IMPROVEMENT IN TONGUE COLOR IMAGE ANALYSIS FOR DISEASE IDENTIFICATION USING DEEP LEARNING BASED DEPTHWISE SEPARABLE CONVOLUTION MODEL

S. Rajakumaran
Research scholar, Dept. of Computer Science and Engineering,
Annamalai University, Annamalai Nagar-608002
srajakumaransubashini@gmail.com

Dr. J. Sasikala
Research supervisor, Associate Professor, Dept. of Information Technology,
Annamalai university, Annamalai Nagar-608002
sasikala.au@gmail.com

Abstract - Disease diagnosis using tongue color image is a traditional non-invasive method widely employed to determine the status of the patient's internal organ. The elimination of dependencies on subjective and expert knowledge assessment for tongue diagnosis might considerably raise the scope of wide utilization of tongue diagnosis over the globe, including Western medicine. Computer based tongue diagnosis connected to light estimation, color correction, tongue segmentation, image analysis, geometry analysis, etc is a proficient method to diagnose diseases. This paper introduces a new Deep Learning with Depthwise Separable Convolution (Xception) Model called DLXM for tongue color image analysis. The presented model involves data augmentation and bilateral filtering (BF) based noise removal at the preprocessing stage. Additionally, the DLXM is applied for feature extraction process. At last, the bagging classifier (BC)and multilayer perceptron classifier (MLPC) models are employed to categorize the feature vectors into distinct types of diseases. The performance of the presented model is evaluated against benchmark tongue image dataset and the results depicted the effectual classification performance on the applied images. The experimental values notified that the DLXM-MLPC model has outperformed the compared methods by achieving a higher precision, recall, accuracy, and F1-Score of 97.35%, 97.01%, 97.01%, and 96.77% respectively.

***Keywords*:** Tongue color analysis, Deep learning, Machine learning, Feature extraction, Xception.

1. Introduction

Naturally, human tongue is comprised of a maximum number of features. In earlier days, the physicians have examined the tongue using prior medical knowledge [1]. Therefore, uncertainty and partiality are embedded in prognostic outcomes. The qualitative factors can be eliminated by examining the tongue images which is meant to be better way of disease prediction and mitigates the deficiency of a patient. Actually, tongue diagnosis is a significant process in Traditional Chinese Medicine (TCM) for last decades. Some of the tongue features like shape, texture, and color, shows the actual health condition of the patient (organs, qi, blood, temperature, heat) and severity of the diseases. Under the observation of tongue features, TCM users discriminate medical traits and select appropriate recovering procedures. But classical tongue analysis depends upon the physician experience, ecological differences, and so on. Hence, it is essential to create unbiased and assessable tongue analysing model which applies practitioner's diagnosis.

In state-of-the-art methods, the system-based tongue image prediction, color, and texture features are highly recommended and applied. No studies were developed for tongue image diagnosis by applying geometry features while in classical medicines like Traditional Chinese Medicine (TCM), shape of a tongue is applied for predicting the disease. Also, feature extraction is initialized with a collection of estimated data and produces a feature into useful and non-repeated enhances the learning process with generalization steps to human interpretation. In order to gain maximum accuracy, feature extraction has been applied for developing variable unification. Hence, classification task is related to classification in which principles and objects are examined, differentiated, and understood.

Recently, diverse types of previous automated tongue segmentation methods were presented as a portion of comprehensive application. For instance, Bi-Elliptical Deformable Contour (BEDC) unifies model-reliant approaches, as well as Active Contour Models (ACM). This model has attained promising segmentation outcomes when the quality of segmentation depends upon the former experiences and sensitive to location as well as primary curves produced from the tongue. In order to resolve this problem, the model-reliant schemes with region merging principle in order to gain coarse segmentation outcomes. Moreover, ACM is applied as a post-processing phase to accomplish considerable segmentation process when related to BEDC. Also, developers in [2] depend upon the previous experience as the position details of primary marker has to be defined manually. Evolved from the efficiency of region combining principle, a combination technology with the help of region-based as well as edge-based which again eliminates the influences of noises in tongue image and maximizes the segmentation outcome and efficiency. Consequently, the model suffers from insufficient efficacy in various scenarios.

A 3-stage method has been developed in [3] for using the concavity data to identify the malicious regions. In Convolutional Neural Network (CNN) is helpful in deriving a deep feature. Regardless, the region generating approach still remains the same with no improvement. It is composed of maximum true tooth-marked regions and minimum non-tooth-marked regions. Under the application of Tongue Images (TI), new technology has been developed in [4] for the purpose of identifying a constitution. The tongue coating prediction, calibration, and constitution analysis can be processed using deep CNN (DCNN). Therefore, multi-label learning is considered to be infeasible with noisy results. A Conceptual Alignment Deep Autoencoder (CADAE) was deployed in [5] to predict the TIs which points the varied Body Constitution (BC) types with TCM models.

Also, non-invasive framework was introduced in [6] for evaluating DM and Non-Proliferative Diabetic Retinopathy (NPDR) reliant on features gathered TIs. But non-invasive application has exhibited low accuracy in anatomical resolution. In order to perform diagnostic feature extraction, in-depth analysis has been deployed in [7]. The utilization of tongue color space is not applicable in systematic TI examination. CIELAB-reliant K-means clustering model was employed in [8] for investigating the color variations in 3D space. TI is attained from DS01-B tongue color data acquisition mechanism. Thus, the scalability of tongue estimation method still remains ideal with no advancements. It is also unable to measure the K-value. A non-invasive application is deployed in [9] to implement the auxiliary prediction in order to retain global constraints in medical sector. Therefore, the inclusion of optical units makes the process more costly.

This paper designs a novel Deep Learning (DL) with Depthwise Separable Convolution (Xception) Model called DLXM for tongue color image analysis. The presented model involves data augmentation and bilateral filtering (BF) based noise removal at the preprocessing stage. Moreover, the DLXM is applied for feature extraction process. Finally, the bagging classifier (BC) and multilayer perceptron classifier (MLPC) models are employed to categorize the feature vectors into distinct types of diseases. The classification results of the presented model are evaluated against benchmark tongue image dataset and the results depicted the effectual classification performance on the applied images.

2. The Proposed Model

The working process involved in the presented model is depicted in Fig. 1. The figure portrays that the input tongue image is primarily preprocessed to augment the data and remove noise. Followed by, the DLXM based feature extraction model is employed for extracting a useful set of feature vectors. Eventually, the BC and MLPC models are utilized to recognize the respective class labels of the input image.

2.1. *Image Preprocessing*

At this point, the input images undergo data augmentation process to enlarge the size of the training dataset. Followed by, BF technique is applied as a tool to remove the noise exist in it. Assume F implies a multichannel image and suppose W is a sliding window of definite size $n \times n$. Let a pixel in W implied in Cartesian Coordinates and implied $\mathrm{u} = (u_1, u_2) \in Y^2$ the location of the pixel $\mathrm{F_u}$ in W where $Y = \{0, 1, \ldots, n-1\}$ is endowed in conjunction with normal order. Based on [10], BF substitutes the middle pixel of a filtering window by weighted average of corresponding neighbor color pixels. A weighting function is developed for smoothing of same colors where the edges are maintained by heavy weight of pixels which are spatially identical and photometrically same as the middle pixel.

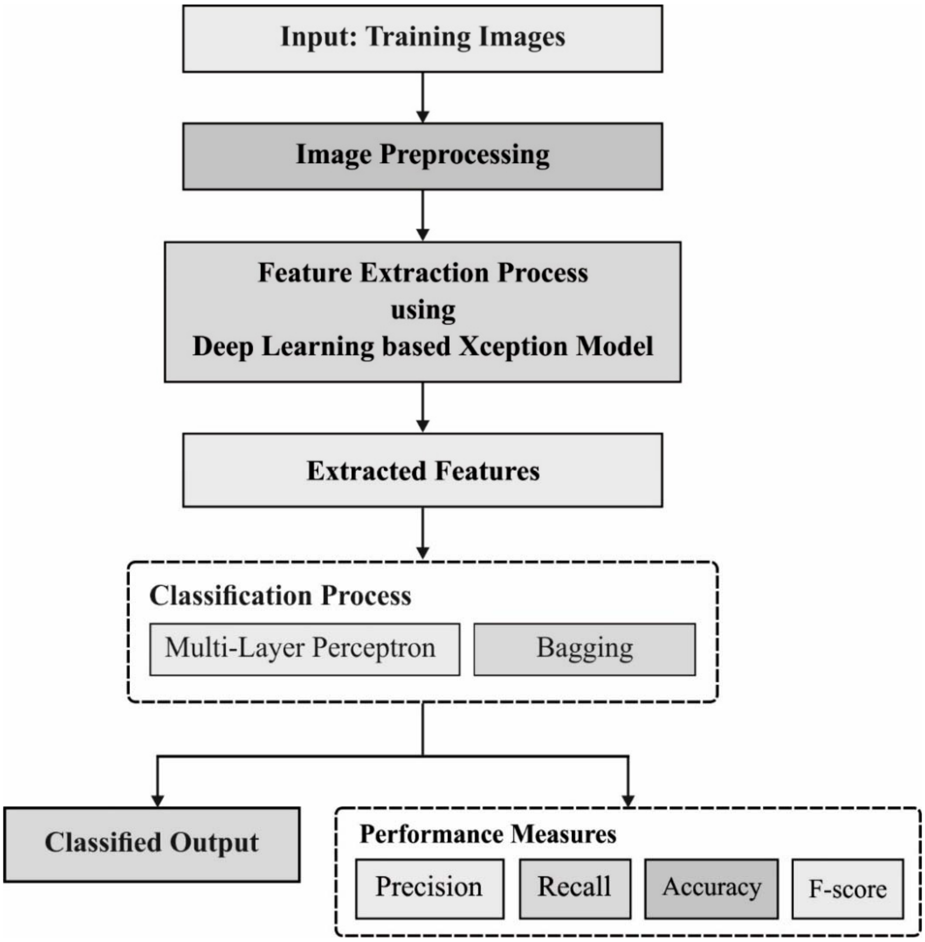


Fig. 1. Overall working process of proposed method.

With the application of $||\cdot||_2$, Euclidean term and $\mathrm{F_u}$ is a middle pixel. Followed by, weight $\mathcal{W}(\mathrm{F_u}, \mathrm{F_v})$ corresponds the pixel $\mathrm{F_v}$ by means of $\mathrm{F_u}$ is a product of 2 units, a spatial and photometrical objective,

$$\mathcal{W}(\mathrm{F_u}, \mathrm{F_v}) = \mathcal{W}_s(\mathrm{F_u}\,, \mathrm{F}_v)\mathcal{W}_p(\mathrm{F_u}, \mathrm{F_v}) \tag{1}$$

In spatial component $\mathcal{W}_s\ (\mathrm{F}_u\,, \mathrm{F_v})$ is depicted by

$$\mathcal{W}_s(F_u, F_v) = e^{-\frac{||\mathrm{u-v}||_2^2}{2\sigma_S^2}} \tag{2}$$

and photometrical component $\mathcal{W}_p(\mathrm{F_u}, \mathrm{F_v})$ is depicted using,

$$\mathcal{W}_p(\mathrm{F_u}, \mathrm{F_v}) = e^{-\frac{\Delta E_{Lab}(\mathrm{F_u}, \mathrm{F_v})^2}{2\sigma_p^2}} \tag{3}$$

where $\Delta E_{Lab} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{\frac{1}{2}}$ indicates the perceptual color error in $L^*a^*b^*$ color space as well as $\sigma_s, \sigma_p > 0$. A color vector result $\widetilde{Fu}$ of a filter is processed under the application of normalized weights and it is demonstrated by

$$\overline{\mathrm{F_u}} = \frac{\sum_{F_V \in w} \mathcal{W}\,(\mathrm{F_u}, \mathrm{F_v})\mathrm{F_v}}{\sum_{F_V \in w} \mathcal{W}\,(\mathrm{F_u}, \mathrm{F_v})} \tag{4}$$

In $\mathcal{W}_s$ weighting function reduces the spatial distance from u and v, and $\mathcal{W}_p$ weighting function limits the perceptual color variations among color vectors enhances. In spatial unit mitigates the control of future pixels limiting the blurring whereas photometric component decreases influence of pixels that are perceptually applicable. Followed by, the perceptual areas of pixels are collected and sharpness of edges are conserved. The attributes σ_s and σ_p are applied for adjusting influence of spatial as well as photometric units, correspondingly. It is assumed as a rough threshold to find pixels which are identical to middle one. It is pointed as $\sigma_p \to \infty$ the BF models Gaussian filter and if $\sigma_s \to \infty$ the filter frameworks a range filter without spatial function. For the expressions $\sigma_p \to \infty$ and $\sigma_s \to \infty$ a BF acts as an AMF.

2.2. *DLXM based Feature Extraction*

Once the input tongue image is preprocessed, the DLXM model is applied as a feature vector to derive the actual set of feature vectors. The frequent deployment of DL of CNN has enhanced the structure for accomplishing précised image classification models. Likewise, Xception structure is introduced under various concepts such as convolutional, depth-wise separable convolution layer, inception method, and residual connections.

Xception [11] is meant to be a hypothesis reliant Inception module used in developing correlations of cross-channel as well as spatial relations inside feature maps of CNN that is capable of isolating the model. The typical Inception module from Inception v3, a module which has employed cross-channel relations by isolating the input data in 4 phases for convolution size of 1 x 1, average pooling, maps correlations of convolution size 3 x 3 and send them for combination. Based on the Inception module, the principle is to convert Xception technology. Once the input is attained, data by applying 1 x 1 convolution develops unique convolution layer with no average pooling and computed in non-overlapping of output channels to induce the combination. Hence, Xception method is effective when compared with the Inception module and computed correlations of cross-channel as well as spatial correlations using fully decoupled. Once the module is attained, existing principle of depth-wise separable convolution has been employed to develop the Neural Network (NN) and the composition within Xception structure as described in the following.

Convolutional Layer

In using convolutional layers within the Xception structure, a layer placed next input layer which generates convolutional kernels for estimating diverse feature maps to exhibit the features of input data. A novel feature map is gathered by primary convolution task with prediction outcomes from convolutional kernels and feds the outcomes to estimation of an activation function. In order to generate the feature map, convolution kernels are classified as input details. The various convolution kernels develop the exact outcomes of feature maps, the position (i, j) upon feature measures in a feature map as kth layer computes the lth, is estimated as,

$$S_{i,j,k}^{l} = \mathrm{W}v_k^l C_{i,j}^l + Bv_k^l \tag{5}$$

In which, weight vector is described as Wv_k^l as well as Bv_k^l, is set of bias value of kth filter of lth layer, for $c_{i,j}^l$ as middle of input patch on (i, j) location of lth layer. In distributing a feature map of $S_{i,j,k}^l$, it develops the estimation of Wv_k^l kernel. The merits of weight sharing operation which limiting the complexities and enhance the network performance of the scheme. The convolutional layer of Xception is included with Batch Normalization (BN) and activation function, and actual activation function is ReLU in applied function:

$$ReLu(d) = \max(d, O) \tag{6}$$

Where d indicates the input data. It is linear under positive and zero for negative measures. ReLU is not complicated math with nonlinearity of a system which is significant in CNN to find the nonlinear features which make robust convergences and optimal predictions with minimum less overfitting.

Depth-Wise Separable Convolution Layer

An important layer of Xception is depth-wise separable convolutions. It reduces the processing and process variables that are arranged in spatial dimensions as well as depth dimensions of colors. It can be processed by dividing from classical convolution procedure with depth-wise convolution connected to point-wise convolution by developing a convolution kernel size that is operated with depth-wise separable convolution. The feature map used for determining $D_F \times D_F \times M$ as well as depth-wise convolution under the application of filter of an input channel processed by subsequent expression:

$$\hat{G} = \sum_{i,j,m} \hat{K}_{i,j,m} \times F_{k+i-1,\mathrm{p+j-1,m}} \tag{7}$$

where $\hat{G}$ replaces the resultant of feature maps produced by F that is a feature map input $\hat{K}$ refers to the depth-wise convolution kernel. The mth filter in $\hat{K}$ is applied for channel of mth in F to evaluate the feature map result. The pixel location of convolution kernel induces into i, j, and pixel location of feature map describes k, p.

Fig. 2 illustrates the 3 color channels of Red, Blue, and Green (RBG) have been gathered by isolation of depth-wise convolution filters [12]. Once the convolution is completed, image is displayed from various channels, and image is interpreted in all color channels. Next, point-wise convolution provides the result for upcoming layer process. For Xception, once the depth-wise separable convolution layer applies BN, next layers apply max - pooling layer for limiting the expense of processing and assist for interpreting invariance by allocating function as:

$$F_m = MaxPooling(F_i, v) \tag{8}$$

Where v signifies a filter of max -pooling. The resultant feature map describe F_m that is organized in shape size, in which F_m records the maximum score of F_i in input feature map.

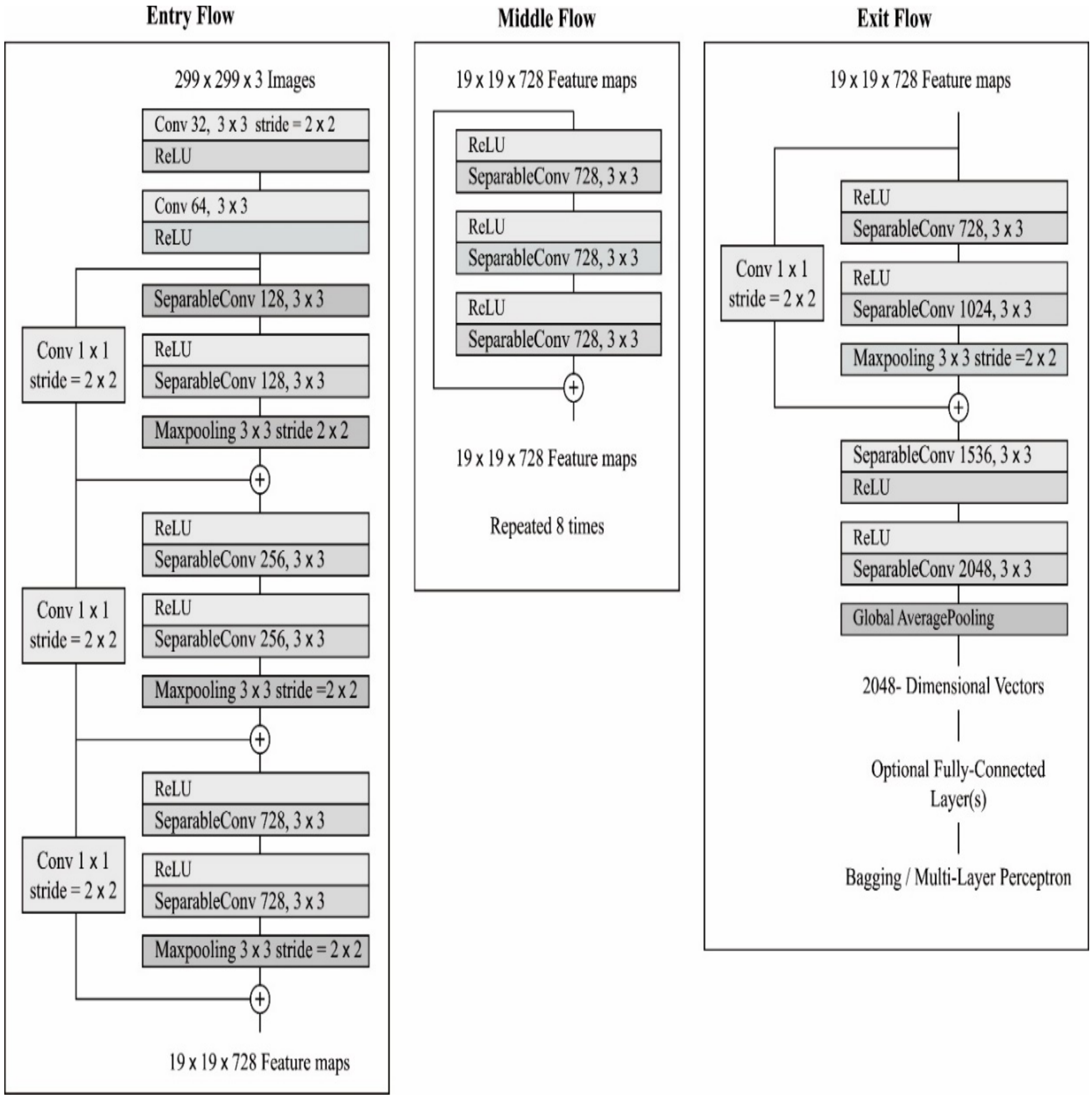

Entry Flow
299 x 299 x 3 Images
Conv 32, 3 x 3 stride = 2 x 2
ReLU
Conv 64, 3 x 3
ReLU
SeparableConv 128, 3 x 3
ReLU
SeparableConv 128, 3 x 3
Maxpooling 3 x 3 stride 2 x 2
Conv 1 x 1 stride = 2 x 2
ReLU
SeparableConv 256, 3 x 3
ReLU
SeparableConv 256, 3 x 3
Maxpooling 3 x 3 stride =2 x 2
Conv 1 x 1 stride = 2 x 2
ReLU
SeparableConv 728, 3 x 3
ReLU
SeparableConv 728, 3 x 3
Maxpooling 3 x 3 stride = 2 x 2
Conv 1 x 1 stride = 2 x 2
19 x 19 x 728 Feature maps
Middle Flow
19 x 19 x 728 Feature maps
ReLU
SeparableConv 728, 3 x 3
ReLU
SeparableConv 728, 3 x 3
ReLU
SeparableConv 728, 3 x 3
19 x 19 x 728 Feature maps
Repeated 8 times
Exit Flow
19 x 19 x 728 Feature maps
ReLU
SeparableConv 728, 3 x 3
ReLU
SeparableConv 1024, 3 x 3
Maxpooling 3 x 3 stride =2 x 2
Conv 1 x 1 stride = 2 x 2
SeparableConv 1536, 3 x 3
ReLU
ReLU
SeparableConv 2048, 3 x 3
Global AveragePooling
2048- Dimensional Vectors
Optional Fully-Connected Layer(s)
Bagging / Multi-Layer Perceptron

Fig. 2. Structure of DLXM.

2.3. *Image Classification*

Finally, the extracted feature vectors are fed into the BC and MLPC models to determine the actual class labels of the applied test images.

2.3.1. MLPC Model

The function applied in this study presents a fully connected (FC) and feedforward MLP system with a single hidden layer as depicted in Fig. 3. The MLP is composed of input layer, resultant layer, and hidden layer. It undergoes training with the help of Backpropagation (BP) learning technology. Consider that n is the count of input nodes, m implies a count of hidden nodes, and l refers the count of final nodes. Assume an input weights $w_{i,j}$ links the ith input to jth hidden unit while resultant weights $w_{\mathrm{out}(j,k)}$ connects the jth hidden unit with kth output. Hence, weighted sums of inputs are measured by the given function:

$$s_j = \sum_{i=1}^{n} \left(w_{ij} x_i \right) - \theta_j, j = \ 1, 2, \cdots \ , m, \tag{9}$$

Where n implies the count of input nodes, w_i denotes the connection weight from ith node in input layer to jthnode in a hidden layer, x_j represents the ith input, and θ_j stands for threshold of jth hidden node. The simulation of a hidden node is determined below [13]:

$$f(j) = \frac{1}{\left(1 + \ \exp\left(-s_j\right)\right)} \ j = 1, 2, \cdots, m. \tag{10}$$

Once the output of hidden nodes is measured, the consequent output is described in the following:

$$o_k = \sum_{j=1}^{m} W_{jk} \cdot f(j) - \theta'_k \; k = 1, 2, \cdots, l, \tag{11}$$

Where W_{jk} refers a connection weight from jth hidden node to kth resultant node as well as θ'_k indicates the bias of kth resultant node. A learning error E (fitness function (FF)) is determined in the following:

$$E_k = \sum_{i=1}^{l} \left(o_i^k - d_i^k\right)^2 \tag{12}$$

$$E = \sum_{k=1}^{q} \frac{E_k}{q},$$

where q refers the count of training instances, l implies the count of results, d_i^k stands for required output of ith input unit while kth training sample has been applied, and o_i^k denotes the original result of ith input unit while using kth training sample. From the given functions, it is observed that consequent value of output in MLPs is reliant on the variables of linking weights and biases. Therefore, training MLP is described as the task of identifying best measures of weights and biases of connections for accomplishing required outputs from specific inputs.

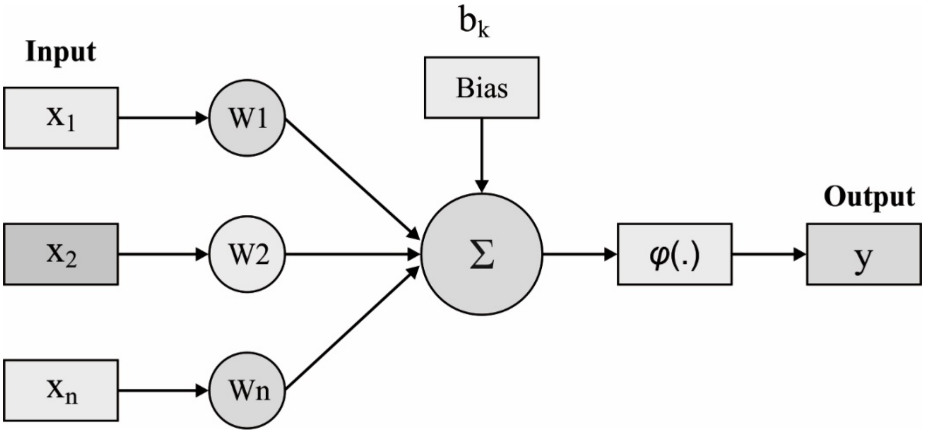


Fig. 3. Structure of MLP.

2.3.2. BC Model

Ensemble learning means the unification of various weak classification model for generating robust classifier, which makes sure the diversity of weak classifiers and enhances the generalization capability. Bagging is defined as a fundamental method of ensemble learning that examines the merits of ensemble approach. The classical bagging approach is comprised of 2 major modules like Bootstrap and Aggregation. Initially, the count of subsets is sampled in random fashion from actual training set with the help of bootstrap sampling principle [14] with substitution. Alternatively, bagging method is used in collecting the outputs of base methods with the help of voting principle for classification process. The algorithm for classical bagging is defined as Algorithm 1. Assume that a training set for C-class classification issues are applied as $D = \{(x_j, y_i) \,|x_j \in R^d, y_i \in \{1, 2, , C\}, i = 1, 2, , N\}$, where (x_i, y_i) indicates a sample encoded by d-dimensional feature vector x_i along with a class label y_i, and N represents the count of samples from a training set. Moreover, ES is considered as actual ensemble size that makes identical sampled subsets and base classifiers, L indicates the base classifier, B means an ensemble approach developed with bagging scheme, and *Bootstrap* (*D*) provides a bootstrapped subset produced from actual training set D. Fig. 4 shows the process of bagging model.

Algorithm 1: Traditional bagging algorithm

Input: D-training set, ES- count of sampled subsets, L- base learner

Output: M-a set of base models, B- bagging ensemble

```
Initialize $M = \emptyset$.
for $i \in \{1, 2, \ldots, ES\}$ do:
    Randomly produce a subset $D_j = Bootstrap(D)$
    Base model $m_i = L(D_i)$ is developed under the application of base classifier $L$ trained on subset $D_i$
    $M = M \cup \{m_j\}$
The result $B(x)$ of a test sample, $x$ examined by ensemble method $B$ is provided in the following:
    $B(x) =$ majority class in $\{m_j(x)\}_{i=1,2,\ldots,ES}$
```

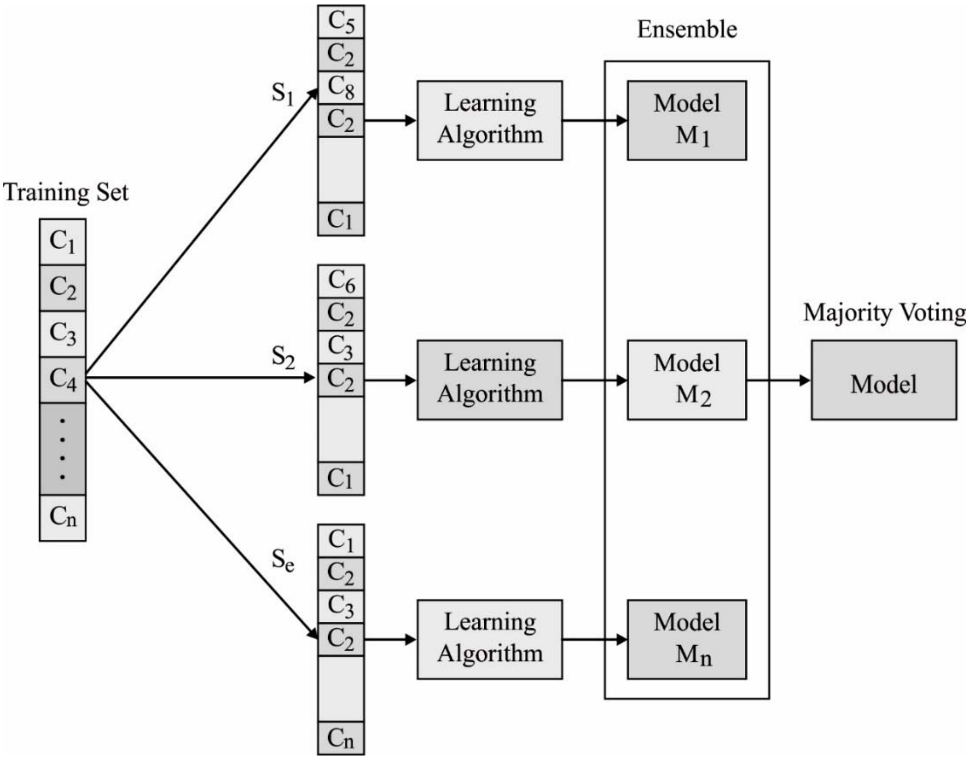


Fig 4. Bagging Model.

3. Experimental Evaluation

The proposed model is simulated on a PC i5-8600k processor, GeForce 1050Ti, 4GB RAM, 16GB OS Storage, and 250GB SSD File Storage. The simulation tool used is Python 3.6.5 tool along with some packages namely tensorflow, keras, numpy, pickle, matplotlib, sklearn, pillow, and opencv-python. The results generated during the simulation process are displayed in Appendix. The performance of the presented models is assessed against the benchmark tongue image dataset, which comprises a total of 78 images under 12 classes. The details relevant to the dataset are tabulated in Table 1 and few of the sample test images are displayed in Fig. 5.

Table 1. Dataset Description.

Diseases Name	**No. of Samples**
Chronic Kidney Disease (CKD)	78
Nephritis (NH)	78
Verrucous Gastritis (VG)	78
Pneumonia (PN)	78
Nephritis Syndrome (NSP	78
Chronic Cerebral Circulation Insufficiency (CCCI)	78
Upper Respiratory Tract Infection (URTI)	78
Erosive Gastritis (EG)	78
Coronary Heart Disease (CHD)	78
Chronic Bronchitis (CB)	78
Mixed Hemorrhoid (MH)	78
Healthy	78
Total Images	**936**

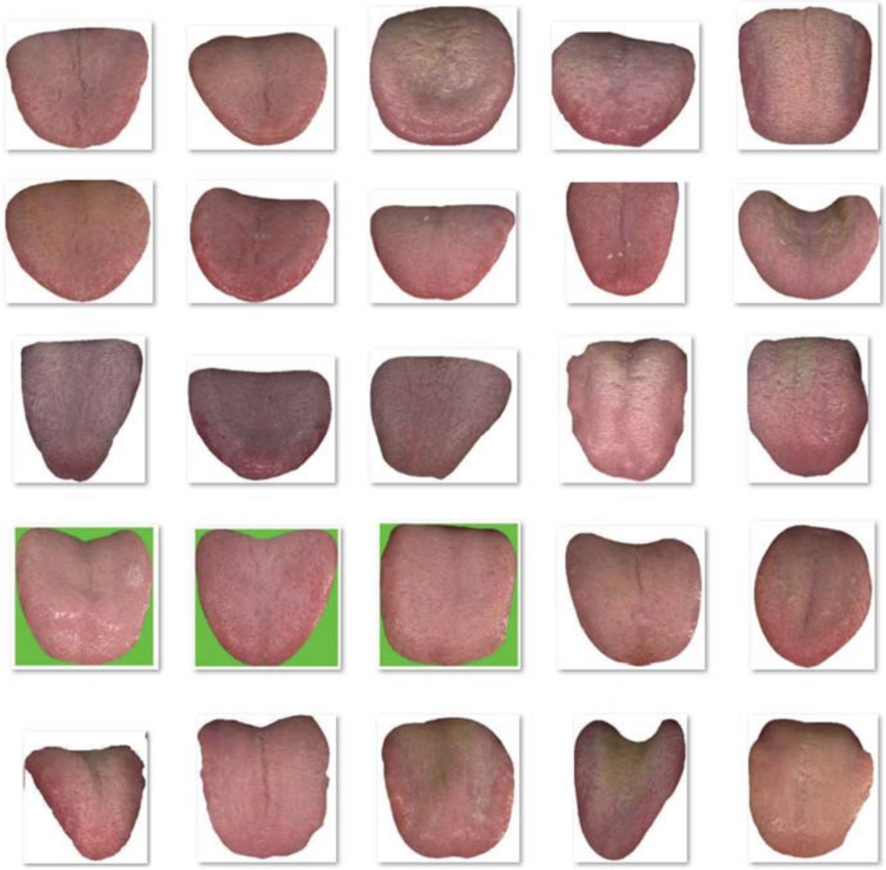


Fig 5. Sample Tongue Images.

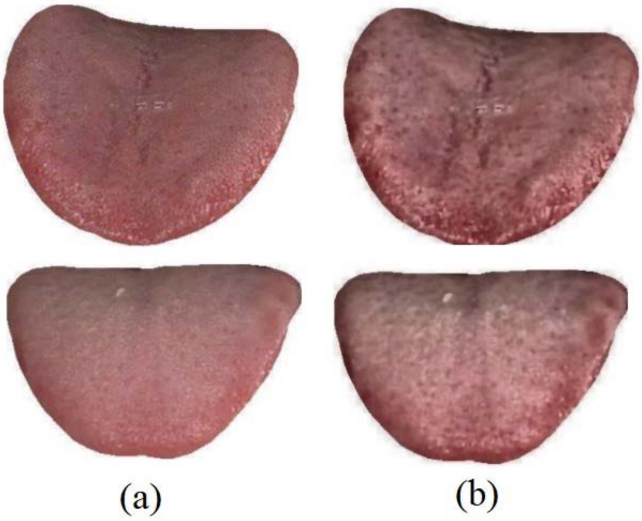
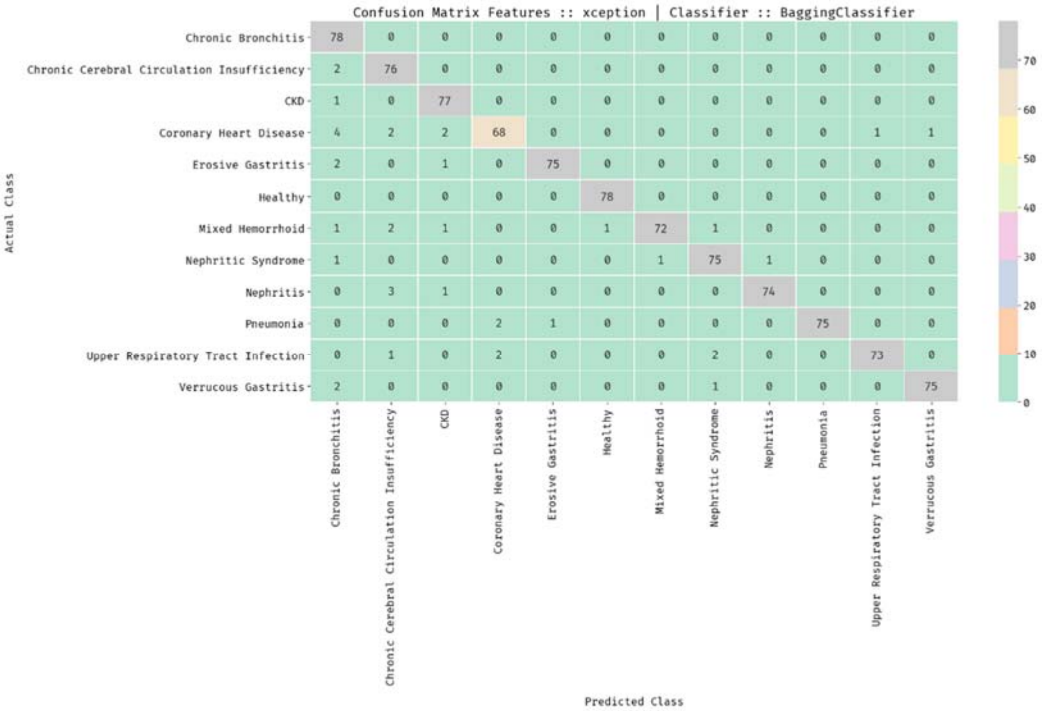


Fig. 6. a) Original Image b) Preprocessed Image.

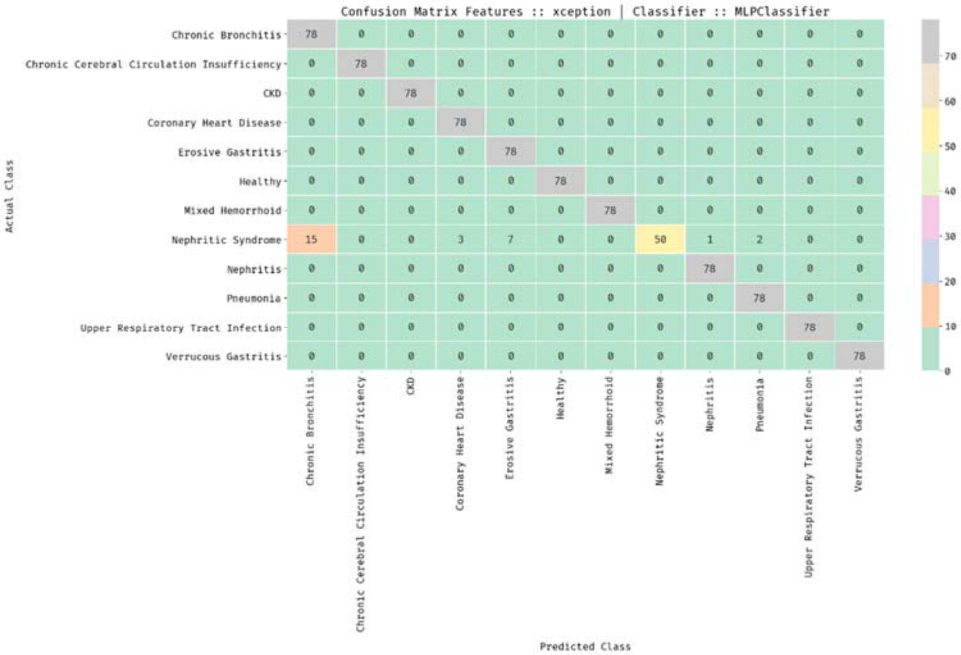
Fig. 6 shows the qualitative analysis of the presented model on the applied test images. Fig. 6a demonstrates the actual input tongue color image and its preprocessed version is depicted in Fig. 6b. The confusion matrix generated by the DLXM-BC model at the time of simulation is depicted in Fig. 7. The outcome of the figure showcased that the DLXM-BC model has effectually classified a set of 78 images into CB class, 76 images into CCCI class, 77 images into CKD class, 68 images into CHD, 75 images into EG class, 78 images into healthy class, 72 images into MH class, 75 images into NS class, 74 images into NH class, 75 images into PN class, 73 images into URTI class, and 75 images into VG class.



	Chronic Bronchitis	Chronic Cerebral Circulation Insufficiency	CKD	Coronary Heart Disease	Erosive Gastritis	Healthy	Mixed Hemorrhoid	Nephritic Syndrome	Nephritis	Pneumonia	Upper Respiratory Tract Infection	Verrucous Gastritis
Chronic Bronchitis	78	0	0	0	0	0	0	0	0	0	0	0
Chronic Cerebral Circulation Insufficiency	2	76	0	0	0	0	0	0	0	0	0	0
CKD	1	0	77	0	0	0	0	0	0	0	0	0
Coronary Heart Disease	4	2	2	68	0	0	0	0	0	0	1	1
Erosive Gastritis	2	0	1	0	75	0	0	0	0	0	0	0
Healthy	0	0	0	0	0	78	0	0	0	0	0	0
Mixed Hemorrhoid	1	2	1	0	0	1	72	1	0	0	0	0
Nephritic Syndrome	1	0	0	0	0	0	1	75	1	0	0	0
Nephritis	0	3	1	0	0	0	0	0	74	0	0	0
Pneumonia	0	0	0	2	1	0	0	0	0	75	0	0
Upper Respiratory Tract Infection	0	1	0	2	0	0	0	2	0	0	73	0
Verrucous Gastritis	2	0	0	0	0	0	0	1	0	0	0	75

Fig. 7. Confusion Matrix of DLXM-BC.

The confusion matrix produced by the DLXM-MLPC method in simulation is demonstrated in Fig. 8. The result of the figure portrayed that the DLXM-MLPC scheme has efficiently categorized a set of 78 images into CB class, 78 images as CCCI class, 78 images as CKD class, 78 images into CHD, 78 images into EG class, 78 images into healthy class, 78 images into MH class, 50 images into NS class, 78 images into NH class, 78 images into PN class, 78 images into URTI class, and 78 images into VG class.



	Chronic Bronchitis	Chronic Cerebral Circulation Insufficiency	CKD	Coronary Heart Disease	Erosive Gastritis	Healthy	Mixed Hemorrhoid	Nephritic Syndrome	Nephritis	Pneumonia	Upper Respiratory Tract Infection	Verrucous Gastritis
Chronic Bronchitis	78	0	0	0	0	0	0	0	0	0	0	0
Chronic Cerebral Circulation Insufficiency	0	78	0	0	0	0	0	0	0	0	0	0
CKD	0	0	78	0	0	0	0	0	0	0	0	0
Coronary Heart Disease	0	0	0	78	0	0	0	0	0	0	0	0
Erosive Gastritis	0	0	0	0	78	0	0	0	0	0	0	0
Healthy	0	0	0	0	0	78	0	0	0	0	0	0
Mixed Hemorrhoid	0	0	0	0	0	0	78	0	0	0	0	0
Nephritic Syndrome	15	0	0	3	7	0	0	50	1	2	0	0
Nephritis	0	0	0	0	0	0	0	0	78	0	0	0
Pneumonia	0	0	0	0	0	0	0	0	0	78	0	0
Upper Respiratory Tract Infection	0	0	0	0	0	0	0	0	0	0	78	0
Verrucous Gastritis	0	0	0	0	0	0	0	0	0	0	0	78

Fig. 8. Confusion Matrix of DLXM-MLPC.

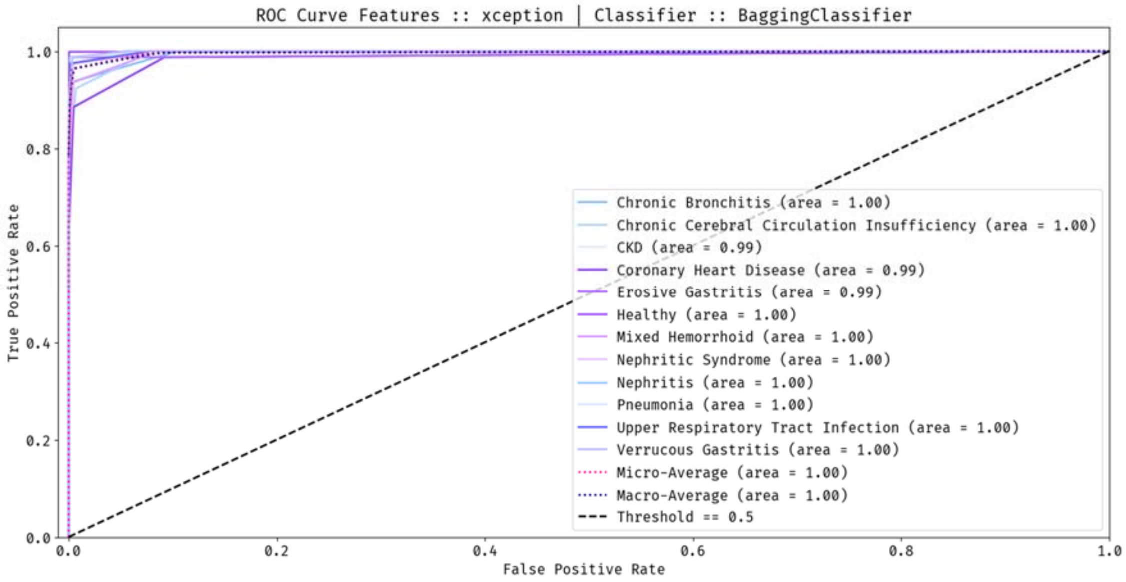


Fig. 9. ROC Analysis of DLXM-BC.

Fig. 9 showcases the ROC analysis of the presented DLXM-BC model on the classification of distinct classes on the tongue image dataset. The resultant values demonstrated the effective performance by attaining a higher ROC value of 1.0 under CB class, 1.0 under CCCI class, 0.99 under CKD class, 0.99 under CHD, 0.99 images under EG class, 1.0 images under healthy class, 1.0 under MH class, 1.0 under NS class, 1.0 under NH class, 1.0 under PN class, 1.0 under URTI class, and 1.0 under VG class.

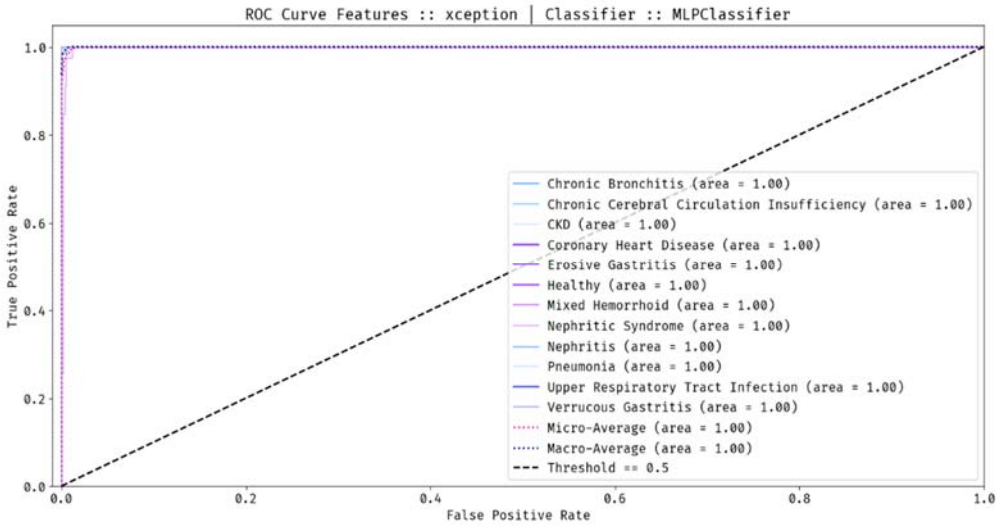


Fig. 10. ROC Analysis of DLXM-MLPC.

Fig. 10 exhibited the ROC examination of the proposed DLXM-MLPC scheme on the categorization of various classes on the tongue image dataset. The final values depicted the efficient function by gaining maximum ROC value of 1.0 under CB class, 1.0 under CCCI class, 1.0 under CKD class, 1.0 under CHD, 1.0 images under EG class, 1.0 images under healthy class, 1.0 under MH class, 1.0 under NS class, 1.0 under NH class, 1.0 under PN class, 1.0 under URTI class, and 1.0 under VG class.

Table 2. Performance Analysis of Proposed Methods in terms of Different Measures.

Methods	Precision	Recall	Accuracy	F1-Score
DLXM-MLPC	97.35	97.01	97.01	96.77
DLXM-BC	95.96	95.73	95.73	95.75

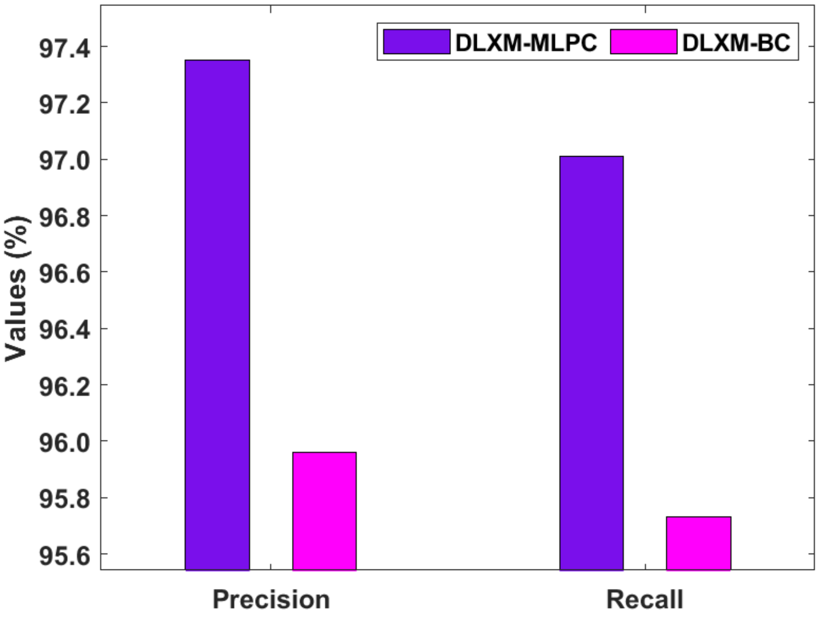


Fig. 11. Precision and recall analysis of proposed method.

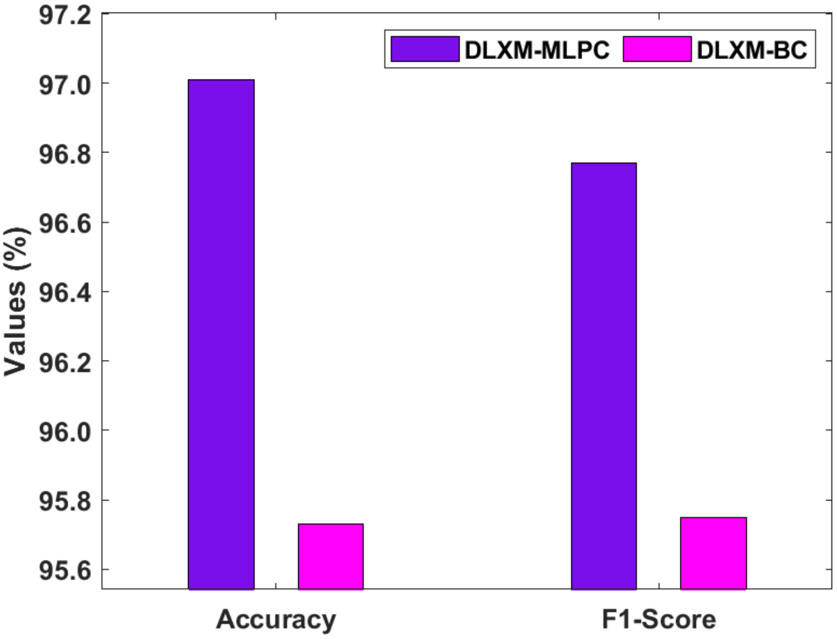


Fig. 12. Accuracy and F1-score analysis of proposed method.

Table 2 and Figs. 11-12 exhibited the results offered by the DLXM-MLPC and DLXM-BC models on the classification of tongue image dataset. The table values signified that the DLXM-MLPC model has reached a maximum precision of 97.35%, recall of 97.01%, accuracy of 97.01%, and F1-score of 96.77%. Concurrently, the DLXM-BC model has resulted in effective outcomes with a precision of 95.96%, recall of 95.73%, accuracy of 95.73%, and F1-score of 95.75%.

Table 3 and Fig. 13 provides an extensive comparative results analysis with the current state of art models [15-21] like Geometry Features+Sparse Representation Classifier (GF+SRC), Conceptual Alignment Deep Autoencoder (CADAE), KNN, GA-SVM, SVM, and VGG-SVM. The figure has implied that the VGG-SVM scheme has demonstrated inefficient classifier results with a minimum accuracy of 59.44%. Simultaneously, the KNN scheme has managed to gain moderate results over the VGG-SVM technology with an accuracy of 73.38%. Additionally, the Bayesian, Geometry features, SVM, and CADAE technologies have illustrated considerable and identical accuracy values of 75%, 76.24%, 76.46%, and 77%. In addition, GF+SRC framework has obtained acceptable accuracy measures of 79.23%. Moreover, the GA-SVM method has illustrated reasonable results with accuracy of 83.06%. However, the newly developed DLXM-MLPC and DLXM-BC approaches have achieved supreme results with accuracy of 93.7% and 92 .52%.

Table 3. Performance Analysis of Proposed Methods with Existing Methods in terms of Different Measures.

Methods	Accuracy
Proposed DLXM-MLPC	93.70
Proposed DLXM-BC	92.52
Geometry Features	76.24
Bayesian	75.00
K-NN	73.38
SVM	76.46
GA-SVM	83.06
VGG-SVM	59.44
GF+SRC	79.23
CADAE	77.00

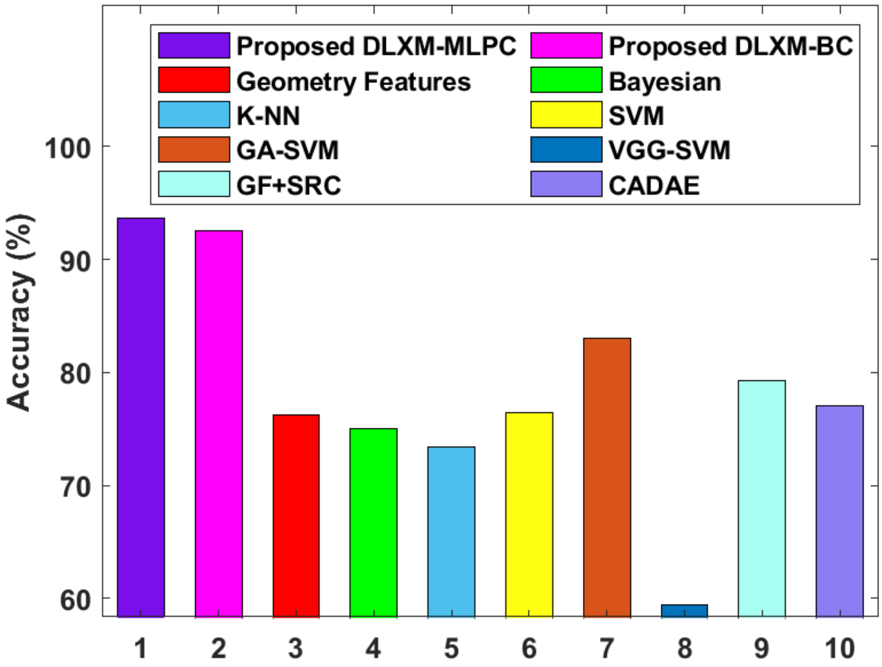


Fig. 13. Accuracy analysis of proposed method with existing methods.

4. Conclusion

This paper has designed a novel DLXM model for tongue color image analysis. Firstly, the input tongue image is preprocessed to augment the data and remove noise. Followed by, the DLXM based feature extraction technique is employed for extracting a useful set of feature vectors. Eventually, the BC and MLPC models are utilized to recognize the respective class labels of the input image. The classification results of the presented model are evaluated against benchmark tongue image dataset and the results depicted the effectual classification performance on the applied images. The experimental values notified that the DLXM-MLPC model has outperformed the compared methods by achieving a higher precision, recall, accuracy, and F1-Score of 97.35%, 97.01%, 97.01%, and 96.77% respectively. Therefore, the DLXM-MLPC model can be considered as a proper tool to examine the tongue image.

References

[1] Zhao, Q.; Zhang, D.; Zhang, B. Digital tongue image analysis in medical applications using a new tongue ColorChecker. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 803–807.

[2] Ning, J.; Zhang, L.; Zhang, D.; Wu, C. Interactive image segmentation by maximal similarity based region merging. Pattern Recognit. 2010, 43, 445–456.

[3] X. Li, Y. Zhang, Q. Cui, X. Yi, Y. Zhang, Tooth-marked tongue recognition using multiple instance learning and CNN features. IEEE Trans. Cybern 49(2), 380–387 (2019)

[4] J. Ma, G. Wen, C. Wang, L. Jiang, Complexity perception classification method for tongue constitution recognition. Artif Intell. Med. Elsevier 96, 123–133 (2019)

[5] Y. Dai, G. Wang, Analyzing tongue images using a conceptual alignment deep autoencoder. IEEE Access. 6:5962–5972

[6] B. Zhang, B.V.K. Vijaya Kumar, D. Zhang, Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. IEEE Trans. Biomed. Eng. 61(2), 491–501 (2014)

[7] X. Wang, B. Zhang, Z. Yang, H. Wang, D. Zhang, Statistical analysis of tongue images for feature extraction and diagnostics. IEEE Trans. Image Process. 22(12), 5336–5347 (2013)

[8] T. Kawanabe, N.D. Kamarudin, C.Y. Ooi, F. Kobayashi, X. Mi, M. Sekine, A. Wakasugi, H. Odaguchi, T. Hanawa, Quantification of tongue color using machine learning in Kampo medicines. Eur. J. Integr. Med. 8, 932–941 (2016)

[9] M.H. Tania, K. Lwin, M.A. Hossain, Advances in automated tongue diagnosis techniques. Integr. Med. Res. Elsevier 8, 42–56 (2019)

[10] Morillas, S., Gregori, V. and Sapena, A., 2006, September. Fuzzy bilateral filtering for color images. In International Conference Image Analysis and Recognition (pp. 138-145). Springer, Berlin, Heidelberg.

[11] Jinsakul, N., Tsai, C.F., Tsai, C.E. and Wu, P., 2019. Enhancement of Deep Learning in Image Classification Performance Using Xception with the Swish Activation Function for Colorectal Polyp Preliminary Screening. Mathematics, 7(12), p.1170.

[12] https://machinelearningtokyo.com/2020/04/06/cnn-architectures-xception/

[13] Pu, X., Chen, S., Yu, X. and Zhang, L., 2018. Developing a novel hybrid biogeography-based optimization algorithm for multilayer perceptron training under big data challenge. Scientific Programming, 2018.

[14] Zhang, H., Song, Y., Jiang, B., Chen, B. and Shan, G., 2019. Two-stage bagging pruning for reducing the ensemble size and improving the classification performance. Mathematical Problems in Engineering, 2019.

[15] Zhang, B. and Zhang, H., 2015. Significant geometry features in tongue image analysis. Evidence-based Complementary and Alternative Medicine: eCAM, 2015.

[16] Tania, M.H., Lwin, K. and Hossain, M.A., 2019. Advances in automated tongue diagnosis techniques. Integrative Medicine Research, 8(1), pp.42-56.

[17] Zhang, B., Wang, X., You, J. and Zhang, D., 2013. Tongue color analysis for medical application. Evidence-Based Complementary and Alternative Medicine, 2013.

[18] Zhang, J., Xu, J., Hu, X., Chen, Q., Tu, L., Huang, J. and Cui, J., 2017. Diagnostic method of diabetes based on support vector machine and tongue images. BioMed Research International, 2017.

[19] Ma, J., Wen, G., Hu, Y., Chang, T., Zeng, H., Jiang, L. and Qin, J., 2018. Tongue image constitution recognition based on Complexity Perception method. arXiv preprint arXiv:1803.00219.

[20] Zhang, H. and Zhang, B., 2014, May. Disease detection using tongue geometry features with sparse representation classifier. In 2014 International Conference on Medical Biometrics (pp. 102-107). IEEE.

[21] Balu, S. and Jeyakumar, V., 2020. A Study on Feature Extraction and Classification for Tongue Disease Diagnosis. In Intelligence in Big Data Technologies—Beyond the Hype (pp. 341-351). Springer, Singapore.