

Article

Benign and Malignant Oral Lesion Image Classification Using Fine-Tuned Transfer Learning Techniques

Md. Monirul Islam ¹, K. M. Rafiqul Alam ², Jia Uddin ³, Imran Ashraf ^{4,*} and Md Abdus Samad ^{4,*}

¹ Department of Software Engineering, Daffodil International University, Daffodil Smart City (DSC), Birulia, Savar, Dhaka 1216, Bangladesh

² Department of Statistics, Jahangirnagar University, Dhaka 1342, Bangladesh

³ AI and Big Data Department, Endicott College, Woosong University, Daejeon 34606, Republic of Korea

⁴ Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, Republic of Korea

* Correspondence: imranashraf@ynu.ac.kr (I.A.); masamad@yu.ac.kr (M.A.S.)

Abstract: Oral lesions are a prevalent manifestation of oral disease, and the timely identification of oral lesions is imperative for effective intervention. Fortunately, deep learning algorithms have shown great potential for automated lesion detection. The primary aim of this study was to employ deep learning-based image classification algorithms to identify oral lesions. We used three deep learning models, namely VGG19, DeIT, and MobileNet, to assess the efficacy of various categorization methods. To evaluate the accuracy and reliability of the models, we employed a dataset consisting of oral pictures encompassing two distinct categories: benign and malignant lesions. The experimental findings indicate that VGG19 and MobileNet attained an almost perfect accuracy rate of 100%, while DeIT achieved a slightly lower accuracy rate of 98.73%. The results of this study indicate that deep learning algorithms for picture classification demonstrate a high level of effectiveness in detecting oral lesions by achieving 100% for VGG19 and MobileNet and 98.73% for DeIT. Specifically, the VGG19 and MobileNet models exhibit notable suitability for this particular task.

Keywords: data-efficient image transformer (DeiT); VGG19; oral lesions; MobileNet; benign; transfer learning; malignant



Citation: Islam, M.M.; Alam, K.M.R.; Uddin, J.; Ashraf, I.; Samad, M.A. Benign and Malignant Oral Lesion Image Classification Using Fine-Tuned Transfer Learning Techniques. *Diagnostics* **2023**, *13*, 3360. <https://doi.org/10.3390/diagnostics13213360>

Academic Editors: Francesco Inchingolo and Gianna Dipalma

Received: 28 September 2023

Revised: 23 October 2023

Accepted: 30 October 2023

Published: 1 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Oral disease is a major global health concern affecting millions of people worldwide. It was the reason for 15,511 deaths across Bangladesh in 2020. More than 50,000 cases of oral cancer were detected during the same time period [1]. Therefore, oral disease is a substantial problem for us in Bangladesh, as we are in the financial region of a developing country [1]. Tobacco is considered the most common risk factor for oral cancer [2]. Consuming betel nuts has been linked to an increased risk of oral lesions and oral cancer in the Indian subcontinent. According to the World Cancer Research Fund (WCRF) [3], this particular cancer ranks 16th in terms of prevalence, with men being affected at the 11th highest rate and women at the 18th highest rate. The primary stage of oral cancer is oral infection. Moreover, it can be characterized by late diagnosis only. So, if it is possible to detect the type of infection automatically, detecting cancer will not be time-consuming and will also help start treatment at the early stage, reducing the death rate due to oral cancer.

The mucosal tissue or oral area can be altered, which is manifested in several lesions, but comparatively localized overbuilding of oral mucosal tissue is considered reactive rather than neoplastic [4]. The most common oral cavities, infections, or lesions are benign and malignant lesions. Benign lesions, or non-cancerous skin lesions, are less harmful than malignant lesions. Malignant lesions contain cancerous cells and are very harmful to human beings. The improper treatment of malignant lesions can cause death. However, both lesions are the results of tumors. These lesions can be detected by their shape, form,

or size. Depending on the weather and geographic position, the characteristics of lesions may vary [5].

Unfortunately, sometimes there is a misunderstanding between cancerous and non-cancerous lesions. Some lesions are not cancerous, depending on their growth, which may occur due to low-grade trauma or a lack of oral hygiene and, in some cases, hormonal reasons; pyrogenic granuloma (PG) is one of the cancerous cases. Similarly, benign lesions are rarely considered non-cancerous lesions, which are the first stage of malignant lesions, and proper treatment is not provided, so the patient must face many difficulties, such as physical problems and sometimes death [4]. Oral pathologies can encompass a wide range of conditions, including both malignant and non-malignant conditions. There are some common malignant and non-malignant oral pathologies associated with various factors, such as oral hygiene, hormonal conditions, trauma, and others. Oral squamous cell carcinoma (OSCC), oral verrucous carcinoma, and mucoepidermoid carcinoma are the common malignant oral cancers [6]. Dental caries, gingivitis, and periodontitis are associated with oral hygiene. Oral contraceptive-associated gingivitis and pregnancy gingivitis occur due to hormonal conditions. Accidents, falls, or physical injuries can lead to various oral injuries, including broken teeth, lip lacerations, or jaw fractures, called oral trauma. Oral thrush (Candidiasis) and herpes labialis (cold sores) are infectious oral conditions. Oral cancer may also be affected by salivary gland disorders, like sialadenitis and Sjogren's Syndrome [7,8].

Automating the detection of these lesions will reduce costs and allow the disease to be diagnosed earlier. Furthermore, perfect detection will help physicians provide the correct treatment and reduce the suffering of patients, as treatment at the last stage is very costly. Artificial intelligence (AI), machine learning (ML), deep learning (DL), transfer learning (TL), and digital image processing can be used to build an automated system for detecting lesions. There are several ways to automate the diagnosis of oral cancer. AI is a computer science field that uses computers and machines to simulate the human mind's problem-solving and decision-making skills. ML and DL are subfields of AI, but the two terms have different ways of working. ML learns from previous data without any explicit programs and predicts the outcome accurately. DL works like a human brain structure. There are many data types for classification using AI, ML, and DL. Some of them are text data, image data, structured data, unstructured data, sensor data, social data, and so on. The particular problem at hand, the characteristics of the data, and the employed methods all influence the choice of data type. Models for machine learning and deep learning are trained using data that are most pertinent to the purpose for which they were created. We used image data types in this paper. For predicting malignant lesions or benign lesions, a microscopic image can be used [9,10]; either a real-time high-regulated image or hyperspectral mouth image can also be used [11], and even a computed tomographic (CT) image [12] or fluorescence image can create a good dataset [13]. The field of computer vision is gaining popularity with convolutional neural networks (CNNs) [9–13]. However, large-scale, reliable clinical data are required for employing deep learning in the early diagnosis of oral cancer. The researchers examined a machine [14,15] for classifying oral cancer and dental disease. The contributions of this work are as follows.

- We added extra patch tokens and distillation tokens to the DeIT method for performing the classification effectively and detecting the lesions by providing 98.73% correct information for malignant lesions.
- We modified the structure of VGG19 and MobileNet by changing the fully connected (FC) layer, and both models give 100% accuracy.
- We established a standard comparison table with similar works. Among them, our models outperform on the basis of accuracy.

For benign and malignant lesion detection using transfer learning approaches, we formulate the following hypotheses:

- Null Hypothesis (H_0): there are no significant differences between the performances of fine-tuned transfer learning models;

- Alternative Hypothesis (H_a): the performances of fine-tuned transfer learning models have statistically significant differences.

A comparison between the results available in the literature and the result proposed in this study is shown in Table 1. The results show that MobileNet and VGG19 have much potential for medical use. On the other hand, DeIT offers an option that might be better at generalization.

Table 1. Comparison among the existing works.

Reference	Year	Model	Accuracy
Welikala et al. [16]	2020	ResNet-101	87.07%
Prabhakar and Rajaguru [17]	2017	TNM	90.88%
Nanditha et al. [18]	2020	BPNN	95%
Thomas et al. [19]	2013	ANN11	97.92%
Nanditha [20]	2021	VGG16	96.2%
Shamim et al. [21]	2019	VGG19	96.7%
Wu et al. [22]	2019	AlexNet	91%
Proposed	2023	VGG19 and MobileNet	100%

2. Literature Review

Many researchers proposed different methods for finding the type of oral lesions using machine learning, deep learning, or transfer learning based on images of the lesions. The authors examined deep learning models for classifying and detecting oral lesions based on the images of the MeMoSA project. MeMoSA (Mobile Mouth Screening Anywhere) is a mobile phone app for collecting oral lesion images developed by Cancer Research Malaysia. They used ResNet-101 for classification, and for object detection, R-CNN was used. Furthermore, the accuracy (F1-score) was 87.07% for the lesion class, and for referral, it was 78.30% [16]. In [17], the authors analyzed oral mucosa disorders and their diagnosis through digital imaging and compared TNM (Tumor–Node–Metastasis) staging with a neural network for 75 patients, achieving high accuracy: 100% for T1, 85.19% for T2, 84.21% for T3, and 94.12% for T4. In [23], a study utilized deep transfer learning algorithms, like ResNet50, MobileNetV2, VGG19, VGG16, and DenseNet, on oral cancer images from histopathologic and real-time datasets. The most effective results were obtained with DenseNet using hybrid optimization techniques, achieving accuracies of 92.41% for real-time, 95.41% for oral cancer, and 92.41% for non-cancerous histopathologic images. Marzouk et al. [24] introduced the AIDTL-OCCM (Artificial Intelligence with Deep Transfer Learning Oral Cancer Detection and Classification Model), which used AI and image processing for oral cancer detection. It employs fuzzy-based contrast enhancement and DenseNet-169 for feature extraction, along with the Chimp Optimization Algorithm (COA) and an autoencoder (AE) for detection and parameter optimization. The results show that AIDTL-OCCM outperforms other methods, achieving a maximum accuracy of 90.08% in the early detection of oral cancer. Nanditha et al. [18] used texture and fractal features to analyze 200 oral lesion images from Karnataka medical institutions. A back-propagation neural network achieved 95% accuracy in detecting malignant and benign lesions, showing promise for improving oral cancer diagnosis. Rahman et al. [25] used transfer learning with AlexNet on oral squamous cell carcinoma biopsy images, achieving impressive classification accuracy of 97.66% for training and 90.06% for testing. In [20], an automated system for early oral lesion diagnosis was created using an ensemble deep learning model (Resnet-50 and VGG-16). Trained on an augmented dataset, it outperforms other models with 96.2% accuracy, 98.14% sensitivity, and 94.23% specificity. In [21], the authors applied six deep convolutional neural network (DCNN) models with transfer learning to automate the screening for oral cavity cancer (OCC) using a small dataset of tongue lesion images. The models successfully differentiated between benign and precancerous lesions

and identified various tongue lesion types with an average accuracy of 93%. The authors proposed machine learning models with an accuracy of 86% for classifying oral cancers based on electronic health record (EHR) data [22]. In the study [25], it was observed that the highest achieved classification accuracy was around 97.66%. Therefore, there are research opportunities to improve the accuracy of the classification of lesion images in regard to the distinction between benign and malignant kinds. The primary objective of this work is to improve the precision of the classification of lesion images by distinguishing between benign and malignant kinds.

The problems and questions dealt with in our work are as follows:

- The efficacy of fine-tuned TL methods: The purpose of this work is to determine how successfully fine-tuned transfer learning models can be adapted to the job of categorizing oral lesions based on picture data.
- Classification accuracy: This work focuses on the classification process's correctness, especially in discriminating between benign and malignant oral lesions.
- Comparison to existing works: This study compares the performance of fine-tuned transfer learning approaches to classic or conventional methods for oral lesion categorization.

In our work, we proposed three deep transfer learning (DTL) models, DeIT, VGG19, and MobileNet. VGG19 and MobileNet perform the best with an accuracy of 100% among the models. Therefore, the deep transfer learning (TL) model is proposed here to detect oral lesions or oral cancer and to classify benign and malignant lesions.

3. Proposed Methodology

Figure 1 deals with the block diagram of the methodology. First, the dataset is divided into training test and validation datasets, and then those data are preprocessed. Preprocessing is essential for achieving better accuracy. During splitting, the real data are used as a test dataset, the augmented data are used as a training dataset, and the same is used for validation. Then, we applied a deep transfer learning model for classifying the oral disease with fine-tuning models and not. The deep transfer learning model is a pre-trained model where knowledge of a model is applied to another model.

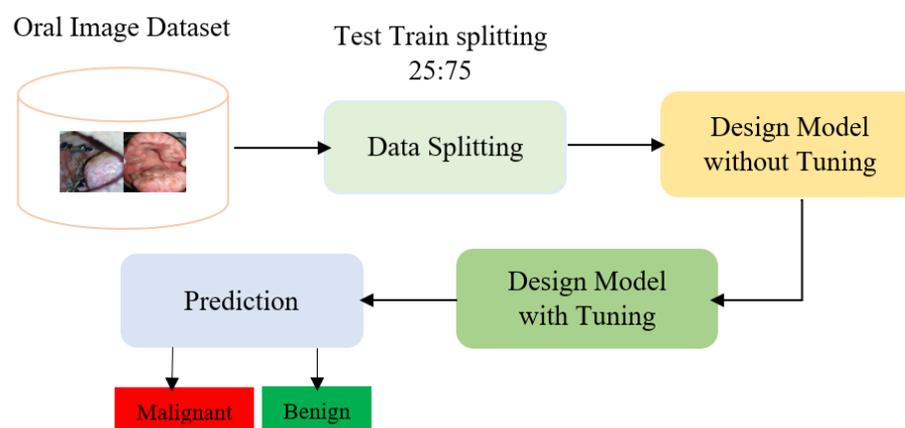


Figure 1. The proposed methodology.

3.1. Dataset Description

The dataset is collected from the Mendeley platform [26], which contains color images of two kinds or types of oral lesions. The dataset was created using mobile cameras and intraoral cameras. After that, these images were consulted on with doctors from different hospitals in Karnataka, India. There were 165 benign lesions and 158 malignant lesions images in this dataset originally, and the augmentation technique was applied to increase this amount by up to 1320 training images for benign lesions and 1273 training images for

malignant lesions. Three augmentation techniques named flipping, rotation, and resizing were used to extend the dataset.

3.2. Dataset Splitting and Enhancement

The dataset contains 2593 images in total. Among them, 1320 images are benign lesions, and the rest 1273 are malignant lesions. This study collected 165 images of benign lesions and 158 of malignant lesions. Later, flipping, rotation, resizing, and cropping in word augmentation increased the data in the dataset. Figure 2 reflects the sample data from the dataset. We considered the image resizing, normalization, augmentation, color, and shape as features.

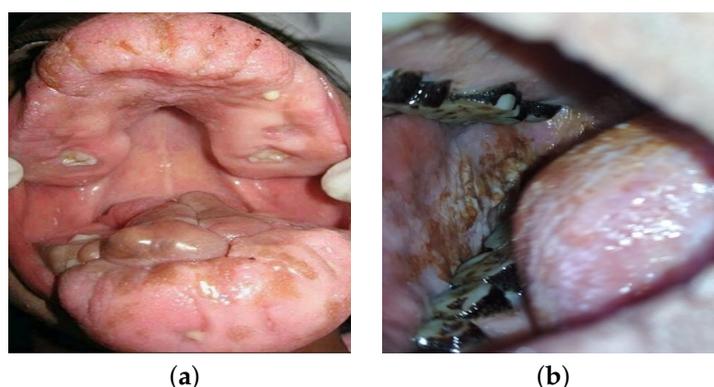


Figure 2. Oral lesion classes: (a) benign lesion; (b) malignant lesion.

3.3. Transfer Learning Model Design

It is imperative at present in each sector of science, especially in medical science, to classify images depending on different features and properties. Transfer learning is becoming popular nowadays, as deep learning requires massive amounts of data and is time-consuming. Transfer learning gives opportunities to use the knowledge of previously used data. It helps to solve problems efficiently and reduces the demand for data. From a considerable dataset, many classes can be recognized using transfer learning. Even in computer vision problems, achieving expected accuracy using less data and computing resources is possible using the mentioned technique. Several transfer learning models are popular, like VGG16, ALEXNet, ResNet, and so on [27–29]. Transfer learning or deep transfer learning (DTL) are related concepts using pre-trained models but have little difference. DTL is a more specialized transfer learning [30]. In this paper, we modified VGG19, MobileNet, and DeIT, which are well-performing transfer learning algorithms.

3.3.1. VGG19 Model

VGG19 is a convolutional neural network model proposed in [31]. The model demonstrated a top-5 test accuracy of 92.7% on the ImageNet dataset, which included more than 14 million images dispersed across 1000 different classes. VGG19 is a variant of the VGG model, which uses 19 layers, including 16 convolutional layers, 5 max-pooling layers, and 3 fully connected layers. Figure 3 shows the regular architecture of the VGG19 model. VGG19 is widely used in applications such as object detection, image classification, and image segmentation [32].

Each convolutional layer uses a 3×3 filter and a ReLU (Rectified Linear Unit) activation function. Max-pooling layers use a 2×2 filter and a stride of 2. The input to the network is an image of size $224 \times 224 \times 3$, where the first number is the height, the second is the width, and the third is the number of channels (RGB).

The convolutional layers apply a 3×3 filter to the image, which is then multiplied by a set of weights and passed through a ReLU function to produce an activation map. The max-pooling layers then combine the output of the convolutional layers into a single

feature map, which is then fed into the fully connected layers. The fully connected layers consist of neurons with weights and biases that are used to classify the image.

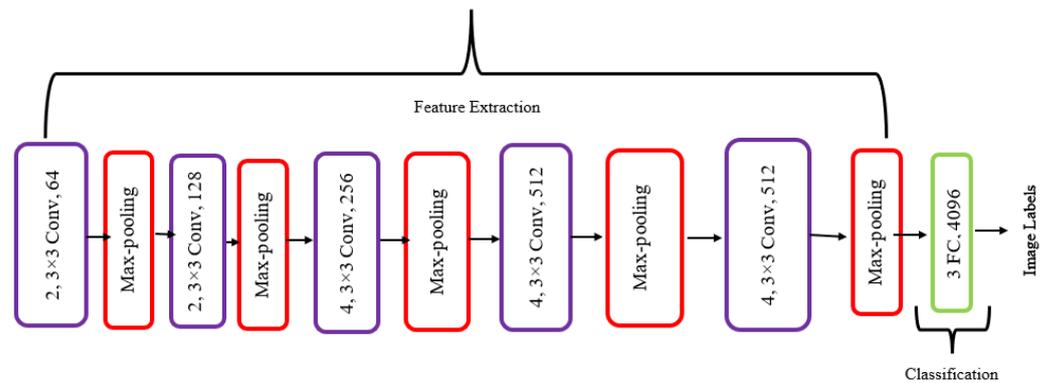


Figure 3. General VGG19 architecture.

In our architecture, we replaced the classification part with dimensional reduction and then added the classification, which is visually shown in Figure 4. Moreover, the hyper-tuning we have performed in the code is given in Table 2.

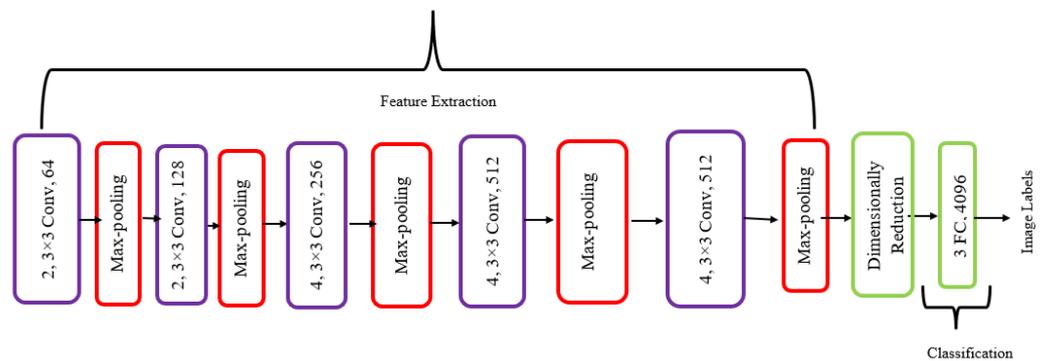


Figure 4. Modified VGG19 architecture.

Table 2. Hyper-tuning in VGG19.

Parameter	Parameter Value
Batch size	32
Weight matrix	imagenet
Activation function	softmax
Criterion	sparse_categorical_crossentropy
Optimizer	Adam

3.3.2. MobileNet Model

MobileNet is a deep learning-based computer vision algorithm for efficient on-device image classification. Figure 5 shows that MobileNet combines depthwise separable convolutions and linear bottlenecks. The depthwise separable convolutions split the traditional convolution operation into two separate operations, reducing the number of computations and the number of parameters, which helps reduce the network’s size and computational complexity. The linear bottlenecks further reduce the number of parameters in the network and make it more computationally efficient [33].

In the case of MobileNet, the convolutional layers are replaced with depthwise separable convolutions. A depthwise separable convolution is a two-step operation that first

applies a single filter to each input channel and then concatenates the results. This approach reduces the number of parameters and computations compared to a traditional convolution, making the network more efficient.

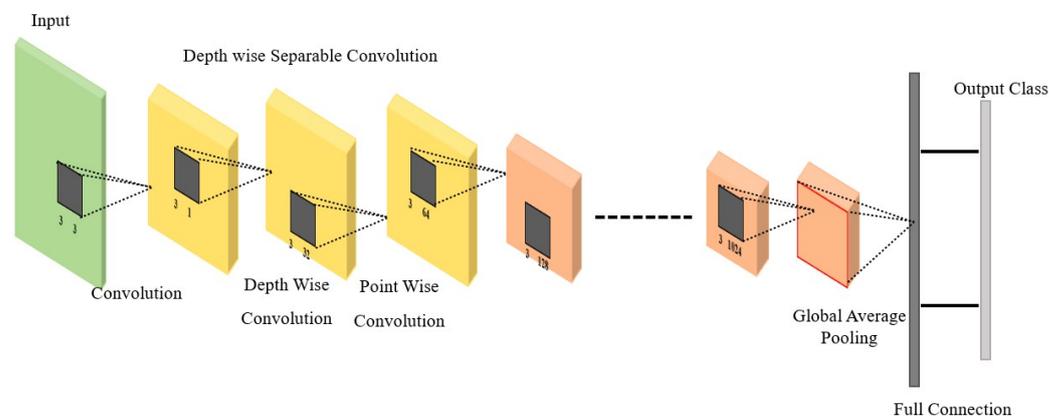


Figure 5. General MobileNet architecture.

The linear bottlenecks in MobileNet refer to 1×1 convolutional layers added between the depthwise separable convolutions. These bottlenecks reduce the number of input channels, reducing the network's computational and memory requirements. In the modified architecture, we used global max pooling instead of global average pooling. Our hyper-tuning for the model is shown in Table 3.

Table 3. Hyper-tuning in MobileNet.

Parameter	Parameter Value
Batch size	32
Weight matrix	Imagenet
Activation function (dense: 1024)	ReLU
Activation function (dense: 512)	ReLU
Activation function (dense: 2)	softmax
Criterion	categorical_crossentropy
Optimizer	Adam
Learning Rate	0.02

3.3.3. DeIT Model

Image transformation techniques are used to improve the appearance of digital images. The improving techniques include improving the clarity, sharpness, contrast, brightness, and color. DeIT is a technique used to improve the quality of digital images automatically. The improving technique is performed using a deep learning algorithm to learn the features of a digital image and then apply appropriate transformations to improve the image. This technique can enhance various medical, industrial, artistic, and recreational images. DeIT can also automatically enhance the quality of images taken with mobile devices, making them look better without manual intervention [34]. The real architecture of this model is shown in Figure 6, and our modified architecture is shown in Figure 7. Table 4 shows the hyper-tuning for the model.

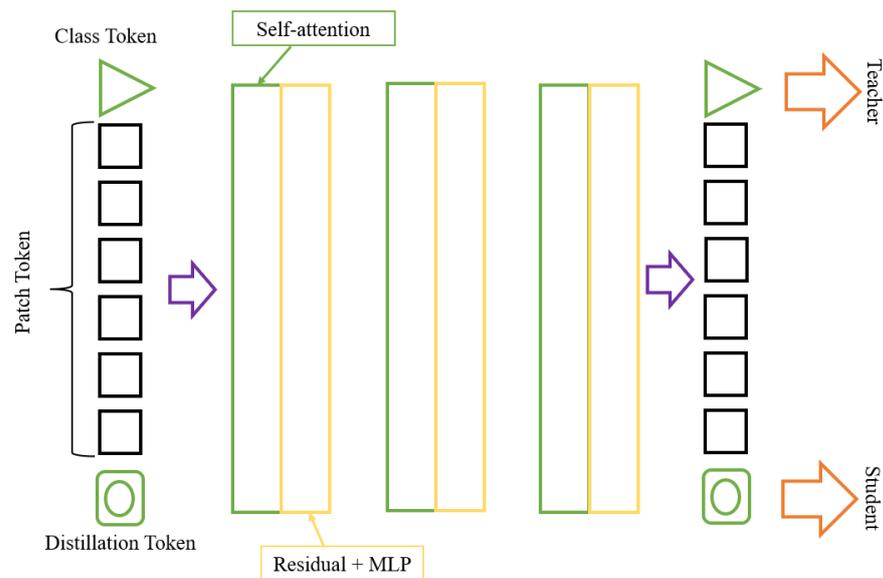


Figure 6. General DeiT architecture.

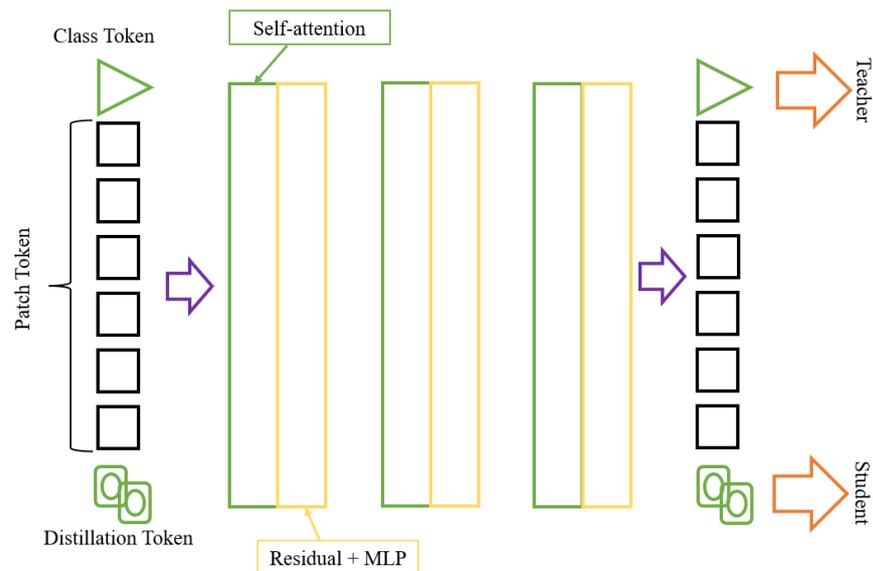


Figure 7. Modified DeiT architecture.

Some important terms regarding DeiT are the following:

- Interpolation: used to estimate the value of unknown points using known nearby points.
- Convolution: a mathematical operation that combines two functions to create a third function.
- Fourier transform: a mathematical tool representing a signal or image as a sum of sine and cosine waves of different frequencies.
- Resampling: the process of changing the resolution of an image by altering the number of pixels.
- Normalization: the process of adjusting the data values to a common scale.
- Histogram equalization: a contrast adjustment technique used to improve the contrast of an image by redistributing the pixel intensities.
- Edge detection: the process of locating sharp changes in an image.
- Affine transformation: a mathematical tool to transform an image by modifying its shape, size, and orientation.

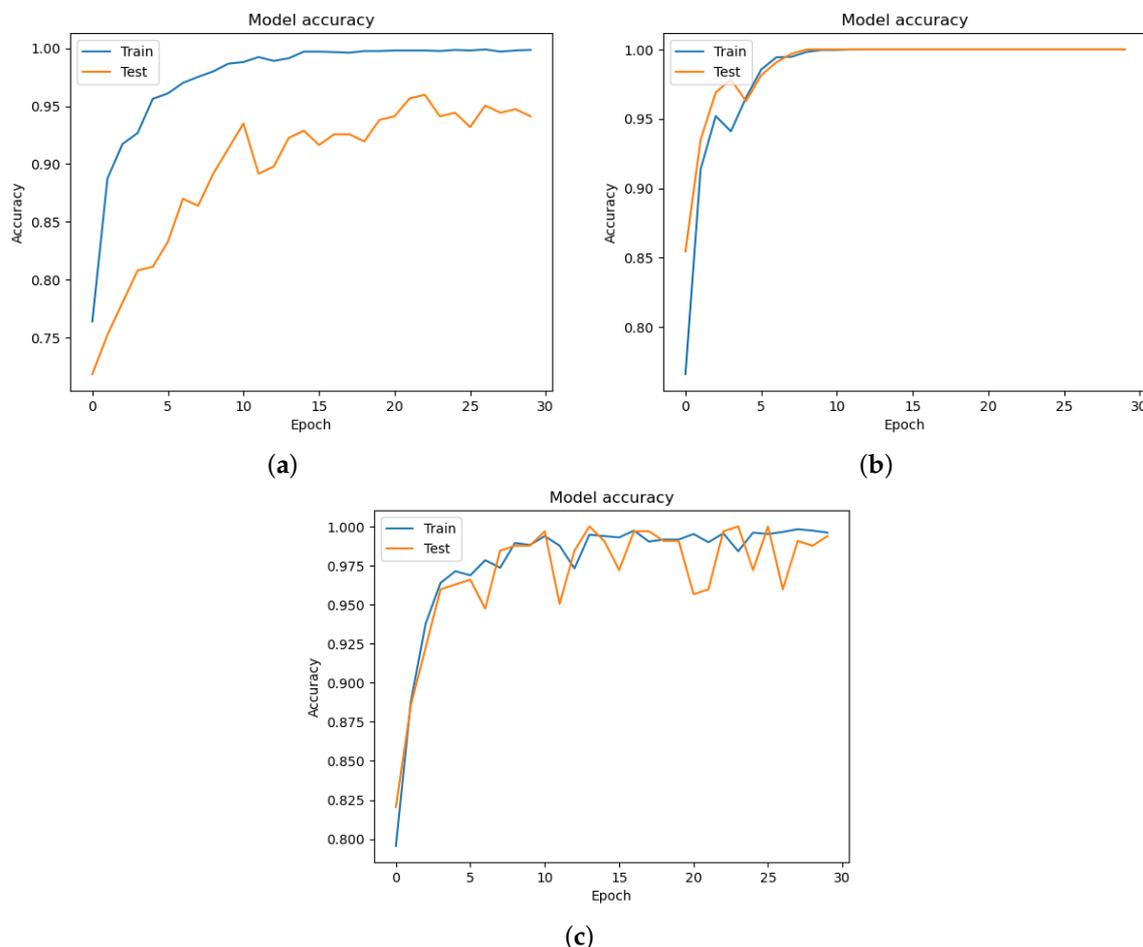
Table 4. Hyper-tuning in DeIT.

Parameter	Parameter Value
Batch size	128
Weight matrix	Linear
Activation function	ReLU
Regularization	Dropout
Criterion	LabelSmoothingCrossEntropy
Optimizer	Adam
Learning rate	0.001

3.4. Experimental Preparation and Assessment

Our dataset is a collection of substantial image sets. So, for training or testing purposes, a high-configuration machine is required. Jupyter Notebook is a way to run the program, as our dataset is split into training, testing, and validation before preprocessing. Our proposed model combines OpenCV, pickle, numpy, seaborn, matplotlib, Keras, TensorFlow, and sklearn.

The cross-entropy loss is calculated among the training and valid datasets at each epoch. The cross-entropy was categorical for our proposed models. The Adam optimizer uses the learning rate of 0.1% [35]. All our models (DeIT, VGG19, ResNet) have been trained for 30 epochs. The accuracy of each model is described in Figure 8, where Figure 8a is of DeIT, Figure 8b is of VGG19, and Figure 8c is of MobileNet. The loss function of the proposed models is presented in Figure 9. The loss graphs of DeIT in Figure 9a, VGG19 in Figure 9b, and MobileNet in Figure 9c illustrate the points of the validation and training dataset.

**Figure 8.** Accuracy of the models: (a) DeIT accuracy; (b) VGG19 accuracy; (c) MobileNet accuracy.

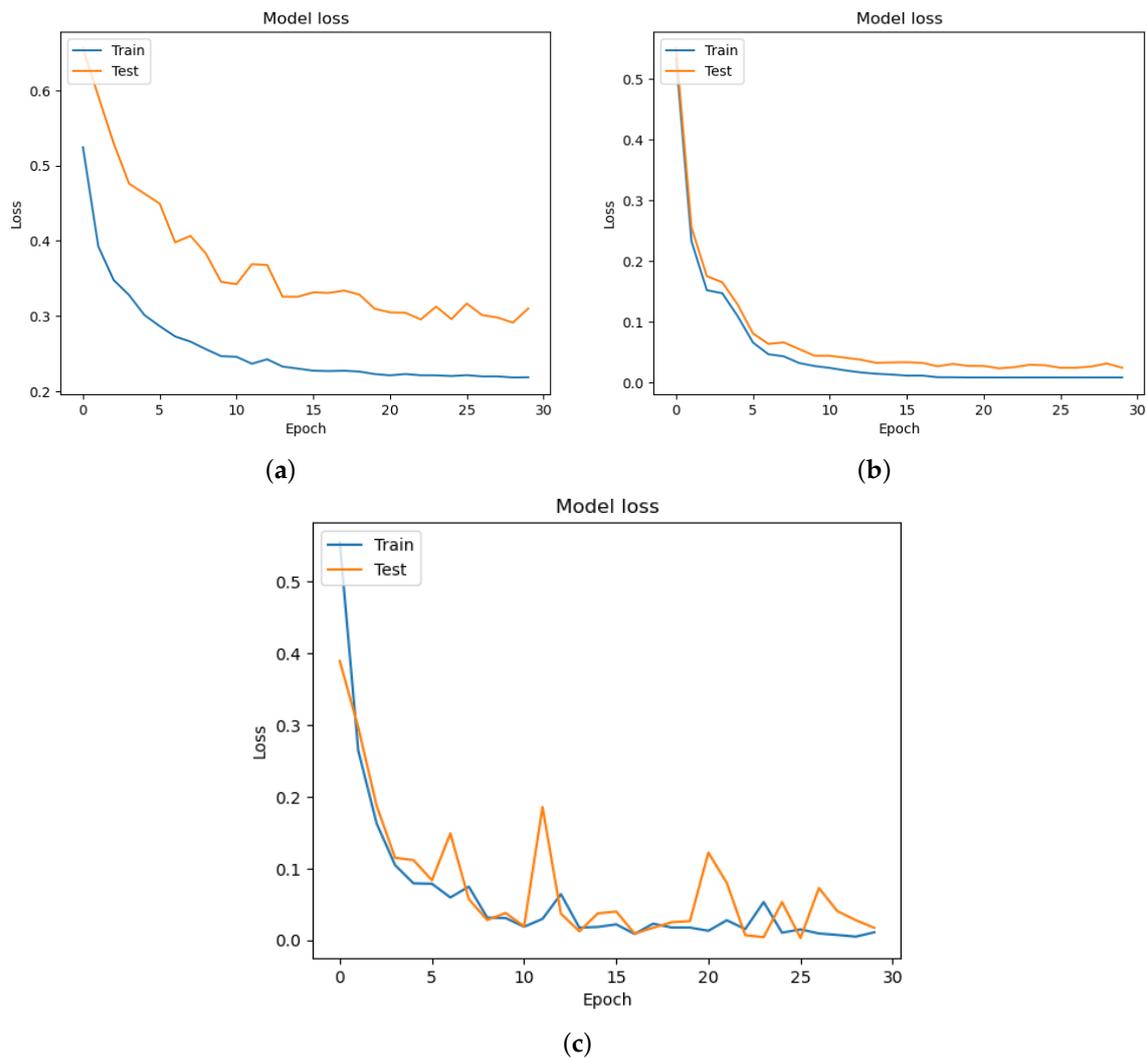


Figure 9. Loss function characteristics of the models: (a) DeIT; (b) VGG19; (c) MobileNet.

4. Result Analysis and Discussion

We evaluated five performance metrics including the precision, recall, F1-score, MCC, and accuracy for all DTL models. All these performance parameters can be measured using the confusion matrix. There are four keywords in the confusion matrix named true positive (TP—it means when the actual value and the prediction value are positive), true negative (TN—it means when the actual value and the prediction value are negative), false positive (FP—it means when the actual value is negative but the prediction value is positive), and false negative (FN—it means when the actual value is positive but the prediction value is negative). All performance metrics are mathematically defined in Equations (1), (2), (3), (4), and (5), respectively.

Recall is the ratio of the total number of true positive classes to the predicted true positive class.

$$\text{Recall}, R = \frac{TP}{TP + FN} \quad (1)$$

Precision is the ratio of correct positive guesses to total (correct and false) positive guesses.

$$\text{Precision}, P = \frac{TP}{TP + FP} \quad (2)$$

The F1 measure is the harmonic average of precision and recall.

$$F1Score, F1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

Accuracy is the ratio between the number of correct predictions and the total number of predictions.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions made}} \quad (4)$$

The Matthews Correlation Coefficient (MCC), also known as the phi coefficient, is a measure used in machine learning and statistics to evaluate the quality of binary classification models, particularly when dealing with imbalanced datasets.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (5)$$

The results of our study showed that all three models, DeIT, VGG19, and MobileNet, achieved high levels of accuracy in detecting oral lesions. Accuracy graphs for all the models are shown in Figure 8. In particular, the VGG19 and MobileNet models had a perfect accuracy rate of 100%, while the DeIT model achieved a slightly lower accuracy rate of 98.86%. From precision, recall, and F1-score metrics for each model, as shown in Table 5, we found that the DeIT model had a precision rate of 96.81%, recall rate of 96.68%, and F1-score of 96.68%. Meanwhile, the VGG19 and MobileNet models had perfect precision, recall, and F1-score values of 100%. We also calculated the Matthews Correlation Coefficient (MCC) for each model to evaluate its performance and found that the VGG19 and MobileNet models had perfect MCCs of 100%, while the DeIT model had an MCC of 93.49%.

Table 5. Summary of the result where P = precision and R = recall.

Model	P (%)	R (%)	F1 (%)	MCC (%)	Accuracy (%)
DeIT	98.73	94.24	96.38	93.49	96.40
VGG19	100	100	100	100	100
MobileNet	100	100	100	100	100

These results suggest that all three models effectively detect oral lesions, with the VGG19 and MobileNet models achieving perfect accuracy rates. The DeIT model also performed well, with high precision, recall, and F1-score values and a relatively high MCC value.

We have shown the performance of each model through the use of loss, as shown in Figure 9, and accuracy graphs, as shown in Figure 8, which indicated that all three models were able to converge to a low loss and high accuracy after several epochs of training. Additionally, we created confusion matrices for each model to visualize their performances in identifying true positives, true negatives, false positives, and false negatives, as shown in Figure 10.

The confusion matrices for the models show that false negatives were more common than false positives, as shown in Figure 10. The figure shows that models are better at detecting benign/malignant lesions than correctly classifying them. The false negatives could be due to the difficulty distinguishing between benign and malignant lesions, especially for images with ambiguous features.

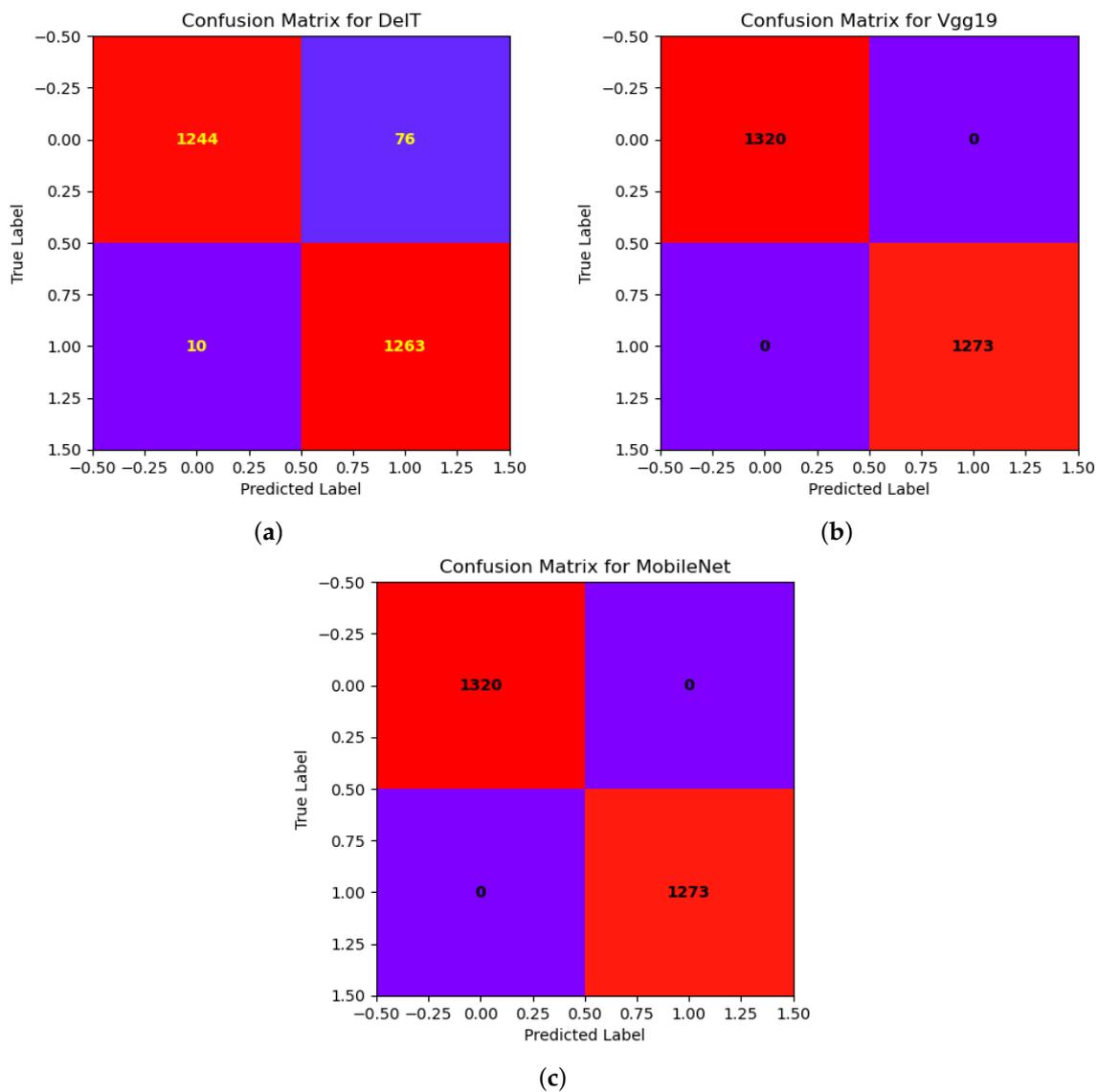


Figure 10. Confusion matrices of the models: (a) DeIT; (b) VGG19; (c) MobileNet.

4.1. Cross-Validation

We applied stratified cross-validation techniques to evaluate the performance of the models. The results of all models using stratified cross-validation are shown in Table 6. This table demonstrates almost the same results as the one without validation.

Table 6. Performance analysis of all models using cross-validation where P = precision and R = recall.

Model	P (%)	R (%)	F1 (%)	MCC (%)	Accuracy (%)
DeIT	98.75	94.5	96.54	93.58	96.70
VGG19	100	100	100	100	100
MobileNet	100	100	100	100	100

4.2. Discussions

The results of our study indicate that deep learning models can effectively detect oral lesions with high accuracy. The 100% accuracy achieved by MobileNet and VGG19 highlights the potential of using these models in clinical settings. However, it should be noted that DeIT achieved a high accuracy of 98.73%, which is still very promising for practical use. The superior performances of MobileNet and VGG19 can be attributed to

their deeper architectures and the use of pre-trained weights. These models could learn and distinguish between different lesion features with high accuracy, even with a relatively small dataset. Due to its transformer-based architecture, DeIT may have a greater capacity to generalize to unseen data despite showing slightly less accuracy. The utilized models have some drawbacks. They are as follows:

- VGG19 has some significant drawbacks. These include its high memory requirements, inefficiency for installations with limited resources, and high computational complexity. It is made just for picture data, frequently has to be adjusted, and is ineffective in capturing spatial hierarchies. Furthermore, the quality and quantity of training data have a significant impact on its performance, which raises issues with fairness when biased data are present.
- Despite its effectiveness, MobileNet is not without its restrictions. It is not well adapted for tasks like natural language processing or textual data. It may struggle with fine detail detection, be inaccurate in complicated picture recognition, and perform best in generic image categorization. Its small size could make it less capable of handling high resolutions or bigger photos. Finally, understanding its decision-making process might be difficult.

We have some direction for the future research's scope. The first is dataset scarcity. There are only 165 benign lesions and 158 malignant lesions. This could be increased in the future. The second is to experiment with more DTL models, such as the three models applied to classify oral lesions in this paper. It could be a combination of two or more models.

We covered all the mentioned questioning problems in this paper. We did fine-tune all the models. Tables 2–4 show the hyper-parameters of all models for fine-tuning. The accuracy of all models is shown in Table 5. Table 1 shows the comparison with existing works.

5. Conclusions

This study showed the efficiency of deep learning models in identifying benign and malignant oral lesions. The comparative analysis among the three techniques—DeIT, VGG19, and MobileNet models—showed promising capabilities for detecting oral lesions. MobileNet and VGG19 have much potential for medical use, as the experiment results showed they had the highest accuracies. However, DeIT offers an option that might be better at generalization. This study also showed the importance of using “data augmentation” methods to make the datasets bigger and make the models work better. In addition, the results show that loss and accuracy graphs and confusion matrices are performance-indicating parameters for oral lesion classifications. Several metrics were used to evaluate how well the models worked, including their precision, recall, F1-score, MCC, and accuracy. The evaluation metrics show that all three models performed well in classifying oral lesions. Significantly, the VGG19 technique achieved perfect performance parameters—its accuracy, precision, recall, F1-score, and MCC. The dataset included in this research is of a limited size (total images). It comprises images of patients from a specific geographic region inside the area. Hence, it is imperative for the approach to undergo further examination using a diverse set of lesion images. For future work, this study recommends further exploring more advanced data augmentation techniques and deep learning architectures to improve the models' performance. Additionally, this study suggests exploring transfer learning techniques to fine-tune pre-trained models on a specific oral lesion dataset. Furthermore, this study suggests assessing the effectiveness of the models on large datasets that cover diverse geographical regions to ascertain their effectiveness.

Author Contributions: Conceptualization, M.M.I.; methodology, M.M.I. and K.M.R.A.; software, M.M.I.; validation, K.M.R.A. and J.U.; formal analysis, M.M.I. and K.M.R.A.; investigation, M.M.I. and K.M.R.A.; resources, M.M.I. and I.A.; data curation, M.M.I.; writing—original draft preparation, M.M.I.; writing—review and editing, M.M.I., J.U., I.A. and M.A.S.; funding acquisition, J.U. and

M.A.S.; visualization, M.M.I. and K.M.R.A.; supervision, J.U.; project administration, J.U., I.A. and M.A.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Woosong University Academic Research 2023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets utilized in this article were obtained from the Mendeley Data webpage, which is freely accessible for all scientists and investigators to conduct experiments and can be accessed at the following website: <https://data.mendeley.com/datasets/mhjym35p4/2>, (accessed on 7 August 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Life expectancy, W. Bangladesh: Oral Cancer. 2020. Available online: <https://www.worldlifeexpectancy.com/bangladesh-oral-cancer> (accessed on 10 August 2023).
- Chaturvedi, P.; Singh, A.; Chien, C.Y.; Warnakulasuriya, S. Tobacco related oral cancer. *BMJ* **2019**, *365*, l2142. [[CrossRef](#)] [[PubMed](#)]
- WCRF. Mouth and Oral Cancer Statistics. 2022. Available online: <https://www.wcrf.org/cancer-trends/mouth-and-oral-cancer-statistics/> (accessed on 15 August 2023).
- Permi, H.S.; Agrawal, T.; Sajitha, K.; Bhat, S.P.; Prasad, H.K.; Srinivas, T. A histomorphological study of benign lesions of oral cavity in a tertiary care hospital. *J. Evol. Med. Dent. Sci.* **2015**, *4*, 16495–16499. [[CrossRef](#)]
- Reid, D.E. Benign Lesions. 2022. Available online: <https://www.tricityderm.com/general-dermatology/benign-lesions/> (accessed on 15 August 2023).
- Givony, S. Oral squamous cell carcinoma (OSCC) an overview. *J. Med. Sci* **2020**, *8*, 67–74.
- Prachi, S.; Jitender, S.; Rahul, C.; Jitendra, K.; Priyanka, M.; Disha, S. Impact of oral contraceptives on periodontal health. *Afr. Health Sci.* **2019**, *19*, 1795–1800. [[CrossRef](#)] [[PubMed](#)]
- McKinney, R.; Olmo, H. Physical and chemical lesions of the oral mucosa. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2023.
- Folmsbee, J.; Liu, X.; Brandwein-Weber, M.; Doyle, S. Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 770–773.
- Gupta, R.K.; Kaur, M.; Manhas, J. Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium. *J. Multimed. Inf. Syst.* **2019**, *6*, 81–86. [[CrossRef](#)]
- Jeyaraj, P.R.; Samuel Nadar, E.R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J. Cancer Res. Clin. Oncol.* **2019**, *145*, 829–837. [[CrossRef](#)]
- Xu, S.; Liu, C.; Zong, Y.; Chen, S.; Lu, Y.; Yang, L.; Ng, E.Y.; Wang, Y.; Wang, Y.; Liu, Y.; et al. An early diagnosis of oral cancer based on three-dimensional convolutional neural networks. *IEEE Access* **2019**, *7*, 158603–158611. [[CrossRef](#)]
- Rana, A.; Yauney, G.; Wong, L.C.; Gupta, O.; Muftu, A.; Shah, P. Automated segmentation of gingival diseases from oral images. In Proceedings of the 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), Bethesda, MD, USA, 6–8 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 144–147.
- Warin, K.; Limprasert, W.; Suebnukarn, S.; Jinaporntham, S.; Jantana, P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J. Oral Pathol. Med.* **2021**, *50*, 911–918. [[CrossRef](#)]
- Prajapati, S.A.; Nagaraj, R.; Mitra, S. Classification of dental diseases using CNN and transfer learning. In Proceedings of the 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, United Arab Emirates, 11–14 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 70–74.
- Welikala, R.A.; Remagnino, P.; Lim, J.H.; Chan, C.S.; Rajendran, S.; Kallarakkal, T.G.; Zain, R.B.; Jayasinghe, R.D.; Rimal, J.; Kerr, A.R.; et al. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access* **2020**, *8*, 132677–132693. [[CrossRef](#)]
- Prabhakar, S.K.; Rajaguru, H. Performance analysis of linear layer neural networks for oral cancer classification. In Proceedings of the 2017 6th ICT International Student Project Conference (ICT-ISPC), Johor, Malaysia, 23–24 May 2017; IEEE: Piscataway, NJ, USA, 2017.
- Nanditha, B.; Geetha Kiran, A.; Dinesh, M.; Murali, S.; Chandrasheka, H. Texture analysis of color oral images for lesion detection. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 067–072.
- Thomas, B.; Kumar, V.; Saini, S. Texture analysis based segmentation and classification of oral cancer lesions in color images using ANN. In Proceedings of the 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 26–28 September 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–5.

20. Nanditha, B.R.; Geetha, A.; Chandrashekar, H.S.; Dinesh, M.S.; Murali, S. An Ensemble Deep Neural Network Approach for Oral Cancer Screening. *Int. J. Online Biomed. Eng.* **2021**, *17*, 121–134. [[CrossRef](#)]
21. Shamim, M.Z.M.; Sadatullah, S.; Shiblee, M.; Usman, M.; Ali, S.K. Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer. *Comput. J.* **2022**, *65*, 91–104. [[CrossRef](#)]
22. Wu, M.P.; Hsu, G.; Varvares, M.A.; Crowson, M.G. Predicting Progression of Oral Lesions to Malignancy Using Machine Learning. *Laryngoscope* **2022**, *133*, 1156–1162. [[CrossRef](#)] [[PubMed](#)]
23. Bansal, K.; Bathla, R.; Kumar, Y. Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer. *Soft Comput.* **2022**, *26*, 11153–11184. [[CrossRef](#)]
24. Marzouk, R.; Alabdulkreem, E.; Dhahbi, S.; Nour, M.K.; Duhayyim, M.A.; Othman, M.; Hamza, M.A.; Motwakel, A.; Yaseen, I.; Rizwanullah, M. Deep Transfer Learning Driven Oral Cancer Detection and Classification Model. *Comput. Mater. Contin.* **2022**, *73*, 3905–3920. [[CrossRef](#)]
25. Rahman, A.U.; Alqahtani, A.; Aldhafferi, N.; Nasir, M.U.; Khan, M.F.; Khan, M.A.; Mosavi, A. Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. *Sensors* **2022**, *22*, 3833. [[CrossRef](#)] [[PubMed](#)]
26. Chandrashekar, H.S.; Geetha Kiran, A.; Murali, S.; Dinesh, M.S.; Nanditha, B.R. Oral Images Dataset, Mendeley Data, V2. 2021. Available online: <https://data.mendeley.com/datasets/mhjyrn35p4/2> (accessed on 10 July 2023).
27. Islam, M.M.; Hossain, M.B.; Akhtar, M.N.; Moni, M.A.; Hasan, K.F. CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. *Algorithms* **2022**, *15*, 287. [[CrossRef](#)]
28. Islam, M.M.; Uddin, M.R.; Ferdous, M.J.; Akter, S.; Akhtar, M.N. BdSLW-11: Dataset of Bangladeshi sign language words for recognizing 11 daily useful BdSL words. *Data Brief* **2022**, *45*, 108747. [[CrossRef](#)]
29. Hossain, M.B.; Iqbal, S.H.S.; Islam, M.M.; Akhtar, M.N.; Sarker, I.H. Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. *Inform. Med. Unlocked* **2022**, *30*, 100916. [[CrossRef](#)]
30. Iman, M.; Arabnia, H.R.; Rasheed, K. A Review of Deep Transfer Learning and Recent Advancements. *Technologies* **2023**, *11*, 40. [[CrossRef](#)]
31. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:cs.CV/1409.1556.
32. Islam, M.M.; Uddin, M.R.; AKhtar, M.N.; Alam, K.R. Recognizing multiclass Static Sign Language words for deaf and dumb people of Bangladesh based on transfer learning techniques. *Inform. Med. Unlocked* **2022**, *33*, 101077. [[CrossRef](#)]
33. Wang, S.; Zheng, R.; Wang, L.; Zhang, M. A MobileNet Based Model for Tongue Shape Classification. In Proceedings of the Cognitive Systems and Information Processing: 7th International Conference, ICCSIP 2022, Fuzhou, China, 17–18 December 2022; Revised Selected Papers; Springer: Berlin/Heidelberg, Germany, 2023; pp. 605–616.
34. Matsoukas, C.; Haslum, J.F.; Sorkhei, M.; Söderberg, M.; Smith, K. What makes transfer learning work for medical images: Feature reuse & other factors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9225–9234.
35. Zabin, M.; Choi, H.J.; Islam, M.M.; Uddin, J. Impact of Tuning Parameters in Deep Convolutional Neural Network Using a Crack Image Dataset. In Proceedings of the 15th KIPS International Conference on Ubiquitous Information Technologies and Applications (CUTE 2021), Jeju, Republic of Korea, 15–17 December 2021.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.