2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM 2025)
27-28 June, 2025, Gazipur-1707, Bangladesh

Optimized Hybrid Cascaded Approach for Accurate Oral Cancer Detection in Histopathology Images Using Deep CNNs

1st Eshat Ahmad Shuvo
Computer Science & Engineering
Bangladesh University
Dhaka, Bangladesh
mdshuvoslk012@gmail.com

2nd Md. Shovon
Computer Science & Engineering
Jahangirnagar University
Dhaka, Bangladesh
mdshovon802@gmail.com

3rd Md. Nazmul Sarkar
Software Engineering
Daffodil International University
Dhaka, Bangladesh
mdnazmulsarkar701@gmail.com

4th Md. Sabbir Hossen
Computer Science & Engineering
Bangladesh University
Dhaka, Bangladesh
sabbirhossen5622@gmail.com

5th Pabon Shaha
Computer Science & Engineering
Mawlana Bhashani Science and Technology
University, Tangail, Bangladesh
pabonshahacse15@gmail.com

6th Wahidur Rahman
Computer Science & Engineering
Uttara University
Dhaka, Bangladesh
wahidtuhin0@gmail.com

***Abstract*—Oral Squamous Cell Carcinoma (OSCC) is a prevalent and deadly form of oral cancer, responsible for approximately 3% of all cancer cases worldwide and over 330,000 deaths annually. Early detection is critical for effective treatment and improved outcomes. Traditional diagnosis through manual examination of histopathological images is time-consuming and subjective. Recent advances in AI-guided computer vision have shown promise, though many approaches require extensive computational resources and large datasets. This paper introduces a cascaded network that integrates deep learning and traditional machine learning techniques for detecting Oral Squamous Cell Carcinoma (OSCC) from histopathology images. The proposed framework employs a comprehensive approach that includes three feature selection strategies—Principal Component Analysis (PCA), Bacterial Foraging Optimization (BFO), and a no-optimization baseline—to assess the impact of feature refinement on classification performance. An exhaustive search strategy was used to evaluate five widely adopted Convolutional Neural Network (CNN) models for feature extraction. The extracted features were subsequently either optimized using PCA or BFO or left unoptimized to identify the most informative subsets. These feature sets were then fed into seven traditional machine learning classifiers to perform OSCC detection. The MobileNetV2-PCA-LR cascaded network demonstrated the best overall performance among the configurations evaluated. This model achieved near-perfect results with an accuracy, precision, recall, and F1-score of 99.70%, while also offering faster image processing times. The proposed cascaded framework thus delivers a balanced solution that combines efficiency, high accuracy, scalability, and robustness, making it a strong candidate for practical and automated OSCC screening in clinical environments.**

***Index Terms*—OSCC, principle component analysis Cascaded Network, Hybrid Model, Deep Learning, and Machine Learning**

I. Introduction

Oral cancer, also known as mouth cancer, is a type of cancer that develops in the tissues of the mouth. It can occur on the lips, tongue, gums, inner cheeks, roof of the mouth, and floor of the mouth. Of the total number of patients with all types of cancer, about 3% particularly belong to oral cancer worldwide [1]. It is ranked as the 6th most dominant cancer worldwide, with an annual global incidence of 657,000 people and approximately 330,000 deaths [2]. The primary reason anyone suffers from oral cancer is their bad habits, including smoking, chewing betel nuts or tobacco, poor oral hygiene, and infection with human papillomavirus (HPV), which affect the oral cavity, nasopharynx, and pharynx regions [3]. It is very common, especially in South Asian developing countries, due to the lack of awareness [4]. Among the different types of oral cancer, Oral Squamous Cell Carcinoma (OSCC) is the most common oral malignancy that originates in the oral cavity, and occurs through multiple mutations happening in the cell [3] and disrupts the epithelium. Though OSCC is a fatal variant of oral cancer it doesn't have any specific vital signs that will help the expert to identify the OSCC accurately. However, it can be predicted by many indicators, such as the location of the lesion inside the mouth, its color, size, appearance, and the history of tobacco and alcohol use [5]. The OSCC is life-threatening; therefore an early diagnosis of Oral malignant lesions is required before it transforms into OSCC. The OSCC has a high chance of recurrence [6]. As As a result, to have a proper prognosis, an in-depth analysis of its occurrence and progression must be performed. According to the oral cancer survival statistics for the past five years, the survival rate is approximated to be between 35% and 50% but includes better prognostic indicators of different populations [7]. Consequently, apart from the pathologist's perspective, a proper and precise classification of histopathological images of Oral lesions is essential.

In traditional approaches, pathologist reports generally rely on various types of factors, including microscope model, staining quality, time taken by a pathologist for analysis of each slide, and experience in the related field [8]. As

979-8-3315-5542-9/25/31.00

a result, there is a big chance of errors in the diagnostic report due to the large number of procedures that have to be followed to make a report. On the other hand, a digital hybrid system combining deep and machine learning can provide high classification accuracy without any kind of help [9]. For building a system that can automatically classify biopsy images, there can be two possible ways: one is deep learning-based, and the other is machine learning-based. Both have their pros and cons, if the system is built on the basics of ML then it can operate on minimal data, but the accuracy might not be satisfactory. On the other hand, a deep learning model may provide a good amount of accuracy, but that requires a big dataset. To overcome these two issues, transfer learning may be applicable where the feature will be extracted directly from the images (OSCC) by applying deep CNN models, and on that feature, traditional machine learning models can be applied to classify which will provide the best output on a limited dataset. Transfer learning can reduce the training time that a deep learning model normally takes to build from scratch on a complex problem [10]. In this study, we applied different pre-trained CNN models and machine learning models to obtain efficient accuracy.

Some of the common machine learning models used for the detection of OSCC include the Support Vector Classifier (SVC), Random Forest (RF), Knearest neighbor (KNN), Naıve Bayes (NB), Extreme Gradient Boosting (XGNB), Logistic Regression (LR) and Decision Tree (DT). These models are primarily used in the shape, texture, color, and structure of the features. LR is the most widely used model for machine learning, which was confirmed in the classification of different stages of OSCC, the classification of various kinds of lesions related to OSCC, and the detection of certain patterns of OSCC. Jiliang Ren et al utilized Machine Learning to detect the various stages of OSCC with RF model [11], Andres M. Bur et al used decision forest to classify the different stages of OSCC in his research [12], Y. Chan Lee et al. detected OSCC using XGBoost in his study [13]. A. Gardner et al used DT to classify the different stages of OSCC in his research [14]. Traditional machine learning (TML) models are best for training on limited datasets and also for maintaining robustness; however, achieving high accuracy with TML is difficult. All these studies above indicate that accuracy usually ranges from 75 to 90% in case of using the machine learning based approach for OSCC detection.

On one hand, the deep learning-based method may be able to provide high accuracy, but it requires a large amount of data, which is quite difficult to obtain in the medical imaging field [15]. Many journals have suggested deep learning, especially CNN-based approaches, for the analysis of histopathological images to detect, segment, and label OSCC. These methods widely use existing CNN models via the transfer learning approach, and some suggest customized CNN models. AlexNet, VGG, Inception, ResNet, GoogleNet, DenseNet, and Xception-based architectures were mostly the models used in these approaches [16]. Sayyada and his team also recommended the deep CNN model DaseNet201 to classify the OSCC in his study [17]. The authors, Suliman et al., proposed the use of a deep CNN-based ResNet18 model in identifying patterns associated with OSCC [18]. Santisudha et al. proposed a DL-based capsule network for OSCC classification of various stages and reported promising results [19]. Madhusmita et al. proposed a generic 10-layer convolutional neural network for binary classification of OSCC [17]. According to the study done by M. Ahmad et al., they were able to detect OSCC-related abnormality patterns using DenseNet201, a Deep CNN-based model [20]. In fact, all these studies above indicate that accuracy usually ranges from 90 to 99% in the case of using the deep learning-based approach for OSCC detection. Those approaches mentioned above were trained using large image datasets, resulting in high accuracy. In most of the studies, the histopathological image was classified into two different classes, such as normal class and OSCC. This study aimed at designing the proposed approach by utilizing machine learning and deep learning techniques together to build a hybrid automated system; not only without optimization (W/O) alone, but also we implemented them in a cascaded network that is fed by bacterial foraging optimization (BFO) and principal component analysis (PCA) optimization technique. The machine learning-based approaches were lightweight and suitable for training using a small dataset. In parallel, deep learning methods have high accuracy when trained with large datasets. The proposed cascaded network was developed with a small dataset but yielded high-level accuracy comparable with other deep learning approaches. This network of different pre-trained CNNs were applied for the automatic extraction of features from the histopathological images. Next, the features were processed by the BFO and PCA to find out the most optimized features. Next, the regular machine learning models are used for binary-class classification with the BFO and PCA features. In this study, we also investigated the role of BFO and PCA in the cascaded network. The major contributions of this paper are:

1) Developing a lightweight and optimal cascaded network for OSCC screening by integrating deep learning-based feature extraction, BFO and PCA for feature optimization, and traditional machine learning for disease detection.
2) Seven machine learning classifiers, including SVM, DT, RF, KNN, LR, XGB, and GNB, are used for classification.
3) The performance comparison of different CNN models without optimization (W/O), with BFO, and with PCA.

The whole process of the study is organized into four related sections. Section II comprehensively explains the proposed model, techniques, and underlying working concepts. In Section III, the experimental data that was used in this model is presented together with the recommended orientations. In the final section, IV, the paper gave the conclusion by highlighting the existing limitations and potential areas for future work.

II. MATERIALS AND METHOD

This section just describes how models for the proposed cascaded network were chosen, how training and testing of the models were done, and the distribution of the dataset.

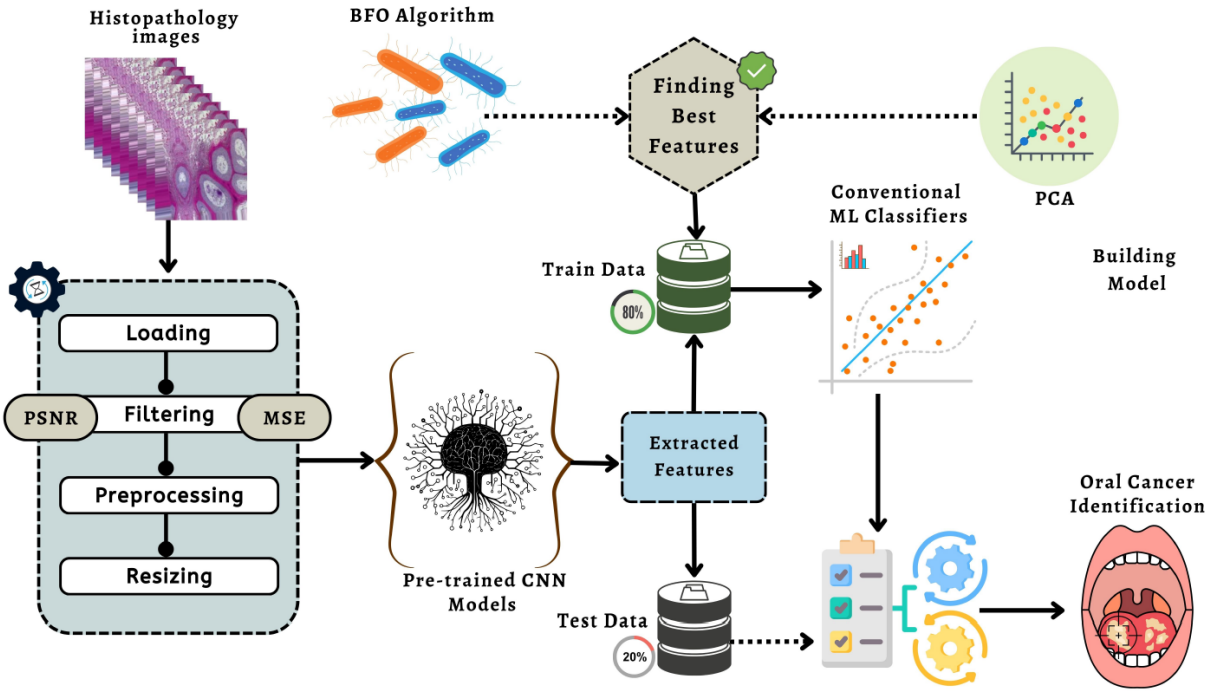
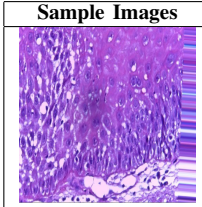
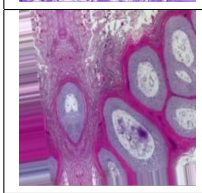


Fig. 1. Workflow Diagram Of the Proposed Model

A. Dataset

The histopathological images were obtained from the dataset available in the public domain [21] and from a local cancer research institute. Then, these images were reviewed by an expert pathologist to identify whether OSCC was present or not in the images. The pathologist assessed the image, rating it from 0 to 1, with 1 representing a high level of OSCC. Then, for our experiment, we narrowed down our images for classification to the ones that had been given a score of 1 or 0. To make sure that all these images were fit for the training of machine learning models, the quality of these images was judged by the image quality assessment method proposed by Yagi et al. on the basis of sharpness and noise [22]. There are about 5001 images for both classes of images, normal and OSCC. In Table I, the authors provided sample images of the given dataset.

TABLE I
SAMPLE ORAL CANCER IMAGES FOR THIS RESEARCH

Sample Images	Class Name	Class No.	Dataset Size
	Oral Normal	1	5001
	Oral_SCC	0	5001

B. Architecture of the proposed system

The diagram of the proposed system is depicted in Figure 1, while the procedure of the proposed method is described in the pseudo-code format as outlined in Algorithm 1. The initial step of the proposed method starts with the preprocessing of the input image, which is essential for the proper functioning of the system. The preprocessing used in this step determines the quality of the image. To assess the quality of an image we set up an image quality assessment cell (IQAC). The IQAC evaluated the quality of the image in its resolution, sharpness, and noise. Applying the formula introduced by Yagi et al [22], the sharpness and noise indices of the image were determined. The quality was obtained using a linear regression model in 1. The coefficients of quality predictors were obtained prior to the experiment in a regression analysis in which mean square errors (MSE) of the images were employed. Consequently, the high Qi value was an indicator of poor quality. In the proposed method, there was a quality value associated with each image denoted by Qi, this was compared to a threshold value Qth; if the quality value Qi was less than the defined threshold value, this was used for further analysis. For the current study, the chosen threshold value was 10. After that, the approved image by the IQAC was converted to the sRGB space to make its color values more normalized. As the last step, the image was resized to the standard size accepted by all AI models. These preprocessing steps, particularly quality verification, are important to make the proposed system usable. After preprocessing, the image is processed using the pre-trained CNN model to extract features automatically. The proposed system relies on the MobileNetV2 network for feature extraction. The MobileNetV2 model was selected among five other CNN models based on its performance and suitability for the proposed system, which is explained in

the following section. Then, the PCA method processed the extracted features to find the suitable one for the classifiers. Finally, the machine learning model were applied on the refined features to predict the class of the image as OSCC or Normal. The proposed method selected the logistic regression model as the classifier.

C. Model Selection for Proposed Network

In this study, we investigated five popular CNN-based models for feature extraction, including MobileNetV2, ResNet50, InceptionV3, DenseNet201 and Xception. For the classification, we examined seven machine learning models: SVC, DT, RF, NB, XGB, KNN, and LR. The proposed cascaded network used the CNN model as the feature extractor and the traditional ML model as the classifier. Therefore, we developed 35 different networks combining five feature extractor models with seven classifier models for the three individual techniques (W/O, BFO, PCA). Each network was trained and validated using 60% and 20% of the dataset, respectively. Finally, these networks were tested using unseen images from the dataset 20% to compare and select the suitable network for the proposed system.

TABLE II
COMPARISON OF DIFFERENT MODELS WITH THE PROPOSED METHOD.

Model	No. of Features	With BFO	With PCA
InceptionV3	2048	1042	800
ResNet50	2048	499	600
XceptionNet	2048	998	800
DenseNet201	1024	518	700
MobileNetV2	1024	514	400

Algorithm 1: Proposed algorithm for OSCC screening pipeline

Initialization: $P \leftarrow \{P_1, P_2, P_3, \ldots, P_n\}$
$B \leftarrow \{B_1, B_2, B_3, \ldots, B_n\}$
$C \leftarrow \{C_1, C_2, C_3, \ldots, C_n\}$
Output: ψ
while $I_{RGB} \neq$ NIL **do**
  $IsRGB \leftarrow$ IQAC(I_{RGB})
  $I_R \leftarrow$ Resize I_{RGB} to 224×224
  $F_{M,i} \leftarrow$ Load MobileNetV2 with parameters P_i
  $f_i \leftarrow F_{M,i}(I_R)$
  $S_{f_i} \leftarrow$ BFO(f_i)
  PCA$_i \leftarrow$ Load PCA with parameters B_i
  $R_{f_i} \leftarrow$ PCA$_i(S_{f_i})$
  $CM_i \leftarrow$ Load classifier LR with parameters C_i
  $y \leftarrow$ CM$_i(R_{f_i})$
  if $y == 1$ **then**
    $\psi \leftarrow$ OSCC
    return ψ
  else
    $\psi \leftarrow$ Normal
  end if
end while
function IQAC(I_{RGB})
  Initialization: $\alpha, \beta, \gamma, Q_{th}$
  Load model with the $n+1$ parameters α, β, γ
  $Q_i = \alpha + \beta \times$ Sharpness $+ \gamma \times$ Noise
  Sharpness and Noise are estimated
  if $Q_i < Q_{th}$ **then**
    return NIL
  end if
  if $C_{\text{linear}} < 0.0031$ **then**
    $IsRGB = 12.99 \times C_{\text{linear}}$
  else
    $IsRGB = 1.0552 \times C_{\text{linear}}^{1/2.4}$
  end if
return $IsRGB$
end function
function BFO(f_i)
  Initialization: $a \leftarrow 2A - 1$, Where $A = 1, 2, 3, 4, \ldots$
  $P \leftarrow$ Input Image
  $R_a \leftarrow$ Resize Feature Vector
  BestFilter $\leftarrow$ applyFilter(P)
  $Q_a \leftarrow$ applyFilterToImage(P, BestFilter)
  $A \leftarrow$ imageToArray(P, Q_a)
  for each A **do**
    Use (P, Q_a) to get $R_a|_{R_a\{V_0, V_1, \ldots, V_{14}\}}$
  end for
return R_a
end function
function PCA(S_{f_i})
  Initialization: Standardize the data S_{f_i}
  $S' \leftarrow$ Compute the covariance matrix of S_{f_i}
  $E \leftarrow$ Compute the eigenvalues and eigenvectors of the covaria
  $E_{\text{sorted}} \leftarrow$ Sort the eigenvalues in descending order
  $V_k \leftarrow$ Select the top k eigenvectors
  $R_{f_i} \leftarrow$ Transform the data S_{f_i} using the selected eigenvectors
return R_{f_i}
end function

III. EXPERIMENTAL RESULT ANALYSIS

In this study, we experimented with 105 cascaded networks designed using three different approaches. The first 35 cascaded networks were developed without any optimization; later on, the other two sets of networks were created by using two different optimization techniques including PCA and BFO to refine the features which were obtained by the CNN models before applying the ML classifier. In our study, we trained, validated, and tested all three types of cascaded networks using the same dataset. Then, they were compared in terms of their accuracy, precision, recall, and F1 score. These networks were trained, validated, and tested using a personal computer without a GPU. Table III shows the comparison among the networks for the test dataset. Here, the results have revealed that most cascaded networks are improved with the presence of PCA. Nevertheless, the ResNet 50 and Xeption-based networks performed well when the BFO was used instead of PCA. As seen in Table II, 2048 features were employed for both the ResNet50 and Xeption-based models. For ResNet50, the number of features was reduced to 499 in the presence of BFO and 600 for PCA. On the other hand, the number of features for model Xeption

TABLE III
PERFORMANCE METRICS FOR DIFFERENT PRE-TRAINED MODELS AND CLASSIFIERS

Pre-trained model	Classifier	Accuracy (%)			Precision (%)			Recall (%)			F1-Score (%)		
		W/O	PCA	BFO	W/O	PCA	BFO	W/O	PCA	BFO	W/O	PCA	BFO
DenseNet201	SVC	91.16	97.40	78.41	45.58	97.39	78.74	50.00	97.39	77.78	47.69	97.39	77.98
	XGB	88.65	94.64	94.05	46.55	94.74	94.20	48.91	94.54	93.85	47.48	94.61	93.99
	RF	88.70	89.50	89.56	45.47	90.10	90.01	48.65	89.17	89.13	47.48	89.38	89.39
	DT	86.74	78.79	77.06	54.86	78.76	76.91	53.86	78.64	76.88	54.23	78.68	76.89
	KNN	90.44	97.02	81.93	56.54	97.13	82.89	51.03	96.93	81.14	50.21	97.01	81.44
	LR	90.94	91.77	74.51	53.93	91.75	74.50	50.16	91.75	74.01	48.24	91.75	74.13
	GNB	70.73	72.40	72.89	57.76	72.32	72.71	64.16	72.32	72.61	62.60	72.32	72.65
InceptionV3	SVC	70.13	78.05	80.85	69.65	78.17	80.92	70.41	78.07	80.83	69.72	78.03	80.83
	XGB	74.15	78.90	76.60	73.93	78.95	76.68	74.37	78.91	76.58	74.04	78.90	76.57
	RF	68.22	77.75	71.70	67.92	77.85	71.70	68.48	77.77	71.70	68.14	77.74	71.70
	DT	58.69	68.20	61.30	58.82	68.20	61.30	58.86	68.20	61.29	58.84	68.20	61.29
	KNN	68.64	83.55	77.85	68.83	83.60	77.90	68.87	83.56	77.83	68.52	83.55	77.83
	LR	75.85	74.45	76.80	76.07	74.46	76.82	75.90	74.46	76.79	75.95	74.45	76.80
	GNB	64.62	64.70	59.10	62.98	64.95	59.60	65.16	64.74	59.01	62.60	64.58	58.43
ResNet50	SVC	78.11	74.18	59.34	79.09	60.93	74.61	73.65	74.22	58.80	75.54	74.09	57.01
	XGB	84.93	71.56	74.13	84.56	74.06	71.58	82.98	71.57	74.05	83.69	71.56	74.06
	RF	73.21	69.25	75.36	80.12	70.02	75.68	74.49	69.32	75.21	76.38	69.00	75.20
	DT	66.78	57.60	63.60	64.37	57.59	63.58	64.81	57.59	63.57	64.48	57.59	63.57
	KNN	77.78	73.56	73.72	76.87	73.74	73.95	75.61	73.59	73.57	75.90	73.53	73.56
	LR	83.72	71.61	68.38	82.67	71.70	68.49	82.88	71.63	68.24	82.77	71.60	68.21
	GNB	54.90	52.05	52.36	58.93	52.84	56.41	62.01	52.22	53.23	55.33	49.38	46.14
Xeption	SVC	88.98	72.69	89.99	88.33	72.71	90.01	87.01	72.50	90.00	87.60	72.54	89.99
	XGB	90.85	70.94	86.64	90.05	70.90	86.66	90.91	70.92	86.64	91.00	70.90	86.64
	RF	88.75	67.20	80.44	87.85	67.17	80.63	87.06	66.98	80.44	87.43	66.99	80.41
	DT	57.50	56.01	68.63	57.30	56.01	68.64	57.20	56.02	68.63	57.25	55.99	68.63
	KNN	75.80	74.54	88.64	75.70	74.63	88.82	75.60	74.32	88.65	75.65	74.37	88.63
	LR	72.40	70.53	85.04	72.30	70.47	85.04	72.20	70.44	85.04	72.25	70.46	85.04
	GNB	56.90	54.93	68.58	56.70	54.67	69.27	56.60	54.48	68.59	56.65	54.20	68.31
MobileNetV2	SVC	96.35	98.00	98.01	97.65	98.02	98.06	96.24	98.01	98.01	96.43	98.01	98.01
	XGB	94.77	95.89	95.85	94.05	95.88	95.87	94.54	95.88	95.88	94.78	95.88	95.89
	RF	92.30	94.76	94.78	92.99	94.73	94.76	92.62	94.72	94.72	92.19	94.73	94.73
	DT	96.69	98.15	98.05	96.23	98.09	98.05	96.21	98.09	98.05	96.22	98.08	98.05
	LR	**96.03**	**99.70**	**98.90**	**96.19**	**99.69**	**98.89**	**96.12**	**99.70**	**98.89**	**96.15**	**99.70**	**98.90**
	KNN	97.09	98.90	98.88	97.09	98.91	98.87	98.91	98.89	98.88	97.00	98.89	98.88
	GNB	97.96	97.95	97.90	97.97	97.98	97.90	97.95	97.96	97.90	97.95	97.97	97.90

was about 998 and 800 for BFO and PCA, respectively. This indicated that ResNet 50 and Xeption-based cascaded networks are more suitable when using the BFO method. For DaseNet and Inception, the accuracy was marginally higher for PCA-enabled cascaded networks. For the MoblineNetV2 model, PCA was beneficial. Incorporating PCA into the MobileNetV2-based cascaded network reduced the number of features to 400 from 1024; however, it increased the classifier accuracy. The MobileNetV2-PCA-LR network achieved comparatively higher accuracy using a significantly lower number of features compared to other networks. The accuracy, precision, and recall were 99.70% for the MobileNetV2-PCA-LR cascaded network using only 400 optimally selected features. This architecture provides a low-weight network (400 features) with significant accuracy (99.70%) for the proposed work compared to the BFO and W/O. Thus, the cascaded MobileNetV2-PCA-LR network was selected for the proposed method. Thus, the MobileNetV2-PCA-LR cascaded network was selected for the proposed method. In addition, Figures 2 and 3 illustrate the ROC curve and Confusion matrix for the proposed model, respectively.

Furthermore, Table IV compares our proposed model with other existing work on OSCC over the past few years. Most of the present studies on OSCC are based on Deep learning, and their accuracy lies between the ranges of 91% and 99%. On the other hand, the proposed method, integrating MobileNetV2 with PCA and Logistic Regression, stands out with perfect scores in accuracy, precision, recall, and F1-score (all 99.70%) and a processing time of 189 ms ± 28.3 ms per image, demonstrating superior performance and efficiency on the selective work.

Fig. 2. ROC curves of MobileNetV2-PCA approach for different classifiers

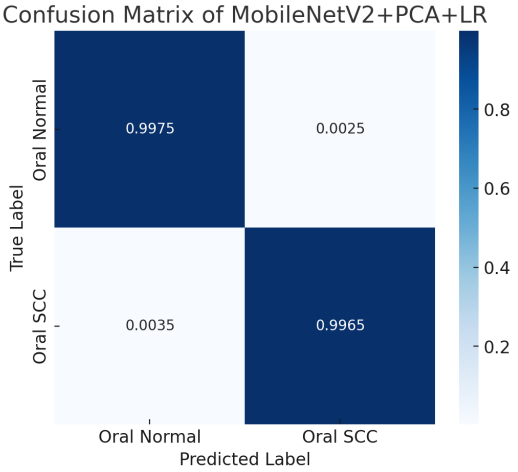


Fig. 3. Confusion Matrix With of MobileNetV2-PCA-LR network

TABLE IV
COMPARISON OF OTHER EXISTING MODEL AND OUR PROPOSED MODEL

Techniques	A (%)	P (%)	R (%)	F1 (%)	Time Complexity	Space Complexity
DenseNet201 [18]	91.25	93.00	93.00	93.00	-	-
DenseNet201 + DL+ML [20]	97.00	96.77	-	93.74	-	-
Capsule network [19]	97.35	96.90	-	97.33	-	-
10-Layer CNN [17]	97.82	97.00	-	97.00	-	-
ResNet18 [18]	99.30	99.31	-	-	-	-
MobileNetV2 + PCA + LR	**99.70**	**99.70**	**99.70**	**99.70**	**189 ms $\pm$ 28.3 ms**	**PM: 2369.77 MiB, INC: 929.07 MiB**

IV. CONCLUSION

In this study, we developed a hybrid automated system for detecting Oral Squamous Cell Carcinoma (OSCC) by integrating deep learning and machine learning techniques, enhanced with Bacterial Foraging Optimization (BFO) and Principal Component Analysis (PCA) for feature selection and reduction. Our proposed approach outperformed existing models on the same topic, achieving 99.70% accuracy, precision, recall, and F1 scores, with an average processing time of 189 ms ± 28.3 ms per image. The key contributions of this study include the development of a lightweight cascaded network combining deep learning-based feature extraction with traditional machine learning classification, a comprehensive performance comparison of CNN models, and an ablation analysis highlighting PCA's role in achieving high accuracy. While the results are promising, model performance may vary across different datasets, and further validation is needed for real-time clinical use. Future work will focus on improving generalizability across diverse datasets, integrating the model into clinical workflows, and optimizing inference speed for deployment on resource-constrained devices.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] K. Dhanuthai, S. Rojanawatsirivej, W. Thosaporn, S. Kintarak, A. Subarnbhesaj, M. Darling, E. Kryshtalskyj, C.-P. Chiang, H.-I. Shin, S.-Y. Choi, *et al.*, "Oral cancer: A multicenter study," *Medicina oral, patologia oral y cirugia bucal*, vol. 23, no. 1, p. e23, 2017.

[3] S. Warnakulasuriya and J. S. Greenspan, "Epidemiology of oral and oropharyngeal cancers," *Textbook of Oral Cancer: Prevention, Diagnosis and Management*, pp. 5–21, 2020.

[4] V. N. Nayak, M. Donoghue, and M. Selvamani, "Oral squamous cell carcinoma: a 5 years institutional study," *Journal of Medicine, Radiology, Pathology and Surgery*, vol. 1, no. 5, pp. 3–6, 2015.

[5] D. Chakraborty, C. Natarajan, and A. Mukherjee, "Advances in oral cancer detection," *Advances in clinical chemistry*, vol. 91, pp. 181–200, 2019.

[6] A. Weckx, M. Riekert, A. Grandoch, V. Schick, J. E. Zoeller, and M. Kreppel, "Time to recurrence and patient survival in recurrent oral squamous cell carcinoma," *Oral oncology*, vol. 94, pp. 8–13, 2019.

[7] S. L. Goldenberg, G. Nir, and S. E. Salcudean, "A new era: artificial intelligence and machine learning in prostate cancer," *Nature Reviews Urology*, vol. 16, no. 7, pp. 391–403, 2019.

[8] L. Li, "Principal component analysis for dimensionality reduction," *Towards Data Science*, 2019.

[9] S. Dobilas, "Lda: Linear discriminant analysis—how to improve your models with supervised dimensionality reduction," 2021.

[10] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "Ga-svm based feature selection and parameter optimization in hospitalization expense modeling," *Applied soft computing*, vol. 75, pp. 323–332, 2019.

[11] J. Ren, M. Qi, Y. Yuan, S. Duan, and X. Tao, "Machine learning–based mri texture analysis to predict the histologic grade of oral squamous cell carcinoma," *American Journal of Roentgenology*, vol. 215, no. 5, pp. 1184–1190, 2020.

[12] A. M. Bur, A. Holcomb, S. Goodwin, J. Woodroof, O. Karadaghy, Y. Shnayder, K. Kakarala, J. Brant, and M. Shew, "Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma," *Oral oncology*, vol. 92, pp. 20–25, 2019.

[13] M. S. Kwak, Y.-G. Eun, J.-W. Lee, and Y. C. Lee, "Development of a machine learning model for the prediction of nodal metastasis in early t classification oral squamous cell carcinoma: Seer-based population study," *Head & Neck*, vol. 43, no. 8, pp. 2316–2324, 2021.

[14] H. Alkhadar, M. Macluskey, S. White, I. Ellis, and A. Gardner, "Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma," *Journal of Oral Pathology & Medicine*, vol. 50, no. 4, pp. 378–384, 2021.

[15] H. Shakhawat, S. Hossain, A. Kabir, S. H. Mahmud, M. M. Islam, and F. Tariq, "Review of artifact detection methods for automated analysis and diagnosis in digital pathology," in *Artificial Intelligence For Disease Diagnosis And Prognosis In Smart Healthcare*, pp. 177–202, CRC Press, 2023.

[16] S. Dixit, A. Kumar, and K. Srinivasan, "A current review of machine learning and deep learning models in oral cancer diagnosis: Recent technologies, open challenges, and future research directions," *Diagnostics*, vol. 13, no. 7, p. 1353, 2023.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[18] S. M. Fati, E. M. Senan, and Y. Javed, "Early diagnosis of oral squamous cell carcinoma based on histopathological images using deep and hybrid learning approaches," *Diagnostics*, vol. 12, no. 8, p. 1899, 2022.

[19] S. Panigrahi, J. Das, and T. Swarnkar, "Capsule network based analysis of histopathological images of oral squamous cell carcinoma," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4546–4553, 2022.

[20] M. Ahmad, M. A. Irfan, U. Sadique, I. u. Haq, A. Jan, M. I. Khattak, Y. Y. Ghadi, and H. Aljuaid, "Multi-method analysis of histopathological image for early diagnosis of oral squamous cell carcinoma using deep learning and hybrid techniques," *Cancers*, vol. 15, no. 21, p. 5247, 2023.

[21] O. S. Naren, "Multi cancer dataset," 2022.

[22] H. M. Shakhawat, T. Nakamura, F. Kimura, Y. Yagi, and M. Yamaguchi, "Automatic quality evaluation of whole slide images for the practical use of whole slide imaging scanner," *ITE Transactions On Media Technology And Applications*, vol. 8, no. 4, pp. 252–268, 2020.