INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 31-10-2023

Accepted: 22-01-2024

Published: 14-02-2024

Citation: Shah R, Pareek J (2024) Non-invasive Primary Screening of Oral Lesions into Binary and Multi Class using Convolutional Neural Network, Stratified K-fold Validation and Transfer Learning. Indian Journal of Science and Technology 17(7): 651-659. https://doi.org/10.17485/IJST/v17i7.2670

* **Corresponding author**.

rinkalshah@gujaratuniversity.ac.in

Funding: None

Competing Interests: None

Copyright: © 2024 Shah & Pareek. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Non-invasive Primary Screening of Oral Lesions into Binary and Multi Class using Convolutional Neural Network, Stratified K-fold Validation and Transfer Learning

Rinkal Shah[1]*, Jyoti Pareek[2]

1 Research Scholar, Department of Computer Science, Gujarat University, Ahmedabad, Gujarat, India
2 Professor and Head, Department of Computer Science, Gujarat University, Ahmedabad, Gujarat, India

Abstract

Objectives: To develop a deep learning method using camera images that can effectively detect the preliminary phase of oral cancer, which has a high rate of morbidity and mortality and is a significant public health concern. If left untreated, it can result in severe deformities and negatively affect the patient's quality of life, both physically and mentally. Early detection is crucial owing to the rapid spread of the disease, where biopsy is the only option left. Therefore, it is essential to identify malignancies swiftly to prevent disease progression non-invasively. **Methods:** Three different scenarios are used in this study to analyze samples: CNN architecture, stratified K-fold validation, and transfer learning. For automated disease identification on binary datasets (normal vs. ulcer) and multiclass datasets (normal vs. ulcer vs. Leukoplakia), camera images are pre-processed with data augmentation. As a feature extractor in the model, transfer learning is used with pre-defined networks such as VGG19, InceptionNET, EfficientNET, and MobileNET weights. **Findings:** Using the proposed CNN architecture, the F1 score for image classification was 78% and 74% for photos showing hygienic mouths or ulcers, and 83%, 87%, and 84% for images showing normal mouths, ulcers, and leukoplakia. Using stratified 3-fold validation, the results were improved to 97%, and an EfficientNET achieves the highest results in a binary F1 score of 98% and a classification with multiple classes F1 scores of 98%, 87%, and 91%, respectively. **Novelty:** Previous studies have mostly concentrated on differentiating oral potentially malignant diseases (OPMD) from oral squamous cell carcinoma (OSCC) or on discriminating between cancerous and non-cancerous tissues. The objective is to diagnose patients with non-invasive procedures to classify ulcers, healthy mouths, or precancerous type "Leukoplakia" without requiring them to visit a doctor.

Keywords: CNN; Transfer Learning; Oral Cancer; Ulcer; Leukoplakia; Stratified K-fold validation

1 Introduction

Ulcers and other types of mouth sores can develop within the mouth. Small, painful sores, called mouth ulcers, often appear on or near gum tissue. Eating, drinking, and conversing with others may be uncomfortable. Ongoing oral ulcers may occasionally be an indication of oral cancer. Oral disease-related ulcers frequently develop on or beneath the tongue; however, they can also develop elsewhere in the mouth. Because oral cancer gradually develops from mouth cancer, early identification is essential[(1)]. Initially, small lesions developed on the surface of the mouth, cheeks, or gums. This emerges because of psychological strain and minor wounds. Leukoplakia, oral squamous cell carcinoma, and other cancers eventually develop as they progress into severe wounds with large areas. A type of cancer known as leukoplakia manifests in the mouth as white or gray patches because of uncontrolled cell proliferation. These lesions are typically either precancerous or malignant. Surgery is the only option left at this stage, and it is sometimes irreparable.

To treat infections, doctors have traditionally examined patients, conducted tests, and studied clinical guidelines. However, recent developments in AI, machine learning, and deep learning have enabled computers to detect or diagnose diseases with an accuracy comparable to that of humans. Occasionally, imaging methods, such as computed tomography (CT) and magnetic resonance imaging (MRI), can be used to determine the severity of the condition. Artificial intelligence (AI) is being utilized in healthcare to diagnose oral cancer more quickly. Nowadays, medical applications are made easier due to machine learning, particularly Deep Learning (DL). Learning from incorrectly recognized photos offers a significant challenge in the field of medical image analysis that can be resolved via DL because the process of image annotation is time- and resource-intensive. For image identification and classification, deep learning typically uses convolutional neural networks (CNNs), which are a subclass of deep neural networks. However, DL requires many medical photographs to function properly. For this, transfer learning (TL) was suggested as a viable solution to this difficult problem.

Some key findings from numerous studies have used deep learning to identify oral cancers and ulcers.

Many cases are of oral potentially malignant disease (OPMD)[(2)] and oral squamous cell carcinoma (OSCC)[(3–9)], two of the most common types of malignant mucosal lesions. The detection of these cancers is from biopsy images or histopathological images[(6–8)], which are invasive in nature. Only a small number of studies have used AI algorithms to examine clinical (MRI, CT) images and assess the early stages of oral cancer[(10–13)]. Some of them have taken photographic images into consideration, but only for OSCC and other types.[(2,14)] Many researchers have concentrated on oral malignancies, either 'suspicious' or 'normal'[(15)]. But whether it is an ulcer or the beginning of oral cancer, it cannot be detected and can only be found after it has spread. To make an ordinary individual informed, there is not enough thorough research about mouth ulcers and normal mouth detection. Contrary to computer vision for natural images, there are significant issues with the absence of exact annotations and training data in the field of medical imaging because they depend on the knowledge and collaboration of medical specialists. Acquisition processes can vary based on time and money consumption. In our work, we emphasized measuring the binary and multiclass classification of normal mucosa, oral ulcer, and Leukoplakia types of cancer on camera images, which are non-invasive.

Transfer learning (TL) is the secret to the success of many DL models. These models are initially pretrained on a source dataset, and then they are fine-tuned. With the limited amount of data available, it has been found to be a successful method. Due to the challenges of collecting medical datasets, this technique is also often used in medical imaging. In this research work, we have primarily investigated these two issues, which will help patients get a primary idea of their oral disease:

- Does the patient have an ulcer or a healthy mouth?
- Does the patient have a normal mouth, leukoplakia, or an ulcer?

2 Methodology

This section highlights the prerequisites and methodology used in the classification of oral cancer types using Deep Learning.

2.1 Dataset Preparation

Determining if a person has a healthy mouth or is experiencing any abnormalities connected to the mouth, such as ulcers or leukoplakia, is the main objective of dataset construction. Following the dentist's approval, the dataset was prepared and pre-processed before the architecture was applied. Nothing more had been gathered about the patient except the pictures and their labels. Two datasets, one based on binary classification (DS1: Normal and Ulcer) and the other on multiclass classification (DS2: Normal, Ulcer, and Leukoplakia), were created after the data had been cleaned. Table 1 displays the picture bifurcation for both datasets. Figure 1 displays the sample image.

Table 1. Image distribution for Binary and Multi-Class classification

	DS1 (Binary Classification)		DS2 (Multi-class Classification)		
	Normal	**Ulcer**	**Normal**	**Ulcer**	**Leukoplakia**
Train	109	103	109	103	108
Validation	45	45	40	40	45
Test	25	25	20	20	20
Total images	**352**		**505**		

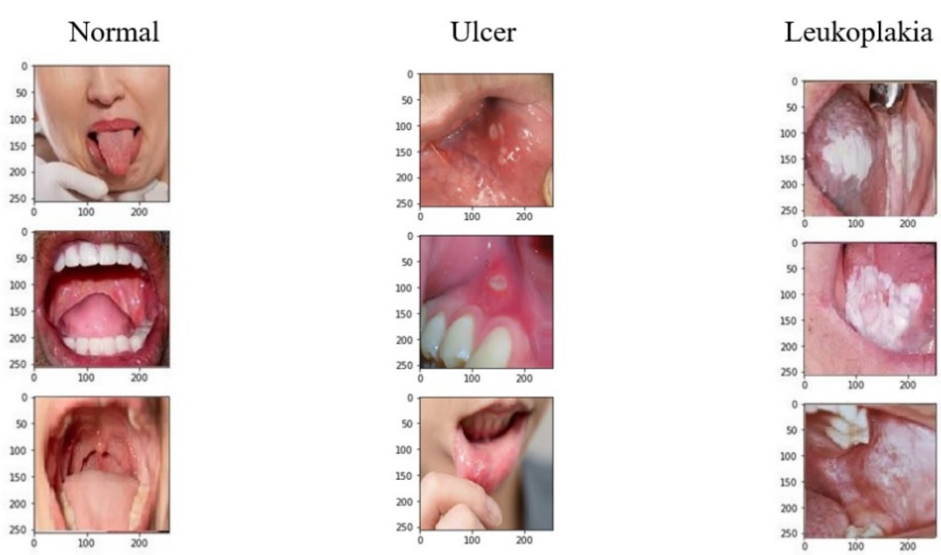


Fig 1. Sample images of Normal mouth, ulcer, and Leukoplakia type of cancer

2.2 Data Augmentation and image pre-processing

It is thought that giving the model a range of images will improve its capacity for learning and aid in its successful generalization. The transformation feature included in the Karas package is the ImageDataGenerator function. To determine the true performance of the models, the test data were not changed. Deep learning algorithms benefit from the feature scaling technique since it accelerates gradient descent convergence. A large range of values, from 0 to 255, are possible for each pixel in the input photo feature. After applying min-max normalization, each pixel is divided by 255 and rescaled between 0 and 1. To get over the issue of low contrast images in the collection, rescaling is done. After the photographs were reduced in size to 256 x 256, roughly 10,000 images were created by using data augmentation elements such as rotation, scaling, zooming, shearing, flips, width shifts, and height shifts. Photos with a perfect match of values were fed into the CNN model after these parameters were adjusted.

2.3 Development of CNN architecture and Stratified K-Fold validation

CNN is frequently utilized in healthcare diagnosis and its applications to simplify medical services. In this study, trials were carried out using the Google Collab platform with a GPU. TensorFlow is the framework, and Python 3:7 is the software programming environment that is utilized. As backend libraries for a deep learning framework, Karas and TensorFlow were used. We have developed a multi-layer CNN model that uses filters to perform convolutional operations based on how an input image is scanned. Since it is difficult to connect hundreds of neurons with different levels of the CNN structure, the image size must be kept to a minimum. Photos are pre-processed with data augmentation. The architecture (Figure 2) consists of three convolution layers with two 32, one 64, and one 128 output channels using the same kernel (3,3) and pool size (2,2) utilizing the maxpooling layer. A flattening layer with a 0.5 dropout is used to prevent overfitting during training. The installation of a dense layer utilizing the SoftMax activation function is the last phase. There are 50 iterations carried out, with a batch size of 64.

The following settings have been established for the other CNN model for multi-class classification: With ImageDataGenerator, the pre-processing has been carried out with the settings for shearing, zooming, horizontal flipping, width, and height shift, as well as rescaling. With two layers of 32 filters, two additional layers of 64 filters, and (2,2) poolsize, the multi-layer CNN architecture is created with an input size of 256*256. The last layer is a "softmax" activation function with a dense layer, and the dropout utilized is 0.5 as displayed in Figure 2. The number of iterations used were 50.

Threefold stratified cross-validation was used in another experiment. By keeping the percentage of samples for each class, the samples are reorganized using the stratification technique so that every fold has a consistent representation of the entire dataset. Six convolution layers, three fully connected layers, and a flattening layer make up the model's architecture. "Relu" was the activation function that was employed, while "Adam" was the optimizer with 50 epoches. The same strategy was implied for both types of classification.

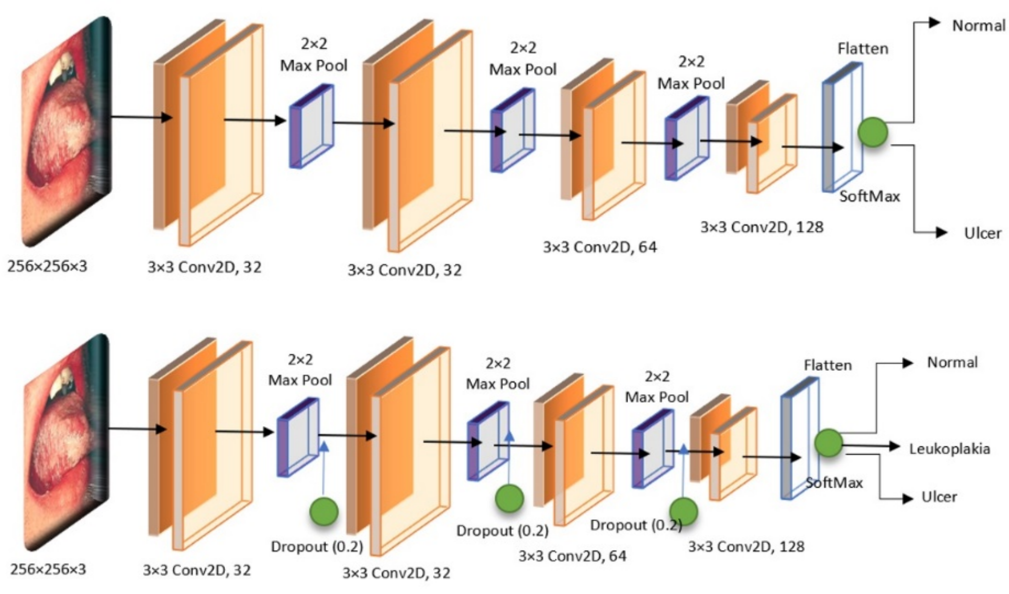


Fig 2. CNN architecture for Binary & Multi class classification

3 Results and Discussion

Three experimental scenarios have been built to compare the performance of customized networks. A customized CNN was used in the first instance, and the outcomes are explained. Higher results were obtained in the second case with the application of stratified k fold validation. Pre-defined architectures have been used in the third instance.

3.1 Binary and multi class classification results using CNN

Confusion matrix, Precision, Recall, and F1 score are the evaluation matrices that we utilised to evaluate the findings. With 50 epochs and the CNN architecture for binary classification, the results were 85.71% for training accuracy, 81.11% for validation accuracy, 76% for testing accuracy, and 78% F1 score to distinguish between a normal mouth and one that has an ulcer. Figure 3 shows the confusion matrix for the same. Findings of multiclass classification differ from binary findings. The results obtained with the adjustment of the hyperparameters are 80% training accuracy, 81.6% validation accuracy, and 85% testing accuracy on 50 epochs. The corresponding F1 score is 83%, 87% and 84% for Normal mouth, Ulceration and Leukoplakia.

3.2 Binary and multi class classification results using Stratified K-Fold validation

Results of applying stratified K Fold validation to binary and multiclass classification are reported. Consideration has been given to K= {1,2,3}. F1 score was approximately 97% after 50 iterations, with 100% training, 97.26% validation, and 86.99% testing accuracy recorded. The results showed that the classification accuracy for normal, ulcer, and leukoplakia was 98.91% during training, 94.11% during validation, and 81% during testing, with an F1 score range of 89%-98%.

3.3 Binary and multi class classification results Transfer Learning

Knowledge transfer from one domain to another is known as transfer learning. Deep learning requires a lot of data to learn certain patterns, making it challenging to train and time-consuming, especially in the field of medical imaging. The weights of pre-trained models are changed to address this issue, minimize training time, and accommodate sparse data. Once the top layers are frozen, fully connected layers are fine-tuned for categorization. In both types of challenges, we used transfer learning with VGG19, EfficientNET, InceptionNET, and MobileNET as pre-trained feature vectors. The pre-trained model can be used as a stand-alone feature extraction program, with the input pre-processed and used as an output for the new training model. Transfer learning is implemented as a feature extractor preprocessor in accordance with the data augmentation specifications in ”VGG19”, ”InceptionNet”, ”EfficientNet”, and ”MobileNet”. The images are pre-processed using ImageDataGenerator with some parameters like rotation, zoom, shift, and shear range, as well as horizontal flips, and pre-trained model weights by specifying the include_top argument as “false” with an input shape of 256 by 256 image size. While making trainable false, the model includes a flattening layer and a dense layer at the end, with the activation function ”SoftMax” or ”Relu.” The model is then compiled with binary or categorical cross-entropy. The optimizers are used as “Rmsprop” or “Adam”. Their hyperparameters are chosen depending on the highest accuracies obtained. The model is then trained over 50 epochs, and model accuracies and losses are calculated using a Matplotlib function graph and a confusion matrix. Each testing input is predicted via a bar chart displaying its expected and true values. (Figure 5)

3.3.1 Binary-classification results

Table 3 shows the transfer learning performance using several models. With 100% accuracy for training and 98% accuracy for testing, VGG19 and EfficientNET offer the best accuracy we have been able to manage. Recall and precision were 100% in certain instances. Table 3 displays the results of all the experiments that were conducted on 50 distinct epochs. Using different preconfigured networks results in varying processing durations, with VGG19 taking the longest. Together with recall, precision, and F1-score, Table 3 displays the accuracy of the training, validation, and testing phases. Considering this, Figure 3 displays the confusion matrix for the optimal result determined by using the test set's images. Our F1-score for binary classification ranges from 96% to 98%, while EfficientNET produces results that are compatible when accuracy and training time are considered. The Guo J et al. research's transfer learning results to distinguish between normal and ulcer photos are approximately 95%, which is less than the results achieved here using TL.[16] on addition, we obtained 100% training accuracy and a balanced dataset on our instance. Using ResNet50, Mimi Zhou et al. obtained an F1 score of 92.24%, recall of 91.84%, and precision of 92.86%, which is still less than what we obtained.[17]

3.3.2 Multi-class classification results

CNN was used to assess accuracy and recall for the categorization within "Normal," "Ulcer," and "Leukoplakia," and the F1-score was utilized to obtain the expected result. Additionally, the quick and precise pre-defined networks were used. The highest training accuracy was achieved by the architecture using the MobileNET model for transfer learning. It trains the model accurately. When testing parameters are examined, EfficientNET produces the best results. Leukoplakia detection (Recall) increased from 68% to 96% when TL was employed for multi-class classification; normal mouth prediction remained equal, while ulcer prediction declined by 8%. This might happen as a result of the regular white tissues seen in a typical mouth. Due to the excessive tissue development and white patches associated with inflammatory conditions, the model may be mistaken for an ulcer. Examining the precision and recall yields a completely positive result in terms of determining the actual result. The effectiveness of multi-class classification is evaluated using the confusion matrix (Figure 4). The code shown in Figure 5 implements a scaling technique, from which the prediction results are measured.

Oral leukoplakia may result from the ulcer if it swells and spreads. Most of the research on oral cancer is restricted to OSCC and OPMD lesions and focuses on establishing the diagnosis of malignancy.[2–6] This study suggests that patients may undertake initial screening on their own and can determine when they will need to see their doctor again.

Table 2. Results based on CNN and stratified 3-fold validation

Type	K-fold	Epochs	Training Accuracy	Validation Accuracy	Testing Accuracy	Type	Precision	Recall	F1-score
Results based on CNN architecture									
Binary Classification	-	50	85.71%	81.11%	76%	Normal	72%	84%	78%
						Ulcer	81%	68%	74%
Multi-class classification	-	50	80%	81.6%	85%	Normal	94%	75%	83%
						Ulcer	89%	85%	87%
						Leukoplakia	77%	92%	84%
Results based on Stratified k fold validation techniques									
Binary Classification	1	50	93.88%	81.08%	86.99%	Normal	95%	75%	84%
						Ulcer	67%	92%	77%
	2	50	99.32%	91.89%		Normal	89%	94%	92%
						Ulcer	95%	90%	92%
	3	50	100%	97.26%		Normal	97%	97%	97%
						Ulcer	97%	97%	97%
Multi-class classification	1	50	87.55%	71.53%	81%	Normal	77%	66%	71%
						Ulcer	73%	70%	72%
						Leukoplakia	64%	82%	72%
	2	50	97.07%	86.13%		Normal	85%	82%	84%
						Ulcer	82%	88%	85%
						Leukoplakia	91%	89%	90%
	3	50	98.91%	94.11%		Normal	89%	95%	92%
						Ulcer	98%	96%	97%
						Leukoplakia	95%	91%	93%

Table 3. Results of Transfer Learning implementations

Pre-defined network	Epochs	Training accuracy	Validation accuracy	Testing accuracy	Type	Precision	recall	F1 score
Binary Classification								
VGG19	50	100%	97.78%	98%	Normal	96%	100%	98%
					Ulcer	100%	96%	98%
EfficientNET	50	100%	98.89%	98%	Normal	100%	95%	98%
					Ulcer	96%	100%	98%
InceptionNET	50	97.7%	96.67%	92%	Normal	95%	95%	95%

Continued on next page

Table 3 continued

					Ulcer	96%	96%	96%
MobileNET	50	100%	95.56%	96%	Normal	96%	96%	96%
					Ulcer	96%	96%	96%
Multi-class classification								
VGG19	50	96%	87.41%	89%	Normal	96%	100%	98%
					Ulcer	90%	76%	83%
					Leukoplakia	82%	92%	87%
EfficientNET	50	96.31%	91.11%	92%	Normal	96%	100%	98%
					Ulcer	95%	80%	87%
					Leukoplakia	86%	96%	91%
InceptionNET	50	92.62%	85.93%	80%	Normal	75%	72%	73%
					Ulcer	75%	96%	84%
					Leukoplakia	95%	72%	82%
MobileNET	50	98.15%	90.37%	92%	Normal	96%	92%	94%
					Ulcer	88%	92%	90%
					Leukoplakia	92%	92%	92%

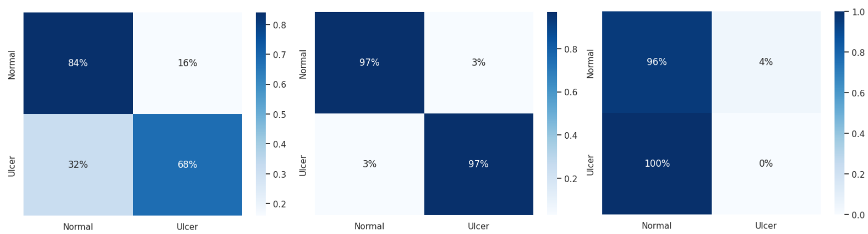


Fig 3. Confusion matrix for binary classification of CNN, 3-fold & Transfer learning

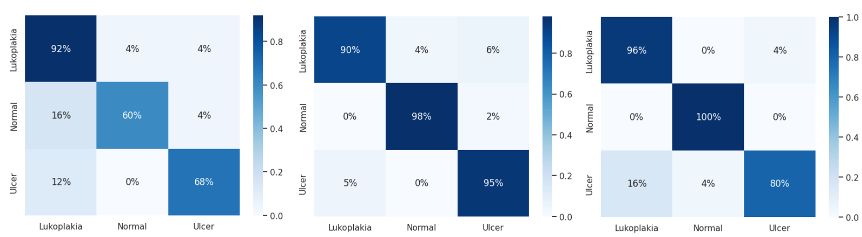


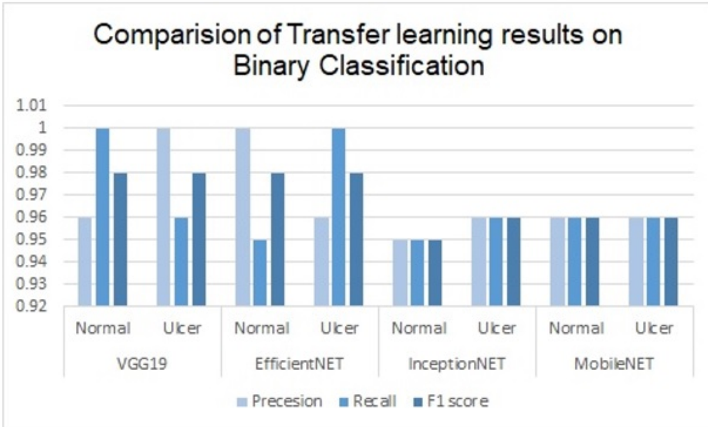
Fig 4. Confusion matrix for Multi class classification of CNN, 3-fold & Transfer learning

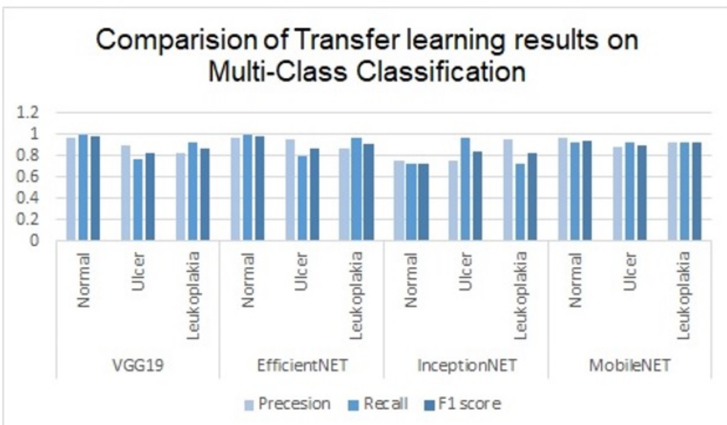
4 Limitations

Obtaining real-time photos is the study's main constraint. Many clinicians were contacted for real-time photos, but the primary limitation was the data gathering process due to concerns about medical ethics and patient privacy. As a result, the dataset utilized in this study is quite little, and it will require the use of various strategies to make it larger. Moreover, additional tagged classes, such as OSCC, OPMD, Erythroplakia, and others, were not employed in our work. An additional constraint is that we were only able to examine a single category of precancerous lesion. Future developments in the field of artificial intelligence may enable the treatment of additional cancer kinds as outlined in section one using more sophisticated methods.

5 Future Developments

Many people remain undiagnosed because of the lack of effective diagnostic tools and professionals in many places, especially in developing and/or remote regions. Consequently, many cancer patients went untreated. Clinical applications that enable patients to receive preliminary recommendations can be developed with the aid of this research. It is possible to design software





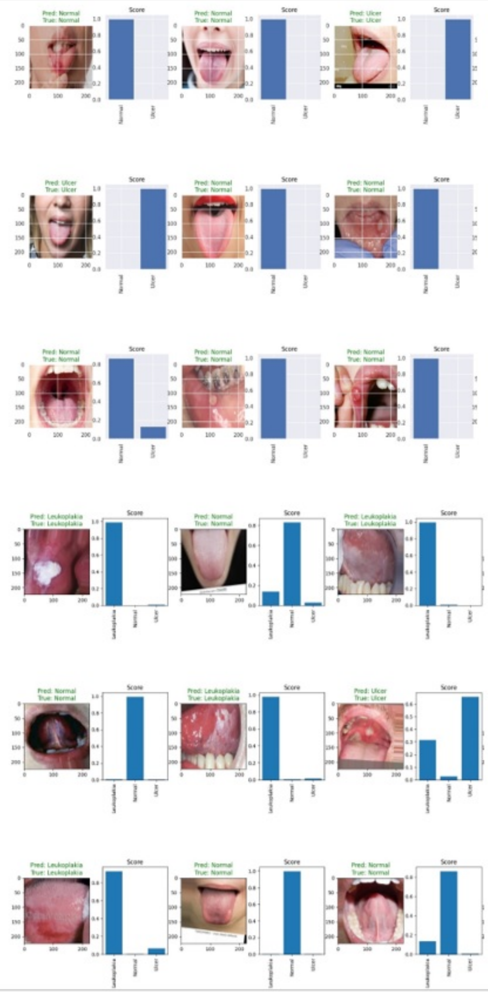


Fig 5. Prediction Result of each image

that will be useful in distant areas with a shortage of medical experts or for those just starting out in their careers. These forecasts can assist citizens in determining whether grown cells are a harmless patch or require more medical care. The development of such an application, which will be extremely beneficial in the field of medical prognosis, is encouraged by this study for future researchers.

6 Conclusion

This study addresses the ambiguity in oral cytology in the context of medical diagnosis, which can be overcome using camera images, and shows how Transfer Learning and Deep Learning can help to eliminate it. The experiment's findings demonstrated the importance of detecting oral cancers as soon as possible without invasive procedures. As a result of our observations, we can now predict how different disease groups will be distinguished on a primary basis without the need for a doctor's visit. Furthermore, using transfer learning can improve accuracy when diagnosing oral ulcer and pre-cancerous conditions although, collecting larger dataset from hospitals is the major limitation. With such research, a strong diagnostic system to detect oral cancer in early screening prior to biopsy could be built in the future.

References

1) Mouth ulcer. 2023. Available from: https://www.nhsinform.scot/illnesses-and-conditions/mouth/mouth-ulcer.

2) Tanriver G, Tekkesin MS, Ergen O. Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders. *Cancers*. 2021;13(11):1–13. Available from: https://doi.org/10.3390/cancers13112766.
3) Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *eClinicalMedicine*. 2020;27:1–7. Available from: https://doi.org/10.1016/j.eclinm.2020.100558.
4) Welikala R, Remagnino P, Lim J, Chan C, Rajendran S, Kallarakkal T, et al. Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer. *IEEE Access*. 2020;8:132677–132693. Available from: https://doi.org/10.1109/ACCESS.2020.3010180.
5) Shamim MZM, Syed S, Shiblee M, Usman M, Ali SJ, Hussein HS, et al. Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer. *The Computer Journal*. 2022;65(1):91–104. Available from: https://doi.org/10.1093/comjnl/bxaa136.
6) Panigrahi S, Nanda BS, Bhuyan R, Kumar K, Ghosh S, Swarnkar T. Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. *Heliyon*. 2023;9(3):1–14. Available from: https://doi.org/10.1016/j.heliyon.2023.e13444.
7) Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Networks*. 2020;128:47–60. Available from: https://doi.org/10.1016/j.neunet.2020.05.003.
8) Das M, Dash R, Mishra SK. Automatic Detection of Oral Squamous Cell Carcinoma from Histopathological Images of Oral Mucosa Using Deep Convolutional Neural Network. *International Journal of Environmental Research and Public Health*. 2023;20(3):1–21. Available from: https://doi.org/10.3390/ijerph20032131.
9) Badawy M, Balaha HM, Maklad AS, Almars AM, Elhosseini MA. Revolutionizing Oral Cancer Detection: An Approach Using Aquila and Gorilla Algorithms Optimized Transfer Learning-Based CNNs. *Biomimetics*. 2023;8(6):1–35. Available from: https://doi.org/10.3390/biomimetics8060499.
10) Xu S, Liu Y, Hu W, Zhang C, Liu C, Zong Y, et al. An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks. *IEEE Access*. 2019;7:158603–158611. Available from: https://doi.org/10.1109/ACCESS.2019.2950286.
11) Bhandari B, Alsadoon A, Prasad PWC, Abdullah S, Haddad S. Deep learning neural network for texture feature extraction in oral cancer: enhanced loss function. *Multimedia Tools and Applications*. 2020;79(37-38):27867–27890. Available from: https://doi.org/10.1007/s11042-020-09384-6.
12) Zhong Y, Yang Y, Fang Y, Wang J, Hu W. A Preliminary Experience of Implementing Deep-Learning Based Auto-Segmentation in Head and Neck Cancer: A Study on Real-World Clinical Cases. *Frontiers in Oncology*. 2021;11:1–9. Available from: https://doi.org/10.3389/fonc.2021.638197.
13) Sumida I, Magome T, Kitamori H, Das IJ, Yamaguchi H, Kizaki H, et al. Deep convolutional neural network for reduction of contrast-enhanced region on CT images. *Journal of Radiation Research*. 2019;60(5):586–594. Available from: https://doi.org/10.1093/jrr/rrz030.
14) Shah R, Pareek J. Pretreatment Identification of Oral Leukoplakia and Oral Erythroplakia Metastasis Using Deep Learning Neural Networks. In: International Conference on Computer Vision and Image Processing- CVIP 2021;vol. 1567 of Communications in Computer and Information Science. Springer, Cham. 2022;p. 306–315. Available from: https://doi.org/10.1007/978-3-031-11346-8_27.
15) Islam MM, Alam KMR, Uddin J, Ashraf I, Samad MA. Benign and Malignant Oral Lesion Image Classification Using Fine-Tuned Transfer Learning Techniques. *Diagnostics*. 2023;13(21):1–15. Available from: https://doi.org/10.3390/diagnostics13213360.
16) Guo J, Wang H, Xue X, Li M, Ma Z. Real-time classification on oral ulcer images with residual network and image enhancement. *IET Image Processing*. 2022;16(3):641–646. Available from: https://doi.org/10.1049/ipr2.12144.
17) Zhou M, Jie W, Tang F, Zhang S, Mao Q, Liu C, et al. Deep learning algorithms for classification and detection of recurrent aphthous ulcerations using oral clinical photographic images. *Journal of Dental Sciences*. 2024;19(1):254–260. Available from: https://doi.org/10.1016/j.jds.2023.04.022.