



## Analysis and Classification of H&E-Stained Oral Cavity Tumour Gradings Using Convolution Neural Network

Prabhakaran Mathialagan<sup>1\*</sup>

Malathy Chidambaranathan<sup>1</sup>

<sup>1</sup>*Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Potheri, Kattankulathur 603203, Chengalpattu District, Tamil Nadu, India*

\* Corresponding author's Email: Prabhakaran.mathialagan@srmuniv.edu.in

---

**Abstract:** Upper Aero Digestive Tract (UADT) cancer is one of the most common cancer types in any gender. Early detection and diagnosing such type of cancer will reduce the risk of death in human. In this paper we addressed the state-of-the-art solutions for two major pathological constraints like, dysplasia type tumour grading and tumour grade classification with artifact's present in biopsy images. A new Whole Slide and patch-based CNN model was proposed which involves different pre-processing techniques to detect and normalize the stain artifacts present in biopsy tissue samples. The proposed CNN model is developed to classify different oral cavity tumour sites and its gradings automatically. It was observed that the proposed CNN model achieved best accuracy of 98 % with zero classification error rate. These findings will throw a light for pathologists to handle digital biopsy images with artifacts and also will classify dysplasia type tumour grades automatically.

**Keywords:** Oral cavity, Cancer, CNN, Classification, Dysplasia.

---

### 1. Introduction

Oral cancer has become one of the major causes of death worldwide in both men and women. As per WHO oral cancer statistics, approximately 6,75,000 new cancer positive cases and maximum of 3,30,000 death occurs every year [1]. As per the article published in cancer country profile 2020 Nasopharynx holds the top position among the other top ten cancers. In the year 2019 cancer observatory database shows that nasopharynx cases have 5.2 % positive cases and 5.4 % death rate occurs globally [2]. Healio, an American oncology journal stated that approximately 1,10,000 peoples are diagnosed in both oral and nasal cavity cancer every year [3]. The above-mentioned reports clearly show the importance of early detection and proper diagnosis of cancer to reduce the death rates. Manual microscopic cancer screening using biopsy tissue samples is tedious process in finding the dysplasia type of tumours. Misclassification of manual screening would lead to metastasis and will put the patients in high risk. In [4, 5] over the past decades machine

learning yields the highest priority in medical domain by playing their major role in automation for disease detection, prediction and classification. The main objective of biopsy tissue analysis is to detect and classify the type of tumours and its gradings. Manual screening is time consuming and also classification of dysplasia type tumours with stain artifacts is challenging. This is due to very less morphological variation that exists in biopsy tissue samples. To overcome this issue an ultimate image enhancement technique called decorrelation contrast stretching is used to enhance the biopsy tissue colour channels and normalize the stain artifacts detected. For proper classification of dysplasia tumours and its gradings with artifacts, a computer aided WS and patch-based Deep Convolution Neural Network (CNN) model is proposed. The Deep CNN model proposed is implemented with high level features extracted from H & E-stained Upper Aero-Digestive Tract (UADT) biopsy slide images. UADT is the combination of oral cavity and nasal cavity whereas the oral cavity anatomy consist of seven different sites and nasal cavity consists of four different sites. UADT biopsy

Whole Slide images were extracted from three microscopic magnifications like, 4X magnified scanner view to show the epithelial and connective tissues, 10X-low power magnified view to show the spread of neoplastic epithelium and 40X-high power magnified view to show the cell and nuclei morphological characteristics. The proposed WS and patch-based CNN model is used here to extract the features from UADT biopsy digital images for ten different oral cavity tumour grading classification. For classification, cancerous, non-cancerous and dysplasia type tumour features from oral cavity is extracted.

Rest of the paper is formulated as follows, challenges and existing related works are outlined in Section 2. Materials and methods of proposed methodology is discussed in Section 3. proposed method for UADT tumour grading classification is elaborated in Section 4. In Section 5 proposed work results were discussed and comparative analysis is done with existing transfer learning models. In Section 6 the objective of proposed work is discussed elaborately and suggestions for future works are given.

## 2. Literature review

In [6] author stated that almost 90 percent of oral cavity squamous cell carcinoma occurs only at the surface of epithelium compared to other squamous cell carcinoma sites. It is been proved that the biotic and morphological behaviour of tumour cell and nuclei are different in every cancer site. Morphological behaviour analysis is important in finding the malignant tumour metastasis level and tissue characteristic analysis is very important in finding the dysplasia type of tumours for malignant transformation. The deep learning model proposed lacks in performance while predicting the low and severe grade dysplasia tumours. In [7]. Author stated that, combination of both pathologist and automated tissue analysis tool support will help in a better way for proper diagnosis. To find the precise metastasis level of primary UADT tumours, patients must undergo a complete inspection of both oral cavity and nasal cavity to discover the chances of occurring secondary tumours in digestive tract. Cancer staging must be evaluated with clinical observation and also tissue analytical tool support must be availed to prevent the misclassification of tumour gradings. In recent years Deep Convolutional Neural Network plays a major role in medical image analysis and proved to show better performances particularly, in cancer grading and classification problem. In [8] a deep CNN model was proposed to classify normal

and abnormal non-alcoholic fatty liver disease. The proposed model was designed with three neural network layers for four class classification using Stochastic Gradient Descent (SGD) optimization technique. The proposed model achieved 95 % classification accuracy but the accuracy is less compared to other pre-trained CNN models like, VGG and Multi-Layer Perceptron (MLP) but fails to break the Alex Net classification performance with 97 % accuracy. In [9] a hybrid residual based deep CNN model was proposed to classify breast biopsy images with pretrained weights generated using two different domain datasets like BACH 2018 and ICIAR 2018. A transfer learning technique is used to finetune the proposed model for multiple class classification to achieve 97 % accuracy. The proposed model has achieved good classification accuracy. Comparative analysis with existing pre trained model was not done and missed to prove the efficiency of proposed model. In [10] author used pre-trained neural network models like inceptionV3 and Recurrent Neural Network model. The study was made with two different cancer gradings like Adenoma and Adenocarcinoma in gastric and colonic epithelial tissue whole slide images and achieved the best AUC rate. In [11] author used four pre-trained deep CNN models like VGG16, VGG19, Inception and ResNet for different melanoma type skin cancer dermoscopy images. In [12] author proposed colon cancer prognosis prediction model using convolutional neural network and gradient boosting machine learning classifier. The proposed model was used to predict stage III colon cancer and to classify either high or low risk for further clinical treatment. In [13] author developed a deep learning-based nuclei segmentation model using H&E-stained whole slide biopsy images from different cancer types. Author proposed a U-Net CNN model using H&E-stained nuclei images extracted from biopsy WSI in ten different cancer types. In [14] GAN based skin cancer classification model was proposed for different skin tissue type using ISIC-2018 skin cancer tissue dataset. The model was tested with four different GAN CNN versions like, GAN, DC-GAN, Style-GAN, SL-Style-GAN. After all the model evaluation was carried out with two different metrics like, inception score and Fréchet inception distance 95.2 % of classification accuracy was achieved. In [15] author developed a deep CNN epithelial segmentation model using three regression network layers called EpithNet. The proposed EpithNet is used for cervical cancer epithelial segmentation using histopathology biopsy slide images. In [16] author proposed two stage head and neck squamous cell carcinoma classification using hyperspectral microscopic

images. The first stage is nuclei extraction using principal component analysis and second stage is nuclei segmentation using hybrid SVM and patch-based CNN model. The proposed automated model achieved 0.94 nuclei segmentation Area Under Curve. In [17] a deep CNN model was proposed for oral cancer classification using cytology images. The classification model is focused on cytology nuclei images extracted from three different datasets like pap smear samples, Whole Slide Images and liquid based images. The dataset was trained using two pre-trained CNN model like DenseNet201 and ResNet50 and achieved 84 % accuracy. In [18] author proposed a deep CNN model for oral squamous cell carcinoma classification. Epithelial tissue segmentation was made using gaussian blurring and Otsu thresholding image processing technique. The proposed CNN model classification accuracy is compared with existing pre-trained transfer learning techniques and show the promising accuracy obtained using proposed CNN model. In [19] a deep CNN model was proposed to classify breast cancer biopsy samples with four different tumour gradings and achieved 95.6 % classification accuracy. To analyse the feasibility of features extracted using CNN an SVM technique is used to classify four different breast cancer gradings and achieved 77.8 % accuracy. The CNN model achieved good classification accuracy and it was reported that the proposed classification model does not include whole slide images for feature extraction and classification. In [20] author proposed a novel two stage step wise fine-tuning deep CNN model to classify the limited gastric tumour biopsy tissue samples. Author stated the step wise fine tuning is training a pre-trained CNN model with different parameter initialization for different CNN layers to acquire unique image features. The model achieved 0.88 AUC with first stage CNN parameter initialization and 0.93 using second stage parameter initialization. In [21] author proposed an oral epithelial dysplasia tumour detection and five class classification model using deep convolution neural network. The proposed model has achieved 91.5 % classification accuracy. Author faced the annotation problem in the biopsy images with inflammatory cells that leads to misclassification of tumour grading's. In [10] author also addressed an oral squamous cell carcinoma tumour grading problem for multiple class classification. The classification experiment was conducted using four different transfer learning techniques like Alex net, VGG16, VGG19 and ResNet 50. The proposed CNN model outperforms other transfer learning models and has achieved 97.5 % classification accuracy. In [22] a deep CNN model was proposed to classify

malignant and non-malignant tumours of both gastric and colonic site with 8164 Whole Slide Images. The model produced a promising accuracy report of 96 %, 97 % and 99 % for three class classification. It was observed that the existing literature shows clearly about the importance of automated systems. At the same time the existing CNN models are capable to work only with the limited UADT tumour cites and gradings. The proposed work moves one step ahead to detect and annotate the artifact's present in the biopsy Whole Slide Images.

### 3. Materials and methods

Now a days deep learning has become the best assisting automated platform for cancer prediction and diagnosis predominantly in different tumour type classification and patient's survival prediction.

Since there was a substantial growth in computational power with very high probabilistic and statistical influence, deep learning has become the benchmark for different automated disease diagnostic systems in medical domain. To provide the best automated platform in assisting the pathologist for decision making the model proposed involves step by step process for UADT tumour grade classification.

The proposed model involves image acquisition, data pre-processing, data augmentation, feature extraction and classification using WS and patch-based CNN model as shown in Fig.1.

#### 3.1 Dataset description

The Upper Aero-Digestive Tract biopsy tissue samples of seven different oral cavity gradings like

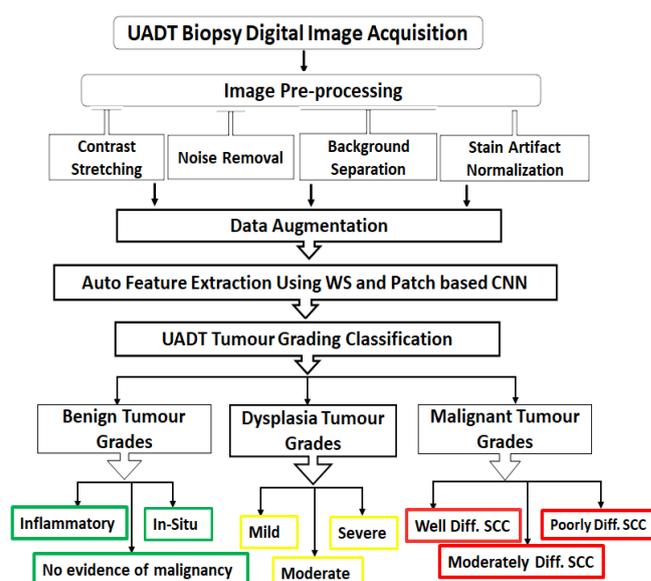


Figure. 1 Proposed workflow for UADT tumour classification

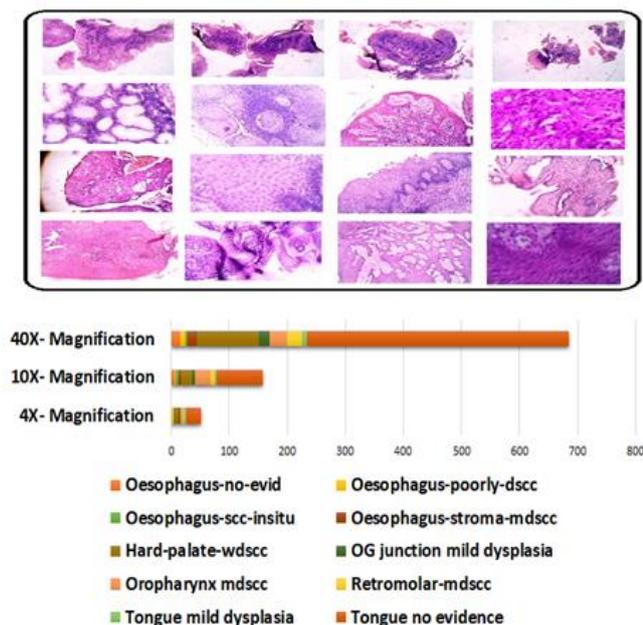


Figure. 2 UADT oral cavity biopsy tissue samples

Oesophagus, Tongue, Oropharynx, Retromolar region, Buccal Mucosa, Hard Palate and OG junction was collected under the guidance of senior pathologist from Government Vellore Medical College Hospital, department of pathology with (Ref. No. 1258/ H&FW (MCA3) / dept. Dt: 18-05-2018). Totally 1425 whole slide images were captured from different oral cavity sites from different oral cavity gradings like Well differentiated, poorly differentiated, moderately differentiated, In-situ, No evidence of malignancy, Severe dysplasia, Mild dysplasia and inflammatory pathology. Sample biopsy tissue images with 4X, 10X and 40X microscopic magnifications are shown in Fig. 2 (a) and its distributions are shown in Fig. 2 (b).

### 3.2 Data pre-processing

To improve the quality of UADT tumour grading classification using WS and patch-based Deep CNN model, the experimental image data is pre-processed:

1. To detect and normalize the stain artifacts.
2. To remove unwanted background details.
3. To enhance the contrast of epithelial and connective tissues for proper feature extraction and data augmentation to avoid over-fitting and less classification accuracy.

#### 3.2.1. Stain normalization and background separation

In every cancer investigation process, histopathology microscopic analysis is still considered as the golden standard for final diagnosis. Due to some improper tissue fixation and tissue glass slide preparation, artifacts might occur and lead to

morphological characteristic loss in tissue samples that leads to misclassification. Decorrelation (décor) stretching is a colour separation technique used to enhance and find the hidden characteristic present in the multi-spectral images [23]. After analysing the internal behaviour of decorrelation stretching it is applied on the H&E-stained biopsy tissue sample to find the stain and bubble artifacts present. Accordingly, the same décor technique is used to remove unwanted background details by applying normalization technique on each RGB colour channel alternately is shown in section 3.2.1.1. The decorrelation stretching technique is used to improve the colour channels with high intensity and to highlight each pixel by stretching its colour contrasts [24]. Sample biopsy image before and after contrast stretching is shown in figure 3.

#### 3.2.1.1. Decorrelation contrast stretching process

Step:1. Consider the input image  $i=1'$ ; and three channels 'C=3' with number of pixels 'n'.  $P_{j,i}$  is the value of jth pixel for channel C.

Step:2. Calculate the sum of pixels of each channel.

$$SUMy_{i,c} = \sum_{j=1}^n (P_{j,c} \times P_{j,i}) \tag{1}$$

Step:3. Calculate the covariance matrix of three channels.

(i). Covariance for single matrix is calculated by the equation shown in Eq. (2).

Where, 'a' and 'b' are the mean value of given image; 'n' is the number of images;  $x_i$ ,  $y_i$  is the individual pixel of an image.

$$cov_{a,b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{2}$$

$$Cov_{c,i} = \frac{1}{n-1} [SUMy_{i,c} - \frac{1}{n} \times SUMy_{i,c} \times SUMy_c] \tag{3}$$

Eq. (2). Shows the covariance calculated for three different channels

Step:4. Calculate the correlation matrix of all the three channels using Eq. (4).

$$Corr_{c,i} = \frac{Cov_{c,i}}{(Cov_{i,i} \times Cov_{c,c})^{1/2}} \tag{4}$$

Step:5. Calculate the Rotation matrix (Eigenvectors calculated from the correlation matrix) ‘R’.

Step:6. Calculate the Stretching Vector (Reciprocal of the square root of each eigenvalue multiplied with standard deviation ‘σ’).

Step:7. Calculate the Transformation matrix ‘T’ (composition of Rotation matrix and Stretching matrix). By following the matrix multiplication as shown in Eq. (5) the transformed matrix ‘T’, is obtained as contrast stretched image.

$$T = R^t S R \tag{5}$$

UADT biopsy whole slide image (Tongue- Mild-Dysplasia 40X magnification) which contains both stain and water blob artifacts is chosen. A contrast stretching technique (3.2.1.1) is applied on the biopsy image to enhance the intensity of the epithelial tissue pixels, to normalize the artifacts, to eliminate the unwanted background details present in the tissue sample. Fig. 3 shows clearly that stain and water blobs artifact edges also enhanced alike epithelium and connective tissues contrast present in the biopsy image. Fig. 3 shows the original biopsy image before and after contrast stretching. The technique eliminates unwanted background details and normalizes the stain artifacts present in the tissue sample. Table 1 shows the comparative analysis of artifact biopsy digital image intensity

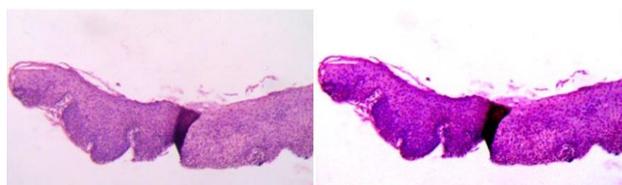


Figure. 3 Decor stretching for stain normalization and background separation

Table 1. Intensity analysis of before and after Decorrelation stretching

ROI	Before decorrelation contrast stretching			After decorrelation contrast stretching	
	Total no of pixels	Min value (0 to 255)	Max value (0 to 255)	Min value (0 to 255)	Max value (0 to 255)
Epithelial Tissue	1712	59	206	93	223
Artifacts	1417	29	208	0	201
Background	2457	218	247	227	255

values for before and after decorrelation contrast stretching.

It is observed that the intensity value for background ranges from 227 to 255 which shows the background is separated clearly from the foreground intensity values. The artifacts are normalized between 0 and 201 with which the total number of pixels gets reduced and intensity gets increased after applying décor stretching. When décor stretching is applied, the intensity increases in artifacts along with the epithelial tissue works even in epithelial tissue by increasing the contrast of every pixel due to which the intensity values increases and threshold falls in the range of 59-206 and 93-223.

### 3.3 UADT Biopsy digital data augmentation

Due to limited set of UADT biopsy samples and imbalance size of input images in each class, training the deep neural network may lead to over-fitting problem which may result in very less classification accuracy [25]. To avoid such CNN training complications, we use different data augmentation techniques, like horizontal rotation with different angles (45° 90°, 135°, 180°), vertical rotation with different angles, horizontal flip, width shifting, shearing and zooming is applied to increase the sample size and to maintain equal number of images in each class. Augmented biopsy samples and each class distribution is shown in Fig. 4.

Fig. 4 Shows the sample UADT biopsy WSI. Different data augmentation technique is applied and augmented sample images are shown with equal distribution in each class.

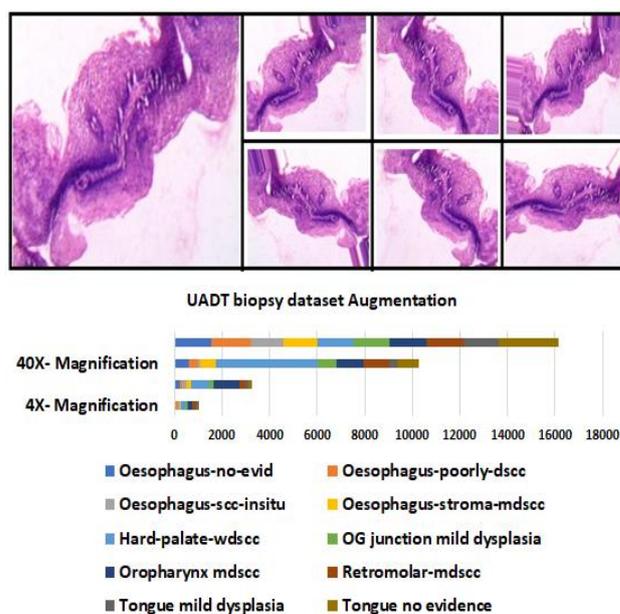


Figure. 4 Biopsy image augmentation

### 4. Proposed methodology

It was observed that the previous state-of-the-art Computer Aided Design (CAD) performances in medical image analysis using deep learning was very promising in assisting the clinicians for automated disease diagnosis [26]. Since our research work is based on cancer grade classification, a deep supervised WS and patch-based CNN Classifier is proposed for both Whole Slide and patches of biopsy images with three different microscopic magnifications like 4X, 10X and 40X magnifications. 4X microscopic magnification is used to analyse and extract the epithelial tissue and connective tissue features, 10X magnification is used to analyse and extract the spread of neoplastic epithelium and 40X magnification is used to extract the normal, abnormal and necrotic cells and nuclei features which are the three key factors to analyse the dysplasia type tumours.

Consider the classifier ‘M’ as  $m(x)=y$ , where, ‘X’ is an input and Y is the predicted class label for the actual class label Y. The ten different classes are oesophagus no evidence of malignancy, oesophagus poorly differentiated squamous cell carcinoma, oesophagus carcinoma in-situ, oesophagus moderately differentiated squamous cell carcinoma, hard palate well differentiated squamous cell carcinoma, oropharynx moderately differentiated squamous cell carcinoma, Retromolar moderately differentiated squamous cell carcinoma, tongue mild dysplasia, tongue no evidence of malignancy and esophago-gastric junction mild dysplasia. The dataset consists of 16172 biopsy images in total extracted from both whole slide (WS) and its patches for ten class. The proposed WS and patch-based CNN model consist of thirteen layers with five types, including an input layer, output layer, convolution layer, fully connected dense layer and output layer as shown in Fig. 5.

Fig. 5 shows step by step feature extraction process with dense layer to classify the images based on class labels. Initially let IL be the input matrix for L layers with convolution kernel K of size  $r \times r$  is applied on the input image  $p \times q$ , with single stride S, to get the curves and edges as low-level features. The input matrix for convolution operation is shown in Eq. (6).

$$L(i, j) = \theta(-\sum IL(i, j) \odot k) \tag{6}$$

The output matrix would be  $(p - k + 1) \times (q - k + 1)$

To get the high-level features of the image data, pooling operation is applied on the output matrix using the Eq. (7).

$$S(i, j) = S_k \odot S(i, j) \tag{7}$$

The summation matrix  $S_k$  stands for the kernel K. In second layer, 16x16 convolution mask is applied on the matrix obtained from Eq. (7). The weights generated in each layer will be passed to the next layer. In our CNN model while applying pooling operation on the convolved image it is observed that data loss has occurred. To avoid such information loss during pooling operation the image pixel resolution is reduced rapidly for each layer in feature extraction process. After pooling operation, the output of each layer is evaluated using non-linear activation function called ReLU as shown in Eq. (8). The ReLU activation function is used to avoid the gradient of the loss with respect to weights generated in each layer as shown in Eq. (8).

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \tag{8}$$

After every convolution layer our model uses ReLU activation function that has the gradient as

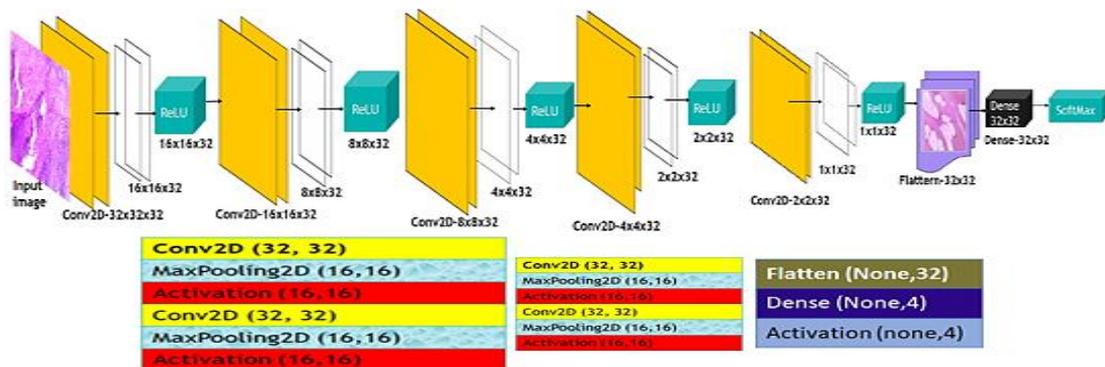


Figure. 5 Proposed WS and patch-based CNN architecture

either 0 or 1 if the weights generated is less than 0 it will be treated as 0 and the weights greater than 0 will be treated as 1 to avoid the weight loss. After the weights generated with five convolution operation the flatten layer is used to convert the 2D matrix in to a single vector to be fed in to fully connected layer for multiple class classification. UADT class labels are classified using SoftMax activation function to calculate the probability distribution of each class as either 0 or 1 as shown in Eq. (9) where z is the input vector at j<sup>th</sup> output vector for the output layer, where j= 1, 2, ..., k.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^k e^z} \quad (9)$$

### 5. Results

In this section we provide the result obtained from the proposed WS and Patch-based CNN classifier for 16172 UADT biopsy images of ten classes. The UADT datasets achieved the best classification accuracy of 98 % with zero validation loss. The dataset is split with 10917 images for training, 4043 images for testing, 1212 images for validation. The model is implemented using TensorFlow with Nvidia Titan XP with 12.0 GB frame buffer, core Clock: 0.954GHz, core Count:2, Memory Size: 2.00GiB and Memory Bandwidth: 13.41GiB/s. The classification report for the proposed CNN model is shown in figure 6. Confusion matrix is calculated for the proposed model which includes precision, recall and F1-score for individual UADT site as shown in Table 2.

Table 2. Performance evaluation of proposed WS and patch-based CNN

UADT oral sites	Precision	Recall	F1-score
Oesophagus-no-evid	0.99	1.00	1.00
Oesophagus-PDSCC	0.97	0.96	0.97
Oesophagus-scc-insitu	0.98	0.98	0.98
Oesophagus-stroma-MDSCC	0.99	0.99	0.99
Hard-palate-wdsc	0.98	0.96	0.97
OG junction mild dysplasia	0.97	0.98	0.97
Oropharynx mdsc	0.98	0.98	0.98
Retromolar-mdsc	0.99	0.98	0.99
Tongue mild dysplasia	0.96	0.97	0.96
Tongue no evid	0.98	0.99	0.99
Average accuracy		0.98	

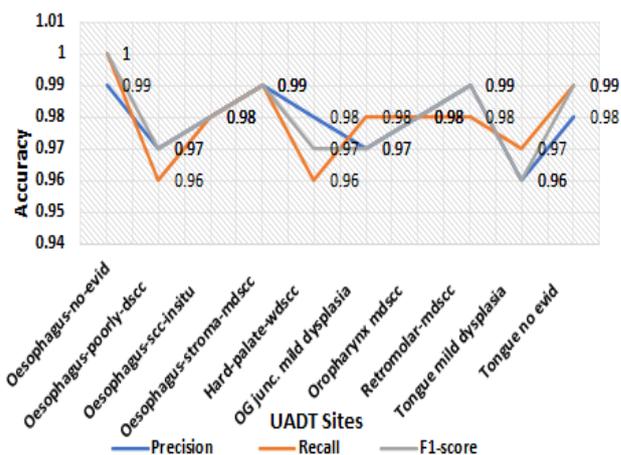


Figure. 6 Proposed CNN model classification comparison report

In Fig. 8 ten UADT tumour grade classification is reported with higher sensitivity of more than 96 % in each class. The weighted average of precision and recall is shown in figure.8 as F1-score with more than 96 % in each class. In Fig. 7 training, validation loss and accuracy are shown for ten class classification with 100 epochs and 405 steps for each epoch. As per the results obtained after execution it is observed that each training epoch has taken 17 seconds to complete.

### 5.1 Comparative analysis of existing and proposed CNN model

Different transfer learning techniques were used to compare and contrast the effectiveness of proposed WS and patch-based CNN model. Transfer learning is a combination of different pre-trained CNN model that allows the change in probability distribution p(x) in both test and train data [27, 28]. Pre-trained model is denoted as the weights generated from very large

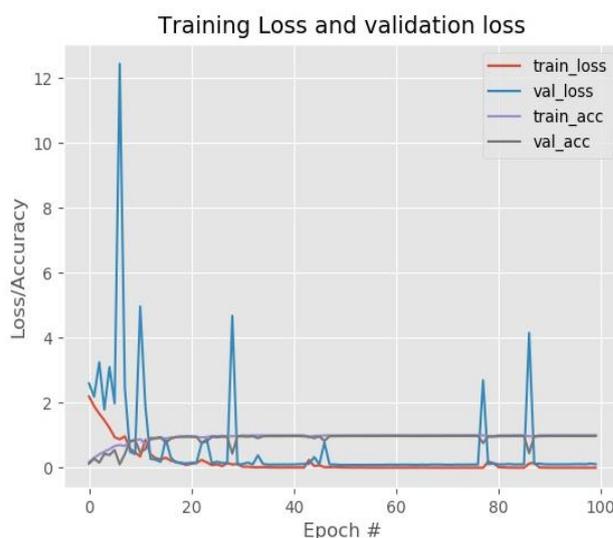


Figure. 7 Proposed CNN model evaluation report

ImageNet dataset with 1000 classes were used to solve the problems associated with the other different dataset.

The proposed WS and patch-based CNN model was compared with five existing transfer learning techniques and the performance of the proposed model was found to be far better than other existing transfer learning models based on ten class F1-score as shown in Fig. 10 and true positive rates calculated as shown in Fig. 11. Existing CNN models like Inception V3, MobileNet V2, VGG 16, VGG19, and Xception were used to validate the proposed model performance.

## 5.2 Discussion on existing pre-trained models

**MobileNet V2:** Mobile architecture model is specially designed for dimensionality reduction using linear convolution operation. Light-weighted convolution filters were used to convert low level features extracted from the image in to high level features. The model was more effective in reducing the computation cost after using 3x3 convolution filter. But it is reported that the classification accuracy falls down with less improvements as drawback [29].

**Inception V3:** Inception V3 is the third version of inception deep learning model which is capable to perform multiple class classification [30].

**VGG:** VGG is known as very deep convolution neural network introduced by Karen Simony and Andrew Zisserman from oxford university in 2014 and it is announced as best deep learning model. VGG network follows very structured model representation like using ReLU in each layer after 3x3 convolution operation and uses gradient or stochastic gradient in each training stack [31]. It is reported that the performance of VGG was outstanding in ImageNet visual recognition challenge held on 2014.

**Xception:** Xception is the advanced version of Inception deep learning model which contains 36 layers for feature extraction. The architecture starts with 1x1 convolution filter followed by 3x3 and so

on. The model achieved maximum of 87 % classification accuracy by applying logistic regression in dense layer [32]. Table 3 shows the comparative analysis of existing and proposed model loss & accuracy validation report.

## 6. Discussion

Our proposed CNN model utilizes the state-of-the-art golden characteristics which exhibits its golden standard performance in three existing pretrained deep learning models like 1). VGG model with SoftMax activation function is used in dense layer to achieve good classification accuracy. 2). 3x3 and 1x1 convolution filters which is used in Xception model to reduce over fitting problem is used in our model also to extract minimum and powerful trainable parameters. 3). MobileNetv2, our proposed model convolution operation is initialized with 32x32 filter size to extract all the curves and edges of the biopsy image as feature map. Using existing pre-trained deep networks characteristics our proposed state-of-the-art CNN model is efficiently upgraded and works well for the new UADT tumour grade dataset with 16172 images and achieved 98 % accuracy for ten class classification. In literature most of the pre-trained models worked well for different applications with large dataset and proved its performances were better. Since our UADT tumour grade dataset consists very less images and it was not efficient to fit the existing models and leads to over fitting problem, except VGG16 and VGG19 network other three pre-trained models' performance is reported with zero F1-score and zero sensitivity in all the ten classes as shown in Fig 10. And also, Inception V3 reported that there are no improvements in accuracy or reduction in training and validation error even after the number of epochs increased to 100. Though VGG's performance was good and acquired 90 % accuracy the evaluation report is recorded with high validation error 0.72 and each epoch execution time is noted with 180 seconds as shown in table 2. Our proposed model works well by

Table 3. Comparison of UADT tumour grade classification results from existing and proposed CNN model

CNN-models	No of Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy in %
Inception V3	100	2.28	0.16	2.28	16 %
MobileNet V2	100	0.71	0.74	0.55	81 %
VGG16	30	0.72	0.73	0.30	93 %
	100	0.78	0.71	0.28	89 %
VGG19	30	0.72	0.75	0.34	91 %
	100	0.83	0.69	0.17	94 %
Xception	100	1.43	0.46	1.03	67 %
Proposed model	100	2.67.e <sup>-4</sup>	0.99	0.10	98 %

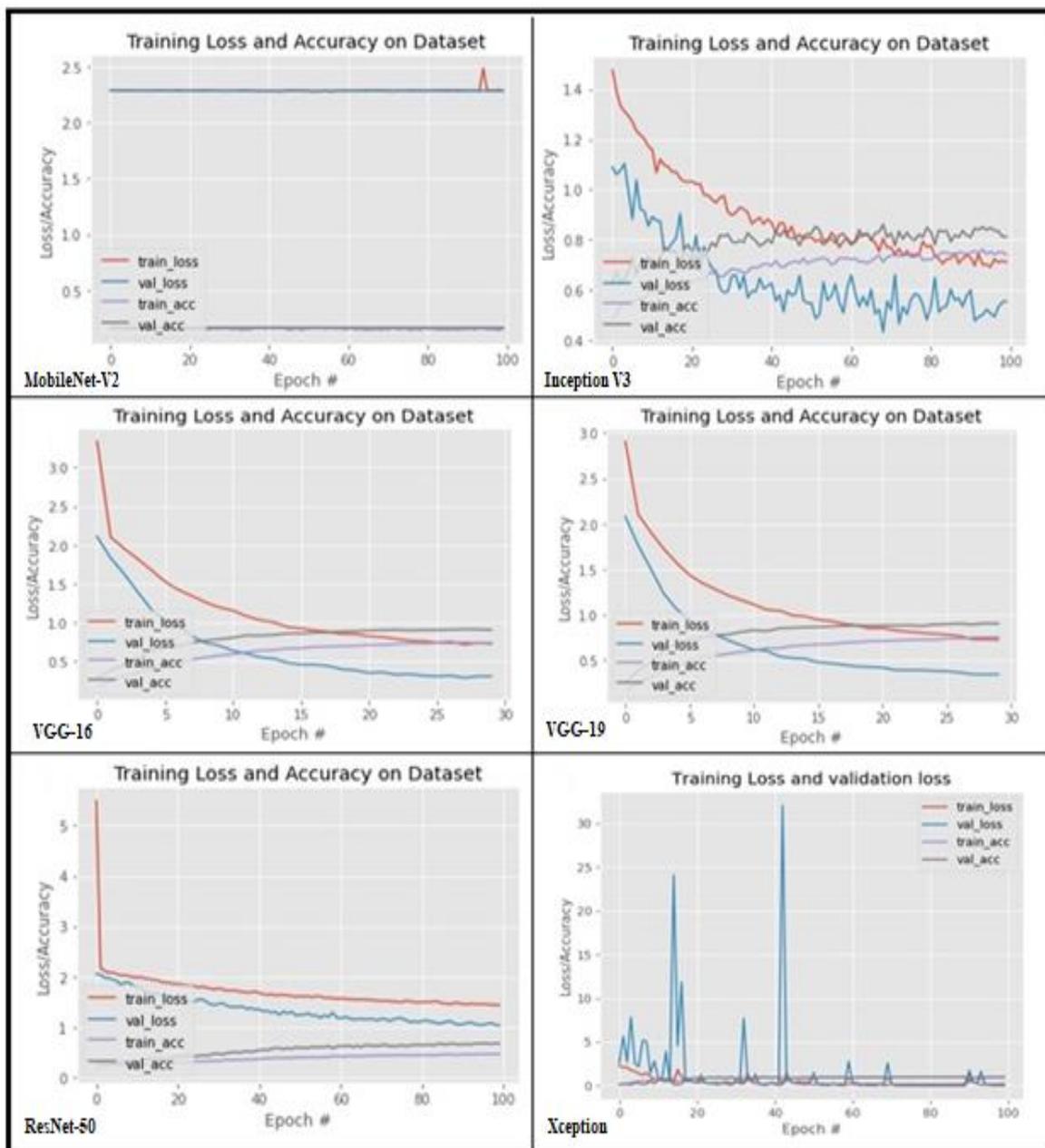


Figure. 8 Performance analysis of existing and proposed CNN models

applying, 1. Decorrelation contrast stretching technique to normalize the noise and unwanted background details like air cracks which occurs during slide preparation and excess of wax present in the slide during tissue fixation. 2. Data augmentation technique is used to equally balance the training data to avoid underfitting problem for UADT biopsy tumour grade classification.

Fig. 8 shows the training and validation performance evaluation graphs of Inception V3, MobileNet V2, VGG16, VGG19, Xception and Proposed WS and patch-based CNN models. The existing transfer learning models took maximum of 100 epochs and 85 steps for each epoch to complete

its entire training process and noted with different training accuracy, validation accuracy with different execution time. In Inception V3 model the experimental results obtained from our UADT dataset shows less performance with zero precision, recall and F1 score on both training and testing data of ten classes. The result analysis clearly shows that the model falls under underfitting problem because the model is not sufficient enough to perform proper training due to lack of parameters or the model trained with huge non-trainable parameters.

Table 4 shows the comparative analysis of proposed model with existing literature and it was observed that the performance of proposed model

Table 4. Comparative analysis of proposed model accuracy with literature

Sl. no	Contribution	CNN model	UADT site / Dataset	Accuracy / F-1 score
1	[33]	ResNet 101	Oral cancer	87.07 %
2	[34]	Alex Net CNN	Head & neck cancer	94 %
3	[35]	Faster R-CNN	Oesophageal cancer	94 - F1 score
4	[36]	CNN with grid	Oesophageal cancer	83 %
5	[37]	CNN with O&P steam	Oesophageal cancer	85.83 %
6	Proposed work	WS & patch-based CNN	Seven UADT oral sites with 16172 images.	98 % & 0.10 validation loss

works well when compared to other existing literature for different UADT cancer images. The proposed model contains 6 different oral sites with different gradings and it is compared with the existing literature sites like, oral cancer (includes all oral cavity sites), head & neck cancer (includes all oral cavity sites) and oesophageal cancer. Out of all the five literatures in table 4, our model performance rate is high for all oral cavity sites along with oesophageal cancer.

## 7. Conclusion

In this paper we proposed state-of-the-art supervised Whole Slide and patch-based convolution neural network model to classify malignant and premalignant oral cavity tumour gradings. The proposed WS and patch-based CNN model will assist the pathologists to solve interobserver variability problem which arise while grading dysplasia type of tumours manually. The WS and patch-based CNN classification model is trained with the best parameters extracted from the pre-processed biopsy images. To design the proposed model, we utilized standard convolution filters, activation functions, pooling operations from different transfer learning models. Oral cavity tumour grade classification experimental results are recorded with high-level performance of 98 % accuracy with less computational time compared to other pre-trained transfer learning models like Inception V3, MobileNet V2, VGG16, VGG19 and Xception. Our proposed automated tumour grade classification model is well trained to classify dysplasia type

tumours with noise and stain artifacts. The proposed model is capable enough to classify benign and malignant tumour gradings automatically. Also, the model will be able to predict the metastasis level for further treatment. The appearance of dysplasia at one site shows the increased chances of metastasis to another sites. To avoid such severe complication, the proposed work must be improved to classify both oral cavity and nasal cavity UADT tumour grading sites. The future work will be focused on building the CNN automated classification system for UADT tumour gradings for both oral cavity and nasal cavity sites.

## Conflicts of Interest

“The authors declare no conflict of interest.”

## Author Contributions

“Conceptualization, Prabhakaran Mathialagan and Malathy Chidambaranathan; methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Prabhakaran Mathialagan; review and editing, Prabhakaran Mathialagan and Malathy Chidambaranathan”.

## Acknowledgments

The authors would like to acknowledge Ministry of Electronics and Information Technology (MeitY's), Government of India for funding the research work under Visvesvaraya PhD scheme. They also acknowledge for the biopsy image dataset collection support received from “Government Vellore Medical College & Hospital, Department of Pathology, Tamil Nadu, India.

## References

- [1] <https://www.who.int/news-room/fact-sheets/detail/cancer>, 2018. Cancer. USA: WHO.
- [2] [https://www.who.int/cancer/country-profiles/Global\\_Cancer\\_Profile\\_2020.pdf](https://www.who.int/cancer/country-profiles/Global_Cancer_Profile_2020.pdf).
- [3] <https://www.healio.com/news/hematology-oncology/2017.eight-updates-in-oral-head-and-neck-cancer>.
- [4] A. B. Krishna, A. Tanveer, P. V. Bhagirath, and A. Gannepalli, “Role of artificial intelligence in diagnostic oral pathology-A modern approach”, *Journal of Oral and Maxillofacial Pathology*, Vol. 24, No.1, pp. 152-156, 2020.
- [5] M. Talo, “Automated classification of histopathology images using transfer learning”, *Artificial Intelligence in Medicine*, Vol. 101, pp. 101743-101751, 2019.

- [6] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, "Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides", *JAMA Network Open*, Vol. 2, No. 11, pp. 1-13, 2019.
- [7] R. D. Coletta, W. A. Yeudall, and T. Salo, "Grand challenges in oral cancers", *Frontiers in Oral Health*, Vol. 1, No. 3, 2020.
- [8] A. Arjmand, C. T. Angelis, V. Christou, A. T. Tzallas, M. G. Tsiouras, E. Glavas, R. Forlano, P. Manousou, and N. Giannakeas, "Training of deep convolutional neural networks to identify critical liver alterations in histopathology image samples", *Applied Sciences*, Vol. 10, No.1, pp.42-60, 2020.
- [9] L. Alzubaidi, O. A. Shamma, L. Farhan, J. Zhang, and Y. Duan, "Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model", *Electronics*, Vol. 9, No. 3, pp. 1-21, 2020.
- [10] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours", *Scientific Reports*, Vol. 10, No. 1, pp. 1-11, 2020.
- [11] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection", *Tissue and Cell*, Vol. 58, pp. 76-83, 2019.
- [12] D. Jiang, J. Liao, H. Duan, Q. Wu, G. Owen, C. Shu, and Z. Wang, "A machine learning-based prognostic predictor for stage III colon cancer", *Scientific Reports*, Vol. 10, No.1, p.1-9, 2020.
- [13] L. Hou, R. Gupta, J. S. V. Arnam, Y. Zhang, K. Sivalenka, D. Samaras, T. M. Kurc, and J. H. Saltz, "Dataset of segmented nuclei in haematoxylin and eosin-stained histopathology images of ten cancer types", *Scientific Data*, Vol. 7, No. 1, pp. 1-12, 2020.
- [14] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification", *Computer Methods and Programs in Biomedicine*, Vol. 195, No.1, pp. 1-19, 2020.
- [15] S. Sornapudi, J. Hagerty, R. J. Stanley, W. Stoecker, R. Long, S. Antani, G. Thoma, R. Zuna, and S. Frazier, "EpithNet: Deep regression for epithelium segmentation in cervical histology images", *Journal of Pathology Informatics*, Vol. 11, No.1, pp. 1-10, 2020.
- [16] L. Ma, M. Halicek, X. Zhou, J. Dormer, and B. Fei, "Hyperspectral microscopic imaging for automatic detection of head and neck squamous cell carcinoma using histologic image and machine learning", *Digital Pathology, International Society for Optics and Photonics*, Vol. 11320, 2020.
- [17] J. Lu, N. Sladoje, C. R. Stark, E. D. Ramqvist, J. M. Hirsch, and J. Lindblad, "A deep learning-based pipeline for efficient oral cancer screening on whole slide images", *International Conference on Image Analysis and Recognition*, Vol. 12132, pp- 249-261, 2020.
- [18] N. Das, E. Hussain, and L. B. Mahanta, "Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network", *Neural Networks*, Vol. 128, pp. 47-60, 2020.
- [19] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning", *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1285-1298, 2016.
- [20] J. Qu, N. Hiruta, K. Terai, H. Nosato, M. Murakawa, and H. Sakanashi, "Gastric pathology image classification using stepwise fine-tuning for deep neural networks", *Journal of Healthcare Engineering*, Vol. 2018, pp. 1-13, 2018.
- [21] R. K. Gupta, M. Kaur, and J. Manhas, "Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium", *Journal of Multimedia Information System*, Vol. 6, No. 2, pp. 81-86, 2019.
- [22] S. A. Taqi, S. A. Sami, L. B. Sami, and S. A. Zaki, "A review of artifacts in histopathology", *Journal of oral and maxillofacial pathology*, Vol. 22, No. 2, pp. 279-295, 2018.
- [23] R. Alley, "Algorithm Theoretical basis document for: decorrelation stretch", 1999.
- [24] M. Kalfon and M. Porat, "A new approach to texture recognition using decorrelation stretching", *International Journal of Future Computer and Communication*, Vol. 2, No.1, pp. 49-53, 2013.
- [25] M. Hägele, P. S. Philipp, S. Lapuschkin, M. Bockmayr, W. Samek, F. Klauschen, K. R. Müller, and A. Binder, "Resolving challenges in deep learning-based analyses of histopathological images using explanation methods", *Scientific Reports*, Vol. 10, No. 1, pp. 1-12, 2020.

- [26] J. E. L. García, C. Y. Márquez, Y. V. Rey, and O. C. Nieto, “A transfer learning method for pneumonia classification and visualization”, *Applied Sciences*, Vol. 10, No. 8, pp. 2908-2925, 2020.
- [27] A. V. Oprobok, M. A. Ikram, M. W. Vernooij, and M. D. Bruijne, “Transfer learning improves supervised image segmentation across imaging protocols”, *IEEE Transactions on Medical Imaging*, Vol. 34, No. 5, pp.1018-1030, 2014.
- [28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks”, In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018.
- [29] Y. A. Alsabahi, L. Fan, and X. Feng, “Image Classification Method in DR Image Based on Transfer Learning”, *International Conference on Image Processing Theory, Tools and Applications*, pp. 1-4, 2018.
- [30] E. Cengil and A. Çinar, “Multiple classification of flower images using transfer learning”, *International Artificial Intelligence and Data Processing Symposium*, pp. 1-6, 2019.
- [31] X. Liu, Q. Hu, Y. Cai, and Z. Cai, “Extreme learning machine-based ensemble transfer learning for hyperspectral image classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, pp. 3892-3902, 2020.
- [32] E. Ayan and H. M. Ünver, “Diagnosis of pneumonia from chest X-ray images using deep learning”, *Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science*, pp. 1-5, 2019.
- [33] R. A. Welikala, P. Remagnino, J. H. Lim, C. S. Chan, S. Rajendran, T. G. Kallarakkal, and R. B. Zain, “Automated detection and classification of oral lesions using deep learning for early detection of oral cancer”, *IEEE Access*, Vol. 8, pp.132677-132693, 2020.
- [34] M. Halicek, J. V. Little, X Wang, M. Patel, C. C. Griffith, M. W. E. Deiry, A. Y. Chen, and B. Fei, “Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks”, *Journal of Biomedical Optics*, Vol. 24, No. 3, pp. 1-9, 2019.
- [35] N. Ghatwary, M. Zolgharni, and X. Ye, “Early esophageal adenocarcinoma detection using deep learning methods”, *International Journal of Computer Assisted Radiology and Surgery*, Vol. 14, No. 4, pp. 611-621, 2019.
- [36] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, “Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides”, *JAMA Network Open*. Vol. 2, No. 11, pp. 1-13, 2019.
- [37] G. Liu, J. Hua, Z. Wu, T. Meng, M. Sun, P. Huang, X. He, W. Sun, X. Li, and Y. Chen, “Automatic classification of esophageal lesions in endoscopic images using a convolutional neural network”, *Annals of Translational Medicine*, Vol. 8, No. 7, pp. 1-10, 2020.