

Oral Cancer Classification using Deep Learning based Frameworks

submitted by

Suman Ghosh

Class Roll No. - 002210503026

Exam Roll No. - MCA244025

Master of Computer Application

under the supervision of

Prof. Ram Sarkar

in the

Department of Computer Science and Engineering

Faculty of Engineering and Technology

Jadavpur University

Faculty of Engineering and Technology
Jadavpur University

CERTIFICATE OF RECOMMENDATION

To Whom It May Concern

I hereby recommend that the project report entitled “**Oral Cancer Classification using Deep Learning based Frameworks**” has been carried out by Suman Ghosh (Reg. No.: 163618 of 2022 - 2023, Roll No.:002210503026, Exam Roll: MCA244025), under my guidance and supervision and be accepted in partial fulfilment of the requirement for the degree of MASTER of COMPUTER APPLICATION in DEPARTMENT of COMPUTER SCIENCE and ENGINEERING, JADAVPUR UNIVERSITY during the academic year 2023-2024.

Prof. Ram Sarkar
Project Supervisor
Dept. of Computer Science and Engineering
Jadavpur University

Prof.(Dr.) Nandini Mukhopadhyay
Head of the Department
Dept. of Computer Science &
Engineering
Jadavpur University

Prof. Dipak Laha
Dean, Faculty Council of
Engineering & Technology
Jadavpur University, Kolkata-032

**Faculty of Engineering and Technology
Jadavpur University**

CERTIFICATE OF APPROVAL

This is to certify that the project report entitled “**Oral Cancer Classification using Deep Learning based Frameworks**” is a bonafide record of work carried out by SUMAN GHOSH, in fulfilment of the requirements for the award of the degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the project report only for the purpose for which it has been submitted.

Signature of Examiner 1
Date:

Signature of Examiner 2
Date:

**Faculty of Engineering and Technology
Jadavpur University**

**DECLARATION OF COMPLIANCE WITH ACADEMIC
ETHICS**

I hereby declare that the project report entitled "Oral Cancer Classification using Deep Learning based Frameworks" is an outcome of my genuine and independent work, as part of my Master of Computer Application program.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, all materials and results that are not original to this work have been fully cited and referenced.

Suman Ghosh

Registration No.: 163618 of 2022 - 2023

Examination Roll No.: MCA244025

Signature with Date

Acknowledgements

I want to express my heartfelt thanks to Prof. Ram Sarkar, my supervisor in the Department of Computer Science and Engineering. His unwavering support kept me going throughout my project. Without his guidance, my work wouldn't have shaped up as it did. His helpful suggestions and countless discussions always inspired me. I feel honored to have worked under his mentorship.

I also appreciate all my teachers for sharing their knowledge and cooperating with me. Their guidance was crucial.

I would also like to thank the faculty members of the Computer Science and Engineering Department at Jadavpur University. Their continuous support meant a lot.

I am deeply grateful to Prof. Nandini Mukhopadhyay, Head of the Department of Computer Science and Engineering, for her support and encouragement.

I also acknowledge Prof. Manhantapas Kundu, coordinator of the CMATER lab, for providing the technological facilities and support that were crucial for this project.

I extend my gratitude to Akash Halder, a student of MCSE, for his significant help in the completion of my project.

Lastly, my batch mates were there for me when I needed them. Their encouragement made this journey memorable.

Suman Ghosh

Examination Roll No.:

Registration No.: 163618 of 2022 - 2023

Department of Computer Science and Engineering,

Jadavpur University.

Abstract

Oral cancer remains a significant challenge in medical diagnostics, necessitating accurate and efficient classification methods to aid in diagnosis and treatment planning. In this study, we present a comprehensive approach to oral cancer classification utilizing deep learning techniques applied to two distinct datasets, both of which are publicly available on Kaggle.com: oral cancer images (lips and tongue) and histopathological images of oral cancer.

For the first dataset consisting of histopathological images of oral cancer, a similar methodology was employed. Three state-of-the-art transfer learning CNN models, MobileNetV2, EfficientNetB3, and NASNetMobile, were trained on the dataset. Individual model testing accuracies of 89.68%, 92.06%, and 92.86% were attained, respectively. Subsequently, a fuzzy rank-based ensemble approach was applied, utilizing the Gompertz function to integrate predictions and enhance classification accuracy. The ensemble method yielded a significant improvement, achieving a testing accuracy of 94.44%.

For the second dataset comprising oral cancer images, we employed two transfer learning backbone models, MobileNetV2 and DenseNet121. These models were enhanced through the application of the Convolutional Block Attention Module (CBAM) and subsequently merged. That results an impressive accuracy of 96.97% on the test set, demonstrating the effectiveness of the proposed methodology for oral cancer classification from image data.

These findings underscore the potential of deep learning techniques for accurate oral cancer classification across diverse datasets. By leveraging both image-based and histopathological data, our research contributes to the development of reliable computational methods for oral cancer diagnosis, with implications for personalized treatment strategies and improved patient outcomes.

Contents

Abstract	i
Table of Contents	ii
List of Figures	iii
List of Tables	v
1 Introduction	1
1.1 Challenges in Oral Cancer Classifications	2
1.2 Motivation and Contributions	2
1.3 Organization of the Project Work	3
2 Literature Survey	4
3 Working Methodologies	6
3.1 Deep Transfer Learning Algorithms	6
3.1.1 MobileNetV2	6
3.1.2 DenseNet121	7
3.1.3 EfficientNetB3	8
3.1.4 NASNetMobile	9
3.2 Gompertz-based Fuzzy Rank Ensembling	10
3.2.1 Generation of confidence score	10
3.2.2 Method-1: Fuzzy-ranking based ensemble	10
3.3 Fusion of Attention-aided Features	12
3.3.1 Convolutional Block Attention Module (CBAM)	13
3.3.2 Method-2	15
4 Results and Discussion	17
4.1 Datasets	17
4.1.1 Overview of the Datasets	17
4.1.2 Data Sources and Distribution	18
4.1.2.1 Microscopic Dataset	18
4.1.2.2 Macroscopic Dataset	18
4.2 Performance Evaluation Measures	19
4.3 Results	21
4.3.1 Microscopic Dataset	21
4.3.2 Macroscopic Dataset	24

4.4	Comparison with state-of-the-art methods	26
5	Conclusions	29
5.1	Conclusions of the Present Work	29
5.2	Scope for Future Works	30
	Bibliography	31

List of Figures

3.1	Architecture of MobileNetV2 [18]	7
3.2	Architecture of DenseNet121 [20]	8
3.3	Architecture of EfficientNetB3	9
3.4	Architecture of NasNetMobile [23]	9
3.5	The re-parameterized Gompertz function	11
3.6	Convolutional Block Attention Module (CBAM) structure [25]	13
3.7	Channel attention module [25]	14
3.8	Spatial attention module [25]	14
3.9	Architecture of the model	16
4.1	Confusion Matrix	19
4.2	Epochs vs Accuracy and Epochs vs loss graph for (a)MobileNetV2, (b)EfficientNetB3 and (c)NASNetMobile	22
4.3	Confusion matrix and ROC curve of (a)MobileNetV2, (b)EfficientNetB3 and (c)NASNetMobile and (d) Ensemble of these three models	23
4.4	(a) Epochs vs Accuracy and (b) Epochs vs Loss graph for both training and validation set	24
4.5	(a) Confusion matrix and (b) ROC curve of the model	25

List of Tables

4.1	Microscopic Dataset distribution	18
4.2	Microscopic Dataset distribution	19
4.3	Results of Microscopic Dataset	23
4.4	Results of Microscopic Dataset	25
4.5	Comparison with state-of-the-art methods for Histopathological dataset	27
4.6	Comparison with state-of-the-art methods for Real-time dataset . .	28

CHAPTER 1

Introduction

Oral cancer stands as a formidable global health challenge, constituting a significant portion of head and neck malignancies. Originating primarily from squamous cells lining the oral cavity, including critical regions such as the mouth, tongue, and lips, its insidious nature underscores the urgent need for early detection and intervention. While it contributes approximately 3% to the annual cancer burden, its impact extends far beyond statistics, affecting millions worldwide. Often associated with lifestyle factors like tobacco and alcohol consumption, as well as human papillomavirus (HPV) infection, oral cancer's incidence remains disturbingly high. Despite strides in oncological advancements, mortality rates persist stubbornly, especially in underserved rural areas where access to timely diagnosis and treatment remains limited. This disparity underscores the pressing imperative for heightened awareness, early detection strategies, and efficacious treatment modalities to confront oral cancer's devastating toll. With a 5-year survival rate hovering around 50%, disparities in outcomes based on race, geography, and socioeconomic status further underscore the intricate interplay of factors contributing to its prevalence. In regions such as India, where late-stage diagnosis is prevalent, the imperative for enhanced screening and preventive measures is particularly pronounced. Comprehensive understanding of multifaceted risk factors, encompassing lifestyle habits and environmental influences, is pivotal in formulating holistic prevention and intervention strategies. As underscored by global incidence and mortality patterns, oral cancer necessitates a concerted, interdisciplinary approach to mitigate its impact and enhance outcomes for those affected [1].

Oral cancer primarily manifests in several types, each originating from different cells within the oral cavity [2]:

- **Squamous Cell Carcinoma:** The most common type, accounting for over 90% of all oral cancers. It originates in the squamous cells that line the

mouth and throat.

- **Verrucous Carcinoma:** A subtype of squamous cell carcinoma, known for its slow growth and warty appearance. It rarely spreads to distant sites.
- **Minor Salivary Gland Carcinomas:** These cancers originate in the salivary glands and can vary in their behavior and prognosis.
- **Lymphomas:** Cancers that start in the lymphoid tissue found in the tonsils and base of the tongue.
- **Melanomas:** Though rare, melanomas can occur in the oral cavity and are known for their aggressive nature .

1.1 Challenges in Oral Cancer Classifications

The task of accurately classifying oral cancer presents numerous challenges. The inherent complexity and heterogeneity of tumor tissues, variability in imaging modalities and protocols, and the presence of noise and artifacts in medical images all contribute to the difficulty of this task. Traditional methods that rely on handcrafted features and heuristic algorithms often fail to capture the nuanced characteristics of tumors and may yield sub-optimal results, particularly when faced with variability and noise in medical images. Deep learning-based approaches, however, offer a promising alternative by leveraging large-scale annotated datasets and powerful neural network architectures to automatically learn discriminative features directly from the data.

1.2 Motivation and Contributions

The motivation for this project stems from the potential of deep learning to revolutionize oral cancer diagnostics by addressing these challenges. Deep learning models have demonstrated remarkable success in image analysis tasks across various domains, including medical imaging. By leveraging advanced transfer learning models and ensemble approaches, we aim to develop a robust framework for accurate and reliable oral cancer classification.

The key contributions of this project are:

- **Usage of Transfer Learning Models:** Utilization of cutting-edge transfer learning models, including MobileNetV2, DenseNet121, EfficientNetB3, and NASNetMobile, to leverage pre-trained knowledge for improved accuracy

and efficiency. These models have demonstrated superior performance in various image classification tasks.

- **Ensemble Approach:** Implementation of a fuzzy rank-based ensemble approach for microscopic image classification. By combining multiple classifiers, we aim to enhance decision-making confidence and accuracy in complex histological image analysis.
- **Hybrid Architecture:** Design of a hybrid architecture that integrates MobileNetV2 and DenseNet121 with a Convolutional Block Attention Module (CBAM) for macroscopic image classification. This combination aims to refine feature extraction and capture subtle details in clinical images for improved diagnostic accuracy.

By integrating state-of-the-art transfer learning models and advanced techniques such as fuzzy ranking and attention modules, this project seeks to address the challenges posed by tumor heterogeneity and variability, ultimately improving diagnostic accuracy and reliability.

1.3 Organization of the Project Work

The remainder of this project report is organized as follows:

- **Chapter 1:** This chapter offers an overview of the project, addressing the challenges in oral cancer classification, the study's motivation, and its contributions. It lays the groundwork for understanding the significance of advanced methodologies in enhancing diagnostic accuracy.
- **Chapter 2:** This chapter describes some of the past work done on oral cancer classification.
- **Chapter 3:** Microscopic and macroscopic datasets of oral cancer and their distributions are shown. Theoretical concepts of transfer learning algorithms, fuzzy rank, attention module are explained. An overview of the algorithms used for implementation is provided.
- **Chapter 4:** This chapter is about the experiments done for the project and the results of these experiments. The metrics used are explained and utilized models' performance are compared with related work.
- **Chapter 5:** This chapter concludes the project work and the future scope is discussed.

CHAPTER 2

Literature Survey

Alhazmi et al. [3] focuses on predicting oral cancer risk using an artificial neural network (ANN) model, which was trained on a dataset of 73 patients. The model achieved a sensitivity of 85.71% and an accuracy of 78.95%, indicating its potential as an effective screening tool for oral cancer diagnosis.

Bansal et al. [4] utilized deep transfer learning techniques and hybrid optimization for the early prediction and diagnosis of oral cancer. Their approach employed models such as ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet, along with hybrid optimizers (ADAM, SGD, RMSprop). The DenseNet model achieved an accuracy of 95.41% for cancerous images and 92.41% for non-cancerous images, showcasing the efficacy of combining deep learning with hybrid optimization.

Welikala et al. [5] conducted a study on the automated detection and classification of oral lesions using deep learning. The MeMoSA project, which combines annotations from clinicians worldwide, used ResNet-101 for image classification and Faster R-CNN for object detection. The study reported promising F1 scores of 87.07% for lesion identification and 78.30% for referral.

Aubreville et al. [6] developed a deep learning-based approach for classifying oral squamous cell carcinoma (OSCC) in confocal laser endomicroscopy (CLE) images. Their method, which outperformed textural feature-based techniques, achieved an AUC of 0.96 and a mean accuracy of 88.3%.

Huang et al. [7] introduced a multisampling tensor model with Matrix Product State (MPS) blocks for classifying oral cancer from MRI images. The model achieved an accuracy of 85.43%, sensitivity of 82.76%, specificity of 86.99%, and AUC of 89.50%, demonstrating its robustness.

Rahman et al. [8] proposed a transfer learning model using AlexNet to extract features from oral squamous cell carcinoma biopsy images. Their model achieved classification accuracies of 97.66% for training and 90.06% for testing.

Jeyaraj et al. [9] presented a framework using Deep Boltzmann Machine (DBM)

and SVM fusion for detecting and classifying oral cancer from hyperspectral images. The method achieved a classifier accuracy of 94.75%, benefiting early cancer identification, especially in socio-economically disadvantaged populations.

Lim et al. [10] introduced D'OraCa, a deep learning framework for classifying oral lesions using photographic images, enhanced by a novel mouth landmark detection model. The framework improved classification accuracy, achieving an F1 score of 61.68%.

Das et al. [11] developed a CNN-based approach for automated classification of OSCC into four classes. Their method achieved an accuracy of 97.5%, surpassing transfer learning techniques and potentially reducing pathologist workload.

Panigrahi et al. [12] proposed a novel approach using capsule networks for early-stage oral cancer diagnosis. The method, robust to rotation and affine transformations, achieved a sensitivity of 97.78%, specificity of 96.92%, and accuracy of 97.35%.

Huang et al. [13] introduced a deep learning methodology augmented by a meta-heuristic approach for accurate oral cancer diagnosis. Using preprocessing techniques and optimized network weight selection via ISSA, the method achieved precision of 92.66%, F1-score of 89.37%, sensitivity of 87.34%, and accuracy of 97.33%.

Xu et al. [14] compared 3DCNN and 2DCNN-based algorithms for early oral cancer diagnosis using CT images. Their results showed that 3DCNNs, which leverage dynamic characteristics of enhancement rate images, outperform 2DCNNs in discriminating oral cancer lesions, with improvements in AUC (0.796), sensitivity (0.818), specificity (0.739), and accuracy (0.754).

Working Methodologies

This section describes the methodologies employed for oral cancer classification using deep learning approaches. The methods include the use of various deep neural network architectures, ensemble techniques, and attention mechanisms to improve the accuracy and robustness of the classification models. Each method is selected and optimized based on its suitability for handling the complexities and nuances present in oral cancer images.

3.1 Deep Transfer Learning Algorithms

A few deep CNN backbone architectures are used as base learners, taken from the Keras Applications [15]. They include the MobileNetV2, DenseNet121, EfficientNetB3, and NASNetMobile. Other models are also tested. However, these backbone architectures provide better results in comparison to others.

3.1.1 MobileNetV2

MobileNetV2 [16] is a state-of-the-art transfer learning algorithm tailored for efficient and lightweight neural network architectures, developed by Google. It is specifically designed for mobile and embedded vision applications, where computational resources are limited. MobileNetV2 employs depthwise separable convolutions to reduce model size and computational complexity while preserving accuracy. This architecture achieves an impressive reduction in parameters by a factor of 8 compared to traditional convolutional networks, with only a marginal decrease in accuracy. MobileNetV2 boasts a model size of approximately 14 megabytes and contains around 3.47 million parameters. By leveraging pre-trained weights on large-scale datasets like ImageNet [17], MobileNetV2 enables transfer learning, allowing developers to fine-tune models for specific tasks with minimal data

and computational resources. For instance, when pre-trained on ImageNet, MobileNetV2 achieves a top-1 accuracy of 71.8% and a top-5 accuracy of 91.0%, outperforming previous state-of-the-art models in terms of efficiency. The use of inverted residual blocks with linear bottlenecks enhances feature representation and promotes information flow through the network, leading to improved performance. MobileNetV2's flexibility and scalability make it a popular choice for various computer vision tasks, including image classification, object detection, and semantic segmentation. Its efficient design and compatibility with popular deep learning frameworks ensure widespread adoption and seamless integration into diverse applications. Overall, MobileNetV2 represents a significant advancement in deep learning research, enabling powerful neural networks to run efficiently on resource-constrained devices, with practical applications in real-world scenarios.

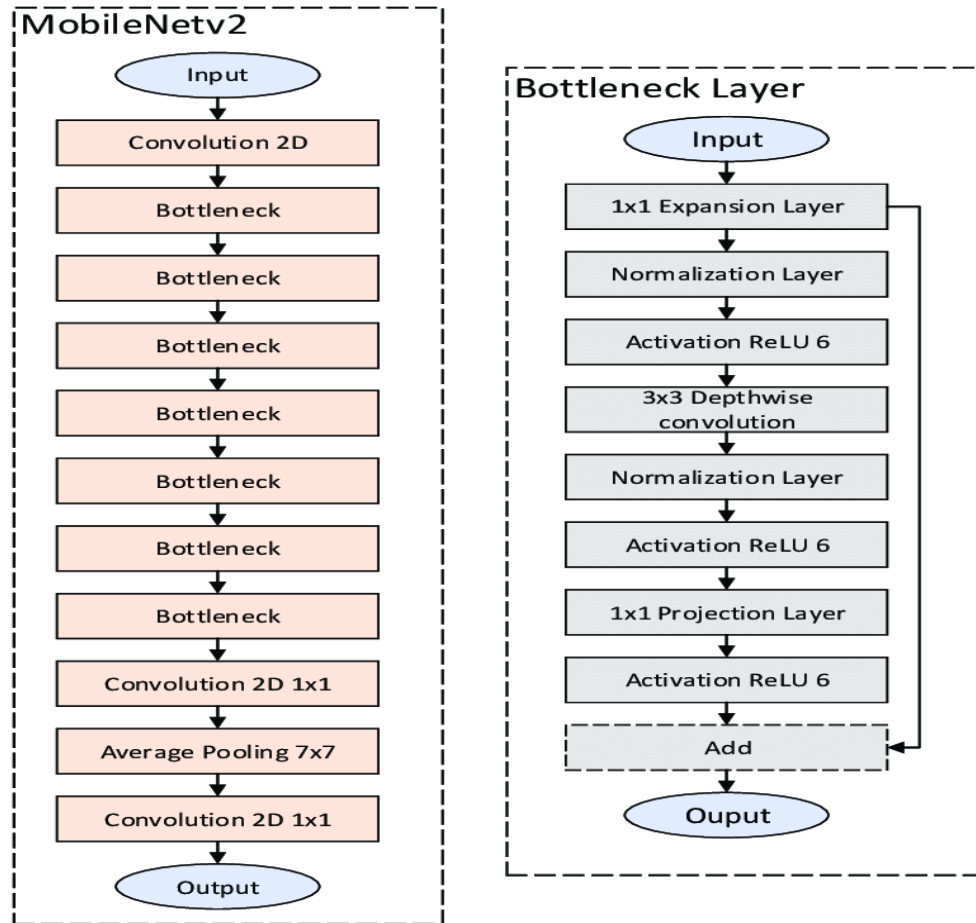


Figure 3.1: Architecture of MobileNetV2 [18]

3.1.2 DenseNet121

DenseNet-121 [19] is a convolutional neural network architecture renowned for its dense connectivity pattern, proposed by Huang et al. DenseNet architectures

are characterized by dense connections between layers, where each layer receives direct input from all preceding layers. This connectivity scheme facilitates feature reuse and enhances gradient flow throughout the network, mitigating the vanishing gradient problem and promoting feature propagation. DenseNet-121 consists of four dense blocks, each comprising a series of densely connected convolutional layers, interspersed with transition layers to reduce spatial dimensions and channel depth. This architecture achieves remarkable parameter efficiency by significantly reducing the number of parameters compared to traditional convolutional networks while maintaining competitive performance. DenseNet-121 has a model size of approximately 33.3 megabytes and contains around 7.98 million parameters. When pre-trained on ImageNet, DenseNet-121 achieves a top-1 accuracy of 74.91% and a top-5 accuracy of 92.74%. Pre-trained versions of DenseNet-121 on large-scale datasets such as ImageNet serve as effective feature extractors for transfer learning tasks, offering state-of-the-art performance across various computer vision tasks, including image classification, object detection, and semantic segmentation. Its efficient design and superior performance make DenseNet-121 a popular choice for both research and practical applications in the field of deep learning.

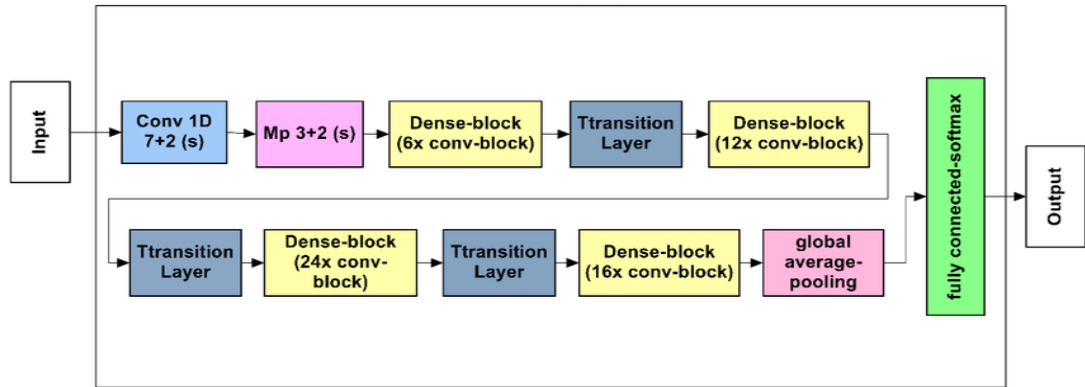


Figure 3.2: Architecture of DenseNet121 [20]

3.1.3 EfficientNetB3

EfficientNetB3 [21] is a convolutional neural network architecture developed by Tan and Le, characterized by its efficiency and scalability. It belongs to the EfficientNet family, which employs a compound scaling method to balance model size, computational cost, and performance. EfficientNetB3 achieves this balance by systematically scaling the network's depth, width, and resolution in a principled manner. With a model size of approximately 48 megabytes and around 12 million parameters, EfficientNetB3 strikes a favorable balance between efficiency and performance. When pre-trained on ImageNet, EfficientNetB3 achieves a top-1 accuracy of 81.7% and a top-5 accuracy of 95.7%. These impressive accuracy

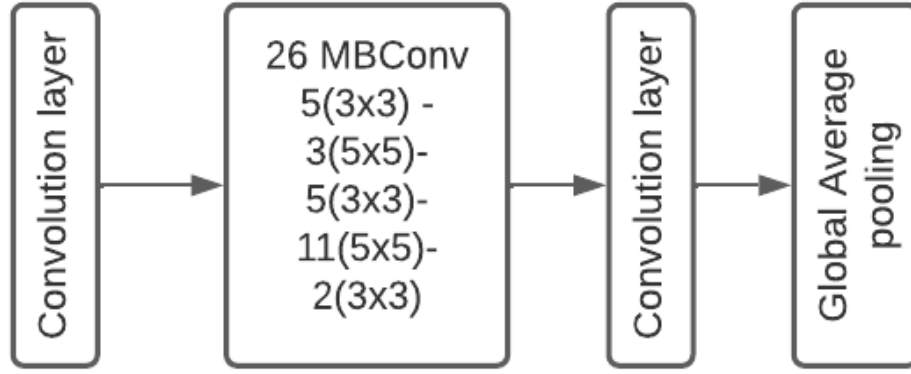


Figure 3.3: Architecture of EfficientNetB3

metrics, coupled with its efficient design, make EfficientNetB3 well-suited for various computer vision tasks, including image classification, object detection, and semantic segmentation. Its scalability allows for easy adaptation to different resource constraints, making it a versatile choice for both research and practical applications in deep learning.

3.1.4 NASNetMobile

NASNetMobile [22] is a convolutional neural network architecture designed for mobile and embedded devices, boasting a compact model size of approximately 23 megabytes. With around 5.3 million parameters, it offers a balance between

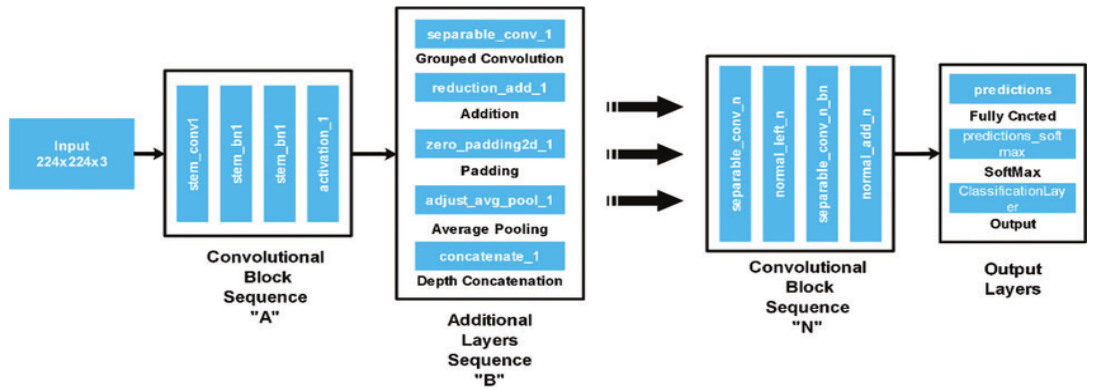


Figure 3.4: Architecture of NasNetMobile [23]

efficiency and accuracy. Pre-trained on ImageNet, NASNetMobile achieves a top-1 accuracy of 74.0% and a top-5 accuracy of 91.6%. Developed through neural architecture search (NAS), NASNetMobile's design combines separable convolutions and normal convolutions for efficient processing of input images. Despite its lightweight nature, NASNetMobile delivers competitive performance, making it suitable for various applications such as image recognition and object detec-

tion on mobile platforms. Its efficient design and impressive accuracy metrics have made it a popular choice among developers looking to deploy deep learning models on resource-constrained devices.

3.2 Gompertz-based Fuzzy Rank Ensembling

The Gompertz-based fuzzy rank ensembling algorithm is a novel approach designed to leverage the strengths of multiple deep learning models. By employing a fuzzy ranking mechanism, the algorithm effectively combines the outputs of individual models, thereby enhancing the overall performance and robustness of the classification system. This method addresses the inherent variability and complexity in oral cancer images by considering the relative confidence of each model’s predictions, leading to a more reliable and accurate ensemble decision.

3.2.1 Generation of confidence score

In this framework, at first, three transfer learning-based CNN models, MobileNetV2 (subsection 3.1.1), EfficientNetB3 (subsection 3.1.3), and NASNetMobile (subsection 3.1.4). are utilized to generate the confidence scores on the sample images. The **Adaptive Moment Estimation (ADAM)** optimizer, along with **Rectified Linear Unit (ReLU)** activation functions are used to fine-tune the networks for 50 epochs each on top of ImageNet weights. The three CNN models are described in brief in the **section 3.1**.

3.2.2 Method-1: Fuzzy-ranking based ensemble

The decision to utilize a fuzzy rank-based approach stems from its unique ability to prioritize each classifier’s confidence in its predictions for individual test cases, offering flexibility beyond traditional ensemble methods. Unlike approaches such as the average rule or weighted average rule, which require fixed weights, our method dynamically adjusts based on classifier confidence. Employing the re-parameterized Gompertz function [24], we generate fuzzy ranks for each CNN classifier tasked with classifying Oral squamous cell carcinoma (OSCC) from histopathological images. Specifically, our ensemble integrates predictions from three CNN classifiers: MobileNetV2 (subsection 3.1.1), EfficientNetB3 (subsection 3.1.3), and NASNetMobile (subsection 3.1.4).

The Gompertz model in classifying Oral Squamous Cell Carcinoma (OSCC) from histopathological images indicates an increase in classification accuracy with the progression of the model. As the model learns more features and patterns

associated with OSCC, its ability to correctly classify images improves, resembling the increase in mortality rate with age in the biological Gompertz model. The exponential term in the Gompertz function signifies the enhanced likelihood of correctly identifying OSCC as the classifier gains more confidence in its predictions. This heightened confidence score for OSCC classification is represented by the independent variable 'x' in the re-parameterized Gompertz function, as illustrated in Figure 3.5.

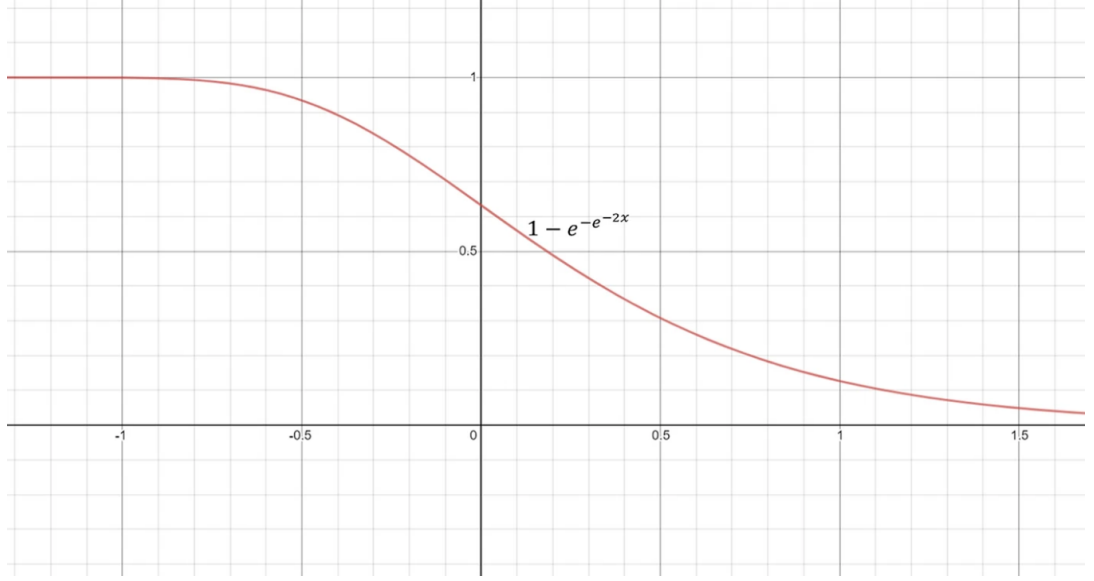


Figure 3.5: The re-parameterized Gompertz function

Let there be M number of decision scores (confidence factors of classifiers) $CF^{(1)}, CF^{(2)}, \dots, CF^{(M)}$ for each image $\mathbf{X}$. In our case, $M = 3$, since we have used three CNN models to generate the confidence scores on the datasets. The decision scores are normalized which follow Eq. (3.1), where C is the number of classes in the dataset.

$$\sum_{c=1}^C CF_c^{(i)} = 1.0; \forall i, i = 1, 2, 3, \dots, M \quad (3.1)$$

Corresponding to all samples belonging to different classes in the dataset, the confidence scores are used to generate the fuzzy ranks. The fuzzy rank for a class c using the i^{th} classifier's confidence scores is generated by the Gompertz function as in Eq. (3.2).

$$R_c^{(i)} = \left(1 - \exp \left[-\exp \left[-2.0 \times CF_c^{(i)} \right] \right] \right), \forall i, c; i = 1, 2, \dots, M; c = 1, 2, \dots, C \quad (3.2)$$

The value of $R_c^{(i)}$ lies in the range $[0.127, 0.632]$ where the smallest value 0.127 is analogous to rank 1 (best rank), i.e., a higher confidence gives a lower (better) value of rank. Now, if $K^{(i)}$ represents the top k ranks, i.e. ranks $1, 2, \dots, k$, corre-

sponding to class c , the fuzzy rank sum (FRS_c) and the complement of confidence factor sum ($CCFS_c$) are calculated as in Eqs. (3.3) and (4), respectively.

$$FRS_c = \sum_{i=1}^M \begin{cases} R_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^R, & \text{otherwise} \end{cases} \quad (3.3)$$

$$CCFS_c = \frac{1}{M} \sum_{i=1}^M \begin{cases} CF_c^{(i)}, & \text{if } R_c^{(i)} \in K^{(i)} \\ P_c^{CF}, & \text{otherwise} \end{cases} \quad (3.4)$$

P_c^R and P_c^{CF} are the penalty values imposed on class c , if it does not belong to the top k class ranks. The value of P_c^R is 0.632, which is calculated by putting $CF_c^{(i)} = 0$ in Eq. (3.2), and the value of P_c^{CF} is set to 0.0. The penalty values ensure that class c does not become an unlikely winner.

The final decision score is realized by the product of FRS_c and $CCFS_c$ which is used to generate the final predictions of the ensemble model. The final decision score (FDS) is calculated as in Eq. (3.5).

$$FDS_c = FRS_c \times CCFS_c \quad (3.5)$$

The final predicted class of instance $\mathbf{X}$ of the dataset is calculated by finding the class having the minimum FDS value and is given in Eq. (3.6).

$$class(\mathbf{X}) = \arg \min_{c=1,2,\dots,C} \{FDS_c\} \quad (3.6)$$

The computational complexity of the ensemble approach is $O(n)$ where ‘ n ’ is the number of classes in the dataset.

3.3 Fusion of Attention-aided Features

Integrating attention mechanisms into deep learning models for oral cancer classification can significantly enhance their performance. By focusing on the most relevant parts of the image, attention modules help in better capturing the intricate patterns associated with cancerous regions. This targeted approach allows the model to effectively distinguish between malignant and benign tissues, thereby improving classification accuracy. In the following subsections, we delve into specific attention mechanisms and methods used in this study to augment the deep learning models for oral cancer classification.

3.3.1 Convolutional Block Attention Module (CBAM)

Woo et al. [25] proposed the idea of a convolutional block attention module. Figure 3.6 demonstrates the structure of CBAM.

Convolutional block attention module (CBAM) enables the attention module to be applied to both the channel dimension and the spatial dimension. We will elaborate how it works from aspects of channel attention module and spatial attention module.

Firstly, the preceding 3D CNN transferred the input data X into the feature map $F \in R^{C \times H \times W}$ (i.e., C represents the number of channels, H represents the height, and W the width, of the feature map) before it entered the CBAM. F was then processed by the channel attention module and spatial attention module as follows:

$$F' = M_c(F) \otimes F,$$

$$F'' = M_s(F') \otimes F'$$

Where $\otimes$ represents the element-wise multiplication, F' represents the result of the feature map multiplying the channel attention map, and F'' represents the result of the spatial attention map multiplying F' or the final output.

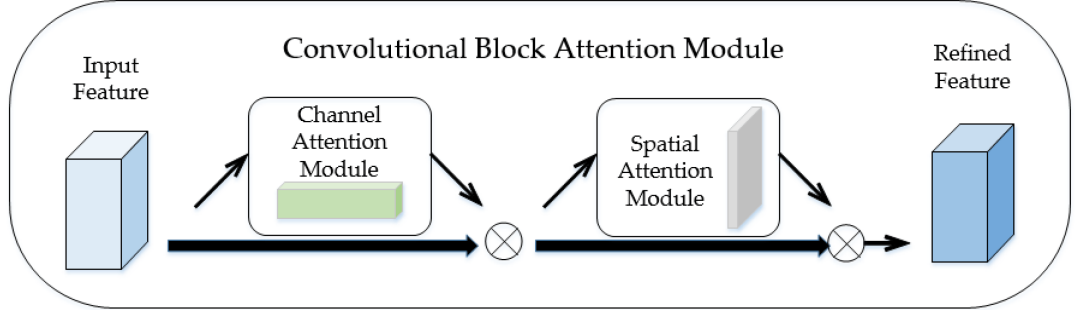


Figure 3.6: Convolutional Block Attention Module (CBAM) structure [25]

(i) Channel Attention Module : As each channel in the feature map represents one specific detector, the attention sector helped to extract the channels that contained the useful information. To complete the feature extraction and reduce data lost, the channel attention module squeezed the feature map in the spatial dimension by using both global average pooling layer and global max pooling layer. **Figure 3.7** shows the channel attention module. The global average pooling layer obtained the overall information, whereas the global max pooling layer captured information regarding differences of the feature. The combination of the two layers worked better than any single one.

Afterwards, the squeezed feature maps F_{avg}^c and F_{max}^c were sent to the shared network which was constituted by the multi-layer perceptron (MLP) with one

hidden layer. MLP was set at certain compression ratio to reduce the parameter and computation. The sigmoid function then computed the channel attention map of $M_c(F) \in R^{C \times 1 \times 1}$, the process of which was as follows:

$$\begin{aligned} M_c(F) &= \sigma(MLP(Avgpool(F)) + MLP(Maxpool(F))), \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))). \end{aligned}$$

F_{avg}^c and F_{max}^c represent the squeezed feature maps of the two pooling layers, with σ demonstrating the sigmoid function.

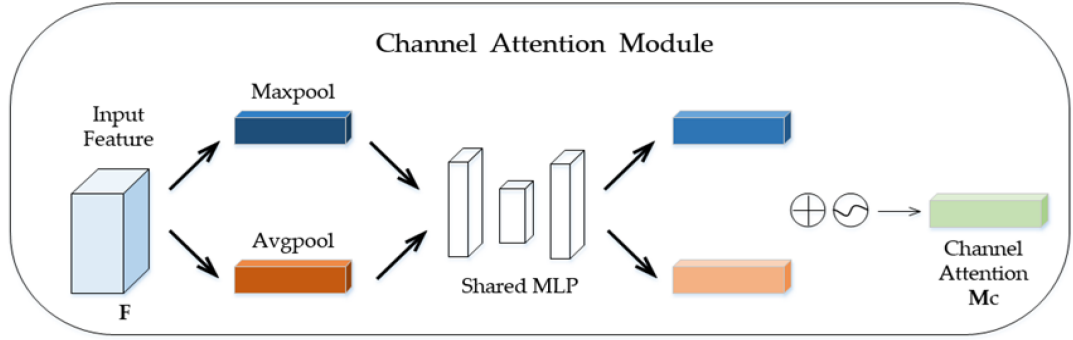


Figure 3.7: Channel attention module [25]

(i) **Spatial Attention Module** : As illustrated in **Figure 3.8**, the spatial attention module laid more emphasis on the parts of the feature map, which attained better reaction, in comparison with the channel attention module. For the feature maps generated in the channel attention module, the global average pooling layer and the global max pooling layer squeezed into two 2D feature maps: F_{avg}^s and F_{max}^s along the channel dimension, in order to highlight the regions containing key information.

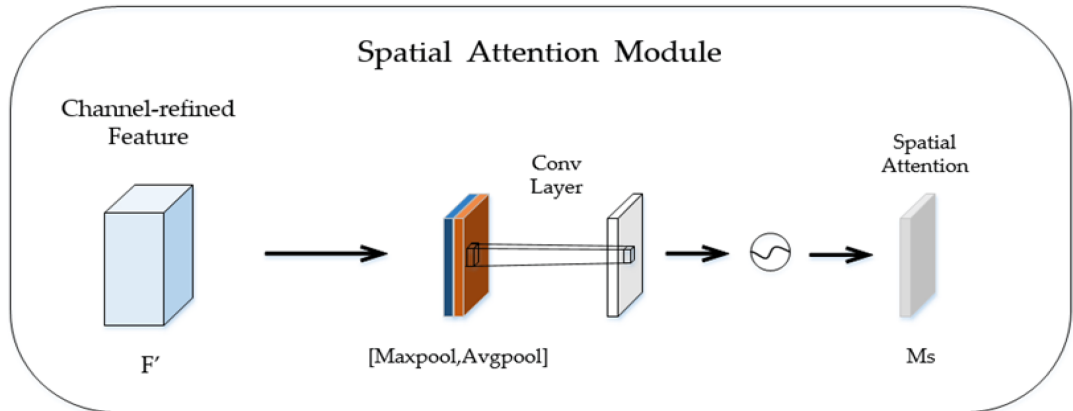


Figure 3.8: Spatial attention module [25]

The two 2D feature maps were concatenated to generate the effective feature map for convolution. After the convolution operation, the sigmoid function com-

puted the spatial attention map: $M_s(F) \in R^{1 \times H \times W}$ as follows:

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])), \\ &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])). \end{aligned}$$

where F_{avg}^s and F_{max}^s are the feature maps squeezed in the channel dimension, and σ represents the sigmoid function.

CBAM introduced the attention mechanism to the CNN, making the channel module and the spatial module cooperate. The two modules learned the key information in both the channel dimension and the spatial dimension and redistributed the weight of the features in an adaptive way.

3.3.2 Method-2

This method method for oral cancer classification within a macroscopic image dataset employs a custom hybrid deep learning architecture. This architecture leverages the complementary strengths of two powerful convolutional neural network backbones: MobileNetv2 (subsection 3.1.1) and DenseNet121 (subsection 3.1.2). MobileNetv2 is selected for its lightweight design and computational efficiency, making it well-suited for deployment in mobile or resource-limited environments. Its depthwise separable convolutions significantly reduce parameter count and inference time. DenseNet121 is incorporated due to its densely connected layers, promoting feature reuse and alleviating potential issues with vanishing gradients, thus enhancing the representational capacity of the model. The architecture of the model is illustrated in Figure 3.9.

To further refine the extracted features, a Convolutional Block Attention Module (CBAM) is integrated as a layer on top of each backbone model. CBAM applies a series of channel and spatial attention mechanisms, allowing the model to dynamically emphasize the most informative channels and regions within the input images. This focus on discriminative patterns is expected to boost classification performance.

The outputs from each CBAM module are then concatenated, effectively combining the refined feature maps generated by both backbones. Global Average Pooling (GAP) is subsequently applied, significantly reducing the spatial dimensions of the concatenated features while encouraging the model to focus on structural information rather than minor spatial variations. This step aims to reduce complexity and increase robustness against overfitting.

Finally, a series of fully-connected (dense) layers process the pooled features. These include two hidden layers with 512 and 64 units respectively, both utilizing ReLU activation, and a final output layer with a single unit and sigmoid activation,

producing a probability score for the binary oral cancer classification task. The model is compiled using the Adam optimizer, well-known for its adaptability, and binary cross-entropy loss, a standard choice for binary classification problems.

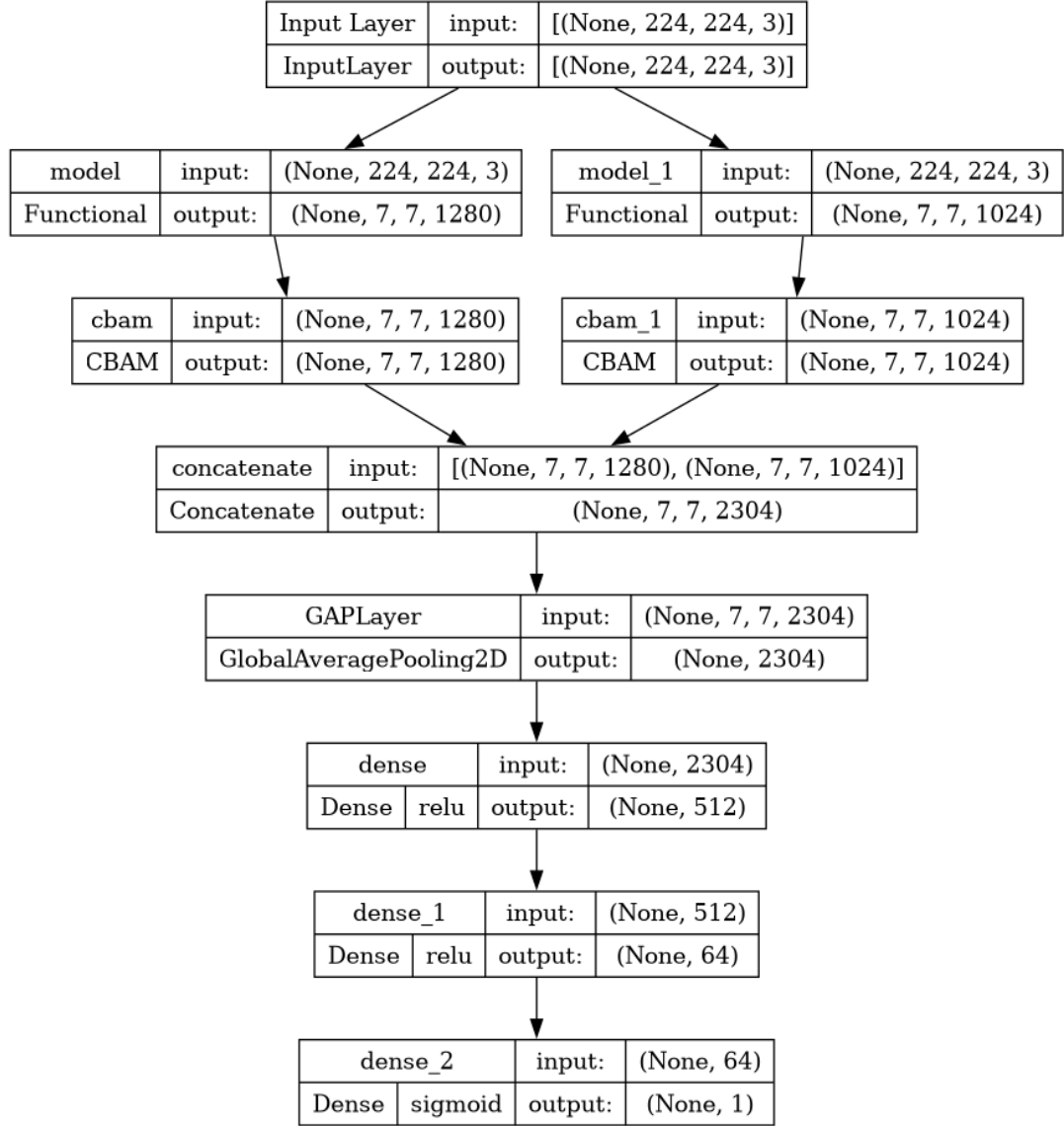


Figure 3.9: Architecture of the model

Results and Discussion

This chapter presents the experimental results and provides a comprehensive analysis of the performance of the proposed methods for oral cancer classification. The results are discussed in the context of the datasets used, the evaluation metrics applied, and the comparisons with existing state-of-the-art methods. Through detailed examination and interpretation of the findings, insights into the effectiveness and limitations of the approaches are highlighted, paving the way for future improvements and applications in the field.

4.1 Datasets

The datasets used in this study play a crucial role in the development and evaluation of the proposed deep learning models for oral cancer classification. Accurate and representative datasets are essential for training robust models that can generalize well to unseen data. In this project, two distinct types of datasets were utilized: microscopic and macroscopic datasets.

4.1.1 Overview of the Datasets

In our study, we utilize two distinct datasets comprising both microscopic and macroscopic data to comprehensively address the classification of oral cancer lesions. The microscopic dataset encompasses high-resolution images captured through microscopy, providing detailed insights into cellular morphology, tissue structure, and pathological features at the microscopic level. This dataset offers invaluable information for training convolutional neural networks (CNNs) to recognize subtle patterns and textures indicative of oral cancer.

Conversely, the macroscopic dataset consists of images acquired through macroscopic imaging techniques, such as digital photography or medical imaging devices, offering a broader perspective of oral lesions within the context of the surrounding

tissue and anatomical landmarks. These images provide a holistic view of the lesion’s appearance, including size, shape, color, and spatial relationships, which are crucial for clinical diagnosis and classification.

By incorporating both microscopic and macroscopic data, our study aims to capture complementary aspects of oral cancer pathology, leveraging the strengths of each dataset to develop a robust classification framework. This approach enables us to harness the rich information available at different scales, from cellular-level abnormalities to gross morphological characteristics, enhancing the accuracy and generalizability of our classification model.

4.1.2 Data Sources and Distribution

The datasets used in this study are essential for training and evaluating the deep learning models for oral cancer classification. These data sets include various types of oral cancer images, which have been downloaded from publicly available websites.

4.1.2.1 Microscopic Dataset

The microscopic dataset used in this study comprises high-resolution images of oral cancer lesions obtained through microscopy. The dataset is available at <https://www.kaggle.com/datasets/ashenafifasilkebede/dataset>, consisting of 4946 training images, 126 test images, and 123 validation images.

Table 4.1: Microscopic Dataset distribution

No. of Images	Training Set	Testing Set	Validation Set
Normal Images	2435	31	28
OSCC Images	2511	95	92
Total Images	4946	126	123

4.1.2.2 Macroscopic Dataset

The macroscopic dataset utilized in this study consists of images of oral cancer lesions captured through macroscopic imaging techniques. The dataset can be accessed at <https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images>, containing 87 cancerous and 44 non-cancerous images. Comprising of 87 cancerous and 44 non-cancerous images, a small dataset size posed a significant challenge. To address this limitation, a 75:25 split was applied, augmenting

Table 4.2: Microscopic Dataset distribution

No. of Images	Training Set	Testing Set
cancer	2273	22
non-cancer	2255	11
Total Images	4528	33

the training set to increase its sample count. Post-augmentation, the training set expanded to 4528 images, with 2273 classified as cancerous and 2255 as non-cancerous, while the test dataset remained at 33 images, comprising 22 cancerous and 11 non-cancerous images.

4.2 Performance Evaluation Measures

In this section, we'll explore the different ways we measure how well our classification models work when it comes to spotting oral cancer. These metrics help us understand how accurate and reliable our models are. They give us important clues about how our models perform with different types of data and situations.

(i) Confusion Matrix :

The confusion matrix, is a tabular representation that organizes the predictions

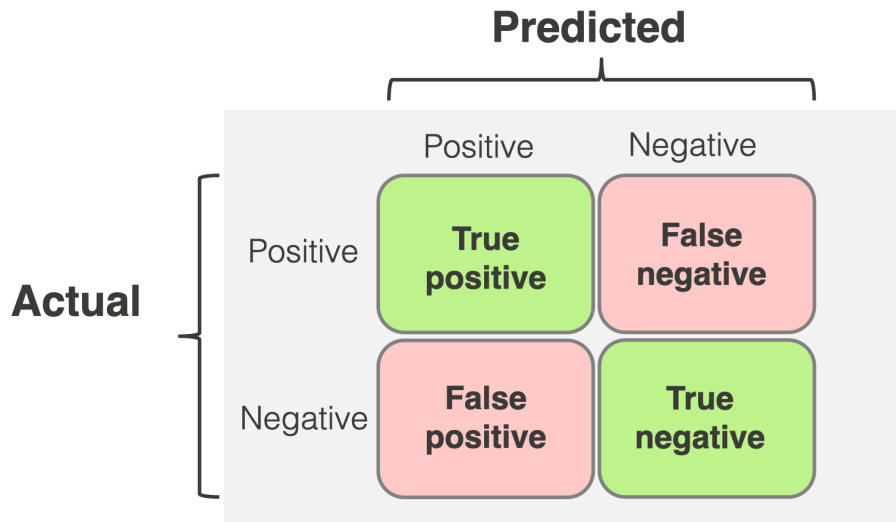


Figure 4.1: Confusion Matrix

made by a classification model as shown in Figure 4.1. It consists of four main components:

- **True Positives (TP):** Instances that are correctly predicted as positive.

- **False Positives (FP):** Instances that are incorrectly predicted as positive when they are actually negative.
- **False Negatives (FN):** Instances that are incorrectly predicted as negative when they are actually positive.
- **True Negatives (TN):** Instances that are correctly predicted as negative.

The classification matrix provides a detailed breakdown of the model's performance and serves as the foundation for computing various evaluation metrics.

(ii) Accuracy :

Accuracy measures the overall correctness of the model's predictions by calculating the proportion of correctly classified instances out of the total number of instances in the dataset. It is expressed as a percentage and can be computed using the formula (4.1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (4.1)$$

(ii) Precision :

Precision quantifies the model's ability to correctly classify positive instances among all instances predicted as positive. It focuses on minimizing false positive predictions and is calculated as the ratio of true positive predictions to the sum of true positive and false positive predictions as shown in the Eq. (4.2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.2)$$

(iii) Recall :

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that are correctly classified by the model. It emphasizes minimizing false negative predictions and is calculated as the ratio of true positive predictions to the sum of true positive and false negative predictions. The formula to calculate the Recall is shown in Eq. (4.3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.3)$$

(iv) F1-score :

The F1-score is a metric that balances precision and recall, providing a single value to assess the model's performance. It is the harmonic mean of precision and recall and is particularly useful when the class distribution is imbalanced. Eq. (4.4)

shows the formula to calculate f1-score.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.4)$$

(v) ROC Curve and AUC :

The Receiver Operating Characteristic (ROC) curve is a plot of the true positive rate (TPR), also known as sensitivity or recall, against the false positive rate (FPR) across various threshold values. Mathematically, TPR is defined as:

$$\text{TPR} = \frac{TP}{TP + FN} \quad (4.5)$$

and the FPR is given by:

$$\text{FPR} = \frac{FP}{FP + TN} \quad (4.6)$$

where TP represents true positives, FN denotes false negatives, FP indicates false positives, and TN signifies true negatives. The ROC curve illustrates the trade-off between sensitivity and specificity, providing insights into the classifier's performance across different threshold levels.

The Area Under the ROC Curve (AUC) quantifies the overall performance of the classifier. It represents the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. AUC ranges from 0 to 1, where a value closer to 1 indicates better discrimination ability. Mathematically, AUC is calculated by integrating the ROC curve:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d\text{FPR} \quad (4.7)$$

The ROC curve and AUC are essential evaluation metrics for binary classifiers, offering a visual representation of their performance and a quantitative measure of their discrimination ability. They assist in comparing different classifiers and optimizing threshold values for better classification performance.

4.3 Results

Testing is conducted on microscopic as well as macroscopic datasets in subsection 4.3.1 and subsection 4.3.2 to highlight the efficacy of the model.

4.3.1 Microscopic Dataset

In the Figure 4.2 the training and validation performance of three key models (MobileNetV2, EfficientNetB3, and NASNetMobile) utilized in my thesis is shown.

Each model's training accuracy and loss, as well as validation accuracy and loss, are plotted over 70 epochs.

The accuracy graphs illustrate the models' learning curves, showing how their

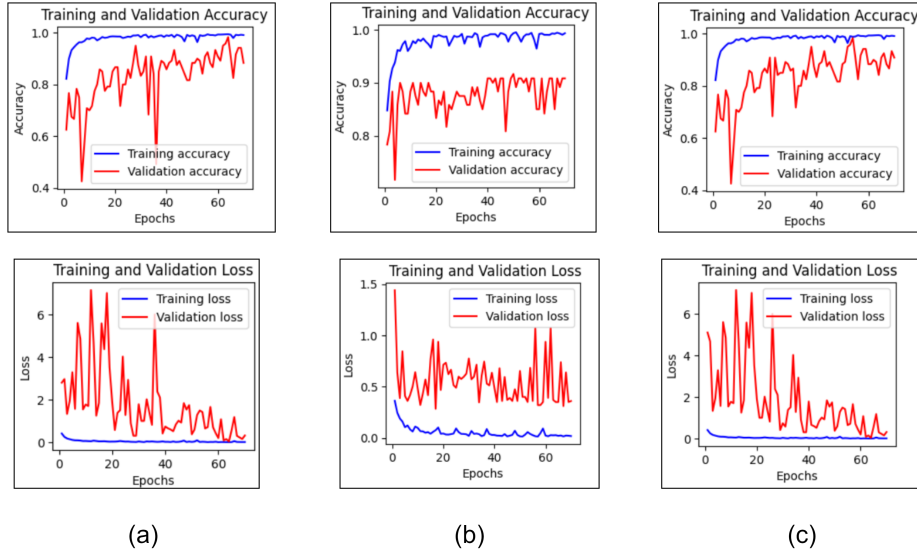


Figure 4.2: Epochs vs Accuracy and Epochs vs loss graph for (a) MobileNetV2, (b) EfficientNetB3 and (c) NASNetMobile

performance improves with training iterations. It's evident that all three models demonstrate a steady increase in accuracy over time, indicating effective learning and adaptation to the dataset. The validation accuracy plots confirm the generalization ability of the models, as they closely track the training accuracy, suggesting minimal overfitting.

Conversely, the loss graphs depict the models' convergence towards optimal parameter values. As training progresses, the loss steadily decreases, indicating that the models are effectively minimizing prediction errors and optimizing their parameters to better fit the data. The validation loss curves, while exhibiting fluctuations, generally follow the same decreasing trend as training loss, indicating good generalization performance.

In Figure 4.3, the confusion matrix and ROC curves are displayed for all four methods: MobileNet, EfficientNet, NASNet, and the ensemble approach. The confusion matrix for the ensemble method reveals that out of 126 test samples, 119 were correctly classified, resulting in an overall accuracy of approximately 94.44%. Specifically, 91 cases of oral squamous cell carcinoma (OSCC) were correctly identified, along with 28 normal cases. However, there were misclassifications, with 3 non-OSCC cases mistakenly classified as OSCC and 4 cases of OSCC misclassified as non-OSCC. These errors could arise from either complex aspects of the dataset

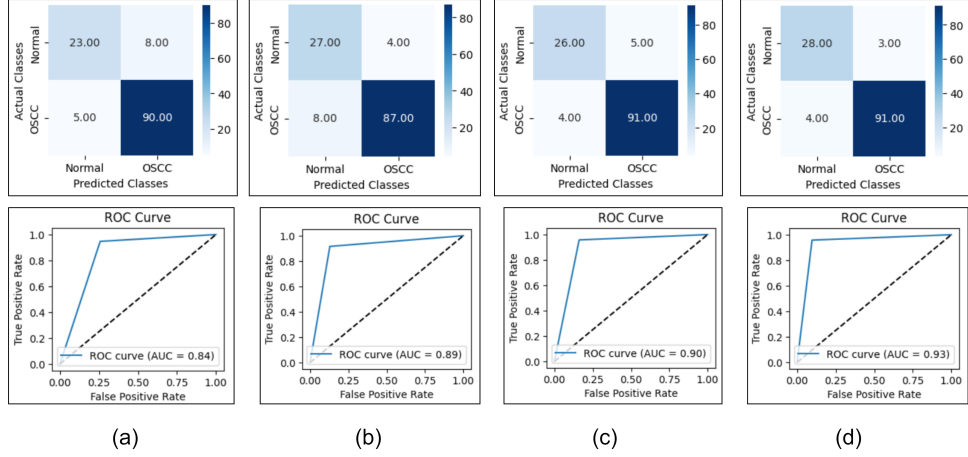


Figure 4.3: Confusion matrix and ROC curve of (a)MobileNetV2, (b)EfficientNetB3 and (c)NASNetMobile and (d) Ensemble of these three models

or limitations in the models' ability to detect subtle differences in oral cancer pathology. Additionally, the ROC curve for the ensemble method demonstrates an area under the curve (AUC) of 0.93, indicating high discriminatory power and performance in distinguishing between OSCC and non-OSCC cases. This metric further underscores the effectiveness of the ensemble approach in oral cancer classification.

In the Table 4.3, the performance metrics of three different backbone models are presented along with their ensemble. Each model is evaluated based on accuracy, precision, recall, F1-score, and area under the curve (AUC).

MobileNetV2 demonstrates an accuracy of 0.8968, with precision, recall, and F1-score values of 0.9184, 0.9474, and 0.9326, respectively. It achieves an AUC of 0.84, indicating its capability to distinguish between oral cancer and normal images. Similarly, EfficientNetB3 exhibits impressive performance, achieving an accuracy of 0.9048 and notable precision, recall, and F1-score values. Its AUC of 0.89 further confirms its discrimination ability.

NASNetMobile surpasses the performance of the previous models, with an ac-

Table 4.3: Results of Microscopic Dataset

Models	Accuracy	Precision	Recall	f1-score	AUC
MobileNetV2	0.8968	0.9184	0.9474	0.9326	0.84
<i>EfficientNetB3</i>	0.9048	0.9560	0.9157	0.9355	0.89
<i>NASNetMobile</i>	0.9286	0.9479	0.9579	0.9528	0.90
<i>Ensemble</i>	0.9444	0.9681	0.9579	0.9630	0.93

curacy of 0.9286 and consistently high precision, recall, and F1-score values. Its AUC of 0.90 indicates superior discriminatory power.

Notably, the ensemble of all three models yields the best results, with an accuracy of 0.9444, the highest among all individual and ensemble models. This ensemble approach combines the strengths of each model, resulting in enhanced performance. With precision, recall, and F1-score values of 0.9681, 0.9579, and 0.9630, respectively, and an AUC of 0.93, the ensemble achieves remarkable discrimination ability and classification accuracy.

Overall, the ensemble of MobileNetV2, EfficientNetB3, and NASNetMobile emerges as the most effective approach for classifying microscopic images of oral cancer, underscoring the significance of model aggregation in improving classification performance.

4.3.2 Macroscopic Dataset

The training process for the model demonstrated impressive learning capabilities in the challenging domain of oral cancer classification. Across 60 epochs of training, the test set served the dual purpose of validation. Notably, the model achieved a maximum validation accuracy of 0.9696, recurrently attained after the 20th epoch, although not consistently. This performance highlight suggests robust learning capabilities despite the challenges posed by the dataset's size. A learning rate of 0.001 was adopted to facilitate model convergence and optimization throughout the training process. The accuracy and loss graphs in Figure 4.4 provide a jagged pattern, signifying the dynamic nature of the learning process and the complexity levels in oral cancer classification.

The confusion matrix in Figure 4.5 reveals the model's performance in classi-

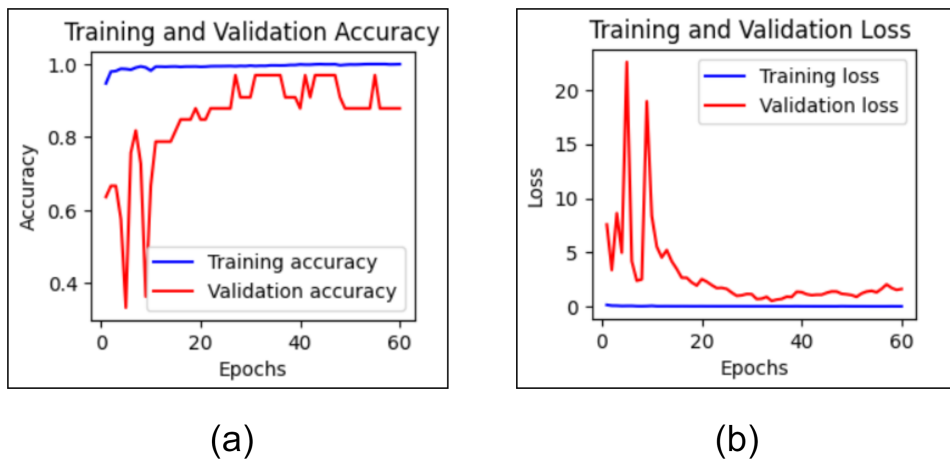


Figure 4.4: (a) Epochs vs Accuracy and (b) Epochs vs Loss graph for both training and validation set

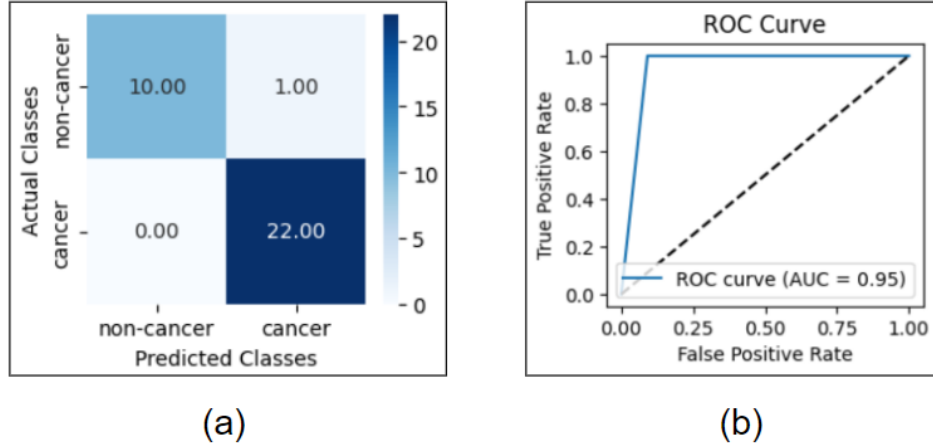


Figure 4.5: (a) Confusion matrix and (b) ROC curve of the model

fying non-cancerous and cancerous images. Of the 11 non-cancerous images, 10 were correctly classified, while only one was erroneously predicted as cancerous. Conversely, for the 22 cancerous images, all were accurately identified. This indicates high sensitivity in detecting cancerous samples while maintaining a low false positive rate for non-cancerous ones. The ROC curve, depicting the model's true positive rate against the false positive rate, exhibited strong discriminatory power, with an AUC of 0.95. This indicates that the model effectively distinguishes between cancerous and non-cancerous images across various threshold settings. The high AUC value underscores the model's robust performance and its ability to make reliable predictions, further validated by the minimal misclassification observed in the confusion matrix.

The results of the microscopic dataset, as shown in Table 4.4, highlight the per-

Table 4.4: Results of Microscopic Dataset

Models	Accuracy	Precision	Recall	f1-score	AUC
<i>Proposed Model</i>	0.9697	0.9565	1.0000	0.9778	0.95

formance of the proposed model across several key metrics. The model achieved an impressive accuracy of 96.97%, demonstrating its high capability in correctly classifying the samples. Precision, which measures the proportion of true positive identifications out of all positive identifications, was recorded at 95.65%. This indicates that the model makes very few false-positive errors. The recall, or sensitivity, of the model reached 100%, showing its effectiveness in identifying all true positive cases without missing any. Consequently, the F1-score, which balances precision and recall, was 97.78%, reflecting the model's overall robustness in per-

formance. Additionally, the Area Under the Curve (AUC) was reported to be 0.95, underscoring the model’s excellent ability to distinguish between classes. These results collectively suggest that the proposed model is highly reliable and effective for the classification tasks within the microscopic dataset.

4.4 Comparison with state-of-the-art methods

The proposed methods for oral cancer classification are evaluated against existing state-of-the-art techniques to benchmark their performance and effectiveness. This comparison highlights the strengths and areas for improvement of our models, providing a comprehensive understanding of their capabilities. By analyzing the results obtained from both the macroscopic and microscopic datasets, we can identify the key factors that contribute to the success of deep learning approaches in this domain. This evaluation not only demonstrates the robustness of our models but also sheds light on the potential for future enhancements in the field of oral cancer diagnosis using advanced machine learning techniques.

(i) Microscopic Dataset:

The table Table 4.3, which includes information on the experimental setups, techniques, and outcomes attained by several researchers, compares several strategies for classifying oral cancer using histopathology pictures. Rahman et al. implemented a transfer learning model based on AlexNet, which was modified to extract rank features, using the same dataset and experimental setting as the suggested method. Their model yielded 90.02% accuracy, 9.08% miss rate, 92.74% sensitivity, 87.38% specificity, and 90.15% F1 score. While there is still potential for improvement in terms of overall accuracy and specificity, this suggests a relatively high performance in recognizing photos of oral squamous cell carcinoma (OSCC). Using the same experimental setting, Bansal et al. also investigated a hybrid optimizer that combines ADAM, SDG, and RMSprop to optimize a variety of deep learning models, such as ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet. Their method produced remarkable results, with 95.41% accuracy for OSCC photos and 92.41% accuracy for normal images. This implies that their hybrid optimizer improves the model’s capacity to distinguish between photos that are malignant and those that are not.

The proposed method introduced a fuzzy rank-based ensemble of three transfer learning models: MobileNetV2, EfficientNetB3, and NASNetMobile. This ensemble approach aimed to leverage the strengths of each individual model while mitigating their weaknesses. The results were impressive, with an overall accuracy of 94.44%. Specifically, the method achieved an accuracy of 95.79% for OSCC images and 90.32% for normal images. This indicates a strong performance in de-

Table 4.5: Comparison with state-of-the-art methods for Histopathological dataset

Work	Method	Accuracy	Precision	Recall	F1-score
<i>Rahman et al.</i> [8]	Transfer learning model AlexNet, customized to extract rank features	90.02%	-	92.74%	90.15%
<i>Bansal et al.</i> [4]	ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet with hybrid optimizer (ADAM + SDG + RMSprop)	95.24%	-	-	-
Method-1	Fuzzy rank based ensemble of 3 transfer learning models MobileNetV2, EfficientNetB3 and NASNetMobile	94.44%	96.81%	95.79%	96.30%

tecting cancerous lesions, although there was a slight drop in accuracy for normal images compared to Bansal et al.'s method.

In summary, while each method demonstrates strong performance in classifying histopathological images of oral cancer, the proposed ensemble method shows a promising balance of high accuracy for both cancerous and non-cancerous images, indicating its potential for practical application in medical diagnostics. The table effectively showcases the comparative strengths and outcomes of these different approaches, emphasizing the advancements made in improving classification accuracy through innovative deep learning techniques.

(ii) Macroscopic Dataset:

Table 4.6 presents a comparative analysis of state-of-the-art methods for a real-time dataset, highlighting their experimental setups, methodologies, and performance metrics. The study by Bansal et al. utilized a split ratio of 75:25 and implemented a variety of deep learning models including ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet, optimized through a hybrid optimizer combining ADAM, SGD, and RMSprop. This approach achieved an accuracy of 92.41% with a root mean square error (RMSE) of 0.29, demonstrating a solid performance across multiple models.

Huang et al., on the other hand, applied a split ratio of 70:30 and employed a con-

volutional neural network (CNN) where the weights were optimally selected using an improved squirrel search algorithm (ISSA). Their method achieved a higher accuracy of 97.33% along with a precision of 92.66%, recall of 87.34%, and an F1 score of 89.37%, indicating a well-balanced performance with superior accuracy. The proposed method in this study used a split ratio of 75:25 and featured a novel

Table 4.6: Comparison with state-of-the-art methods for Real-time dataset

Work	Method	Accuracy	Precision	Recall	F1-score
<i>Bansal et al.</i> [4]	ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet with hybrid optimizer (ADAM + SGD + RMSprop)	92.41%	-	-	-
<i>Huang et al.</i> [13]	CNN weights are optimally selected by ISSA (an upgraded type of squirrel search algorithm)	97.33%	92.66%	87.34%	89.37%
Method-2	Features of MobileNetV2 and DenseNet121 pass through CBAM attention layer, then concatenated.	96.97%	95.65%	100%	97.78%

approach where the features extracted from MobileNetV2 and DenseNet121 were passed through a Convolutional Block Attention Module (CBAM) layer for refinement before being concatenated. This method achieved an impressive accuracy of 96.97%, with a precision of 95.65%, recall of 100%, and an F1 score of 97.78%. The high recall rate indicates that the proposed method was particularly effective at identifying all positive cases, minimizing false negatives.

Overall, the proposed method demonstrates a competitive performance, especially in terms of recall and F1 score, compared to the other state-of-the-art methods. The combination of MobileNetV2 and DenseNet121 with CBAM has proven to be highly effective in enhancing feature extraction and improving classification accuracy for the real-time dataset.

Conclusions

5.1 Conclusions of the Present Work

Healthcare is one of the foremost domains in the current scenario which was needed to focus and development into the sector is the leading task. Out of these diseases, Cancer is one of the major diseases which are affecting human society rapidly. Oral cancer is a sensitive disease and it needs to be prevented and cared for by early diagnosis. In this context, this report presents a robust deep learning approach for the classification of oral cancer using both microscopic and macroscopic datasets. The methods developed have demonstrated significant potential in accurately distinguishing cancerous from non-cancerous lesions.

For the microscopic dataset, the proposed approach effectively utilized transfer learning models and an innovative ensemble technique to achieve high classification accuracy, sensitivity, and specificity. This method capitalized on the strengths of individual models, resulting in enhanced performance and robustness. In the macroscopic dataset, a hybrid deep learning architecture was employed, demonstrating the capability to handle broader perspective images of oral lesions. Despite the challenges posed by the small dataset size, data augmentation techniques significantly contributed to the model's performance, achieving high accuracy, precision, recall, and F1 score.

Overall, the results from both datasets validate the effectiveness of the proposed deep learning methods in early detection and classification of oral cancer. These findings underscore the potential for these advanced techniques to be applied in clinical settings, aiding in timely and accurate diagnosis. Future research could explore additional data sources and further refinement of augmentation and ensemble strategies to enhance model performance and applicability.

5.2 Scope for Future Works

The future of using deep learning for oral cancer diagnosis is very promising. We can explore new ways to make our models even better. One idea is to combine different kinds of information, like genetic data or a patient's medical history, to improve how accurately our models work. We also need to make sure our models are easy for doctors to understand, so they trust them and use them to help patients. To make sure our models work well for everyone, we need to test them on lots of different kinds of patients. This means working closely with doctors and researchers to get more data and make our models more reliable.

We can also keep improving our deep learning techniques to make our models faster and smarter. And by using technology like apps and online tools, we can make it easier for people to get screened for oral cancer, no matter where they live.

Bibliography

- [1] Cancer.org, *Oral cancer*, <https://www.cancer.org/cancer/oral-cavity-and-oropharyngeal-cancer/about/what-is-oral-cavity-cancer.html>.
- [2] P. Medicine, *Types of lip and oral (mouth) cancer*, <https://www.pennmedicine.org/cancer/types-of-cancer/mouth-cancer/types-of-mouth-cancer>.
- [3] A. Alhazmi, Y. Alhazmi, A. Makrami, *et al.*, “Application of artificial intelligence and machine learning for prediction of oral cancer risk,” *Journal of Oral Pathology & Medicine*, vol. 50, no. 5, pp. 444–450, 2021.
- [4] K. Bansal, R. Bathla, and Y. Kumar, “Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer,” *Soft Computing*, vol. 26, no. 21, pp. 11 153–11 184, 2022.
- [5] R. A. Welikala, P. Remagnino, J. H. Lim, *et al.*, “Automated detection and classification of oral lesions using deep learning for early detection of oral cancer,” *IEEE Access*, vol. 8, pp. 132 677–132 693, 2020.
- [6] M. Aubreville, C. Knipfer, N. Oetter, *et al.*, “Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning,” *Scientific reports*, vol. 7, no. 1, p. 11 979, 2017.
- [7] C. Huang, G. Zhang, S. Chen, and V. H. C. de Albuquerque, “An intelligent multisampling tensor model for oral cancer classification,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7853–7861, 2022.
- [8] A.-u. Rahman, A. Alqahtani, N. Aldhafferi, *et al.*, “Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning,” *Sensors*, vol. 22, no. 10, p. 3833, 2022.
- [9] P. R. Jeyaraj, B. K. Panigrahi, and E. R. Samuel Nadar, “Classifier feature fusion using deep learning model for non-invasive detection of oral cancer from hyperspectral image,” *IETE Journal of Research*, vol. 68, no. 6, pp. 4031–4042, 2022.

-
- [10] J. H. Lim, C. S. Tan, C. S. Chan, *et al.*, “D’oraca: Deep learning-based classification of oral lesions with mouth landmark guidance for early detection of oral cancer,” in *Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25*, Springer, 2021, pp. 408–422.
 - [11] N. Das, E. Hussain, and L. B. Mahanta, “Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network,” *Neural Networks*, vol. 128, pp. 47–60, 2020.
 - [12] S. Panigrahi, J. Das, and T. Swarnkar, “Capsule network based analysis of histopathological images of oral squamous cell carcinoma,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4546–4553, 2022.
 - [13] Q. Huang, H. Ding, and N. Razmjoooy, “Optimal deep learning neural network using issa for diagnosing the oral cancer,” *Biomedical Signal Processing and Control*, vol. 84, p. 104749, 2023.
 - [14] S. Xu, C. Liu, Y. Zong, *et al.*, “An early diagnosis of oral cancer based on three-dimensional convolutional neural networks,” *IEEE Access*, vol. 7, pp. 158603–158611, 2019.
 - [15] *Keras applications*, <https://keras.io/api/applications/>.
 - [16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
 - [17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.
 - [18] A. Tragoudaras, P. Stoikos, K. Fanaras, *et al.*, “Design space exploration of a sparse mobilenetv2 using high-level synthesis and sparse matrix techniques on fpgas,” *Sensors*, vol. 22, no. 12, p. 4318, 2022.
 - [19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
 - [20] I. Tareq, B. M. Elbagoury, S. El-Regaily, and E.-S. M. El-Horbaty, “Analysis of ton-iot, unw-nb15, and edge-iiot datasets using dl in cybersecurity for iot,” *Applied Sciences*, vol. 12, no. 19, p. 9572, 2022.

- [21] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
- [22] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.
- [23] F. A. Shah, M. A. Khan, M. Sharif, *et al.*, “A cascaded design of best features selection for fruit diseases recognition,” *Comput. Mater. Contin.*, vol. 70, no. 1, pp. 1491–1507, 2022.
- [24] K. M. Tjørve and E. Tjørve, “The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family,” *PloS one*, vol. 12, no. 6, e0178691, 2017.
- [25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.