International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.13, No.1 (Apr-23)

http://dx.doi.org/10.12785/ijcds/130170

Deep Learning Model for Automatic Detection of Oral squamous cell carcinoma (OSCC) using Histopathological Images

Sayyada Hajera Begum[1] **and P Vidyullatha**[2]

[1]*Research Scholar,Department of Computer Science and Engineering,Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*

[2]*Associate Professor,Department of Computer Science and Engineering,Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*

Received 2 May. 2022, Revised 21 Dec. 2022, Accepted 6 Feb. 2023, Published 16 Apr. 2023

Abstract: Oral squamous cell carcinoma (OSCC), is a type of cancer that causes the loss of the structural formation of layers and membranes in the oral cavity region. With the recent advent of Deep learning (DL) in biomedical image classification, the automated early diagnosis of oral histopathological images can aid in effective treatment of oral cancer. This work attempts to perform an automated classification of benign and malignant oral biopsy histopathological images by implementing a DL-based convolutional neural network (CNN) model for the initial analysis of OSCC. For this research, four recently developed candidate pre-trained DL-CNN models namely NASNetLarge, InceptionNet, Xception, and DenseNet201 are selected through the approach of transfer learning. These pre-trained models are then modified with additional layers for effective OSCC detection. The efficacy of these modified models is examined on an oral cancer histopathological image database. It is examined that the pre-trained DenseNet201 model with modified structure has surpassed other models in terms of performance parameters by recording an accuracy of 91.25% and is considered as our proposed DL-CNN model.

Keywords: oral cancer detection, Oral squamous cell carcinoma (OSCC), Deep Learning (DL), Convolutional Neural Network (CNN).

1. INTRODUCTION

The rate of oral cancer is known to be highest worldwide and the incidence is lower in women compared to men and nearly 660,000 new incidences of oral cancer are reported each year and more than 340,000 deaths worldwide due to lack of timely diagnosis. In oral cancer, the cancerous tissues can be located in the lips, oral cavity, and pharynx and causes the loss of the structural formation of layers and membranes in the oral cavity region.

Oral cancers are classified into OSCC, salivary gland carcinoma, verrucous carcinoma, and lymphoepithelial carcinoma. The majority of the carcinomas are due to OSCC [1], [2].

Despite applying various treatment modalities, the total mortality rate of OSCC is not declined significantly which is only due to lack of efforts for early detection and diagnosis. The physicians examine the presence of any suspicious lesion which can be cancerous and suggests for biopsy. Slides with the biopsy sections are observed for any deformities which are different from usual cell arrangements like size and shape using microscope [3]. At the histopathological level, malignant squamous cells are bigger compared to the normal cells and are particularly different from each other in shapes. A confirmatory diagnosis of oral cancer from this report is needed to be done by a highly qualified and experienced specialist which is very vital and needs to be accurate [3]. However, the entire manual data interpretation of the cancerous slide is too time-consuming and at the same time is prone to human errors [4].

Because of the above-mentioned reasons, computer-aided diagnostic (CAD) techniques may assist the physicians in reducing both time and bias with improved efficiency in the analysis of the features. The intention is to discover cancer at an early stage which will lead to early treatment, which lowers the risk of morbidity and mortality. Moreover, the oral diagnosis CAD systems will reduce the volume of load in the laboratories and most of the cases may be benign, the pathologist may focus more on malignant cases [5] .

In the development of CAD systems, biomedical imag-

E-mail address: sayyada.hajera07@gmail.com

ing data is widely accepted in modern medicine due to its benefit for disease diagnosis, treatment schedule, and required treatment. It also aids in collecting noninvasive potentially informative details in form of patient explicit disease characteristics. Such imaging data is also rapidly increasing due to the application of advanced hardware, low cost, and increase in population.

2. RELATED WORK

Current developments in artificial intelligence (AI) have started infusing into the healthcare sector. Among these DL techniques (DLTs), the CNN rose to recognition due to its high accuracy for image classification specifically the texture classification tasks. Based on DL, various methods have been proposed and developed on medical data such as breast cancer [5] , lung cancer [6] , and even for covid-19 detection [7]. The DLTs have been proved to offer improved accuracy, specificity, and sensitivity [7]. Moreover, the transfer learning approach is also widely accepted for medical image classification which improves the results in applied DLT [7]. Current research work has also proved the effectiveness of DLTs in the classification of oral lesions from medical images including histopathological or real-time oral cavity images [8] .

Numerous researchers have focused their studies to apply DLTs to detect oral cancer from the histopathological images. This also motivates us to consider the potential of DL to extract the classification features from oral cavity suspicious lesions for early detection of OSCC from histopathological images.

In recent studies, G. Forslid et al [8] proposed that DLTs can be used for the early detection and diagnosis of oral and cervical cancer detection. The experiments results are then evaluated for VGG-16 [9] and ResNet-50 [10]. The authors reported accuracy within a range of 78-82% dependent on the dataset and the model applied. The results specify a high potential for detecting aberrations in the oral cavity.

Fu et al. [11] applied the cascaded DL to classify SCC from 44,409 total biopsy-proven SCC photographic images and normal clinical images. The applied DLT achieved a specificity of 88.70%. and sensitivity of 94.90%.

Das et al. [12] applied the DL to classify OSCC into its four classes first through the transfer learning approach and utilized pre-trained models such as VGG-16 [9] , VGG-19 [9], and Resnet-50 [10] and obtained highest classification accuracy of 92.15% with ResNet-50. Later the proposed CNN model based on VGG-19 architecture is applied to achieve a higher classification accuracy of 97.50%.

Tanriver et al. [13] explored the possible application of DLT for detecting oral malignant disorders (OMD) by proposing a two-stage model to detect oral lesions and classify them into three classes benign, OMD, and carcinoma. The photographic oral dataset with lesions was collected from the department of Tumor Pathology department, Oncology Institute at Istanbul University. The authors reported that the EfficientNet-B7 model achieved the maximum accuracy of 92.90% considering semantic segmentation.

Welikala et al. [14] applied the model ResNet-101 and Fast R-CNN for the classification of OSCC from bounding box annotated images of the oral cavity. The authors testified F1 score of 87.07% for OSCC identification. The authors demonstrated the potential of the DLT for the early detection of oral cancer.

The authors demonstrated the efficacy of DLT and utilized six models using the transfer learning approach to classify pre-cancerous oral lesions from annotated images and detected the initial stage of oral cancer [15]. The authors then reported classification accuracy of 98.00% with VGG-19 and 97.00% with ResNet50 in distinguishing five forms of oral lesions mainly the toung lesions. The results were demonstrated to achieve near-human-level performance for the detection of early-stage oral cancer.

The authors in [16] proposed a new structure of regression-based segregation with DLT on hyperspectral cancerous images and exhibited comparison with other techniques in terms of accuracy, specificity, sensitivity, and reported an accuracy of 91.40% to classify malignancies.

Figueroa et al. [17] adopted the Grad-CAM method of [18] to insert interpretability and applied the GAIN [19] architecture rather than using the simple DL-CNN for classification. The authors optimally linked the GAIN classification and attention map in an endwise mode. The authors used the VGG-19 as the base CNN for training whose output was further passed through the GAIN and an accuracy of 86.38% was estimated.

Xu et al. [20] constructed a 3-D CNN to profile initial stage oral tumors as benign and malignant. A comparison is carried out with conventional 2-D DL-CNN-based techniques and reported better performance. This technique may assist in the designing of CT- based diagnosis systems using 3-D DL-CNN models in the future.

Gupta et al. [21] exploited a dataset of biopsy slides of epithelial squamous tissues. A total of 2688 images were generated with augmentation and pre-processing and supplied to DL-CNN. A training accuracy of 91.65% and testing accuracy of 89.30% have been noticed for the projected system.

Song et al. [22] developed a portable smartphone-based oral inspecting device and demonstrated the effectiveness of of DLTs for dual-modal image classification. An image classification algorithm was presented which uses a fusion of white light and fluorescence images which is fed to DL-CNN. The authors reported a validation accuracy of 86.90% with the VGG-CNN-M network.

Alabi et al. [23] presented a review of DL-CNN applied for the prognosis of OSCC. The DLTs were applied for various medical data such as histopathological, clinic pathological, Raman spectroscopy data, gene expression, and CT images. The review explained that new imaging modalities such as CT or enhanced CT and spectra data could also exhibit significant outcomes.

Santisudha Panigrahi and Tripti Swarnkar [24] reviewed various DLTs for the oral dataset of histopathological images and provided a comparison of various strategies adopted while implementing DL-CNN models for the prognosis of early-stage oral cancer.

Shaban et al. [25] proposed a different technique for the objective calculation of tumor-infiltrating lymphocytes (TILs) profusely present in OSCC images. The TIL value is computed by first partitioning the full OSCC image into primary tissue types such as a tumor, lymphocytes, etc., and later quantifying the location of tumor and lymphocytes regions. The proposed DL technique achieved high accuracy of 96.31%.

Fujima [26] proposed to use the F-fluorodeoxyglucose PET images to predict the infection-free sustenance with OSCC. The ResNet-101 network is applied to FDG-PET images to diagnose parameters such as as h-index, metabolic tumor volume, and overall lesion glycolysis. The highest accuracy of 80% was attained by applying the DL classification.

Das DK et al. [27] proposed to determine the existence of variation in epithelial layers and the keratin pearls from histopathological images. The authors applied a 12-layered (7 × 7×3 channel patches) DL-CNN for segmentation of oral integral layers to detect the keratin pearls from the tissue regions. The proposed method achieved an accuracy of 96.88% for keratin pearls detection.

Chan et al. [28] proposed an innovative DL-CNN method using a texture map of OSCC detection. The network comprises two collective layers, a lower layer to perform segmentation and ROI marking and an upper layer to perform oral cancer detection. ROI marking makes the OSCC regions clearer. The texture maps are computed from the standard deviation of sliding windows. This texture map data is fed as input to the DL-CNN and specificity of 71.29% and sensitivity of 96.87% are reported.

Nandita et al. [29] developed an ensemble DL-CNN model by combining the advantages of Resnet-50 and VGG-16. This ensemble model is trained with a dataset of augmented oral lesion images and 96.20% accuracy was estimated which outperformed other eminent DL-CNN models in OSCC classification.

The authors developed a lightweight DL-CNN via the transfer learning approach [30] aand used EfficientNet-B0 to perform binary classification of 716 real-time clinical images into potentially malignant or benign images. The proposed DL-CNN model attained 85.0% accuracy.

Thus, concerning the above-related work, at histopathological levels, malignant OSC cells are bigger compared to normal cells and varies from one another in their shapes. Confirmatory identification of oral cancer is done by a much skilled and qualified individual. Thus, automating this process can significantly ease the burden of specialists. Few studies have been reported for OSCC prognosis by using DLT at the histopathological levels.

Towards this goal, we proposed to investigate the advantage of DLTs for the early detection of OSCC. For this work, we considered four candidate pre-trained DL-CNN architectures in the framework of transfer learning and modified them with additional layers to effectively identify specific visual patterns of the oral cavity at histopathological levels affected by cellular changes due to cancer. The four pre-trained DL-CNN models such as InceptionV3 [31], NASNetLarge [32] , Xception [33], and DensNet201 [34] are applied to the histopathological dataset and analyzed their performance

In particular, we classified histopathological images into benign and malignant classes using a DL-CNN-based classification. We also studied the analysis of the model performance in detail using various metrics by applying transfer learning and data pre-processing to find the best performing DL-CNN model from the four candidate models. Later, the selected best model is considered as the proposed model for further analysis and comparison.

The additional part of the paper is arranged as follows. Section 3 explores the details of the database utilized, model formulation, and the OSCC detection process. Section 4 explains the experimental work conducted on four candidates' modified DL-CNN models and proposes the best suitable DL-CNN model for the OSCC detection task. Finally, in section 5, the paper is concluded.

3. MATERIALS AND METHODS

The next section outlines the dataset selection, briefs the theory of transfer learning, and also provides the details of model formulation and the proposed model modification for the four candidate pre-trained DL-CNN models.

A. *Dataset*

Computational approaches including the DLTs are applied effectively to obtain the solution for OSCC detection. The effectiveness of the detection is dependent on the dataset applied. There are few publicly available datasets as such were published by Tabassum et al. [35]. The dataset contains 1224 oral histopathological images out of which 934 are cancerous and 290 are non-cancerous. The histopathological images of the dataset used are taken from biopsy slides and examined using various cytological measures under a microscope. Thus, the dataset used is clinically proven and can be used for working with various DL

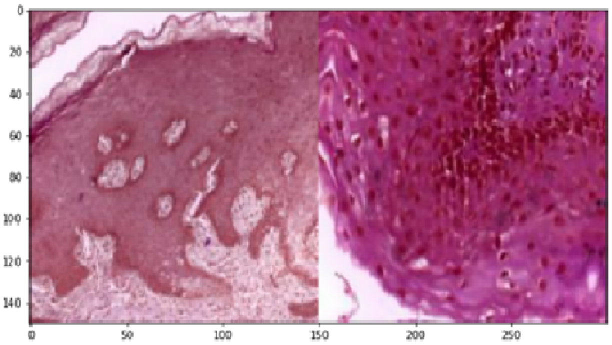

Figure 1. Sample Histopathological Images with Benign Tissue (Left), Malignant Tissue (Right)

models. To demonstrate the efficacy of the proposed work on a larger dataset, we considered images from different sources and combined the images from the dataset [35] with images from [36]. This in turn creates a large dataset of 2000 images in total for benign and malignant lesion images. We randomly split this data into a ratio of 80:20 of validation and testing datasets with each comprising similar class distribution of benign and malignant images.

Moreover, DL-CNN models are computationally expensive and requires all the input images to be of identical size. To bring uniformity in the size and shape of images, we resized the images to the same dimension of 224×224×3. The images are further normalized with the range between 0 and 1 to avoid the difference in magnitude of various pixels which will aid the deep learning. To also avoid the data imbalance which affects the generalization ability of the model we have also applied the data augmentation technique such as vertical-flip and horizontal-flip. The difference between a sample histopathological image of benign and malignant tissue infected with oral cancer is depicted in Figure 1.

B. Transfer Learning using Pre-trained DL-CNN Models

Transfer learning applies knowledge gained by a model from one task to another relevant task. Thus, instead of training a model from scratch, the information gained by the model previously is adjusted to the new problem. Transfer learning reduces the training time of the model and also enables it to work with small data. Models are usually trained on freely available datasets like ImageNet, CIFAR etc.

DL-CNN models have considerably enhanced the contemporary techniques applied in many image-based issues such as object detection and recognition. CNN is a type of DL network comprising of architecture where one layer is connected to the subsequent layer [37]. The layers are constructed by neurons and the spatial architecture of a layer creates a volume of these neurons with a width, height, and depth. The width and height determine the size of the neuron and the depth determines the number of neurons. The depth of the network can be understood in terms of the number of stacked layers in the whole network. The architecture of a CNN varies depending on the usage the architect chooses on endless combinations of layers and constructs each layer in endless ways. The most significant layers are convolution, pooling, and fully connected. Other layers such as ReLU, batch normalization, and dropout layers are also making the DL-CNN model complete alongside input and output layers as depicted in Figure 2.

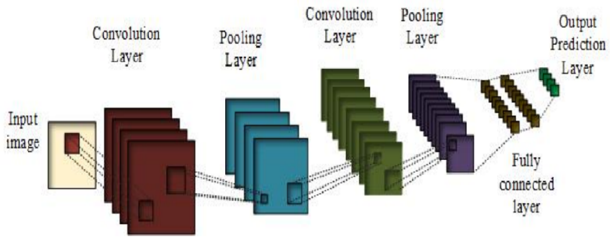

Figure 2. Generalized Structure of DL-CNN

These layers facilitates effectual learning of features from the input images. When the input image is fed to a typical CNN, the convolution layers which are made up of various filters with a width, height, and depth extract different kinds of features. The width and height determine the size of the filter kernel and the depth is the number of kernels. Each kernel is built up by learnable parameters which are convolved over the input Image and perform a dot product to extract features. The convolutional layer also has parameters such as size, stride, and padding. The stride determines how many steps the kernel takes before performing a convolution operation. The padding controls the size of output from the layer and boundary pixels. The extracted features are further provided as input to the pooling layers for further efficient processing. The feature map produced from the convolution layer is still large and needs to be reduced. The pooling layers perform operations similar to the convolution layer but it serves to reduce the feature map. Average pooling layers and max-pooling layers are most frequently applied. Thus, by reducing the feature map size the CNN becomes less computationally challenging. Later, the Batch normalization layer along with ReLU normalizes the shifts in the middle layers thus allowing better network convergence. Dropout layers helps in avoiding overfitting of the model. Finally, the reduced feature map is forwarded towards the fully connected layer with the SoftMax function to perform classification into corresponding classes.

There are few famous pre-trained DL-CNN models available for image classification. They are VGG-16 [9], ResNet50 [10], Inception-V3 [31], NASNetLarge [32] , Xception [33], and DensNet201 [34]. With the help of transfer learning, the modified DL-CNN models are also able to demonstrate a strong ability to generalize the images external to the ImageNet dataset. Among these, we considered Inception-V3, NASNetLarge, Xception, and

DensNet201 as four candidates pre-trained models and modified them with additional layers for effective OSCC detection. Table 1 depicts the various CNN models and their parameter specifications.

1) InceptionV3

InceptionV3 is an improvement to the previous Inception CNNs. It concentrates on being more computationally efficient by incorporating smaller and factorized convolution filters termed inception modules. These modules are built to handle the issues due to computational cost, and overfitting. Similarly, the filter banks are broadened to deal with the representational block. To do this, inception model applies filters of different sizes and later combines all the outputs at the end.

2) NASNetLarge

Neural Search Architecture network NASNet [32] was created using a neural architectural search procedure that uses reinforcement learning and a control neural net to find the best CNN model. The parent procedure improves the efficiency of the model by making modifications based on the number of layers, weights, regularization methods, etc. The NASNet architecture is trained with two different size input images of 331×331 and 224×224 and two new architectures NASNetLarge and NASNetMobile were created. The resulting architectures achieved excellent performance particularly in ImageNet datasets, as several computer vision applications derive features from its classification models.

3) Xception

This network is an alteration of the InceptionNet DL-CNN where the inception modules are replaced with separable convolutions arranged in a depth-wise manner. In this model, output of specific layers is added to the output from the preceding layers. Its parameter size is similar but performs slightly better than InceptionNet. Due to this, the Xception achieves a comparably outperforming classification accuracy compared to InceptionV3.

4) DenseNet201

DenseNet is comparable in architecture to ResNet with fewer variations. In the DenseNet model, feature maps of all the earlier layers are concatenated and used as input to the forthcoming layer. For L Layers there will be L(L+1)/2 straight connections. Dense blocks are bound together using transition layers. The transition layer reduces the spatial extents of the inputs, and also "compresses" the feature maps to a smaller number. Figure 3 depicts the DenseNet architecture.

C. Model Formulation

This section focuses on the model formulation for binary classification OSCC images. To consider the transfer learning approach, we have applied four pre-trained DL-CNN models such as InceptionV3, NASNetLarge, Xception, and DensNet201 for the OSCC detection. These models have achieved success in the field of computer vision and medical imaging and thus preferred in our study for the classification of benign and malignant cases from oral lesion images. These models are already trained on large-scale labeled dataset called ImageNet [37] and is now fine-tuned over the oral lesion image dataset. These models were modified by adding appropriate layers to achieve great performance for OSCC detection. Later the best performing model is selected and considered for further comparison.

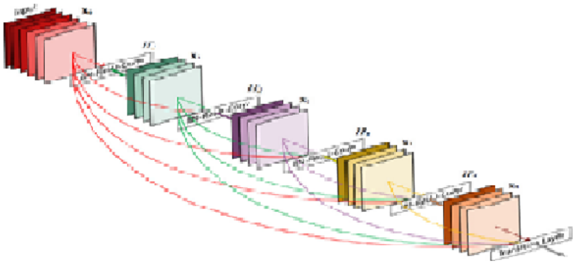

Figure 3. DenseNet Architecture with 5 layers, Courtesy of model [34]

Our proposed DL-CNN model accepts an input image of size 224× 224×3 and gives a binary decision on malignant or benign classes. For the binary classification case, we propose a DL-CNN for the detection of OSCC. The proposed DL-CNN model consists of (1) an input layer with the images of the size 224×224 ×3; (2) transferred convolutional and pooling layers of any of the four pre-trained models (3) a single convolutional layer with a filter size of 32 and kernel size 4×4, (4) a ReLU activation function (5) a MaxPooling layer for down-sampling the image (6) a flatten layer (7) a dropout out layer with 0.5 dropout rate (8) and a final dense layer with SoftMax activation function for classification of a binary output using the binary cross-entropy function. Max-pooling layers decreases the number of trainable parameters to reduce the image representation. Figure 4 highlights the proposed model without the pre-trained DL-CNN model layers. In the convolutional layers, filters of size 3×3 with stride [1 1] and padding "same" have been applied to the image. Max-Pooling has been performed over a 2×2 pixel window with stride [2 2]. The ReLU function accomplishes the non-linear transformation of inputs present in the model. The dropout layer with a rate of 0.5 drops some units to prevent the model from overfitting.

4. EXPERIMENTAL RESULTS

In our experimental study we have selected four recently developed candidate pre-trained DL-CNN models namely InceptionNet, NASNetLarge, Xception, and DenseNet201 through the approach of transfer learning for the oral lesion biopsy histopathological image dataset. These candidate models were modified with additional layers for effective OSCC detection. We have considered the total size of 2000 images combining both malignant and benign images which

TABLE I. Different DL-CNN models and Parameter Specifications

DL-CNN Models	Depth in terms of No. of Layers	Image Dimension(w,h,d)	No. of Parameters(millions)
VGG-16 [9]	16	224×224×3	138.4
ResNet50 [10]	107	224×224×3	25.6
Inception-V3 [31]	189	299×299×3	24.0
NASNetLarge [32]	533	224×224×3	88.90
Xception [33]	81	224×224×3	22.91
DenseNet201 [34]	16402	224×224×3	20.20

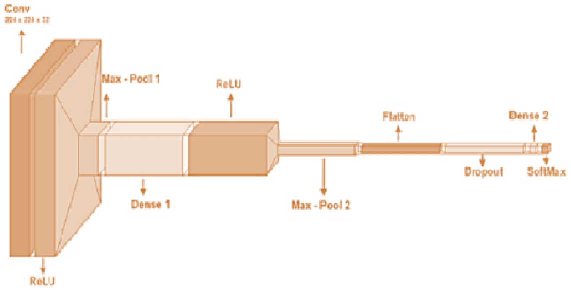

Figure 4. The architecture of the Proposed DL-CNN model

is still less number compared to the ImageNet dataset.

We considered the dataset from [35] and [36] and created a subset of 2000 images. The collection and usage of the images is addressed properly through citation. A set of 1200 malignant images and 800 benign images are considered to evaluate the performance of the four candidate pre-trained DL-CNN models. All the images are resized to 224×224 pixels to solve compatibility issues before providing them as input to the DL-CNN models.

The total number of 2000 images are split into testing and training datasets in the ratio of 20% and 80% based on the train-test split strategy. For the training and testing processes, the images from both the classes i.e., malignant and benign are selected. The training set data is used for training the model and the test data set is used for validating the model on previously unexamined data, after training and performing the hyper-parameter selection. The work is carried out with the open-source Keras framework and tthe TensorFlow backend using Google Colab with Python. All of the experiments were performed using a laptop with a dual-core I5 processor and 8 GB RAM. All of the experiments were performed using a desktop computer equipped with a dual-core I3 processor with 6 GB of DDR4 RAM. During experimentation, we considered the most suitable functions and hyperparameters heuristically as depicted in Table 2.

A. Evaluation Measures

The evaluation measures for the results are based on the overall number of truly classified and misclassified detections. This can be depicted by a confusion matrix. A confusion matrix is an outline of detection results for a classification process. The overall number of correctly identified and misidentified detections are represented by sum values and divided into two categories: Predicted Labels and True Labels. The parameters associated with the confusion matrix are depicted in Table 3.

The performance of the proposed model is evaluated from the following factors: True positive represents those numbers where the subjects are predicted with OSCC and the subjects actually have OSCC. True negatives represent those numbers where the subjects are predicted healthy (benign) and the subjects actually are healthy.

False positives represent those numbers where the subjects are predicted with OSCC when the subjects are actually healthy. False negatives represent those numbers where the subjects are predicted healthy when the subjects are actually having OSCC. As a result of the experiments, the confusion matrix parameters are also utilized to discuss other classification parameters such as Sensitivity, recall, precision, F1-Score, and accuracy. A classification test's recall parameter is specified in (1) as,

$$Recall = \frac{t_p}{t_p + f_n} * 100 \tag{1}$$

The count of true positives is given by t_p, whereas the count of false negatives is given by f_n. The classification test's precision parameter is specified as,

$$Precision = \frac{t_p}{t_p + f_p} * 100 \tag{2}$$

The correctness of the classification task is also indicated by the F1-Score. The F1 score may be a special measure to apply when there is an irregular class division due to the presence of a significant count of Actual labels. The Recall and Precision values as established in (3) can be used to estimate this value.

$$F1Score = 2 * \frac{[Recall * Precision]}{[Recall + Precision]} \tag{3}$$

The accuracy of the experiment in terms of confusion matrix parameters is computed as the ratio of true findings $(t_p + t_n)$ and all findings $(t_p + f_n + f_p + t_n)$ specified in (4),

$$Accuracy = \frac{(t_p + t_n)}{(t_p + f_n + f_p + t_n)} * 100 \tag{4}$$

TABLE II. Functions and Hyperparameter Value Settings

Hyperparameters	values
Input size	224×224 pixels
Train Test Split ratio	80:20
Batch Size	8 Samples
Epochs	10
Optimizer	Adam
Learning Rate	1e-4
Dropout	0.5
Rotation range	15

TABLE III. A 2×2 Confusion Matrix

Total No of Subjects	Label Predicted (Yes)	Label Predicted (No)
True positive Label	true positive (t_p)	false positive (f_p)
True Negative Label	false negative (f_n)	true negative (t_n)

The area under the ROC Curve (AUC) fuses the receiver operating characteristics (ROC) curve from (0,0) to (1,1). It provides the collective measure of classification performance at numerous threshold values. AUC has a range from 0 to 1. The higher the AUC, the better the model is at classification.

B. *Comparison of various CNN-DL models*

The results of the four candidates' pre-trained DL-CNN models modified with additional layers have been compared. We presented the confusion matrices results of these four modified pre-trained DL-CNN models on the selected dataset. The training performance can be evaluated from training loss, validation loss, and validation accuracy obtained by the selected DL-CNNs for the selected number of epochs. Table 4 shows the Recall, Precision F1 Score, and accuracy values for the applied pre-trained models. All the values here reported are from the results of the 10th epoch.

It has been observed that the proposed model along with the pre-trained DenseNet201 DL-CNN model has attained the top outcomes with a precision of 93.00%, recall of 93.00%, F1-score of 93.00%, and accuracy of 91.25%. The sensitivity and specificity of the proposed model are 88.75%, and 92.92% respectively. The best performing model is built by the DenseNet201 DL-CNN model which has layers of densely connected CNN [34]. The special structure of this pre-trained model and the addition of certain layers enhances data flow across the network and relieves the vanishing gradient issues. Additionally, DenseNet201 improves the parameter efficacy and offers each layer shared learning of the network. A further significant feature of this model is its regularization effect which offers reduced over-fitting on training with reduced data sets [34].

Also, it is observed that the acquired precision values are surprisingly good with 92.00% with the modified InceptionNet model and 90.00% with the modified NASNetLarge model but with fewer recall values. Xception is the second-best performer which obtained a precision of 87.00%, and recall of 90.00% with an accuracy of 86.50% at the 10th epoch.

Figure 5 illustrates the detection results of the modified DL-CNN model with DenseNet201 as the pre-trained model. The indicator true value=0 indicates that the detection=0 for OSCC true detection i.e., malignant case. Whereas true value=1 indicates that the detection=1 for normal patients i.e., the benign case with correct outcome.

The training accuracy, training loss, validation accuracy, and validation loss graphs are shown in Figure 6 to Figure 9 for InceptionNet, NASNetLarge, Xception, and DenseNet201 respectively. The significance of the proposed model along with the pre- DenseNet201 pre-trained model is that it automatically avoids overfitting issues due to the inclusion of drop-out layers. From these results, we can deduce that any patient who is having benign results (true negatives) can be diagnosed as normal with high accuracy by considering the histopathological images and applying the modified DenseNet201 DL-CNN model. Figure 10 represents the confusion matrices for different modified DL-CNN models. It represents the true labels (benign and malignant) accordingly with predicted labels (benign and malignant) for different models. This gives a clear assessment of the t_n, t_p, f_p, and f_n values. While it is always desirable to have the large values of the t_n and t_p, consequently the values of f_p and f_n are also likewise significant in the medical field. When a person is having a benign lesion but is considered as having a malignant lesion then we have an f_p value, and this means undesirable psychological distress and harmful health side effects due to cancer therapy.

The confusion matrix for the modified DenseNet201

TABLE IV. Evaluation Metrics results (%)

CNN-DL Models	Precision	Recall	F1-Score	Accuracy
DenseNet201	93.00	93.00	93.00	91.25
InceptionNetV3	92.00	80.00	86.00	84.50
NASNetLarge	90.00	81.00	87.00	85.25
Xception	87.00	90.00	89.00	86.50

TABLE V. Evaluation Metrics results (%)

CNN-DL Models	optimizer	Precision	Recall	F1-Score	Accuracy
Xception	SGD	87.00	89.00	87.00	85.80
Xception	Adadelta	86.00	89.00	88.00	86.00
Xception	Adam	87.00	90.00	89.00	86.50
DenseNet201	SGD	91.00	90.00	92.00	90.45
DenseNet201	Adadelta	92.00	92.00	91.00	90.10
DenseNet201	Adam	93.00	93.00	93.00	91.25

TABLE VI. Performance Comparison with Notable work

DL-CNN	Accuracy (%)
G. Forslid et al. [8]	82.39
Rutwik et.al. [38]	89.52
Welikala et. al. [14]	88.20
Gupta et. al. [21]	89.30
Song et al. [22]	86.90
Rahman et. al. [35]	89.70
Kim et al. [39]	78.10
M. Aubreville et. al. [40]	88.30
Proposed DL-CNN Model- Modified DenseNet201	91.25

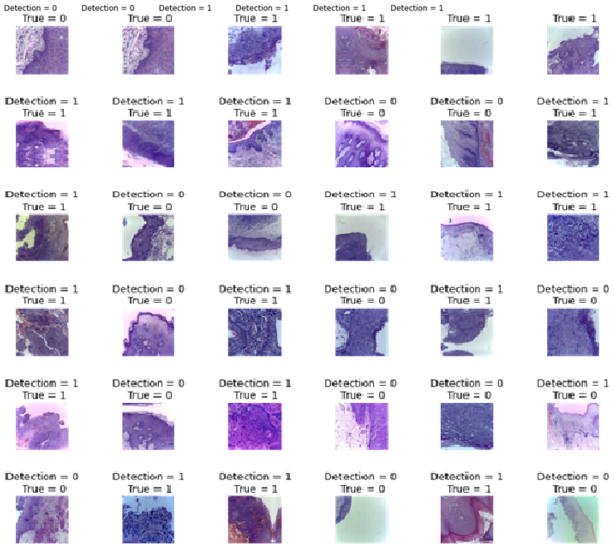

Figure 5. Detection Results for sample Histopathological Images of DenseNet201

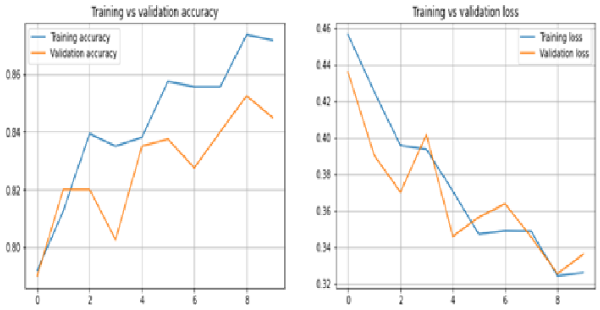

Figure 6. Training Loss and Accuracy plots of InceptionNet

DL-CNN model displays the high values of t_ns and the low values for f_ns. The confusion matrices allow for visual assessment of the modified DL-CNN models for correctly classifying each of the 400 test images into their respective target class. Higher labels of f_ns are produced by rest of the modified DL-CNN models including InceptionNet, NASNetLarge, and Xception can be considered critical as the misclassified lesions are all malignant. We also exploited the ROC curve to estimate the performance of the modified DL-CNN models for effective OSCC detection. The ROC

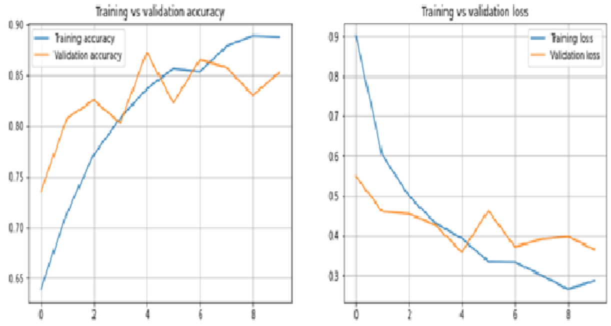

Figure 7. Training Loss and Accuracy plots of NASNetLarge

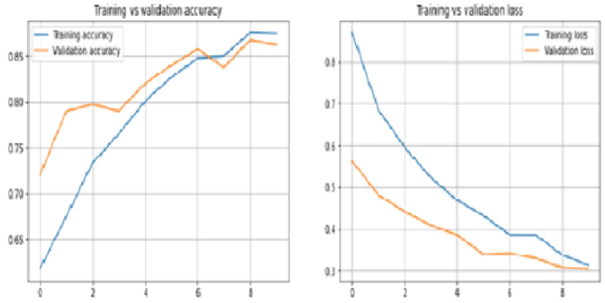

Figure 8. Training Loss and Accuracy plots of Xception

curve plots the true positive rate (sensitivity) and the false positive rate (1-specificity) with various threshold values and also computes the AUC value. Figure 11 depicts the ROC curves of the modified DL-CNN models.

Here, the modified DenseNet201 DL-CNN model was able to achieve the highest AUC of 0.908. The other models InceptionV3, NASNetLarge, and Xception achieved the AUC of 0.855, 0.862, and 0.852 respectively. Sensitivity measures the number of images of malignant patients who are correctly classified whereas specificity measure the number of images of benign patients who are correctly classified. All statistical computations were performed with scipy and scikit-learn python packages.

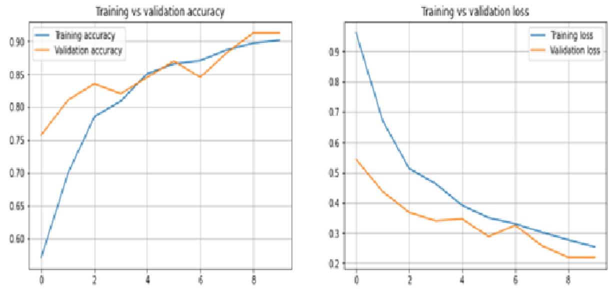

Figure 9. Training Loss and Accuracy plots of DenseNet201

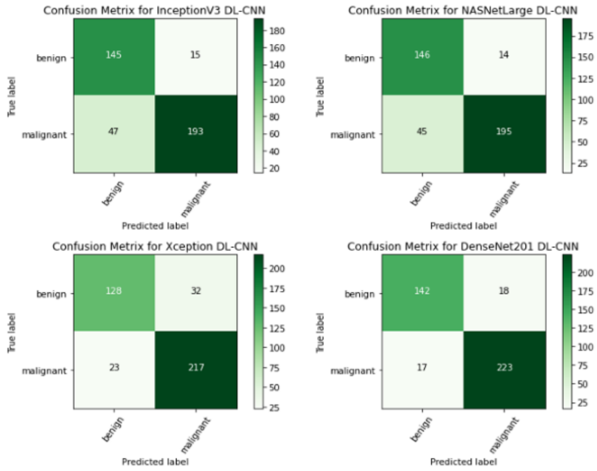

Figure 10. Confusion Matrices for modified DL-CNN models

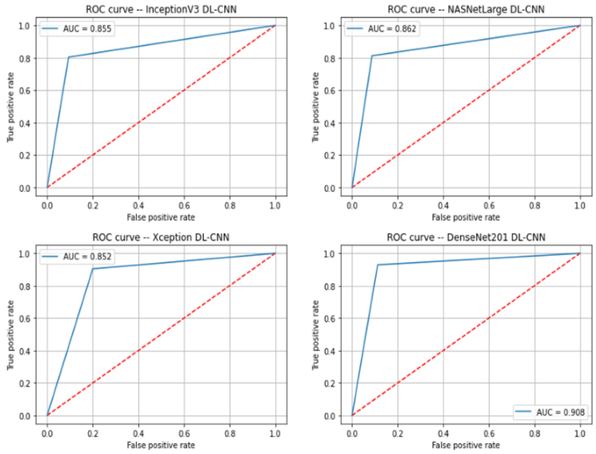

Figure 11. ROC Curves of modified DL-CNN models

C. Comparison with Other Optimizers

In this work, the Adam optimizer [41] is selected for the training of all the modified DL-CNN models. A comparison with other optimizers is also presented here with SGD [42] and Adadelta [43]. Table 5 lists the confusion matrix results for the two best performing modified DL-CNN models Xception and DenseNet201. The outcome of the experiments indicates the efficacy of Adam optimizer against other optimizers. The selection of other optimizers does not affect much to the performance parameters, still Adam optimizer surpasses these values.

D. Comparison with Contemporary Methods

Here, we compared the modified DenseNet201 DL-CNN model results with other DL models for OSCC detection using histopathological images as tabulated in Table 6. It is observed that the proposed modified model with the pre-trained DenseNet201 has attained better outcomes compared with other candidate modified models and also with other existing methods for OSCC detection. Compared to all these notable works [8], [38], [14], and [30] we considered a dataset with a relatively larger number of histopathological images for OSCC detection. D. W. Kim et al. [39], in their work achieved an accuracy of 78.10% for OSCC detection from comparatively large size database. M. Aubreville et al used the laser endomicroscopy images of the oral cavity for OSCC detection and achieved an accuracy of 88.30%. The proposed model in this study which has DenseNet201 DL-CNN modified with additional layers achieved excellent results compared with methods for OSCC detection.

5. CONCLUSION

Recently DLTs have offered ample opportunities for automatically detecting OSCC with the performance matching or even better than that of human experts. The DL-CNN-based detection of oral lesion images provides a non-invasive and cost-effective method to detect OSCC lesions in early-stage and thus enables early treatment. This work aimed to perform an automated classification of benign and malignant oral histopathological images by implementing modified DL-CNN models. The work proposes an application of the best suitable DL-CNN model for fully automated OSCC detection and examined the performance of the proposed DL-CNN model for OSCC classification. For this, four recently developed candidate pre-trained DL-CNN models namely InceptionNet, NASNetLarge, Xception, and DenseNet201 were selected through the approach of transfer learning. A proposed DL-CNN model is constructed with suitable additional layers and the candidate models were modified with this architecture for effective OSCC detection. A suitable dataset is constructed from the 2000 histopathological images including benign and malignant images. Among these, the DenseNet201 DL-CNN model with modified architecture outperforms other modified models and achieved an accuracy of 91.25%. The proposed work was also found to be significantly superior in results to some of the notable work. Thus, the proposed DL-CNN model has achieved substantial performance for binary classification of benign versus malignant biopsy histopathological images.

References

[1] Bray F and Ferlay J, et al., "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 68, p. 394–424, 2018.

[2] Coletta RD, Yeudall WA, and Salo T, "Grand challenges in oral cancers," *Front Oral Health*, vol. 1, pp. 1–3, 2020.

[3] Sahanaz Praveen Ahmed and Lekshmy Jayan, et al., "Oral squamous cell carcinoma under microscopic vision: A review of histological variants and its prognostic indicators," *SRM Journal of Research in Dental Sciences*, vol. 10, pp. 90–97, 2019.

[4] Gigliotti J, Madathil S, and Makhoul N, "Delays in oral cavity cancer," *Int J Oral Maxillofac Surg*, vol. 48, pp. 1131–1137, 2019.

[5] A. Duggento and A. Conti, et al., "Deep computational pathology in breast cancer," *Seminars in cancer biology*, vol. 72, pp. 226–237, 2020.

[6] S. Wang and D. M. Yang, et al., "Artificial intelligence in lung cancer pathology image analysis," *Cancers*, vol. 11, p. 1111–1163, 2019.

[7] Muqeet M.A. and Quadri M.U., et al., "Deep learning-based prediction of covid-19 disease using chest x-ray images (cxris)," *Contactless Healthcare Facilitation and Commodity Delivery Management During COVID 19 Pandemic. Advanced Technologies and Societal Change.*, vol. 1, pp. 15–25, 2021.

[8] Hakan Wieslander and Gustav Forslid, et al., "Deep convolutional neural networks for detecting cellular changes due to malignancy," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 82–89, 2017.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Vision and Pattern Recognition*, 2015.

[10] Kaiming He and Xiangyu Zhang, et al., "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[11] Qiuyun Fu and Yehansen Chen, et al., "A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study," *E-Clinical Medicine*, vol. 27, 2020.

[12] Navarun Das and Elima Hussain, et al, "Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network," *Neural Networks*, vol. 128, pp. 47–60, 2020.

[13] Tanriver G and Soluk Tekkesin, et al., "Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders," *Cancers*, vol. 11, 2021.

[14] Welikala and R. A, Remagnino, et al., "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *IEEE Access*, vol. 8, pp. 132 677–132 693, 2020.

[15] Shamim and M.Z.M., et al., "Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer," *The Computer Journal*, vol. 65, p. 91–104, 2022.

[16] Jeyaraj P. R. and Nadar E. R. S, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of Cancer Research and Clinical Oncology*, vol. 145, p. 829–837, 2019.

[17] Kevin Figueroa and Bofan Song, et al., "Interpretable deep learning approach for oral cancer classification using guided attention inference network," *Journal of Biomedical Optics*, vol. 27, 2022.

[18] Ramprasaath R. and Selvaraju, et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," *IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

[19] Kunpeng Li and Ziyan Wu, et al., "Tell me where to look: Guided attention inference network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9215–9223, 2018.

[20] Shipu Xu and Chang Liu, et al., "An early diagnosis of oral cancer based on three-dimensional convolutional neural networks," *IEEE Access*, vol. 7, pp. 158 603–158 611, 2019.

[21] R. K. Gupta and M. Kaur, et al., "Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium," *Journal of Multimedia Information System*, vol. 6, pp. 81–86, 2019.

[22] Bofan Song and Sumsum Sunny, et al., "Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomedical Optics Express*, vol. 9, pp. 5318–5329, 2018.

[23] Alabi RO and Almangush A, et al., "Deep machine learning for oral cancer: From precise diagnosis to precision medicine," *Frontiers in Oral Health*, vol. 2, 2022.

[24] Panigrahi S and Swarnkar T, "Machine learning techniques used for the histopathological image analysis of oral cancer-a review," *Journal of Multimedia Information System*, vol. 13, pp. 106–118, 2020.

[25] Shaban M and Khurram SA, et al., "A novel digital score for abundance of tumor infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma," *Scientific Reports*, vol. 9, 2019.

[26] Fujima N and Andreu-Arasa VC, et al., "Deep learning analysis using fdg-pet to predict treatment outcome in patients with oral cavity squamous cell carcinoma," *Eur Radiol*, vol. 30, p. 6322–6330, 2020.

[27] Das DK and Bose S, et al., "Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis," *Tissue Cell*, vol. 53, pp. 111–119, 2018.

[28] Chih-Hung Chan and Tze-Ta Huang, et al., "Texture-map-based branch-collaborative network for oral cancer detection," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, pp. 766–780, 2019.

[29] Nanditha BR and Geetha A, et al., "An ensemble deep neural network approach for oral cancer screening," *International Association of Online Engineering*, 2021.

[30] Jubair F and Al-karadsheh O, et al., "A novel lightweight deep convolutional neural network for early detection of oral cancer," *Oral Diseases*, vol. 28, pp. 1123–1130, 2021.

[31] C. Szegedy and V. Vanhoucke, et al., "Rethinking the inception architecture for computer vision," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 2818–2826, 2016.

[32] Barret Zoph and Vijay Vasudevan, et al., "Learning transferable architectures for scalable image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, 2018.

[33] Chollet F, "Xception: Deep learning with depthwise separable convolutions," *IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[34] Gao Huang and Zhuang Liu, et al., "Densely connected convolutional networks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[35] T. Y. Rahman and L. B. Mahanta, et al., "Histopathological imaging database for oral cancer analysis," *Data in Brief*, vol. 29, pp. 105–114, 2020.

[36] A. F. Kebede, "Histopathological oral cancer [dataset]," *Kaggle*, 2020.

[37] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, p. 84–90, 2017.

[38] Rutwik Palaskar and Renu Vyas, et al., "Transfer learning for oral cancer detection using microscopic images," *Cornell University*, 2020.

[39] Dong Wook Kim and Sanghoon Lee, et al., "Deep learning-based survival prediction of oral cancer patients," *Scientific Reports*, vol. 9, 2019.

[40] Marc Aubreville and Christian Knipfer, et al., "Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning," *Scientific Reports*, vol. 7, 2017.

[41] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *International Conference for Learning Representations*, vol. 3, 2015.

[42] Ilya Sutskever and James Martens, et al., "On the importance of initialization and momentum in deep learning," *International Conference on Machine Learning*, vol. 30, pp. 1139–1147, 2013.

[43] M. D. Zeiler, "Adadelta: An adaptive learning rate method," *Machine Learning*, 2012.

Sayyada Hajera Begum Sayyada Hajera is a research scholar in the Department of Computer Science and Engineering at Koneru Lakshmaiah University, India. Her research area deals with deep learning models in Healthcare. Her area of interests include Machine Learning, Deep Learning and Data Mining.

Vidyullatha Pellakuri Dr Vidyullatha Pellakuri is working as Associate Professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah University, India. She has published 37 papers in reputed journals and conferences. Her area of interest include Data Mining, Soft Computing Techniques, Machine Learning, Big Data, IOT.