantitative Structure-Activity Relationships*, 11(2), pp.195-199.
- Remelli, M., Nurchi, V.M., Lachowicz, J.I., Medici, S., Zoroddu, M.A. and Peana, M., 2016. Competition between Cd (II) and other divalent transition metal ions during complex formation with amino acids, peptides, and chelating agents. *Coordination Chemistry Reviews*, 327, pp.55-69.
- Ribeiro, F., Gallego-Urrea, J.A., Jurkschat, K., Crossley, A., Hassellöv, M., Taylor, C., Soares, A.M. and Loureiro, S., 2014. Silver nanoparticles and silver nitrate induce high toxicity to *Pseudokirchneriella subcapitata*, *Daphnia magna* and *Danio rerio*. *Science of the Total Environment*, 466, pp.232-241.
- Rigatti, S.J., 2017. Random forest. *Journal of Insurance Medicine*, 47(1), pp.31-39.
- Rizk, M.Z., Ali, S.A., Hamed, M.A., El-Rigal, N.S., Aly, H.F. and Salah, H.H., 2017. Toxicity of titanium dioxide nanoparticles: effect of dose and time on biochemical disturbance, oxidative stress and genotoxicity in mice. *Biomedicine & pharmacotherapy*, 90, pp.466-472.

- Rosenfeldt, R.R., Seitz, F., Senn, L., Schilde, C., Schulz, R. and Bundschuh, M., 2015. Nanosized Titanium Dioxide Reduces Copper Toxicity • The Role of Organic Material and the Crystalline Phase. *Environmental science & technology*, 49(3), pp.1815-1822.
- Roy, J. and Roy, K., 2021. Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach. *Environmental Science: Nano*, 8(11), pp.3395-3407.
- Roy, J. and Roy, K., 2022. Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells. *Nanotoxicology*, 16(5), pp.629-644.
- Roy, J. and Roy, K., 2022. Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (Me Ox NPs) used in nanomedicines: A mechanistic QSAR approach. *Environmental Science: Nano*, 9(9), pp.3456-3470.
- Roy, J. and Roy, K., 2023. Evaluating metal oxide nanoparticle (MeOx NP) toxicity with different types of nano descriptors mainly focusing on simple periodic table-based descriptors: A mini-review. *Environmental Science: Nano*.
- Roy, J. and Roy, K., 2023. Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: a nano-QSAR approach. *SAR and QSAR in Environmental Research*, 34(6), pp.459-474.
- Roy, J., Ojha, P.K. and Roy, K., 2019. Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table based descriptors. *Nanotoxicology*, 13(5), pp.701-716.
- Roy, J., Pore, S. and Roy, K., 2023. Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches. *Beilstein Journal of Nanotechnology*, 14(1), pp.939-950.
- Roy, K. and Mitra, I., 2011. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Combinatorial chemistry & high throughput screening*, 14(6), pp.450-474.
- Roy, K., Ambure, P. and Kar, S., 2018a. How precise are our quantitative structure–activity relationship derived predictions for new query chemicals?. *ACS omega*, 3(9), pp.11392-11406.
- Roy, K., Ambure, P., Kar, S. and Ojha, P.K., 2018b. Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models?. *Journal of Chemometrics*, 32(4), p.e2992.
- Roy, K., Das, R.N., Ambure, P. and Aher, R.B., 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 152, pp.18-33.
- Roy, K., Kar, S. and Ambure, P., 2015. On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 145, pp.22-29.

- Roy, K., Kar, S. and Das, R.N., 2015. A primer on QSAR/QSPR modeling: fundamental concepts. Springer.
- Roy, K., Kar, S., Das, R.N., Roy, K., Kar, S. and Das, R.N., 2015. Statistical methods in QSAR/QSPR. A Primer on QSAR/QSPR Modeling: Fundamental Concepts, pp.37-59.
- Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N. and Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. *Journal of chemical information and modeling*, 52(2), pp.396-408.
- Roy, P.P., Leonard, J.T. and Roy, K., 2008. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 90(1), pp.31-42.
- Russell, S. J.; Norvig, P.; Davis, E.; Edwards, D. D.; Forsyth, D.; Hay, N. J.; Malik, J. M.; Mittal, M.; Sahami, M.; Thrun, S. *Artificial intelligence a modern approach*; Pearson Education, Inc.: Upper saddle River, New Jersey, 2010.
- Ryhänen, J., Kallioinen, M., Tuukkanen, J., Lehenkari, P., Junila, J., Niemelä, E., Sandvik, P. and Serlo, W., 1999. Bone modeling and cell–material interface responses induced by nickel–titanium shape memory alloy after periosteal implantation. *Biomaterials*, 20(14), pp.1309-1317.
- Ryter, S.W., Kim, H.P., Hoetzel, A., Park, J.W., Nakahira, K., Wang, X. and Choi, A.M., 2007. Mechanisms of cell death in oxidative stress. *Antioxidants & redox signaling*, 9(1), pp.49-89.
- Saadat, A., Dehghani Varniab, A. and Madani, S.M., 2022. Prediction of the antibacterial activity of the green synthesized silver nanoparticles against gram negative and positive bacteria by using machine learning algorithms. *Journal of Nanomaterials*, 2022.
- Sang, L., Wang, Y., Zong, C., Wang, P., Zhang, H., Guo, D., Yuan, B. and Pan, Y., 2022. Machine learning for evaluating the cytotoxicity of mixtures of nano-TiO₂ and heavy metals: qSAR model apply random forest algorithm after clustering analysis. *Molecules*, 27(18), p.6125.
- Sayes, C.M., Reed, K.L. and Warheit, D.B., 2007. Assessing toxicity of fine and nanoparticles: comparing in vitro measurements to in vivo pulmonary toxicity profiles. *Toxicological sciences*, 97(1), pp.163-180.
- Schaming, D. and Remita, H., 2015. Nanotechnology: from the ancient time to nowadays. *Foundations of Chemistry*, 17, pp.187-205.
- Schrand, A.M., Rahman, M.F., Hussain, S.M., Schlager, J.J., Smith, D.A. and Syed, A.F., 2010. Metal-based nanoparticles and their toxicity assessment. *Wiley interdisciplinary reviews: Nanomedicine and Nanobiotechnology*, 2(5), pp.544-568.
- Schultz, T.W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D.J., Mahony, C., Schwarz, M., White, A. and Cronin, M.T.D., 2015. A strategy for structuring and reporting a read-across prediction of toxicity. *Regulatory Toxicology and Pharmacology*, 72(3), pp.586-601.

- Schultz, T.W., Cronin, M.T., Walker, J.D. and Aptula, A.O., 2003. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular structure: THEOCHEM*, 622(1-2), pp.1-22.
- Segerstrom, S.C. and Miller, G.E., 2004. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin*, 130(4), p.601.
- Sengul, A.B. and Asmatulu, E., 2020. Toxicity of metal and metal oxide nanoparticles: a review. *Environmental Chemistry Letters*, 18(5), pp.1659-1683.
- Shah, A. and Dobrovolskaia, M.A., 2018. Immunological effects of iron oxide nanoparticles and iron-based complex drug formulations: Therapeutic benefits, toxicity, mechanistic insights, and translational considerations. *Nanomedicine: Nanotechnology, Biology and Medicine*, 14(3), pp.977-990.
- Sheehan, D., Meade, G., Foley, V.M. and Dowd, C.A., 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical journal*, 360(1), pp.1-16.
- Shin, H.K., Kim, S. and Yoon, S., 2021. Use of size-dependent electron configuration fingerprint to develop general prediction models for nanomaterials. *NanoImpact*, 21, p.100298.
- Simon-Vazquez, R., Lozano-Fernandez, T., Davila-Grana, A. and Gonzalez-Fernandez, A., 2016. Metal oxide nanoparticles interact with immune cells and activate different cellular responses. *International journal of nanomedicine*, pp.4657-4668.
- Singh, K.P. and Gupta, S., 2014. Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. *RSC Advances*, 4(26), pp.13215-13230.
- Sivanandam, S.N., Deepa, S.N., Sivanandam, S.N. and Deepa, S.N., 2008. Genetic algorithms (pp. 15-37). Springer Berlin Heidelberg.
- Sizochenko, N., Leszczynska, D. and Leszczynski, J., 2017. Modeling of interactions between the zebrafish hatching enzyme ZHE1 and a series of metal oxide nanoparticles: Nano-QSAR and causal analysis of inactivation mechanisms. *Nanomaterials*, 7(10), p.330.
- Sizochenko, N., Rasulev, B., Gajewicz, A., Kuz'min, V., Puzyn, T. and Leszczynski, J., 2014. From basic physics to mechanisms of toxicity: The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. *Nanoscale*, 6(22), pp.13986-13993.
- Soenen, S.J., Rivera-Gil, P., Montenegro, J.M., Parak, W.J., De Smedt, S.C. and Braeckmans, K., 2011. Cellular toxicity of inorganic nanoparticles: common aspects and guidelines for improved nanotoxicity evaluation. *Nano today*, 6(5), pp.446-465.
- STATISTICA is a statistical software of STATSOFT Inc., USA, <http://www.statsoft.com/>
- Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A. and Folke, C., 2015. Planetary boundaries: Guiding human development on a changing planet. *science*, 347(6223), p.1259855.
- Subhan, M.A., Choudhury, K.P. and Neogi, N., 2021. Advances with molecular nanomaterials in industrial manufacturing applications. *Nanomanufacturing*, 1(2), pp.75-97.

- Suhendra, E., Chang, C.H., Hou, W.C. and Hsieh, Y.C., 2020. A review on the environmental fate models for predicting the distribution of engineered nanomaterials in surface waters. *International journal of molecular sciences*, 21(12), p.4554.
- Sun, H., Jiang, C., Wu, L., Bai, X. and Zhai, S., 2019. Cytotoxicity-related bioeffects induced by nanoparticles: the role of surface chemistry. *Frontiers in Bioengineering and Biotechnology*, 7, p.414.
- Tantra, R., Oksel, C., Puzyn, T., Wang, J., Robinson, K.N., Wang, X.Z., Ma, C.Y. and Wilkins, T., 2015. Nano (Q) SAR: Challenges, pitfalls and perspectives. *Nanotoxicology*, 9(5), pp.636-642.
- Taylor, P.J., 1990. In Hansch, C., Sammes, PG and Taylor, JB (Eds.) *Comprehensive Medicinal Chemistry*, Vol. 4.
- Tiwari, J.N., Tiwari, R.N. and Kim, K.S., 2012. Zero-dimensional, one-dimensional, two-dimensional and three-dimensional nanostructured materials for advanced electrochemical energy devices. *Progress in Materials Science*, 57(4), pp.724-803.
- Todeschini, R.O.B.E.R.T.O. and Consonni, V., 2009. *Molecular Descriptors for Chemoinformatics*. Vol. 1. Alphabetical Listing; Vol. 2. Appendices, References (pp. 1-1257). Wiley-VCH.
- Tong, W., Hong, H., Xie, Q., Shi, L., Fang, H. and Perkins, R., 2005. Assessing QSAR limitations-A regulatory perspective. *Current Computer-Aided Drug Design*, 1(2), pp.195-205.
- Toropov, A.A., Sizochenko, N., Toropova, A.P. and Leszczynski, J., 2018. Towards the development of global nano-quantitative structure–property relationship models: Zeta potentials of metal oxide nanoparticles. *Nanomaterials*, 8(4), p.243.
- Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D. and Leszczynski, J., 2012. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli*. *Chemosphere*, 89(9), pp.1098-1102.
- Toropova, A.P., Toropov, A.A., Benfenati, E., Korenstein, R., Leszczynska, D. and Leszczynski, J., 2015. Optimal nano-descriptors as translators of eclectic data into prediction of the cell membrane damage by means of nano metal-oxides. *Environmental Science and Pollution Research*, 22, pp.745-757.
- Tortella, G.R., Rubilar, O., Durán, N., Diez, M.C., Martínez, M., Parada, J. and Seabra, A.B., 2020. Silver nanoparticles: Toxicity in model organisms as an overview of its hazard for human health and the environment. *Journal of hazardous materials*, 390, p.121974.
- Tripathi, G.K., 2019. Engineered Nanomaterials And Their Properties: A Review. *Biosci. Biotechnol. Res. Commun*, 12, pp.764-771.
- Tute, M.S., 1990. History and objectives of quantitative drug design, in: Ramsden, C.A. (Ed.), *Quantitative Drug Design*. Pergamon Press, Oxford, pp. 1–31.
- UMETRICS SIMCA-P 10.0, info@umetrics.com:www.umetrics.com 2002 Umea, Sweden.

- Valic, M.S., Halim, M., Schimmer, P. and Zheng, G., 2020. Guidelines for the experimental design of pharmacokinetic studies with nanomaterials in preclinical animal models. *Journal of controlled release*, 323, pp.83-101.
- van De Waterbeemd, H., 1992. The history of drug research: from Hansch to the present. *Quantitative Structure-Activity Relationships*, 11(2), pp.200-204.
- Varsou, D.D., Afantitis, A., Tsoumanis, A., Papadiamantis, A., Valsami-Jones, E., Lynch, I. and Melagraki, G., 2020. Zeta-potential read-across model utilizing nanodescriptors extracted via the nanextract image analysis tool available on the enalos nanoinformatics cloud platform. *Small*, 16(21), p.1906588.
- Vedani, A., Dobler, M. and Lill, M.A., 2005. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of medicinal chemistry*, 48(11), pp.3700-3703.
- Venkatasubramanian, V. and Sundaram, A., 2002. Genetic algorithms: introduction and applications. *Encyclopedia of Computational Chemistry*, 2.
- Walker, J.D., Enache, M. and Dearden, J.C., 2003. Quantitative cationic-activity relationships for predicting toxicity of metals. *Environmental Toxicology and Chemistry: An International Journal*, 22(8), pp.1916-1935.
- Walker, J.D., Enache, M. and Dearden, J.C., 2007. Quantitative cationic activity relationships for predicting toxicity of metal ions from physicochemical properties and natural occurrence levels. *QSAR & Combinatorial Science*, 26(4), pp.522-527.
- Wang, T.Y., Libardo, M.D.J., Angeles-Boza, A.M. and Pellois, J.P., 2017. Membrane oxidation in cell delivery and cell killing applications. *ACS chemical biology*, 12(5), pp.1170-1182.
- Whitehead, T.M., Strickland, J., Conduit, G.J., Borrel, A., Mucs, D. and Baskerville-Abraham, I., 2023. Quantifying the Benefits of Imputation over QSAR Methods in Toxicology Data Modeling. *Journal of Chemical Information and Modeling*.
- Winkler, D.A., Mombelli, E., Pietroiusti, A., Tran, L., Worth, A., Fadeel, B. and McCall, M.J., 2013. Applying quantitative structure–activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology*, 313(1), pp.15-23.
- Wold, S., 1995. PLS for multivariate linear modeling. *Chemometric methods in molecular design*, pp.195-218.
- Wold, S., Sjöström, M. and Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), pp.109-130.
- Wold, S., Trygg, J., Berglund, A. and Antti, H., 2001. Some recent developments in PLS modeling. *Chemometrics and intelligent laboratory systems*, 58(2), pp.131-150.
- Wu, Q., Miao, W., Zhang, Y., Gao, H. and Hui, D., 2020. Mechanical properties of nanomaterials: a review. *Nanotechnol Rev* 9: 259–273.
- Wyrzykowska, E., Mikołajczyk, A., Sikorska, C. and Puzyn, T., 2016. Development of a novel in silico model of zeta potential for metal oxide nanoparticles: a nano-QSPR approach. *Nanotechnology*, 27(44), p.445702.

- Xia, T., Kovochich, M., Brant, J., Hotze, M., Sempf, J., Oberley, T., Sioutas, C., Yeh, J.I., Wiesner, M.R. and Nel, A.E., 2006. Comparison of the abilities of ambient and manufactured nanoparticles to induce cellular toxicity according to an oxidative stress paradigm. *Nano letters*, 6(8), pp.1794-1807.
- Xia, T., Kovochich, M., Liong, M., Madler, L., Gilbert, B., Shi, H., Yeh, J.I., Zink, J.I. and Nel, A.E., 2008. Comparison of the mechanism of toxicity of zinc oxide and cerium oxide nanoparticles based on dissolution and oxidative stress properties. *ACS nano*, 2(10), pp.2121-2134.
- Yan, X., Sedykh, A., Wang, W., Yan, B. and Zhu, H., 2020. Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. *Nature communications*, 11(1), p.2519.
- Yan, X., Yue, T., Winkler, D.A., Yin, Y., Zhu, H., Jiang, G. and Yan, B., 2023. Converting nanotoxicity data to information using artificial intelligence and simulation. *Chemical Reviews*, 123(13), pp.8575-8637.
- Yang, K., Zhu, L. and Xing, B., 2006. Adsorption of polycyclic aromatic hydrocarbons by carbon nanomaterials. *Environmental science & technology*, 40(6), pp.1855-1861.
- Zahradnik, R., 1962. Correlation of the biological activity of organic compounds by means of the linear free energy relationships. *Experientia*, 18(11), pp.534-536.
- Zahradnik, R., 1962. Influence of the structure of aliphatic substituents on the magnitude of the biological effect of substances. *Archives Internationales de Pharmacodynamie et de Thérapie*, 135, pp.311-329.
- Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., Pokhrel, S., Lin, S., Wang, X., Liao, Y.P. and Wang, M., 2012. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS nano*, 6(5), pp.4349-4368.
- Zhang, F., Wang, Z., Peijnenburg, W.J. and Vijver, M.G., 2023. Machine learning-driven QSAR models for predicting the mixture toxicity of nanoparticles. *Environment International*, 177, p.108025.
- Zhou, J., Foster, D.P., Stine, R.A., Ungar, L.H. and Guyon, I., 2006. Streamwise feature selection. *Journal of Machine Learning Research*, 7(9).
- Zhou, Z., Tang, X., Dai, W., Shi, J. and Chen, H., 2017. Nano-QSAR models for predicting cytotoxicity of metal oxide nanoparticles (MONPs) to *E. coli*. *Canadian Journal of Chemistry*, 95(8), pp.863-866.
- Zhu, M., Nie, G., Meng, H., Xia, T., Nel, A. and Zhao, Y., 2013. Physicochemical properties determine nanomaterial cellular uptake, transport, and fate. *Accounts of chemical research*, 46(3), pp.622-631.

APPENDIX

REPRINTS

PAPER

View Article Online
View Journal | View Issue



Cite this: *Environ. Sci.: Nano*, 2021, 8, 3395

Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach†

Joyita Roy and Kunal Roy *

Nowadays, nanotechnology is a promising area of research. Therefore, a detailed understanding of comprehensive risk assessment of nanomaterials and their behavior to the biological system is required before working with any metallic oxide (hydroxide) nanoparticles (MeOx NPs). Empirical testing of all toxicity endpoints of MeOx NPs is impossible with limited resources; therefore, an efficient alternative method complementary to experimentation is very essential. Quantitative structure–activity relationship (QSAR) models are widely adopted to predict the toxicity of untested nanomaterials. In the present study, QSAR models were developed to computationally evaluate the cytotoxicity of MeOx NPs towards RAW 264.7 cells. Both MLR and PLS models were developed with 25 NPs (nanoparticles) from 6 metal (hydro) oxide families (SiO_2 , TiO_2 , CeO_2 , AlOOH , ZnO , and Ni(OH)_2) using the release of lactate dehydrogenase (LDH_R) from the cell as the endpoint. Multiple linear regression (MLR) models were developed using the best subset selection method followed by enhancement of model derived predictions by the application of the intelligent consensus predictor (ICP) tool. On the other hand, a partial least squares (PLS) model was developed using the Small Dataset Modeler software which can handle a limited number of data points. Periodic table descriptors and some physicochemical descriptors have been used to characterize the nanostructure of MeOx NPs. All models were validated meticulously by employing internal and external validation metrics followed by the randomization test and applicability domain (AD) study. The statistical metric values suggested the robustness of the models. The insights from both models suggested that electronegativity, solubility, formation of a metal cation increase the cytotoxicity whereas non-metal NPs and an increase of the oxidation degree contribute to low toxicity towards RAW 264.7 cells. A short proposed mechanism has also been explained to determine the cause of cellular toxicity by the MeOx NPs.

Received 10th August 2021,
Accepted 30th September 2021

DOI: 10.1039/d1en00733e

rsc.li/es-nano

Environmental significance

The fate and dynamic behavior of metal oxide and hydroxide nanoparticles (MeOx NPs) depend on their interaction with the different biological compartments of the environment. Considering the increased synthesis of metallic NPs and lack of experimental toxicity data, it is essential to develop toxicity assessment methods to predict their possible environmental toxicity impact. This study involves sophisticated computational tools to provide a comprehensive mechanistic view of how electronegativity, solubility, formation of a metal cation and an increase of the oxidation degree collectively influence the cytotoxicity of MeOx NPs to RAW 264.7 cells. This approach aims to develop safer and effective MeOx NPs for commercial, environmental and medicinal use.

1. Introduction

In 1959, the term ‘Nanotechnology’ was coined by the Nobel laureate Richard P. Feynman in his well-known lecture ‘There’s

Plenty of Room at the Bottom”,¹ and since then there has been a lot of developments in this field. Nanotechnology has been applied in diverse areas, and at the same time, the probability that it could adversely affect humans and the environment has also expanded rapidly.² Today, nanotechnology materials have become the prime focus of many researchers due to the unique properties and diversity of such materials. One of the important classes of nanomaterials (NMs) is metal oxide nanoparticles (MeOx NPs). They have been included in different products such as paints, additives in food, electronic devices, cosmetics, biological products, and nanomedicines.³ Due to this fact, the

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India.

E-mail: kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in;

Fax: +91 33 2837 1078; Tel: +91 98315 94140

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1en00733e

interaction of nanoparticles with humans and the environment is inevitable. Their potential threat to human life and the environment has also drawn much attention.⁴ MeOx NPs venture the human system through breathing, gastrointestinal tract, and skin, and then circulate *via* the blood or lymphatic system and eventually accumulate in various organs.⁵ Several researchers have found that MeOx NPs can decrease cell longevity through production of ROS (reactive oxygen species) which would cause mitochondrial damage by modulating the DNA strand that further leads to other ways of cell death *via* autophagy, pyroptosis and apoptosis. In some cases, an enzyme may not produce superoxide but could accelerate free radical production, and one such candidate is the lactate dehydrogenase enzyme (LDH).

LDH, a cytoplasmic enzyme present in almost all kinds of cells, leaks out into the extracellular space when the plasma membrane becomes distorted.⁶ It is believed that the ROS produced from mitochondria can activate LDH-bound NADH and enhance the free radical reaction chain in injured cells through other sources (NO_3^- , O_2 , H_2O_2 , and superoxide ($\text{O}^{\cdot-}$))⁷ that might eventually cause toxicity towards the cells.

Considering the complexity of nanoparticles (NPs) and their diverse applications to mankind, it is required to determine their expected toxicity economically and more easily. There is a desperate need to develop a technical method to extract the essential properties of NPs.⁸ Since the investigational assessment of hazardous characteristics of NPs is laborious as well as expensive, and most importantly to avoid animal sacrifice, alternative computational methods can give reliable predictions before their mass production. *In silico* approaches are extensively accepted by regulatory agencies such as the United States Environmental Protection Agency (U.S. E.P.A) to fill the data gap of toxicity with no experimental details recorded.

QSAR (quantitative structure–activity relationship) is the front runner among the alternative *in silico* methods. The implication of QSAR in the field of risk assessment and management of NPs is recommended by regulatory agencies like the European Union Commission's Scientific Committee on Toxicity, European Centre for the Validation of Alternative Methods (ECVAM), and the Agency for Toxic Substances and Disease Registry (ATSDR). When QSAR is applied to nanostructures, it is then termed as nano-QSAR.⁹ *In silico* predictive tools like nano-QSAR can be useful to facilitate better understanding of certain mechanisms.

The application of computational approaches in the nanomaterials field has been increasing as demonstrated by some studies. Puzyn *et al.*¹⁰ have applied nano-QSAR modeling on a dataset of 17 metallic oxides having different chemical compositions with *Escherichia coli* toxicity and concluded that the cause of toxicity was due to the electronegativity and cation charge. Further, this was corroborated by Gajewicz *et al.*¹¹ while working with the HaCaT cell line and 18 MeOx NPs. Forest *et al.*¹² have also performed a nano-QSAR study with 25 nanoparticles from 6 metal hydro (oxide) families to determine the cytotoxicity to RAW 267.4 cells. A classification decision tree

along with a partial least squares (PLS) model was developed with a limited number of data points. However, the PLS models developed by Forest *et al.* were not predictive¹² as evidenced by unacceptable leave-one-out Q^2 statistics. On the other hand, classification-based models give only graded response prediction instead of quantitative values. This made us interested to develop robust and properly validated regression-based models for quantitative predictions of cytotoxicity data to RAW 267.4 cells.

In the current paper, the relationships between physicochemical properties and the cytotoxicity of 25 nanoparticles from 6 metal hydro (oxide) families (CeO_2 , SiO_2 , TiO_2 , AlOOH , ZnO , and Ni(OH)_2) to RAW 264.7 cells (collected from ref. 12) were examined using multiple linear regression (MLR) and partial least squares (PLS) models with lactate dehydrogenase release (ldh_R) used as the endpoint. Nano-QSAR modeling was done using periodic table-based descriptors and physico-chemical descriptors.¹³ The Organization for Economic Co-operation and Development (OECD) guidelines were followed during determination of the quality of the models.¹⁴

2. Methods and materials

2.1 Dataset

In this work, data on toxic potential of 25 NPs (nanoparticles) from 6 metal (hydro) oxides families (CeO_2 , SiO_2 , TiO_2 , AlOOH , ZnO , and Ni(OH)_2) were obtained from laboratory experiments reported in the literature¹² and were expressed in terms of LDH release (ldh_R) from macrophage RAW 264.7 cells. A total of forty descriptors were calculated covering both physicochemical and periodic table based descriptors.¹³ The details of the dataset are presented in the ESI† (Table S1).

2.2 Descriptor calculation

To develop nano-QSAR models, we have gathered information related to MeOx NPs (hydroxides) to encrypt their toxicity towards macrophage RAW 264.7 cells. The complexity and versatile behavior of NPs complicate the nanodescriptor calculation. The nano-QSAR models were developed using descriptors derived from the periodic table and some physicochemical descriptors derived from the experimental procedure as already reported in a previous study.¹² Periodic table descriptors were calculated from the Elemental Descriptor Calculator software developed in our laboratory,¹³ while some additional information of the metals was added directly from the periodic table as descriptors. The selection of periodic table descriptors was advantageous because of their ease of calculation.

2.3 Data set division

Division of a dataset is crucial for the development of statistically significant and predictive models. For the development of MLR models, the toxicity dataset was divided into a training set (containing 18 compounds) and a test set (containing 7 compounds) with the help of a modified

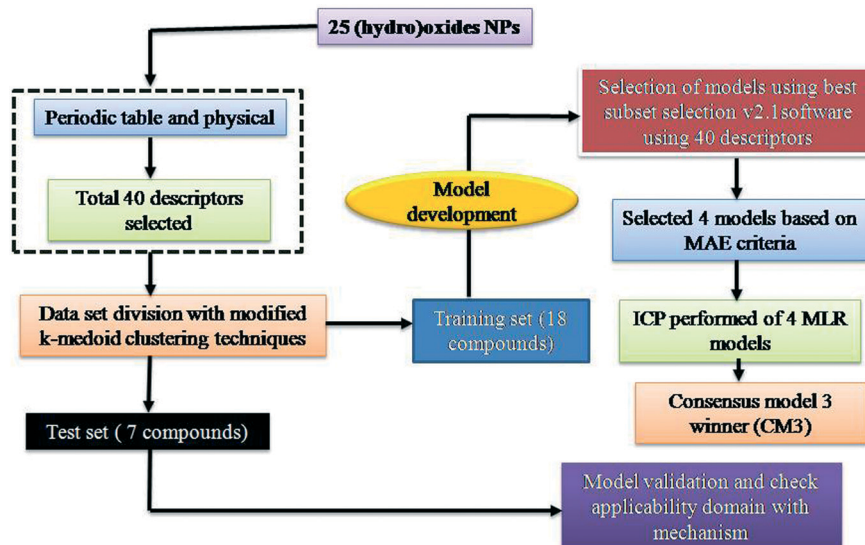


Fig. 1 Flowchart of the work.

k-medoids clustering technique keeping the highest and lowest toxic data of NPs in the training set. The training set was utilized for model development whereas the test set was used to assess the predictivity of the MLR models. Note that the partial least squares (PLS) model was also obtained with the Small Dataset Modeler considering the whole dataset without any dataset division.¹⁵

2.4 Model development

Managing small datasets due to non-availability of experimental data for a particular endpoint is a pivotal work

for QSAR researchers. As the existing dataset contains limited data, we have applied Small Dataset Modeler version 1.0.0 on the whole dataset. This software performs dataset curation, “exhaustive” double cross-validation and consensus predictions as optimal model selection techniques. It also involves the DCV (double cross-validation) (as shown in Fig. 2) method for model development for the small dataset without the need of dividing the data into training and test sets. The inner loop used as the modeling set is constantly split into calibration and validation sets containing n and $n-r$ compounds respectively for deriving all possible “ k ” combinations. Based on the defined calibration and

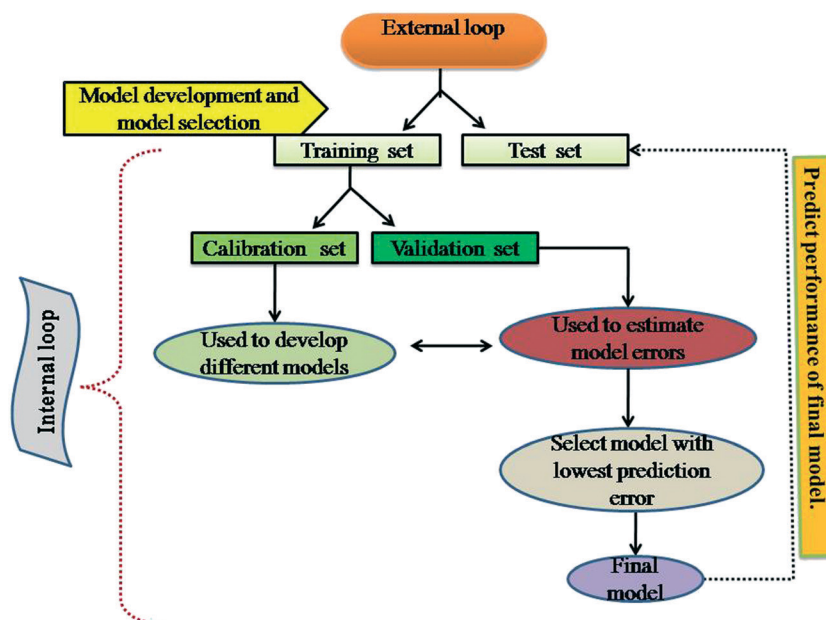


Fig. 2 Double cross-validation methodology.

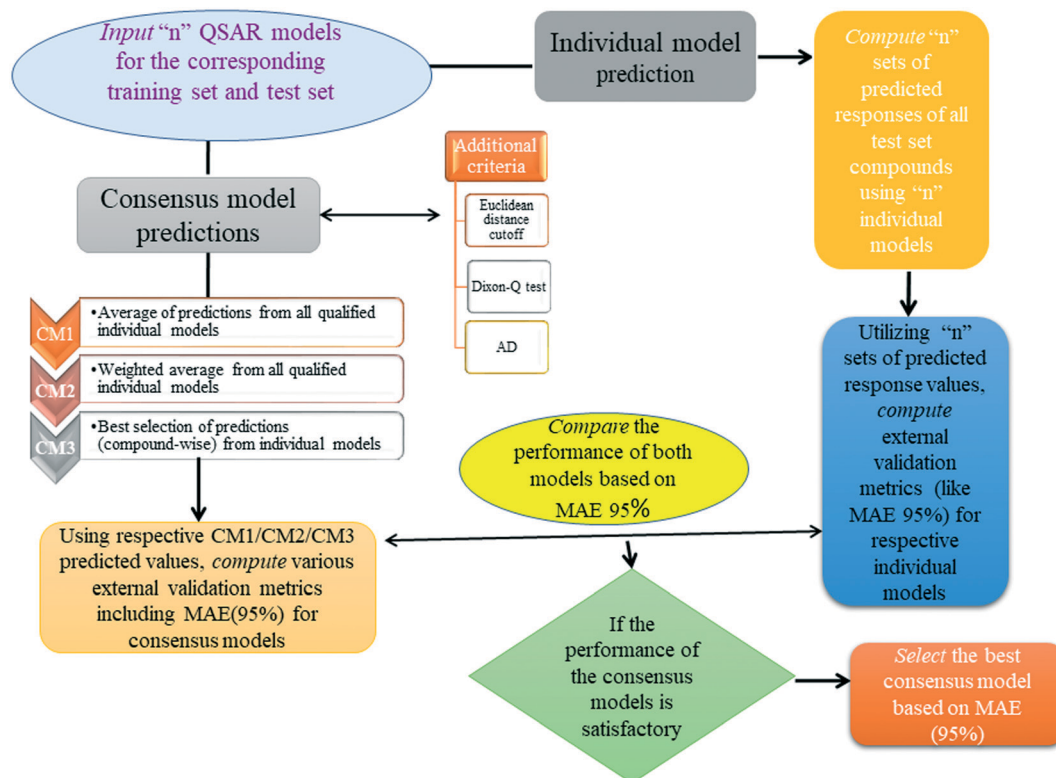


Fig. 3 Algorithm of the ICP tool.

validation sets, the tool allows the management of validation set (r) compounds. The calibration set uses the GA-MLR (genetic algorithm-MLR) technique to develop MLR models whereas the quality of the developed MLR models is predicted with the validation set. All validation metrics (both internal and external metrics) were calculated with the help of exhaustive double cross-validation techniques for the respective models.

Further, we tried dividing the dataset using the modified k -medoids clustering technique. Best Subset Selection v2.1 software was used for model development. The training set (with forty descriptors) was used for development of models from which four best MLR models were selected based on the mean absolute error (MAE) criteria.¹⁶ Further, ICP (intelligent consensus prediction) was performed with the four MLR models to enhance the predictability of the models¹⁷ through intelligent selection of multiple models (Fig. 3). The whole work flow is well depicted in Fig. 1.

2.4.1 Double cross validation (DCV). Double cross validation (DCV) consists of two nested loops named as internal and external loops. The dataset in the external cross validation loop is split into a training set and a test set. The training set in the internal loop is used for model selection and development; on the other hand, the test set is used for assessment of the developed models. The training set (in the internal loop) is repeatedly divided into calibration and validation sets. The calibration data set is utilized to generate various models whereas the validation set is used to

determine the error of the models. Lastly, the model having the least prediction error (in the internal loop) is chosen. The test dataset in the outer loop is then employed to assess the predictability of the selected model in the inner loop. This method of splitting the training set into calibration and validation obviates the bias caused due to usage of a single training set of selected composition.

2.4.2 ICP (intelligent consensus predictions)¹⁷. This software evaluates the performance of the consensus models and relates them with the individual MLR model derived prediction quality based on the MAE criteria (95%). It is assumed that an individual model might not predict all individual test set compounds with the same accuracy; in other words, one QSAR model may be best for prediction of one test set compound while other QSAR models may be best for prediction for some other test set compounds. So, here in our current study, we have selected four models (M1–M4) and performed ICP to explore the quality of predictions of test set compounds being enhanced through an “intelligent” selection of multiple MLR models. Four different ways of consensus prediction are employed they are:

- Consensus model 0 (CM0): Ordinary consensus.
- Consensus model 1 (CM1): Average of predictions from all qualified individual models.
- Consensus model 2 (CM2): Weighted average predictions (WAP) from all qualified individual models.
- Consensus model 3 (CM3): Best selection of predictions (compound-wise) from individual models.

Table 1 Statistical results of MLR models

Statistical parameters	Training set					Test set						
	R^2	R_{adj}^2	Q_{LOO}^2	$\Delta r_{m(LOO)}^2$	$r_{m(LOO)}^2$	Q_{F1}^2	Q_{F2}^2	$\Delta r_{m(LOO)}^2$	$r_{m(LOO)}^2$	CCC	MAE (100%)	MAE based criteria
M1	0.749	0.695	0.611	0.166	0.665	0.722	0.682	0.130	0.632	0.855	0.235	Good
M2	0.750	0.696	0.612	0.161	0.666	0.720	0.679	0.133	0.628	0.853	0.233	Good
M3	0.750	0.697	0.624	0.161	0.667	0.712	0.670	0.140	0.615	0.847	0.249	Good
M4	0.747	0.693	0.620	0.169	0.661	0.707	0.665	0.147	0.596	0.842	0.241	Good
CM0	—	—	—	—	—	0.717	0.675	0.137	0.618	0.850	0.236	Good
CM1	—	—	—	—	—	0.714	0.672	0.140	0.611	0.848	0.242	Good
CM2	—	—	—	—	—	0.714	0.672	0.140	0.610	0.848	0.240	Good
CM3 ^a	—	—	—	—	—	0.711	0.669	0.143	0.606	0.845	0.228	Good

^a Winner model.

Consensus models are more reliable due to the following reasons:

1. Consensus predictions are precise and reliable as they do not rely on a single model but on multiple models and the predictions are not biased.
2. Consensus predictions may consider the WAPs from the individual models along with compound wise prediction of the individual models thus selecting the model that is best for a particular test compound.
3. They also check prediction taking the AD criteria, Dixon Q-test, and Euclidean distance into consideration. Checking AD before prediction makes the prediction value more acceptable.
4. For limited data points, we can rely more on consensus models than individual models as it enhances the applicability domain.

2.4.3 Applicability domain¹⁸. The applicability domain (AD) is the physico-chemical, structural or biological space of a QSAR model surrounded by model descriptors and modeled response based on which training sets have been developed.¹⁸ Relying on this knowledge or information, the prediction of new compounds is made. The motive of computing the AD is to state whether the required assumptions of the models are met and for which type of chemical the model is reliably applicable. In the development of a QSAR model, the AD has a significant role for evaluating the unpredictability of the new compounds depending upon how close it is to the compounds used in development of the model. Therefore, the prediction of any response with the help of QSAR is possible if the new predicted compounds fall in the space of the modeled AD. Two methods namely the standardization approach (for MLR models) and DModX, *i.e.*, distance to model in the X-space (for the PLS model) were used to check any outlier(s) in the dataset. Both methods were capable of determining the AD correctly. The standardization approach assumes that the descriptor values follow a normal distribution, but it does not consider the intercorrelation among descriptors nor its contribution to the model unlike the *k*-nearest neighbour approach proposed by Consonni and co-workers some years ago.^{19–21} Therefore, along with the above-mentioned methods, we have also applied the

k-nearest neighbour approach and found that all the test set compounds were inside the AD since the test set reference number matched the reference number of the training set. It was found that no metal (hydro) oxides were outside the region of the AD.

2.5 Model validation

Validation of a QSAR model is crucial for nano-QSAR models. Rigorous assessment of the model quality for the measurement of stability, fitness, robustness and predictivity of the developed models was carried out. Here, all the models were evaluated based on the OECD proposed criteria.¹⁴ In the present work, we have calculated R^2 , R_{adj}^2 , the variance ratio (F), and the standard error of estimate (s) and the leave-one-out squared correlation coefficient (Q_{LOO}^2) for internal model quality and validation.²² We have also computed the leave-many-out squared correlation coefficient ($Q_{LMO(20\%)}^2$) for the final PLS model.²³ Additionally, we have also calculated the r_m^2 metrics²⁴ and mean absolute error (MAE).¹⁶ Further, other external validation metrics (R_{pred}^2 and Q_{F2}^2) were also calculated for the test set in the case of the MLR models.

3. Results and discussion

Linear regression models have been developed for the cytotoxicity data of 25 nano oxides (hydroxides) towards macrophage RAW 267 cells using release of LDH (ldh_R) as the endpoint. The present work was divided into two sections; first, the whole set was subjected to the small data set modeler due to the limited number of experimental data being available, and secondly, the same dataset was further divided into a training set and a test set to check the predictivity of the developed models (Tables 1 and 2). The statistical values ($R^2 = 0.75$, $Q^2 = 0.664$) of the PLS model (N1) developed with the small dataset modeler were satisfactory,

Table 2 PLS model results

Model	R^2	Q^2	$Q_{LMO(20\%)}^2$	$\bar{r}_m^2$	Δr_m^2
N1	0.75	0.662	0.662	0.579	0.077

Box 1 MLR models

M1

$\text{ldh}_R = 1.2782(\pm 0.4431) + 0.00134(\pm 0.00055) \text{ Metals_SumIP} + 0.0322(\pm 0.0066) \text{ SumElectrons_Active_M_SM} + 0.20123(\pm 0.03801) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.749, R_{\text{adj}}^2 = 0.695, Q^2 = 0.611, S = 0.421658, F = 13.92, \text{PRESS} = 3.85635, \text{MAE based criteria} = \text{moderate}$
 $n_{\text{test}} = 7, Q_{F1}^2 = 0.722, Q_{F2}^2 = 0.682, \text{MAE based criteria} = \text{good}$

M2

$\text{ldh}_R = -0.488(\pm 0.6306) - 0.00367(\pm 0.00058) \text{ Metals_SumIP} + 0.0406(\pm 0.0083) \text{ Cat_rad} + 0.197(\pm 0.0375) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.750, R_{\text{adj}}^2 = 0.696, Q^2 = 0.612, S = 0.421001, F = 13.98, \text{PRESS} = 3.84633, \text{MAE based criteria} = \text{moderate}$
 $n_{\text{test}} = 7, Q_{F1}^2 = 0.720, Q_{F2}^2 = 0.679, \text{MAE based criteria} = \text{good}$

M3

$\text{ldh}_R = -0.150(\pm 0.6904) + 0.041596(\pm 0.0079) \text{ Cat_rad} - 0.091(\pm 0.0137) \text{ Oxidation degree of metal(N_oxy_M)} + 0.1834(\pm 0.045) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.750, R_{\text{adj}}^2 = 0.697, Q^2 = 0.624, S = 0.420392, F = 14.03, \text{PRESS} = 3.72620, \text{MAE based criteria} = \text{moderate}$
 $n_{\text{test}} = 7, Q_{F1}^2 = 0.712, Q_{F2}^2 = 0.670, \text{MAE based criteria} = \text{good}$

M4

$\text{ldh}_R = -0.436(\pm 0.6208) - 0.246(\pm 0.5132) \text{ D1_Metals} + 0.0405(\pm 0.0088) \text{ Cat_rad} + 0.0204(\pm 0.0377) \text{ Solubility}$
 $n_{\text{training}} = 18, R^2 = 0.747, R_{\text{adj}}^2 = 0.693, Q^2 = 0.620, S = 0.423426, F = 13.77, \text{MAE based criteria} = \text{moderate}$
 $n_{\text{test}} = 7, Q_{F1}^2 = 0.707, Q_{F2}^2 = 0.665, \text{MAE based criteria} = \text{good}$

PLS model (N1)

$\text{ldh}_R = 1.44 + 0.19 \times \text{Solubility} + 0.01 \times \text{Metals_SumIP} + 0.029 \times \text{SumElectrons_Active_M_SM}$
 $R^2 = 0.75, Q^2 = 0.664, \bar{r}_m^2 = 0.579, \Delta r_m^2 = 0.077$

and it was also noted that a few descriptors obtained from the MLR models (M1, M2, M3, M4; Box 1) were common to the PLS (N1) model thus justifying the importance of such descriptors. The MLR models (M1, M2, M3, M4) were obtained through the best subset selection method and the models were subjected to the intelligent consensus prediction technique. The descriptors obtained from both the PLS model (N1) and MLR models (M1, M2, M3, M4) were elaborately discussed (Table 3) below with their possible mechanisms. The scatter plots of MLR models are also shown in Fig. 4. The descriptors obtained from the developed models (MLR and PLS) were not highly correlated as can be checked from the cross-correlation matrix.²⁵ Besides these, we have presented additional statistical results (normal distribution of residuals, homoscedasticity, no autocorrelation of the residuals, and non-multicollinearity among the descriptors) in the ESI.†

The AD was checked using the standardization approach for MLR models and DModX, *i.e.*, distance to model in the X-space with SIMCA 16.0.2 software for the PLS model to calculate the AD. At 95% confidence level ($D\text{-crit} = 0.009998$), the AD of the PLS model was checked (Fig. 5) and it was

found that no metal (hydro) oxides were outside the region of the AD.

The variable importance plot of the PLS model highlights the influential descriptors contributing to the cytotoxicity of MeOx NPs to RAW 267.2 cells. The descriptors of the N1 model appearing as per the VIP scores are solubility, Metals_SumIP, and SumElectrons_Active_M_SM. The descriptors from higher to lower contributions are shown in Fig. 6.

3.1 Loading plot of the PLS model

The loading plot gives the relationship between the X-variables (descriptors) and the Y-variable (LDH) of the PLS model as shown in Fig. 7. The development of the loading plot was done using two PLS components (first and second components). This plot gives insights into how different variables influence the model. During the evaluation, the distance from the origin was taken into consideration in the loading plot. The descriptors located far away from the origin in the plot are considered to have a greater influence on the model and on the Y response. The solubility descriptor is situated furthest from the origin, and hence it has more

Table 3 Description and contribution of the descriptors obtained from MLR and PLS models

Sl. no.	Descriptor	Description	Contribution
1	Metals_SumIP	Sum of ionization potential energy (kJ mol^{-1}) of metals	+ve
2	D1_Metals	Total number of metals	+ve
3	Cat_rad	Distance between the nucleus and the electron in the outermost shell of an ion	+ve
4	SumElectrons_Active_M_SM	Summation of electrons of active metals and semi-metals	+ve
5	N_oxy_M	Oxidation degree or state of the metal	−ve
6	Solubility	The ability of a solid, liquid, or gaseous chemical substance (referred to as the solute) to be dissolved in solvent (usually a liquid) and form a solution	−ve

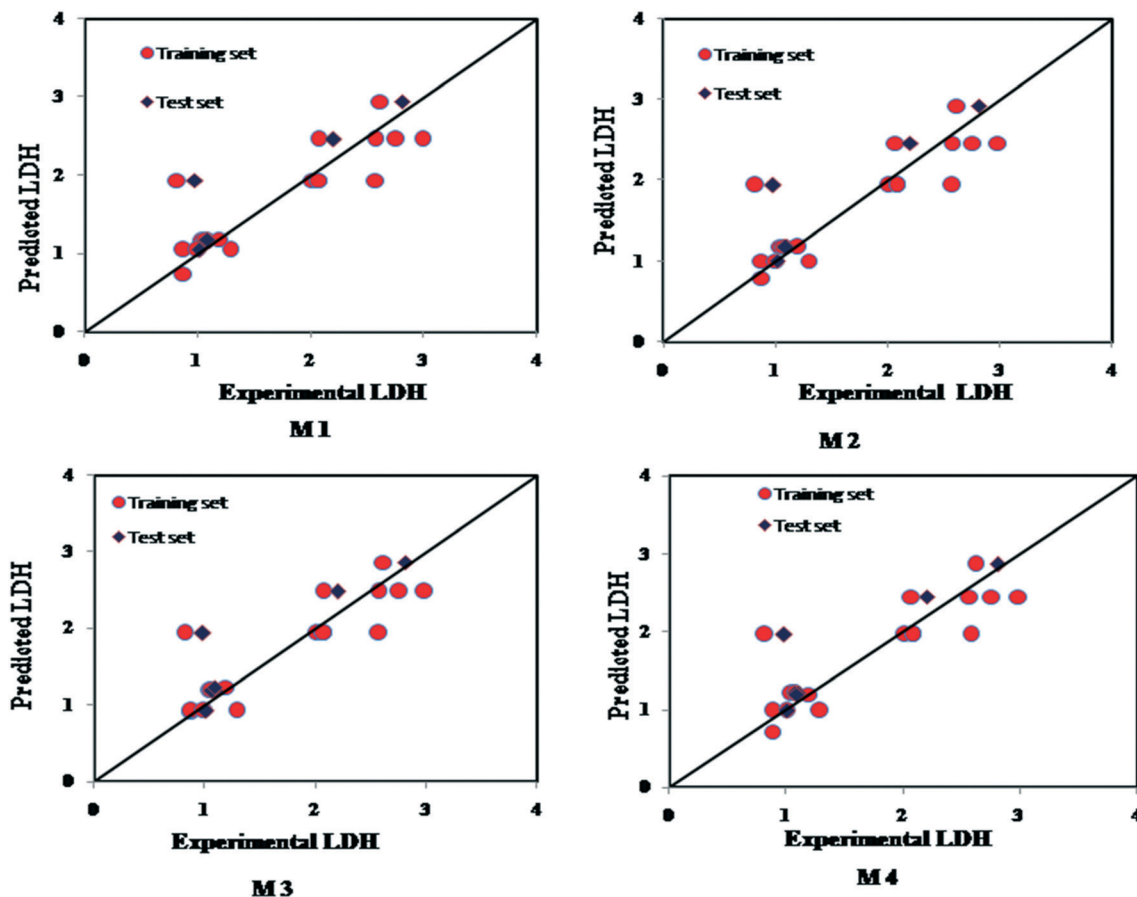


Fig. 4 Scatter plots of MLR models.

impact with respect to the other two descriptors. It can also be confirmed from the VIP plot.

3.2 Model randomization

Randomization of a model authenticates the robustness of the model. This signifies that the developed model is not obtained by chance correlation. The randomized model is developed by shuffling different combinations of X or Y variables leading to multiple models generated. Here, we

have used 100 permutations, which can be randomized depending on the user's choice. The random models should have poor statistics to prove that the nonrandom model is not formed by chance correlation. We have provided the correlation between original Y -vector and permuted Y -vector *versus* cumulative R_y^2 and cumulative Q_y^2 plots as represented in Fig. 8. The plot shows that the model is nonrandom as well as robust since $R_y^2 = 0.025$ and $Q_y^2 = -0.258$ and one can easily predict the cytotoxicity of the MeOx NPs from the developed PLS model.

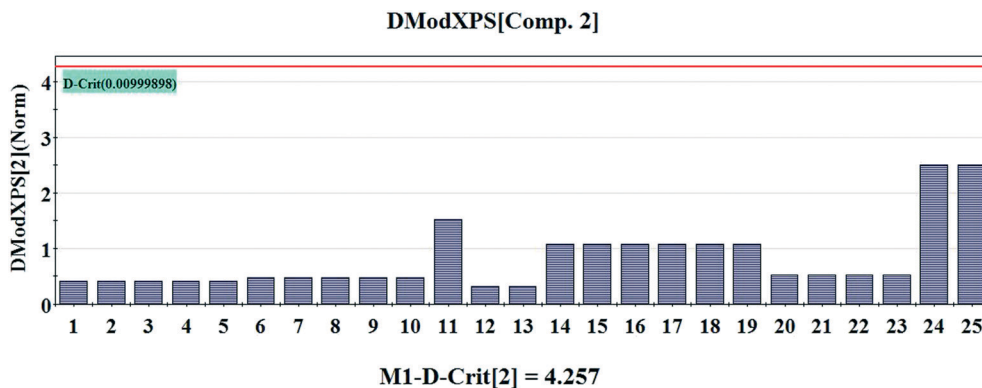


Fig. 5 Applicability domain of model N1 (PLS model).

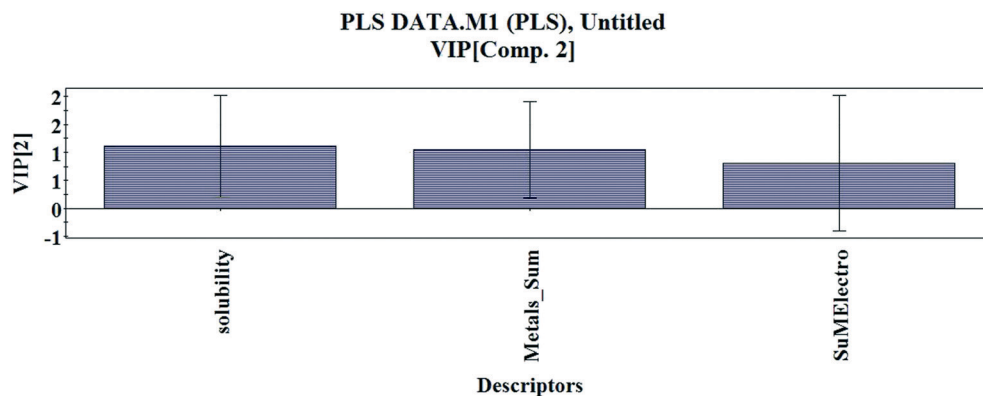


Fig. 6 Variable importance plot (VIP) of the N1 model.

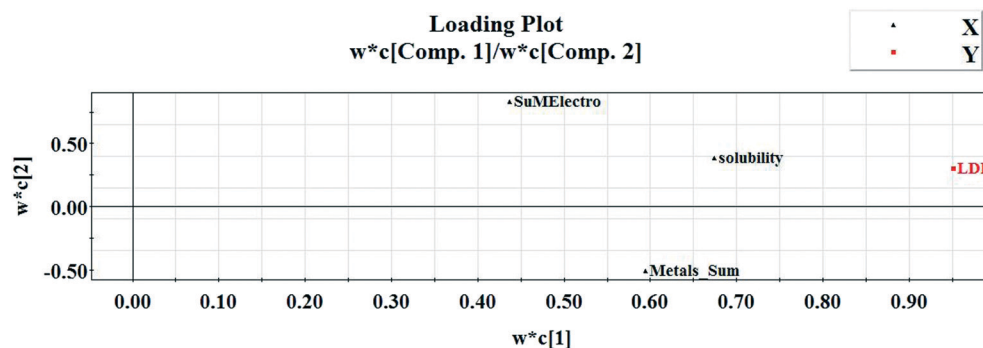


Fig. 7 Loading plot of the PLS model (N1).

3.3 Description of the descriptors

The Metals_SumIP descriptor stands for the sum of ionization potential energy (kJ mol^{-1}) of metals. The ionization potential energy is the quantity of energy that an isolated gaseous atom in the ground electronic state must absorb to discharge an electron resulting in a cation. The positive contribution of the descriptor suggests that with an increase of the ionization potential, the toxicity also increases as found in the case of nanoparticles 12 (ZnO-1) and 6 ($\text{Ni}(\text{OH})_2$ -1) whereas the reverse happens for nanoparticles 24

(SiO_2 -1) and 3 (AlOOH -3) where with a decrease in ionization potential the toxicity also decreases. Ionization potential is related to electronegativity and electron affinity. If a molecule has more ionization potential that means that it is difficult to remove the valence electrons easily to form a cation which suggests that the molecule is more electronegative. The increase of electronegativity of the chemicals may increase the production of reactive oxygen species (ROS).

D1_Metals stands for the total number of metals. The presence of metals in the nanoparticles tends to increase the release of LDH from the cells as indicated by nanoparticles 8

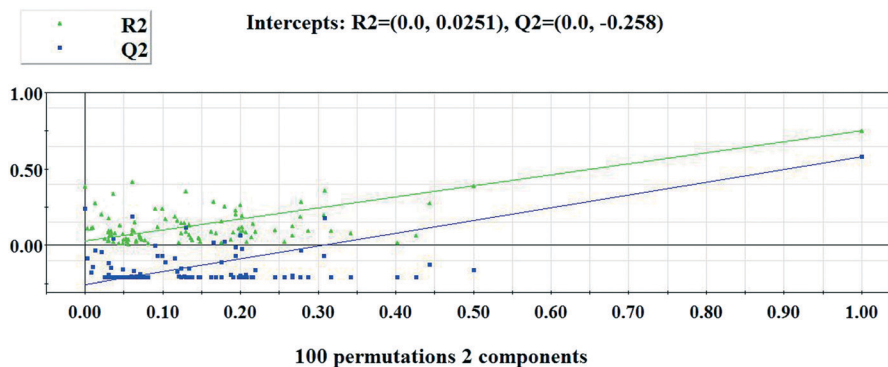


Fig. 8 Randomization plot of the PLS model (N1).

(Ni(OH)₂-3), 12(ZnO-1) while nanoparticle 24 (SiO₂-1) shows a low LDH release from the RAW 264.7 cell line for being non-metallic in nature. This is correctly indicated by the positive contribution of the descriptor. Metal-based NPs can show tremendous toxicity and can pose detrimental biological interaction to mammalian cells.²⁶ Metal oxide nanoparticles with a larger surface area can penetrate deeply into the body due to their smaller size and possess surface charge which causes greater inflammatory response.²⁷ Studies revealed that different nanoparticles modulate inflammatory responses of macrophages in different ways.

The periodic table descriptor, Cat_rad, is the distance between the nucleus and the electron present in the outermost shell of an ion. The positive contribution of the descriptor suggests that with an increase of the cationic radius, the LDH release from the cell increases as shown in nanoparticles 14 (CeO₂-1) and 8 (Ni (OH)₂-3). In contrast, a decrease in the cationic radius leads to less LDH release as indicated by nanoparticles 3 (AlOOH-3) and 24 (SiO₂-1). The addition of electrons increases the size of an atom through repelling of the electrons within the shells. When the cationic radius increases, it becomes easier for the metal to lose electrons to form a cation. This happens because as the size of the atom increases, the outermost electrons feel a decrease of shielding by the inner-shell electron and as a result electrons present in the outer shell do not “feel” the entire positive charge of the nucleus. Consequently, metal cations released from the metal oxides (MeOx) may catalyze the formation of hydroxyl radicals to produce ROS *via* the so-called Haber–Weiss–Fenton cycle^{28,29} leading to apoptosis to the cell through LDH leakage.

SumElectrons_Active_M_SM stands for the summation of electrons of active metals and semi-metals. This descriptor positively contributes towards cell toxicity. The presence of electrons can also be correlated to higher electronegativity of that atom. The increase in electronegativity is responsible for the catalytic properties of the metal cationic form as well as for the cytotoxicity to RAW 264.7 cells. This descriptor positively contributes as seen in nanoparticles 15 (CeO₂-2) and 8 (Ni(OH)₂-3) where an increase of electronegativity increases the release of LDH from the cell, but on the other hand a decrease in electronegativity decreases the cytotoxicity and release of LDH from the cell as evidenced by nanoparticles 23 (TiO₂-4) and 3 (AlOOH-3). The presence of electronegativity is responsible for initiating cation formation in metal oxides.³⁰

N_ox_M stands for the oxidation degree or state of the metal. This descriptor is negatively correlated with the cytotoxicity of the cell which means that an increase of the oxidation degree decreases the cytotoxicity to the cell as evidenced by nanoparticles 19 (CeO₂-6) and 23 (TiO₂-4). On the other hand, the lower degree of oxidation enhances the cytotoxicity effect towards the cell and promotes the release of LDH as seen in nanoparticles 8 (Ni(OH)₂-3) and 12 (ZnO-1). The lower degree of oxidation makes the metal unstable due to the unfilled valence shell and therefore the metal tries

to gain a higher oxidation state or a stable one. A transition metal has a different oxidation state due to partially filled-d and -f orbital shells.³¹ When a transition metal tends to have a lower oxidation state, it means that the metal easily donates its valence electrons or can gain electrons to form metal cations or enrich itself with electrons respectively but this is not so easily done if the oxidation state of the metal is increased; thus, a lower oxidation state of the metal is detrimental to the cell.

Another descriptor, Solubility stands for the ability of a solid, liquid, or gaseous chemical substance (referred to as the solute) to be dissolved in a solvent (usually a liquid) and form a solution. The solubility depends on three factors such as temperature, pressure as and the solvent used. Solubility is one of the pivotal parameters that is involved in nanomaterials-induced toxicity.³² The positive contribution of the descriptor suggests that an increase in solubility increases the cytotoxicity as in nanoparticles 12 (ZnO-1) and 8 (Ni (OH)₂-3), and the opposite happens for nanoparticles 11 (NiO) and 19 (CeO₂-6) with decreased solubility. The NP toxicity is related to its solubility, the dissolution of NPs under acidic conditions and as well as the presence of lipids and proteins (amino acids) which can induce toxicity by binding with them and interfering with the cell functions. Few studies of Xia *et al.* and Li *et al.*^{33,34} confirmed that solubility of NPs has a strong relation with cellular toxicity.

The descriptors selected in our models are in conformity with the descriptors reported in the study of Forest *et al.*¹² such as Cat_rad and N_ox_M which show a high impact in the determination of the toxicity of MeOx NPs. The selection of appropriate descriptors is also necessary for a robust model. The descriptors used by Forest *et al.*¹² were physicochemical and dimensional descriptors whereas the present work not only included the physicochemical descriptors but also the periodic table descriptors. The advantage of periodic table descriptors is that they are not specific family dependent and show a good impact towards the determination of cytotoxicity.

3.4 Mechanism of toxicity

The wide applications of metallic NPs with their unique properties and structural alteration have led to the modification of biological functions along with potential ROS production as well as the change of cellular signaling linked with cell death, proliferation, and differentiation. Reactive oxygen species (ROS) are small, short lived and reactive molecules that are continuously produced and destroyed by a wide variety of antioxidant defense mechanisms in the human body. The generation of ROS by the MeOx NPs has been considered as the major contributor and originator of cytotoxicity. Excessive production of ROS by MeOx NPs causes oxidative damage, by destroying the cell function through acting on the biomolecules (lipids, proteins and DNA) hampering their functionality, ultimately causing death of the cell. The biological toxicity of MeOx NPs occurs when

they interact with solubilized metallic ions. The derived information from the various descriptors gave us insights that among the various features of the nanoparticles, electronegativity is the most prevailing feature responsible for the cytotoxicity. Due to the higher electronegativity, the dipole charge distribution takes place between the atoms in the MeOx NPs. The surface charged NPs then cause electrostatic interaction with the cellular membrane (CM) of the cell which enhances the membrane-wrapping phenomena. This may also increase the surface tension and facilitate the entry of NPs through generation of pores. Hydrophobic interaction is also possible with oxides of NPs. The cationic form of MeOx NPs has a more disruptive influence on the membrane bilayer, and through membrane phase change they enter into the cell.³⁵ MeOx NPs can interfere with the electron transport chain and depolarize the mitochondrial membrane through activation of NADPH-related enzymes (nicotinamide adenine dinucleotide phosphate). The interaction of NPs blocks the electron transfer chain, thus elevating the cellular $O_2^{\cdot -}$ (oxygen radical) level through electron transfer from the respiratory carriers to O_2 .³⁶ The oxidative stress generated further affects the DNA by formation of 8-hydroxyl-20-deoxyguanosine (8-OHdG) that ultimately leads to DNA damage, arrest of the cell cycle causing alteration of the apoptosis mechanism and damage to the cellular membrane which promotes the leakage of LDH from the membrane.³⁷ Inflammatory responses by ROS

are also enhanced by metallic NPs like TiO_2 NPs and SiO_2 NPs.³⁸ ZnO NPs cause cytotoxicity by triggering oxidative injury by generating ROS and releasing inflammatory mediators that finally lead to cell death in phagocytic RAW 264.7 cells.

3.5 Comparison with previously published models

The present work was performed using periodic table descriptors in addition to a few physicochemical descriptors, as the periodic table-based descriptors were easy to calculate without involving much complicated calculations. One of the problems with the dimensional descriptors as used by Forest *et al.*¹² is that they are family dependent whereas the other descriptors are not. Therefore, it might be difficult to obtain predictive results with dimensional descriptors for new nanoparticles.

Forest *et al.*¹² considered physicochemical and dimensionality related descriptors for the development of PLS regression models first. The developed PLS models were no way statistically acceptable as they show leave-one-out Q^2 values of 0.0 and -0.3. To be qualified for statistical acceptability and predictability, all QSAR models should satisfy some requirements like statistical quality metrics, internal validation metrics and external validation metrics, *etc.* as per the OECD guidelines (<https://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm>).

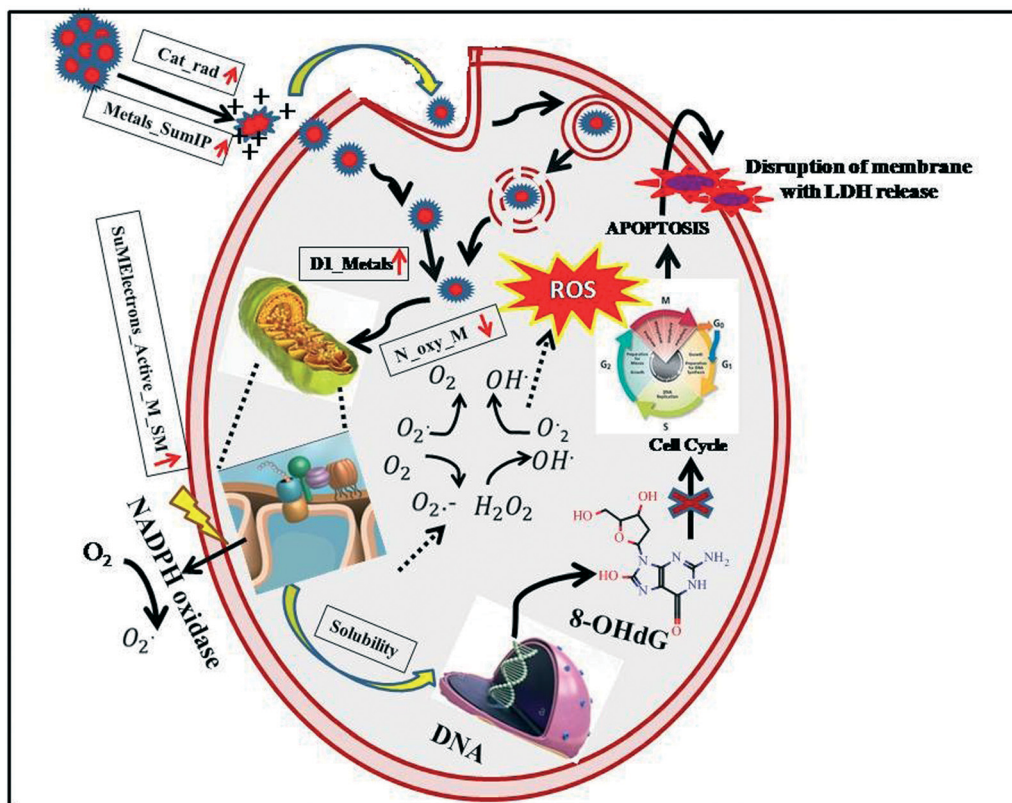


Fig. 9 Mechanism of toxicity.

A check of simple statistical measures suggests that previously reported PLS models are not presentable as they do not have any predictive quality. This was also acknowledged in the paper of Forest *et al.*¹² The leave-one-out Q^2 values of their models are no way acceptable (0.0 and -0.3) while the statistical metric values of our regression models are indeed acceptable as per the OECD guidelines. In addition, the previous paper also does not report external validation, which is considered as the gold standard of validation of QSAR models.²⁰

We would also like to add that Forest *et al.* performed leave-one-out cross-validation, but not double cross-validation. To know more about the latter technique one can refer to ref. 15.

However, Forest *et al.* then developed classification based QSAR models which were indeed acceptable. Classification based QSAR models give graded or semi-quantitative predictions while regression-based models give exact quantitative predictions. In our current work, we have developed regression-based models for quantitative predictions. Obviously these two categories of models are not comparable. Thus, we will not compare our models with classification-based predictions which are only semi-quantitative and give graded values. We have also performed external validation of our models and finally applied consensus predictions which are considered good practices in QSAR.²³ We have also reported a model for the whole undivided data set with leave-many-out cross validation in addition to models obtained from data set splitting.

We have also proposed the mechanism of toxicity for better understanding.

4. Conclusion

Nanotechnology has potential applications in various fields. Recent advances in NP engineering are believed to have potential impact in diverse areas in the near future. Production of metal-based NPs for commercial use and other purposes has increased exponentially, and as a result, assessment for their safety has also gained attention. Therefore, in this work, we have made efforts to develop interpretative nano-QSAR models that are reliable and capable of predicting not only the investigated NPs but also unexplored NPs if present within the applicability domain. As we know working with small datasets is a critical job, here we have applied the Small Dataset Modeler version 1.0.0 to develop a PLS model without applying the conventional splitting approach. Further, MLR models were also generated through best subset selection, and then ICP was applied to obtain reliable predictions. After performing ICP, CM3 was found to be the winner model. The insights obtained from the various descriptors from both PLS and MLR models pointed out that electronegativity, solubility, formation of a metal cation and an increase in the number of metal NPs increase the cytotoxicity whereas non-metal NPs and an increase of the oxidation degree contribute to low toxicity to

RAW 267.4 cells. A short-proposed mechanism has also been explained in Fig. 9 to determine the cause of cellular toxicity by the MeOx NPs. We hope that the present work would be useful in establishing the cause of toxicity of unexplored NPs within the applicability domain, and the information obtained about the features of MeOx NPs would help in designing new NPs entities with reduced toxicity to living cells. Considering the limited experimental data available for toxicity of metal oxide and hydroxide nanoparticles, the developed models might be used for prediction of toxicity of untested nanoparticles that may help in prioritizing further experiments.

Conflicts of interest

None.

Acknowledgements

Financial assistance from Indian Council of Medical Research, New Delhi, in the form of a Senior Research Fellowship to JR is thankfully acknowledged. Financial assistance from CSIR, New Delhi to KR under a Major Research project (CSIR Project No. 01(2895)/17/EMR-II) is thankfully acknowledged.

References

- 1 R. P. Feynman, Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics, *Handbook of Nanoscience, Engineering and Technology*, 2003.
- 2 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y. P. Liao and M. Wang, Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation, *ACS Nano*, 2012, **6**(5), 4349–4368.
- 3 J. Parada, O. Rubilar and M. A. Fernández-Baldo, The nanotechnology among US: are metal and metal oxides nanoparticles a nano or mega risk for soil microbial communities?, *Crit. Rev. Biotechnol.*, 2018, **39**(2), 157–172.
- 4 M. Amde, J. F. Liu, Z. Q. Tan and D. Bekana, Transformation and bioavailability of metal oxide nanoparticles in aquatic and terrestrial environments. A review, *Environ. Pollut.*, 2017, **230**, 250–267.
- 5 C. S. Yah, S. E. Iyuke and G. S. Simate, A review of nanoparticles toxicity and their routes of exposures, *Iran. J. Pharm. Res.*, 2012, **8**(1), 299–314.
- 6 J. F. Burd and M. Usategui-Gomez, A colorimetric assay for serum lactate dehydrogenase, *Clin. Chim. Acta*, 1973, **46**, 223–227.
- 7 F. Petrat, T. Bramey, M. Kirsch and H. De Groot, Initiation of a superoxide-dependent chain oxidation of lactate dehydrogenase-bound NADH by oxidants of low and high reactivity, *Free Radical Res.*, 2005, **39**, 1043–1057.
- 8 M. T. Cronin, Quantitative structure–activity relationships (QSARs)–applications and methodology. In Recent Advances in QSAR Studies, in *Recent Advances in QSAR Studies*, Springer, Dordrecht, 2010, pp. 3–11.

- 9 T. Puzyn, J. Leszczynski and M. T. D. Cronin, *Recent Advances in QSAR Studies: Methods and Applications*, Springer Science+Business Media BV, 2010.
- 10 T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynska and J. Leszczynski, Using nano-QSAR to predict the cytotoxicity of metaloxide nanoparticles, *Nat. Nanotechnol.*, 2011, **6**(3), 175–178.
- 11 A. Gajewicz, B. Rasulev, T. C. Dinadayalane, P. Urbaszek, T. Puzyn, D. Leszczynska and J. Leszczynski, Advancing risk assessment of engineered nanomaterials: application of computational approaches, *Adv. Drug Delivery Rev.*, 2012, **64**, 1663–1693.
- 12 V. Forest, J. F. Hocheplied, L. Leclerc, A. Trouvé, K. Abdelkebir, G. Sarry, V. Augusto and J. Pourchez, Towards an alternative to nano-QSAR for nanoparticle toxicity ranking in case of small datasets, *J. Nanopart. Res.*, 2019, **21**(5), 1–14.
- 13 http://teqip.jdvu.ac.in/QSAR_Tools.
- 14 OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69. GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS, 2007, accessed from [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=en v / j m / mono\(2007\)2&doclanguage=en](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=en v / j m / mono(2007)2&doclanguage=en), (13.02.13).
- 15 P. Ambure, A. Gajewicz, M. Natália, D. S. Cordeiro and K. Roy, A New Workflow for QSAR Model Development from Small Data Sets: Integration of Data Curation, Exhaustive Double Cross-Validation and A Set of Optimal Model Selection Techniques, *J. Chem. Inf. Model.*, 2019, **59**(10), 4070–4076.
- 16 K. Roy, R. N. Das, P. Ambure and R. B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, *Chemom. Intell. Lab. Syst.*, 2016, **152**, 18–33.
- 17 K. Roy, P. Ambure, S. Kar and P. K. Ojha, Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models?, *J. Chemom.*, 2018, **32**(4), 2992.
- 18 D. Gadaleta, G. F. Mangiatordi, M. Catto, A. Carotti and O. Nicolotti, Applicability domain for QSAR models: where theory meets reality, *Int. J. Quant. Struct.-Prop. Relat.*, 2016, **1**, 45–63.
- 19 F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni and R. Todeschini, Comparison of different approaches to define the applicability domain of QSAR models, *Molecules*, 2012, **17**(5), 4791–4810.
- 20 F. Sahigara, D. Ballabio, R. Todeschini and V. Consonni, Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions, *Aust. J. Chem.*, 2013, **5**(1), 1–9.
- 21 D. Gadaleta, G. F. Mangiatordi, M. Catto, A. Carotti and O. Nicolotti, Applicability domain for QSAR models: where theory meets reality, *Int. J. Quant. Struct.-Prop. Relat.*, 2016, **1**(1), 45–63.
- 22 N. Chirico and P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, *J. Chem. Inf. Model.*, 2011, **51**(9), 2320–2335.
- 23 K. Roy, S. Kar and R. N. Das, Statistical methods in QSAR/QSPR, in *A primer on QSAR/QSPR modeling*, Springer, Cham, 2015, pp. 37–59.
- 24 K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das and H. Kabir, Comparative studies on some metrics for external validation of QSPR models, *J. Chem. Inf. Model.*, 2012, **52**, 396–408.
- 25 A. Morales Helguera, M. Perez Gonzalez, M. N. Dias Soeiro Cordeiro and M. A. Cabrera Perez, Quantitative Structure–Carcinogenicity Relationship for Detecting Structural Alerts in Nitroso Compounds: Species, Rat; Sex, Female; Route of Administration, Gavage, *Chem. Res. Toxicol.*, 2008, **21**(3), 633–642.
- 26 H. A. Jeng and J. Swanson, Toxicity of metal oxide nanoparticles in mammalian cells, *J. Environ. Sci. Health, Part A: Toxic/Hazard. Subst. Environ. Eng.*, 2006, **41**(12), 2699–2711.
- 27 A. Nemmar, P. M. Hoet, B. Vanquickenborne, D. Dinsdale, M. Thomeer, M. F. Hoylaerts and B. Nemery, Passage of inhaled particles into the blood circulation in humans, *Circulation*, 2002, **105**(4), 411–414.
- 28 J. Roy, P. K. Ojha and K. Roy, Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table-based descriptors, *Nanotoxicology*, 2019, **13**(5), 701–716.
- 29 W. H. Koppenol, The Haber-Weiss cycle–70 years later, *Redox Rep.*, 2001, **6**(4), 229–234.
- 30 T. Ozben, Oxidative Stress and Apoptosis: impact on Cancer Therapy, *J. Pharm. Sci.*, 2007, **96**(9), 2181–2196.
- 31 J. D. Walker, M. Enache and J. C. Dearden, Quantitative Cationic-Activity Relationships for Predicting Toxicity of Metals, *Environ. Toxicol. Chem.*, 2003, **22**(8), 1916–1935.
- 32 A. M. Schrand, M. F. Rahman, S. M. Hussain, J. J. Schlager, D. A. Smith and A. F. Syed, Metal-based nanoparticles and their toxicity assessment, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2010, **2**(5), 544–568.
- 33 T. Xia, M. Kovochich, M. Liong, L. Madler, B. Gilbert, H. Shi, J. I. Yeh, J. I. Zink and A. E. Nel, Comparison of the mechanism of toxicity of zinc oxide and cerium oxide nanoparticles based on dissolution and oxidative stress properties, *ACS Nano*, 2008, **2**, 2121–2134.
- 34 K. G. Li, J. T. Chen, S. S. Bai, X. Wen, S. Y. Song, Q. Yu, J. Li and Y. Q. Wang, Intracellular oxidative stress and cadmium ions release induce cytotoxicity of unmodified cadmium sulfide quantum dots, *Toxicol. In Vitro*, 2009, **23**(6), 1007–1013.
- 35 S. Behzadi, V. Serpooshan, W. Tao, M. A. Hamaly, M. Y. Alkawareek, E. C. Dreaden, D. Brown, A. M. Alkilany, O. C. Farokhzad and M. Mahmoudi, Cellular uptake of nanoparticles: journey inside the cell, *Chem. Soc. Rev.*, 2017, **46**(14), 4218–4244.
- 36 S. J. Soenen, P. Rivera-Gil, J.-M. Montenegro, W. J. ParakDe, S. C. Smedt and K. Braeckmans, Cellular toxicity of inorganic nanoparticles: Common aspects and guidelines

- for improved nanotoxicity evaluation, *Nano Today*, 2011, **6**, 446–465.
- 37 T. Xia, M. Kovichich, J. Brant, M. Hotze, J. Sempf, T. Oberley, C. Sioutas, J. I. Yeh, M. R. Wiesner and A. E. Nel, Comparison of the abilities of ambient and manufactured nanoparticles to induce cellular toxicity according to an oxidative stress paradigm, *Nano Lett.*, 2006, **6**, 1794–1807.
- 38 E.-J. Park, J. Yoon, K. Choi, J. Yi and K. Park, Induction of chronic inflammation in mice treated with titanium dioxide nanoparticles by intratracheal instillation, *Toxicology*, 2009, **260**, 37–46.

Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides

Joyita Roy & Kunal Roy

To cite this article: Joyita Roy & Kunal Roy (2022) Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeOxNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides, *Nanotoxicology*, 16:2, 152-164, DOI: [10.1080/17435390.2022.2038299](https://doi.org/10.1080/17435390.2022.2038299)

To link to this article: <https://doi.org/10.1080/17435390.2022.2038299>

 View supplementary material 

 Published online: 15 Feb 2022.

 Submit your article to this journal 

 Article views: 188

 View related articles 

 View Crossmark data 

 Citing articles: 5 View citing articles 

ARTICLE



Modeling and mechanistic understanding of cytotoxicity of metal oxide nanoparticles (MeO_xNPs) to *Escherichia coli*: categorization and data gap filling for untested metal oxides

Joyita Roy and Kunal Roy 

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

ABSTRACT

Metal oxide nanoparticles (MeO_xNPs) production is expected to increase every year exponentially, and their potential to cause adverse effect to the environment and human health will also expand rapidly. Hence, risk assessment of nanoparticles (NPs) is necessary to design ecosafe products. However, experimental ecotoxicological assessments are time-consuming requiring a lot of resources. Therefore, researchers rely on alternative *in silico* approaches to predict the behavior of NPs in the biological system. Quantitative structure – toxicity relationship (QSTR) has been adopted as a potential method to predict the cytotoxicity of untested NPs. Hence, in the present study, multiple linear regression (MLR) models were developed using 17 MeO_xNPs on *Escherichia coli* (*E. coli*) bacteria cells under both light and dark conditions. The models were developed applying Small Dataset Modeler software, version 1.0.0 (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab/) which generates models with a limited number of data points. Periodic table-based descriptors (both 1st and 2nd generation) were used for the modeling purpose. Two statistically significant MLR models based on photo-induced toxicity ($Q^2_{(LOO)} = 0.612$, $R^2 = 0.726$) and dark-based toxicity ($Q^2_{(LOO)} = 0.627$, $R^2 = 0.770$) were developed. From the developed models, we interpreted that increase in valency and oxidation state of the metal will decrease the cytotoxicity whereas the atomic radius of the metal and electronegativity of MeO_xNPs influence the toxicity toward *E. coli* cells. The MLR models were validated using different internal validation metrics. Additionally, we have collected 42 MeO_xNPs as an external set to observe the predictive power of the two developed MLR models and categorize them into toxic and non-toxic classes. The chemical features selected in the developed models are important for understanding the mechanisms of nanotoxicity. Thus, the developed models can be a scientific basis for designing safer NPs.

ARTICLE HISTORY

Received 15 November 2021
Revised 25 January 2022
Accepted 31 January 2022

KEYWORDS

QSTR; Small Dataset Modeler; *Escherichia coli*; MeO_xNPs

1. Introduction

Metal oxide nanoparticles (MeO_xNPs) due to their unique inherent physicochemical properties have been applied to different consumer products like cosmetics, dental fillings, self-cleaning coatings, textiles, etc. (Puzyn et al. 2011). They are also used in solar batteries, in automobiles as catalytic converter. Owing to the photocatalyst property, the MeO_xNPs are used as water-treatment agents to degrade various organic pollutants. However, it has been observed that these nano-sized particles with their extraordinary properties may be toxic to some organisms. Cosmetics containing such nanoparticles may cause more harm than the UV rays they are

intended to protect from (Dreher 2004). The recent report from Research and Markets warned that the nanoparticle drug market may reach up to US\$200 Billion by 2024 which is approximately equal to 10% of growth according to per compound annual growth rate (CAGR) (Global Nanoparticle Drug Delivery Market, Dosage, Price and Clinical Pipeline Outlook 2024). The significant data gap and uncertainties of MeO_xNPs have led to obstruction to their potential risk assessment which has doubted our concern regarding their adequate regulations. A prior study of 24 MO_xNPs in mammalian cells revealed that some of these materials show hazard potential (Kaweeteerawat et al. 2015). Therefore,

CONTACT Kunal Roy  kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in  Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India

 Supplemental data for this article can be accessed [here](#).

© 2022 Informa UK Limited, trading as Taylor & Francis Group

developing rapid methods for predicting their toxicity profile to the living organism in a timely fashion is very essential. As it is impossible to evaluate many MeOxNP based commercial goods *via in vivo* and *in vitro* studies, *in silico* approaches are the best alternatives (Fourches et al. 2010). Computational techniques may speed up the process and reduce time and cost in the development of new NPs with reduced toxicity. Chemoinformatics along with data statistics, machine learning, and artificial intelligence (AI) have gained a lot of attention in recent times. A study on the toxicity of NPs has been reported by Sizochenko et al. (2018) against various species. Furthermore, Gonzalez-Diaz et al. (2013) introduced various theories including the perturbation-theory (PT), machine learning (ML), and information fusion (IF) approach known as the IFPTML algorithm. The basic idea of this approach is to predict the output values of the function $f(V_{ij})_{calc}$ for the j th multiple properties of the query system (S_i) under study with a single model (Ortega-Tenezaca and González-Díaz 2021). The developed models by this process can be used to predict the biological activity of NPs, proteins, drugs, and complex molecules. Nanoparticles with different surface coating popularly known as Nanoparticle Drug Delivery Systems (DDNS) are also of high interest in biomaterial sciences. But the data are dispersed and not as large as the DDNS components. Therefore, Santana et al. combined Perturbation Theory and Machine Learning (PTML algorithm) to predict the score function $(f(V_{ij}))_{calc}$ and predict the best component for Nanoparticle Drug Delivery Systems (DDNS) designing. In another work, Urista et al. performed perturbation theory in combination with ML and IF (PTMLIF) to predict the antimalarial activity probability of nanoparticle compounds or drug complexes (Urista et al. 2020). Perturbation theory was also applied in the computational modeling of nanomedicine (Speck-Planche et al. 2015). The perturbation theory was also used by Kleandrova et al. (2014) to probe the ecotoxicity and cytotoxicity of NPs under various experimental conditions, sizes, compositions, coating agents, and time to which the biological targets were exposed to the NPs.

The *in-silico* approach has been adopted by the United States Environmental Protection Agency (U.S. EPA) for toxicity data gap filling for those

MeOxNPs that do not have available experimental data. Among the *in-silico* approaches, quantitative structure-activity relationship (QSAR) is most commonly used to predict and screen toxicity of nanoparticles (NPs) at a low cost and in less time. QSAR is a mathematical method that finds correlations among the properties of interest to the nanoparticle structures with the help of certain descriptors. QSAR finds the complex relationship of the property with the nanostructures without having detailed knowledge about their interaction mechanism. The recent development in the field of nanoparticles has raised interest in extending the traditional QSAR paradigm to develop nano-QSAR for predicting the toxicity of newly generated or synthesized MeOxNPs which can also work as a blueprint for experimental design of new MeOxNPs with enhanced activity (Puzyn, Leszczynska, and Leszczynskii 2009). One of the biggest challenges of QSAR is assessing the reliability of the identified relationships with the interpreted mechanism (Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models 2007). Although QSAR has made great progress in modeling traditional chemicals, its application in the field of NPs is very limited. The reason may be attributed due to the less amendable features or due to highly complex large three-dimensional nanostructures. The lack of appropriate descriptors also limits the nano-QSAR modeling. In most cases, the determination of experimental descriptors is laborious and time-consuming. For the computation of the quantum mechanical descriptors, a certain quantum-chemical background is essential to be able to calculate such descriptors; these descriptors are also computationally demanding whereas periodic table-based descriptors are easy to calculate, and they do not require any extensive computational knowledge. Periodic tables descriptors were used in Kar et al. (2014) work which showed the efficiency and the predictive of the easily computable descriptors. Moreover, the main problem with NPs is the availability of a sufficient number of experimental data with high quality. During handling of limited data, much useful information may get lost due to the hold-out samples. Again, the division of such datasets into training and test sets may cause a bias in descriptor selection with the fixed composition of

the training set. The Small Dataset Modeler tool (Ambure et al. 2019) can be a solution to such problems. This tool is designed to handle small datasets and provide components like 'Double cross-validation' which have options set for model selection techniques to generate different models.

In the present study, we have tried to develop cytotoxicity models for 17 MeO_xNPs ranging from 15 to 90 nm and validated the models to find the correlation of chemical features with the cytotoxicity of MeO_xNPs to *E. coli* cells [$-\log LC_{50}$, which is a dose required for a fixed response type of endpoint (Roy, Kar, and Das 2015)] using the Small Dataset Modeler tool. The cytotoxicity of NPs was investigated in *E. coli* cells because they are single-celled organisms and can be correlated to other higher-level organisms. Single-celled organisms, due to their versatile metabolism, are considered excellent environmental indicators to determine the impact of NPs on environmental health as well as human health (Kertesz and Mirleau 2004; Osler and Sommerkorn 2007). During model development, we have avoided using physicochemical descriptors, such as size, shape, zeta potential, etc., as the use of such descriptors will make the model less usable for new data without such information. In addition, without using these descriptors also, we have got statistically sound models. In addition, we may mention here that Adams, Lyon, and Alvarez (2006) in one of their papers mentioned that particle size of NPs does not affect the antibacterial activity against *E. coli* for the investigated sizes as all the nanopowders in water suspension result in similar size aggregate. So, the effect of the powder size of NPs gets nullified over here. Because of the limited experimental data available, this study uses the Small Dataset Modeler tool with an application of the Double Cross-validation tool. Although the availability of toxicity data sets for metal oxides is very limited, this is also the inspiration for the development of new robust models for filling data gaps.

2. Methods and materials

2.1. The dataset

The cytotoxicity data of 17 MeO_xNPs to *E. coli* was obtained from the previously published research

paper by Pathakoti et al. 2014 and was expressed in terms of negative logarithm of lethal concentration causing 50% growth inhibition of *E. coli* ($-\log LC_{50}$). The dataset was considered for 2D-QSAR modeling under two experimental conditions, i.e. in the dark and in presence of light.

2.2. Molecular descriptors

The dynamic behavior of MeO_xNPs complicates the descriptor calculation. So, descriptor calculation for metal oxide nanoparticles is a crucial job for researchers. Therefore, in the present work, we have calculated sixteen 2nd generation periodic table-based descriptors (De and Roy 2021) along with forty-four descriptors using Elemental Descriptor Software freely available from our laboratory (<https://dtclab.webs.com/software-tools>). The rest periodic table descriptors were taken directly from the periodic table based on the nature of the metals. Total sixty-one periodic table descriptors were calculated for modeling of the MeO_xNPs.

2.3. Model development: application of the small dataset modeler

Before performing QSAR model development, usually division of a dataset into training and test sets is performed. However, as the present dataset contains a limited number of data points, division of the dataset is not desirable. Here, the model development has been done using the Small Dataset Modeler, version 1.0.0 software (Ambure et al. 2019) without any data set division. Small dataset modeler works on the double cross-validation (Ambure et al. 2019) strategy as illustrated in Figure 1. The double cross-validation consists of two nested loops: internal (inner) and external (outer) cross-validation loops. The external loop divides the dataset into calibration and validation sets while in the internal loop, the calibration set is used for model development and model selection. The predictivity assessment of the models is entirely done by the validation set. The Small Dataset Modeler software (Ambure et al. 2019) avoids the dataset splitting step in the inner loop; instead, the whole dataset is used as the 'modeling set.' The modeling set is constantly divided into calibration and validation sets and all possible combinations ('k') of both sets are

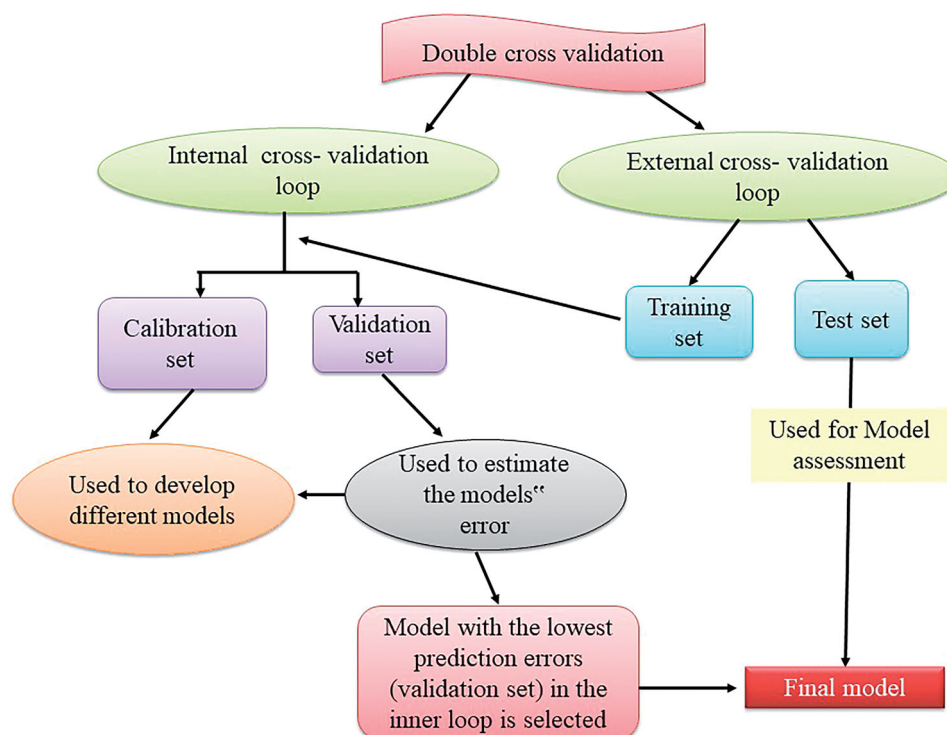


Figure 1. The double cross validation (DCV) methodology.

derived. The number of compounds in the validation set is user-defined and based on it, different combinations of calibration sets and validation sets are derived. The generated calibration sets are used to develop multiple linear regression (MLR) models applying the genetic algorithm-MLR (GA-MLR) technique (Devillers 1996; Venkatasubramanian and Sundaram 2002). On the other hand, the validation set is used to obtain the prediction quality of the respective developed MLR models. The whole workflow is illustrated in Figure 2.

Additionally, internal and external validation metrics are calculated for all the selected models. The final model could be selected out of any of the five recommended methods.

1. MLR/PLS model with the least absolute MAE (95%) value in the validation set is selected.
2. MLR/PLS model with least MAE (95%) value in the modeling set.
3. MLR/PLS model with highest $Q^2_{\text{Leave-many-out}}$ in the modeling set.
4. In the case of consensus predictions, the top-ranking model should be selected based on the least MAE (95%) values in the respective validation set. It works on two variants namely

- a. Simple arithmetic average of prediction containing all the selected top models
 - b. The weighted average of prediction works by assigning appropriate weight to the top selected models based on the mean absolute error gathered from leave-one-cross validation MAE_{cv} (95%).
5. A pool of unique descriptors is selected from the top three models with the least MAE (95%) of the validation set. Further, these descriptors are used for model development purposes. The best subset selection (BSS) method is used in the case of the MLR model to find the best descriptor combination, whereas, in the case of PLS models, the models are formed by the descriptors selected from the top models.

The small dataset modeler software also combines data curation, exhaustive double cross-validation, along consensus predictions in the case of model development which is essential for small datasets.

2.4. Reliability of predictions of untested NPs

The estimation of checking the reliability of predictions of the generated model for an external set is

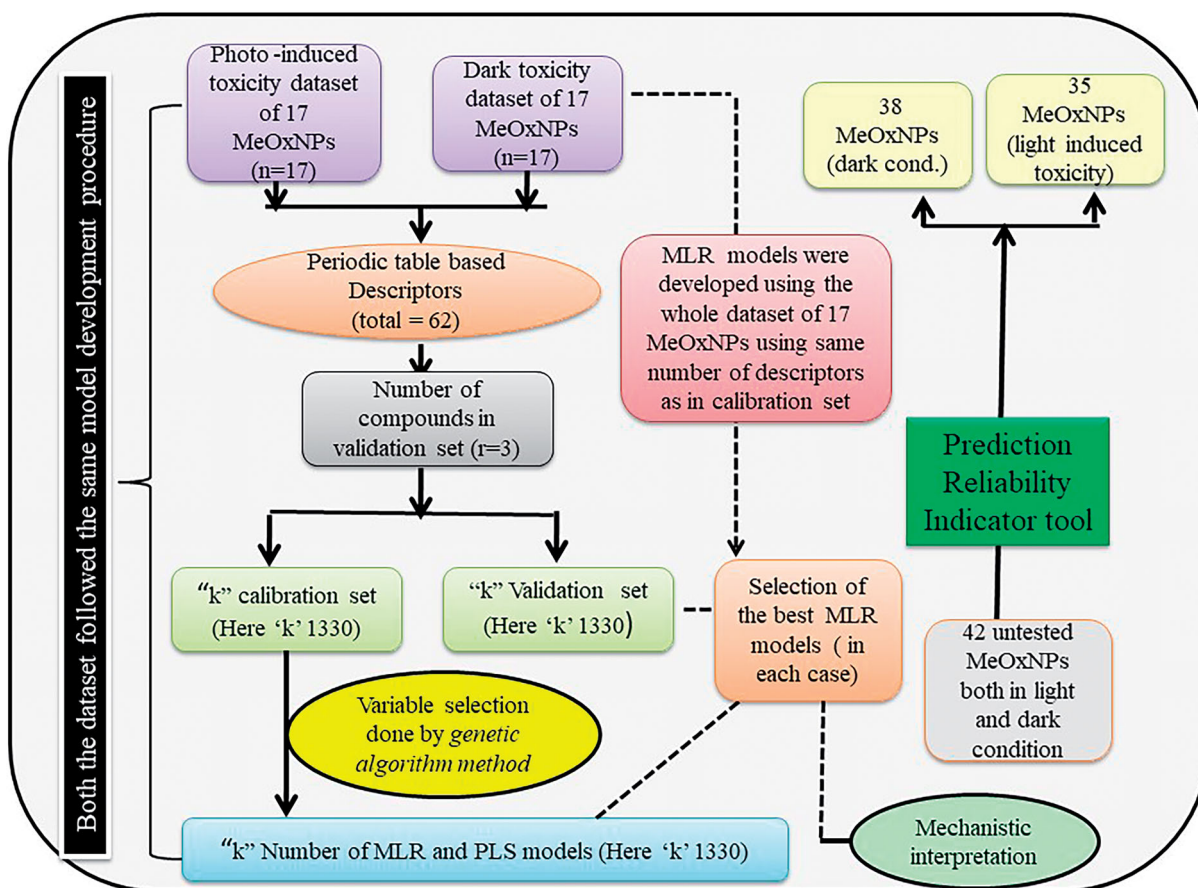


Figure 2. Flowchart of the approach used by the Small Dataset Modeler.

an interesting job when the data points are within the AD. A robust QSAR model can help in predicting the external dataset depending on the molecular features. Several tools are available for estimating the confidence of a QSAR model. Among them, one of the tools which can rely on for the prediction of the response data of the external set is the 'Prediction Reliability Indicator' tool available from <http://dtclab.webs.com/software-tools>. This tool categorizes the quality of predictions for the test set or a true external set as good, moderate, and bad based on the three rules of scoring. Rule 1: scoring depending on the quality of leave-one-out prediction of the 10 closest training set compounds to the test/external compounds, Rule 2: scoring based on the AD in terms of similarity using the standardization method (Roy, Kar, and Das 2015), and the Rule 3: scoring based on the proximity of predicted value of the query compound to the experimental mean of the training set responses. Thus, in the present study we have

collected 42 MeO_xNPs as an external set (Table S3) to evaluate the predictive power of the developed MLR models and found that out of 42 MeO_xNPs, 38 MeO_xNPs showed good predictions with the dark condition toxicity data model and 35 MeO_xNPs showed good predictions with the photo-induced toxicity data model. We were able to classify the untested metal oxides into toxic and non-toxic categories based on the mean toxicity value of the training set as the threshold (Figure 3).

2.5. Statistical validation metrics

Validation of the developed models is an important part of any QSAR study. Multiple approaches can be applied to the developed MLR models to rigorously measure and validate their fitness, stability, robustness, and predictivity. In the present study, we have computed statistical quality metrics like determination coefficient (R^2) and variance ratio (F) and validation metrics like leave-one-out determination coefficient

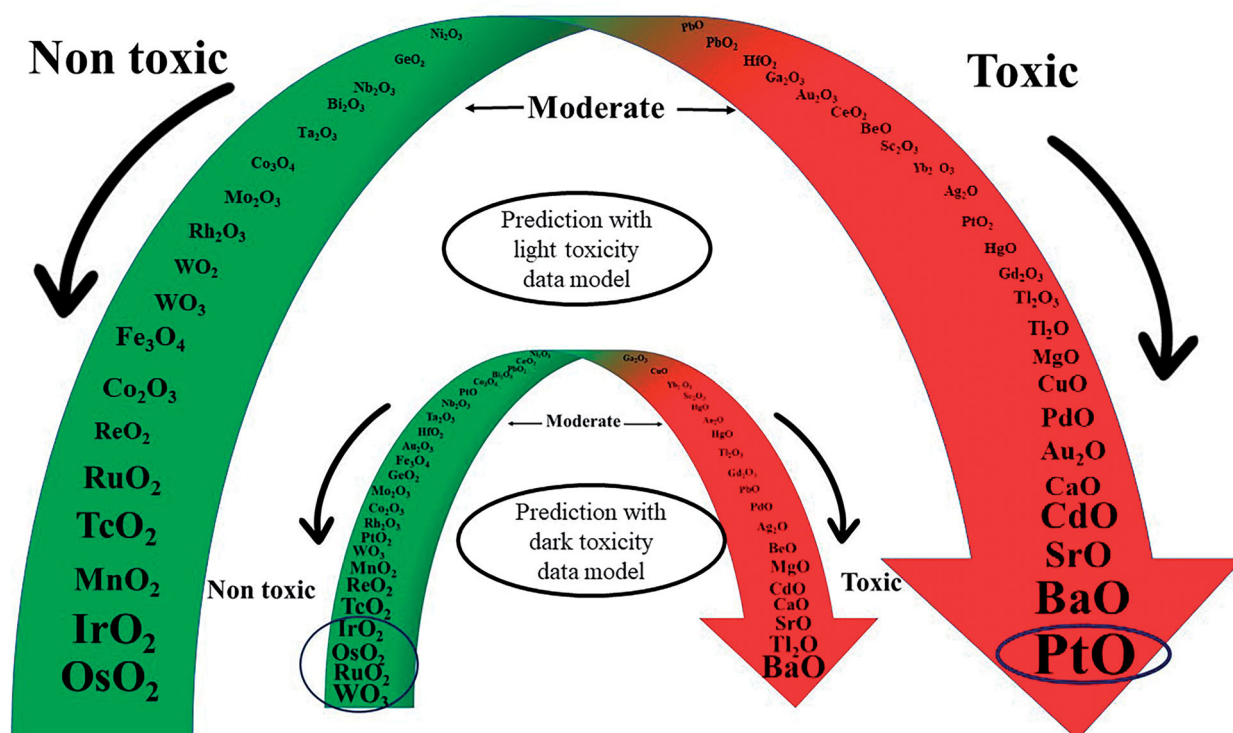


Figure 3. Prediction of the untested NPs with the developed MLR models shown in the increasing order of toxicity. The NPs within the circle represent the NPs which were not correctly predicted by the MLR models as per the 'PRI tool.'

(Q^2_{LOO}) and predicted residual sum of squares (PRESS). Additionally, r_m^2 metrics (Roy et al. 2012), root mean square error (RMSE), and mean absolute error (MAE) were also calculated (Roy et al. 2016).

3. Applicability domain

The applicability domain, also known as AD, is defined as the chemical space in which the QSAR model makes predictions for new compounds based on the knowledge of the training set with given reliability. The test set compounds should be structurally similar to the training set compounds so that the models work fine based on the similarity-based principle. The AD assessment of the proposed models was performed using a standardization approach (Roy, Kar, and Das 2015). After performing AD, it was found that compounds 1 (CuO) and 9 (Al_2O_3) were outliers in the modeling set, and compounds 2 (Au_2O), 19 (IrO_2), 25 (OsO_2), 29 (PtO), 30 (PtO_2), and 33 (RuO_2) were outside AD in the external prediction set in the case of light-induced toxicity data whereas in the case of dark toxicity data it was observed that compound 9 (Al_2O_3) was an outlier in the modeling set and compounds 25

(OsO_2), 33 (RuO_2), and 41(Yb_2O_3) were outside AD in the external set.

4. Result and discussion

Periodic table-based descriptors were used for modeling 17 MeO_xNPs using the Small Dataset Modeler software (http://teqip.jdvu.ac.in/QSAR_Tools/). The details of the descriptors used in the QSAR modeling are shown in Table 1 in the molecular descriptors section. The statistical quality and validation metrics of both the models have shown acceptable values for R^2 , Q^2_{LOO} , RMSE, MAE, r_m^2 metrics as per Table 2. The descriptors obtained from the MLR models of the two datasets (dark and light-induced toxicity) are explained in Table 3. The experimental and predicted values of all MeO_xNPs are given in Tables S1, S2 in the Supplementary Section. The scatter plots (Figure 4) of the MLR models are also shown in this section.

4.1. Interpretation of the obtained descriptors

4.1.1. Model 1: photo-induced toxicity

$$\log \frac{1}{\text{LC}_{50}} = 4.686 + 0.015 \text{ atomic rad (pm)} - 0.725 \text{ Valence} + 0.134 \text{ sq_sum_epsilon/N} \quad (1)$$

Table 1. Description of the descriptors generation wise.

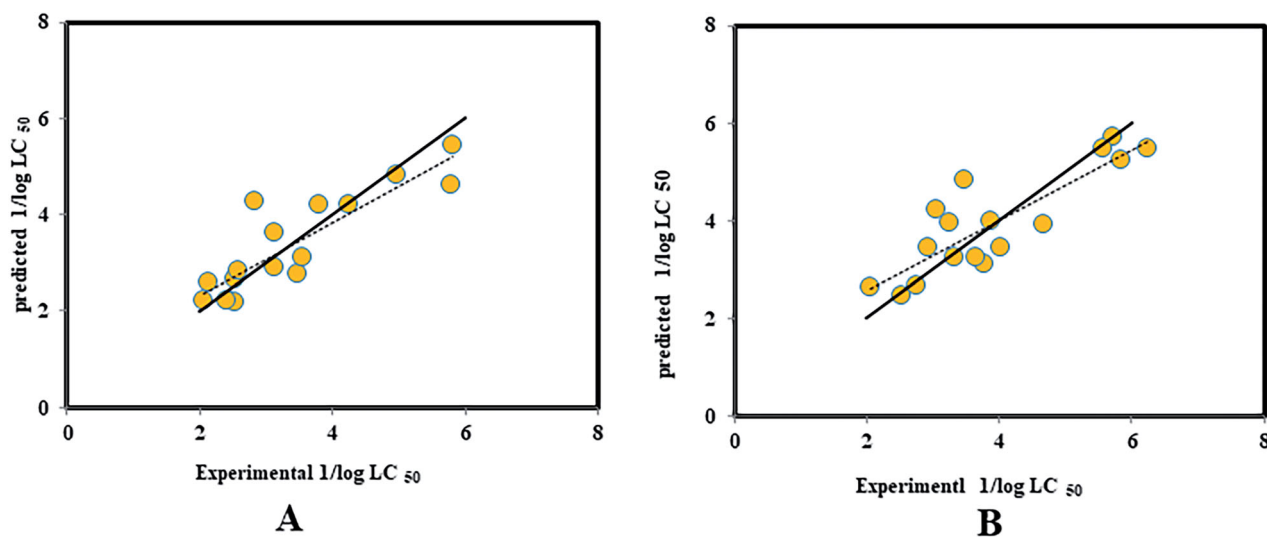
No. generation of PT descriptors	Mathematical expression	Description
1st generation		
1	D1_Metals	Total no. of metals
2	D2_SemiMetal	Total no. of semi-metals
3	D3_HeteroNonMetals	Total no. of non-metallic heteroatoms (N, O, F, P, S, Cl, Se, Br, I, At) excluding hydrogen
4	Metals_SumIP	Sum of ionization potential energy (kJ/mol) of metals
5	SemiMetals_SumIP	Sum of ionization potential energy (kJ/mol) of semi-metals
6	Electrons_ActiveM	No. of electrons of active metal
7	AtWt_ActiveM	Atomic wt. of active metal
8	MassNo_ActiveM	Mass no. of active metal
9	Neutons_ActiveM	Number of neutrons of active metal
10	IP_ActivM	Ionization potential energy (kJ/mol) of active metal
11	X_ActivM	Electronegativity of active metal
12	VWR_ActivM	van der Waal radius of active metal
13	Electrons_Activ_SM	No. of electrons of active metalloid/semi-metal
14	AtWt_Activ_SM	Atomic wt. of active metalloid/semi-metal
15	MassNo_Activ_SM	Mass no. of active metalloid/semi-metal
16	Neutons_Activ_SM	Number of neutrons of active metalloid/semi-metal
17	IP_Activ_SM	Ionization potential energy of active metalloid/semi-metal
18	X_Activ_SM	Electronegativity of active metalloid/semi-metal
19	VWR_Activ_SM	van der Waals radius of active metalloid/semi-metal
20	Electrons_Activ_NM	No. of electrons of active non-metal
21	AtWt_Activ_NM	Atomic wt. of active non-metal
22	MassNo_Activ_NM	Mass no. of active non-metal
23	Neutons_Activ_NM	Number of neutrons of active non-metal
24	IP_Activ_NM	Ionization potential energy of active non-metal
25	X_Activ_NM	Electronegativity of active non-metal
26	VWR_Activ_NM	van der Waals radius of active non-metal
27	SuM_Active_M_SM	Summation of number of metals and semi-metal
28	SuMElectrons_Active_M_SM	Summation of electrons of active metal and semi-metal
29	SuMatWt_Active_M_SM	Summation of At. wt of active metal and semi-met
30	SuMMassNo_Active_M_SM	Summation of mass no. of active metal and semi-metal
31	SuMNeutons_Active_M_SM	Summation of neutrons of active metal and semi-metal
32	χ_{Ox}	Oxidation number
33	N oxygen	Number of oxygen atoms present in the metal oxide.
34	Molar volume	
35	Atomic rad	Atomic radius
36	Electronegativity	—
37	Thermal conductivity	—
38	Melting point	—
39	Electron affinity	—
40	Density	—
41	Mol wt	Molecular weight
42	Cationic charge	—
43	$\sum \chi$	Total metal electronegativity in specific metal oxide
44	$\sum \chi / no$	Total metal electronegativity in specific metal oxide relative to the number of oxygens.
2nd generation		
1	Z metal	The atomic number of metals
2	Z vmetal	Valence electron of metal
3	PNmetal	Period number of metals
4	$\lambda = (Z_{\text{metal}} - Z_{\text{vmetal}}) / Z_{\text{vmetal}}$	The core environment of metal is defined by the ratio of the number of core electrons to the number of valence electrons
5	$\mu = 1 / (PN_{\text{metal}})$	—
6	V metal	Valence of metal
7	$\alpha_{\text{metal}} = \lambda * \mu$	—
8	$\sum \alpha_{\text{metal}} = \alpha_{\text{metal}} * N_{\text{metal}}$	—
9	$\sum \alpha_{\text{oxy}} = N_{\text{oxy}} * 0.33$	—
10	$\sum \alpha = \sum \alpha_{\text{metal}} + \sum \alpha_{\text{oxy}}$	The core count gives a measure of the molecular bulk
11	$\varepsilon_{\text{metal}} = -\alpha_{\text{metal}} + (0.3 * Z_{\text{vmetal}})$	Electronegativity count of metal
12	$\varepsilon_{\text{oxy}} = -\alpha_{\text{oxy}} + (0.3 * Z_{\text{oxyv}})$	Electronegativity count of oxygen
13	$\sum \varepsilon = \varepsilon_{\text{metal}} * N_{\text{metal}} + \varepsilon_{\text{oxy}} * N_{\text{oxy}}$	Electronegativity count of total metal oxide
14	$\sum \varepsilon / N$	Sum epsilon relative to the number of atoms in the molecule
15	$(\sum \alpha)^2$	Square of the summation of alpha, gives a measure of the molecular bulk
16	$(\sum \varepsilon / N)^2$	Squared sum epsilon relative to the number of atoms

Table 2. Details of statistical parameters were obtained from both MLR models.

Statistical parameters	R^2	R^2_{adj}	Q^2_{LOO}	$\bar{r}^2_m$	Δr^2_m	MAE (95%)	PRESS value	F
Model 1	0.726	0.663	0.612	0.497	0.166	0.587	9.747	11.48
Model 2	0.770	0.717	0.627	0.543	0.054	0.482	8.494	14.48

Table 3. A detailed description of the obtained descriptors from two developed MLR models (A and B).

Serial No.	Descriptors	Description
1.	Atomic rad (pm)	This descriptor stands for the measure of the size of the atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons.
2.	Valence	The valence of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules
3.	$(\sum \varepsilon/N)^2$	It stands for the sum of electronegativity of the atoms of the metal oxide scaled by the number of atoms present which is calculated based on the electronegativity count ($\sum \varepsilon$) of the oxides.
4.	χ_{ox}	It represents the oxidation number of the metal atom.

**Figure 4.** Scatter plots for the dark toxicity dataset (A) and the photo induced dataset (B).

$$Q_{LOO}^2 = 0.612, R^2 = 0.726, R_{adj}^2 = 0.663, \\ \overline{r}_m^2 = 0.497, \Delta r_m^2 = 0.166$$

The first descriptor influencing the toxicity is Atomic rad (pm). This descriptor stands for the measure of the size of an atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons. This descriptor is positively correlated with the response which means that with an increase of the descriptor value, the toxicity data also increases as in nanoparticles 11 (La_2O_3) and 8 (Y_2O_3); in contrast, a decrease of descriptor values as represented in nanoparticles 13 (SiO_2) and 9 (Al_2O_3) decreases the toxicity value. The reason for such phenomenon is that due to the increase of the atomic radii, the pull of the electrons by the neutron reduces since the number of the shells in the atom increases which is also known as the shielding effect. This makes the removal of the electrons from the atom easier and leads to the formation of metal cation.

The next descriptor, Valence or valency of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules. This also indicates the presence of the number of valence electrons in the outermost shells. The negative contribution of the descriptor suggests that with an increase of valence of the metal, the toxicity decreases as in nanoparticles 16 (Cr_2O_3) and 17 (Fe_2O_3) whereas with a decrease of valence the toxicity increases as observed in nanoparticles 9 (Al_2O_3) and 2 (ZnO). The presence of few electrons in the outermost shells makes it easier for the metal to exchange the valence electron with its neighboring element and form a covalent or electrovalent bond with it. On the other hand, it is also possible to lose its electron easily to get converted into its cationic form and accelerate the toxic reaction.

This second-generation descriptor $(\sum \varepsilon/N)^2$ stands for the sum of electronegativity of the atoms of the metal oxide scaled by the number of atoms present, which is calculated based on the electronegativity count ($\sum \varepsilon$) of the oxides.

$$\sum \varepsilon = \varepsilon_{\text{metal}} * N_{\text{metal}} + \varepsilon_{\text{oxy}} * N_{\text{oxy}}$$

Here, $\varepsilon_{\text{metal}}$ and ε_{oxy} are the electronegativity count of metal and oxygen atoms, respectively and N_{metal} and N_{oxy} are the number of metal and oxygen atoms. The positive coefficient of the descriptor in Equation 1 infers that an increase in electronegativity favors the rise in toxicity as in nanoparticles 1 (CuO) and 2 (ZnO) whereas the nanoparticles with the presence of low electronegativity have less toxicity as observed in nanoparticles 14 (SnO₂) and 10 (Bi₂O₃). It has been observed that due to the presence of high electronegativity, the atoms pull electrons from their neighboring element or molecules leading to the development of an electrostatic bond with it.

4.1.2. Model 2: dark-induced toxicity

$$\log \frac{1}{LC_{50}} = 6.433 - 0.838 \chi_{\text{ox}} + 0.014 \text{ atomic rad (pm)} - 0.607 \text{ Valence} \quad (2)$$

$$Q_{\text{LOO}}^2 = 0.627, R^2 = 0.770, R_{\text{adj}}^2 = 0.717, \\ \overline{r}_m^2 = 0.543, \Delta r_m^2 = 0.054$$

The first descriptor in this equation contributing to the toxicity is χ_{ox} , which stands for oxidation number of the metal atom. This descriptor contributes negatively toward the toxicity which is well-observed in nanoparticles 3 (TiO₂) and 14 (SnO₂) where an increased oxidation number leads to a decreased toxicity value and vice versa as exemplified in nanoparticles 1 (CuO) and 2 (ZnO). The oxidation number reflects the ability of an atom to acquire, donate or share electrons, and it is quite like the valence electrons. If the oxidation number is positive, then it means that the element has lost electrons and has fewer electrons in the outermost shell. On the other hand, the negative oxidation number suggests that the element has gained electrons and has more electronegativity. As described earlier in the case of the valence descriptor, here also, the loss of electrons can accelerate the formation of the cationic form of the metal that leads to the toxic reaction.

Other descriptors having an influencing effect on toxicity to *E. coli* are discussed previously in the case of Model 1.

The atomic rad and valence descriptors appearing in both Model 1 and Model 2 play an important role in influencing the toxicity toward *E. coli*. Comparing both the models, we can infer that the presence of light triggers the impact of toxicity of the metal oxides. It was noted that with an increase in atomic rad and valence descriptors values, the respective toxicity value ($\log 1/LC_{50}$) was found to be more in presence of light than in dark. The positively contributing descriptor, atomic rad, lead to $\log 1/LC_{50}$ values of 4.96 and 5.79 for nanoparticles 11 (La₂O₃) and 8 (Y₂O₃), respectively in dark conditions whereas in presence of light, the toxicity value increases to 5.56 and 5.84 for the same respective nanoparticles with same descriptor values. The same observation was seen in the case of negatively contributing descriptor valence; in this case too, with the same descriptor value, the $\log 1/LC_{50}$ values for nanoparticle 9 (Al₂O₃) and 2 (ZnO) in dark condition were 2.42 and 5.8, respectively which became 2.75 and 6.23 in presence of light.

4.2. Proposed mechanism of toxicity to *E. coli* cell

The potential mechanism of cytotoxicity of MeO_xNPs to *E. coli* cells was deduced with the help of information extracted from the descriptors of the nano-QSTR models. The insights obtained from the descriptors pointed that atomic radius, the electronegativity of the metal in metal oxides, and the number of electrons in the valence shell have a substantial effect on the mortality of *E. coli*. The cluster of MeO_xNPs after forming metal cations due to the loss of electrons can easily interact with the proteins present on the surface of the membrane and can enter into the *E. coli* cells. An increase in electronegativity raises the catalytic properties of the metal oxides and enhances their intervention process. After the entry of the metal oxides in the plasma region through physiochemical interaction, it breaks the Me-O bond due to the presence of acidic conditions within the cells. The detached metal cation can lead to macromolecular complexation and destruct the molecular structure (Mu et al. 2016). On the other hand, the metal cation instigates the formation of free radicals ($\cdot\text{OH}$) while interacting with the ribosomes present in the *E. coli* cell more than the metals naturally do (Roy, Ojha, and Roy 2019). The elevated levels of radicals

interact with the molecules in the cells and produce more reactive oxygen species. ROS production causes nucleoid damage that causes DNA fragmentation and death of the cell due to blockade of the synthesis of proper protein structure. The whole process of the toxicity mechanism is well-depicted in Figure 5.

4.3. Comparison with the previous work

A comprehensive comparison is not possible with the previous work since the models were developed using different compositions of modeling sets. Pathakoti et al. (2014) used DFT (density functional theory) descriptors which require computationally intensive quantum chemical calculations whereas

the work performed in the current paper used periodic table-based descriptors which are simple, easy, and do not require any expert knowledge. It is also noteworthy that in the previous paper, the authors reported only a few statistical quality metrics. They have kept aside a few data points for predictions and developed models using a lower number of data points. But in the present paper, we have used all available data points to avoid information loss and reported more extensively the quality metrics, such as R^2 , F , PRESS value, R^2_{adj} , Q^2 , $\bar{r}^2_m$, Δr^2_m metrics (Tables 4, 5). We believe that handling a small dataset is a crucial job for a nano-QSAR modeler and information may be lost while modeling a small dataset due to holding out samples. This problem has been overcome in this work by using the small

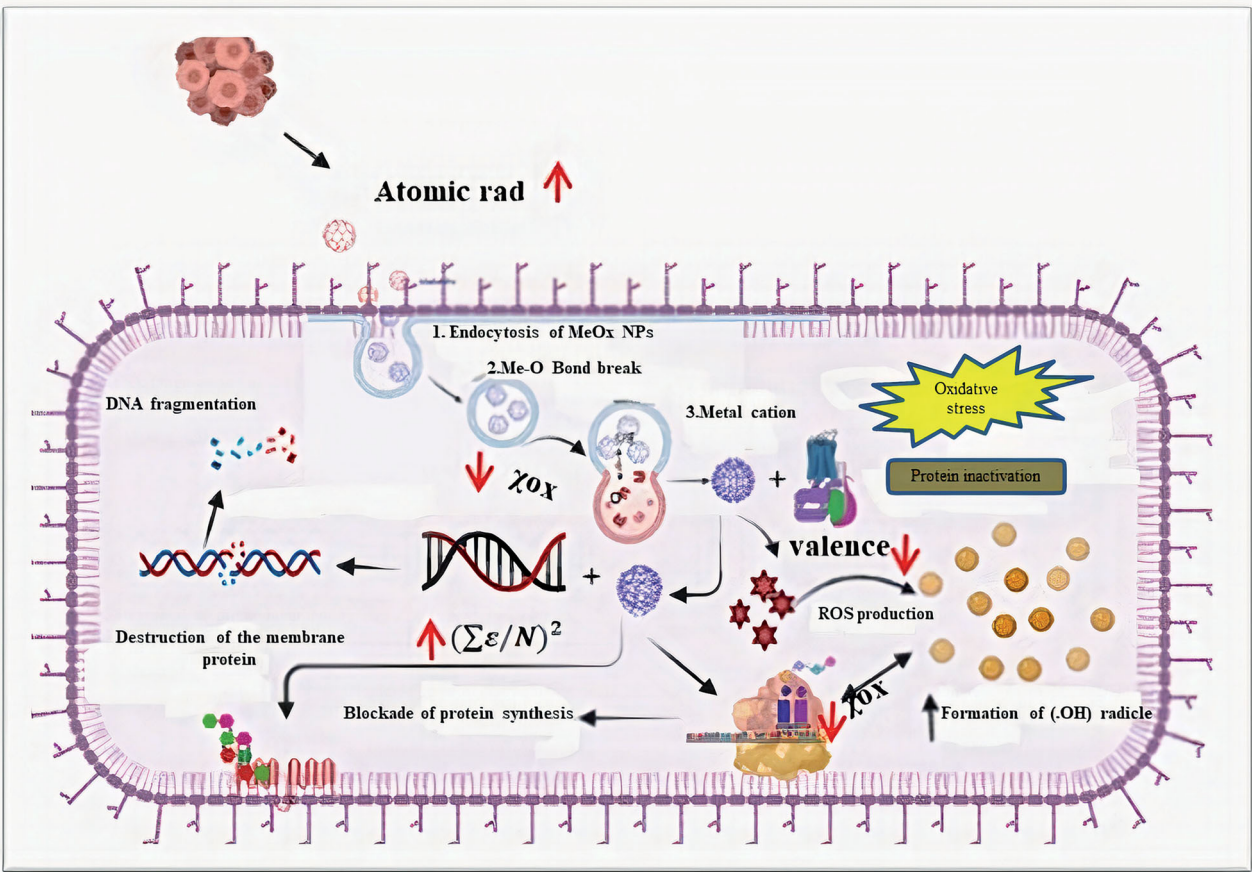


Figure 5. Toxicity mechanism of the MeOx NPs to *E. coli* cell.

Table 4. Comparison table for the photo-induced dataset.

Statistical parameters	Number of data points (<i>n</i>)	R^2	F	Press value	R^2_{adj}	Q^2	$\bar{r}^2_m$	Δr^2_m	MAE (95%)	SD
Previous* work	Training set = 13 Test set = 4	0.804	20.51	–	–	–	–	–	–	0.63
Current work	Whole dataset (<i>n</i> = 17)	0.726	11.48	9.747	0.663	0.612	0.497	0.166	0.587	–

* R^2_{pred} calculated to be a negative value.

Table 5. Comparison table for the dark dataset.

Statistical parameters	Number of data points	R^2	F	Press value	R^2_{adj}	Q^2	T^2_m	Δr^2_m	MAE (95%)	SD
Previous* work	Training set = 13 Test set = 4	0.87	33.83	–	–	–	–	–	–	0.48
Current work	Whole dataset ($n = 17$)	0.77	14.48	8.494	0.717	0.627	0.543	0.054	0.482	–

* R^2_{pred} calculated to be a negative value.

dataset modeler tool which uses the DCV method. By using the DCV method, we also omit the bias in description selection with a constant composition training set. Furthermore, the developed models were used to predict the toxicity of 42 untested metal oxides which do not have any reported toxicity data. The developed models successfully predicted most of the query NPs which were found to be within the AD, and with the help of the Prediction Reliability Indicator tool (PRI), their prediction quality was also tested. Additionally, we have also tried to provide a conceptual toxicity mechanism diagram for a better understanding of toxicity to *E. coli* cells (Figure 5).

5. Conclusion

In the study, we have developed 2D MLR QSAR models to investigate the essential features in 17 MeO_xNPs responsible for the cytotoxicity of *E. coli* cells with $-\log LC_{50}$ as the endpoint. This predominant nano-structure information has helped in identifying the cause of toxicity of MeO_xNPs and has also suggested some mechanisms through which it can affect the *E. coli* cells or other biological cells. The statistical results of the developed models are robust and can be used to extrapolate the endpoint if the experimental information of new MeO_xNPs is unavailable. The periodic table descriptors used in this work are can be easily obtained either using the Elemental Descriptor Calculator software (available from <https://dtclab.webs.com/software-tools>) or from the Periodic Table information. Now, knowing the regression coefficients of different descriptors in the reported model, one can easily compute the toxicity in an excel sheet without requiring the use of any codes.

The major findings of the present study are summarized below:

1. Only periodic table descriptors were used which can be calculated easily without any computational calculations. The reproducible

descriptors can easily encode the cytotoxicity with good interpretability.

2. The application of the ‘Small Dataset Modeler’ software was emphasized in the current work. Division of the small dataset (17 compounds) into training and test sets is not desirable here to generate and validate QSAR models to avoid the loss of information and introduce bias in descriptor selection.
3. The developed MLR models were conceptually simple, easy to apply, computationally inexpensive as well as can be used to reduce the data gap in untested NPs.
4. Here, we interpreted that increase in valency and oxidation state of the metal in MeO_xNPs will decrease the cytotoxicity whereas the atomic radius of the metal and electronegativity of MeO_xNPs has a substantial influence on the toxicity toward *E. coli* cells.
5. This study can be used for predicting the cytotoxicity of untested MeO_xNPs for which the experimental data are not known. The developed models have can be used in designing eco-safe NPs with better properties as evidenced by the predictions of 42 MeO_xNPs, out of which, 38 MeO_xNPs showed good predictions with dark condition toxicity data model and 35 MeO_xNPs showed good predictions with the photo-induced toxicity data model. Therefore, these models could be used to predict the cytotoxicity of the untested NPs before their utilization in commercial products.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

JR thankfully acknowledges the Indian Council of Medical Research, New Delhi for supporting financially the work in the form of a Senior Research Fellowship. KR also thankfully acknowledged financial assistance from CSIR, New Delhi to

KR under a Major Research project (CSIR Project No. 01IJ2895/17/EMR-II).

ORCID

Kunal Roy  <http://orcid.org/0000-0003-4486-8074>

References

- Adams, L. K., D. Y. Lyon, and P. J. Alvarez. 2006. "Comparative Eco-Toxicity of Nanoscale TiO_2 , SiO_2 , and ZnO Water Suspensions." *Water Research* 40 (19): 3527–3532. doi:10.1016/j.watres.2006.08.004.
- Ambure, P., A. Gajewicz-Skretna, M. N. D. Cordeiro, and K. Roy. 2019. "New Workflow for QSAR Model Development from Small Data Sets: Small Dataset Curator and Small Dataset Modeler. Integration of Data Curation, Exhaustive Double Cross-Validation, and a Set of Optimal Model Selection Techniques." *Journal of Chemical Information and Modeling* 59 (10): 4070–4076. doi:10.1021/acs.jcim.9b00476.
- De, P., and K. Roy. 2021. "QSAR and QSAAR Modeling of Nitroimidazole Sulfonamide Radiosensitizers: Application of Small Dataset Modeling." *Structural Chemistry* 32 (2): 631–642. doi:10.1007/s11224-021-01734-w.
- Devillers, J. 1996. *Genetic Algorithms in Molecular Modeling*. Academic Press.
- Dreher, K. L. 2004. "Health and Environmental Impact of Nanotechnology: toxicological Assessment of Manufactured Nanoparticles." *Toxicological Sciences* 77 (1): 3–5. doi:10.1093/toxsci/kfh041.
- Fourches, D., D. Pu, C. Tassa, R. Weissleder, S. Y. Shaw, R. J. Mumper, and A. Tropsha. 2010. "Quantitative Nanostructure-Activity Relationship Modeling." *ACS Nano* 4 (10): 5703–5712. doi:10.1021/nn1013484.
- Global Nanoparticle Drug Delivery Market, Dosage, Price and Clinical Pipeline Outlook 2024. Accessed 15 June 2020. https://www.researchandmarkets.com/research/kn78v5/global_200?w1/45
- Gonzalez-Diaz, H., S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto, and J. M. Ruso. 2013. "General Theory for Multiple Input-Output Perturbations in Complex Molecular Systems. 1. Linear QSPR Electronegativity Models in Physical, Organic, and Medicinal Chemistry." *Current Topics in Medicinal Chemistry* 13 (14): 1713–1741. doi:10.2174/1568026611313140011.
- Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, Paris, 2007, vol. 69, ch. 5. http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab/
- Kar, S., A. Gajewicz, T. Puzyn, K. Roy, and J. Leszczynski. 2014. "Periodic Table-Based Descriptors to Encode Cytotoxicity Profile of Metal Oxide Nanoparticles: A Mechanistic QSTR Approach." *Ecotoxicology and Environmental Safety* 107: 162–169. doi:10.1016/j.ecoenv.2014.05.026.
- Kaweeteerawat, Chitrada, Angela Ivask, Rong Liu, Haiyuan Zhang, Chong Hyun Chang, Cecile Low-Kam, Heidi Fischer, et al. 2015. "Toxicity of Metal Oxide Nanoparticles in *Escherichia coli* Correlates with Conduction Band and Hydration Energies." *Environmental Science & Technology* 49 (2): 1105–1112. doi:10.1021/es504259s.
- Kertesz, M. A., and P. Mirleau. 2004. "The Role of Soil Microbes in Plant Sulphur Nutrition." *Journal of Experimental Botany* 55 (404): 1939–1945. doi:10.1093/jxb/erh176.
- Kleandrova, V. V., F. Luan, H. González-Díaz, J. M. Ruso, A. Melo, A. Speck-Planche, and M. N. D. Cordeiro. 2014. "Computational Ecotoxicology: Simultaneous Prediction of Ecotoxic Effects of Nanoparticles under Different Experimental Conditions." *Environment International* 73: 288–294. doi:10.1016/j.envint.2014.08.009.
- Mu, Y., F. Wu, Q. Zhao, R. Ji, Y. Qie, Y. Zhou, Y. Hu, et al. 2016. "Predicting Toxic Potencies of Metal Oxide Nanoparticles by Means of nano-QSARs." *Nanotoxicology* 10 (9): 1207–1214. doi:10.1080/17435390.2016.1202352.
- Ortega-Tenezaca, B., and H. González-Díaz. 2021. "IFPTML Mapping of Nanoparticle Antibacterial Activity vs. Pathogen Metabolic Networks." *Nanoscale* 13 (2): 1318–1330. doi:10.1039/D0NR07588D.
- Osler, G. H., and M. Sommerkorn. 2007. "Toward a Complete Soil C and N Cycle: Incorporating the Soil Fauna." *Ecology* 88 (7): 1611–1621. doi:10.1890/06-1357.1.
- Pathakoti, K., M. J. Huang, J. D. Watts, X. He, and H. M. Hwang. 2014. "Using Experimental Data of *Escherichia coli* to Develop a QSAR Model for Predicting the Photo-Induced Cytotoxicity of Metal Oxide Nanoparticles." *Journal of Photochemistry and Photobiology. B, Biology* 130: 234–240. doi:10.1016/j.jphotobiol.2013.11.023.
- Puzyn, T., B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski. 2011. "Using Nano-QSAR to Predict the Cytotoxicity of Metal Oxide Nanoparticles." *Nature Nanotechnology* 6 (3): 175–178. doi:10.1038/nnano.2011.10.
- Puzyn, T., D. Leszczynska, and J. Leszczynski. 2009. "Toward the Development of "nano-QSARs": Advances and Challenges." *Small* 5 (22): 2494–2509. doi:10.1002/smll.200900179.
- Roy, J., P. K. Ojha, and K. Roy. 2019. "Risk Assessment of Heterogeneous TiO_2 -Based Engineered Nanoparticles (NPs): A QSTR Approach using Simple Periodic Table Based Descriptors." *Nanotoxicology* 13 (5): 701–716. doi:10.1080/17435390.2019.1593543.
- Roy, K., I. Mitra, S. Kar, P. K. Ojha, R. N. Das, and H. Kabir. 2012. "Comparative Studies on Some Metrics for External Validation of QSPR Models." *Journal of Chemical Information and Modeling* 52 (2): 396–408. doi:10.1021/ci200520g.

- Roy, K., S. Kar, and R. N. Das. 2015. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Academic Press.
- Roy, K., R. N. Das, P. Ambure, and R. B. Aher. 2016. "Be Aware of Error Measures. Further Studies on Validation of Predictive QSAR Models." *Chemometrics and Intelligent Laboratory Systems* 152: 18–33. doi:[10.1016/j.chemolab.2016.01.008](https://doi.org/10.1016/j.chemolab.2016.01.008).
- Roy, K., S. Kar, and P. Ambure. 2015. "On a Simple Approach for Determining Applicability Domain of QSAR Models." *Chemometrics and Intelligent Laboratory Systems* 145: 22–29. doi:[10.1016/j.chemolab.2015.04.013](https://doi.org/10.1016/j.chemolab.2015.04.013).
- Sizochenko, N., A. Mikolajczyk, K. Jagiello, T. Puzyn, J. Leszczynski, and B. Rasulev. 2018. "How the Toxicity of Nanomaterials towards Different Species Could Be Simultaneously Evaluated: A Novel Multi-Nano-Read-across Approach." *Nanoscale* 10 (2): 582–591. doi:[10.1039/c7nr05618d](https://doi.org/10.1039/c7nr05618d).
- Speck-Planche, A., V. V. Kleandrova, F. Luan, and M. N. D. S. Cordeiro. 2015. "Computational Modeling in Nanomedicine: Prediction of Multiple Antibacterial Profiles of Nanoparticles Using a Quantitative Structure-Activity Relationship Perturbation Model." *Nanomedicine* 10 (2): 193–204. doi:[10.2217/nnm.14.96](https://doi.org/10.2217/nnm.14.96).
- Urista, D. V., D. B. Carru , I. Otero, S. Arrasate, V. F. Quevedo-Tumaili, M. Gestal, H. Gonz lez-D  az, and C. R. Munteanu. 2020. "Prediction of Antimalarial Drug-Decorated Nanoparticle Delivery Systems with Random Forest Models." *Biology* 9 (8): 198. doi:[10.3390/biology9080198](https://doi.org/10.3390/biology9080198).
- Venkatasubramanian, V., and A. Sundaram. 2002. *Genetic Algorithms: Introduction and Applications*. Weinheim: Wiley. doi:[10.1002/0470845015.cga003](https://doi.org/10.1002/0470845015.cga003).

PAPER

[View Article Online](#)
[View Journal](#)

Cite this: DOI: 10.1039/d2en00303a

Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (Me_{OX} NPs) used in nanomedicines: a mechanistic QSAR approach†

Joyita Roy and Kunal Roy *

Nanotechnology has become a fast-growing area of science. Recently, there has been an increase in the use of metal oxide nanoparticles (Me_{OX} NPs) due to their unique morphology and physical properties. The application of Me_{OX} NPs in nanomedicines has opened up new perspectives for medical devices and development in health care. Nanomedicines and nanotoxicity are complementary to each other. Lack of sufficient data and low-adequacy experimental setups obstruct the assessment of their potential risk to human cells. Regulatory authorities have also encouraged the application of the quantitative structure–activity relationship (QSAR) approach towards the safe design of nano-based products. Therefore, in the present computational study, multiple linear regression (MLR) modeling with 16 Me_{OX} NP datasets and partial least squares (PLS) regression modeling with 15 Me_{OX} NP datasets (117 data points with different physical features and conditions) were performed to determine the cytotoxicity of NPs towards *HaCaT* cells. Easily interpretable periodic table descriptors (1st and 2nd generation) and some previously reported physical descriptors were employed for modeling. The genetic algorithm (GA) followed by best subset selection (BSS) technique was applied before the model development. The final MLR model (4 descriptors) and PLS (2 LVs) model for the two datasets was rigorously validated with various stringent validation criteria following the strict Organisation for Economic Co-operation and Development (OECD) guidelines. The statistical data of the MLR model ($Q_{(LOO)}^2 = 0.83$, $R^2 = 0.95$) and PLS model ($Q_{(LOO)}^2 = 0.73$, $R^2 = 0.76$) were robust and reliable. Features like electronegativity, formation of a metal cation, increase in solubility, and dose were important for the mechanism of toxicity. Therefore, these models can be efficient tools for designing modern materials with optimal properties and minimal potential toxicity.

Received 29th March 2022,
Accepted 18th July 2022

DOI: 10.1039/d2en00303a

rsc.li/es-nano

Environmental significance

Nano-technological advancements in biomedical applications have vast and promising future avenues. However, it is believed that nano-sized metal oxide nanoparticles (Me_{OX}NPs) can facilitate unfavorable absorption through natural barriers, resulting in toxicological changes that may differ from those of their pristine forms. Despite these properties, Me_{OX}NPs, which are considered boons to mankind, can also present potential risks. Exposure to Me_{OX}NPs can induce adverse and undesirable effects to non-targeted organisms in the environment, raising more concerns regarding potential environmental risks from nanomedicines. This work highlights the potential features that influence Me_{OX}NP toxicity and explains possible mechanisms, which may elucidate their risks to environmental and human health.

1. Introduction

The past few decades have witnessed the evaluation of “nano” species in the field of science and technology. Recently, there has been increasing application of metal oxide nanoparticles due to their compelling properties. A diverse array of

nanomaterial-based goods are produced every year. Approximately 1015 different types of products are produced and marketed as per a report in 2009,¹ and this number is expected to increase steadily. According to the project on Emerging Nanotechnologies [<https://www.nanotechproject.org/inventories/consumer/>], many new nanotechnology-based products are introduced in the market on a weekly basis, and the annual turnover from such products would be \$2.6 trillion. According to previous studies, nanoparticles (NPs) have different useful properties than the materials in their conventional forms, which has greatly increased their applications. Notably, nanomedicines² are an application of nanotechnology that has

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India.

E-mail: kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in;

Fax: +91 33 2837 1078; Tel: +91 98315 94140

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2en00303a>

ignited hope in the field of medical sciences. The implication of nanostructures in this field has not only created advanced opportunities, but also generated perspectives on issues such as targeting tissues or cells, controlled release of a drug at a specific target, and detection of diseases at an early stage. Nanotechnology, nanomedicines, and nanotoxicology go hand in hand. The increased use of nanoparticles (NPs) in pharmaceutical sciences has imposed major concerns about their possible nanotoxicity and has emerged as a major sub-discipline of nanotechnology. Nanotoxicology refers to the study of the interaction of nanostructures with biological systems to elucidate the physical and chemical properties of nanoparticles that cause toxic biological responses. Determination of the long-term toxic effects of NPs may consume a lot of time in clinical trials involving animal models and human volunteers. Toxicity and biokinetics information for nano-products, along with information regarding unintended human exposure or intended delivery for medical purposes, is necessary to determine the perceived risks of nanomaterials. The dose of NPs should also be taken into consideration for the determination of their toxicity. It was shown in the study by Pisanic II *et al.*³ that when PC12 cells were exposed to different doses of Fe₂O₃ MNPs, there was a decrease in cell viabilities. Vanadium oxide (V₂O₃) was ten times more potent in the reduction of cell viability even at a concentration of 10 µg mL⁻¹. The induction of heme oxygenase 1 (HO-1) in ECV304 cells and reduction of protein levels by nanoscale V₂O₃ NPs in epithelial cells (A549) was found to occur in a dose-dependent manner.

Considering these facts, the determination of toxicity for nanomedicines is very important. Hence, regulatory bodies are searching for ways to assess their risk before their application as pharmaceuticals.⁴ The use of nanoinformatics in combination with nanomaterials (NMs) has attracted the attention of many scientists, leading them to create comprehensive databases comprising the physical and chemical features of the materials, based on which quantitative structure–activity relationship (QSAR) models are generated.⁵ *In silico* toxicological methods such as QSARs play an efficient role to complement and reduce animal testing. QSARs can also be used as an excellent tool for determining the physicochemical properties of NPs computationally, involving various types of descriptors. These studies can be utilized to forecast the properties or toxicity of NMs that are yet to be explored or synthesized.⁶ In the field of nanotoxicity, only a few studies have been carried out using QSAR models, and only a few reviews concerning the risks of NPs have been reported to date.^{7–9} Furthermore, to date, all QSAR models have predicted the toxicity of NPs without giving proper information regarding important factors such as dose, hydrodynamic size, *etc.*, which can be determining factors in influencing toxicity.

Therefore, in this context, here we have tried to develop nano-QSAR models¹⁰ that refer to the statistically significant relationships of the toxicity of metal oxide nanoparticles (MeOx-NPs) in a human keratinocyte cell line (*HaCaT* cells) with the nanoparticle structures. Most of the existing works on nanotoxicity modeling have focused mainly on prokaryotic

systems (*e.g.*, *Escherichia coli*), while very few toxicity models are available for eukaryotic systems. Information related to the toxicity of metal oxide (MeOx) NPs to the *HaCaT* cell line is also limited.^{11–13} Therefore, the present work mainly focuses on modeling the toxicity of MeOx toward the *HaCaT* cell line. The *HaCaT* cell line (skin cells), an immortalized epithelial cell line, was chosen to evaluate the mechanism of toxicity because it exhibits a similar nature to human keratinocytes. Skin also plays a role of prime importance against toxins or pathogens, and it is the largest organ of the body with a large surface area. Thus, it is ideal for studying dermal toxicity. Exploring the *HaCaT* cell toxicity of nanoparticles through a QSAR modeling approach appears to be of importance due to its potential application in developing safe nanomedicines.

2. Methods and materials

2.1 Dataset

In the present work, two different cytotoxicity datasets of metal oxides were employed for performing nano-QSAR modeling towards the *HaCaT* cell line, which are presented as Case study 1 and Case study 2. The cytotoxicity data were collected from previously published papers.^{14,15}

Case study 1. The first dataset comprises 18 nano-metal oxides taken from a previously published paper.¹⁴ Their cytotoxicity is expressed the logarithm of the inverse molar lethal concentration *i.e.*, $\log(1/LC_{50})$. Here, we removed two MeOx-NPs because of their outlier behavior in the initial analysis (Table S1†). Furthermore, another dataset was used for modeling in Case study 2 to observe the influence of the dose on the cytotoxicity endpoint.

Case study 2. The second dataset consists of 15 MeOx NPs (117 data points total) taken from the Safe and Sustainable Nanotechnology (S2NANO) database built by a research group in South Korea.¹⁵ This database deals with toxicity information for various nanomaterials that has been gathered from different experimental reports or literature. The cytotoxicity is expressed as percentage (%) cell death in Table S2 and S3†. The *HaCaT* cells were exposed to NPs with various physicochemical properties such as core size, hydrodynamic size, surface charge, and dose ranging from 0–1500 ppm.

2.2 Molecular descriptors

Nano-QSAR modeling was performed using the basic information for the metal oxides taken from the periodic table to encode the cytotoxicity and percentage cell death against the *HaCaT* cell line. For Case study 1, a total of 69 descriptors were calculated using the basic information of the metal oxides from the periodic table directly, and other descriptors (2nd generation periodic table descriptors) were calculated from the primary descriptors.¹⁶ Additionally, a few periodic table descriptors were calculated with the help of the Elemental Descriptor Calculator software available from (<https://dtclab.webs.com/software-tools>). For Case Study 2, a total of 73 descriptors were calculated, including periodic-table-based

descriptors as well as some physicochemical descriptors taken from the experimental conditions.¹⁴ The list of collected and derived periodic table descriptors is given in Table 1. The advantage of using periodic-table-based descriptors over other descriptors is that they can be easily calculated without using any sophisticated software. The details of the data sets are given in ESI-1 and 2.†

2.3 Data set splitting

Selecting a proper training set and a test set is vital for the development of a robust QSAR model. The combination of metal oxides is different for each data point (Case study 2). So, in the present study, after pretreatment and the removal of noisy descriptors from both datasets, the dataset for each case was divided with the dataset division software freely available from the DTC Lab site (<https://dtclab.webs.com/software-tools>). In Case study 1: eleven (11) metal oxides were kept in the training set and five (5) metal oxides were present in the test set. In Case study 2, one hundred and six (94) metal oxide data points were present in the training set and twenty-six (23) metal oxide data points were present in the test set for model development and validation respectively.

2.4 Chemometric tools

Model development is a crucial job for the QSAR modeler. Therefore, after data pretreatment, the datasets were subjected to feature selection using the genetic algorithm (GA)¹⁷ followed by model development using the Best Subset Selection (BSS) software (<https://dtclab.webs.com/software-tools>) based on the MAE criteria.¹⁸ The dataset containing sixteen metal oxides containing the cytotoxicity data towards *HaCaT* cell line used periodic table descriptors, and the final model with the best combination of four descriptors was selected. In the second case, along with the periodic-table-based descriptors, some physicochemical descriptors were also included. Partial least squares (PLS) regression was performed using the four descriptors obtained in the MLR model. The PLS model was selected as the final model with two latent variables. The whole workflow is represented in Fig. 1.

2.5 Validation metrics

Validation of the developed models was done using both internal and external validation metrics. The internal quality metric R^2 determines the fitting potential of the models, whereas the Q_{LOO}^2 metric¹⁹ measures the predictive ability of the model based on the training set compounds. Although Q_{LOO}^2 measures the robustness of the model, it may not be sufficient to determine the performance of the model on a new set of compounds. The external validation metrics ($Q_{F_1}^2$, $Q_{F_2}^2$) deal with the quality of the predictions for the test set compounds. Additionally, the r_m^2 metrics and root mean squared error (RMSE) were also computed. The accepted threshold value for both the internal and external metrics is 0.5. The results obtained from both datasets confirmed the robustness of the developed models, as depicted in Table 2.

2.6 Applicability domain (AD)²⁰

The applicability domain (AD) in QSAR indicates the physicochemical, biological, and chemical space knowledge of the training set based on which the model was developed. The model is capable of predicting new compounds when they belong to the same chemical domain.²⁰ The purpose of the AD is to confirm the model assumptions. To investigate the AD of the training set, one can analyze the features of the multivariate descriptor information of the training set compounds, or the similarity-based metrics. When a compound is highly dissimilar to the training set compounds, the reliability of that particular compound is low. The concept of AD was used in the final model to avoid unjustified extrapolation of the predictions. In this study, we checked the applicability domain for the MLR model with the leverage approach^{21,22} using STATISTICA software,²³ and for the PLS model, we used the DModX (distance to model) approach²⁴ using SIMCA-P software²⁵ at the 99% confidence level. Through the leverage approach, it is possible to determine whether a compound is within the chemical space or outside it. If the leverage values (h) of the compounds are less than the critical value (h^*), then the prediction can be considered reliable, and *vice versa*. The critical value (h^*) is calculated as $h^* = 3p'/n$, where p' represents the modeled variables plus one and n is for the number of compounds used in model development. Outliers can easily be detected by plotting the leverages (Hat diagonal) against the cross-validated standardized residuals (R), *i.e.*, a William's plot. Compounds with a cross-validation standardized residual greater than three standard deviation units, *i.e.*, $>2.5\sigma$ can be considered to be outliers (Fig. 2).

3. Results and discussion

Based on the toxicity data and calculated descriptors of the two datasets, we tried to develop statistically significant QSAR models based on the regression approach to successfully predict cytotoxicity toward the *HaCaT* cell line. The MLR model generated from the first dataset (model 1) consists of only periodic table descriptors, which were able to describe the cytotoxicity successfully without the need to calculate quantum chemical descriptors. In the case of the second dataset (model 2), the PLS model having two latent variables was obtained using periodic-table-based descriptors and physicochemical descriptors.

Model 1 (MLR)

$$\log(1/\text{LC}_{50}) = 1.46 + 0.0117 \text{ Amt rad (pm)} + 0.00520 \text{ Thermal cond (W mk}^{-1}\text{)} - 0.000123 \text{ M.P (K)} - 0.301\chi_{\text{ox}}$$

$$n_{\text{training}} = 11, R^2 = 0.95, R_{\text{adj}}^2 = 0.92, Q^2 = 0.83, S = 0.0953388, F = 29.03, \text{PRESS} = 0.186201$$

$$n_{\text{test}} = 5, Q_{F_1}^2 = 0.73, Q_{F_2}^2 = 0.62, \text{MAE based criteria} = \text{Good}$$

Table 1 List of the descriptors and their generation method

No.	Mathematical expression	Description
1st generation descriptors		
1	D1_Metals	Total no. of metals
2	D2_SemiMetal	Total no. of semi-metals
3	D3_HeteroNonMetals	Total no. of non-metallic heteroatoms (N, O, F, P, S, Cl, Se, Br, I, At) excluding hydrogen
4	Metals_SumIP	Sum of ionization potential energy (kJ mol ⁻¹) of metals
5	SemiMetals_SumIP	Sum of ionization potential energy (kJ mol ⁻¹) of semi-metals
6	Electrons_ActiveM	No. of electrons of active metal
7	AtWt_ActiveM	Atomic wt. of active metal
8	MassNo_ActiveM	Mass no. of active metal
9	Neutons_ActiveM	Number of neutrons of active metal
10	IP_ActiveM	Ionization potential energy (kJ mol ⁻¹) of active metal
11	X_ActiveM	Electronegativity of active metal
12	VWR_ActiveM	van der Waals radius of active metal
13	Electrons_Active_SM	No. of electrons of active metalloid/semi-metal
14	AtWt_Active_SM	Atomic wt. of active metalloid/semi-metal
15	MassNo_Active_SM	Mass no. of active metalloid/semi-metal
16	Neutons_Active_SM	Number of neutrons of active metalloid/semi-metal
17	IP_Active_SM	Ionization potential energy of active metalloid/semi-metal
18	X_Active_SM	Electronegativity of active metalloid/semi-metal
19	VWR_Active_SM	van der Waals radius of active metalloid/semi-metal
20	Electrons_Active_NM	No. of electrons of active non-metal
21	AtWt_Active_NM	Atomic wt. of active non-metal
22	MassNo_Active_NM	Mass no. of active non-metal
23	Neutons_Active_NM	Number of neutrons of active non-metal
24	IP_Active_NM	Ionization potential energy of active non-metal
25	X_Active_NM	Electronegativity of active non-metal
26	VWR_Active_NM	van der Waals radius of active non-metal
27	SuM_Active_M_SM	Summation of number of metals and semi-metals
28	SuMElectrons_Active_M_SM	Summation of electrons of active metal and semi-metal
29	SuMATWt_Active_M_SM	Summation of at. wt. of active metal and semi-metal
30	SuMMassNo_Active_M_SM	Summation of mass no. of active metal and semi-metal
31	SuMNeutons_Active_M_SM	Summation of neutrons of active metal and semi-metal
32	χ_{Ox}	Oxidation number
33	N_{oxygen}	Several oxygen atoms are present in the metal oxide
34	Molar volume	
35	Atomic rad	Atomic radius
36	Electronegativity	—
37	Thermal conductivity	—
38	Melting point	—
39	Electron affinity	—
40	Density	—
41	Mol wt	Molecular weight
42	Cationic charge	—
43	$\sum \chi$	Total metal electronegativity in specific metal oxide
44	$\sum \chi/n\text{O}$	Total metal electronegativity in specific metal oxide relative to oxygens
2nd generation descriptors		
1	Z_{metal}	Atomic number of metals
2	Z_{vmetal}	Valence electron of metal
3	PN_{metal}	Period number of metals
4	$\lambda = (Z_{\text{metal}} - Z_{\text{vmetal}})/Z_{\text{vmetal}}$	Core environment of metal defined by the ratio of the number of core electrons to the number of valence electrons
5	$\mu = 1/(\text{PN}_{\text{metal}})$	—
6	V_{metal}	Valence of metal
7	$\alpha_{\text{metal}} = \lambda \times \mu$	—
8	$\sum \alpha_{\text{metal}} = \alpha_{\text{metal}} \times N_{\text{metal}}$	—
9	$\sum \alpha_{\text{oxy}} = N_{\text{oxy}} \times 0.33$	—
10	$\sum \alpha = \sum \alpha_{\text{metal}} + \sum \alpha_{\text{oxy}}$	The core count gives a measure of the molecular bulk
11	$\epsilon_{\text{metal}} = -\alpha_{\text{metal}} + (0.3 \times Z_{\text{vmetal}})$	Electronegativity count of metal
12	$\epsilon_{\text{oxy}} = -\alpha_{\text{oxy}} + (0.3 \times Z_{\text{oxyv}})$	Electronegativity count of oxygen
13	$\sum \epsilon = \epsilon_{\text{metal}} \times N_{\text{metal}} + \epsilon_{\text{oxy}} \times N_{\text{oxy}}$	Electronegativity count of total metal oxide
14	$\sum \epsilon/N$	Sum of epsilon relative to the number of atoms in the molecule
15	$(\sum \alpha)^2$	Square of the summation of alpha, gives a measure of the molecular bulk
16	$(\sum \epsilon/N)^2$	Squared sum of epsilon relative to the number of atoms

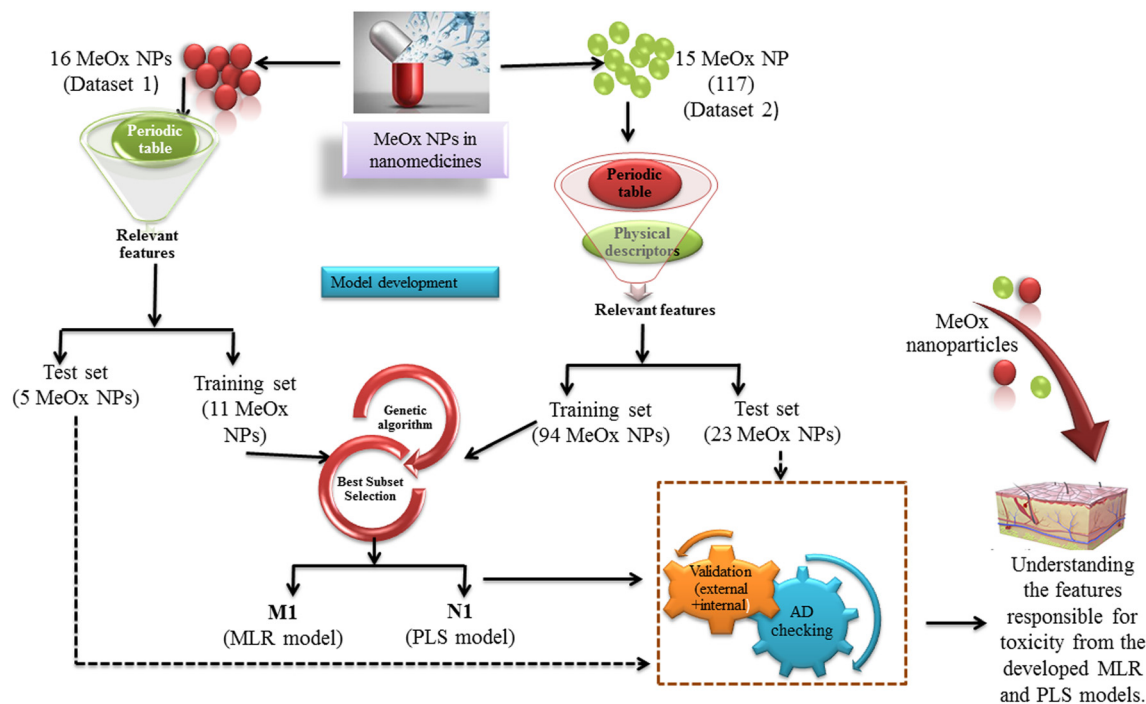


Fig. 1 Flowchart of the work.

Table 2 Statistical quality and validation parameters obtained from the developed MLR and PLS models

Types of models	Descriptors	Division	Model R^2	Training set statistics					Test set statistics				
				Model R^2	$Q_{(LOO)}^2$	$r_{m(LOO)}^2$	$\Delta r_{m(LOO)}^2$	F value	R_{pred}^2 or $Q_{F_1}^2$	$Q_{F_2}^2$	CCC	$r_{m(test)}^2$	$\Delta r_{m(test)}^2$
MLR	Case 1 (16 MeOx NPs)	4 (four)	Train: 11 Test: 5	0.95	0.83	0.773	0.037	29.03	0.73	0.62	0.79	0.271	-0.505
	Case 2 (15 MeOx NPs)	4 (four)	Train: 94 Test: 23	0.76	0.73	0.63	0.038	71.69	0.69	0.63	0.81	0.61	0.123
PLS	P1 (case 2)	2 (two)	Train: 94 Test: 23	0.76	0.73	0.63	0.038	71.69	0.69	0.63	0.81	0.61	0.123

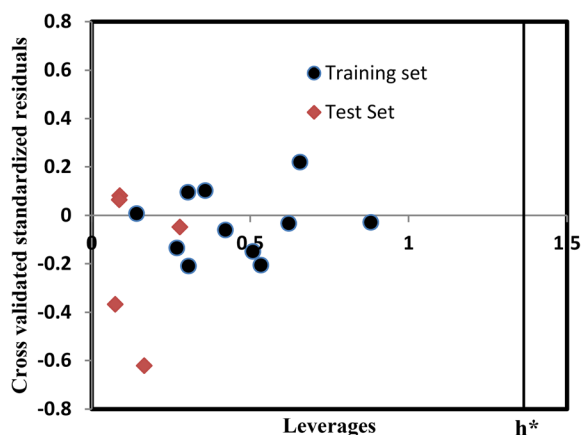


Fig. 2 William's plot of the MLR model.

Model 2 (PLS)

$$\% \text{ Cell death} = 113.295 + 0.039 \text{ Dose} - 21.297 \text{ SuM_Active_M_SM} - 15.400 \text{ cationic charge} - 1.194 \text{ Heat of Fusion (kJ mol}^{-1}\text{)}$$

$$n_{\text{training}} = 94, R^2 = 0.76, Q^2 = 0.73, LV = 2, n_{\text{test}} = 23, Q_{F_1}^2 = 0.69$$

Taking the *HaCaT* cell cytotoxicity as the response variable, the regression models (MLR and PLS) yielded determination coefficients (R^2) of 0.95 and 0.76 respectively. The whole work was divided into two parts; in the first part, we tried to develop regression-based models (MLR) using only periodic table descriptors and determine the features responsible for the toxicity of the particular metal oxides, and in the second part, another set of 15 metal oxides (117 data points total) was utilized to develop a PLS model using periodic-table-based descriptors along with physicochemical descriptors to determine the influence of these descriptors on the toxicity to living cells. The PLS model was constructed with all four descriptors from the obtained MLR model. A PLS model can handle data that are strongly correlated or noisy. The model finds new variables (*i.e.*, latent variables) that are linear combinations of the original variables.

The models were validated according to the Organisation for Economic Co-operation and Development (OECD) principles, as properly validated models can be used for

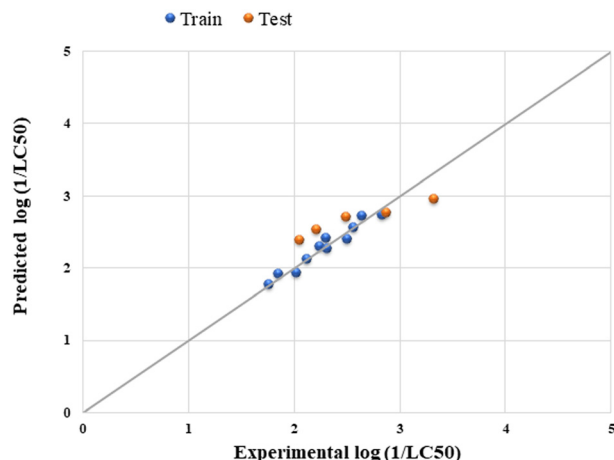


Fig. 3 Scatter plot of model 1.

mechanistic interpretation.²⁶ The values of all classical metrics were within the acceptable limits for both cases, as presented in Table 2. The descriptors obtained from both models are listed in Table 3 and explained in detail below. For Model 1, the William's plot is shown in Fig. 2 while the scatter plot is shown in Fig. 3. The AD for both the models (MLR and PLS) was checked using the Leverage approach²¹ and DModX, *i.e.*, the distance to model in the X-space with

SIMCA 16.0.2 software, respectively. No compound was outside the AD as per the William's plot in Fig. 2 and the DModX approach for the training and test set in Fig. 4 and 5.

3.1 Interpretations of different plots

The variable importance plot (VIP) of the PLS model (model 2) highlights the most influential descriptors of the Me_{Ox} in terms of contributing to the cytotoxicity of the *HaCaT* cells. In the case of the PLS model, the most contributing descriptors according to the plot were dose, cationic charge, and SuM_Active_M_SM , followed by the descriptor heat of fusion (Fig. 6).

The loading plot gives detailed information regarding the X-variable and Y-variables of the PLS plot, as represented in Fig. 7. The loading plot was developed with two PLS components. From Fig. 7, we can infer that the dose of the Me_{Ox} plays a dominant role in determining the toxicity to *HaCaT* cells. While evaluating the plot, it was observed that the descriptors located far away from the origin have more impact on the Y response of the model.

The coefficient plot shows the contributions of the modeled descriptors. Except for the dose descriptors, all other descriptors (heat of fusion, SuM_Active_M_SM , Cationic charge) were found to contribute negatively to the response in the PLS model of the 15 Me_{Ox} NP dataset (Fig. 8).

Table 3 Detailed descriptions of the obtained descriptors from the two developed MLR models (model 1 and model 2)

Serial no.	Descriptor	Description
Model 1		
1.	Atomic rad (pm)	This descriptor represents the size of the atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons
2.	Thermal cond (W mK^{-1})	The transfer of energy is due to the random motion of molecules across a temperature gradient
3.	Melting point	The temperature at which a material changes state from solid to liquid
4.	χ_{ox}	This represents the oxidation number of the metal atom
Model 2		
1.	Heat of fusion	The heat of enthalpy is the change of enthalpy of a fixed amount of metal resulting from the supply of energy, typically in the form of heat
2.	Dose	A quantity of a substance taken or recommended to be taken at a particular time
3.	SuM_Active_M_SM	Summation of metals and semi-metals in the nanoparticles
4.	Cationic charge	Ion that bears a positive charge formed when an atom loses its electron

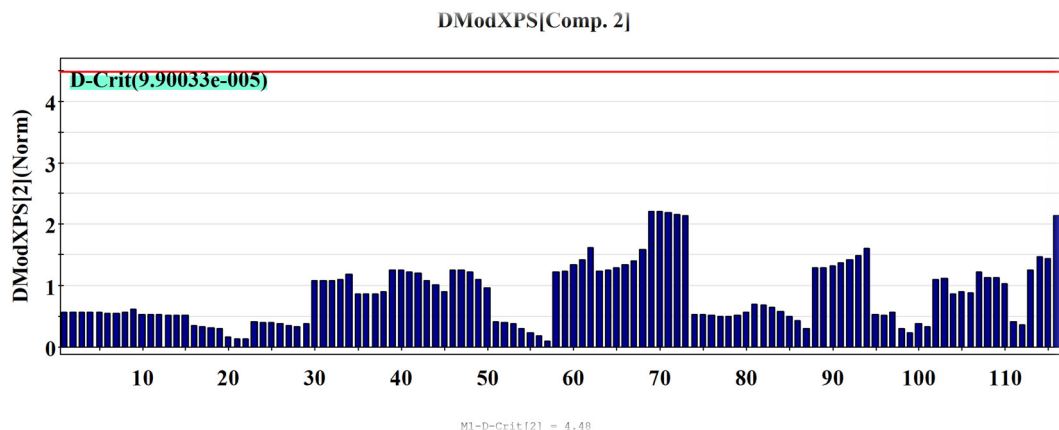


Fig. 4 Applicability domain of model 2 for the training set (PLS model).

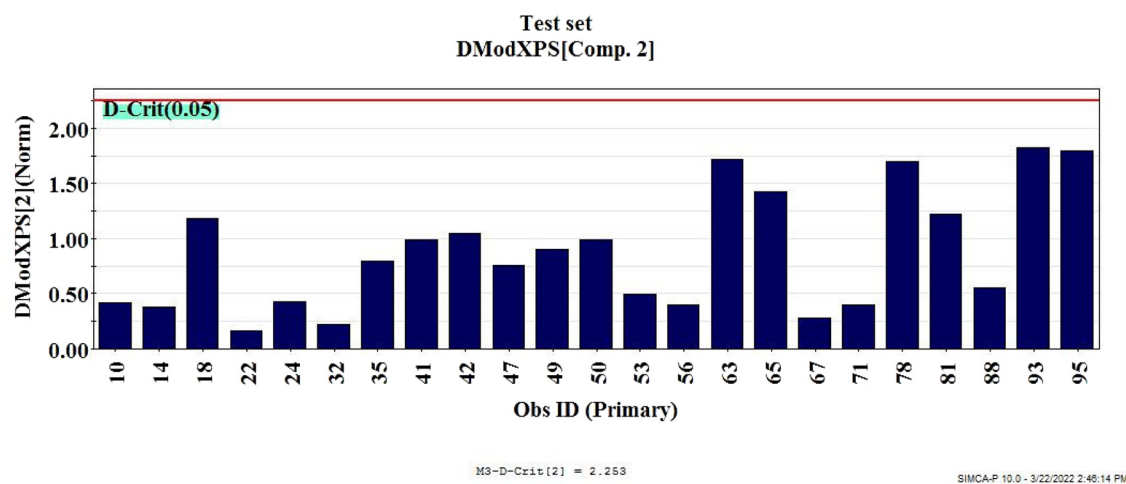


Fig. 5 Applicability domain of model 2 for the test set (PLS model).

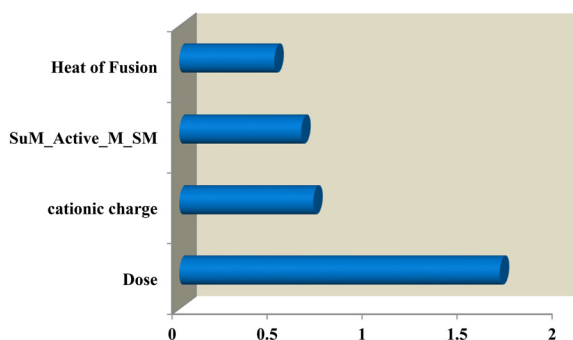


Fig. 6 Variable importance plot of model 2 (PLS model).

Randomization of the model authenticates the robustness of the developed model and ensures that it was not developed by chance correlation. Randomized models are developed by multiple shuffling of X and Y variables. Here, 100 permutations were carried out to generate the models. The

developed random models should have poor statistics to show it is a nonrandom model. The correlation between the original Y -vector and permuted Y -vector *versus* the cumulative R_y^2 and cumulative Q_y^2 plots is represented in Fig. 9. The plot shows that the model is nonrandom as well as robust since R_y^2 intercept = -0.0211 and Q_y^2 intercept = -0.179 , and one can easily predict the cytotoxicity of the MeO_xNPs from the developed PLS model.

3.2 Descriptor interpretation

1st case study. The 1st generation periodic-table-based descriptor **atomic radius (atomic rad, pm)** stands for the distance between the nucleus and the valence electron. The **positive** coefficient of the descriptor suggests that with increased atomic radius, the cytotoxicity towards the *HaCaT* cell line increases, as observed for the nanoparticle WO_3 (atom rad = 210 (pm), $\log(1/\text{LC}_{50}) = 2.56$) and *vice versa* in the case of the nanoparticle Al_2O_3 (atom rad = 125 (pm),

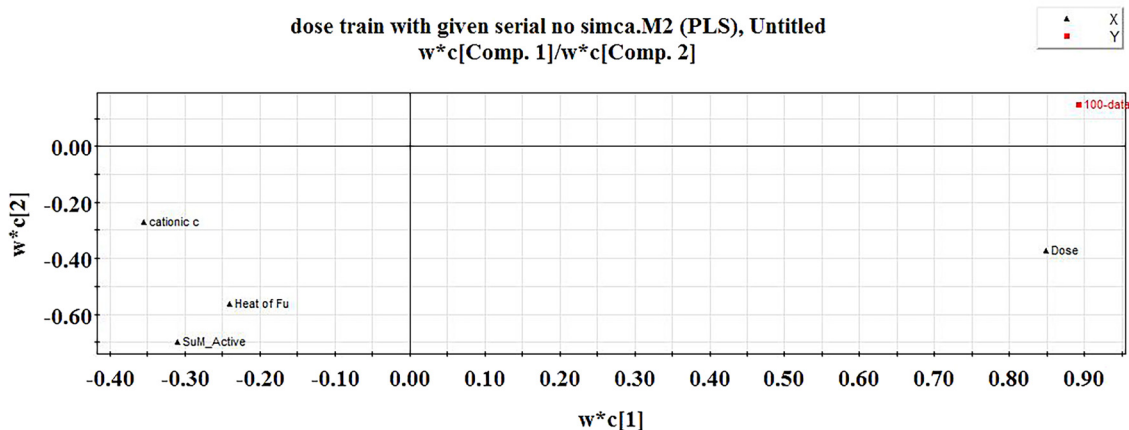


Fig. 7 Loading plot of the PLS model.

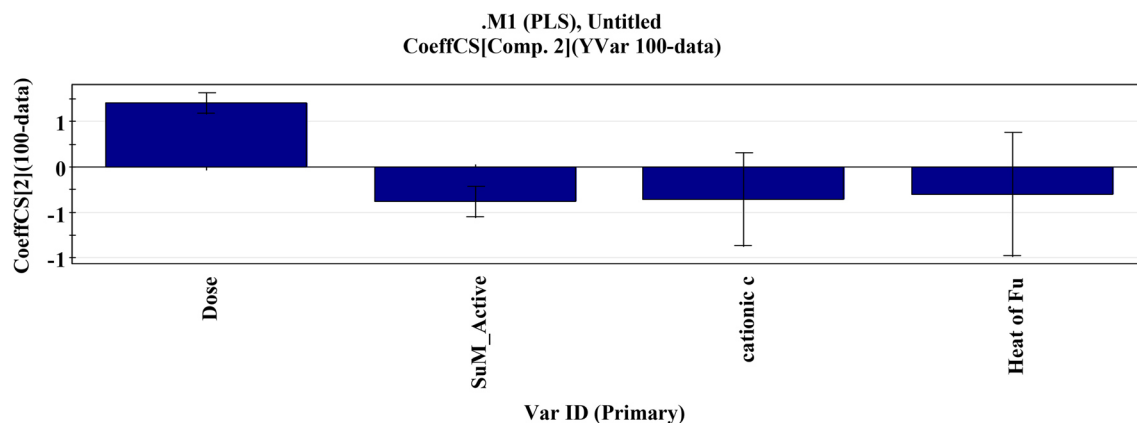


Fig. 8 Coefficient plot of PLS model.

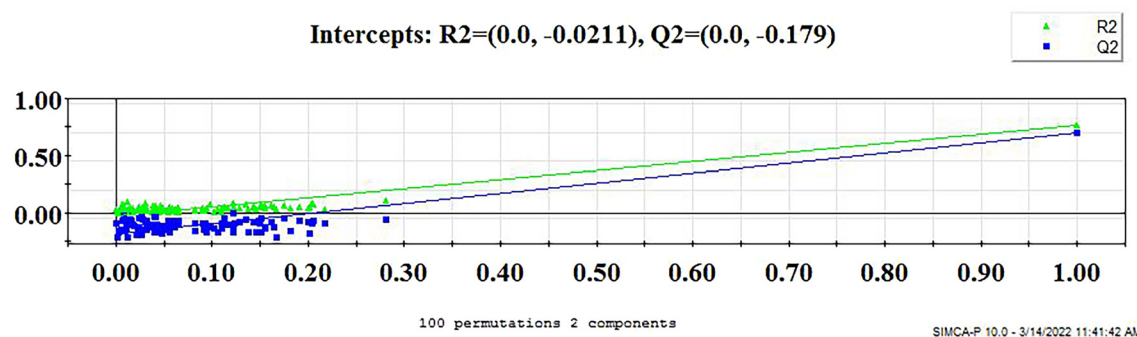


Fig. 9 Randomization plot of the PLS model.

$\log(1/LC_{50}) = 1.85$). This happens because as the atomic radius increases, the valence electron in the outermost shell continually moves away from the nucleus, which means that more atomic shell is being added. The addition of shells

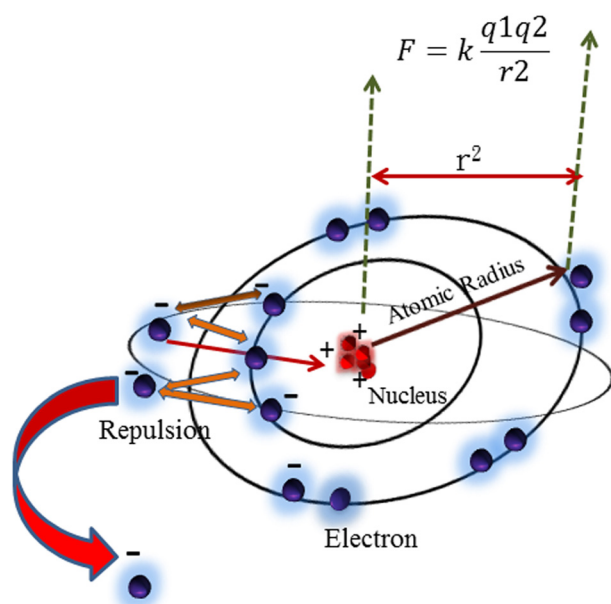


Fig. 10 Electron removal as the atomic radius increases.

shields the valence electron in the outermost shell. The electron in the outermost shell does not feel the counter-pull of the nucleus, and the presence of electrons in the inner shells repels the similarly charged electron, which makes it easier to isolate and remove from its shell (Fig. 10). The removal of the electron converts the metal into its cationic form. It is this cationic form of the metal that initiates the production of oxidative radicals in the cell, which is detrimental to cell viability. Studies^{27,28} on ROS-mediated mechanisms have shown that cellular uptake of cationic NPs accelerates the depolarization of the cell membrane and facilitates calcium channel opening, which stimulates the mitochondria to generate more oxidative stress.

Another intrinsic property of metals is **thermal conductivity** (thermal cond, $W\ mK^{-1}$), defined as the transfer of energy due to the random motion of molecules across a temperature gradient. It is the measurement of the amount of heat that passes across a material per unit area of the material per unit time. According to band theory, the atoms in the metal crystal are close to each other, which leads to overlapping of their molecular orbitals. This phenomenon, which is depicted in Fig. 11, suggests the presence of a large number of electrons in a small piece of metal, and their closeness prefers them as bands.²⁹ In band theory, there are two types of bands, one partially filled with electrons, known as the conduction band, and the other filled with electrons, known as the valence

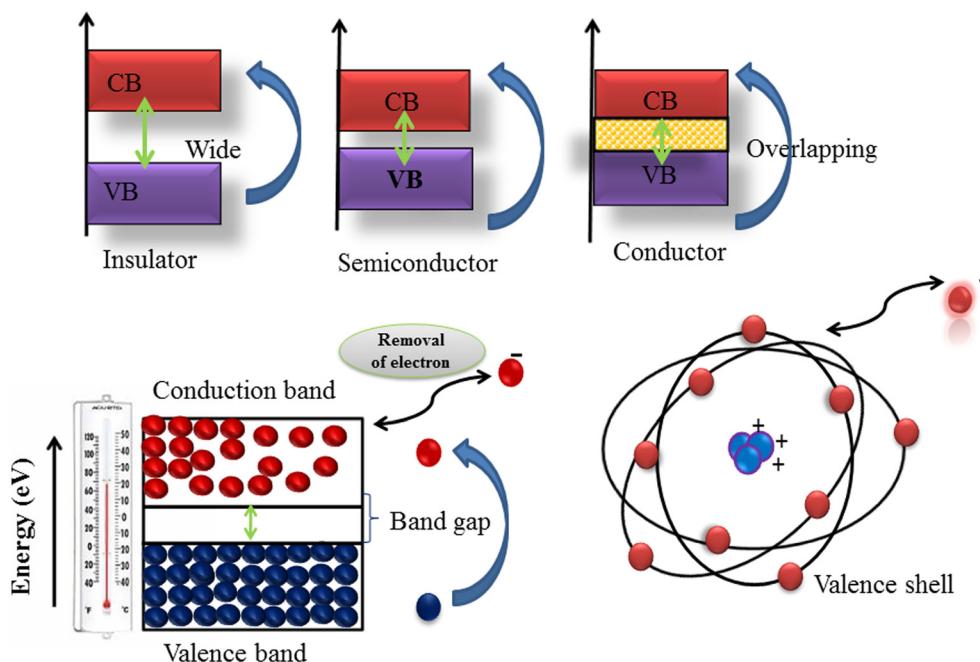


Fig. 11 Electron removal through band theory.

band. The ability of electricity to pass through the metal depends on the proximity of the valence and conduction bands. The **positive** regression coefficient of the descriptor suggests that with increased thermal conductivity, the cytotoxicity also increases, as in the nanoparticle CoO (thermal cond. = 100 W mK^{-1} , $\log(1/\text{LC}_{50}) = 2.85$). The reverse happens for the nanoparticle Al_2O_3 (thermal cond. = 2.37 W mK^{-1} , $\log(1/\text{LC}_{50}) = 1.85$). This happens because with increasing thermal conductivity, the distance between the two bands decreases, and it becomes easier for the electrons to move faster from

the valence band to the conduction band. This redox potential occurring naturally in the cell may be the reason for the toxicity of nanoparticles.

Another property of metals is the **melting point (MP)**, defined as the temperature at which a material changes state from solid to liquid. The melting point of a metal is primarily related to the bond strength. Materials having weaker bonds between atoms tend to have low melting points. The **negative** coefficient of the descriptor suggests that with decreasing melting point, the toxicity of the nanoparticle increases, as

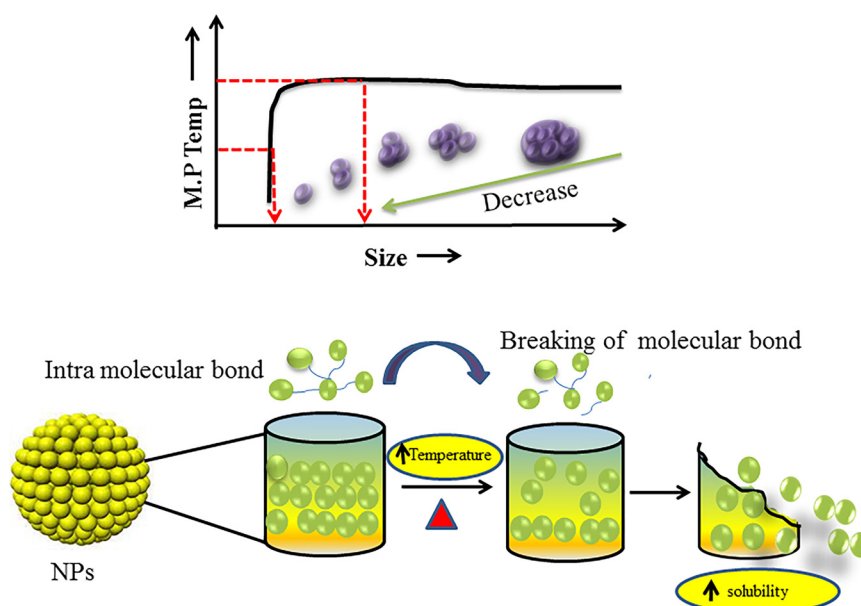


Fig. 12 Increase in solubility with melting point.

observed for the TiO_2 nanoparticle, for which the MP is 1941 (K) and $\log(1/\text{LC}_{50})$ is 1.76. On the contrary, the V_2O_3 nanoparticle with a lower MP (30.7 K) has a $\log(1/\text{LC}_{50})$ value of 2.24. This is because the melting point is associated with the solubility of the metal. With decreasing MP, the solubility increases. The solubility determines its transport to the active site within the cell and thus, the toxicity. Low-melting Me_{Ox} NPs easily dissociate at body temperature (Fig. 12). The presence of heat makes the molecules vibrate faster and break their intra-molecular bonds. Since nanoparticles have fewer bonds compared to their bulk counterparts, it becomes easier to break them.

The periodic table descriptor χ_{ox} represents the oxidation number of the metal. The oxidation number reflects the hypothetical charges in an atom. Electronegativity indicates the distribution of electrons in a molecule, and it can also determine the oxidation number of the atom. The oxidation number indicates the oxidation power of the metal in a particular environment. The transition metals may exist in different oxidation states as they have partially filled d- and f- shells (Fig. 13). An atom with fewer valence electrons will have a lower oxidation state, and atoms with a lower oxidation state are more harmful than those with a higher or stable oxidation state due to their tendency to become oxidized.³⁰ An atom bearing a negative oxidation number has higher electronegativity, whereas an atom with a positive oxidative number will have low electronegativity. The electronegativity is responsible for the catalytic properties of

the cationic form of the metal. The **negative** coefficient of the descriptor indicates that with decreasing descriptor value, the electronegativity increases, and hence, cytotoxicity to the *HaCaT* cells increases as in the case of the nanoparticle TiO_2 ($\chi_{\text{ox}} = 4$, $\log(1/\text{LC}_{50}) = 1.76$), and *vice versa* in case of the nanoparticle CoO ($\chi_{\text{ox}} = 2$, $\log(1/\text{LC}_{50})$ is 2.83). The low-oxidation-state metals are more toxic due to their tendency to be oxidized to a higher oxidation state.

2nd case study. The periodic table descriptor **heat of fusion** (ΔH_f), also known as the heat of enthalpy, is the change in the enthalpy of a fixed amount of metal resulting from the supply of energy, typically in the form of heat. Latent heat is the amount of heat added or removed from the system to cause a phase change in the system (Fig. 14). The latter breaks the intermolecular forces and helps the system to expand by providing the necessary energy. The temperature at which this takes place is called the melting point. Based on the negative coefficient of the descriptor, a decrease in the heat of fusion leads to an increase in cytotoxicity and percentage cell death. For the nanoparticle Sb_2O_3 , the heat of fusion was $19.87 \text{ (kJ mol}^{-1}\text{)}$ and the % cell death was 15.3%. On the other hand, the opposite was observed for the nanoparticle In_2O_3 , for which the heat of fusion was $3.26 \text{ (kJ mol}^{-1}\text{)}$ and the % cell death was 72.7%. The heat of fusion descriptor is inversely proportional to the solubility, which means that with decreasing heat of fusion, the solubility increases, thus increasing the toxicity.

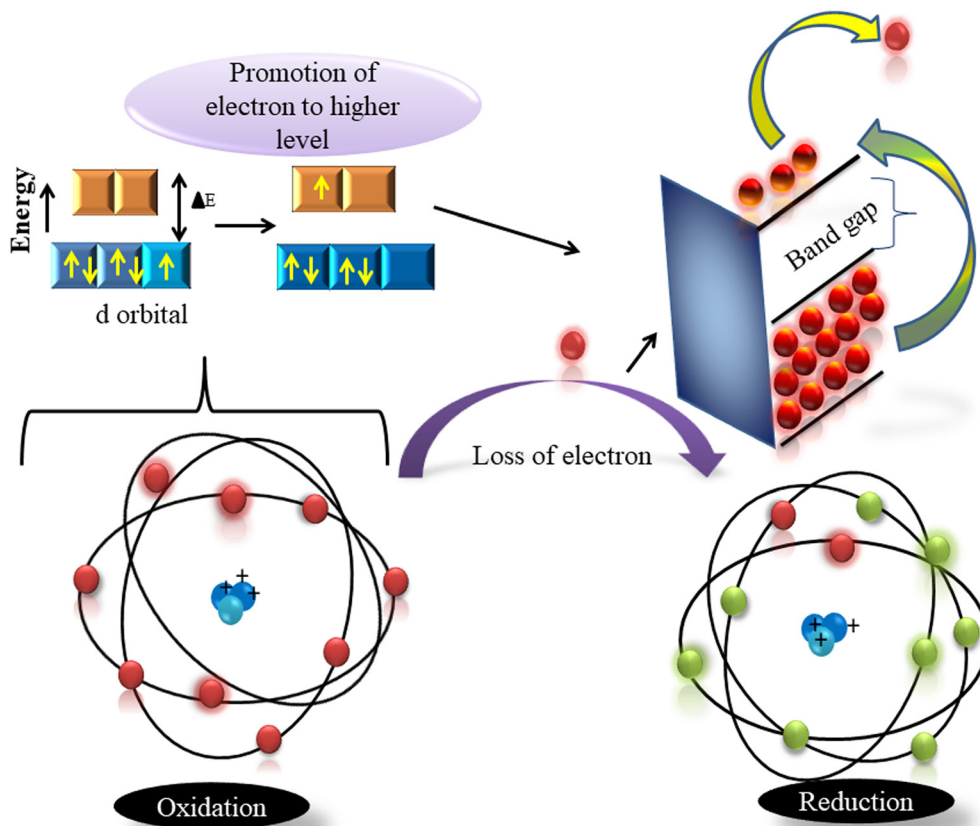


Fig. 13 Electron removal with electronegativity.

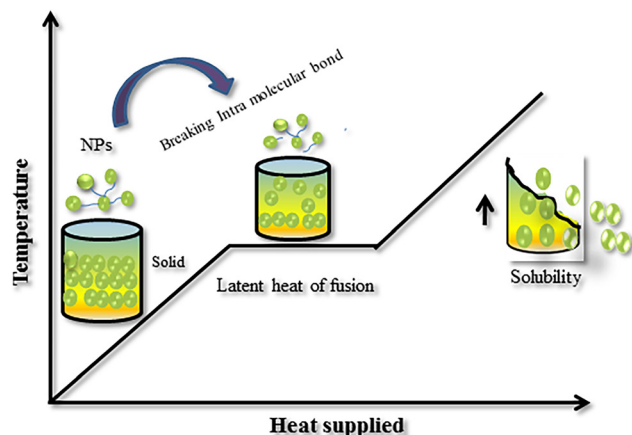


Fig. 14 Mechanism of the heat of fusion.

Another descriptor influencing the cytotoxicity is the **dose**. Understanding of the dose is very critical. Drugs that are beneficial at a low dose can be detrimental at a high dose. The **positive** contribution of the descriptors indicates that increasing the dose of nanoparticles with the same metal core can decrease the cell viability, as observed for the nanoparticle SnO_2 (dose 1500, 82% cell death). On the contrary, with a decrease of the dose for the same metal core, the cell viability increases, as seen for the nanoparticle Al_2O_3 (dose 50, cell death 5.1%). Rizk *et al.*³¹ demonstrated that oxidative stress and liver function were greatly dependent on the dose and the exposure time of the cells to the nanoparticles. He also mentioned that genetic disturbance occurs at a higher dose and long duration. Studies

have also supported the hypothesis that NPs increase free radicals inside the cell in a dose-dependent manner. The catalytic processes can initiate free radical generation until the NPs are removed from the body or degraded.

SuM_Active_M_SM is the summation of the number of metals and semi-metals in the nanoparticles. The toxicity phenomena of metallic NPs are complicated and depend on various physicochemical parameters. These NPs, with their unique natures, may have different mechanisms and the amount of toxicity may vary. It has also been observed that different nanomaterials stimulate certain mechanisms. The presence of metal does not clearly correspond to an increase or decrease in toxicity. The toxicity depends on the solubility, oxidation number, *etc.* This descriptor contributes **negatively** to the toxicity, which can be interpreted as an increase in the number of metals or semimetals in the metal oxides leading to an increase in the molecular size. Monovalent metal oxides are smaller in size than the divalent or trivalent metal oxides. Therefore, with increasing the number of metals in metal oxides, the cytotoxicity increases, as in the case of the nanoparticle Sb_2O_3 , for which the descriptor value is 2 and the % cell death is 11.6%; on the other hand, the nanoparticle NiO having descriptor value 1 has a % cell death of 87.4%. It can be assumed that an increase in molecular size increases the size of the MeO_x NPs and hence interferes with the penetration of the NPs through the pores of the *HaCaT* cells, leading to a less toxic effect.

The **cationic charge (Cat charge)** descriptor corresponds to ions that bear a positive charge formed when the atom loses an electron. The loss of electrons results in an overall positive charge. Cationic charges are related to cytotoxicity due to

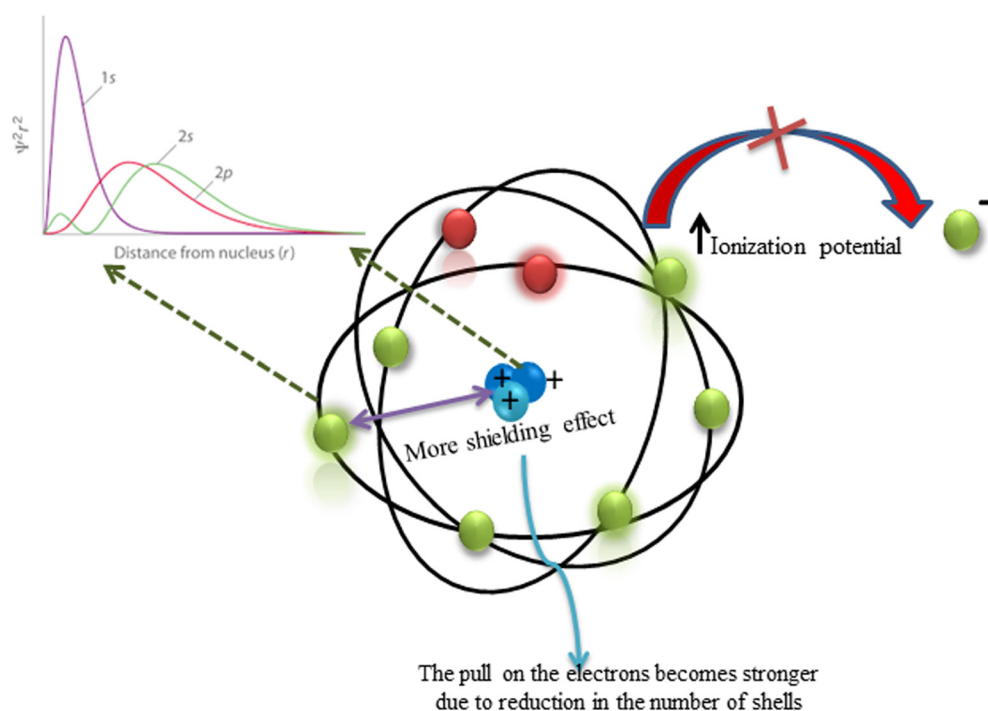


Fig. 15 Mechanism of reduced toxicity due to increased cationic charge.

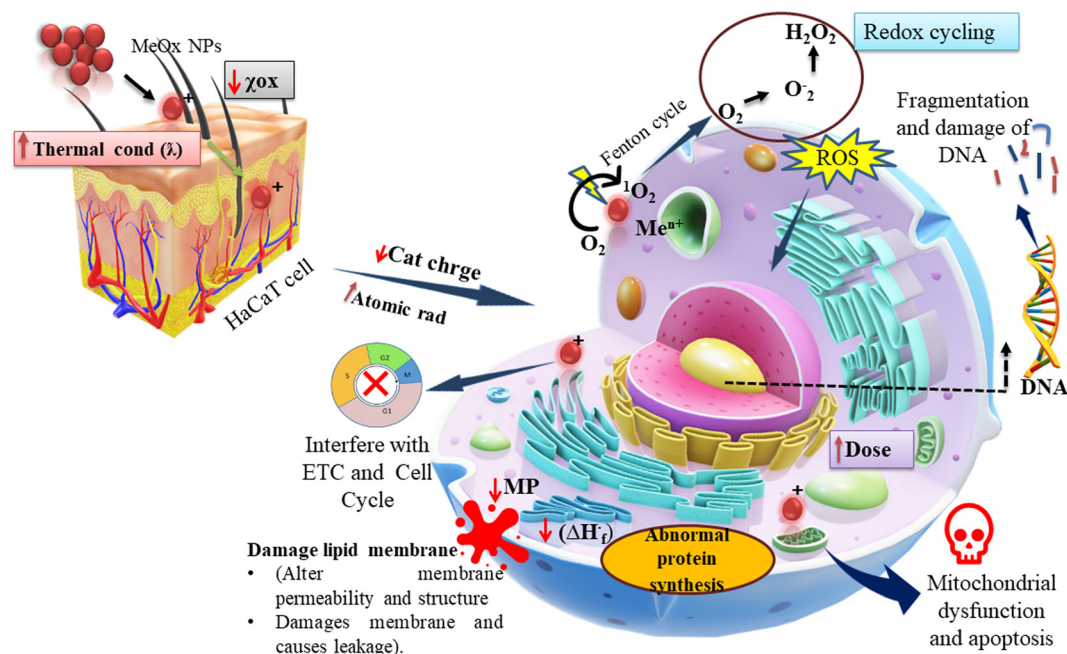


Fig. 16 Mechanism of MeOx NP toxicity via various pathways. *ETC – Electron transport chain.

their electrostatic interaction with the negatively charged membrane. The **negative** coefficient of the descriptor indicated that an increase in the cationic charge decreases the cytotoxicity, as in the case of the nanoparticle ZrO_2 (cationic charge = 4, % cell death = 2.9). On the other hand, decrease in the cationic charge to a certain level increases the toxicity towards the *HaCaT* cell line, as observed for the nanoparticle ZnO (cationic charge = 2, % cell death = 99.07). Electronegativity is the tendency to attract electrons. Weakly electronegative atoms tend to lose electrons and form cations, whereas strongly electronegative atoms acquire electrons and become anions. Electronegativity decreases as the electrons in the valence shell decrease. The decreased electronegativity makes the atom electropositive. Electropositive atoms have high ionization potentials, *i.e.*, they require more energy to remove the electron from the valence shell. As the cationic charge increases, the ionization energy moves from the 1st ionization potential to a higher form. It is well known that the energy required for the 2nd and 3rd ionization potential is much higher than that of the 1st ionization potential. Electronegativity is a function of the nuclear charge, *i.e.*, the effective positive charge from the nucleus that an electron experiences. A large effective nuclear charge corresponds to a strong attraction between the electron and nucleus. The effective nuclear pull increases as the number of shells decreases, as witnessed for the atomic radius descriptor in case 1. Thus, it becomes tougher for the electron to leave its valence shell against the strong nuclear pull (Fig. 15).

3.3 Mechanism of the toxicity

The increase in production and use of NPs increases the exposure of humans and the environment to NPs. Humans are most

prone to NP exposure through inhalation, ingestion, dermal contact, or injection. When a human comes into contact with NPs, the NPs cause toxicity *via* various mechanisms. The features obtained from the modeled descriptors can explain the mechanism of toxicity well. Periodic-table-based descriptors such as ionic charge affect the interaction of NPs with the cell membrane, which has a great impact on the induced toxicity mechanism. The charged surfaces of the NPs can be described by the electronegativity, thermal conductivity, and atomic radius of the MeO_x NPs. The presence of these descriptors promotes the formation of the cationic form of the metal. The physical interaction of the NPs with the cell membrane is mainly governed by the cationic charge of the NPs. Due to the presence of a cationic charge, the NPs can easily enter the cell through endocytosis. The electrostatic interaction of the NPs with the cellular membrane produces cellular responses that have been identified as a paradigm for cytotoxicity. The presence of a highly electronegative ion also increases the catalytic properties of the metal cation. Thus, electronegativity increases the toxicity of nanoparticles through the Haber–Weiss–Fenton cycle.³² Oxidative stress depends on the concentration of the NPs, or in other words, on the dose of the NPs. The VIP plot for the PLS model also demonstrated the importance of the dose descriptor in influencing cytotoxicity. The intracellular generation of oxidative stress in the targeted cells may form a malondialdehyde molecule, which triggers deterioration of the lipid membrane (lipid peroxidation).³³ Excessive production of ROS by NPs can damage RNA and DNA (nuclear components) and interfere with cellular signaling pathways, which may alter gene expression, which ultimately stops the cell growth mechanism (Fig. 16). The presence of metallic NPs also damages cells through cell membrane permeabilization in a spatial manner. Metallic NPs bind with the cell wall and lead to cell death. Co NPs (cobalt

nanoparticles) produce DNA damage that leads to ataxia-telangiectasia mutated (ATM) protein activation, which further increases the phosphorylation of p53 and Rad 51 protein expression. The heat of the fusion and melting point descriptors increase the solubility of the Me_{Ox} NPs. The NPs release metallic ions to the cytoplasm; this is directly proportional to the dissolution rate of NPs in a particular size range.

3.4 Comparison with the previous work

This study focused on identifying the variables that govern the toxicity of Me_{Ox} NPs toward *HaCaT* cells. Here, two datasets of Me_{Ox} NPs having cytotoxicity toward *HaCaT* cells were used. In the study by Puzyn *et al.*,⁶ quantum chemical descriptors (Table 4) were used; these require a computational chemistry background to perform, are also computationally challenging. Generally, they are difficult to use due to their mathematical and computational complexities. In every case in which a different metal oxide is used, energy minimization and changing of the orientation of the Me_{Ox} is required before calculating the quantum descriptor. The use of quantum descriptors also has some drawbacks. For instance, determining the quantum descriptors for mixture nanoparticles can be challenging, but it can easily be accomplished with periodic-table-based descriptors. Quantum chemical descriptors have failed to directly address the bulk effect of nanoparticles.³⁴ The periodic table descriptors obtained in our model corroborate the findings of Puzyn *et al.*⁶ obtained from quantum chemical descriptors. It is worth mentioning here that fundamental properties such as cation charge formation of a metal oxide and LUMO and HOMO energies can be easily explained from the information acquired from periodic table descriptors in no time. Additionally, such information is independent of the size range and does not require energy minimization for descriptor calculation. The other work performed by Choi *et al.*¹⁴ used two cell lines in the same model. According to the OECD guidelines, to perform any QSAR study, the experimental conditions must be same. Thus, here we have selected the data corresponding to *HaCaT* cells (only one cell line) with a specific assay method (CytoTox-Glo). The developed models were found to be statistically significant. We have also changed the endpoint to % cell death for a better understanding of metal oxide toxicity. A strict comparison was not possible in case of the 2nd dataset due to differences in the model development procedure, the use of a different cell line, and different assay methods.

4. Conclusion

The ever-increasing use of NMs in the form of nanoparticles (NPs) in biomedical applications and consumer products has raised concern about their safety to humans and the environment. Nanosafety has become an important part of nano-education for future researchers and a high-demand research area in recent times. The potential of NPs to cause undesirable health effects in the environment and the susceptible part of the human population could be seen as an alarm and a call for preventive measures. The concept of nanotoxicology in the medical sciences is needed for the safety evaluation of nanomedicines for nanotechnology-based therapies and diagnostics to identify the potential hazards they may cause after application. The potential toxicity of nanoparticles, as well as their human toxicity, has prompted major scientific organizations to assess the risks that may be associated with new nanoparticles. Therefore, the present work was performed to identify the main features governing the cytotoxicity in human keratinous cells (*HaCaT*). Combining the calculated periodic table descriptors with available experimental data allowed us to reveal and explain the features responsible for toxicity to the corresponding cell line through the development of a predictive and validated quantitative structure-toxicity relationship (QSTR) model. This work demonstrated the toxicity profile of NPs toward *HaCaT* cells derived from two datasets. The insights obtained from the two datasets demonstrated that periodic table descriptors can also predict the relevant features responsible for toxicity toward these cells. The key factors influencing the toxicity of the NPs were determined from the developed models (MLR and PLS). From the MLR model, the atomic radius and thermal conductivity were found to increase toxicity through metal cation formation. On the other hand, decreased melting point and increased electronegativity of the atom were responsible for cell death through ROS production and damage to DNA, mitochondria, and lysosomes. The same observations were noticed in the case of the PLS model, in which the heat of fusion, increased cationic charge, and dose have a substantial effect on the cytotoxicity. From the developed QSAR models, it was concluded that major factors for toxicity were cation formation, increased solubility, and the dose of the NPs. These features are responsible for controlling the physicochemical properties of metallic NPs to achieve safe and reliable Me_{Ox} NPs. This work opens a new horizon in terms of awareness and knowledge for understanding NP toxicity. The nanotoxicology information can help to design modern materials within a safe threshold with optimal properties and minimal potential toxicity.

Table 4 Comparison with previously published work (dataset 1)

Statistical parameters	Number of data points (<i>n</i>)	R^2	Q_{LOO}^2	F value	$\bar{r}_m^2$, train	Δr_m^2 train	R_{adj}^2	$Q_{F_1}^2$	$Q_{F_2}^2$	$\bar{r}_m^2$, test	Δr_m^2 test
Previous work (quantum descriptors)	Training set = 10 Test set = 8	0.93	0.86	44.6	—	—	—	0.83	—	—	—
Current work (periodic table descriptors)	Training set = 11 Test set = 5	0.95	0.83	29.03	0.773	0.037	0.92	0.73	0.62	0.27	-0.505

Conflicts of interest

None.

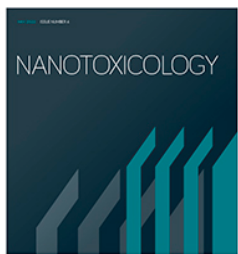
Acknowledgements

JR gratefully acknowledges the Indian Council of Medical Research, New Delhi (ICMR) for financial assistance in the form of a Senior Research Fellowship.

References

- 1 *Analysis of the Nanotechnology Consumer Inventory of August 25, 2009*, Woodrow Wilson International Centre for Scholars, 2009, available at: <https://www.nanotechproject.org/inventories/consumer/>.
- 2 J. A. Damasco, S. Ravi, J. D. Perez, D. E. Hagaman and M. P. Melancon, Understanding nanoparticle toxicity to direct a safe-by-design approach in cancer nanomedicine, *Nanomaterials*, 2020, **10**(11), 2186.
- 3 T. R. Pisanic II, J. D. Blackwell, V. I. Shubayev, R. R. Fiñones and S. Jin, Nanotoxicity of iron oxide nanoparticle internalization in growing neurons, *Biomaterials*, 2007, **28**(16), 2572–2581.
- 4 D. A. Winkler, E. Mombelli, A. Pietroiusti, L. Tran, A. Worth, B. Fadeel and M. J. McCall, Applying quantitative structure-activity relationship approaches to nanotoxicology: current status and future potential, *Toxicology*, 2013, **313**(1), 15–23.
- 5 S. Venigalla, S. Dhail, P. Ranjan, S. Jain and T. Chakraborty, Computational study about cytotoxicity of metal oxide nanoparticles invoking nano-QSAR technique, *Front. Chem.*, 2014, **23**(2), 123.
- 6 T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynska and J. Leszczynski, Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles, *Nat. Nanotechnol.*, 2011, **6**(3), 175–178.
- 7 E. Burello and A. P. Worth, QSAR modeling of nanomaterials, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2011, **3**, 298–306.
- 8 T. Puzyn, D. Leszczynska and J. Leszczynski, Toward the development of “nano-QSARs”: advances and challenges, *Small*, 2009, **5**, 2494–2509.
- 9 A. G. Cattaneo, R. Gornati, E. Sabbioni, M. Chiriva-Internati, E. Cobos, M. R. Jenkins and G. Bernardini, Nanotechnology and human health: risks and benefits, *J. Appl. Toxicol.*, 2010, **30**, 730–744.
- 10 K. Roy, S. Kar and R. N. Das, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, 2015.
- 11 N. Sizochenko, B. Rasulev, A. Gajewicz, V. Kuz'min, T. Puzyn and J. Leszczynski, From basic physics to mechanisms of toxicity: The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles, *Nanoscale*, 2014, **6**, 13986–13993.
- 12 S. Kar, A. Gajewicz, T. Puzyn, K. Roy and J. Leszczynski, Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach, *Ecotoxicol. Environ. Saf.*, 2014, **107**, 162–169.
- 13 A. Gajewicz, N. Schaeublin, B. Rasulev, S. Hussain, D. Leszczynska, T. Puzyn and J. Leszczynski, Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies, *Nanotoxicology*, 2015, **9**(3), 313–325.
- 14 J. S. Choi, T. X. Trinh, T. H. Yoon, J. Kim and H. G. Byun, Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials, *Chemosphere*, 2019, **217**, 243–249.
- 15 P. De, S. Kar, K. Roy and J. Leszczynski, Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modeling for exploration of toxicity mechanisms, *Environ. Sci.: Nano*, 2018, **5**(11), 2742–2760.
- 16 J. Devillers, *Genetic algorithms in computer-aided molecular design*, Academic Press, 1996.
- 17 V. Venkatasubramanian and A. Sundaram, Genetic Algorithms: Introduction and Applications, *Encyclopedia of Computational Chemistry*, 2002.
- 18 K. Roy, R. N. Das, P. Ambure and R. B. Aher, Be Aware of Error Measures, Further Studies on Validation of Predictive QSAR Models, *Chemom. Intell. Lab. Syst.*, 2016, **152**, 18–33.
- 19 K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das and H. Kabir, Comparative Studies on Some Metrics for External Validation of QSPR Models, *J. Chem. Inf. Model.*, 2012, **52**(2), 396–408.
- 20 K. Roy, S. Kar and P. Ambure, On a Simple Approach for Determining Applicability Domain of QSAR Models, *Chemom. Intell. Lab. Syst.*, 2015, **145**, 22–29.
- 21 P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb. Sci.*, 2007, **26**, 694–701.
- 22 L. Eriksson, J. Jaworska, A. P. Worth, M. T. Cronin, R. M. McDowell and P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs, *Environ. Health Perspect.*, 2003, **111**(10), 1361–1375.
- 23 STATISTICA is a statistical software of STATSOFT Inc., USA, <https://www.statsoft.com/>.
- 24 S. Wold, M. Sjostrom and L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.
- 25 UMETRICS SIMCA-P 10.0, Umea, Sweden, 2002, info@umetrics.com, <https://www.umetrics.com>.
- 26 OECD Principles for the validation, for regulatory purposes, of (Quantitative) Structure Activity Relationship models, 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides, and Biotechnology, Organisation for economic co-operation and development, Paris, France, 2004, available at: <https://www.oecd.org/env/ehs/riskassessment/validationofqsarmodels.html>.
- 27 H. Sun, C. Jiang, L. Wu, X. Bai and S. Zhai, Cytotoxicity-related bioeffects induced by nanoparticles: the role of surface chemistry, *Front. Bioeng. Biotechnol.*, 2019, 414.
- 28 J. Roy, P. K. Ojha and K. Roy, Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a

- QSTR approach using simple periodic table based descriptors, *Nanotoxicology*, 2019, **13**(5), 701–716.
- 29 J. D. Walker, M. Enache and J. C. Dearden, Quantitative cationic activity relationships for predicting toxicity of metal ions from physicochemical properties and natural occurrence levels, *QSAR Comb. Sci.*, 2007, **26**(4), 522–527.
- 30 A. B. Sengul and E. Asmatulu, Toxicity of metal and metal oxide nanoparticles: a review, *Environ. Chem. Lett.*, 2020, **18**(5), 1659–1683.
- 31 M. Z. Rizk, S. A. Ali, M. A. Hamed, N. S. H. F. El-Rigal Aly and H. H. Salah, Toxicity of titanium dioxide nanoparticles: effect of dose and time on biochemical disturbance, oxidative stress and genotoxicity in mice, *Biomed. Pharmacother.*, 2017, **90**, 466–472.
- 32 W. H. Koppenol, The Haber-Weiss cycle–70 years later, *Redox Rep.*, 2001, **6**(4), 229–234.
- 33 T. Y. Wang, M. D. J. Libardo, A. M. Angeles-Boza and J. P. Pellois, Membrane oxidation in cell delivery and cell killing applications, *ACS Chem. Biol.*, 2017, **12**(5), 1170–1182.
- 34 M. Karelson, V. S. Lobanov and A. R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.*, 1996, **96**(3), 1027–1044.



Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells

Joyita Roy & Kunal Roy

To cite this article: Joyita Roy & Kunal Roy (2022): Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPS) used in nanopesticides to BEAS-2B and RAW 264.7 cells, Nanotoxicology, DOI: [10.1080/17435390.2022.2132887](https://doi.org/10.1080/17435390.2022.2132887)

To link to this article: <https://doi.org/10.1080/17435390.2022.2132887>



View supplementary material [↗](#)



Published online: 19 Oct 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

ARTICLE



Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPs) used in nanopesticides to BEAS-2B and RAW 264.7 cells

Joyita Roy and Kunal Roy

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

ABSTRACT

The demand for nutrients and new technologies has increased with population growth. The agro-technological revolution with metal oxide engineered nanoparticles (MeOx ENPs) has the potential to reform the resilient agricultural system while maintaining the security of food. When utilized extensively, MeOx ENPs may have unintended toxicological effects on both target and non-targeted species. Since limited information about nanopesticides' pernicious effects is available, *in silico* modeling can be done to explore these issues. Hence, in the present work, we have applied computational modeling to explore the influence of metal oxide nanoparticles on the toxicity of bronchial epithelial (BEAS-2B) and murine myeloid (RAW 264.7) cells to bridge the data gap relating to the toxicity of MeOx NPs. Initially, partial least squares (PLS) regression models were developed applying the Small Dataset Modeler software (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab/) using four datasets having effective concentration (EC₅₀%) as the endpoints and employing only periodic table descriptors. To further explore the predictions, we applied a read-across approach using the descriptors selected in the QSAR models. Also, the inter-endpoint cytotoxicity relationship modeling (quantitative toxicity-toxicity relationship or QTTR) was conducted. It was found that the result obtained by nano-read-across provided a similar level of accuracy as provided by QSAR. The information derived from the PLS models of both the cell lines suggested that metal cation formation, and bond-forming capacity influence the toxicity whereas the presence of metal has an influential impact on the ecotoxicological effects. Thus, it is feasible to design safe nano-pesticides that could be more effective than conventional analogs.

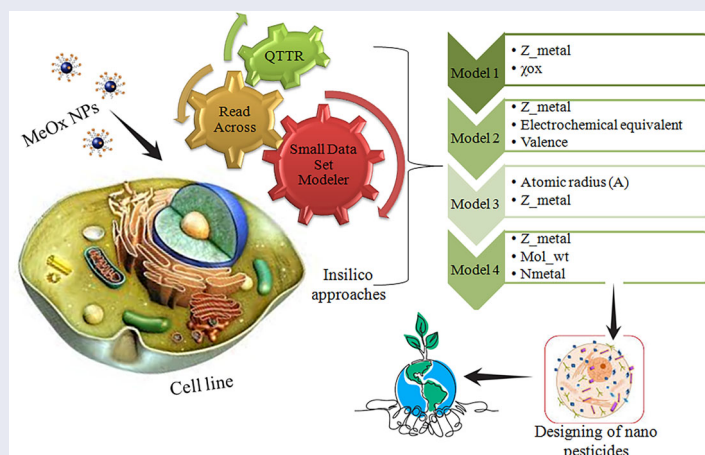
ARTICLE HISTORY

Received 6 July 2022
Revised 29 September 2022
Accepted 30 September 2022

KEYWORDS

Nanopesticides; QSTR; read-across; small dataset modeler; interspecies

GRAPHICAL ABSTRACT



1. Introduction

Nanotechnological breakthroughs in the last 10 years had a potential revolution in the agricultural

sector. The application of nanotechnology in the field of agriculture has emphasized developing nano-agrochemicals (Kah 2015). The decrease in

CONTACT Kunal Roy kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/17435390.2022.2132887>

© 2022 Informa UK Limited, trading as Taylor & Francis Group

agricultural areas due to globalization has also demanded large-scale food production. The applications of innovative and unique nano-sized particles as nanopesticides are used extensively to improve crop production. These nanopesticides and nanofertilizers in different forms in agriculture enable direct exposure of metal oxide engineered nanoparticles (MeOx ENPs) to the environment (Colman et al. 2018b). The extensive use (nearly 4 million tons) of a large number of metal oxide nanoparticles in the form of nanopesticides annually may enter the agro-ecosystem through a direct or indirect application (Simonin et al. 2018). Only a small portion of nanopesticides may reach the targeted organisms leaving a large portion in the ecosystem as potential hazards. The higher availability and low degradation of nanopesticides may result in unwanted toxicity to the non-targeted organisms. Therefore, several parameters need to be investigated in depth that may be a strong drive on an exploration of environmental fate and ecotoxicity of nanopesticides. The European Union (EU) Regulation 1107/2009 characterizes the novel nanopesticides for plant protection products in Europe (EFSA 2011; Kookana et al. 2014). The rate of metal oxide application in nanopesticides far exceeds the rate at which the toxicity studies are performed. The conventional risk assessment methods are time-consuming for the vast application of newly approved materials. Hence, an easy method that can predict toxicity is required to keep up with the high and low rates of tests. Computational methods are in the fore line to investigate nanopesticides which have the potential to cause ecotoxicity and provide a viable tool to address experimental issues observed during pesticide risk assessment. Chemical legislative frameworks such as Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) have considered such non-experimental methods reliable. Another Regulation (EC) No 396/2005 covers matters relating to allowed limits of residues in food and the official controls. Regarding the safety aspect of human and animal health, the European Regulations (EU) No 283/2013 and (EU) No, 284/2013 were set up for plant production products i.e. PPP, and for data requirement of the active products used in nanopesticides.

Over the last years, many chemoinformatics predictive models (Nano-QSAR) have become a notion to address the limitation of knowledge of nanomaterial behavior. But these predictive models cannot efficiently address the problem of missing data for nanoparticle toxicity. Hence, there is a lack of models that can fill the data gaps. For, instance nano-read across approach is the technique to fill data gaps based on external prediction for target compounds (or untested) using source compounds. In other words, the read-across method can predict information of endpoints for one nanoparticle using an experiment or calculated data from another nanoparticle.

Thus, in absence of appropriate validated QSAR models, the nano-read across approach can be an attractive technique to fill the existing data gap for toxicity analysis from a regulatory perspective (OECD, 2007). Read-across has been applied successfully to predict the toxicity of many compounds (Gajewicz et al. 2015; Sizochenko et al. 2018; Ball et al. 2016). Therefore, the current study focuses on the application of an efficient approach that is capable of predicting different properties of unknown nanoparticles based on information derived from limited data sources. It is known that metal oxide nanoparticles are capable of generating inflammatory responses (Buchman et al. 2019). The current paper considers the toxicity of nanopesticides due to metal oxide nanoparticles; therefore, bronchial epithelial cell, that is, BEAS-2B data was selected to determine the inflammatory response in the lung. Further, a murine myeloid (RAW 264.7) cell line was selected to determine the toxicity of the murine mice, as we know that macrophages are resident in many organs (e.g. lungs) and responsible for the clearance of NPs, and murine myeloid macrophage (RAW 264.7) cell lines act as excellent models for mimicking human cell lines (Ganguly, Breen, and Pillai 2018) during nanopesticides exposure. These cell lines are commonly included in determining cytotoxicity of engineered nanoparticles. Both cell lines have been used extensively in the literature for NP and ambient ultrafine particle (UFP) toxicity studies, including assessment of cellular uptake routes and study of a cellular stress response (Batista Caixeta et al. 2020; Liu et al. 2013; Uboldi et al. 2019).

Most of the predictive models reported in the literature use quantum chemical descriptors which require computational resources and expertise; however, despite these, these calculations may not be very reproducible and transferable due to several assumptions and approximations. The present work uses periodic table descriptors which are easily obtainable, transferable, and reproducible.

2. Methods and materials

2.1. The dataset

The cytotoxicity data on RAW 264.7 and BEAS-2B cell lines were taken from previously published literature (Zhang et al. 2012). The cytotoxicity data were obtained from two different assay methods (i.e. MTS and ATP) for a single cell line. A total of four datasets were used for modeling purposes with the half-maximal effective concentration in the logarithmic ($\log(1/EC_{50})$) scale as the endpoint. Here, four datasets containing 9 (Dataset 1), 10 (Dataset 2), 10 (Dataset 3), and 12 (Dataset 4) MeOx ENPs respectively were taken for modeling purposes. Merging compounds from two datasets (ATP assay data) out of the four datasets from RAW 264.7 and BEAS-2B cell lines, common eight MeOx ENPs were used to develop the nano-quantitative toxicity-toxicity relationship (nano-QTTR) model. The same four datasets were also further utilized for predictions using the similarity-based approach.

2.2. Molecular descriptors

Molecular descriptors are used to encode the properties of MeOx ENPs. The complex behavior of metal oxides complicates their descriptor calculations making it a difficult job for the QSAR researchers. Therefore, in the present study, 58 periodic table descriptors were calculated which include twenty-three second-generation periodic table descriptors calculated from the primary descriptors (De et al. 2018), and 30 descriptors were taken from the Elemental Descriptor Calculator software available from <https://dtclab.webs.com/software-tools>. Additionally, some basic information about the metals was taken from the periodic table to encode the cytotoxicity of RAW 264.7 and BEAS-2B cell lines.

The details of descriptors are given in Table S1 of the supplementary section.

2.3. Small dataset modeler for feature selection

The choice of appropriate structural characteristics that can quantitatively describe the variation of the MeOx ENPs' structures is essential for computational modeling. This is because the general concept of the read-across method is based on the assumption that similar compounds will bear similar responses to toxicity. Hence, to identify the structural features that can relate to the toxic features of MeOx NPs to RAW 264.7 and BEAS-2B cell lines, the Small Dataset Modeler software version 1.0.0 (Ambure et al. 2019) was utilized. The existing dataset contains limited data for modeling, and handling datasets with limited data points (less than 20) is problematic for statistical model development as lower degrees of freedom lead to statistically less interesting and unreliable models. The small dataset modeler software (http://teqip.jdvu.ac.in/QSAR_Tools/) can handle the situation as it works with the whole data set without training-test splitting and cross-validates the models using the double cross-validation technique. This software performs dataset curation, 'exhaustive' double cross-validation, and performs consensus predictions as optimal model selection techniques. In the Double Cross-Validation (DCV) strategy, the internal loop (training set) is divided into a calibration set and a validation set where the calibration set is used for model development and model selection, and predictivity is entirely checked by the validation set. In the case of Small Dataset Modeler software (Ambure et al. 2019), the whole dataset is used as modeling set in the inner loop (avoiding the data splitting step in the inner loop). The modeling set is constantly divided into all possible 'k' combinations of calibration and validation sets. The developed calibration set is used to develop multiple linear regression (MLR) or partial least squares models (PLS) using the genetic algorithm-MLR (GA-MLR) technique (Devillers 1996; Venkatasubramanian and Sundaram 2002). Again, the validation set is used to derive the prediction quality of the respective developed models. The results obtained from the PLS models of BEAS-2B cell and RAW 264.7 cell datasets were found to be

statistically significant. The descriptors obtained from the four PLS models with significant outcomes were used in the read-across technique.

2.4. Read-across based predictions

After selecting appropriate descriptors from the four PLS models, a machine learning approach based on similarity measures was performed. This approach uses the Quantitative Read-Across v4.0 tool available from our laboratory website (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home>). The Read-across approach uses Euclidean distance, Gaussian kernel function, and Laplacian kernel function-based similarity techniques. This method requires optimization of hyperparameters, that is, sigma and gamma values; distance, and similarity thresholds. The best result obtained with the specific hyperparameter was selected as the final set to get the desired outcome. All four datasets were divided into a training set

(source compounds) and a test set (target compounds). Since the dataset was small (less than 20), conventional dataset division, that is, 70% data points in the training set and 30% in the test set does not appear to be satisfactory. In such a case, the number of test set compounds would be very limited to check the quality of predictions. For read-across predictions, 'many' source compounds are not usually required as the limited number of close source compounds would be sufficient to make efficient predictions. Therefore, 50% division was opted (Gajewicz 2017), and the division was done randomly many times to further avoid any bias during modeling.

The read-across method is well illustrated in Figure 1.

2.5. Nano-QTTR model development

The common eight compounds were taken from the ATP assay of both BEAS-2B cell and RAW 264.7

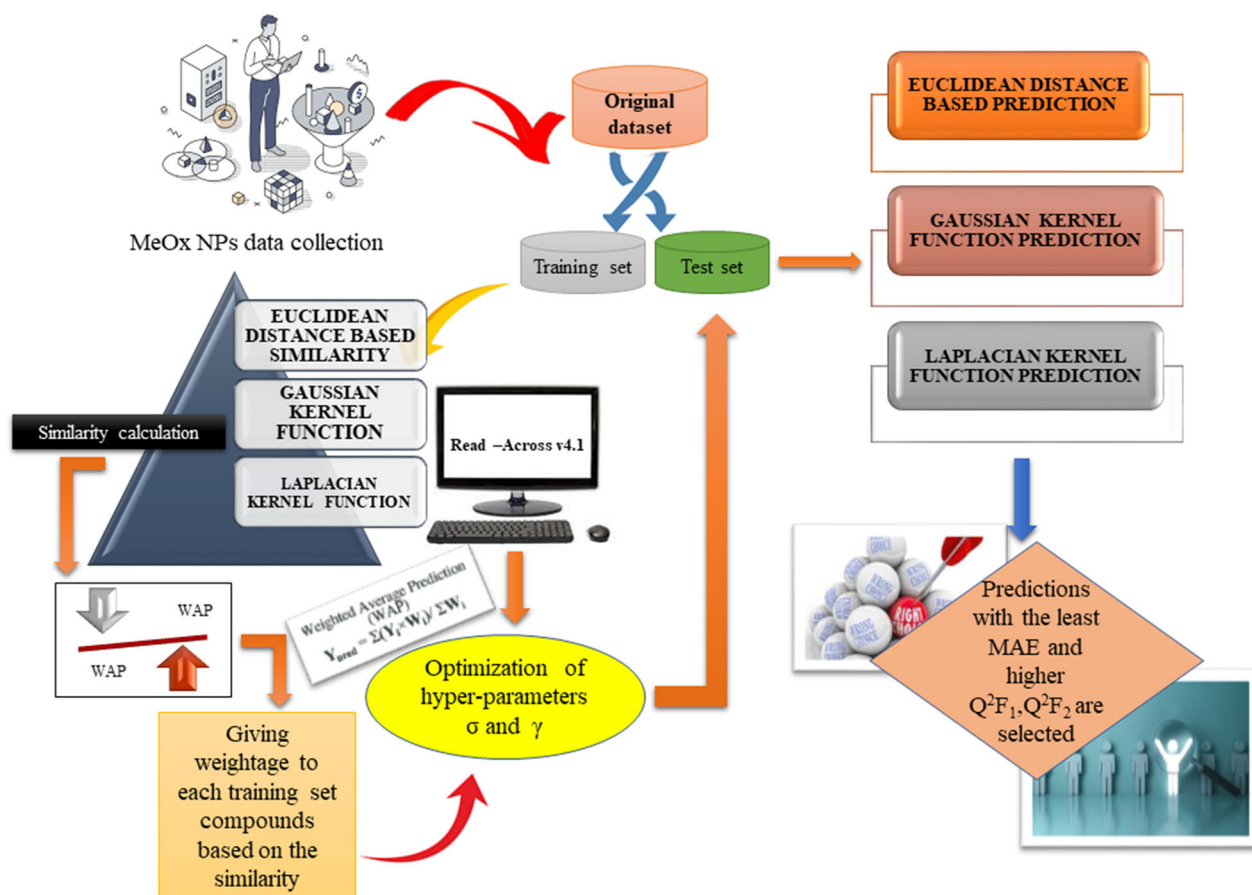


Figure 1. TheRead-across flow diagram.

cell datasets with cytotoxicity values ($EC_{50}\%$), for model development. It is interesting to note that initially without any chemical descriptors, just simple linear regression was performed between the cytotoxicity data of both BEAS-2B cells and RAW 264.7 cells. The encouraging results from the MLR model inspired us to develop the nano-QTTR model. Subsequently, identification of the most relevant descriptors from the pool of fifty-eight descriptors was done employing the stepwise regression technique (stepwise-MLR) (Darlington 1990). Along with the selected descriptors, the cytotoxicity of one of the cells was used as the predictor variable to predict the cytotoxicity of the other cell using the partial least squares (PLS) method (Wold 1995) (Figure 7). As we know that finding a large number of data points relevant to MeOx ENPs is difficult, therefore QTTR models may give us the insight to predict MeOx ENPs' toxicity to various species despite a limited number of available datapoints and help in data gap filling. The experimental values of the eight metals were available for the ATP assay methods (Dataset 1 and Dataset 4) and not for the MTS assay methods (Dataset 2 and Dataset 3). Therefore, these were not applied in QTTR modeling.

2.6. Statistical validation metrics

The validation of the developed model is crucial for the QSAR study. Multiple rigorous assessments were performed to determine the predictability, stability, and robustness of the developed models. All the models were evaluated as per the OECD guidelines. Here, we have computed coefficient (R^2), leave-one-out determination coefficient (Q^2_{LOO}), and mean absolute error (MAE) (Roy et al. 2016). Additionally, r_m^2 metrics (Roy et al. 2012) for the final PLS models (Wold 1995) were also calculated. Furthermore, Q^2_{F1} and Q^2_{F2} , RMSEP, and MAE were also calculated for read-across prediction.

For compliance of the developed models with the OECD guidelines for applicability in regulatory purposes, the following principles were applied:

Principle 1 (a defined endpoint): The responses being modeled in the present work for the different datasets were measured under identical

conditions. Thus, the QSAR models were developed following the 1st OECD principle.

Principle 2 (an unambiguous algorithm): Various chemometric tools based on specific algorithms were employed for the calculation of the different categories of descriptors and subsequent QSAR model development using specific software packages. Thus, the model development pathway employed for the present studies follows a definite algorithm.

Principle 3 (a defined domain of applicability): The domain of applicability of all the statistically significant PLS QSAR models was analyzed using the DModX method. Thus, the selection of the best QSAR model was done in corroboration with this principle.

Principle 4 (appropriate measures of goodness-of-fit, robustness, and predictivity): All the developed models were rigorously validated using internal, external, and overall validation approaches. The quality of fitness and the predictive potential of the developed models were assessed based on the different validation metrics while the robustness of the models was judged using the randomization approach. The selection of the most significant models based on the acceptable values of the various validation metrics accounts for the compliance of the models with the 4th guideline.

Principle 5 (a mechanistic interpretation): All the descriptors appearing in the developed QSAR models could aptly define the essential structural attributes of the molecules imparting optimum endpoint values thus signifying suitable mechanistic interpretation of the developed models.

3. Applicability domain

The applicability domain (AD) in QSAR (Gadaleta et al. 2016) is the physicochemical, structural, and biochemical space knowledge of the training set compounds based on which the model has been developed. Prediction of the new compounds is possible when the new entity falls in the same chemical domain of the modeled descriptors. AD confirms the model assumption for which type of compounds the model is reliably applicable. Therefore, in this study, the AD assessment for the developed PLS models was done by the DModX

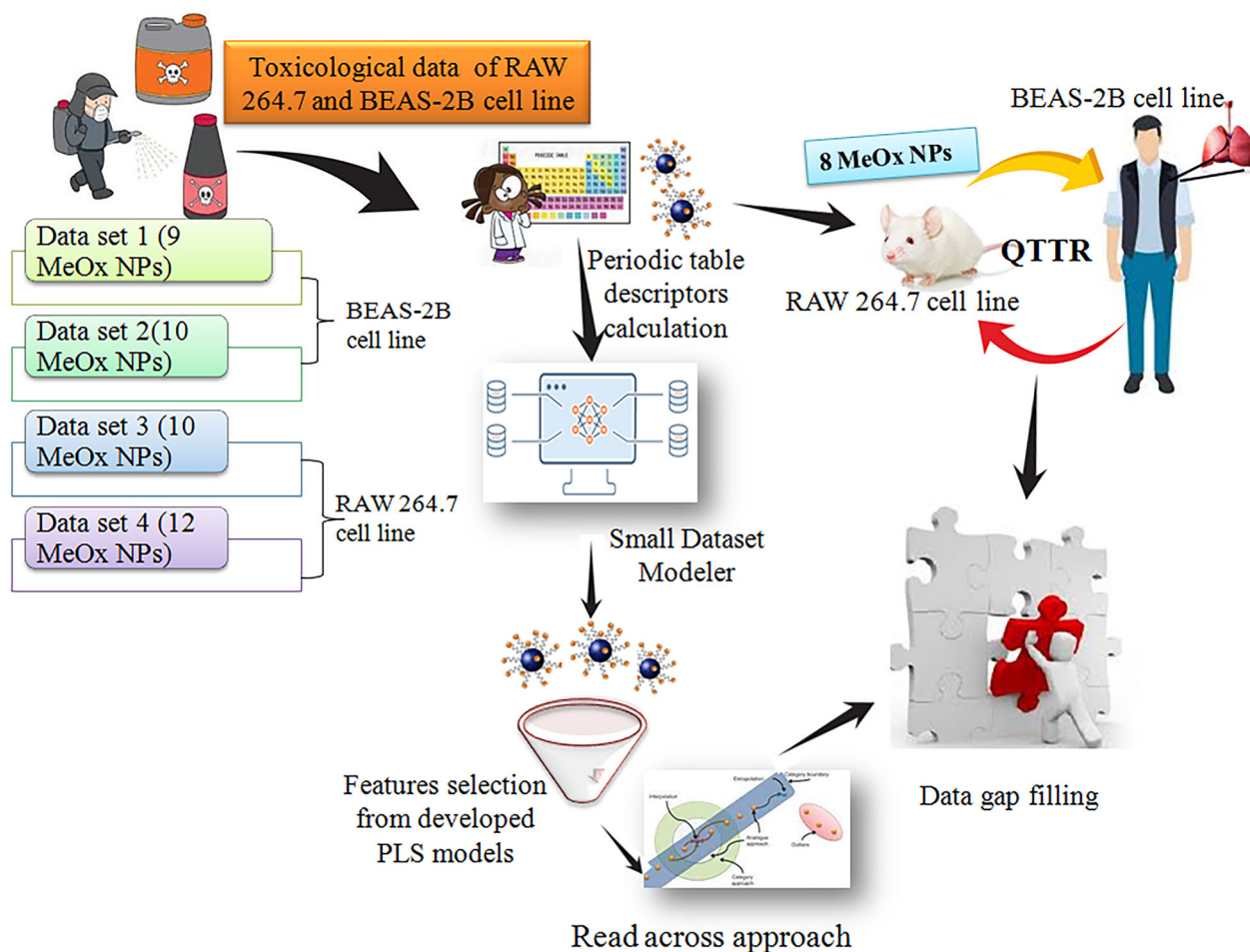


Figure 2. Flowchart of the present work.

approach, that is, distance to model in the X-space with SIMCA 16.0.2 software. The AD was checked at a 99% confidence level ($D\text{-crit} = 0.009998$). The whole workflow is well explained in Figure 2.

4. Result and discussion

Periodic table descriptors were used for modeling four datasets (M1-M4) comprising 10, 9, 10, and 12 MeOx ENPs respectively (the details are given in [supplementary Tables S2–S5](#)). First, the small dataset modeler software (Ambure et al. 2019) was employed to find out whether the datasets were modelable or not. After developing statistically significant models with the small dataset modeler software, we have selected the modeled descriptors to perform read-across to fill the data gap. Furthermore, the inter-endpoint cytotoxicity comparison between BEAS-2B and RAW 264.7 cells was performed to evaluate the toxicological characteristics of MeOx ENPs. The models were validated according

to the OECD principles (OECD, Guidance of grouping of chemicals, Organisation of Economic Cooperation and Development, Paris, France 2007). The statistical results are given in [Table 1](#), and the values are within the acceptable criteria. The AD of the four PLS models was checked using DModX, that is, distance to model in the X-space with SIMCA 16.0.2 software. At a 99% confidence level, the AD test was performed and all the compounds were found to be within the critical value. No compound was outside the AD as per the AD plot in [Figure 3](#).

The order of importance of the descriptors of four PLS models toward the cellular uptake was determined from the Variable Importance Projections (VIPs) ([Figure 4](#)). The descriptors are ranked in decreasing order according to the VIP values. The developed PLS models were also validated using randomization tests through randomly shuffling (100 permutations) response data using SIMCA 10.0. software. Each random shuffle subsequently generates a new set of Q^2 values which are plotted

Table 1. Statistical parameters of the developed PLS models with Small Dataset Modeler.

Statistical parameters										
CELL	ASSAY	MODEL	MeOx NPs	Modeled descriptors	R ²	Q ² _{LOO}	$\bar{r}_m^2$	Δr_m^2	MAE (95%)	Standardized contribution of descriptors
BEAS-2B cell	ATP	Model 1 (M1) (LV1)	9	• Z_metal • Xox	0.931	0.871	0.849	0.027	2.154	
	MTS	Model 2 (M2) (LV2)	10	• Z_metal • Electrochemical equivalent • Valence	0.945	0.651	0.277	0.465	4.187	
RAW 264.7 Cell	MTS	Model 3 (M3) (LV1)	10	○ Atomic radius (A) ○ Z_metal	0.982	0.969	0.961	0.006	0.861	
	ATP	Model 4 (M4) (LV2)	12	○ Mol_wt ○ Z_metal ○ Nmetal	0.964	0.949	0.933	0.022	0.458	

LV: Latent variable.

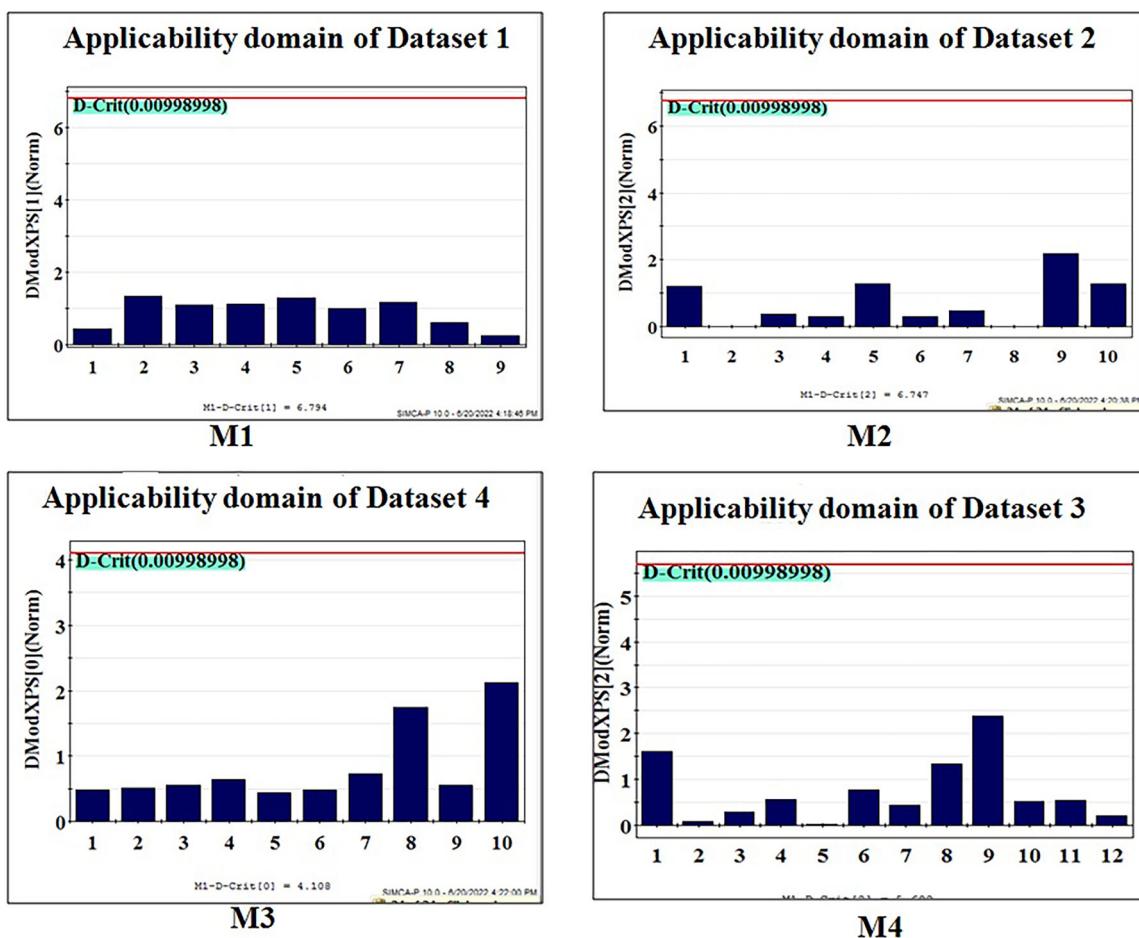


Figure 3. Applicability domain of the PLS models (M1-M4).

against the correlation coefficient with the original and permuted Y values. The Q^2 and R^2 values should yield an intercept that is below zero, indicating the robustness and applicability of the model [Figure 5](#). The obtained result showed that the models are not obtained by chance.

The loading plots were checked for the developed PLS models. The loading describes the relationship between the descriptors, that is, X variable with that of the response, that is, Y . Through the loading plot ([Figure 6](#)), the impact of descriptors on the response can be easily understood. The descriptors with a higher impact on the developed model will be situated far away from the origin.

4.1. Prediction by the small dataset modeler

PLS regression models were developed using the four cytotoxicity datasets of BEAS-2B and RAW 264.7 cells performed using the two assay methods (MTS and ATP). The small dataset modeler version 1.0.0 software was employed to check the

modelability of all the datasets. The availability of limited MeOx NP data restricted data division. The whole dataset was used for modeling by the software. The statistical results and modeled descriptors along with their contribution are given in [Table 1](#). The results obtained from the PLS models of BEAS-2B cell ($Q^2_{\text{LOO}} = 0.655-0.871$, $R^2 = 0.931-0.956$) and RAW 264.7 cell ($Q^2_{\text{LOO}} = 0.946-0.969$, $R^2 = 0.964-0.982$) were statistically significant. The definitions of the descriptors obtained from the PLS models are given in [Table 2](#).

4.2. Statistical aspects of the nano-QTTR models

The linear regression performed between cytotoxicity data for the common eight compounds from BEAS-2B cells and RAW 264.7 cells yielded a squared correlation coefficient (R^2) of 0.5 without using any kind of calculated descriptors. To get a more robust model, along with the response value of one of the cells, some periodic table descriptors were also included. Using the RAW 264.7 cell

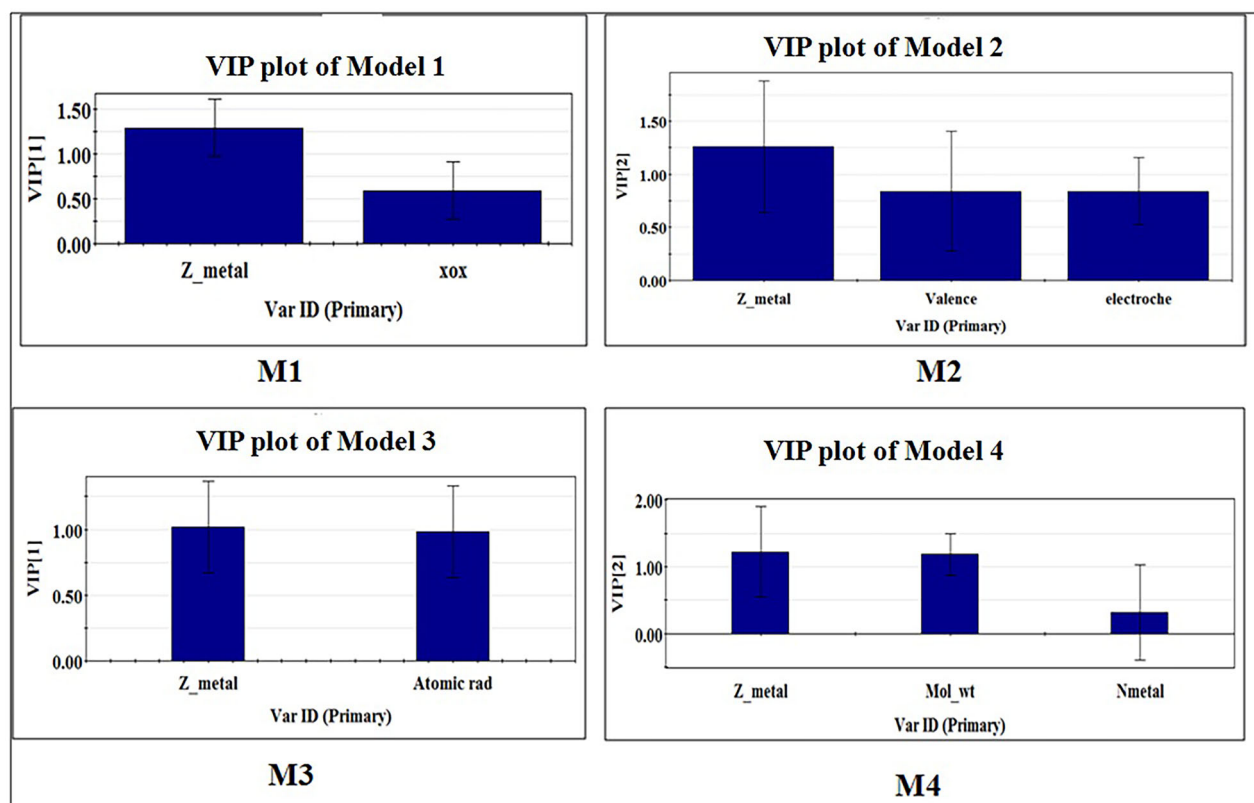


Figure 4. Variable Importance Projections (VIPs) of the PLS models (M1-M4).

response variable, the BEAS-2B cell cytotoxicity was used as a predictor variable and vice versa. The values of the leave-one-out cross-validation metrics (Q^2_{LOO}) were 0.37 and 0.50 respectively. Therefore, to get an interpretable and statistically significant model, along with some periodic table descriptors, the cytotoxicity of one cell type can be used as the predictor variable for determining the cytotoxicity of other cells.

4.2.1. RAW 264.7 cells and BEAS-2B cell toxicity correlation

Considering RAW 264.7 cell toxicity as the response and BEAS-2B cell cytotoxicity as a predictor variable along with some calculated periodic table descriptors, we have constructed a PLS model as below:

$$\log_{EC_{50}} \text{Rawcell} = -8.69557 - 0.22261 \log_{EC_{50}} \text{Beas cell} + 4.4442 X_{\text{ActivM}} \quad (1)$$

$$n = 8, R^2 = 0.90, R^2_{adj} = 0.90, Q^2_{LOO} = 0.86, \\ LV = 1, PRESS = 6.60254$$

In Equation (1), X_{ActivM} is a second-generation periodic descriptor, which stands for the electro-negativity of active metalloid/semi-metal.

4.2.2. BEAS-2B cell and RAW 264.7 cell toxicity correlation

Similarly, in this case, BEAS-2B cell toxicity was taken as the response, and RAW 264.7 cell toxicity was considered as a predictor variable along with a structural descriptor to develop a PLS model represented as Equation (2).

$$\log_{EC_{50}} \text{Beas cell} = 13.6745 - 1.2606 \log_{EC_{50}} \text{Rawcell} - 8.0142 \text{ Atomic radius} \quad (2)$$

$$n = 8, R^2 = 0.98, R^2_{adj} = 0.99, Q^2_{LOO} = 0.66, LV \\ = 1, PRESS = 100.7$$

In Equation (2), the atomic radius is a periodic table descriptor that signifies the measure of the size of the atom. It typically measures the distance from the center of the nucleus to the outermost shell containing electrons (Roy and Roy 2022). Thus, it is evident that the toxicity values of the cell lines are interrelated to each other. The MeOx NPs responsible for murine cell line toxicity are also toxic to human cells. The nano-QTTR models can be employed for extrapolating cytotoxicity to BEAS-2B and murine myeloid RAW 264.7 cells for metal

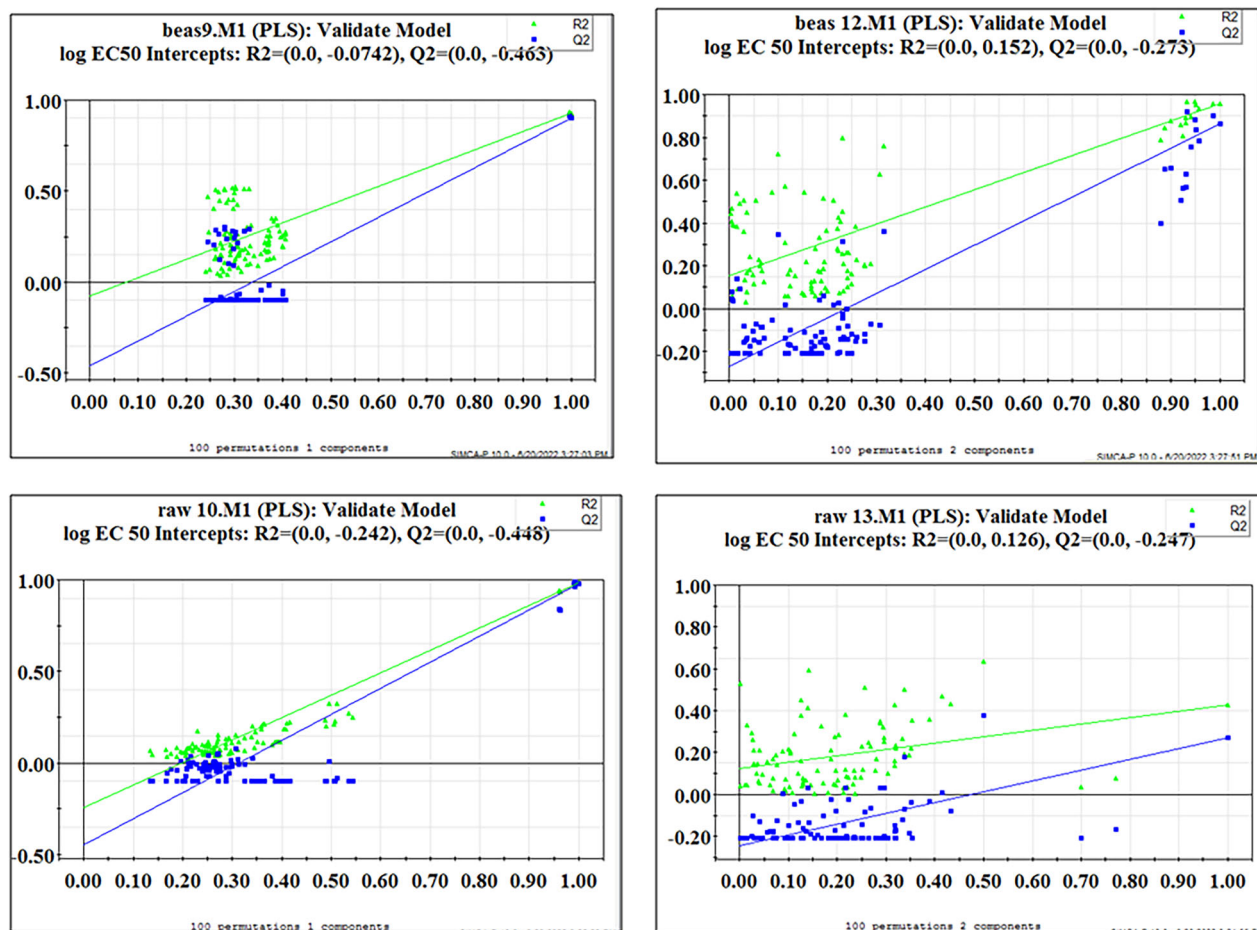


Figure 5. Y randomization plot of the PLS models (M1-M4).

nanoparticles when the data for the other species are available.

In the scarcity of experimental nanotoxicity data, these results would indicate a new approach and contribute to more possibilities for comprehensive risk assessment of nanoparticles when cytotoxicity data is missing for a specific cell line. The predicted values are given in Table 3.

4.3. Chemical read-across analysis

Read-across (RA) is a similarity-based approach for the prediction of a new compound referred to as 'target compound' based on the endpoint of other similar compounds known as the 'source' compounds. The similarity-based prediction method has gained enormous attention owing to its success in data gap-filling issues. In the present work, the quantitative-read across approach based on similarity for the four datasets was implemented using the

descriptors obtained from the four PLS models for the four datasets. The datasets were divided into training and test sets for performing the read-across technique. In this work, we tried to maintain an equal number of compounds in training and test sets. Three similarities-based measures were involved in this work namely Euclidean Distance (ED)-based Gaussian Kernel (GK) Similarity-based, and Laplacian Kernel (LK) Similarity-based prediction. After optimizing the hyperparameters, that is, σ (sigma value) and γ (gamma value) for all the datasets separately, it was observed that the read-across prediction was within acceptable criteria as given in Table 4. The work was performed using Read-Across-v4.0 (Banerjee and Roy 2022; Banerjee et al. 2022; Chatterjee et al. 2022). The threshold values for distance can range from 1 to 0, similarity threshold values from 0 to 1, and the numbers of most similar training compounds from 2 to 10. The sigma (σ) and gamma value (γ) range from 0–2.

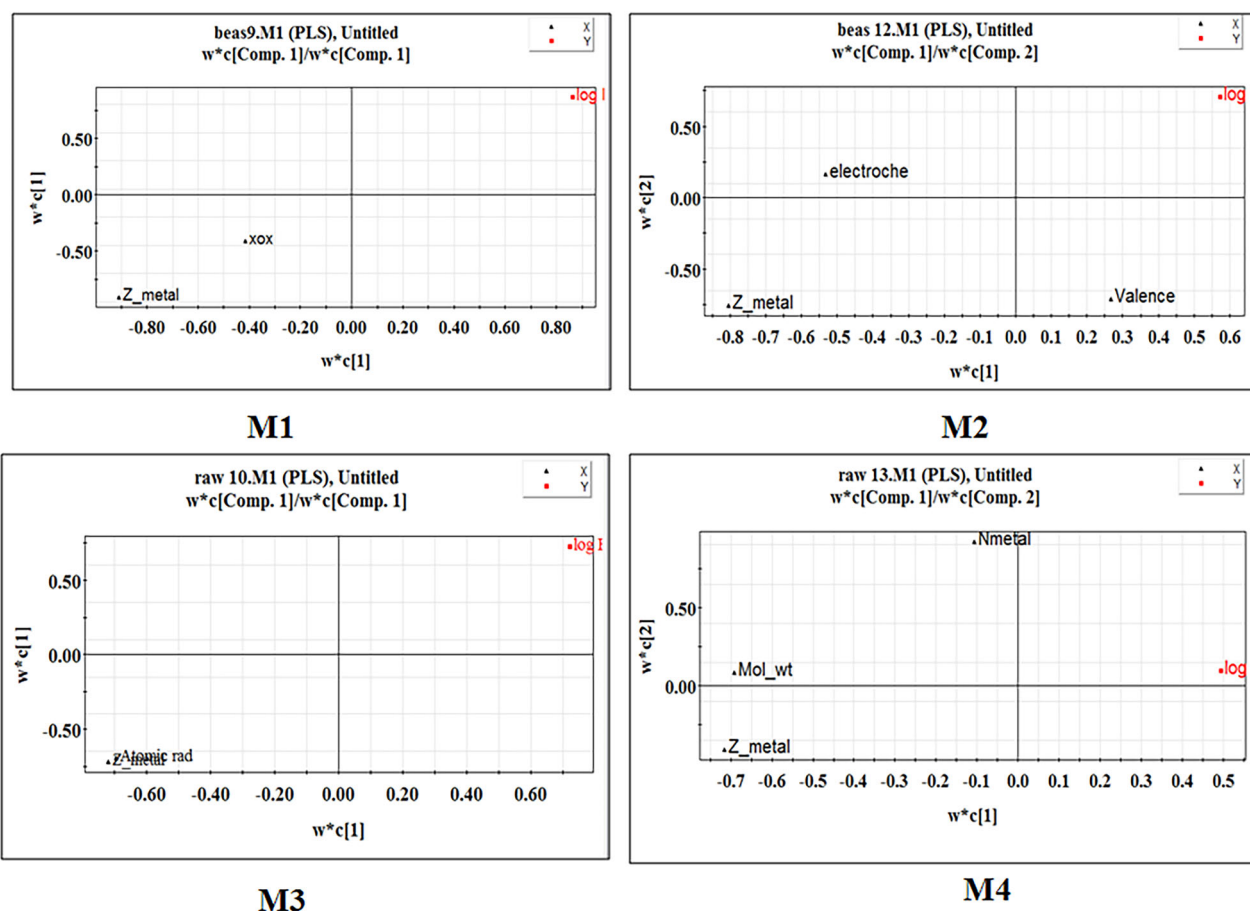


Figure 6. Loading plot of the PLS models (M1-M4).

4.4. Interpretation of the descriptors for the BEAS-2B cell toxicity model

	Model	Descriptors	Interpretation	Example
BEAS-2B cell toxicity	Model 1	Z _{metal}	The atomic number of an atom is the charge number of the nucleus. The increased electropositivity of the nucleus has a stronger pull on the valence electrons preventing them to leave the outermost shell. As a result, the valence electrons cannot form a bond or influence ROS initiation since the electrons cannot move out due to the strong electropositive attraction of the nucleus.	La ₂ O ₃ and Sb ₂ O ₃ MeOx NPs with an increase in atomic number show a decrease in the toxic effect on the cell. On the other hand, the reverse phenomenon happens in Cr ₂ O ₃ and Mn ₂ O ₃ MeOx NPs, where a decrease in atomic number elevates the toxicity of the cell.
		Xox	The oxidation number or state reflects the electronic configuration. A reducible oxide contains metal ions with a maximum oxidation state, yet can be reduced to a lower oxidation state within certain limits. Such metals contain vacant oxygen vacancies after the loss of oxygen in the crystal lattice that may exist in the bulk or on or subsurface of nanomaterials (Baek et al. 2015). These vacancies result in presence of reduced cations on the surface of nanoparticles with Lewis base character. The surface of Oxidizable oxides prevalently exhibits oxygen vacancies associated with the oxidized metals that behave like a Lewis acid, that is, they accept electrons. Concerning their Lewis acid-base behavior, this nanoparticle vacancy acts as a binding site, particularly for biomolecules that may compose bio-corona responsible for adverse pathways	The increase in the oxidation number suggests that with the increase of the oxidation number, more electrons are removed from the metal oxide as in MeOx NPs CoO and CuO, and the metal results in the formation of metal cation. Again, the opposite happens in MeOx NPs Sb ₂ O ₃ and Ni ₂ O ₃ .
	Model 2	Electro Eqv	The electrochemical equivalent (Electro Eqv) of a chemical element indicates the mass of that element (in grams) transported by 1 coulomb of electric charge. It indicates the kinetics of corrosion rate and estimates the oxidizing	The Electro Eqv descriptor negatively influences cytotoxicity in BEAS-2B, and as the descriptor value decreases, cytotoxicity increases just

(continued)

Continued.

	Model	Descriptors	Interpretation	Example
RAW 264.7 cell toxicity			power of metal in a specific environment. The number of valence electrons plays an important role here. This means that with an increase in the number of valence electrons, the descriptor value decreases. The lower oxidation state elements are found to be more harmful than the compounds with higher or stable oxidation states (Walker, Enache, and Dearden 2003). This metal ion can then radically combine with the macromolecules present inside the body and enhance toxicity through the Haber–Weiss–Fenton cycle (Koppenol 2001).	as it does in $\text{MeOxCr}_2\text{O}_3$. On the contrary, an increase of the descriptor value decreases the cytotoxicity as the wearing of the metal reduces the toxicity from its oxide as in $\text{MeOxNPFe}_2\text{O}_3$.
		Valence	The first-generation periodic table descriptor Valence measures the combining capacity of the metal with other atoms when it forms chemical compounds or molecules. The different valence states of metal oxide exhibit different characteristics and patterns of toxicity.	The negative contribution of this descriptor suggests that an increase of valence as in CoO MeOx NP decreases the toxicity whereas the decrease of the descriptor value as observed in CuO increases the cytotoxicity of the cell.
	Model 3	Atm rad (A)	The size of the atom is associated with the distribution of electrons around the nucleus. When electrons are closer to the nucleus, they are more tightly bound, thus creating a higher electronegativity in the metal. Therefore, when the atomic radius is increased, the electrons are not bound tightly, which results in a reduction in the oxide's electronegativity, thus reducing the metal's toxic effect.	This descriptor contributes negatively to the cytotoxicity which means that the latter explanation is justified as observed in MeOx NPs Yb2O3 and HfO2 where an increase of the radius causes a less toxic effect than the MeOx NPs CuO and ZnO having a larger atomic radius.
	Model 4	Mol_wt	Mol_wt describes the molecular weight of the metal. This descriptor contributes negatively to RAW 264.7 cell cytotoxicity. This happens because as the weight increases, the size of the metal oxide will also increase, and the penetration of the metal oxide through the cell will become limited. Reducing the size of the metal oxides increases the surface area, making the active size of the metal more accessible for chemical, electrical, magnetic, structural, and morphological interaction.	MeOx NPs Gd2O3 and La2O3, bearing high molecular weight has a low cytotoxicity whereas MeOx NPs CuO and CoO with low molecular weight possess higher toxicity to RAW 264.7 cells.
		N _{metal}	N _{metal} stands for the presence of metals in the metal oxide nanoparticles. The higher number of metals in the NPs can trigger more ROS production through the formation of metal cation. The higher cation formation may further accelerate bond formation with the protein in the biological system which may increase the chance of metals accumulating in the body making excretion difficult. The persistent nature of metals can cause carcinogenic, mutagenic, and teratogenic effects on the living cell that may consequently cause disease and abnormal conditions in the body (Briffa, Sinagra, and Blundell 2020).	As this descriptor contributes positively toward RAW 264.7 cell cytotoxicity, an increase of the descriptor value increases the toxicity of the cell as observed in MeOx NPs Co3O4, whereas the opposite is observed in MeOx NPs HfO2.

*The common descriptors from all the models are discussed.

4.5. Mechanism of toxicity

The assessment of the detrimental effects of nano-pesticides on living organisms demands a comprehensive understanding of their potential impact on the environment. Despite the benefits and the requirement of an agro-food section, the enormous use of agrochemicals has led to the emission of a huge runway of pesticides into the environment. Further challenges in toxicity assessment of nano-pesticides arise from other sectors through contamination of terrestrial and aquatic ecosystems and incorporation of toxic ingredients in food for animal and human consumption. Each of the individual sectors has its complexity and should be dealt with specific consideration, especially when the

pesticides are present in low concentrations and dispersed formed. Biotransformation of nanomaterials in the ecosphere may interact with various bio-films or biological organisms that can determine the fate, behavior, and persistence in the environment (Mu et al. 2014). It is noticeable that some terrestrial species come in direct contact with pesticides and hence different interactions can occur with or within the organism. The metal cation released from the metal oxide exhibit higher binding and cellular uptake to the cell than the negatively charged NPs. The charged particles induce different perturbations to the cellular membrane potential. The lower valence suggests the release of metal cation through corrosion would depolarize the cellular membrane and form holes. These holes

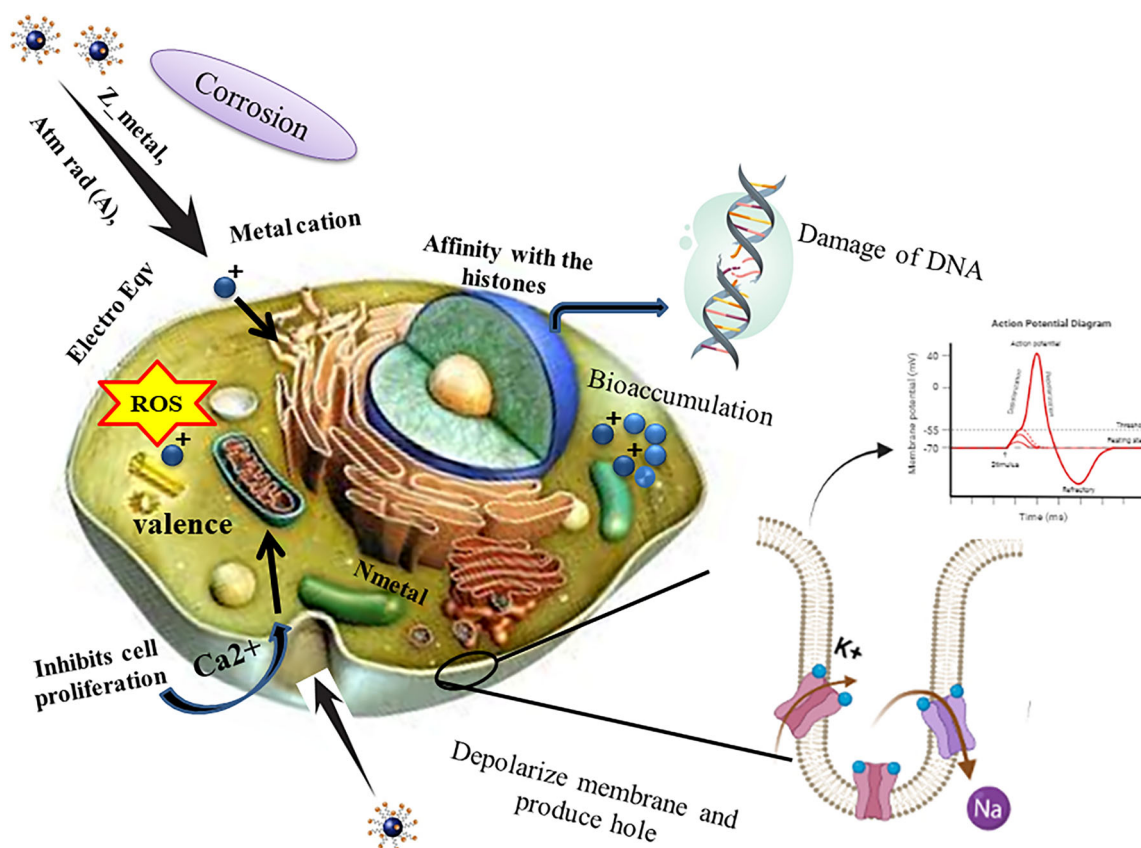


Figure 7. Mechanism of toxicity to BEAS-2B cell and RAW 264.7 cells.

Table 2. Details of the modeled descriptors from developed PLS models (M1-M4).

SL. no.	Descriptors	Description	Contribution
1	Z _{metal}	The atomic number of metals	–ve
2	χ _{ox}	Oxidation number	–ve
3	Electrochemical equivalent (Eq)	Mass of that element (in grams) transported by 1 coulomb of electric charge.	–ve
4	Valence	The valence of an element is the measure of its combining capacity with other atoms when it forms chemical compounds or molecules.	–ve
5	Mol _{wt}	Relative molecular mass	–ve
6	N _{metal}	Number of metals	+ve

Table 3. Observed and predicted values for individual compounds for BEAS-2B and murine myeloid RAW 264.7 cells lines from QTR models.

ID	Metal oxide	Observed log (1/EC ₅₀)BEAS-2B cell	Predicted log (1/EC ₅₀)BEAS-2B cell	Observed log (1/EC ₅₀)RAW 264.7 cell	Predicted log (1/EC ₅₀)RAW 264.7 cell
1	Co ₃ O ₄	–1.37	–1.41	–1.35	–1.00
2	Cr ₂ O ₃	–1.55	–2.50	–1.07	–1.93
3	Ni ₂ O ₃	–1.8	–1.75	–1.94	–1.10
4	CuO	–1.38	–0.37	–1.16	–0.56
5	Mn ₂ O ₃	–1.12	–2.06	–1.1	–2.28
6	CoO	–1.64	–1.83	–1.68	–1.06
7	ZnO	–0.56	0.27	–0.91	–1.71
8	La ₂ O ₃	–19.26	–19.02	–8.52	–8.09

lead to the initialization of metal cation inside the cell as well as diffusion of the cytosolic content out of the cell. The manifestation of the NPs is supposed to be dependent on the size and morphology of NPs. MeOx NPs can enter the cell through various paths which in return depends on the size

and possibility of interaction with various surface proteins and lipids, but the initial contact between NPs and cell (Djurišić et al. 2015) is usually through electrostatics as shown by the Electro Eqv descriptor. The decrease in size as indicated by the lower Mol_{wt} descriptor is thought to be correlated with

Table 4. Prediction quality of the features obtained from the PLS (M1-M4) models using the read-across approach.

			Hyperparameters				Statistical parameters			
			σ	γ	Distance threshold	Similarity threshold	Q^2_{F1}	Q^2_{F2}	MAE	RMSEP
Feature combinations		Hypothesis								
BEAS-2B cell	M 1 (9 MeOx NPs)	RA-ED	0.5	0.5	1	0	0.779	0.777	1.859	3.428
		RA-GK					0.993	0.993	0.506	0.622
		RA-LK					0.583	0.580	3.485	4.707
	M 2 (10 MeOx NPs)	RA-ED	1	1	1	0	0.912	0.531	3.169	3.733
		RA-GK					0.713	−0.525	5.962	6.734
		RA-LK					0.703	−0.575	6.317	6.844
RAW 264. 7 cell	M 3 (10 MeOx NPs)	RA-ED	1.5	1.5	1	0	0.953	0.895	0.889	1.067
		RA-GK					0.784	0.519	1.432	2.284
		RA-LK					−0.062	−1.365	2.572	5.064
	M 4 (12 MeOx NPs)	RA-ED	1	1	1	0	0.833	0.832	1.875	3.113
		RA-GK					0.937	0.936	1.307	1.916
		RA-LK					0.914	0.913	1.432	2.236

The best model quality is shown in the bold face.

higher uptake into the cell. The toxicity of the NPs inside the cell depends on their stability in the cellular environment (Zhu et al. 2013). The acidic surrounding of lysosomes can cause a redox reaction as demonstrated by the decrease of χ_{ox} descriptor. The binding properties of the NPs are explained by the Z_{metal} , as the decrease of this descriptor increases the cellular interaction leading to damage to the membrane and the production of ROS (reactive oxygen species). NPs can enter the mitochondria (the powerhouse of the cell) and affect their function. The nuclear envelope inhibits the entry of large NPs into the nucleus. The small-sized NPs enter the nucleus and exhibit strong affinity with the histones or other cell organelles through electronegativity. The NPs may cause decondensation of nuclear chromatin in mice which is driven by the interaction of NPs with the nucleus. The nano-specific behavior also intrinsically depends on the metallic nature of NPs used in pesticides as the descriptor N_{metal} . Their toxicity is associated with internalization and bioaccumulation by the non-targeted organism (Firdaus et al. 2018; Fojtová et al. 2019). Such interaction and bioaccumulation may alter oxidative stress-dependent and -independent pathways that affect proliferation and differentiation. Other signaling pathways such as MAPK, NF- κ B, and DNA damage are responsible for nanoparticle-induced oxidative stress. All these events as in Figure 7 eventually cause the death of the respective cell (Fraceto et al. 2020).

5. Conclusion

Advancement in nanotechnology has paved innovations in the agro-food sector. Research work has

focused to develop a nanoscale formulation of pesticides to enhance the efficacy and production of food. Nanotechnology has provided us with enormous opportunities while imposing a potential threat to human beings and living organisms. The small size particulates have raised a question about their safety and their sustainability in the environment. It has been seen that the small-sized particulates with a large surface area led to unintended ecotoxicological effects on non-target organisms when applied to the agricultural field. Therefore, understanding the benefits and potential risks of MeOx ENPs used in food as pesticides are urgent and critical. Thus, in the present work, we have tried to identify the main features governing the toxicity of BEAS-2B and RAW 264.7 cell lines derived from four datasets. Periodic table-based descriptors were used to reveal the main features responsible for toxicity to the corresponding cells through the development of predictive QSTR models. Considering the knowledge and current ecotoxicological status of MeOx ENPs, PLS regression and the read-across approaches were used. Furthermore, the interspecies toxicity correlation study was performed between BEAS-2B and RAW 264.7 cells to fill the data gap that needs to be addressed for safer and sustainable use of pesticides. Particular emphasis was given to determining the fate and toxicological behavior of nanopesticides in the environmental compartments. The studies examined the role of metal cation formation, electropositivity, binding capacity, and bioaccumulation of nanoparticles in contributing to toxicity. The periodic table descriptors that can be easily computed can successfully explain the cause of toxicity. The

information derived from the PLS models of both the cell lines suggested that metal cation formation, and bond-forming capacity influence the toxicity, whereas the presence of metal has an influential impact on the ecotoxicological effects. All the descriptors in the present study were found to have a negative impact on the toxicity except number of metals which has a positive impact. It is also worth mentioning here that the toxicity depends on the organism type. Therefore, the knowledge derived from the work will not only help in data gap filling but also help in substantial enhancement in efficacy and stability of formulation of nanopesticides.

Acknowledgement

JR thankfully acknowledges Indian Council of Medical Research, (ICMR) New Delhi for supporting financially to the work in the form of Senior Research Fellowship. KR acknowledges financial assistance from SERB, New Delhi to KR under a MATRICS project (Project No. MTR/2019/000008).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China (82070699) and Clinical Application Research Program of the Xijing Hospital (JSYXZ05). This work was also supported by the Nephrology Department of Xijing Hospital and the Fourth Military Medical University.

ORCID

Kunal Roy  <http://orcid.org/0000-0003-4486-8074>

References

- Ambure, P., A. Gajewicz-Skretna, M. N. D. Cordeiro, and K. Roy. 2019. "New Workflow for QSAR Model Development from Small Data Sets: small Dataset Curator and Small Dataset Modeler. integration of Data Curation, Exhaustive Double Cross-Validation, and a Set of Optimal Model Selection Techniques." *Journal of Chemical Information and Modeling* 59 (10): 4070–4076. doi:10.1021/acs.jcim.9b00476.
- Baek, S., R. K. Singh, D. Khanal, K. D. Patel, E. J. Lee, K. W. Leong, W. Chrzanowski, and H. W. Kim. 2015. "Smart Multifunctional Drug Delivery towards Anticancer Therapy Harmonized in Mesoporous Nanoparticles." *Nanoscale* 7 (34): 14191–14216. doi:10.1039/C5NR02730F.
- Ball, Nicholas, Mark T. D. Cronin, Jie Shen, Karen Blackburn, Ewan. D. Booth, Mounir Bouhifd, Elizabeth Donley, et al. 2016. "T4 Report: Toward Good Read-across Practice (GRAP) Guidance." *ALTEX* 33 (2): 149–166. doi:10.14573/altex.1601251.
- Banerjee, Arkaprava, Mainak Chatterjee, Priyanka De, and Kunal Roy. 2022. "Quantitative Predictions from Chemical Read-Across and Their Confidence Measures." *Chemometrics and Intelligent Laboratory Systems* 227: 104613. doi:10.1016/j.chemolab.2022.104613.
- Banerjee, A, and K. Roy. 2022. "First Report of q-RASAR Modeling toward an Approach of Easy Interpretability and Efficient Transferability." *Molecular Diversity* 26 (5): 2847–2862. doi:10.1007/s11030-022-10478-6.
- Batista Caixeta, M., P. S. Araújo, B. Gonçalves, L. Damacena Silva, M. I. Grano-Maldonado, and T. Lopes Rocha. 2020. "Toxicity of Engineered Nanomaterials to Aquatic and Land Snails: A Scientometric and Systematic Review." *Chemosphere* 260: 127654. doi:10.1016/j.chemosphere.2020.127654.
- Briffa, J., E. Sinagra, and R. Blundell. 2020. "Heavy Metal Pollution in the Environment and Their Toxicological Effects on Humans." *Heliyon* 6(9): e04691. September 2020, doi:10.1016/j.heliyon.2020.e04691.
- Buchman, J. T., N. V. Hudson-Smith, K. M. Landy, and C. L. Haynes. 2019. "Understanding Nanoparticle Toxicity Mechanisms to Inform Redesign Strategies to Reduce Environmental Impact." *Accounts of Chemical Research* 52 (6): 1632–1642. doi:10.1021/acs.accounts.9b00053.
- Chatterjee, M., A. Banerjee, P. De, A. Gajewicz-Skretna, and K. Roy. 2022. "A Novel Quantitative Read-across Tool Designed Purposefully to Fill the Existing Gaps in Nanosafety Data." *Environmental Science: Nano* 9 (1): 189–203. doi:10.1039/D1EN00725D.
- Colman, M., B. P. Tang, W. Judy, J. D. Anderson, S. M. Bergemann, and E. S. C.M. Bernhardt. 2018b. "Plant and Microbial Responses to Repeated Cu(OH)₂ nanopesticide Exposures under Different Fertilization Levels in an Agroecosystem Front." *Journal of Microbiology and Biotechnology* 9: 1769. doi:10.3389/fmicb.2018.01769.
- Darlington, R. B. 1990. *Regression and Linear Models*. McGrawHill, NewYork.
- De, P., S. Kar, K. Roy, and J. Leszczynski. 2018. "Second Generation Periodic Table-Based Descriptors to Encode Toxicity of Metal Oxide Nanoparticles to Multiple Species: QSTR Modeling for Exploration of Toxicity Mechanisms." *Environmental Science: Nano* 5 (11): 2742–2760. doi:10.1039/C8EN00809D.
- Devillers, J. 1996. *Genetic algorithms in molecular modeling*. NY: Academic Press.
- Djurišić, A. B., Y. H. Leung, A. M. Ng, X. Y. Xu, P. K. Lee, N. Degger, and R. S. S. Wu. 2015. "Toxicity of Metal Oxide Nanoparticles: mechanisms, Characterization, and Avoiding Experimental Artefacts." *Small (Weinheim an Der Bergstrasse, Germany)* 11 (1): 26–44. doi:10.1002/smll.201303947.
- EFSA Scientific Committee, 2011. "Guidance on the Risk Assessment of the Application of Nanoscience and Nanotechnologies in the Food and Feed Chain." *European*

- Food Safety Authority* 9 (5): 2140. doi:[10.2903/j.efsa.2011.2140](https://doi.org/10.2903/j.efsa.2011.2140).
- Firdaus, M. A. M., A. Agatz, M. E. Hodson, O. S. A. Al-Khazrajy, and A. B. A. Boxall. 2018. "Fate, Uptake, and Distribution of Nanoencapsulated Pesticides in Soil-Earthworm Systems and Implications for Environmental Risk Assessment." *Environmental Toxicology and Chemistry* 37 (5): 1420–1429. doi:[10.1002/etc.4094](https://doi.org/10.1002/etc.4094).
- Fojtová, D., J. Vašíčková, R. Grillo, Z. Bílková, Z. Šimek, N. Neuwirthová, M. Kah, and J. Hofman. 2019. "Nanoformulations Can Significantly Affect Pesticide Degradation and Uptake by Earthworms and Plants." *Environmental Chemistry* 16 (6): 470–481. doi:[10.1071/EN19057](https://doi.org/10.1071/EN19057).
- Fraceto, L. F., V. L. S. De Castro, R. Grillo, D. Ávila, H. C. Oliveira, and R. Lima. 2020. *Nanopesticides*. Cham, Switzerland: Springer International Publishing.
- Gadaleta, D., G. F. Mangiatordi, M. Catto, A. Carotti, and O. Nicolotti. 2016. "Applicability Domain for QSAR Models: where Theory Meets Reality." *International Journal of Quantitative Structure-Property Relationships* 1 (1): 45–63. doi:[10.4018/IJQSPR.2016010102](https://doi.org/10.4018/IJQSPR.2016010102).
- Gajewicz, A. 2017. "Development of Valuable Predictive Read-across Models Based on "Real-Life"(Sparse) Nanotoxicity Data." *Environmental Science: Nano* 4 (6): 1389–1403. doi:[10.1039/C7EN00102A](https://doi.org/10.1039/C7EN00102A).
- Gajewicz, A., M. T. Cronin, B. Rasulev, J. Leszczynski, and T. Puzyn. 2015. "Novel Approach for Efficient Predictions Properties of Large Pool of Nanomaterials Based on Limited Set of Species: nano-Read-across." *Nanotechnology* 26 (1): 015701. doi:[10.1088/0957-4484/26/1/015701](https://doi.org/10.1088/0957-4484/26/1/015701).
- Ganguly, P., A. Breen, and S. C. Pillai. 2018. "Toxicity of Nanomaterials: Exposure, Pathways, Assessment, and Recent Advances." *ACS Biomaterials Science & Engineering* 4 (7): 2237–2275. doi:[10.1021/acsbomaterials.8b00068](https://doi.org/10.1021/acsbomaterials.8b00068).
- Kah, M. 2015. "Nanopesticides and Nanofertilizers: emerging Contaminants or Opportunities for Risk Mitigation?" *Frontiers in Chemistry* 3: 64. doi:[10.3389/fchem.2015.00064](https://doi.org/10.3389/fchem.2015.00064).
- Kookana, Rai. S., Alistair B. A. Boxall, Philip. T. Reeves, Roman Ashauer, Sabine Beulke, Qasim Chaudhry, Geert Cornelis, et al. 2014. "Nanopesticides: guiding Principles for Regulatory Evaluation of Environmental Risks." *Journal of Agricultural and Food Chemistry* 62 (19): 4227–4240. doi:[10.1021/jf500232f](https://doi.org/10.1021/jf500232f).
- Koppenol, W. H. 2001. "The Haber-Weiss Cycle–70 Years Later." *Redox Report : communications in Free Radical Research* 6 (4): 229–234. doi:[10.1179/135100001101536373](https://doi.org/10.1179/135100001101536373).
- Liu, R., H. Y. Zhang, Z. X. Ji, R. Rallo, T. Xia, C. H. Chang, A. Nel, and Y. Cohen. 2013. "Development of Structure–Activity Relationship for Metal Oxide Nanoparticles." *Nanoscale* 5 (12): 5644–5653. doi:[10.1039/C3NR01533E](https://doi.org/10.1039/C3NR01533E).
- Mu, Q., G. Jiang, L. Chen, H. Zhou, D. Fourches, A. Tropsha, and B. Yan. 2014. "Chemical Basis of Interactions between Engineered Nanoparticles and Biological Systems." *Chemical Reviews* 114 (15): 7740–7781. doi:[10.1021/cr400295a](https://doi.org/10.1021/cr400295a).
- OECD, Guidance of grouping of chemicals, Organisation of Economic Cooperation and Development, Paris, France 2007. 26 *OECD, Guidance on Grouping of Chemicals. Series on Testing and Assessment No. 194* (second ed.), Organisation of Economic Cooperation and Development, Paris, France, 2014.
- Roy, J, and K. Roy. 2021. "Assessment of Toxicity of Metal Oxide and Hydroxide Nanoparticles Using the QSAR Modeling Approach." *Environmental Science: Nano* 8 (11): 3395–3407. doi:[10.1039/D1EN00733E](https://doi.org/10.1039/D1EN00733E).
- Roy, J, and K. Roy. 2022. "Modeling and Mechanistic Understanding of Cytotoxicity of Metal Oxide Nanoparticles (MeOxNPs) to Escherichia coli: Categorization and Data Gap Filling for Untested Metal Oxides." *Nanotoxicology* 16 (2): 152–164. doi:[10.1080/17435390.2022.2038299](https://doi.org/10.1080/17435390.2022.2038299).
- Roy, K., I. Mitra, S. Kar, P. K. Ojha, R. N. Das, and H. Kabir. 2012. "Comparative Studies on Some Metrics for External Validation of QSPR Models." *Journal of Chemical Information and Modeling* 52 (2): 396–408. doi:[10.1021/ci200520g](https://doi.org/10.1021/ci200520g).
- Roy, K., R. N. Das, P. Ambure, and R. B. Aher. 2016. "Be Aware of Error Measures. Further Studies on Validation of Predictive QSAR Models." *Chemometrics and Intelligent Laboratory Systems* 152: 18–33. doi:[10.1016/j.chemolab.2016.01.008](https://doi.org/10.1016/j.chemolab.2016.01.008).
- Simonin, M., B. P. Colman, W. Tang, J. D. Judy, S. M. Anderson, C. M. Bergemann, J. D. Rocca, J. M. Unrine, N. Cassarand and E. S. Bernhardt. 2018. "Plant and microbial responses to repeated Cu (OH) 2 nanopesticide exposures under different fertilization levels in an agro-ecosystem." *Frontiers in microbiology* 9, 176. doi:[10.3389/fmicb.2018.01769](https://doi.org/10.3389/fmicb.2018.01769).
- Sizochenko, N., A. Mikolajczyk, K. Jagiello, T. Puzyn, J. Leszczynski, and B. Rasulev. 2018. "How the Toxicity of Nanomaterials towards Different Species Could Be Simultaneously Evaluated: A Novel Multi-Nano-Read-across Approach." *Nanoscale* 10 (2): 582–591. doi:[10.1039/C7NR05618D](https://doi.org/10.1039/C7NR05618D).
- Uboldi, Chiara, Marcos Sanles Sobrido, Elodie Bernard, Virginie Tassistro, Nathalie Herlin-Boime, Dominique Vrel, Sébastien Garcia-Argote, et al. 2019. "In Vitro H.analysis of the Effects of ITER-like Tungsten Nanoparticles: Cytotoxicity and Epigenotoxicity in BEAS-2B Cells." *Nanomaterials* 9 (9): 1233. doi:[10.3390/nano9091233](https://doi.org/10.3390/nano9091233).
- Venkatasubramanian, V, and A. Sundaram. 2002. "Genetic Algorithms: introduction and Applications." *Encyclopedia of Computational Chemistry* 2, Wiley. doi:[10.1002/0470845015.cga003](https://doi.org/10.1002/0470845015.cga003).
- Walker, J. D., M. Enache, and J. C. Dearden. 2003. "Quantitative Cationic-Activity Relationships for Predicting Toxicity of Metals." *Environmental Toxicology and Chemistry* 22 (8): 1916–1935. doi:[10.1897/02-568](https://doi.org/10.1897/02-568).

- Wold, S. 1995. "PLS for Multivariate Linear Modelling." In *Chemometric Methods in Molecular Design*, edited by vandeWaterbeemd, H. 195–218. VCH: Weinheim, Germany.
- Zhang, Haiyuan, Zhaoxia Ji, Tian Xia, Huan Meng, Cecile Low-Kam, Rong Liu, Suman Pokhrel, et al. 2012. "Use of Metal Oxide Nanoparticle Band Gap to Develop a Predictive Paradigm for Oxidative Stress and Acute Pulmonary Inflammation." *ACS Nano* 6 (5): 4349–4368. doi: [10.1021/nn3010087](https://doi.org/10.1021/nn3010087).
- Zhu, M., G. Nie, H. Meng, T. Xia, A. Nel, and Y. Zhao. 2013. "Physicochemical Properties Determine Nanomaterial Cellular Uptake, Transport, and Fate." *Accounts of Chemical Research* 46 (3): 622–631. doi: [10.1021/ar300031y](https://doi.org/10.1021/ar300031y).



Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: a nano-QSAR approach

J. Roy & K. Roy

To cite this article: J. Roy & K. Roy (2023): Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: a nano-QSAR approach, SAR and QSAR in Environmental Research, DOI: [10.1080/1062936X.2023.2227557](https://doi.org/10.1080/1062936X.2023.2227557)

To link to this article: <https://doi.org/10.1080/1062936X.2023.2227557>

 View supplementary material 

 Published online: 23 Jun 2023.

 Submit your article to this journal 

 View related articles 

 View Crossmark data 



Understanding mechanism governing the inflammatory potential of metal oxide nanoparticles using periodic table-based descriptors: a nano-QSAR approach

J. Roy and K. Roy

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

ABSTRACT

Metal oxide nanoparticles (MeOxNPs) can be made safer by understanding the interaction between the immune system and nanoparticles. A nano-quantitative structure–activity relationship (nano-QSAR) model can be used to evaluate nanoparticle risk quickly and conveniently. The present work attempts to develop nano-QSAR models to determine the inflammatory potential of MeOxNPs based on the THP-1 cell line. A comprehensive dataset comprising 32 MeOxNPs was used to develop a regression model with fold change (FC) of pro-inflammatory cytokine interleukin (IL)-1 β (IL-1 β) release in the THP-1 cell line as the endpoint. Further, the same number of MeOx NPs with varying doses was modelled for the cell viability [$-\ln(p/(1-p))$] endpoint. The results obtained from regression models were statistically significant. The descriptors obtained from the developed predictive models inferred that dose, electronegativity and the presence of metal ions and oxygen can be responsible for IL-1 β leakage from the THP-1 cell line. Based on our results, we can conclude that periodic table-based descriptors, incorporated into the QSAR models, are reliable for modelling pro-inflammatory potential. Researchers can use these comprehensive results to design metal oxide nanoparticles with lower toxicity and determine the cause of pro-inflammatory conditions induced by MeOxNPs.

ARTICLE HISTORY

Received 29 May 2023
Accepted 15 June 2023

KEYWORDS

Nano-QSAR; periodic table-based descriptors; immune system; MeOx NPs

Introduction

Nanoparticles provide an immense opportunity to prepare new materials with desired properties due to their unique attributes. The rapid introduction of nanomaterial products to the market makes nanomaterial research one of the fastest-growing contemporary fields in research areas. It is believed that most of the products now available may contain nanomaterials in them in one or the other form. The global nanotechnology market was valued at approximately \$14 in 2014 [1] with an annual growth rate of 20%. There is a continuous demand for new materials with useful properties but less deleterious to the ecosystem and human health [2]. The pace at which these nanomaterials are commercialized has left the regulatory agencies struggling for proper models for their safety

CONTACT K. Roy ✉ kunalroy_in@yahoo.com; kunal.roy@jadavpuruniversity.in

Supplemental data for this article can be accessed at: <https://doi.org/10.1080/1062936X.2023.2227557>

© 2023 Informa UK Limited, trading as Taylor & Francis Group

prediction. The 'precautionary principles' applied for veterinary chemicals and pharmaceuticals fail for nanomaterials which make serious consequences for their premature translocation to the market. Therefore, a toxicity check of nanomaterials should be a prerequisite for any kind of application. Nanomaterials are highly heterogeneous which represents challenges for the assessment of their possible adverse biological effects, which is more complex than drugs or other industrial chemicals. Different types of nanoparticles are being used by industries but little has been known about their possible toxicity (mutagenic and/or carcinogenic/toxic effect) [3]. Some recent finding reports the toxicity of selected NPs and highlights the potential risk associated with the engineering of such NPs. There is insufficient information regarding the harmful interaction of NPs with biological systems and the environment. Regarding nanomaterials, limitations stem from the absence of proper databases on their structures, physiological properties and endpoints [4]. In addition, the toxic effect varies according to the physicochemical properties and the same type of nanomaterials may exhibit diverse toxicity to various biochemical conditions. Hence, knowledge gaps prevail due to a lack of databases, rules for complement activation and inhibition and immune response predictions.

Organic chemicals are identical with predictable properties that can easily be encoded with mathematical objects, i.e. molecular descriptors. On the other hand, nanomaterials exist as a population of different forms and their effect upon interaction with the biological system is context-dependent. As a result, a complete characterization of nanomaterials is required to evaluate their toxicity profile. A subtle change in the characteristics of NPs can lead to a change in the property of interest. Thus, selecting the appropriate descriptors is crucial. At present, nano-descriptors can be obtained either from experimental characterization or theoretical calculations. Ideally, nano-descriptors should reflect the chemical composition, chemical structures as well as physicochemical characteristics related to NPs [5,6]. In experimental characteristics, there can be the constant implementation of new parameters that can change with varying experimental conditions for different NPs. Thus, nano-descriptors derived from experimentation conducted with diverse conditions or different protocols can introduce additional sources of uncertainty for new NP descriptor calculation. However, size-independent descriptors can be a good alternative for analysing NPs toxicity. The achievement of such descriptors can exhibit toxicological responses without the involvement of any kind of conditions. Periodic-table descriptors such as electronegativity and the charge of the metal cation corresponding to a given oxide (χ_{ox}) can be easily calculated, and these basic properties are independent of the size range of NPs nullifying a typical problem that many physical properties of NPs change for different size ranges.

The 'nanotoxicology' term was first introduced by Oberdörster [7] in 2005, and much modification has evolved ever since. It is now well referred to as 'nanosafety' which studies the influence of nanomaterials on biological responses. Modern modelling methods, specifically, *in silico* methods that were developed to predict the biological effects of drugs or chemicals, play an important part in leveraging the limited data. If a model can exhibit the adverse biological effects of nanomaterials, then extrapolation of properties of untested materials can be inferred from the developed models. Such approaches are well adopted by regulatory agencies for prioritizing nanomaterials for mammalian and environmental toxicity testing.

Statistical methods such as Quantitative Structure–Activity Relationship (QSAR) are extensively used for the toxicity prediction of nanomaterials. Several researchers have shown that nano-QSAR modelling can directly predict the toxicity of nanomaterials [8–10]. However, only a few studies related to metal oxide nanoparticles have been reported till now. Puzyn et al. [10] developed a model to describe the cytotoxicity of 17 metal oxide nanoparticles (MeOxNPs) to bacteria *Escherichia coli* using quantum chemical-based descriptors to obtain a model with $r^2 = 0.862$. Another group of researchers [11] established a nano-QSAR model for assessing the cytotoxicity of MeOxNPs using metrics based on the dose-response analysis along with consensus self-organizing map clustering for human bronchial epithelial (BEAS-2B) and murine myeloid (RAW264.7) cells. Further using the same dataset, Toropov et al. [12] reported QSAR models using CORAL software (web-based) with SMILES-based optimal descriptors. Good internal and external results in terms of r^2 (Training) and r^2 (Test) values falling within the ranges 0.740–0.837 and 0.83–0.96, respectively, were observed after six random splits of data into training and test sets. However, most of the published models were established with traditional toxicity endpoints, i.e. cell death. Beyond merely killing living organisms, nanoparticles (MeOxNPs) cause subtle effects on the immune system [13,14]. Some articles have considered the toxicity of MeOxNPs to immune cells, but there is hardly any in silico prediction model for MeOxNPs involving the immune system. The understanding of the effect of nanomaterials on the immune system is still vague. It is believed that the reticuloendothelial system (RES) eliminates NPs, but many studies have suggested the interaction of nanoscale particles with the defence activities of immune cells, inducing strong immune responses via inflammation and toll-like receptor activation as evident from the release of a substantial amount of the pro-inflammatory cytokine interleukin-1 beta (IL-1b) [13,15]. Therefore, from the nano-safety perspective and for advancing toxicity assessment, predictive modelling of the inflammatory potential of MeOxNPs on the immune system should be performed.

Hence, we have developed here QSAR models to predict the inflammatory potential of 32 MeOxNPs. The whole work includes 1) developing interpretable QSAR models for predicting the inflammatory potential of MeOxNPs, 2) validating and predicting the models with independent MeOxNPs and 3) exploring MeOxNP induced inflammatory mechanism.

Materials and methods

Dataset

The toxicity data used for model development were obtained from the literature [16]. Thirty-two MeOxNPs were analysed for fold change ($FC_{IL-1\beta}$) and cell viability. The doses for the cell viability dataset were 100–200 $\mu\text{g/mL}$, which resulted in 54 data points. Here, two datasets were used: dataset 1 with 32 MeOxNPs with fold change ($FC_{IL-1\beta}$) as the endpoint and dataset 2 with 54 MeOxNPs that have cell viability [$-\ln(p/(1-p))$, p being the per cent cell viability] as the endpoint and considering 100–200 $\mu\text{g/mL}$ doses.

We have changed the cell viability dataset endpoint to [$-\ln(p/(1-p))$] for a better understanding of the toxicity (Supplementary Materials S1). The release of cytokine

interleukin (IL)-1beta (IL-1b) from the THP-1 cells and negative logit transformed per cent cell viability $[-\ln(p/(1-p))]$ were used as indicators for the toxicity assessment and thus were used for further modelling purposes. The THP-1 cell line can be used in immune system disorder research, immunology research and toxicology research.

The Fold Changes of IL-1b production ($FC_{IL-1\beta}$) were expressed as the level of MeOxNPs induced IL-1b release in THP-1 cells. It was calculated using the following equation:

$$FC_{IL-1\beta} = \frac{avg(Sample)}{avg(Control)}$$

Here, avg (Sample) stands for the average level of IL-1b ($\mu\text{g/mL}$) induced by specific MeOx NPs at each dose, and avg (Control) stands for the average level of IL-1b ($\mu\text{g/mL}$) for untreated THP-1 cell lines.

Periodic table descriptor calculations

We have taken some of the basic pieces of information from the periodic table to calculate descriptors for MeOx NPs. The metal's information was taken directly from the periodic table and calculated with the Elemental Descriptor Calculator software available from (<https://dtclab.webs.com/software-tools>) are known as the first-generation periodic table descriptors, and those calculated using some relevant formulas are known as the second-generation periodic table descriptors [17]. A total of 44 descriptors were calculated out of which 28 were first generation and 16 were second generation periodic table-based descriptors. The list of second-generation descriptors and a few first-generation descriptors are provided in Supplementary Materials S2 (Table S1). These descriptors were calculated in no time which is convenient as well as cost-effective. It is worth mentioning here that these descriptors are independent of nanoparticle size, and these descriptors can be utilized in the mixture calculation of MeOxNPs which is problematic with quantum chemical descriptors.

Model development

We have developed MLR (multiple linear regression) models with $FC_{IL-1\beta}$ (dataset 1) and PLS model with $[-\ln(p/(1-p))]$ (dataset 2) datasets. For regression modelling at first, linear regression was performed with the first set of data comprising 32 MeOxNPs (dataset 1). After randomly dividing the training and test sets (70:30), feature selection was carried out using the Genetic algorithm (GA). With the selected features, multiple linear regression was performed using the MINITAB Software (MINITAB is a Statistical Software of Minitab Inc., USA; <http://www.minitab.com>) for model development. A schematic diagram of the whole process is illustrated in Figure 1. For PLS regression modelling with the cell viability $[-\ln(p/(1-p))]$ endpoint (dataset 2), we have performed stepwise regression in MINITAB software. Further, partial linear regression (PLS) analysis was performed with the features obtained from the stepwise regression procedure followed by best subset selection (BSS) [18].

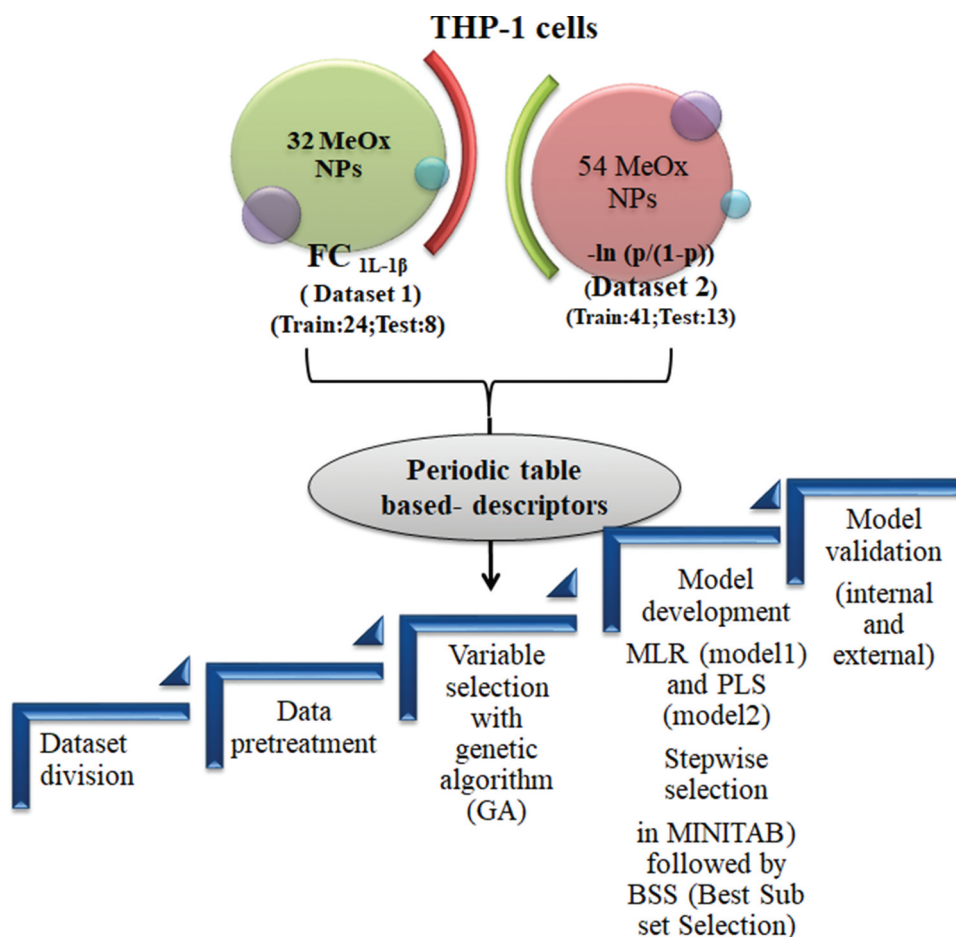


Figure 1. Workflow of the present work.

Validation metrics

All the requisite internal and external validation metrics were computed using the training and test sets, respectively, for the MLR models. The fitting potential of regression-based models was determined based on some statistical metrics like the determination coefficient (r^2), the correlation coefficient of the model (Q^2) and root-mean-square error (RMSE) for internal prediction. External validation was checked by the validation metric Q^2_{F1} . The r_m^2 metrics namely r_m^2 and Δr_m^2 developed by Roy [19] were also employed for the present work. Further, the reliability of the developed MLR model developed using 32 data points was checked on a true external set using the PRI (Prediction Reliability Indicator) tool from the DTC website (<https://dtclab.webs.com/software-tools>). The reliability of predictions was checked on 37 MeOx NPs that lacked experiment data. This tool categorizes the quality of prediction of true external set into three groups i) good (with composite score 3), ii) moderate (with composite score 2) and iii) bad (with composite score 1) [20].

Applicability domain (AD)

A QSAR model's applicability domain (AD) defines the space of responses and chemical structures in which it makes predictions with a given degree of accuracy. The idea was introduced to avoid unjustified extrapolation of predictions. In this study, we have checked the AD for the MLR model with the standardization approach [21] along with the leverage approach [22,23]. Both approaches help to determine whether the compound is within the chemical domain of the model or outside. In the standardization approach, all descriptor values should follow a normal distribution pattern. According to this distribution, most of the compounds should be present within the mean ± 3 standard deviation (SD). Any compound outside this zone is considered dissimilar. On the other hand, the leverage approach considers the leverage values (h) of the compounds; the prediction is said to be reliable if the ' h ' is less than the critical value (h^*). The ' h^* ' is calculated as $h^* = 3p'/n$, where p' stands for the modelled variables plus one and ' n ' stands for the number of compounds used in model development. The plot (William's plot) of leverages (Hat diagonal) vs. cross-validated standardized residuals (R) helps to detect the outlier. Compounds with cross-validation standardized residual greater than 2.5 standard deviation units, i.e. $>2.5\sigma$ can be considered as outliers.

Results and discussions

Successful development of the QSAR model depends on the quality of biological data. However, the available MeOxNPs data presented here are limited and are sufficient for developing predictive models. Many authors have developed and published QSAR models with a small number of data points [10,24,25]. The criterion of a useful model is the difference between r^2 and Q^2_{loo} being less than 0.3. With this, the MLR model ($FC_{IL-1\beta}$ dataset) satisfies this criterion ($r^2 = 0.893$, $Q^2_{loo} = 0.86$, $Q^2_{F1} = 0.86$). Equation (1) suggested that the developed MLR model could be applied to evaluate the inflammatory potential of new, untested MeOxNPs (Table 1). The results obtained from the cell viability $[-\ln(p/(1-p))]$ dataset ($r^2 = 0.72$, $Q^2_{loo} = 0.53$, $Q^2_{F1} = 0.66$) with 54 data points and latent variable 4 Equation (2), could explain 72% of variance while it could predict 53% of leaving-one-out (LOO) predicted variance. The scatter plots (Figure 2) suggest that the models have good predictions.

The AD checked with the standardization approach showed that for Model 1, compound no. 28 was an outlier in the training compounds, whereas there was no outlier compound in the case of Model 2. No compound was found to be outside the applicability domain for the test sets for models, as per the William plot (Figure 3).

Table 1. Results of MLR and PLS models with $FC_{IL-1\beta}$ and $-\ln(p/(1-p))$ endpoints, respectively.

Model	Descriptors	Number of data points	Endpoints	r^2	Q^2_{loo}	Q^2_{F1}
MLR (M1)	Nmetal, $\Sigma\chi$, atomic radius	32	$FC_{IL-1\beta}$	0.893	0.86	0.86
PLS (M2)	$\frac{\Sigma\chi}{nO}$, χ , $Vp\phi$, atomic radius, dose, $\Sigma\chi$	54	$-\ln(p/(1-p))$	0.72	0.53	0.66

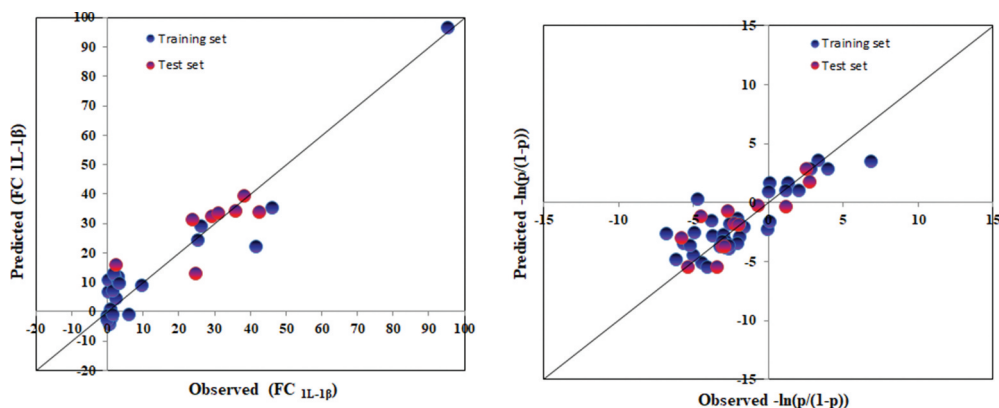


Figure 2. Scatter plots for the developed models [FC_{IL-1β} (M1) and -ln (p/ (1-p)) (M2)].

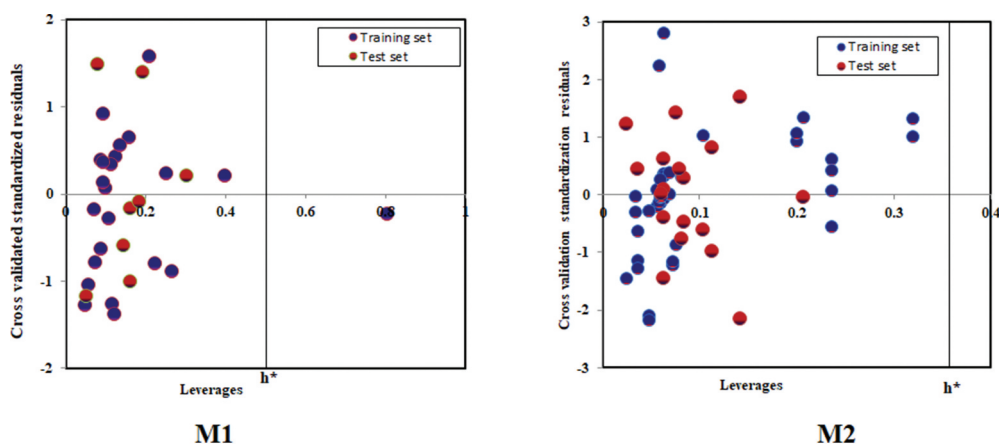


Figure 3. William's plot of the FC_{IL-1β} (M1) and -ln (p/ (1-p)) (M2) models.

Model 1 (Fold change)

$$FC_{IL-1\beta} = -32.2 + 30 \text{ } N_metal -13.4 \Sigma\chi + 14.9 \text{ atomic radius} \quad (1)$$

$$n_{\text{train}} = 24, r^2 = 0.893, Q_{\text{LOO}}^2 = 0.86, \text{PRESS} = 1637.56, F = 55.60, s = 7.80424$$

$$n_{\text{test}} = 8, Q_{F1}^2 = 0.86$$

N_{metal} is first generation periodic table descriptor. This descriptor indicates the number of metals present in the MeOx NP. For example, if the MeOx NP contains one metal in the MeOx NP, then the descriptor value would be one and so on. This descriptor contributes positively towards the Fold Change (FC_{IL-1β}) endpoint which suggests that with an increase in metal number, the FC_{IL-1β} increases. The aforesaid statement can be well understood from the Pr₆O₁₁ MeOx NP where the descriptor N_{metal} value is 6 and the FC_{IL-1β} value is 95. In contrary, ZnO MeOx NP bearing

N_{metal} value 1 represents an FC_{IL-1 β} value of 0.29. Thus, it can be clearly depicted that the N_{metal} descriptor has a significant effect to the toxicity. Certain metals have been reported to seriously affect the immune system potentially causing ill effects to health. The presence of metal, depending on its concentration, route of entry and duration of exposure may lead to immunosuppressant or stimulation of immune cell activity. The metal oxides alter the expression level of adhesion molecules along with chemokine receptors, i.e. the key factor associated with immune response. This alteration affects the T-cell proliferative response to mitogens and chemotaxis [26]. The induction of metals can form metal–protein aggregates through metal co-ordinate bond formation, eventually causing pro-inflammatory function by the assembly of NLRP3 (NACHT, LRR and PYD domain-containing protein) and inflammasome initiating IL-1 β release from the cell [27].

$\Sigma\chi$ stands for the summation of electronegativity of the metals present within the MeOx NP. This descriptor contributes negatively towards the inflammatory effects which suggest that with the increase of the descriptor, the toxicity decreases whereas the opposite takes place on the decrease of the descriptor value. This observation also was in agreement with the result that Co₃O₄ MeOx NP with low electronegativity tends to induce more inflammatory potential than Sm₂O₃ MeOx NPs with higher electronegativity. The MeOx NPs with less electronegativity will have less valence shell electrons, which turn means that its oxidation number is less. From the previous studies, we know that metals with less oxidation number are more harmful than higher oxidation number [28]. The tendency of a metal to gain its higher oxidation state disrupts cellular processing through the generation reactive oxygen species (ROS) which reduces the antioxidants of the cell and increases stress within the cell [29]. The MeOx NPs with less electronegativity can easily enter the cell through phagocytosis. The increased cell stress can further modify the immune system and down regulate the cortisol receptors. This down regulation, in turn, enhances the cytokine-mediated inflammatory processes by reducing respond of cell to anti-inflammatory signal or stimuli [30].

The third period descriptor of the model is atomic radius. This descriptor measures the size of the element, usually, it is the typical measurement of distance from the nucleus to the outermost isolated electron. With an increase of the atomic radius, the formation of metal ion becomes easier since with the increase in the number of shells, there is an increase of shielding effect and the outermost electrons in the valence shell do not experience the counter-positive pull from the nucleus. The positive contribution of the descriptor suggests that MeOx NPs with a high descriptor value will have high potential to cause inflammatory effect as in La₂O₃ MeOx NPs with descriptor value 2.74 have a FC_{IL-1 β} value 38.63, whereas CoO MeOx NPs with less atomic radius value 1.53 tends to cause minimum effect on the inflammatory potential with 0.31. The metal ion releases through corrosion, binds with the serum protein and forms hapten or hapten-like complexes which trigger the immune system. Some metal ions induce an increased level of superoxide anions that indicate neutrophil stimulation [31,32] and may bind to phosphorylated cellular molecules which play a major role in cell signalling. By binding to such messenger pathway, MeOx NPs may interfere with the signalling pathways and the immune function.

Model 2 (Cell viability)

$$\begin{aligned}
-\ln(p/(1-p)) = & -6.85457 + 0.01647 \text{ dose} - 4.51478 \chi + 8.48 \frac{\sum \chi}{nO} \\
& - 0.05770 \text{ atomic radius} + 0.02652 \text{ valence electron potential} \\
& - 0.42535 \sum \chi
\end{aligned}
\tag{2}$$

$$n_{\text{train}} = 41, r^2 = 0.72, Q_{\text{LOO}}^2 = 0.53, \text{PRESS} = 138.486, F = 16.62, s = 1.67892, \\
\Delta r_{\text{m(LOO)}}^2 = 0.0791, \bar{r}_{\text{m(LOO)}}^2 = 0.5056, \text{LV} = 4$$

$$n_{\text{test}} = 13, Q_{\text{F1}}^2 = 0.66, \Delta r_{\text{m(LOO)}}^2 = 0.1866, \bar{r}_{\text{m(LOO)}}^2 = 0.553$$

Dose of any metal oxide plays an important role in causing the inflammatory potential. This descriptor positivity contributes towards the release of cytokine from the cell. Higher the dose, greater is the release of cytokines from the respective cell as observed in La_2O_3 -7 MeOx NP (where dose is 200 $\mu\text{g/mL}$ and $\log_2(\text{FC}_{\text{IL-1}\beta})$ is 5.27), whereas a low dose is not that much influential as observed in Sm_2O_3 -1 MeOx NP (where dose is 3.1 $\mu\text{g/mL}$ and $\log_2(\text{FC}_{\text{IL-1}\beta})$ is 0.07). The effect on T-cell depends on the route of administration and also on dose of the NPs. Increased expression of cells and adverse effects have been observed at higher doses of MeOx NP. This observation is supported by some literature report such as that CeO_2 , TiO_2 NPs and Al_2O_3 activate ERK (1, 2); this activation is dose-dependent and does not activate at lower dose. ZnO NPs or Zn^{2+} salt was also dose-dependent, and phosphorylation of ERK (extracellular signal-regulated kinases) is only detected at higher concentration of Zn^{2+} ions [33]. Chen et al. [34] have also demonstrated that TiO_2 nanoparticles not only dose-dependently increased histamine secretion but also increased cytosolic Ca^{2+} concentration in rat mast cells.

The valence electron potential (Vp ϕ) descriptor provides a quantitative indication of elements reactivity and is based on the charge of the valence electrons and the ionic radius. This descriptor contributes positively towards $-\ln(p/(1-p))$. An increase of the descriptor value (220 eV) as observed in MnO_2 MeOxNP causes an increase in cell death (4.027) compared to the lower descriptor value (40.71 eV) in La_2O_3 MeOxNP where cell death is -1.57 . The charge of the atom describes the detachment of the metal from the metal oxides. It is determined by the equation.

$$-ev = kn/r$$

Here, 'k' expresses the force exerted by the valence electron in volts which is equated to 14.33, 'n' is the valence and 'r' is the ionic radius. Valence electron is furthest from the nucleus and is removed by spending less energy. Therefore, the electron relatively away has a smaller (1/r) value so the energy is larger. Decreased interaction of the nucleus is the principal reason why electrons away from the nucleus are higher in energy than the electrons near to the nucleus [33–36]. The charge and radius ratio of the metal ion indicate the distance between the ion and another charge present, and the size of the electrostatic force created thus defines the index for ionic bond formation.

The $(\Sigma\chi)/nO$ descriptor stands for total metal electronegativity in a specific metal oxide relative to several oxygen. This ratio gives information on the electronegativity of the atom with respective number of oxygens present in the MeOx NP. With an increase of electronegativity, there is a decrease of count cell death. The same observation was also observed in the case of the fold change ($FC_{IL-1\beta}$) endpoint. The positive coefficient of this descriptor suggests that with an increase of the ratio as in CuO (1.9) there is an increase of cell inflammatory potential (6.9), whereas the reduction of $(\Sigma\chi)/nO$ ratio (0.56) decreases cell inflammatory potential (-3.37) as in CeO_2 MeOx NP. This descriptor also considers the presence of oxygen and indicates the effect of oxygen towards the inflammatory potential. The increase of the number of oxygens relative to electronegativity enhances the change of reactive oxygen species which led to stress in the cell. Inflammation and environment oxygen levels are linked with each other. Low level of oxygen count gives a high ratio; therefore, a low level of oxygen or hypoxia can lead to inflammatory condition to the cell.

It is to be noted that electronegativity (χ) has a positive contribution but summation of electronegativity ($\Sigma\chi$) has an opposite contribution. Therefore, it can be inferred that electronegativity describes the tendency to attract electrons. Metals with low χ tend to attract electron towards itself or metals with low χ tend to lose electron to oxygen atom. Oxygen with more electrons traps protons (positively charged ions). Therefore, metals with low χ have high affinity towards protons than with MeOx NPs with higher χ . As per the proton sponge effect [37,38] Figure 4, some nanoparticles enter the liposomal compartment and sequester protons supplied by proton pump (v-ATPase). This retains the proton functions which lead to retention of one Cl^- molecule and two molecules of water. Thus, it causes the rupture of lysosomal content, triggering excess IL-1b production and inflammation. Therefore, this hypothesis is supported by the developed MLR and classification models. Other common descriptors are described previously for other models.

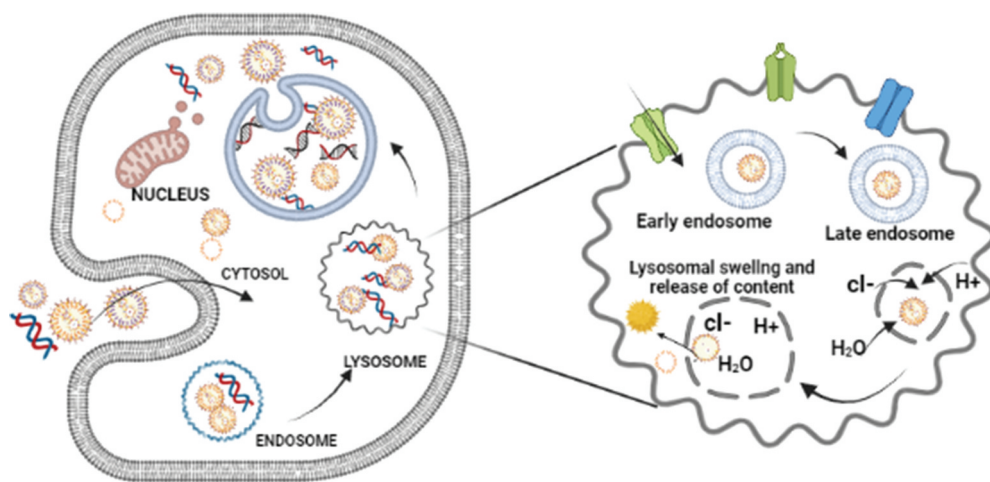


Figure 4. Proton sponge effect.

Evaluation of predictive ability of the developed MLR model for fold change

To further evaluate the performance of the developed MLR model (Equation (1)), new MeOxNPs were selected to verify the performance of the MLR model that lacked inflammatory potential response experimental information. These MeOxNPs were not included in the training set or test set. The 'Prediction Reliability Indicator' [39] tool was used to categorize the quality of predictions for the external set into three groups i) good (with composite score 3), ii) moderate (with composite score 2) and iii) bad (with composite score 1). Three criteria were used: 1) mean absolute error of leave-one-out predictions for 10 most close training compounds for each query molecule [40]; 2) applicability domain in terms of similarity based on the standardization approach [21]; 3) proximity of the predicted value of the query compound to the mean training response [40] in different weightage schemes for making a composite score of predictions depending on the molecular features. Thus, we have collected 37 MeOx NPs as an external set (Table S2 in Supplementary Material S2) to evaluate the predictive power of the developed MLR models and found that out of 37 MeOxNPs, 34 MeOxNPs showed good predictions except one (Tm_2O_3) which had moderate prediction, while CeO_2 was found to be an outlier.

Features responsible for the inflammatory effect to THP-1 cells

To explore the hazardous effects of the MeOxNPs, the THP-1 cells were exploited to examine the toxic impact of MeOxNPs on cell viability and inflammatory responses. This was done since the behaviour of differentiated THP-1 cells is similar to monocyte-derived macrophages [41] and phagocytic uptake is the prime pathway to remove non-soluble particulates [42]. We present a schematic diagram to collectively determine the inflammatory response of MeOxNPs. The regression models pinpointed that N_{metal} , $\Sigma\chi$, atomic radius, valence electron potential, $(\Sigma\chi)/nO$ and dose are the most determining factors for inflammatory potential in THP-1 cells. The main feature derived from the developed models, i.e. presence of metals, has been found to manifest allergic reactions towards orthopaedic implants [43]. Another indication is the increased cytokines in the serum with orthopaedic implants. The immune reactivity of metal ions is considered to be a type IV hypersensitive reaction [44]. The presence of electronegative atoms triggers IL-1 β production and tends to rupture lysosomal content from the cell. IL-1 β is involved in a variety of cellular activities such as cell proliferation, differentiation and cell death. The higher dose of the MeOxNPs also activates IL-1 β productions which activates the NALP-1 caspase-1 inflammasome cascade and participates in the process of cell pyroptosis, characterized by the release of massive cytokine levels [45]. Caspase-1 is generally present in inactive form as procaspase-1 but can get activated by the inflammasome. The activation of inflammasome further stimulates the cleavage of pro-IL-1 β by caspase-1, resulting in mature IL-1 β that is released from the cells. This cascade of reactions further activates the receptor in a paracrine manner producing pro-inflammatory gene products.

Comparison with the previous study

This study focuses on identifying the features governing the inflammatory potential of MeOx NPs to THP-1 cells. The previous study by Huang et al. [16] performed the

Table 2. Comparison of the previous work with present study.

<i>Model</i>	<i>Descriptors</i>	<i>Method</i>	<i>Endpoint</i>	<i>No. of compounds</i>	r^2	Q^2_{loo}	Q^2_{F1}
Previous model Huang et al. [16]	Mainly Quantum descriptors	Regression (PLS)	$FC_{IL-1\beta}$	26	0.89	0.87	0.80
Present work	Periodic table-based descriptors	Regression (MLR- M1) and (PLS-M2)	$FC_{IL-1\beta}$	32	0.893	0.86	0.86
			$-\ln(p/(1-p))$	54	0.72	0.53	0.66

same study with 26 MeOx NPs ($FC_{IL-1\beta}$) applying machine learning techniques with quantum-mechanical computations, experimental characterization and 5 periodic table descriptors to generate QSAR models (partial least squares). The developed model in the previous work showed a result of $r^2 = 0.89$, $Q^2 = 0.87$ and $Q^2_{F1} = 0.80$. On the other hand, the present study used 32 MeOx NPs for modelling using multi-linear regression method for generating a regression model using the in-house QSAR modelling tools. Again, the per cent cell viability data was converted to logit form ($-\ln(p/(1-p))$) to evaluate the toxicity profile of the reported metal oxides. This endpoint was not modelled by the previous authors. The main novelty of the present work is that the whole work was performed using only periodic based-table descriptors that are computationally less demanding, independent of nanoparticle size and do not require computational resources unlike quantum chemical descriptors. The result obtained in the present model with periodic table-based descriptor showed similar model quality ($r^2 = 0.89$, $Q^2_{loo} = 0.86$, $Q^2_{F1} = 0.86$) with the previous work Table 2. The features obtained from the MLR model ($FC_{IL-1\beta}$) were also akin to that of the previous study involving machine learning technique and quantum-mechanical descriptors which are more complicated.

Conclusion

Safe implementation of MeOx NPs requires an intimate knowledge of biological and nanomaterial interactions. Only a few nanotoxicological studies have investigated pro-inflammatory effects. The pro-inflammatory cytokine release of THP-1 cells can be a good indicator for MeOx NPs. The relationship of metal with the inflammation potential has been focused on in the present study. Models for fold change ($FC_{IL-1\beta}$) and cell viability [$-\ln(p/(1-p))$] using QSAR MLR and PLS methods, respectively, were developed that included the dose of MeOx NPs as a factor. The developed models were predicted and validated rigorously following the OECD (Organization for Economic Co-operation and Development) guidelines. The descriptors obtained from the developed models suggested that the number of metal ions or an increased number of metals influences inflammatory potential. Conversely, low electronegativity was found to be more detrimental than high electronegativity. We found that dose, low electronegativity and the presence of metal ions and oxygen are connected to IL-1 β leakage from the cell. The metal ions can trigger inflammasome that leads to a cascade of reactions and activates IL-1 β productions as demonstrated in Figure 5. We suggest that MeOx NPs with lower electronegativity ($\chi_{me} \leq 1.55$) have more inflammatory potential. Summarizing our results, one can conclude that periodic table-based descriptors are efficient for modelling pro-inflammatory potential giving rise to statistically sound QSAR models. Our results are in

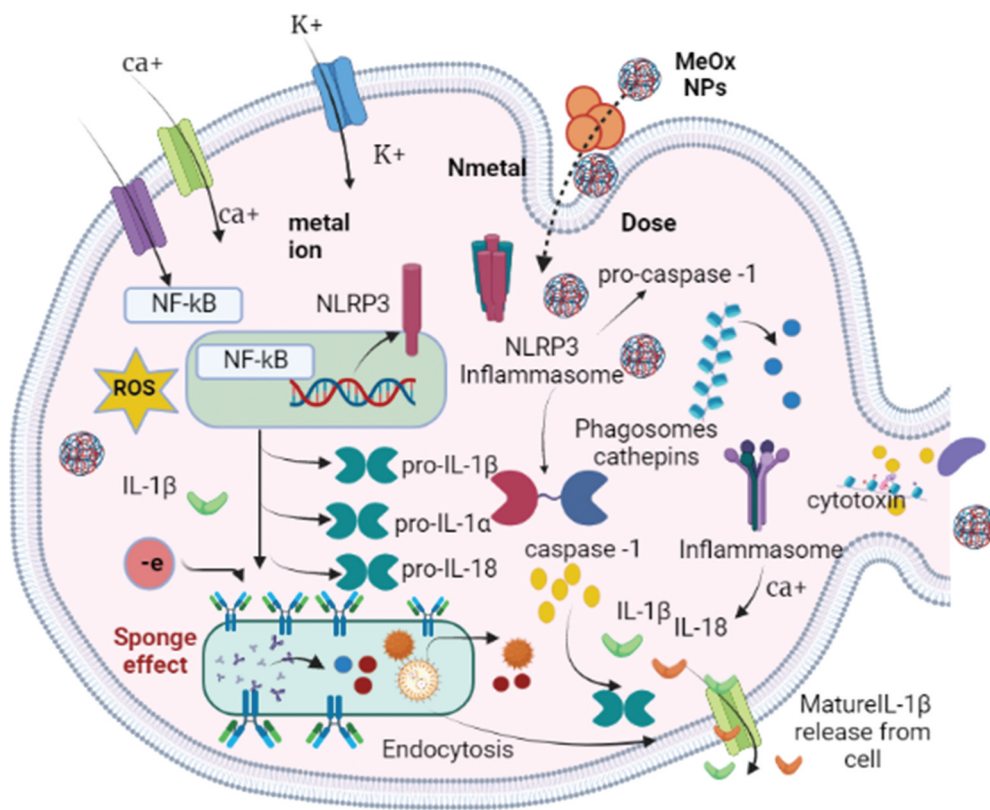


Figure 5. Mechanism of inflammatory potential with sponge effect, release of cytotoxins and lysosomal leakage.

agreement with the literature sources. These comprehensive results and information can be helpful for further research in designing NPs with lower toxicity and deriving mechanistic information on the cause of pro-inflammatory conditions due to MeOx NPs exposure.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded by Indian Council of Medical Research (ICMR), New Delhi, in the form of a Senior Research Fellowship to JR.

References

- [1] A. McWilliams, *Nanotechnology: A Realistic Market Assessment*, BCC Research, Wellesley, MA, 2010.
- [2] G. Forloni, *Responsible nanotechnology development*, J. Nanopart. Res. 14 (2012), pp. 1–7. doi:10.1007/s11051-012-1007-1.

- [3] K.L. Dreher, *Health and environmental impact of nanotechnology: Toxicological assessment of manufactured nanoparticles*, *Toxicol. Sci.* 77 (2004), pp. 3–5. doi:[10.1093/toxsci/kfh041](https://doi.org/10.1093/toxsci/kfh041).
- [4] C. Oksel, C.Y. Ma, J.J. Liu, T. Wilkins, and X.Z. Wang, (Q) SAR modelling of nanomaterial toxicity: A critical review, *Particuology* 21 (2015), pp. 1–9. doi:[10.1016/j.partic.2014.12.001](https://doi.org/10.1016/j.partic.2014.12.001).
- [5] T. Puzyn, N. Jeliaskova, H. Sarimveis, R.L. Robinson, V. Lobaskin, R. Rallo, A.N. Richarz, A. Gajewicz, M.G. Papadopoulos, J. Hastings, and M.T. Cronin, *Perspectives from the nanosafety modelling cluster on the validation criteria for (Q) SAR models used in nanotechnology*, *Food Chem. Toxicol.* 112 (2018), pp. 478–494. doi:[10.1016/j.fct.2017.09.037](https://doi.org/10.1016/j.fct.2017.09.037).
- [6] E. Wyrzykowska, A. Mikolajczyk, I. Lynch, N. Jeliaskova, N. Kochev, H. Sarimveis, P. Doganis, P. Karatzas, A. Afantitis, G. Melagraki, and A. Serra, *Representing and describing nanomaterials in predictive nanoinformatics*, *Nat. Nanotechnol.* 17 (2022), pp. 924–932. doi:[10.1038/s41565-022-01173-6](https://doi.org/10.1038/s41565-022-01173-6).
- [7] G. Oberdörster, E. Oberdörster, and J. Oberdörster, *Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles*, *Environ. Health Perspect.* 113 (2005), pp. 823–839. doi:[10.1289/ehp.7339](https://doi.org/10.1289/ehp.7339).
- [8] P. Rivera Gil, G. Oberdörster, A. Elder, V. Puentes, and W.J. Parak, *Correlating physico-chemical with toxicological properties of nanoparticles: The present and the future*, *ACS Nano* 4 (2010), pp. 5527–5531. doi:[10.1021/nn1025687](https://doi.org/10.1021/nn1025687).
- [9] T. Puzyn, D. Leszczynska, and J. Leszczynski, *Toward the development of “nano-QSARs”: Advances and challenges*, *Small* 5 (2009), pp. 2494–2509. doi:[10.1002/smll.200900179](https://doi.org/10.1002/smll.200900179).
- [10] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T.P. Dasari, A. Michalkova, H.M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski, *Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles*, *Nat. Nanotechnol.* 6 (2011), pp. 175–178. doi:[10.1038/nnano.2011.10](https://doi.org/10.1038/nnano.2011.10).
- [11] R. Liu, H.Y. Zhang, Z.X. Ji, R. Rallo, T. Xia, C.H. Chang, A. Nel, and Y. Cohen, *Development of structure–activity relationship for metal oxide nanoparticles*, *Nanoscale* 5 (2013), pp. 5644–5653. doi:[10.1039/c3nr01533e](https://doi.org/10.1039/c3nr01533e).
- [12] A.A. Toropov, A.P. Toropova, E. Benfenati, G. Gini, T. Puzyn, D. Leszczynska, and J. Leszczynski, *Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli*, *Chemosphere* 89 (2012), pp. 1098–1102. doi:[10.1016/j.chemosphere.2012.05.077](https://doi.org/10.1016/j.chemosphere.2012.05.077).
- [13] W.S. Cho, R. Duffin, C.A. Poland, S.E. Howie, W. MacNee, M. Bradley, I.L. Megson, and K. Donaldson, *Metal oxide nanoparticles induce unique inflammatory footprints in the lung: Important implications for nanoparticle testing*, *Environ. Health Perspect.* 118 (2010), pp. 1699–1706. doi:[10.1289/ehp.1002201](https://doi.org/10.1289/ehp.1002201).
- [14] M. Zhu, Y. Li, J. Shi, W. Feng, G. Nie, and Y. Zhao, *Exosomes as extrapulmonary signaling conveyors for nanoparticle-induced systemic immune activation*, *Small* 8 (2012), pp. 404–412. doi:[10.1002/smll.201101708](https://doi.org/10.1002/smll.201101708).
- [15] R. Li, Z. Ji, C.H. Chang, D.R. Dunphy, X. Cai, H. Meng, H. Zhang, B. Sun, X. Wang, J. Dong, and S. Lin, *Surface interactions with compartmentalized cellular phosphates explain rare earth oxide nanoparticle hazard and provide opportunities for safer design*, *ACS Nano* 8 (2014), pp. 1771–1783. doi:[10.1021/nn406166n](https://doi.org/10.1021/nn406166n).
- [16] Y. Huang, X. Li, S. Xu, H. Zheng, L. Zhang, J. Chen, H. Hong, R. Kusko, and R. Li, *Quantitative structure–activity relationship models for predicting inflammatory potential of metal oxide nanoparticles*, *Environ. Health Perspect.* 128 (2020), pp. 067010. doi:[10.1289/EHP6508](https://doi.org/10.1289/EHP6508).
- [17] P. De, S. Kar, K. Roy, and J. Leszczynski, *Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modeling for exploration of toxicity mechanisms*, *Environ. Sci. Nano* 5 (2018), pp. 2742–2760. doi:[10.1039/C8EN00809D](https://doi.org/10.1039/C8EN00809D).
- [18] P. Mitteroecker and F. Bookstein, *Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics*, *Evol. Biol.* 38 (2011), pp. 100–114. doi:[10.1007/s11692-011-9109-8](https://doi.org/10.1007/s11692-011-9109-8).
- [19] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, and H. Kabir, *Comparative studies on some metrics for external validation of QSPR models*, *J. Chem. Inf. Model.* 52 (2012), pp. 396–408. doi:[10.1021/ci200520g](https://doi.org/10.1021/ci200520g).

- [20] K. Roy, P. Ambure, and S. Kar, *How precise are predictions from our QSAR models for new query compounds*, ACS Omega 3 (2018), pp. 11392–11406. doi:10.1021/acsomega.8b01647.
- [21] K. Roy, S. Kar, and P. Ambure, *On a simple approach for determining applicability domain of QSAR models*, Chemom. Intell. Lab. Syst. 145 (2015), pp. 22–29. doi:10.1016/j.chemolab.2015.04.013.
- [22] P. Grammatica, *Principles of QSAR models validation: Internal and external*, QSAR, ACS Comb. Sci. 26 (2007), pp. 694–701.
- [23] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, and P. Gramatica, *Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs*, Environ. Health Perspect. 111 (2003), pp. 1361–1375. doi:10.1289/ehp.5758.
- [24] S. Kar, A. Gajewicz, T. Puzyn, K. Roy, and J. Leszczynski, *Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach*, Ecotoxicol. Environ. Saf. 107 (2014), pp. 162–169. doi:10.1016/j.ecoenv.2014.05.026.
- [25] J. Roy and K. Roy, *Nano-read-across predictions of toxicity of metal oxide engineered nanoparticles (MeOx ENPs) used in nanopesticides to BEAS-2B and RAW 264.7 cells*, Nanotoxicology 16 (2022), pp. 629–644. doi:10.1080/17435390.2022.2132887.
- [26] T. Lozano-Fernández, L. Ballester-Antxordoki, N. Pérez-Temprano, E. Rojas, D. Sanz, M. Iglesias-Gaspar, S. Moya, Á. González-Fernández, and M. Rey, *Potential impact of metal oxide nanoparticles on the immune system: The role of integrins, L-selectin and the chemokine receptor CXCR4*, Nanotechnol. Biol. Med. 10 (2014), pp. 1301–1310. doi:10.1016/j.nano.2014.03.007.
- [27] M. Pettersson, S. Almlin, G.E. Romanos, and A. Johansson, *Ti ions induce IL-1 β release by activation of the NLRP3 inflammasome in a human macrophage cell line*, Inflammation 45 (2022), pp. 2027–2037. doi:10.1007/s10753-022-01672-7.
- [28] J. Roy and K. Roy, *Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (Me Ox NPs) used in nanomedicines: A mechanistic QSAR approach*, Environ. Sci. Nano 9 (2022), pp. 3456–3470. doi:10.1039/D2EN00303A.
- [29] S. Kang, M. Pinault, L.D. Pfefferle, and M. Elimelech, *Single-walled carbon nanotubes exhibit strong antimicrobial activity*, Langmuir 23 (2007), pp. 8670–8673. doi:10.1021/la701067r.
- [30] S.C. Segerstrom and G.E. Miller, *Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry*, Psychol. Bull. 130 (2004), pp. 601–630. doi:10.1037/0033-2909.130.4.601.
- [31] R. Kumazawa, F. Watari, N. Takashi, Y. Tanimura, M. Uo, and Y. Totsuka, *Effects of Ti ions and particles on neutrophil function and morphology*, Biomaterials 23 (2002), pp. 3757–3764. doi:10.1016/s0142-9612(02)00115-1.
- [32] Y. Mu, T. Kobayashi, M. Sumita, A. Yamamoto, and T. Hanawa, *Metal ion release from titanium with active oxygen species generated by rat macrophages in vitro*, J. Biomed. Mater. Res. 49 (2000), pp. 238–243. doi:10.1002/(SICI)1097-4636(200002)49:2<238::AID-JBM12>3.0.CO;2-J.
- [33] A. Shah and M.A. Dobrovolskaia, *Immunological effects of iron oxide nanoparticles and iron-based complex drug formulations: Therapeutic benefits, toxicity, mechanistic insights, and translational considerations*, Nanotechnol. Biol. Med. 14 (2018), pp. 977–990. doi:10.1016/j.nano.2018.01.014.
- [34] R. Simon-Vazquez, T. Lozano-Fernandez, A. Davila-Grana, and A. Gonzalez-Fernandez, *Metal oxide nanoparticles interact with immune cells and activate different cellular responses*, Int. J. Nanomed. 11 (2016), pp. 4657. doi:10.2147/IJN.S110465.
- [35] E.Y. Chen, M. Garnica, Y.C. Wang, A.J. Mintz, C.S. Chen, and W.C. Chin, *A mixture of anatase and rutile TiO₂ nanoparticles induces histamine secretion in mast cells*, Part. Fibre Toxicol. 9 (2012), pp. 1–10. doi:10.1186/1743-8977-9-2.
- [36] J.D. Walker, M. Enache, and J.C. Dearden, *Quantitative cationic-activity relationships for predicting toxicity of metals*, Environ. Toxicol. Chem. 22 (2003), pp. 1916–1935. doi:10.1897/02-568.
- [37] P. Samir and T.D. Kanneganti, *Hidden aspects of valency in immune system regulation*, Trends Immunol. 40 (2019), pp. 1082–1094. doi:10.1016/j.it.2019.10.005.
- [38] A. Nel, T. Xia, L. Madler, and N. Lin, *Ti is an abundant element in soil and appears in a variety of forms including primary and secondary minerals as well as organically bound and amorphous compounds*, Science 311 (2006), pp. 622–627. doi:10.1126/science.1114397.

- [39] K. Roy, P. Ambure, S. Kar, and P.K. Ojha, *Is it possible to improve the quality of predictions from an "intelligent" use of multiple QSAR/QSPR/QSTR models?* J. Chemom. 32 (2018), pp. 2992. doi:[10.1002/cem.2992](https://doi.org/10.1002/cem.2992).
- [40] K. Roy, P. Ambure, and R.B. Aher, *How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models*, Chemometr. Intell. Lab. Sys. 162 (2017), pp. 44–54. doi:[10.1016/j.chemolab.2017.01.010](https://doi.org/10.1016/j.chemolab.2017.01.010).
- [41] J. Auwerx, *The human leukemia cell line, THP-1: A multi-facetted model for the study of monocyte-macrophage differentiation*, Experientia 47 (1991), pp. 22–31. doi:[10.1007/BF02041244](https://doi.org/10.1007/BF02041244).
- [42] M. Geiser, *Update on macrophage clearance of inhaled micro-and nanoparticles*, J. Aerosol Med. Pulm. 23 (2010), pp. 207–217. doi:[10.1089/jamp.2009.0797](https://doi.org/10.1089/jamp.2009.0797).
- [43] J. Foussereau and P. Laugier, *Allergic eczemas from metallic foreign bodies*, St. John's Hosp. Dermatol. Soc. 52 (1966), pp. 220–225.
- [44] L. Büdinger, M. Hertl, and L. Büdinger, *Immunologic mechanisms in hypersensitivity reactions to metal ions: An overview*, Allergy 55 (2000), pp. 108–115. doi:[10.1034/j.1398-9995.2000.00107.x](https://doi.org/10.1034/j.1398-9995.2000.00107.x).
- [45] P. Bonaventura, A. Lamboux, F. Albarède, and P. Miossec, *Differential effects of TNF- α and IL-1 β on the control of metal metabolism and cadmium-induced cell death in chronic inflammation*, PLoS One 13 (2018), pp. 0196285.



Prediction of cytotoxicity of heavy metals adsorbed on nano-TiO₂ with periodic table descriptors using machine learning approaches

Joyita Roy, Souvik Pore and Kunal Roy^{*§}

Full Research Paper

Open Access

Address:

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India

Email:

Kunal Roy^{*} - kunalroy_in@yahoo.com

^{*} Corresponding author

[§] Phone: +91 98315 94140; Fax: +91-33-2837-1078; kunal.roy@jadavpuruniversity.in

Keywords:

heavy metals; HK-2 cell; ML algorithm; periodic table descriptors; QSAR

Beilstein J. Nanotechnol. **2023**, *14*, 939–950.

<https://doi.org/10.3762/bjnano.14.77>

Received: 23 June 2023

Accepted: 30 August 2023

Published: 12 September 2023

This article is part of the thematic issue "Nanoinformatics: spanning scales, systems and solutions".

Associate Editor: A. Salvati



© 2023 Roy et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

Nanoparticles with their unique features have attracted researchers over the past decades. Heavy metals, upon release and emission, may interact with different environmental components, which may lead to co-exposure to living organisms. Nanoscale titanium dioxide (nano-TiO₂) can adsorb heavy metals. The current idea is that nanoparticles (NPs) may act as carriers and facilitate the entry of heavy metals into organisms. Thus, the present study reports nanoscale quantitative structure–activity relationship (nano-QSAR) models, which are based on an ensemble learning approach, for predicting the cytotoxicity of heavy metals adsorbed on nano-TiO₂ to human renal cortex proximal tubule epithelial (HK-2) cells. The ensemble learning approach implements gradient boosting and bagging algorithms; that is, random forest, AdaBoost, Gradient Boost, and Extreme Gradient Boost were constructed and utilized to establish statistically significant relationships between the structural properties of NPs and the cause of cytotoxicity. To demonstrate the predictive ability of the developed nano-QSAR models, simple periodic table descriptors requiring low computational resources were utilized. The nano-QSAR models generated good R^2 values (0.99–0.89), Q^2 values (0.64–0.77), and Q^2F_1 values (0.99–0.71). Thus, the present work manifests that ML in conjunction with periodic table descriptors can be used to explore the features and predict unknown compounds with similar properties.

Introduction

Nanoparticles (NPs) have gained much attention due to their widespread applications in different areas, and they are continually designed to yield certain desired properties [1]. With the

uninterrupted development of new NPs, engineered nanoparticles in the form of metal oxide nanoparticles are becoming a new area of research. Metal oxides have been used in different

industries, and the market is developing rapidly [2]. According to a recent analysis, approximately 1300 consumer products containing NPs were marketed in 2012. As a common metal oxide nanoparticle material, nanoscale titanium dioxide (nano-TiO₂) has been evaluated for diverse applications. TiO₂ has been shown to be a promising material for practical applications because it is highly photoreactive, inexpensive, non-toxic, chemically and biologically inert, and photostable. Also, nano-TiO₂ exhibits high specific surface area and anti-corrosion and photocatalytic properties [3]. It absorbs UV radiation and shows self-cleaning ability. Nanoparticles have a susceptibility to adsorb other substances to form a mixture leading to a shift of toxicity to living organisms [4]. Hence, many studies have reported cytotoxic characteristics of TiO₂ [5,6].

Some NPs are fatal to living cells, and their cytotoxicity may inhibit cell growth cycles, leading to death of organisms. Considering this fact, the cytotoxicity of TiO₂ in combination with other pollutants has been evaluated. TiO₂ is the most commonly manufactured nanoparticle material. It is assumed that because of the considerably high exposure TiO₂ NPs may enter the food chain. Because of current industrialization processes, organisms are also exposed to heavy metal pollutants [7]. Emitted NPs may interact with the pollutants, and this may subsequently lead to bioaccumulation. The contamination of water and soil with heavy metals has increased with anthropogenic and industrial activities [8,9]. TiO₂ NPs commonly co-exist with different heavy metals as they are released from wastewater treatment facilities to freshwater bodies, affecting the mode of action and the fate of the contamination. Studies have reported the ability of TiO₂ NPs to adsorb heavy metals and to increase their transport rate into hosts, increasing their concentration in the cell. Hu et al. [10] investigated the joint effect of TiO₂ NPs and humic acid (HA) on Cd²⁺ bioaccumulation in zebrafish. In another study, Yang et al. [11] showed that TiO₂ NPs increased the accumulation of Cd²⁺ in the ciliate *Tetrahymen thermophila*. Further, Tan et al. [12] showed increased uptake and retention of Cd²⁺ and Zn²⁺ adsorbed on TiO₂ NPs in *Daphnia magna*. Heavy metal contamination affects plant growth and indirectly affects human health via the food chain. Heavy metals have become an important factor limiting crop yields and, thus, threatening food security. Therefore, to improve crop yields, heavy metals need to be removed.

The toxicity of single-substance NPs has been tested extensively; however, the combination of single-substance NPs with other NPs or metals may cause co-exposure effects on living organisms. The extensive use of heavy metals in areas such as medicine and agriculture increased the negative impact of heavy metals on environment and living organisms, raising the need for risk assessment. Unlike other pollutants, heavy metals do

not decompose, leading to bioaccumulation and biological hazards [13]. Heavy metals enter the human body through the consumption of fish and plants [14]. To date, heavy metals are removed through various methods. Among all methods available for removing heavy metals and toxic pollutants from waters, adsorption is the most widely used. Therefore, the joint organismal toxicity should be assessed.

Recently, nanoscale quantitative structure–activity relationship (nano-QSAR) models have been successfully applied to investigate the toxicity of NPs. QSAR models for predicting the biological activity of 48 fullerene derivatives [15], 51 manufactured nanoparticles with varying core metals, coatings, and surface attachments [16], and 80 surface-modified multiwall carbon nanotubes have been reported. Another approach, namely nano-read-across (nano-RA) [17], has been used to determine the cytotoxicity of unknown nanomaterials based on structure similarities with known substances. Materials with similar structures are likely to produce similar toxicity through comparable mechanisms. The development of machine learning (ML) approaches, such as artificial neural networks (ANNs), decision trees, logistic regression (LR), support vector machines (SVM), Naïve Bayes (NB), random forest (RF), and *k*-nearest neighbor (*k*-NN), can be used to construct models that simulate complex relationships [18] and make predictions based on training data.

Using ensemble learning (EL) [19] methods, one can determine the relationship between the response and the predictor as well as solve regression problems. Additionally, such methods overcome problems with weak predictors and can be used to reduce the overfitting of the training data by averaging and incorporating multiple models. Ensemble learning is established with multiple algorithms and is divided into bagging and boosting algorithms. The boosting algorithm is an iterative algorithm that uses a weak model to build a strong model. Both bagging and boosting improve the prediction accuracy of weaker learners. A boosting algorithm combines many models linearly, with each new model depending on the previous one. In the bagging algorithm, replica data sets are generated that minimize prediction variance in machine learning. An iterative algorithm performs a series of repeated steps to gradually improve the model's performance or to optimize a specific parameter. The algorithms continue to update the model's parameters based on the training data until a certain stopping criterion is met, such as reaching an optimal solution, or a predefined number of steps are completed. This process is performed during the training of the model, where the model learns from the data by adjusting its parameters to minimize a specific cost or error function. These algorithms play a crucial role in training machine learning models and are fundamental to many optimization and learning techniques. Fine-tuning the model parameters through iterations

helps to improve the model's performance and makes it more suitable for making accurate predictions for new, unseen data.

The boosting algorithm is an ensemble method that works sequentially by adding predictors to an ensemble, each one correcting its predecessors. In the boosting algorithm, at first, an initial model is developed with the dataset and then the algorithm tries to adjust the model parameters and again develops a model that tries to correct or minimize errors present in the previous model. This process is repeated until a satisfactory model is obtained or the error function is significantly optimized. Through this process, we get a strong learner or model from several weak learners or models by sequentially minimizing the error present in the predecessor models. Here, the weak model represents the models that are developed at an initial stage and contain a significant amount of error. The strong model is indicated by the final model, which contains a significantly low level of error and is able to predict new unknown data more accurately. Bagging (or bootstrap aggregating) is an ensemble method that generates a number of bootstrap datasets by a method called random sampling with replacement, and each dataset is used to train the models separately. The final prediction is obtained by averaging the outcome of each model (for regression models) or by majority voting (for classification models).

The objective of the present study was to construct EL-based regression models (RF, Gradient Boost, Extreme Gradient Boost, and AdaBoost) with periodic table descriptors for predicting the cytotoxicity, in terms of cell viability, of eight heavy metals adsorbed on nano-TiO₂. Also, the best algorithm showing the most contributing features responsible for the toxicity to HK-2 (human kidney 2) cell has been determined. To the best knowledge of the authors, this is the first work on ML models using periodic table descriptors to successfully demonstrate the high potential of the proposed modeling approaches.

Methods and Materials

Dataset

The dataset was collected from previously published literature [20]. A mixture of nano-TiO₂ powders was added to HK-2 cells in Hyclone DMEM medium supplemented with 10% fetal bovine serum (FBS) and 100 mg penicillin/streptomycin and maintained at 37 °C in the presence of 5% carbon dioxide. Nine concentrations of heavy metal salts were added to a constant amount of nano-TiO₂ (25 µmol/L). The details of heavy metal concentrations are given in Table 1.

HK-2 cells were utilized to determine the toxicity in this study using cell viability as the endpoint. HK-2 cells are a sensitive model for examining renal cytotoxicity. They grow in monolayers and are suitable for studying the proximal tubular toxicity of a variety of compounds [21]. The main advantage of HK-2 cells is that they retain the basic morphological and functional properties of proximal tubular epithelial cells [22]. Cell viability was measured by using Equation 1:

$$S = \frac{A_{\text{exp}} - A_{\text{blank}}}{A_{\text{control}} - A_{\text{blank}}} \quad (1)$$

Here, *S* stands for cell survival rate, *A*_{exp} is the absorbance value of the experimental group, *A*_{control} is the absorbance value of the control group, and *A*_{blank} is the absorbance value of the blank control group.

Descriptor calculation

Based on the characteristics of metals, we used easily calculable periodic table descriptors. Simple molecular information was generated time-effectively and cost-effectively. The previously used descriptors by Kar et al. [23] are the metal electronegativity (*χ*), the sum of metal electronegativity for an individual metal oxide (*Σχ*), the sum of metal electronegativity for

Table 1: Different concentrations of heavy metal salt samples in µmol/L.

Sample	CdCl ₂	ZnCl ₂	CuSO ₄	NiCl ₂	Pb (NO ₃) ₂	MnCl ₂	SbCl ₃	CoCl ₂
1	10	60	30	100	100	100	5	10
2	20	90	60	200	200	200	10	20
3	30	120	90	300	300	300	15	30
4	40	150	120	400	400	400	20	40
5	50	180	150	500	500	500	25	50
6	60	210	180	600	600	600	30	60
7	70	240	210	700	700	700	35	70
8	80	270	240	800	800	800	40	80
9	90	300	270	900	900	900	45	90

an individual metal oxide divided by the number of oxygen atoms present in that metal oxide ($\sum \chi/\text{NO}$), the number of metal atoms (N_{Metal}), the number of oxygen atoms (N_{Oxygen}), the charge of the metal cation in a given oxide (χ_{ox}), and the molecular weight (MW). These descriptors are termed “first-generation periodic table descriptors”. The newly introduced sixteen descriptors are denoted as “second-generation periodic table descriptors” [24]. The computed descriptors for all metals are reported in the Excel file in Supporting Information File 1. In addition to being computationally less demanding, periodic table descriptors are size-independent.

Splitting of data set and hyperparameter tuning

The dataset was split into training and test sets before building the model. The training set was mainly used to fit the model, and the test set was used to measure the generalization ability of the developed model. Theoretically speaking, the dataset was divided based on a sorted response-based approach using the in-house dataset division tool (<https://dtclab.webs.com/software-tools>). In this study, the size ratio was set at 3:1 (training set/test set) for dataset division.

In almost any ML algorithm, different models are trained for a dataset and the best-performing model is selected. However, there may be room for improvement, and hyperparameter tuning can significantly improve the model. Here, the optimal values of the hyperparameters of the models were obtained with the GridSearchCV algorithm using the hyperparameter optimizer tool (<https://sites.google.com/jadavpuruniversity.in/dtclab-software/home/machine-learning-model-development-guis?pli=1>). GridSearchCV tests all combinations of values in the dictionary and evaluates the model using the cross-validation method for each combination. Therefore, we choose the hyperparameter combination with the best average MAE results from the validation sets.

Feature selection with random forest

The goal of feature selection techniques is to find the best set of features that allows one to build optimized models. Feature selection using RF is an embedded method. Embedded methods combine the benefits of filter and wrapper techniques. These methods encompass the interaction of features while maintaining reasonable computational cost. In embedded methods, each iteration of the model training process is taken care of, and a few features that contribute the most to the training process are carefully extracted. More precisely, it is measured how much impurity is reduced on averaging (weighted average) through each tree nodes with the selected features. Here, each node is equivalent to the number of training samples associated with it. Through the RF algorithm, we have selected the most

contributing eight periodic table descriptors, namely “conc”, “ $\sum \chi$ ”, “atomic radius”, “IP_ActivM”, “Mol_Wt”, “ χ of metal”, “D3_HeteroNonMetal”, and the total number of atoms in a molecule, from a pool of 43 periodic table descriptors by using the features with the highest Gini importance [25]. The selected first eight descriptors (most contributing features) were further used for modeling using RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost algorithms.

Model development

This section introduces four classification models; all of them are ensemble learning models. ML is a subset of artificial intelligence where the machine learns from data and improves performance from past experiences and makes a prediction based on it [26]. In this study, along with RF, Gradient Boost, Extreme Gradient Boost, and AdaBoost were also performed. In the supervised learning approach, a model is trained on labeled datasets. Regression analysis algorithms are trained and learned from both input features and output labels. Regression analysis seeks a mapping function from the input features for a continuous output function. In this study, there is no intention to categorize the dataset, instead it is to be predicted quantitatively. Hence, the supervised regression method is selected to map the function of heavy metals and predict the cytotoxicity of these metals on HK-2 cells with periodic table descriptors. For the model development, after dataset division and feature selection, different ML algorithms are performed. The overall workflow is illustrated in Figure 1.

Random forest (RF)

In ensemble learning, RF is often used for its flexibility. Whether it is regression or classification, RF is a versatile learning method that can handle both. It works by building several decision trees in the training phase and generates average forecasts of various decision trees involved. In other words, it combines the results of different decision trees to make the best possible decision. Though the goal variable in classification-based issues is categorical, numerical values are present in regression. One advantage of RF is its capacity to analyze large datasets with great efficiency [27]. It can be regarded as a dimensionality reduction method since it analyzes large input data and finds all important variables. While handling RF datasets, the model emphasizes the importance of parameters, which is a highly helpful aspect [28].

Adaptive boosting (AdaBoost)

AdaBoost is one of the best boosting algorithms. It uses an ensemble learning method. This approach of machine learning is based on the idea of creating accurate prediction rules by combining many relatively weaker and inaccurate rules and assisting in alleviating overfitting issues. It is possible to make a

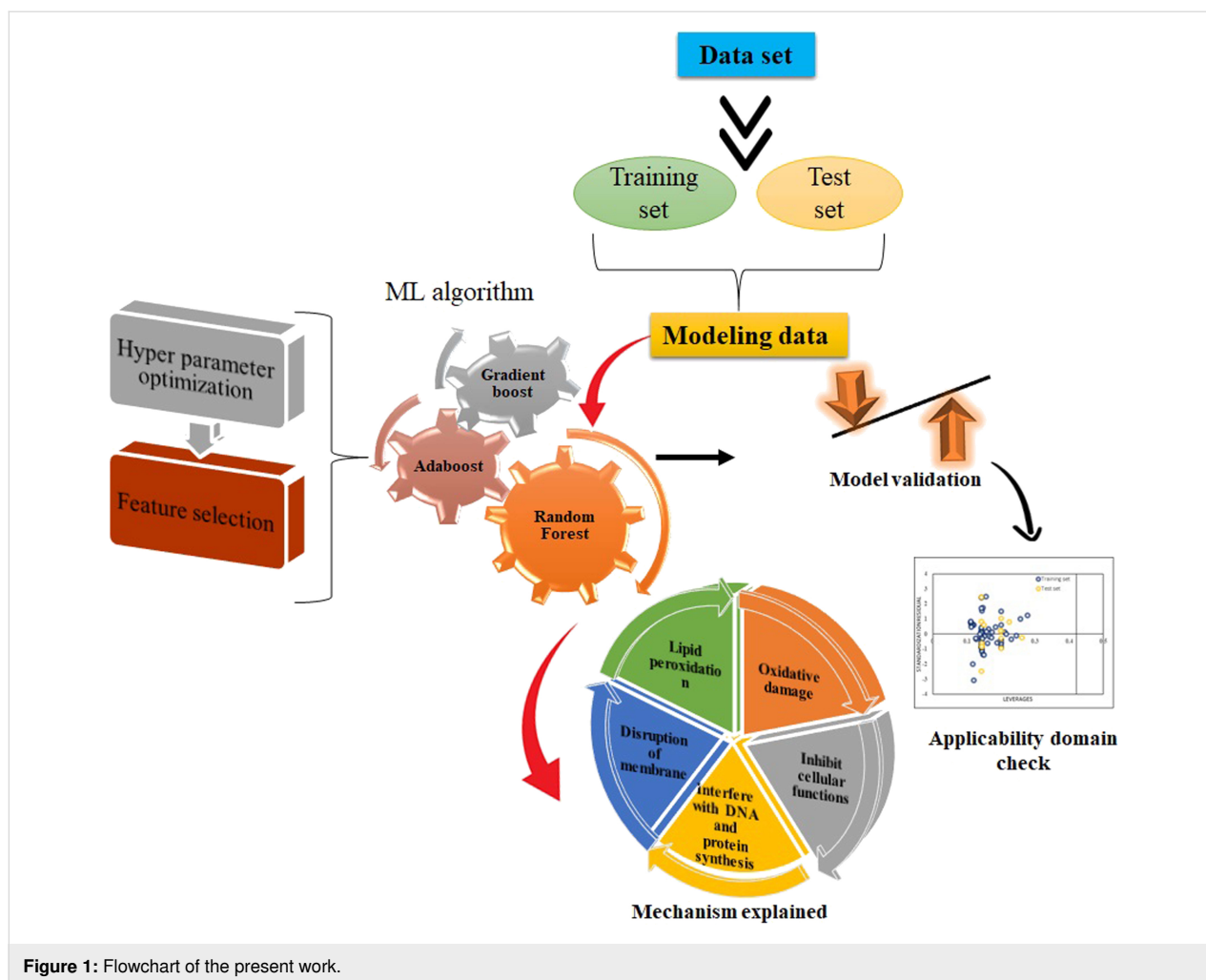


Figure 1: Flowchart of the present work.

smarter learner by altering the training data intelligently and constructing many submodels. It includes an unlimited amount of decision trees for input data throughout the training stage. During the creation of the first decision tree, incorrect data are highlighted inside the primary model.

The identical data serve as input for a separate model. This procedure is repeated until a specific number of base learners is generated. It uses a weighted average relying on the subsets to determine whether it should be included in the finalized model. In reality, some data may include linear predictions, and others may not. Therefore, utilizing the ensemble AdaBoost allows us to capture the nonlinear predictions and make a precise prediction for such data [29].

Gradient Boost (GB)

In 2002, Friedman [30] suggested an ensemble learning algorithm for both regression and classification. The GB method is associated with each repetition of the randomly chosen training data set with the fundamental model. Overfitting is inhibited by

randomly subsampling the training set data; by doing so, the execution time and model accuracy are also improved. Since every repetition of the model must include small data (as a training set) the regression becomes quicker. The GB approach also requires modification or changes in a few parameters. That is, n -trees should not be too small, and the shrinkage aspect, also recognized as the learning rate, must not be kept too high [31].

Extreme Gradient Boost (XGBoost)

In a similar manner as described in [32], another ensemble ML algorithm, XGBoost of tree boosting, uses a gradient-boosting framework for efficient and scalable implementation performance. Ensemble learning uses multiple predictions that are multiple models for gradient enhancement and yields good adaptability to outliers and continuous variables. It is an efficient tool for dealing both regression and classification problems. The basic idea is to build “ N ” regression trees to train each subsequent tree using the residual from the previous tree. Models are built recursively until there is no improvement in

the results obtained. The new models predict the residuals of the prior model and then collectively provide the final predictions [33]. The gradient descent algorithm is used to minimize the loss while adding new models. Then, these individual predictors or classifications are combined to give more strong and more precise predictions. The workflow of the ML algorithm is represented in Figure 2. Tuning can be done using the grid search method.

SHAP analysis

The feature importance in the model was determined using the Shapley Additive exPlanation (SHAP) method, using SHAP version 0.41.0. The SHAP framework takes into account the calculation of Shapley values. These values are calculated from the average marginal contribution of each feature from all conceivable coalitions. First, the dataset is incorporated into the

model, then the SHAP framework assigns a Shapley value to each feature that contributes to the corresponding output of the model. Therefore, SHAP helps to select the features based on a ranking algorithm [34]. We have selected the features having the highest Shapley values for the training set since the standard method tends to overestimate the continuous variables.

Model validation

A reliable model should pass the threshold values for different internal and external validation metrics. Internal validation generates the generalization ability and robustness of the model. In contrast, external prediction is used to validate the model. The most common metrics to measure internal quality are the coefficient of determination (R^2 and Q^2_{LOO}). Besides these, we have also calculated the root mean square error (RMSE) of the training set. The mean absolute error ($\text{MAE}_{(\text{test})}$), the root mean

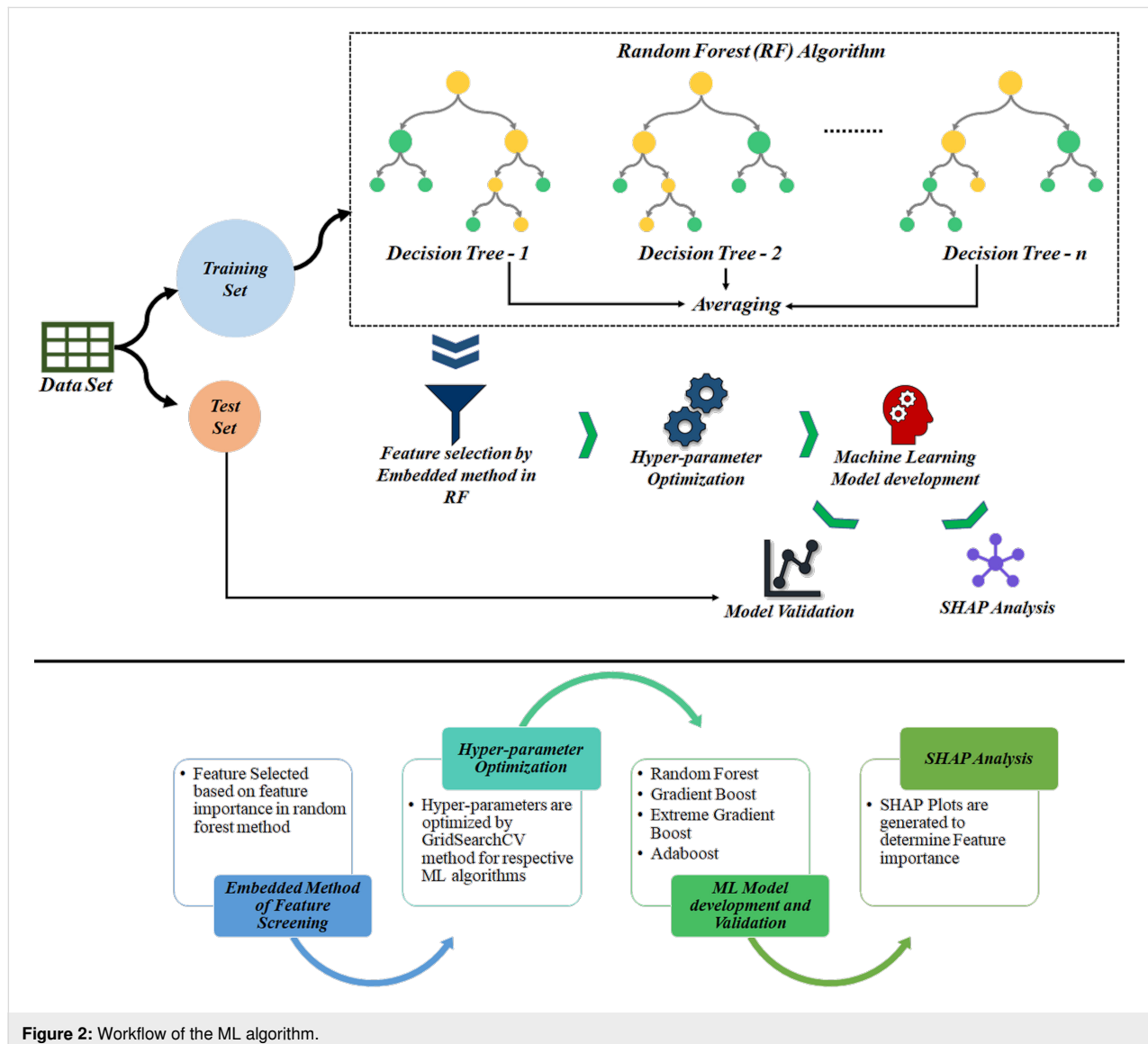


Figure 2: Workflow of the ML algorithm.

square error of prediction (RMSEP), Q^2_{F1} , and Q^2_{F2} were used for external validation or for the test set. The evaluation criteria included are as follows: Q^2_{LOO} and Q^2_{F1} greater than 0.5, RMSE between 0.2 and 0.5, and the closer the value of MAE is to 0, the better [35].

Applicability domain (AD) analysis

After building the model, the applicability domain (AD) must be considered. AD represents the domain that can be effectively predicted by the model that is based on the training set data. The samples within the domain of applicability can only explain the reliability of the predicted values. A William's plot was used to determine the AD of the present work. The leverages were calculated using the in-house Hi_Calculator-v1 Software (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home?pli=1>). The distance between the X value of the i -th observation and all X values is represented by the leverage value. It generally considers $3k'/N$ as the critical value or the standard value (h^*). Here, k' represents the number of descriptors plus 1, and N represents the number of compounds in the training set. If the leverage value is higher than h^* , the corresponding compound is outside the AD.

Results and Discussion

In this research, we have used four ML models, namely RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost to forecast the toxicity of heavy metals adsorbed on nano-TiO₂ to HK-2 cells using periodic table descriptors (Table 2). The ML

models were built using the features selected by the RF algorithm. Model specification and configuration were carried out by optimization of the hyperparameters. The AD was also determined, and all compounds were found to be below the threshold of $h^* = 0.42$, as shown in the Williams plot in Figure 3. The AD is the chemical space formed based on the descriptors of the training set compounds. The compounds in the chemical space are considered reliable for predictions, while those beyond the AD would not guarantee accurate predictions. The AD plays an important role in determining the uncertainty of the predictions

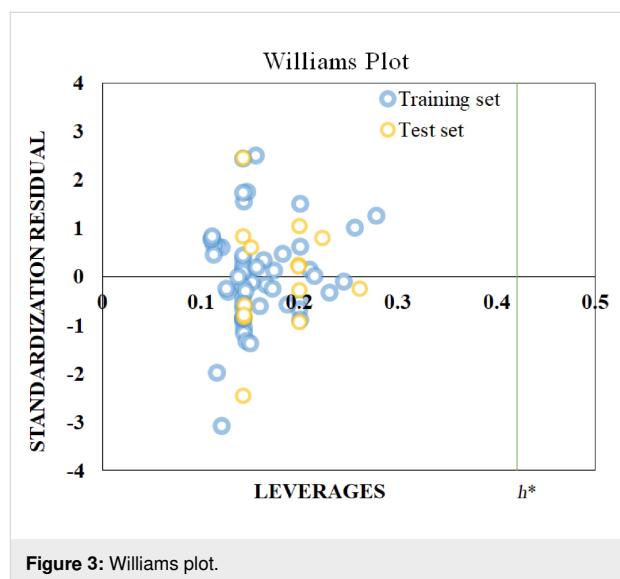


Figure 3: Williams plot.

Table 2: Statistical parameters and selected features from the developed ML models.

Method	R^2	$Q^2_{(LOO)}$	MAE_{train}	$RMSE_C$	Q^2_{F1}	Q^2_{F2}	MAE_{test}	$RMSEP$	Optimized hyperparameters
Random Forest	0.96	0.72	0.13	0.2	0.94	0.94	0.14	0.19	'max_depth': none, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 80, max_features: 1.0, bootstrap: true, random_state: 0
Gradient Boost	0.99	0.77	0.06	0.08	0.82	0.81	0.2	0.34	'max_depth': 6, 'min_samples_leaf': 4, 'min_samples_split': 4, 'n_estimators': 130, max_features: none, random_state: 0
Extreme Gradient Boost	0.94	0.46	0.16	0.26	0.83	0.83	0.25	0.32	'booster': 'gbtree', 'colsample_bytree': 0.3, 'max_depth': none, 'min_child_weight': 1, 'n_estimators': 70, 'subsample': 0.5
AdaBoost	0.88	0.64	0.31	0.37	0.72	0.71	0.33	0.41	'loss': 'linear', 'n_estimators': 170, random_state: none

of specific molecules based on how close they are to the training set compounds used to develop the model. AD is a valuable tool for the characterization of interpolation spaces based on the modeled descriptors and response functions.

The applicability domain developed here is based on the features of some specific heavy metal salts, that is, CdCl_2 , ZnCl_2 , MnCl_2 , CoCl_2 , CuSO_4 , NiCl_2 , $\text{Pb}(\text{NO}_3)_2$, and SbCl_3 . The developed model should be applicable to other closely related heavy metal salts.

Diagnosis based on SHAP value

The goal of SHAP is to explain the prediction of an instance by computing each feature of the prediction. First, the SHAP value is used to calculate the magnitude of the contribution of each feature and then ranked to obtain the importance ranking of features. Features with large absolute Shapley values are important. Here, we have used the kernel method to calculate the SHAP values [36]. The SHAP analysis and hyperparameter tuning (max_depth: “none” min_samples_leaf, min_samples_split, n_estimators) revealed that concentration, followed by atomic radius and IP_ActivM, ranked highest among the eight features (conc, $\Sigma\chi$, atomic radius, IP_ActivM, Mol_Wt, χ of metal, D3_HeteroNonMetal, and atoms in the

molecule) in the RF model. The hyperparameter setting n_estimators was kept at a value of 80 for RF, while it was 130, 70, and 170 for Gradient Boost, Extreme Gradient Boost, and AdaBoost respectively. The relative importance of each descriptor for all ML algorithms can be understood using the SHAP analysis (Figure 4). The SHAP methodology identifies the features contributing most to the model prediction. We can find that the conc (concentration of the heavy metal) descriptor contributes the most to the EL algorithms. The Shapley values reflect the average marginal contribution of a feature value across all possible feature coalitions, both in terms of magnitude and direction.

Results of model validation for all ML methods

In order to determine if heavy metals and TiO_2 nanoparticles had any cytotoxic effects, the selected eight important periodic table-based features were used. The final models developed with RF, AdaBoost, Gradient Boost, and Extreme Gradient Boost were evaluated using $\text{MAE}_{\text{train}}$, $\text{RMSE}_{\text{train}}$, R^2 , and Q^2 for the training set and MAE_{test} , MSE , $\text{RMSE}_{\text{test}}$, Q^2F_1 , Q^2F_2 metrics for the test set, and the results are shown in Table 1. According to the results, the MAE_{test} (0.14) was found to be the least for the test set in the RF method, followed by AdaBoost,

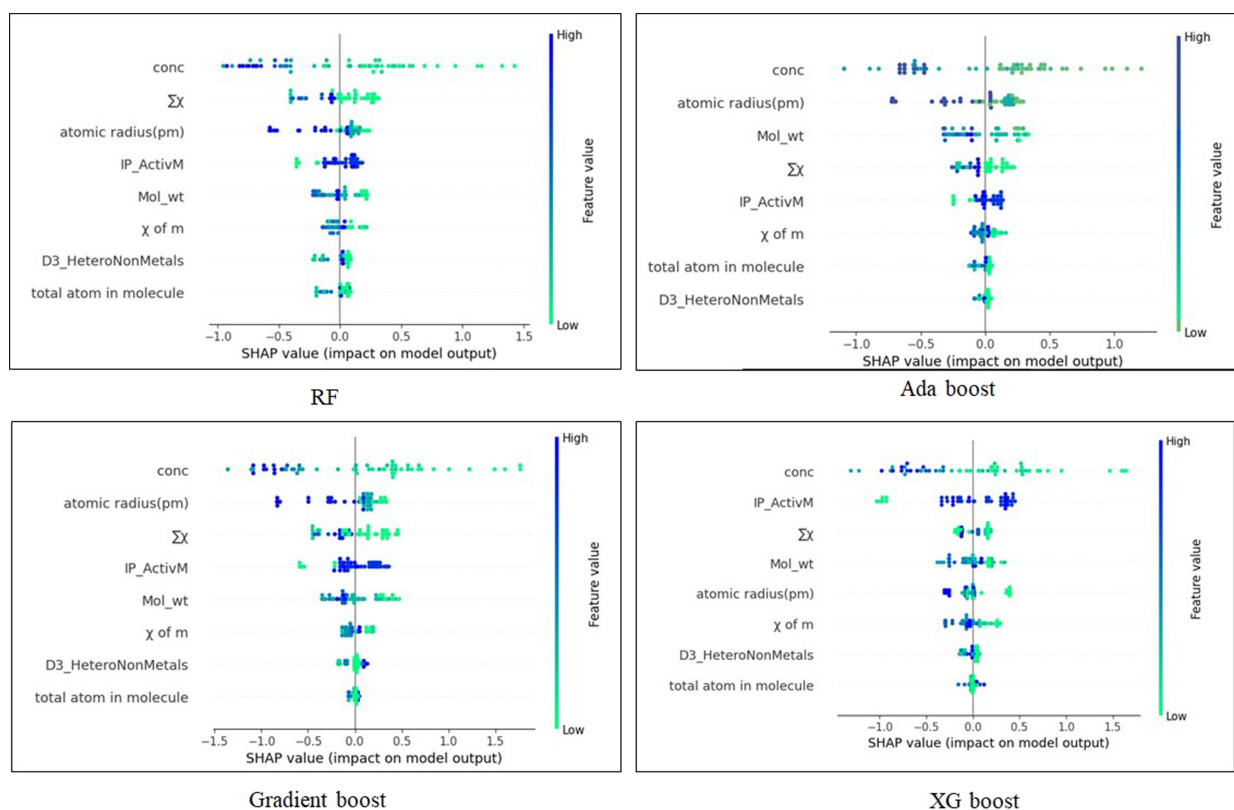


Figure 4: SHAP summary plot.

Gradient Boost, and XGBoost. The XGBoost gives the highest R^2 (0.99) for the training set, while AdaBoost gives the lowest R^2 (0.88) with the highest MAE_{test} (0.33). Cross-validation (CV) statistics were obtained based through 20 times fivefold repetitive CV along with 1000 times shuffle split CV (mean $\pm$ SEM) method. This is done to protect the model from overfitting when the data is limited. The results of the CV indicate clearly that the models do not memorize the correspondence between the descriptors since the outcome of R^2 is highest and the MAE value is lowest for the RF model after the repetitive CV method. This suggests the superiority of the RF model to other models. Figure 5 presents the cross-validation statistics based on 20 times fivefold repetitive CV and 1000 times shuffle split CV on R^2 and MAE for the developed ML model.

General mechanism of toxicity

In the process of screening all descriptors from different ML methods, some common descriptors for heavy metals were discovered that are clear indicators of their importance regarding toxicity to HK-2 cells. We found that the concentration of the heavy metal (conc), the atomic radius of the metal, the electronegativity, and the molecular weight of the heavy

metal influence the survival rate of the HK-2 cells. It was observed that conc, mol wt, atomic radius, and the total number of atoms in the molecule were of high importance in all the models. The increase of conc, mol_wt, and total atoms in a molecule is believed to increase toxicity. The toxicity of the heavy metals is also time- and dose-dependent. Among many other factors, the valence state plays an important role in toxicokinetics and toxicodynamics. Many studies have shown that an increased concentration of heavy metals is correlated with the severity of hepatotoxicity and nephrotoxicity [37]. Lead causes toxicity through an ionic mechanism followed by the generation of reactive oxygen species (ROS). Another, biomarker for ROS is lipid peroxidation [38] as free radicals cause lipid peroxidation inside the cell membrane. The catalytic properties of the metals are also responsible for an increased toxicity of manufactured nanoparticles [39] (Figure 6). Electronegativity and atomic radius influence the catalytic properties of the metal. Metal cations also catalyze the lipid peroxidation process [40] through enhancement of endocytosis and the intrinsic properties of the heavy metal. The toxicity is associated with internalization and bioaccumulation in the HK-2 cells. The increase in the concentration of heavy metals and their adsorption to nano-

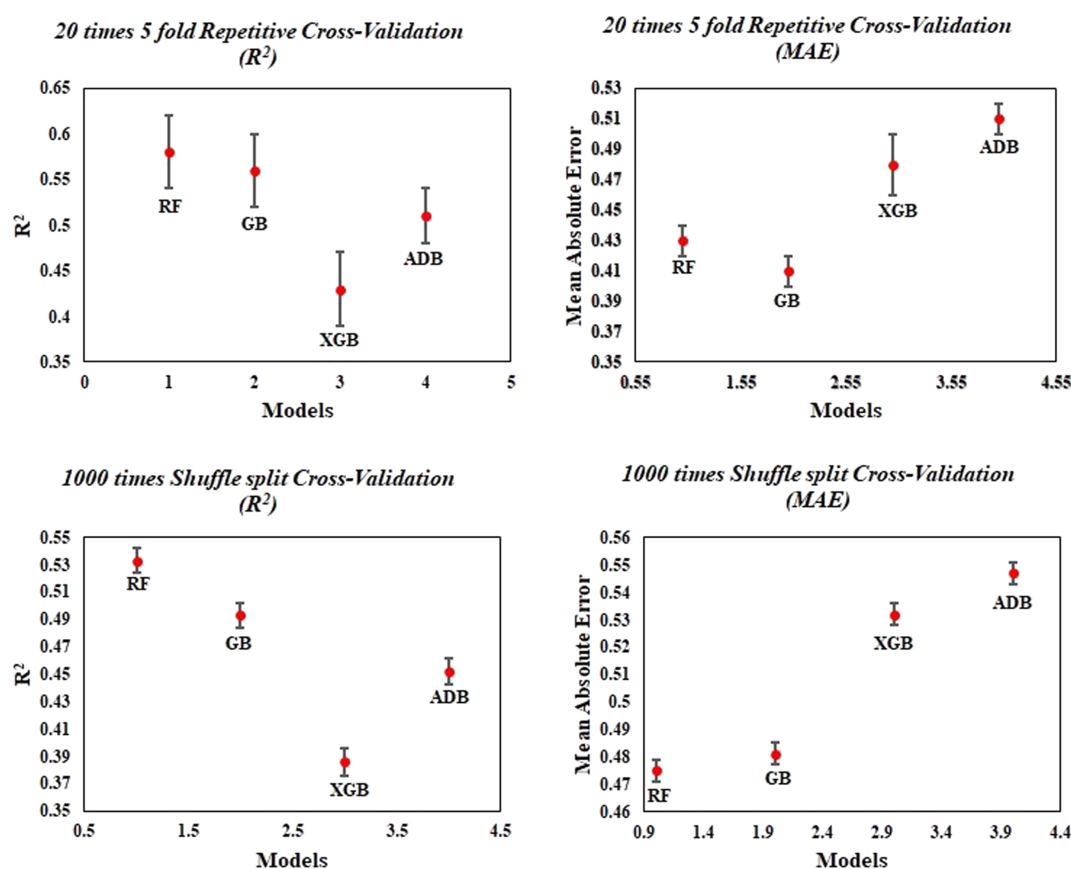
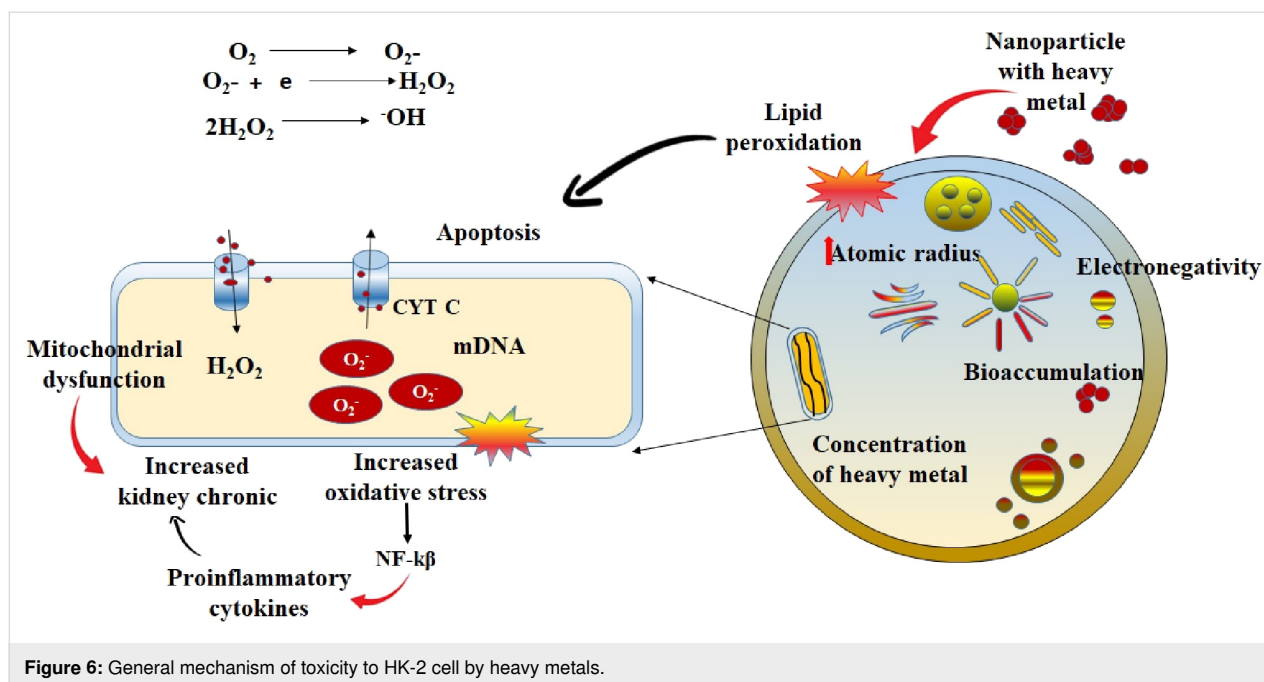


Figure 5: Cross-validation statistics based on 20 times fivefold repetitive CV and 1000 times shuffle split CV method (mean $\pm$ SEM).



TiO₂ induces toxicity by increasing the generation of ROS in the HK-2 cells.

Comparison with the previous work

The present work describes the development of a model for heavy metals with different concentrations through simple periodic table descriptors using various ML methods. The results obtained from the ML method suggest that the models have better predictivity than the models developed previously by Sang et al. [20] as shown in Table 3. Sang et al. [20] applied the random forest algorithm and the AdaBoost algorithm for QSAR modeling using quantum mechanical descriptors. In contrast, the present study involved the random forest algorithm, the AdaBoost algorithm along with Gradient Boost and XGBoost algorithms using simple periodic table descriptors that are easy to interpret and can be calculated quickly without the involvement of expert personnel. These descriptors simplify the nanostructure property calculation and determine the nanoscale interactions without much computational intervention. The use of such descriptors saves time; the descriptors are also cost-effective and have a clear and straightforward physical

meaning, which facilitates the mechanical interpretation of the QSAR models. A direct comparison was not possible due to different dataset division and descriptors but the results obtained in the present work for the RF method was superior to that of the previous work.

Conclusion

We have performed cytotoxicity modeling of eight heavy metal compounds adsorbed on nanoscale TiO₂ regarding HK-2 cells and explored the features responsible for the toxicity mechanism. Many studies have examined the co-exposure of metal and metalloid mixtures with heavy metals. The co-exposure may also be affected by dose variations at the biomarker level. Also, co-exposure in humans was found to lead to more profound renal damage than exposure to each of the elements alone. Hence, to elucidate the features responsible for the toxicity, in the present study, ML algorithms were applied along with periodic table descriptors for QSAR modeling. Experiment-independent periodic table descriptors produced better results than quantum chemical descriptors in previous studies. The periodic table descriptors used in QSAR models have strong

Table 3: Comparison of the current work with the previous study.

Descriptors	Method	R^2	$Q^2_{(\text{LOO})}$	$\text{MAE}_{\text{train}}$	RMSE_C	Q^2_{F1}	Q^2_{F2}	MAE_{test}	RMSE_P
periodic table-based (current study)	random forest	0.96	0.72	0.13	0.2	0.94	0.94	0.14	0.19
quantum mechanical (Sang et al.)	random forest	0.85	0.70	—	0.06	0.86	0.85	—	0.10

theoretical guidance, which can help scientists design new entities with expected properties. As a part of the model development process, periodic table descriptors can be used in conjunction with other descriptors that are compatible with them. The periodic table descriptors are not only less computationally demanding but also independent of the size of the particles. The ML algorithm with periodic table descriptors has helped to evaluate the cell survival rate of HK-2 cells in less time and at less cost than using expensive quantum chemical descriptors and experimental descriptors. Among all algorithms, the random forest model shows the best prediction ability with $Q^2_{F1} = 0.94$ and $MAE_{test} = 0.14$ for the test set. Hence, a good feature selection method reduced the computation time required to train a model. The SHAP analysis also emphasized the most significant features contributing to the model. We have proposed also a generalized mechanism for the most impactful features generated by the model. As a result, periodic table descriptors and machine learning can be used together to decipher features of unknown compounds and predict compounds that are similar.

Supporting Information

Supporting Information File 1

Detailed information regarding heavy metals at different concentrations.

[<https://www.beilstein-journals.org/bjnano/content/supplementary/2190-4286-14-77-S1.xlsx>]

Funding

This study was funded by Indian Council of Medical Research (ICMR), New Delhi, in the form of a Senior Research Fellowship to JR. SP acknowledges AICTE (All India Council for Technical Education), New Delhi for financial assistance.

ORCID® iDs

Joyita Roy - <https://orcid.org/0000-0001-5000-7073>

Kunal Roy - <https://orcid.org/0000-0003-4486-8074>

References

- Kumar, P.; Mahajan, P.; Kaur, R.; Gautam, S. *Mater. Today Chem.* **2020**, *17*, 100332. doi:10.1016/j.mtchem.2020.100332
- Roy, J.; Roy, K. *Environ. Sci.: Nano* **2021**, *8*, 3395–3407. doi:10.1039/d1en00733e
- Chong, M. N.; Tneu, Z. Y.; Poh, P. E.; Jin, B.; Aryal, R. *J. Taiwan Inst. Chem. Eng.* **2015**, *50*, 288–296. doi:10.1016/j.jtice.2014.12.013
- Aruoja, V.; Dubourguier, H.-C.; Kasemets, K.; Kahru, A. *Sci. Total Environ.* **2009**, *407*, 1461–1468. doi:10.1016/j.scitotenv.2008.10.053
- Qi, Y.; Xiang, B.; Zhang, J. *Sol. Energy Mater. Sol. Cells* **2017**, *172*, 34–43. doi:10.1016/j.solmat.2017.07.017
- Zhang, Y.; Tang, Z.-R.; Fu, X.; Xu, Y.-J. *ACS Nano* **2010**, *4*, 7303–7314. doi:10.1021/nn1024219
- Xu, X.; Li, Y.; Wang, Y.; Wang, Y. *Toxicol. In Vitro* **2011**, *25*, 294–300. doi:10.1016/j.tiv.2010.09.007
- Tomno, R. M.; Nzeve, J. K.; Mailu, S. N.; Shitanda, D.; Waswa, F. *Sci. Afr.* **2020**, *9*, e00539. doi:10.1016/j.sciaf.2020.e00539
- Alam, R.; Ahmed, Z.; Howladar, M. F. *Groundwater Sustainable Dev.* **2020**, *10*, 100311. doi:10.1016/j.gsd.2019.100311
- Hu, X.; Chen, Q.; Jiang, L.; Yu, Z.; Jiang, D.; Yin, D. *Environ. Pollut.* **2011**, *159*, 1151–1158. doi:10.1016/j.envpol.2011.02.011
- Yang, W.-W.; Wang, Y.; Huang, B.; Wang, N.-X.; Wei, Z.-B.; Luo, J.; Miao, A.-J.; Yang, L.-Y. *Environ. Sci. Technol.* **2014**, *48*, 7568–7575. doi:10.1021/es500694t
- Tan, C.; Fan, W.-H.; Wang, W.-X. *Environ. Sci. Technol.* **2012**, *46*, 469–476. doi:10.1021/es202110d
- Ahmad, S. Z. N.; Wan Salleh, W. N.; Ismail, A. F.; Yusof, N.; Mohd Yusop, M. Z.; Aziz, F. *Chemosphere* **2020**, *248*, 126008. doi:10.1016/j.chemosphere.2020.126008
- Rajeshkumar, S.; Liu, Y.; Zhang, X.; Ravikumar, B.; Bai, G.; Li, X. *Chemosphere* **2018**, *191*, 626–638. doi:10.1016/j.chemosphere.2017.10.078
- Durdagi, S.; Mavromoustakos, T.; Chronakis, N.; Papadopoulos, M. G. *Bioorg. Med. Chem.* **2008**, *16*, 9957–9974. doi:10.1016/j.bmc.2008.10.039
- Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S. Y.; Mumper, R. J.; Tropsha, A. *ACS Nano* **2010**, *4*, 5703–5712. doi:10.1021/nn1013484
- Chatterjee, M.; Banerjee, A.; De, P.; Gajewicz-Skretna, A.; Roy, K. *Environ. Sci.: Nano* **2022**, *9*, 189–203. doi:10.1039/d1en00725d
- Oh, E.; Liu, R.; Nel, A.; Gemill, K. B.; Bilal, M.; Cohen, Y.; Medintz, I. L. *Nat. Nanotechnol.* **2016**, *11*, 479–486. doi:10.1038/nnano.2015.338
- Opitz, D.; Maclin, R. J. *Artif. Intell. Res.* **1999**, *11*, 169–198. doi:10.1613/jair.614
- Sang, L.; Wang, Y.; Zong, C.; Wang, P.; Zhang, H.; Guo, D.; Yuan, B.; Pan, Y. *Molecules* **2022**, *27*, 6125. doi:10.3390/molecules27186125
- Racusen, L. C.; Monteil, C.; Sgrignoli, A.; Lucskay, M.; Marouillat, S.; Rhim, J. G. S.; Morin, J.-P. *J. Lab. Clin. Med.* **1997**, *129*, 318–329. doi:10.1016/s0022-2143(97)90180-3
- Ryan, M. J.; Johnson, G.; Kirk, J.; Fuerstenberg, S. M.; Zager, R. A.; Torok-Storb, B. *Kidney Int.* **1994**, *45*, 48–57. doi:10.1038/ki.1994.6
- Kar, S.; Gajewicz, A.; Puzyn, T.; Roy, K.; Leszczynski, J. *Ecotoxicol. Environ. Saf.* **2014**, *107*, 162–169. doi:10.1016/j.ecoenv.2014.05.026
- De, P.; Kar, S.; Roy, K.; Leszczynski, J. *Environ. Sci.: Nano* **2018**, *5*, 2742–2760. doi:10.1039/c8en00809d
- Menze, B. H.; Kelm, B. M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F. A. *BMC Bioinf.* **2009**, *10*, 213. doi:10.1186/1471-2105-10-213
- Marcilio, W. E.; Eler, D. M. From explanations to feature selection: assessing SHAP values as feature selection mechanism. *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*; IEEE: Piscataway, NJ, U.S.A., 2020; pp 340–347. doi:10.1109/sibgrapi51738.2020.00053
- El Naqa, I.; Murphy, M. J. What Is Machine Learning?. *Machine Learning in Radiation Oncology*; Springer International Publishing: Cham, Switzerland, 2015; pp 3–11. doi:10.1007/978-3-319-18305-3_1

28. Russell, S. J.; Norvig, P.; Davis, E.; Edwards, D. D.; Forsyth, D.; Hay, N. J.; Malik, J. M.; Mittal, M.; Sahami, M.; Thrun, S. *Artificial intelligence a modern approach*; Pearson Education, Inc.: Upper saddle River, New Jersey, 2010.
29. Breiman, L. *Mach. Learn.* **2001**, *45*, 5–32.
doi:10.1023/a:1010933404324
30. Friedman, J. H. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378.
doi:10.1016/s0167-9473(01)00065-2
31. Freund, Y.; Schapire, R. E. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
doi:10.1006/jcss.1997.1504
32. Saadat, A.; Dehghani Varniab, A.; Madani, S. M. *J. Nanomater.* **2022**, *2022*, 4986826. doi:10.1155/2022/4986826
33. Chen, T.; Guestrin, C. *XGBoost: A Scalable Tree Boosting System*. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016; pp 785–794.
doi:10.1145/2939672.2939785
34. Criminisi, A.; Shotton, J.; Konukoglu, E. *Found. Trends Comput. Graphics Vision* **2012**, *7*, 81–227.
doi:10.1561/06000000035
35. Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. *Risks* **2019**, *7*, 70.
doi:10.3390/risks7020070
36. Liu, Y.; Liu, Z.; Luo, X.; Zhao, H. *Biocybern. Biomed. Eng.* **2022**, *42*, 856–869. doi:10.1016/j.bbe.2022.06.007
37. Lund, B.-O.; Miller, D. M.; Woods, J. S. *Biochem. Pharmacol.* **1991**, *42*, S181–S187. doi:10.1016/0006-2952(91)90408-w
38. Niki, E. *BioFactors* **2008**, *34*, 171–180. doi:10.1002/biof.5520340208
39. Pujalté, I.; Passagne, I.; Brouillaud, B.; Tréguer, M.; Durand, E.; Ohayon-Courtès, C.; L'Azou, B. *Part. Fibre Toxicol.* **2011**, *8*, 10.
doi:10.1186/1743-8977-8-10
40. Roy, J.; Roy, K. *Nanotoxicology* **2022**, *16*, 629–644.
doi:10.1080/17435390.2022.2132887

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjnano/terms>), which is identical to the Creative Commons Attribution 4.0

International License

(<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:

<https://doi.org/10.3762/bjnano.14.77>



Exploring the relationships between physiochemical properties of nanoparticles and cell damage to combat cancer growth using simple periodic table-based descriptors

Joyita Roy and Kunal Roy^{*§}

Full Research Paper

Open Access

Address:

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India

Email:

Kunal Roy^{*} - kunal.roy@jadavpuruniversity.in

* Corresponding author

§ Phone: +91 98315 94140; Fax: +91-33-2837-1078; Email: kunalroy_in@yahoo.com

Keywords:

cancer cell treatment; cell damage; MeOx NMs (metal oxide nanomaterials); nano-QSPR; zeta potential

Beilstein J. Nanotechnol. **2024**, *15*, 297–309.

<https://doi.org/10.3762/bjnano.15.27>

Received: 21 November 2023

Accepted: 23 February 2024

Published: 12 March 2024

Associate Editor: A. Salvati



© 2024 Roy and Roy; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

A comprehensive knowledge of the physical and chemical properties of nanomaterials (NMs) is necessary to design them effectively for regulated use. Although NMs are utilized in therapeutics, their cytotoxicity has attracted great attention. Nanoscale quantitative structure–property relationship (nano-QSPR) models can help in understanding the relationship between NMs and the biological environment and provide new ways for modeling the structural properties and bio-toxic effects of NMs. The goal of the study is to construct fully validated property-based models to extract relevant features for estimating and influencing the zeta potential and obtaining the toxicity profile regarding cell damage in the treatment of cancer cells. To achieve this, QSPR modeling was first performed with 18 metal oxide (MeOx) NMs to measure their materials properties using periodic table-based descriptors. The features obtained were later applied for zeta potential calculation (imputation for sparse data) for MeOx NMs that lack such information. To further clarify the influence of the zeta potential on cell damage, a QSPR model was developed with 132 MeOx NMs to understand the possible mechanisms of cell damage. The results showed that zeta potential, along with seven other descriptors, had the potential to influence oxidative damage through free radical accumulation, which could lead to changes in the survival rate of cancerous cells. The developed QSPR and quantitative structure–activity relationship models also give hints regarding safer design and toxicity assessment of MeOx NMs.

Introduction

Engineered nanoparticles have become an integral part of our daily lives in consumable products and commercial goods. Their versatile tunable properties have made nanomaterials a center of innovation in different areas [1]. However, the innova-

tion of nanomaterials (NMs) is hindered because of potential adverse effects. It is believed that small particles can enter the body through inhalation, ingestion, and skin penetration and have the potency to interact with macromolecules for a long

period. Many studies have demonstrated that metal oxide nanoparticles (MeOx NPs) are toxic and tend to have adverse effects on living organisms and the environment [2–6]. The toxicity of NPs depends on various structural (intrinsic) [7] and extrinsic properties. Depending on the dispersing environment, nanoparticles can easily agglomerate into particles with larger diameter. Upon intake by organisms, depending on the pH value, these agglomerations disintegrate again becoming a source for toxins in the body [8]. The formation of agglomerated NPs depends upon the surface charge of the NPs, which is believed to stabilize and prevent agglomeration of NPs. As no experimental techniques are available to measure the surface charge directly, its value is measured through the zeta potential (ζ) in a given medium [9]. Zeta potential is the electrostatic potential at the electrical double layer surrounding the NPs in solution. It is closely related to suspension stability and morphology. In metals, the zeta potential can be altered by altering pH, concentration, and conductivity of the components of NPs [10]. Zeta potential can provide information regarding the fate, behavior, and toxicity of NPs in the environment as well as in biological systems. Since the cell membrane is negatively charged, the interaction between NPs and cell membrane or organelles can be highly influenced by the zeta potential. There is an increased interest in integrating data on metal oxides in the field of nanotoxicology that would be able to predict toxicity based on measured properties. Indeed, there are several studies related to the zeta potential and its behavior in solutions and biological systems [11]. Comparable zeta potential measurements across various studies may allow one to find correlations regarding the behavior of different types of NMs. These correlations can then enable the prediction of the behavior of novel NMs based on their properties. As the zeta potential is a system-dependent extrinsic property, it depends on both particle and medium. The behavior of NPs can also change depending on the formation of a protein corona. The formation of a protein corona on the surface of NPs, which influences the interaction with cell membranes or proteins, is also associated with zeta potential and surface charge. Very limited studies have reported the influence of zeta potential, surface charge, hydrophobicity, and biocompatibility on NP toxicity. These properties of NPs determine their toxicity and interaction with the cell membrane damaging human health and the environment [12]. The toxic effect of NPs can be used as a medical treatment for diseases at the cellular level, that is, targeting and destroying cancerous cells. To date, few studies have reported on the mechanism of apoptosis of cancerous cells after metal oxide treatment, which still remains unclear. Traditional approaches are very costly, time-consuming, involve a lot of resources and lead to ethical implications; also, they are inadequate in addressing the safety concerns regarding new NPs in this rapidly growing field. Therefore, computational-based approaches are effective methods in risk

assessment. Among them, quantitative structure–property relationship (QSPR) models seem to be the most promising method [13]. However, the physicochemical and structural diversity of metal oxide nanoparticles (MeOx NPs) poses significant challenges in determining their toxic effect on living cells [14,15]. Works related to nanoscale toxicity modeling have been published [16–20] to predict the toxicity profile of MeOx NPs on various cell lines and species. The most important criterion to improve nanoscale toxicity models is the selection of the appropriate structural descriptors of NPs. Periodic table-based descriptors have been a promising tool in predicting toxicity profiles and risk assessment of MeOx NPs with high predictivity and interpretability [21–25]. This type of descriptors can indicate relevant features and mend the mechanism interpretation. Some properties (size, zeta potential, molecular weight, mass percentage of metal elements, and cation charge) are investigated to have a better understanding of the structure of NPs and its influence on toxicity.

Methods and Materials

Dataset

The study is based on two datasets, that is, dataset I (zeta potential) and dataset II (cell membrane damage). Dataset I consists of 18 metal oxide nanoparticles (MeOx NPs) with stoichiometries of MO, MO₂, MO₃, M₂O₃, and M₃O₄. This data was obtained from Cao et al. [26], where the zeta potential of MeOx NPs was measured in a cell culture of 20% fetal bovine complete medium. Dataset II was taken from Toropova et al. [27], where cell damage measurement was performed based on the uptake of propidium iodide (PI). The dataset is related to four doses (50, 100, 150, and 200 µg/mL) and exposure times ranging from 1 to 7 h, which results in 132 MeOx NPs data points. The detailed dataset is provided in Supporting Information File 2, Section S1.

Descriptor calculation

Selecting the appropriate descriptors is crucial for property and toxicity modeling. Quantitative values of chemical features (descriptors) play a significant role in determining the target endpoint. Therefore, in this study, we have calculated periodic table-based descriptors (PT descriptors) for calculating the relevant features contributing to the respective property and toxicity endpoint. Physicochemical features encoding the information of MeOx NPs into PT descriptors were used to build prediction models for zeta potential and cytotoxicity (cell damage). The basic information of MeOx NPs was directly taken from the periodic table and some were calculated with the Elemental Descriptor Calculator software available from (<https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/other-dtc-lab-tools?authuser=0>), termed first-generation periodic table descriptors. Also, second-generation PT descriptors

were calculated using relevant formulas [28]. These descriptors were calculated without any expert intervention and are independent of size variations.

Splitting of the data sets

Splitting of the datasets into training sets and test sets is essential for developing statistically robust nano-QSPR models. Each of the datasets, that is, the zeta potential dataset and cell damage dataset, was divided into training and test sets with a ratio of 7:3 using the dataset division software in the DTC lab software suite (http://teqip.jdvu.ac.in/QSAR_Tools/). Accordingly, thirteen compounds were in the training set and five compounds in the test set for the zeta potential dataset; for the cell damage dataset, 111 compounds were present in the training set, and the remaining 21 compounds were in the test set. The training set compounds were used for feature selection and model development; the test set was utilized for assessing the predictivity of the developed model.

Model development

Zeta potential QSPR model

To develop the property-based QSPR model, the training set was utilized for model development. The training set of 13 compounds was processed through feature selection via stepwise regression and genetic algorithm (GA) [29]. After feature selection, the training set was utilized for model development through stepwise regression using the MINITAB software (Minitab Inc., USA, <https://www.minitab.com>). A multiple linear regression (MLR) model was obtained with three descriptors keeping the F values to enter and remove 4 and 3.9, respectively. Finally, a PLS (partial least squares) model was developed with the selected features from the MLR model. The developed PLS model consisted of 1 LV (latent variable), which was also developed in the MINITAB software.

Cell damage QSPR model

The previously developed QSPR model (dataset I) was utilized to calculate the zeta potential of the MeOx NPs in the cell damage dataset (dataset II), which lacks the zeta potential information (imputation of sparse data). The zeta potential was used as a descriptor in the model development along with the PT descriptors. Although the solvents used for testing metal oxides in both datasets differ, the work involves correlating the zeta potential data (experimental or computed) with the cell damage model as a descriptor. Cao et al. [26] also used zeta potential as one of the determinants for the modeled endpoint. The zeta potential of all data points was determined in the same solvent, and this does not contribute to the variations in zeta potential values due to solvents. This work is similar to imputation in quantitative structure–activity relationship (QSAR) modeling, where a missing value is replaced by a predicted value from

another model [30]. The training set with 111 MeOx NPs after feature selection through GA was further used for model development. The model development was performed with stepwise regression using the MINITAB software followed by the best subset selection method. Further, to enhance the quality of predictions for the test set, we have performed a chemical read-across approach for the developed MLR model with eight descriptors.

Model validation

The validation procedure is the prerequisite for the application of nano-QSPR models. Rigorous validation of the developed models was performed following principles of the Organization for Economic Cooperation and Development [31]. Validation of the model includes both internal and external validation. Internal validation indicates the robustness and fit of the developed model applying the training set, whereas the test set indicates the predictivity of the developed model for new NMs. Common internal validation methods include the leave-one-out cross-validation (Q^2_{LOO}) algorithm and the Y-randomization test [32,33]. The model fit ability is expressed by the determination coefficient (R^2) and mean absolute error (MAE). For judging the external predictivity for the test set, we chose the Q^2_{f1} and Q^2_{f2} metrics. According to Golbraikh and Tropsha [34], R^2 should be greater than 0.6 and Q^2 should be greater than 0.5 to meet the standard requirements of external validation. A true external set was also used to evaluate the predicting power of the model. This was done using the prediction reliability indicator (PRI) tool available from the DTC lab software tools (http://teqip.jdvu.ac.in/QSAR_Tools/). To further validate model 2 for the similarity-based prediction, we have performed chemical read-across analysis.

Prediction reliability indicator (PRI) tool

Ensuring the reliability of predictions for a new set of data is a vital task. By making robust predictions based on molecular features, we can estimate the external set accurately. In this study, we used the Prediction Reliability Indicator tool [35] (http://teqip.jdvu.ac.in/QSAR_Tools/) to predict the response of a true external set comprising 49 MeOx NPs. The tool categorizes the prediction quality as good, moderate, or bad, based on certain scoring rules. To assess the predictive power of the developed QSPR models, we used the QSPR model (model 1 with zeta potential endpoint) to predict the response of the external set. Figure 1 shows the overall workflow of the present work, highlighting our confident approach to the study.

Read-across analysis

The read-across technique is a reliable and scientifically proven method to predict the endpoint of a new compound, also known as the target compound. This technique involves utilizing data

performed considering the zeta potential as the Y-response (model 1). Model 1 was developed with basic periodic table-based descriptors. The different validation metrics showed the models to be robust and of good predictivity. Furthermore, toxicity-based modeling (model 2) was conducted to illustrate the impact of zeta potential on BEAS-2B cell damage. The modeling aimed to create robust and predictive property- and toxicity-based models capable of predicting novel MeOx NPs with enhanced features. Figure 2 shows the bubble plots for both dataset 1 and dataset 2. The green and red colors indicate the positive and negative coefficients of the respective descrip-

tors. The size of the bubble represents the importance of the descriptors; smaller bubbles indicate less contribution to the respective endpoints than larger bubbles. The Y-randomization plot and loading plot are also reported in Supporting Information File 1 and Supporting Information File 2, Figure S1 and Figure S2. The Williams plot in Figure 3 shows that three compounds were outliers in the cell damage dataset. According to the PRI tool estimation on a true external set, out of 49 MeOx NPs, we confidently predicted 39 with good accuracy using this simple tool. This means that we were able to make predictions for untested metal oxides with great confidence.

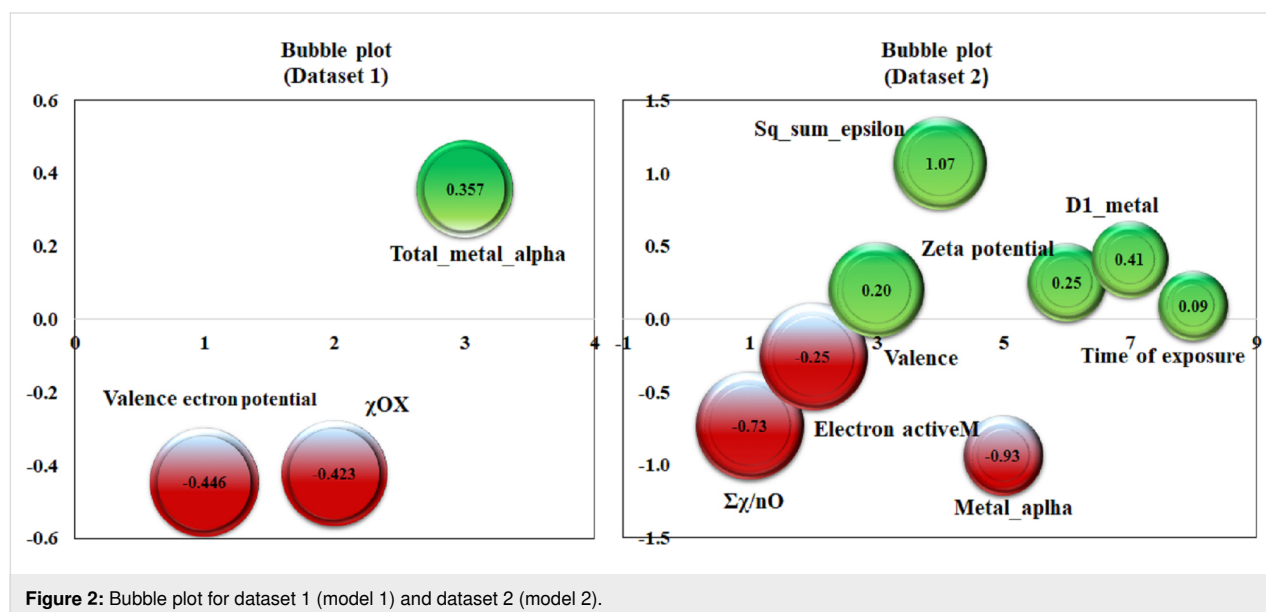


Figure 2: Bubble plot for dataset 1 (model 1) and dataset 2 (model 2).

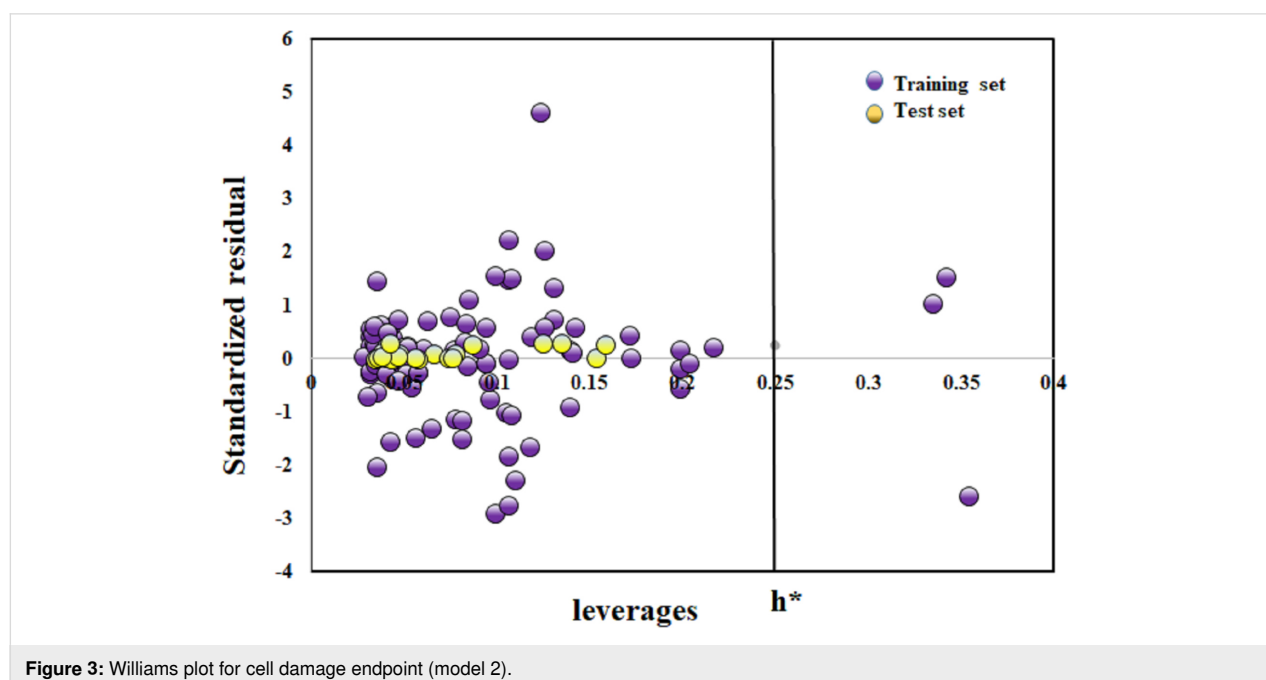


Figure 3: Williams plot for cell damage endpoint (model 2).

QSPR model for zeta potential

The zeta potential is the key parameter from the regulatory point and can directly affect the NPs' behavior in solution and their interaction with biological organisms (Figure 4). 18 MeOx NPs were modeled against the zeta potential endpoint to obtain the partial least squares (PLS) model with one latent variable (LV).

Model 1 (PLS)

$$\zeta = 35.9157 - 8.4317\chi_{\text{ox}} + 2.0002\text{tot_metal_alpha} - 0.1854 \text{ valence electron potential}$$

$$N_{\text{train}} = 13; R^2 = 0.80; Q_{(\text{LOO})}^2 = 0.67; N_{\text{test}} = 5; Q_{f1}^2 = 0.68; (1) \\ Q_{f2}^2 = 0.67; \text{LV} = 1; F = 44.20; p = 0.002$$

Model 1 considers three descriptors to evaluate the influence of the zeta potential based on basic attributes. Here, N_{train} and N_{test} stand for the number of training and test set compounds, respectively. R^2 is the determination coefficient; $Q_{(\text{LOO})}^2$ is the leave-one-out cross validation determination coefficient. Again, Q_{f1}^2 and Q_{f2}^2 were calculated for external data predictions. The model parameters suggest the good predictive ability of the developed model as it passes various statistical criteria [34]. The descriptors depicted in the model also interpret the influence of the zeta potential as discussed below.

The descriptor “ χ_{ox} ” pertains to the oxidation number of the metal, which represents the hypothetical charges within an

atom. The zeta potential decreases as the oxidation number increases, as indicated by the negative coefficient of the descriptor. A lower (negative) oxidation number indicates a higher electronegativity of the metal, which determines the electron distribution in a molecule. The metal's electronegativity also influences the catalytic property of the cationic form and the surface charge formed around the metal oxide surface. The highly electronegative surface of MeOx NPs [39] affects their behavior and stability, thus determining the net charge of ions in a given medium. Certain MeOx NPs are unstable and tend to agglomerate. NPs attract negative or positive ions from the medium to build a diffusion double layer. The electronegativity of the NPs also depends on the pH value of the medium [40]. In colloidal solutions, negatively charged metal oxides decrease the zeta potential, which reflects stability based on the aggregation phenomenon. This is well observed in MeOx NPs, where an increase in the oxidation number (χ_{ox}) decreases the zeta potential. In WO_3 NPs, the χ_{ox} value is 6 and the zeta potential value is -23 mV; for NiO NPs, the χ_{ox} value is 2, and the zeta potential value is 34.4 mV.

The “valence electron potential” ($-eV$) determines the elements' reactivity and is based on the charge of the valence electrons and the ionic radius:

$$-eV = kn/r.$$

Here, k is a proportionality factor expressing the energy of the valence electrons in electronvolts. n is the valence, and r is the ionic radius.

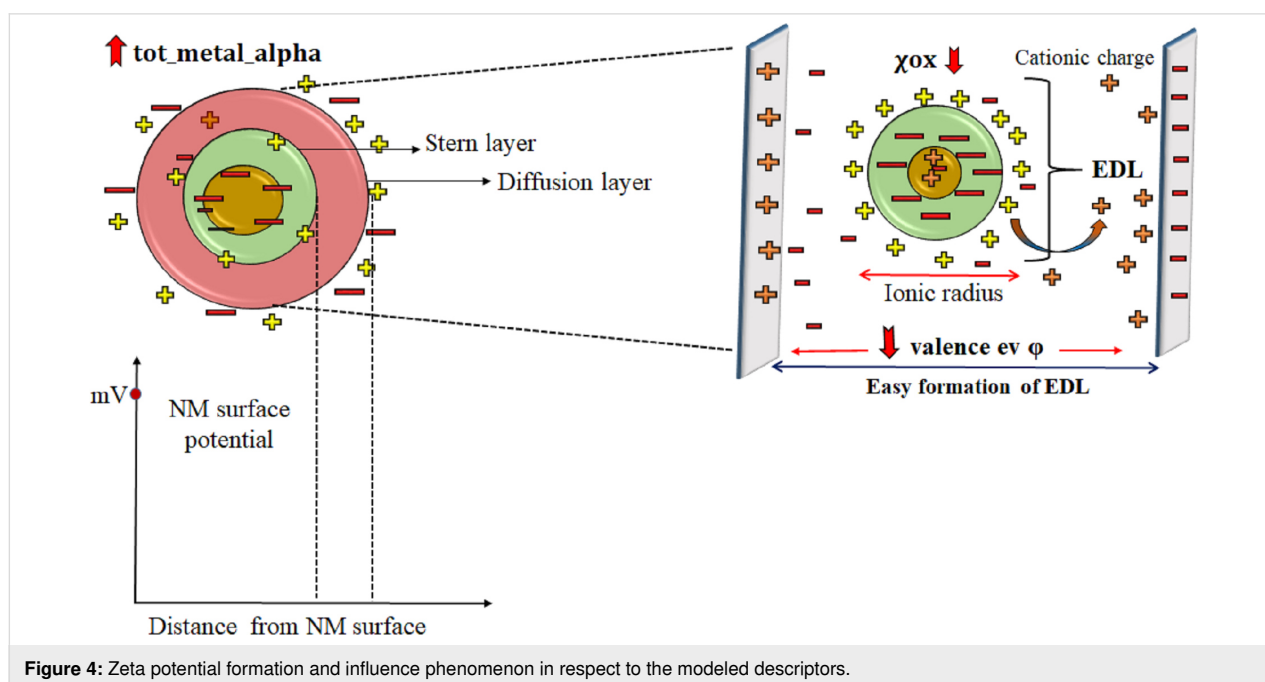


Figure 4: Zeta potential formation and influence phenomenon in respect to the modeled descriptors.

This descriptor negatively contributes to the zeta potential suggesting that with the increase of the valence electron potential of the metal, there will be a decrease in zeta potential value. This has been observed in Mn_2O_3 NPs, which have a valence electron potential value of 220eV and a zeta potential of -15.9 mV . Co_3O_3 NPs show the opposite result; the decrease in the valence electron potential value (38eV) shows an increase in zeta potential value (22.6 mV). MeOx NPs with large ionic radius tend to have low valence electron potential, as it is inversely proportional to the ionic radius of the NPs. NPs with lower valence potential allow for an easier formation of the electrostatic double layer (EDL). If the solution with NPs shifts to lower ionic strength, then the zeta potential increases as the EDL expands to balance the electrostatic force, thus allowing for the dispersion of NPs.

The descriptor “tot_metal_alpha” defines the core environment of the metal. It also defines the molecular bulk of the metal oxide. This descriptor has vital characteristics that are heavily influenced by the number of metals present in the metal oxide. Furthermore, the electronegativity of the metal is a crucial factor in determining the surface charge and stability of the NPs in the solution. The positive regression coefficient suggests that an increase in the surface charge of the metal helps the NPs to remain dispersed in the media and thus avoids flocculation. This phenomenon is observable in Yb_2O_3 NPs with a high tot_metal_alpha value (13.6) and the highest zeta potential (46 mV); in contrast, SnO_2 NPs with a descriptor value of 2.88 have a zeta potential value of -20.5 mV .

QSPR model for cell damage Model 2 (PLS)

$$\begin{aligned} \text{cell damage} = & -1.681 - 1.11 \frac{\sum \chi}{n\text{O}} + 0.295 \frac{\text{sq}_{\text{sumepsilon}}}{N} \\ & + 0.318 D1_{\text{metal}} - 0.263 \text{Metal alpha} \\ & + 0.035 \text{time of exposure} + 0.00057 \text{zeta potential} \\ & + 0.0057 \text{Electrons}_{\text{ActiveM}} - 0.079 \text{valence} \end{aligned} \quad (2)$$

$$\begin{aligned} N_{\text{train}} = 110; R^2 = 0.62; Q_{(\text{LOO})}^2 = 0.54; rm_{\text{LOO}}^2 = 0.389; \\ \Delta rm_{\text{LOO}}^2 = 0.246; N_{\text{test}} = 21; Q_{f1}^2 = 0.653; Q_{f2}^2 = 0.652; \\ rm_{\text{LOO}}^2 = 0.532; \Delta rm_{\text{LOO}}^2 = 0.183; LV = 7; \\ MAE_{\text{test}} = 0.206; F = 19.87 \end{aligned}$$

Model 2 utilizes eight descriptors to evaluate crucial attributes that can impact cell damage. Equation 2 shows the number of compounds used in the training and test sets represented by N_{train} and N_{test} , respectively. Additionally, R^2 and $Q_{(\text{LOO})}^2$, the determination coefficient and leave-one-out (LOO) cross-vali-

dation coefficient, were employed. Furthermore, external data prediction calculations were made using Q_{f1}^2 and Q_{f2}^2 . The model parameters demonstrate exceptional predictive ability, meeting various statistical criteria [34]. The descriptors used in the model were well interpreted and are comprehensively discussed in a later section. Note that the zeta potential has appeared as a significant descriptor in defining the cell damage. On removal of zeta potential as a descriptor, the model quality decreases. Predictions from one model as a descriptor for another model are made to fill the data gap or to determine the missing values. This approach is similar to the imputation methodology, which creates a model embedded within another model. Instead of using dummy variables for quantitative prediction, a useful imputation method can predict various types of inputs. It is worth noting that many existing works utilize imputation techniques [41]. In QSAR studies, it is not unusual to use a model-derived prediction as a descriptor for the development of other models or for prediction when the endpoint has been tested under different experimental or varying conditions (as in the case of interspecies modeling). This approach is reliable and aims to establish a correlation between different conditions to fill the data gap.

Chemical read-across analysis

The developed QSPR (PLS) model for dataset 2 provided eight descriptors that were utilized for read-across predictions. Three similarity-based prediction methods, namely Euclidean distance (ED)-based, Gaussian kernel (GK) similarity-based, and Laplacian kernel (LK) were employed. Upon optimizing the dataset, it was concluded that the read-across based on the Euclidean distance (RA-ED) function outperformed the others, as shown in Table 1. The Read-Across v4.0 software [42] was utilized for this work. After performing RA, the resultant Q_{f1}^2 increased from 0.65 to 0.766.

Interpretation of the descriptors

The periodic table descriptor $\sum \chi/n\text{O}$ stands for the total metal electronegativity in a specific metal oxide relative to the number of oxygen atoms. This descriptor takes into account the crucial role of oxygen atoms in causing cell damage. With regard to the cell damage endpoint, this descriptor has a negative effect, indicating that an increase in the number of oxygen atoms compared to the electronegativity sum results in a lower ratio of the descriptor. Thus, a high concentration of oxygen atoms in the metal oxide can expedite the oxidative damage process, leading to the production of more reactive oxygen species (ROS) and causing more cell damage. CoO NPs show that a high $\sum \chi/n\text{O}$ value (1.88) leads to less cell damage (-4.38), whereas a low value ($\sum \chi/n\text{O} = 0.77$) leads to more cell damage (-2.50) as observed for TiO_2 NPs. The production of ROS can enhance the catalytic activity of Fenton/Fenton-like

Table 1: Results for read across prediction using different similarity-based approaches.

Feature combinations	Hypothesis	Hyper parameters				Statistical parameters			
		σ	γ	Distance threshold	Similarity threshold	Q_{f1}^2	Q_{f2}^2	MAE	RMSEP
Model 2 (132 MeOx NPs) 8 descriptors (7 LVs)	RA-ED	1.75	1.75	1	0	0.766	0.765	0.177	0.252
	RA-GK					0.764	0.763	0.178	0.254
	RA-LK					0.724	0.724	0.195	0.274

reactions, but can also result in cellular damage [16]. ROS can break down the basic components of the cell, including DNA, proteins, and lipids. ROS can cause double-strand breaks in DNA by converting guanine to 8-oxoguanine. This conversion can lead to mispairing with adenine, resulting in transversion mutations. Proteins can also be damaged when their amino acid side chains are oxidized by ROS. Exposure of lipids to ROS can result in lipid peroxidation, which can cause cell damage and generate reactive by-products that further damage the cell.

The second-generation periodic table-based descriptor “sq_sum_epsilon/N” ($\sum \epsilon/N$)² stands for the sum of electronegativity of the atoms of the metal oxide, which is calculated based on the electronegativity count ($\sum \epsilon$) of the oxides, scaled by the number of atoms:

$$\sum \epsilon = \epsilon_{\text{metal}} * N_{\text{metal}} + \epsilon_{\text{oxy}} * N_{\text{oxy}}.$$

Here, ϵ_{metal} and ϵ_{oxy} are the electronegativity count of metal and oxygen atoms, respectively, and N_{metal} and N_{oxy} are, respectively, the number of metal and oxygen atoms. The positive coefficient of the descriptor in Equation 2 indicates that an increase in electronegativity favors the rise in cell damage as in CuO nanoparticles, where a high ($\sum \epsilon/N$)² value (9.93) causes more cell damage (−2.87), whereas Sb₂O₃ nanoparticles with low electronegativity ($(\sum \epsilon/N)^2 = 0.018$) are less toxic (−4.625). Because of the high electronegativity, the atoms pull electrons from their neighboring atoms or molecules, leading to the development of an electrostatic bond with proteins in biological systems. The high electronegativity also influences the formation of metal cations. The increase of catalytic properties of metal cations enhances the toxicity through the generation of ROS, causing damage to cell membranes [16]. The high electronegativity helps in removing electrons from molecules, producing free radicals. Free radicals are unstable and highly reactive. These short-lived radicals are unable to leave the sub-cellular location where they are generated without being reduced, leading to oxidative damage [43]. The presence of high-electronegativity metals in the cellular membrane can lead to the leakage of cellular content [22].

The “D1_{metal}” descriptor signifies the total number of metal atoms in the MeOx NP composition. An increase in the number of metals can have a detrimental effect on cells by impacting ROS generation. The positive coefficient of the D1_{metal} descriptor indicates that an increase in the metal fraction in MeOx NPs causes more cell damage (−2.63) as observed in Fe₃O₄ NPs (D1_{metal} = 3). In contrast, CoO NPs with a low metal fraction (D1_{metal} = 1) nanoparticles cause less cell damage (−4.375). Metal ions can generate reactive hydroxyl radicals, resulting in oxidative damage to proteins. Moreover, they can bind non-specifically to amino acid residues and replace existing metal ions at active sites of enzymes, leading to abnormal protein folding. Protein aggregation diseases are a type of neurodegenerative diseases that occur when proteins lose their structure and are deposited in the brain. These diseases are the most common type of neurodegenerative diseases. Many of these structures are highly toxic to cells [44]. The folding of proteins also causes damage to the immune system, because certain structures do not induce the production of antibodies [45].

The descriptor “Metal alpha” (α_{metal}) defines the core environment of the metal. This descriptor represents the ratio of the number of core electrons to the number of valence electrons. The Metal alpha descriptor describes the electron density of the metal. This descriptor is calculated using Equation 3:

$$\alpha_{\text{metal}} = \lambda * \mu. \quad (3)$$

Here, λ is $(Z_{\text{metal}} - Z_{\text{vmetal}})/Z_{\text{vmetal}}$ and μ is $1/(\text{PN}_{\text{metal}})$, where Z_{metal} is the atomic number, Z_{vmetal} stands for the valence electrons of the metal, and PN_{metal} stands for the periodic number in the periodic table. The negative coefficient of the descriptor signifies the low electron affinity of the metal oxide to accept electrons. This means that the metal has a propensity of having a cationic charge, which leads to the catalytic power of metal cations. For example, in WO₃, the metal alpha value is 7.2 and cell damage is −4.57. In contrast, Al₂O₃ with a metal alpha value of 1.66 causes higher damage to cells (−2.8). Metal cations are more harmful than normal nanoparticles. This is

because their electropositivity and inherent toxicity increase significantly with atomic weight. In addition, the formation of metal–ligand bonds has a direct impact on the metal's toxicity. Furthermore, it is a well-established fact that each metal has an affinity constant for various ligands, which means that most metal cations can form stable complexes with a wide variety of ligands, further increasing their potential toxicity.

In the field of physical chemistry, the zeta potential is a crucial parameter that measures the surface charge of particles relative to their size. In colloidal systems, the zeta potential is widely used as an indicator to reflect the stability. It is important to note that NPs with higher positive charges can be more harmful than those with higher negative charges. Moreover, positively charged NPs interact more significantly with cells, leading to greater cell damage. Another crucial factor to consider is that NPs with a higher zeta potential, regardless of their charge, are more easily absorbed by cells due to the electrostatic interaction between dispersed particles and the effective electric charge on the surface of the NPs [40]. This feature is particularly relevant to their biological activity, especially their ability to bind to and be absorbed by cell membranes. For instance, Cr_2O_3 NPs have a high zeta potential (2130 mV) and a high cell damage propensity, whereas Y_2O_3 NPs with a low zeta potential (−23 mV) cause less damage to cells (−4.5). The increase in zeta potential enhances the accumulation of nanoparticles on the surface of cells. The intensity of accumulation determines the toxicity of the nanoparticles. The concept of zeta potential plays a vital role in adhesion to the hydro–water interface and solid surfaces, providing an idea about the viability and permeability of the cell membrane under stress. As most of the cell surface carries a negative charge, metals with higher zeta potential can easily enter the cell and increase the production of ROS. Also, they can have a mechanical effect on the membrane, leading to depolarization of the membrane and cell damage.

The “Electron Active M” descriptor is a representation of the number of electrons that an active metal possesses. Active metals are known for their quick and robust reactions owing to the electron arrangement in their structure. These metals contain free electrons in their outermost shell that can readily create a cation by interacting with other atoms and initiating a chemical reaction. The delocalized electrons can easily interact with macroproteins, leading to the acceleration of damage to the biological membrane. A positive coefficient of Electron Active M indicates more oxidative stress and more damage to the cell due to an increase in free radicals. WO_3 has a high descriptor value of 74 resulting in high cell damage (−2.8), while Cr_2O_3 NP has a low descriptor value of 24 leading to low cellular damage (−4). Transition metals are capable of forming coordinate complexes with the imidazolyl group of histidine. These metal ions

are redox-active and can play a crucial role in the production of ROS within the cell. The reduced forms of these redox-active metal ions are involved in the Fenton reaction, which generates hydroxyl radicals from hydrogen peroxide. Similarly, the Haber–Weiss reaction involves the oxidized forms of redox-active metal ions and superoxide anions, which generate the reduced form of the metal ion. This reduced form can then be coupled to Fenton chemistry to produce hydroxyl radicals. ROS further accelerate the damage of the cell.

“Valence” (V) is a factor that contributes to cell damage. It indicates the number of electrons in the outermost shell of an atom that are available for chemical bonding and is similar to other descriptors that provide information about free electrons. The insights obtained from the developed model 2 strongly suggest that an increase in valence (7) leads to a decrease in cell damage (−3), as observed in MnO_3 NPs. This is supported by the negative regression coefficient of the descriptor. Conversely, a low valence (2) leads to greater cell damage (−2), as seen in ZnO NPs. Atoms with fewer electrons in their outer shell tend to lose them and become metal cations, which can damage cells [16]. Cations aid in the transportation of metal ions across the cell surface by interacting with its negatively charged surface. Unfortunately, this interaction can lead to DNA damage through processes such as delocalization, redox chemistry, and the generation of ROS.

Our research aimed to examine how the time of exposure to metal oxide affects cell damage, regardless of other physicochemical properties of MeOx NPs. Our findings indicate that exposure time plays a crucial role in cell damage. Prolonged exposure times increase the damaging potential. For instance, exposing cells to WO_3 NP for 7 h resulted in a cell damage score of −2.75. In contrast, exposure to Yb_2O_3 for only 1 h resulted in a score of −3.5. These results demonstrate the significance of considering exposure time when evaluating the potential risks of metal oxide exposure. When living organisms are exposed to NPs for an extended period of time, inflammatory conditions can occur that lead to physical, muscular, and neurological degeneration, or increased intensity of oxidative stress. This happens because longer exposure times enhance the toxicity mechanism of NPs. In contrast, short-term exposure does not affect significantly the cells. NPs can induce oxidative stress by impairing antioxidant defenses in humans when they are chronically exposed to NPs.

Importance of the zeta potential as a descriptor

The developed QSPR model without zeta potential descriptor shows $R^2 = 0.47$ and $Q_{(\text{LOO})}^2 = 0.34$, which is well below the desired acceptance criteria. The obtained results indicate that

the fitting and robustness of the developed model without the presence of the zeta potential descriptor is unsatisfactory. Therefore, to achieve the fit and predictive power of the model, we included the zeta potential descriptor along with the other seven descriptors. In the presence of zeta potential, the statistical quality and internal validation metrics increased ($R^2 = 0.62$ and $Q_{(LOO)}^2 = 0.54$) showing the stability and predictive ability of the model.

Utilization of the metal oxide cell damage knowledge for cancer treatment

NPs have shown immense potential in treating various diseases owing to their small size and high surface-to-volume ratio, which makes them effective drug delivery systems. Metal NPs can lead to greater signal amplification, greater sensitivity, and higher detection. However, NPs with properties that generate ROS can increase cell damage. In cancer cells, rapid proliferation leads to an imbalance of oxygen, abnormal structure, and blood supply, making the tumor microenvironment (TME) prone to hypoxic conditions [46]. Insufficient oxygen reduces ROS generation, which decreases the efficacy of oxygen-dependent therapies, such as photodynamic therapy (PDT), chemodynamic therapy (CDT), and radiation therapy. The information derived from the positive contribution of the D1metal descriptor (model 2) draws attention to the fact that metal oxides are good candidates for generating oxidative stress in cells. The $\sum\chi/nO$ descriptor suggested a higher oxygen

requirement for damaging the cells. It indicates that a higher fraction of oxygen in the metal oxide nanoparticles can increase the sensitivity to PDT. Furthermore, transition metals can catalyze Fenton/Fenton-like reactions [47], generating highly oxidative species that can kill tumor cells. The electronegativity of the metal oxides helps the NPs in crossing the cell membrane. The formation of metal cations can also affect the pH value of the cell and increase the catalytic properties of metal oxides, thereby increasing ROS generation. Tumor cells have a mechanism for dealing with hypoxia, acidosis, and high glutathione (GSH) levels, which promote drug resistance, especially for ROS-dependent drugs (Figure 5). However, metal oxides can change the TME conditions by supplying oxygen and suppressing hypoxia-inducible factor 1 and CD39/CD73 in T cells, which reduces the immunosuppression effect of tumors.

Comparison with previously published literature

This study successfully develops a QSPR model with a cell damage endpoint that uses the zeta potential value as a descriptor. The descriptor was calculated using another model that used the zeta potential as the endpoint (Y-response) for its QSPR model development. The QSPR models were developed with simple periodic table-based descriptors that do not depend on size or any other experimental conditions. These descriptors are easy to calculate, less expensive, and can be calculated by anyone without the need for expert personnel.

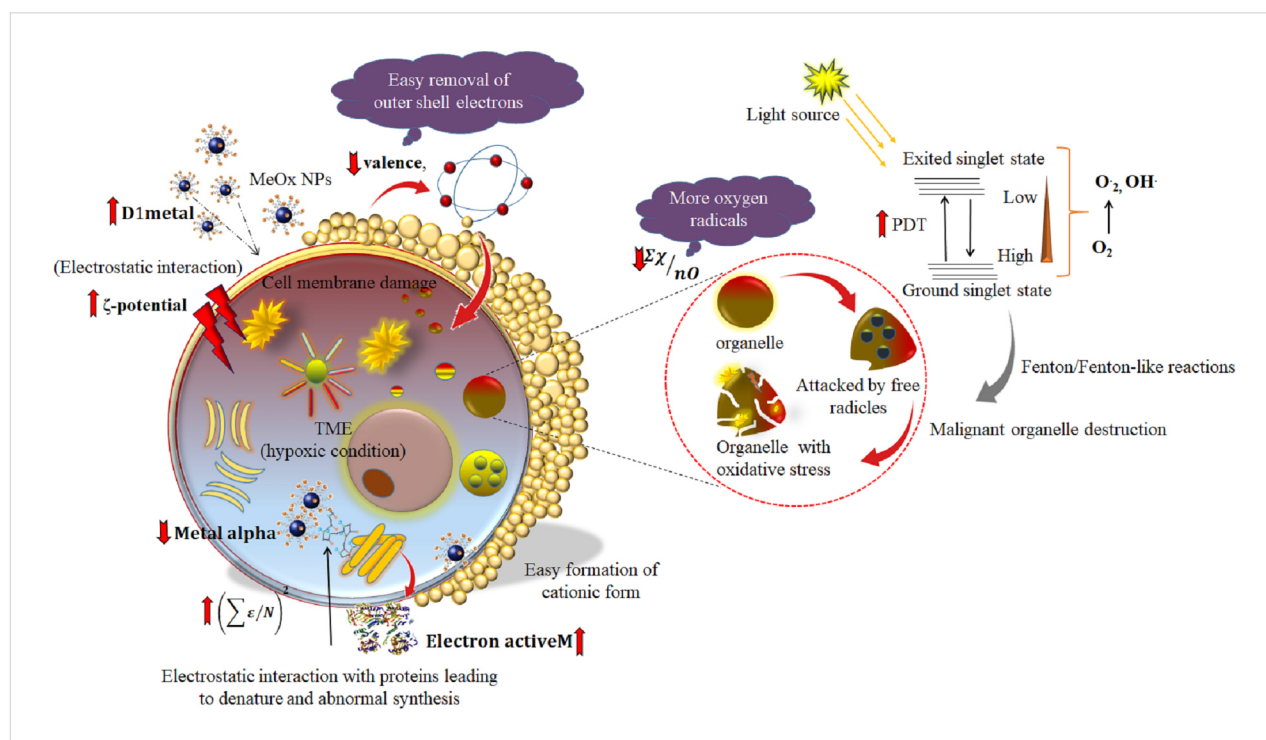


Figure 5: Interpretation of descriptors with respect to cell damage (endpoint) in cancer cells.

The study provides in-depth knowledge about the properties and causes of toxicity of nanoparticles using simple regression-based models. It is important to note that a direct comparison with a previous study by Toropova [27] is not possible because of the different data division methods (five random splits), the use of different types of descriptors (optimal nano-descriptors), and the dissimilar modeling methods (Monte Carlo method). However, it is clear that the statistical metric values for the developed model in the present study are similar to those of the previous study (the best-split results only shown) as presented in Table 2. Furthermore, we have proposed an effective mechanism to treat cancerous cells with the cell-damaging properties of MeOx NPs.

Conclusion

The impact of nanoparticles on cell membranes has been thoroughly examined through nanotoxicological research and in vitro modeling [48,49]. While the toxicity endpoint is a well-explored topic, it is crucial to investigate non-fatal endpoints such as cell damage. The zeta potential is a widely used parameter to characterize the properties of nanoparticles. However, cell membrane damage is influenced by various factors, including exposure time and dose. Thus, this study aimed to establish a relationship between the properties of nanoparticles and their toxicity, with a focus on cell membrane damage.

The study was divided into two parts. The first part involved modeling nanoparticles against the zeta potential to determine the features that can alter their properties. The second part focused on the elements that can influence toxicity and damage to the cell membrane. Both the QSPR model for the zeta potential and another model against cell damage were developed using periodic table-based descriptors. The QSPR model (zeta potential) was able to predict the zeta potential for MeOx NPs without experimental values. The developed models showed good predictivity and robustness, confirming their effectiveness.

The features obtained from the models suggest that surface charge and electronegativity play a role in altering the zeta potential. Additionally, an increase in oxygen count, electronegativity, formation of cationic charge, and an increase in zeta potential can influence cell membrane damage. Based on these

findings, the authors propose that the damaging power of metal oxide nanoparticles can be harnessed in treating cancerous cells. This study not only identifies the features required to enhance the properties of nanoparticles but also provides knowledge for treating cancerous cells through cell damage techniques. The study can pave the way for researchers to use nanoparticles in clinical practice with confidence.

Supporting Information

Supporting Information File 1: The sheet details information on the metal's oxides with zeta potential and cell damage endpoint along with the external set used in the present work. Supporting Information File 2: PLS graphs for both cell damage and zeta potential data.

Supporting Information File 1

Additional experimental data.

[<https://www.beilstein-journals.org/bjnano/content/supplementary/2190-4286-15-27-S1.xlsx>]

Supporting Information File 2

Additional figures.

[<https://www.beilstein-journals.org/bjnano/content/supplementary/2190-4286-15-27-S2.pdf>]

Funding

JR thankfully acknowledges Indian Council of Medical Research (ICMR), New Delhi for supporting financially to the work in the form of Senior Research Fellowship.

Conflict of Interest

No potential conflict of interest was reported by the authors.

ORCID® iDs

Joyita Roy - <https://orcid.org/0000-0001-5000-7073>

Kunal Roy - <https://orcid.org/0000-0003-4486-8074>

Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article.

Table 2: Comparison of the statistical parameters with a previous study.

Sl. no.	Q^2	R^2_{train}	Strain	Q^2_{f1}	R^2_{test}	MAE_{test}	S_{test}
current study	0.538	0.621	0.361	0.768	0.767	0.181	0.377
previous study (best split) [27]	0.486	0.512	0.387	—	0.822	—	0.318

References

- Kraegeloh, A.; Suarez-Merino, B.; Sluijters, T.; Micheletti, C. *Nanomaterials* **2018**, *8*, 239. doi:10.3390/nano8040239
- Service, R. F. *Science* **2003**, *300*, 236–239. doi:10.1126/science.300.5617.236
- Meng, H.; Xia, T.; George, S.; Nel, A. E. *ACS Nano* **2009**, *3*, 1620–1627. doi:10.1021/nn9005973
- Bai, Y.; Zhang, Y.; Zhang, J.; Mu, Q.; Zhang, W.; Butch, E. R.; Snyder, S. E.; Yan, B. *Nat. Nanotechnol.* **2010**, *5*, 683–689. doi:10.1038/nnano.2010.153
- Zhang, H.; Ji, Z.; Xia, T.; Meng, H.; Low-Kam, C.; Liu, R.; Pokhrel, S.; Lin, S.; Wang, X.; Liao, Y.-P.; Wang, M.; Li, L.; Rallo, R.; Damoiseaux, R.; Telesca, D.; Mädlar, L.; Cohen, Y.; Zink, J. I.; Nel, A. E. *ACS Nano* **2012**, *6*, 4349–4368. doi:10.1021/nn3010087
- Roy, J.; Roy, K. *SAR QSAR Environ. Res.* **2023**, *34*, 459–474. doi:10.1080/1062936x.2023.2227557
- Rivera Gil, P.; Oberdörster, G.; Elder, A.; Puentes, V.; Parak, W. J. *ACS Nano* **2010**, *4*, 5527–5531. doi:10.1021/nn1025687
- Sengottayan, S.; Mikolajczyk, A.; Jagiello, K.; Swirog, M.; Puzyn, T. *ACS Nano* **2023**, *17*, 1989–1997. doi:10.1021/acsnano.2c06977
- Lowry, G. V.; Hill, R. J.; Harper, S.; Rawle, A. F.; Hendren, C. O.; Klaessig, F.; Nobbmann, U.; Sayre, P.; Rumble, J. *Environ. Sci.: Nano* **2016**, *3*, 953–965. doi:10.1039/c6en00136j
- Wang, N.; Hsu, C.; Zhu, L.; Tseng, S.; Hsu, J.-P. *J. Colloid Interface Sci.* **2013**, *407*, 22–28. doi:10.1016/j.jcis.2013.05.058
- Salopek, B.; Krasic, D.; Filipovic, S. *Rud.-Geol.-Naftni Zb.* **1992**, *4*, 147–151.
- Escorihuela, L.; Martorell, B.; Rallo, R.; Fernández, A. *Environ. Sci.: Nano* **2018**, *5*, 2241–2251. doi:10.1039/c8en00389k
- Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178–180. doi:10.1038/194178b0
- Sayes, C.; Ivanov, I. *Risk Anal.* **2010**, *30*, 1723–1734. doi:10.1111/j.1539-6924.2010.01438.x
- Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. *Nat. Nanotechnol.* **2011**, *6*, 175–178. doi:10.1038/nnano.2011.10
- Roy, J.; Ojha, P. K.; Roy, K. *Nanotoxicology* **2019**, *13*, 701–716. doi:10.1080/17435390.2019.1593543
- Choi, J.-S.; Trinh, T. X.; Yoon, T.-H.; Kim, J.; Byun, H.-G. *Chemosphere* **2019**, *217*, 243–249. doi:10.1016/j.chemosphere.2018.11.014
- Mikolajczyk, A.; Sizochenko, N.; Mulkiewicz, E.; Malankowska, A.; Rasulev, B.; Puzyn, T. *Nanoscale* **2019**, *11*, 11808–11818. doi:10.1039/c9nr01162e
- Huang, Y.; Li, X.; Xu, S.; Zheng, H.; Zhang, L.; Chen, J.; Hong, H.; Kusko, R.; Li, R. *Environ. Health Perspect.* **2020**, *128*, 067010. doi:10.1289/ehp6508
- Pan, Y.; Li, T.; Cheng, J.; Telesca, D.; Zink, J. I.; Jiang, J. *RSC Adv.* **2016**, *6*, 25766–25775. doi:10.1039/c6ra01298a
- Toropova, A. P.; Toropov, A. A.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. *J. Math. Chem.* **2013**, *51*, 1684–1693. doi:10.1007/s10910-013-0177-0
- Roy, J.; Roy, K. *Environ. Sci.: Nano* **2023**, *10*, 2989–3011. doi:10.1039/d3en00598d
- Kar, S.; Gajewicz, A.; Puzyn, T.; Roy, K.; Leszczynski, J. *Ecotoxicol. Environ. Saf.* **2014**, *107*, 162–169. doi:10.1016/j.ecoenv.2014.05.026
- Roy, J.; Roy, K. *Environ. Sci.: Nano* **2022**, *9*, 3456–3470. doi:10.1039/d2en00303a
- Mikolajczyk, A.; Gajewicz, A.; Mulkiewicz, E.; Rasulev, B.; Marchelek, M.; Diak, M.; Hirano, S.; Zaleska-Medynska, A.; Puzyn, T. *Environ. Sci.: Nano* **2018**, *5*, 1150–1160. doi:10.1039/c8en00085a
- Cao, J.; Pan, Y.; Jiang, Y.; Qi, R.; Yuan, B.; Jia, Z.; Jiang, J.; Wang, Q. *Green Chem.* **2020**, *22*, 3512–3521. doi:10.1039/d0gc00933d
- Toropova, A. P.; Toropov, A. A.; Benfenati, E.; Korenstein, R.; Leszczynska, D.; Leszczynski, J. *Environ. Sci. Pollut. Res.* **2015**, *22*, 745–757. doi:10.1007/s11356-014-3566-4
- De, P.; Kar, S.; Roy, K.; Leszczynski, J. *Environ. Sci.: Nano* **2018**, *5*, 2742–2760. doi:10.1039/c8en00809d
- Sivanandam, S. N.; Deepa, S. N. *Genetic Algorithms. Introduction to Genetic Algorithms*; Springer: Berlin, Heidelberg, 2008; pp 15–37. doi:10.1007/978-3-540-73190-0_2
- Walter, M.; Allen, L. N.; de la Vega de León, A.; Webb, S. J.; Gillet, V. J. *J. Cheminf.* **2022**, *14*, 32. doi:10.1186/s13321-022-00611-w
- OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; OECD Series on Testing and Assessment; OECD Environment health and safety publications, 2014. doi:10.1787/9789264085442-en
- Gütlein, M.; Helma, C.; Karwath, A.; Kramer, S. *Mol. Inf.* **2013**, *32*, 516–528. doi:10.1002/minf.201200134
- Gramatica, P. *QSAR Comb. Sci.* **2007**, *26*, 694–701. doi:10.1002/qsar.200610151
- Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Mod.* **2002**, *20*, 269–276. doi:10.1016/s1093-3263(01)00123-1
- Roy, K.; Ambure, P.; Kar, S. *ACS Omega* **2018**, *3*, 11392–11406. doi:10.1021/acsomega.8b01647
- Chatterjee, M.; Banerjee, A.; De, P.; Gajewicz-Skretna, A.; Roy, K. *Environ. Sci.: Nano* **2022**, *9*, 189–203. doi:10.1039/d1en00725d
- Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; van de Sandt, J. J. M.; Tong, W.; Veith, G.; Yang, C. *ATLA, Altern. Lab. Anim.* **2005**, *33*, 155–173. doi:10.1177/026119290503300209
- Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. *J. Comput. Chem.* **2013**, *34*, 2121–2132. doi:10.1002/jcc.23361
- Kwon, C.-W.; Poquet, A.; Mornet, S.; Campet, G.; Delville, M.-H.; Treguer, M.; Portier, J. *Mater. Lett.* **2001**, *51*, 402–413. doi:10.1016/s0167-577x(01)00328-7
- Toropov, A. A.; Sizochenko, N.; Toropova, A. P.; Leszczynski, J. *Nanomaterials* **2018**, *8*, 243. doi:10.3390/nano8040243
- Whitehead, T. M.; Strickland, J.; Conduit, G. J.; Borrel, A.; Mucs, D.; Baskerville-Abraham, I. *J. Chem. Inf. Model.* **2023**, in press. doi:10.1021/acs.jcim.3c01695
- Banerjee, A.; Chatterjee, M.; De, P.; Roy, K. *Chemom. Intell. Lab. Syst.* **2022**, *227*, 104613. doi:10.1016/j.chemolab.2022.104613
- Ge, Y.; Zhang, Y.; Xia, J.; Ma, M.; He, S.; Nie, F.; Gu, N. *Colloids Surf., B* **2009**, *73*, 294–301. doi:10.1016/j.colsurf.2009.05.031
- Dobson, C. M. *Methods* **2004**, *34*, 4–14. doi:10.1016/j.ymeth.2004.03.002
- Bossy-Wetzel, E.; Schwarzenbacher, R.; Lipton, S. A. *Nat. Med.* **2004**, *10* (Suppl. 7), S2–S9. doi:10.1038/nm1067
- Gilkes, D. M.; Semenza, G. L.; Wirtz, D. *Nat. Rev. Cancer* **2014**, *14*, 430–439. doi:10.1038/nrc3726

47. Cao, W.; Jin, M.; Yang, K.; Chen, B.; Xiong, M.; Li, X.; Cao, G. *J. Nanobiotechnol.* **2021**, *19*, 325. doi:10.1186/s12951-021-01074-1
48. Pathakoti, K.; Huang, M.-J.; Watts, J. D.; He, X.; Hwang, H.-M. *J. Photochem. Photobiol., B* **2014**, *130*, 234–240. doi:10.1016/j.jphotobiol.2013.11.023
49. Qi, R.; Pan, Y.; Cao, J.; Jia, Z.; Jiang, J. *Chemosphere* **2020**, *249*, 126175. doi:10.1016/j.chemosphere.2020.126175

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjnano/terms>), which is identical to the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
<https://doi.org/10.3762/bjnano.15.27>

Joyita Roy.

