

Oral Cancer Detection Using Convolutional Neural Network(CNN)

Lalan Kumar Yadav¹, Aashish Bishokarma², Kanimozhi S³

^{1,2} UG Scholar, Dept. of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore-641407, Tamil Nadu, India

³Assistant Professor, Dept. of Biomedical Engineering, KPR Institute Of Engineering and Technology, Coimbatore-641407, Tamil Nadu, India

Emails: hiilalan405@gmail.com¹, aashishbishokarma0@gmail.com², kanimozhi.s@kpriet.ac.in³

Abstract

Oral squamous cell carcinoma (OSCC), the most widespread type of oral cancer, necessitates prompt biopsy evaluation in order to increase patient survival rates. However, healthcare professionals frequently disagree on subtle cell changes when performing standard histological (H&E)-stained slide evaluations. To overcome this issue, we developed a convolutional neural network (CNN) classifier using EfficientNetB3 pre-trained weights to classify oral samples as either normal or cancerous with high reliability. We created a composite image collection by combining publicly available images into a 5,192-image dataset that was randomly divided into 70%/15%/15% (3,634 training/779 validation/779 test) images, cropped to 224×224 pixels, normalized, and augmented (flipped) to create a robust training set. Our classifier was built upon the core of EfficientNetB3, adding a batch normalization, and using Adamax optimization for 100 epochs. The final performance of the classifier on unseen test image data was 98.33% accuracy and a balanced precision/recall/F1 score of 98% for each class (as shown by the confusion matrix) indicating the classifier is ready for clinical use in providing consistent second opinion results to help extend the capacity of pathologists in heavily burdened hospitals in India.

Keywords: OSCC, histopathology slides, EfficientNetB3 CNN, transfer learning, medical image sorting, data augmentation, Adamax training.

1. Introduction

Oral cancer is one of the biggest health threats globally; it is particularly common in areas where people smoke and are diagnosed with oral cancer late [1] and [3]. When patients are diagnosed with oral malignant tumors, the most common form of diagnosis occurs in patients with oral squamous cell carcinoma (OSCC). If there is an ability to diagnose early enough, there are significantly higher chances of survival according to studies [5]. The most reliable method for diagnosing oral cancer is through the histopathological examination of biopsy samples. Pathologists evaluate the shape and structure of tissues and nuclei under a microscope to make a diagnosis. Even though this method of diagnosing oral cancers is very reliable, it requires specific skill sets, takes time, and may result in different interpretations of diagnosis among pathologists [7] and [31]. Over the past few years, artificial intelligence has dramatically changed how medical

imaging is interpreted. In particular, deep learning techniques (especially CNNs) are able to learn how to recognize meaningful visual features from raw image data without having to use manual techniques [9],[10]. CNNs have also successfully completed applications in medical fields, such as detecting skin lesions and screening for lung cancer, suggesting that there is great potential for them to be used in digital pathology [17], [18], and [22]. Limited availability of annotated medical data sets has led to the use of transfer learning to enhance model accuracy in several medical imaging applications. A network can be adapted to a particular clinical dilemma rather than training the model from scratch, using a pre-trained model. EfficientNet is an updated CNN architecture that improves predictive accuracy through coordinated depth, width, and resolution scaling. Utilising these features, this study creates a convolutional neural network based on

EfficientNetB3 to differentiate between normal and malignant oral histopathological images.

2. Related Work

There has been a gradual shift from manual diagnosis of oral cancer to the use of automated computer-based diagnostic systems through research into the detection of oral cancers. Historically, the diagnosis has been based on the evaluation of tissue samples by expert pathologists through microscopy this represents the best method for diagnosing cancer and is known as the “gold standard.” The drawbacks of this method include time considerations

2.1 Conventional Machine Learning Approaches

The first computational techniques for identifying oral cancer used hand-crafted visual descriptors to analyse microscopy images. Feature extraction from the images was often accomplished using texture, colour histograms, and structural features with classifiers. Although these technologies have shown moderate success in correctly classifying cancerous vs non-cancerous tissue, their reliance on hand-crafted feature extraction means that their performance was significantly impacted by how effectively their features were engineered. Given the complexity of the morphologies exhibited by cancerous tissues (and how different they can appear), there is no guarantee that hand-crafted visual descriptors would account for variability in the underlying morphology; thus, when presented with changes in the following conditions, the performance of these systems would commonly reduce: staining intensity, magnification and/or illumination [9].

2.2 Deep Learning in Medical Imaging

CNNs represented a paradigm shift in the way images were analysed, as they learned hierarchical abstractions from unprocessed input images rather than developing features manually [10], [11]. ResNet [12] and DenseNet [15] architectures allowed for the creation of much deeper networks while avoiding instability when training the network. Deep Learning models have also been applied to medical problems and have performed very well across a wide variety of diagnostic tasks, including: classifying skin lesions [17], detecting lung cancer [18], and analyzing histopathological tumours [22], [30]. They are also

capable of detecting abnormalities in soft tissue that may not be measurable by traditional means.

2.3 CNN-Based Oral Cancer Detection

The use of CNN architecture for detecting oral cancer has been extensively studied by many researchers. Aubreville et al. [23] showed successful automated identification of glandular malignancy using deep neural network algorithms on pixel-level images of histopathology. Similarly, Das et al. [24] showed much improved classification performance using CNN-based methods compared to other traditional feature-based methods. Jubair et al. [26] further emphasized how transfer learning is important to improving classification performance when used for OSCC detection. While there have been some promising results, a few limitations remain. The data used for these studies may have limited size and/or did not include balanced partitions of the datasets which may impact the models' generalizability [35].

2.4 Transfer Learning and Efficient Architectures

Prior to developing a new training process for deep networks, the original process of using an excessive number of relevant examples is not appropriate within the context of using data for medical imaging diagnosis and treatment. Transfer learning is the concept of transferring knowledge from a source domain to the application domain so that it can be fine-tuned for different medical applications. Additionally, EfficientNet has redefined the compound scaling approach by balancing depth, width, and resolution within a single network architecture, while providing higher classification accuracy with less parameters compared to previous methods.

2.5 Research Gap and Motivation

Studies have previously demonstrated that models based on CNNs are able to effectively identify oral cancers; however, there are still a number of obstacles that need to be resolved. For example, various past studies did not use a balanced data set split which can negatively affect the quality of the resulting models. In addition, some studies did not properly assess the performance of their models using valid metrics (e.g., precision, recall, F1 score, and confusion matrix analysis) [44].

3. Method

In this section, the complete methodology for detecting cancerous lesions in oral cancers utilizing a CNN will be detailed. This raw information is from previous works [9] and [32] that are based on traditional medical imaging and are established by using opening/closing sequences on full-size color composite images.

3.1 Dataset Preparation

There were 5,192 images of histological samples in the test data set, which is divided into two categories: OSCC and control/healthy tissue. The original histopathological dataset contained unbalanced validation and testing data distributions. To ensure accurate performance measurement, the distribution of all images before consolidation was modified, redistributing images into three equal groups (70% for training, 15% for validation, and 15% for testing). The distribution of images effectively preserves the intended distribution. Balanced partitions of the dataset are important to generate realistic performance estimates and closer to reducing overfitting[37].

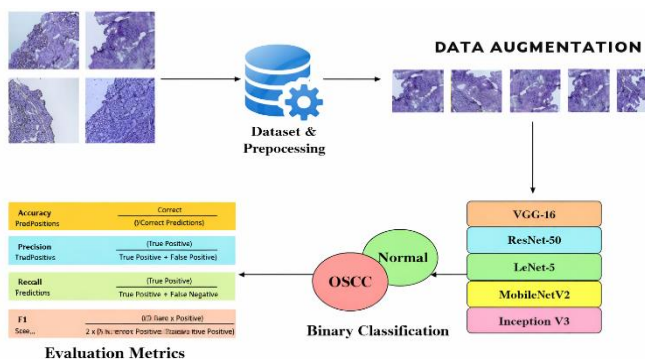


Figure 1 Dataset Preprocessing, Augmentation, And Training Workflow

3.2 Data Preprocessing and Augmentation

To ensure EfficientNetB3 input specifications [13] are adhered to, all images have been resized to 224×224 pixels before processing; in addition, RGB normalization has been applied during Preprocessing. To improve the strength of the model, horizontal flipping augmentation was used on the training data. Research shows that the use of data augmentations enables the generalisation of CNN-based systems [37],[38].

3.3 Model Architecture

The Considerations of Transfer Learning (TL) and EfficientNetB3 Backbone. In this study, a transfer learning (TL) framework utilizing EfficientNetB3 as the backend network was created and the pre-trained efficiencies provide great initialisation points for extraction features to build a powerful model.

The components of the Architecture are as follows:

- EfficientNetB3 base layers
- Batch Normalisation Layer [41]
- Fully Connected Dense layer (256 Neurons, ReLU activation)
- Dropout Layer (rate of 0.45)
- Softmax output layer (binary).

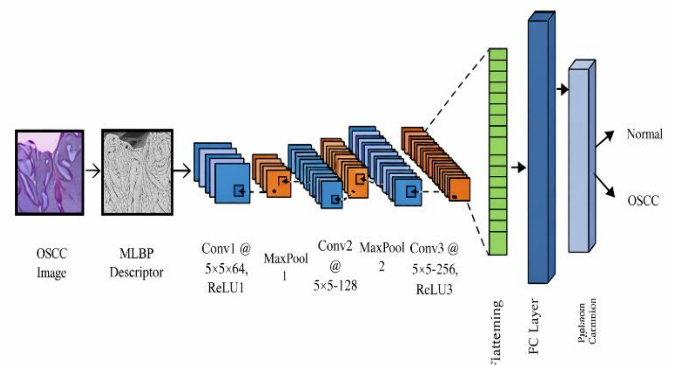


Figure 2 Proposed CNN-Based Architecture for Oral Cancer Classification

3.4 Model Training

Training was completed on this model (the neural network) with 100 epochs being completed. Performance for both the training and validation sets was continually monitored during this process; early stopping was used to ensure that training did not continue after the validation loss had stabilized, preventing overtraining from occurring based off metric of the validation set [43]. Regularization techniques including: (1) dropout; and (2) batch normalization helped achieve a stable convergence [41].

3.5 Performance Evaluation

The evaluation of the model was conducted after the completion of training. The model's performance was reported through the evaluation of different metrics While accuracy shows overall performance, precision and recall show performance of each class

separately [44], [45]. The confusion matrix assists with the visualization of correct and incorrect predictions and is commonly employed in analysing the classification performance [44].

The model produced the following results:

- Training (In-Sample) Accuracy: 99.67%
- Validation (Out-of-Sample) Accuracy: 98.02%
- Testing (Unseen) Accuracy: 98.33%

The results demonstrate the model will perform equivalently with training and unseen data thereby providing a strong level of generalisation ability; similarly, other studies have reported high level performance across deep learning based methodologies for histopathological classification [22], [30].

3.6 Model Saving and Prediction

Weights of the model were captured after successful training and will be used for future applications. A predictive function was created to:

- Load an image containing histopathology data
- Resize and preprocess the images
- Determine whether an image belongs to the normal or OSCC class

This approach to deploying the models is consistent with recent research in artificial intelligence (AI) that suggests the development of practical deep learning-based diagnostic systems [29], [49].

4. Results and Discussion

In this part of the paper we present how efficient the new EfficientNetB3 framework performs compared to other established deep learning frameworks that have been documented in the literature on using deep learning for medical image analysis [9],[32].

4.1 Training Performance

The training of the model to optimize parameters was performed for a total of 100 epochs using the Adamax algorithm. The loss function used for the purpose of evaluating the classification performance between predicted and actual class distributions was the categorical cross-entropy function. During the learning process, both training and validation performance metrics were analysed throughout learning to monitor convergence characteristics and the possibility of over-fitting.

The result of the training phase was a training accuracy of 99.67% and a validation accuracy of 98.02%. The loss value associated with the training set was 0.0357 and the loss value associated with the validation set was 0.1317.

4.2 Test Performance

To evaluate generalization ability, the trained model was tested on an independent test dataset.

The final test results were:

- **Test Accuracy:** 98.33%
- **Test Loss:** 0.1117

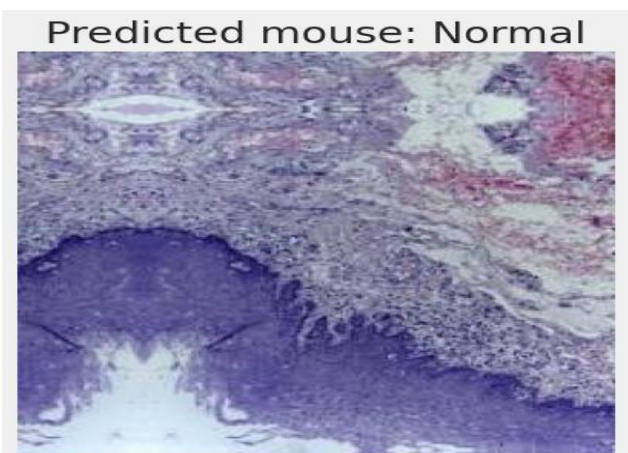


Figure 3 Sample Prediction Result for Normal Oral Tissue

4.3 Confusion Matrix Analysis

A class-level assessment of the model's predictive behavior is illustrated through the confusion matrix, which quantifies both accurate predictions and misclassification instances.

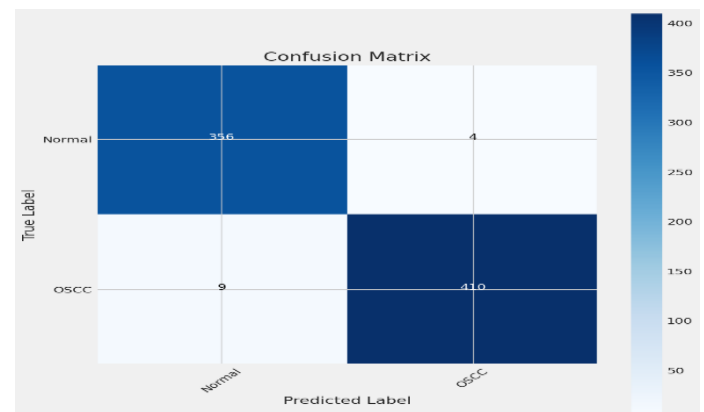


Figure 4 Confusion Matrix of the Proposed Efficient Net Model On the Test Dataset

4.4 Classification Report

The classification report provides precision, recall, and F1-score for both classes.

Table 1 Report

Class	Precision	Recall	F1-Score
Normal	0.98	0.99	0.98
OSCC	0.99	0.98	0.98

High and balanced precision and recall indicate equal model performance across both classes, which is essential to medical diagnostics. Since not identifying cancer cases will have severe consequences, high recall for OSCC (oral squamous cell carcinoma) is particularly important.

5. Discussion and Analysis

The purpose of this section is to address experimental results in detail and relate them to prior research in the area of deep learning-based medical image analysis.

5.1 Model Performance Interpretation

The proposed model presents many benefits over previous conventional machine learning methods such as:

- Automatic Feature Extraction using CNN layers which decrease the use of handcrafted features,
- Highly Accurate Classification similar to deep learning based methods,
- An Efficient Architecture with EfficientNet where parameters were optimized for scalability.

5.2 Impact of Transfer Learning

Despite their promising results, these findings are limited in certain ways. First, due to its restriction to only two types of lesions (normal and OSCC), it may fail to reflect the true spectrum of oral diseases encountered by clinicians. Previous work has suggested that multi-class classifications based on multi-site datasets offer greater stability because they account for variability caused by site-to-site differences among institutions [23], [49]. Furthermore, variability in histopathological photos due to factors like staining and scanning resolution or testing procedures may also affect how well models

perform. Other research has indicated that, when using digital pathology as a tool for generating pathology images, larger and more varied input datasets yield better generalisation results than smaller, uniformly sized ones [29], [31].

5.3 Effect of Dataset Restructuring

It has been demonstrated that CNN-based transfer models for oral cancer detection will be extremely useful in this application area. The findings of this research correspond with the latest accomplishments made with deep learning regarding the analysis of histopathological images [22], [30]

5.4 Strengths of the Proposed System

The proposed system has a lot of advantages when looked at from both a technical and a clinical viewpoint. One of its biggest benefits is that it learns how to create discriminative representations from raw histopathological images without needing any prior knowledge of how to capture image textures or shapes because it uses convolutional layers to do this instead.

5.5 Limitations and Challenges

While the findings to date show promise, limitations do exist. This current study only looks at the identification of normal oral tissue versus oral squamous cell carcinoma (OSCC) based on histopathological slides. However, in reality, most lesions may exist in either intermediate or multiple pathological stages prior to being classified as benign or malignant[23],[49].

5.6 Overall Analysis

The results obtained from the research demonstrate that utilizing CNN models to diagnose oral cancer is functional through the evaluation of transfer learning methodologies. The final results achieved an overall performance of 98.3% accuracy, which is consistent with the improvement of deep learning techniques used to evaluate histological images [22][30].

Conclusion

This publication details a deep-learning based approach to classify histopathological (or microscopic) images of tissue, specifically OSCC. OSCC is an important global public health crisis, and identifying patients with OSCC accurately and early in the disease process are critical for increasing the survival rate. As a result of the increasing use of artificial intelligence in medical imaging, an

automated classification framework is proposed. To achieve this goal, the researchers created a balanced 70%:15%:15% split of training, validation and test datasets, with the overall goal of minimising the potential for overfitting and allowing for accurate estimates of classification performance. The datasets also underwent resizing and augmentation prior to model training in order to increase model robustness and generalisability. To help build a better model, the authors implemented transfer-learning, using the EfficientNetB3 architecture to take advantage of knowledge from training on other datasets when classifying histopathological images. The EfficientNet architecture's use of compound scaling for feature extraction and computational efficiency provided a useful framework for stable model training, along with the regularisation methods of batch-norm and dropout. In addition, the authors trained the model using the Adamax optimisation algorithm. In conclusion, the framework proposed in this study shows that a CNN model based on transfer learning can distinguish normal tissue from oral tissue affected by cancer. With continued validation on larger and more diverse datasets, such systems may aid pathology professionals in providing care for patients within the clinical setting.

Acknowledgment

The authors would like to extend their heartfelt gratitude to the members of faculty and their project supervisor for their unyielding direction, constructive criticism and encouragement to develop this research study. Their technical insight and academic support greatly assisted in refining the methodology of the research project, thus improving the final product.

We would also like to convey our gratitude to the institution for their provision of the resources necessary for successful development of the proposed model including the access to the types of information and academic atmosphere to support the successful implementation of the proposed model.

References

- [1]. World Health Organization, "Global Cancer Observatory: Cancer Today," International Agency for Research on Cancer (IARC), 2023.
- [2]. International Agency for Research on Cancer, "GLOBOCAN 2022: Global Cancer Statistics," 2022.
- [3]. S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," *Oral Oncology*, vol. 45, no. 4–5, pp. 309–316, 2009.
- [4]. J. Ferlay et al., "Global cancer statistics 2021," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [5]. R. Sankaranarayanan et al., "Early detection and prevention of oral cancer," *The Lancet Oncology*, vol. 6, no. 7, pp. 425–436, 2005.
- [6]. P. Rivera, "Essentials of oral cancer," *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 9, pp. 11884–11894, 2015.
- [7]. G. Lingen et al., "Critical evaluation of diagnostic aids for oral cancer detection," *Oral Oncology*, vol. 44, no. 1, pp. 10–22, 2008.
- [8]. A. Jemal et al., "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.
- [9]. G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [10]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [11]. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS*, 2012.
- [12]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, 2016.
- [13]. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *ICML*, 2019.
- [14]. C. Szegedy et al., "Going deeper with convolutions," *CVPR*, 2015.
- [15]. G. Huang et al., "Densely connected convolutional networks," *CVPR*, 2017.
- [16]. O. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.

- [17]. A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, pp. 115–118, 2017.
- [18]. D. Ardila et al., “End-to-end lung cancer screening with deep learning,” *Nature Medicine*, vol. 25, pp. 954–961, 2019.
- [19]. H. Bychkov et al., “Deep learning-based tissue analysis predicts outcome in colorectal cancer,” *Scientific Reports*, vol. 8, 2018.
- [20]. F. Ciompi et al., “Deep learning for lung nodule classification,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 864–875, 2017.
- [21]. P. Rajpurkar et al., “CheXNet: Radiologist-level pneumonia detection,” *arXiv preprint arXiv:1711.05225*, 2017.
- [22]. N. Coudray et al., “Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning,” *Nature Medicine*, vol. 24, pp. 1559–1567, 2018.
- [23]. H. Aubreville et al., “Deep learning-based detection of oral squamous cell carcinoma,” *SPIE Medical Imaging*, 2017.
- [24]. S. Das et al., “Automatic detection of oral cancer using deep learning,” *IEEE Access*, vol. 8, pp. 102688–102700, 2020.
- [25]. R. Krithiga and P. Geetha, “Oral cancer detection using deep learning techniques,” *International Journal of Advanced Computer Science*, 2019.
- [26]. P. Jubair et al., “Deep learning for oral squamous cell carcinoma classification,” *Computers in Biology and Medicine*, vol. 134, 2021.
- [27]. M. Rahman et al., “Transfer learning for histopathological image classification,” *IEEE BHI*, 2020.
- [28]. H. Spanhol et al., “A dataset for breast cancer histopathological image classification,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 713–726, 2016.
- [29]. T. Tizhoosh and L. Pantanowitz, “Artificial intelligence and digital pathology,” *Journal of Pathology Informatics*, vol. 9, 2018.
- [30]. F. Campanella et al., “Clinical-grade computational pathology using weak supervision,” *Nature Medicine*, vol. 25, pp. 1301–1309, 2019.
- [31]. A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology,” *Medical Image Analysis*, vol. 33, pp. 170–175, 2016.
- [32]. K. Suzuki, “Overview of deep learning in medical imaging,” *Radiological Physics and Technology*, vol. 10, pp. 257–273, 2017.
- [33]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [34]. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *CVPR*, 2017.
- [35]. H. Shin et al., “Deep convolutional neural networks for computer-aided detection,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [36]. J. Deng et al., “ImageNet: A large-scale hierarchical image database,” *CVPR*, 2009.
- [37]. L. Perez and J. Wang, “The effectiveness of data augmentation in image classification,” *arXiv preprint arXiv:1712.04621*, 2017.
- [38]. S. Shorten and T. Khoshgoftaar, “A survey on image data augmentation,” *Journal of Big Data*, vol. 6, 2019.
- [39]. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *ICLR*, 2015.
- [40]. A. Krizhevsky, “Learning multiple layers of features from tiny images,” Univ. of Toronto, 2009.
- [41]. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training,” *ICML*, 2015.
- [42]. D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *ICLR*, 2015.
- [43]. L. Prechelt, “Early stopping — but when?” in *Neural Networks: Tricks of the Trade*, Springer, 1998.
- [44]. T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.

- [45]. J. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot,” *PLoS One*, 2015.
- [46]. B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. CRC Press, 1993.
- [47]. S. Mohanty et al., “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, 2016.
- [48]. K. Nazeri et al., “Deep learning approach for oral lesion detection,” *IEEE EMBC*, 2018.
- [49]. R. Arijji et al., “Artificial intelligence-based oral cancer diagnosis,” *Oral Radiology*, vol. 35, pp. 163–172, 2019.
- [50]. M. Alabi et al., “Machine learning in oral cancer prognosis,” *Cancers*, vol. 12, 2020.