Original paper

Histopathological image analysis and enhanced diagnostic accuracy explainability for oral cancer detection

V.P. Gladis Pushparathi[1], Sylaja Vallee Narayan S.R.[2], Pratheeba R.S.[3], V. Naveen[4]

[1]Department of Computer Science and Engineering, R.M.K. College of Engineering and Technology, R.S.M. Nagar, Puduvoyal, Thiruvalur Dist, Tamil Nadu, India
[2]Department of Computer Science and Engineering, GITAM University, Bangalore, Karnataka, India
[3]Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti, Chennai, Tamil Nadu, India
[4]Department of Computer Science and Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India

Deep learning (DL) has transformed medical imaging, particularly in the realm of oral cancer (OC) diagnosis using histopathological images. Timely detection of OC is essential for enhancing precision medicine and saving lives. However, incorrect diagnosis may impede effective treatment. In this study, we have proposed a DL model for OC classification, enhanced diagnosis decision-making and interpretability. We achieve this by starting with colour normalisation of histopathology images using the Vahadane 3-stain parameter normalisation and watershed segmentation method, followed by tiling and augmentation. Key features are selected using the weighted Fisher score (WFS) to address class imbalance. The U-Net classifier has been improved by using feature-based inputs instead of full images, reducing computational complexity and training time. The integration of Vahadane normalisation for consistent preprocessing across samples, WFS, and explainable artificial intelligence (XAI) addresses critical challenges in histopathological image analysis. The proposed model surpasses existing approaches with a classification accuracy of 99.54%, and it outperforms DenseNet201 and VGG10 in precision and reliability. The efficiency in handling imbalanced datasets and explainability features make it suitable for early precise OC detection, which can reduce diagnostic errors and enhance treatment outcomes.

Key words: oral cancer detection, histopathological image analysis, Vahadane 3-stain parameter normalisation, weighted Fisher score, U-Nete.

Introduction

Oral cancer (OC) is characterised by the unchecked growth of cells that invade and damage surrounding tissues within the oral cavity. The three main types of OC are oral squamous cell carcinoma (OSCC), verrucous carcinoma, and minor salivary gland carcinoma [1]. Consequently, many cases remain undiagnosed until they metastasise to other parts of the body, posing challenges and reducing treatment effectiveness [2]. Treatment options for OC typically include surgery, radiotherapy, and chemotherapy, either as standalone therapies or in combination, depending on the disease stage [3]. In recent times, the field of

medicine has undergone a significant transformation due to the emergence of DL applications. These DL models excel in identifying intricate data patterns and making predictions autonomously, without the need for explicit programming [4]. DL has advanced several medical domains, including radiology, dermatology, ophthalmology, neurology, genomics, and cardiology. It has also emerged as a promising technique for analysing histopathological images [5]. Convolutional neural networks (CNN) are DL algorithms that consist of several types of layers, including convolutional layers, pooling layers, and fully connected layers [6, 7]. Applying DL to oral histopathology images can aid clinicians in diagnosing, detecting, and predicting the prognosis of OC in clinical settings. This enables early diagnosis, optimal treatment selection, and improved survival rates [8]. Moreover, robust preprocessing methods and advanced DL algorithms are essential due to the detailed structural information histopathological images offer about underlying tissues [9]. To overcome this, the proposed model developed an automatic prediction model for OC using advanced DL-based models.

- The foremost contributions of the proposed OC prediction model are listed below:
- The proposed Vahadane Three-Stain Parameter Normalization technique allows setting the number of stains to 3, aiming to encompass a broader range of variability.
- XAI technique, namely gradient-weighted class activation mapping (GRAD-CAM), used to offer insights into the model's predictions.
- The proposed methodology addresses the issue of class imbalance by incorporating the WFS, which integrates class weights into its computation. This adjustment ensures that the feature selection process considers the imbalanced class distribution, giving proper importance to minority classes. This potentially enhances the classifier's performance across all classes, particularly those that are underrepresented. Addressing class imbalance entails several implications, including the following:
  - Upgraded categorisation performance:
    – enhanced feature representation balance: utilising WFS increases the likelihood of selecting features critical for distinguishing all classes, including minority classes;
  - equitable model predictions: by mitigating class imbalance during feature selection, the model's predictions exhibit reduced bias towards the majority class, promoting fairer treatment across all classes.

The remaining portion of the paper is arranged as follows: Section *Literature survey* analyses earlier studies related to OC diagnosis, Section *Proposed methodology* outlines the proposed model for detecting OC, Section *Results and discussion* presents the findings, and the last Section summarises the conclusions.

Literature survey

This section analyses the existing works of OC cancer detection and classification. Marzouk *et al*. [10] introduced AI and DL models for diagnosing and classifying OC using a fuzzy-based contrast enhancement technique, DenseNet-169, and a Chip optimisation strategy with an autoencoder. The model improves the performance with the highest precision. However, an advanced DL model could be used to classify medical images for diagnosing OC. To overcome this, Das *et al.* [11] utilised histopathology images of OC to automatically and early OSCC detection using a 10-layer CNN model. The model was analysed with state-of-the-art DL models like VGG16, VGG19, Alexnet, and Inception Net and achieved high accuracy. It could be expanded further to identify distinct stages of OC, which could benefit both the patient and the doctor. For this, to automatically categorise 6 clinical representation types of oral lesion images, Gomes *et al.* [12] used CNN models such as ResNet50, VGG16, InceptionV3, and Xception. InceptionV3's oral elementary classification produced the best results, but it struggled to analyse the images because multiple diseases were present in the same regions.

Myriam *et al*. [13] intended to establish reliable OC diagnosis in medical images based on a CNN and optimised deep belief network, which are tuned by a combination of particle swarm optimisation and Al-Briruni Earth Radius optimisation algorithms. The model could be enhanced by applying the model over a larger dataset and thoroughly examining its advantages and disadvantages. Song *et al*. [14] explored the efficacy of the Vision Transformer (ViT) and Swin transformer architectures for mobile-based image classification of OC. The pre-trained Swin transformer model outperformed better the ViT model in binary classification. Despite this, the transformer-based model provided remarkable performance. It was difficult to comprehend the prediction because of its interpretability. Confer *et al*. [15] created a DL-based multimodal discrete frequency infrared and darkfield microscopy to assist with the label-free histopathological examination of oral potentially malignant disorders. The tissues were segmented into 3 classes using a fully convolutional network with a ResNet50 backbone. Although the model demonstrated potential for high histopathological segmentation accuracy, it was not generally applicable to all acquired tissues.

Thus, by analysing the existing models, we identified several issues, such as inadequate diagnosis, challenges in histopathology image analysis, lack of interpretability, unclear object boundaries, low reliability in prediction, scarcity of diverse datasets, and class imbalance. To address these limitations, our proposed model

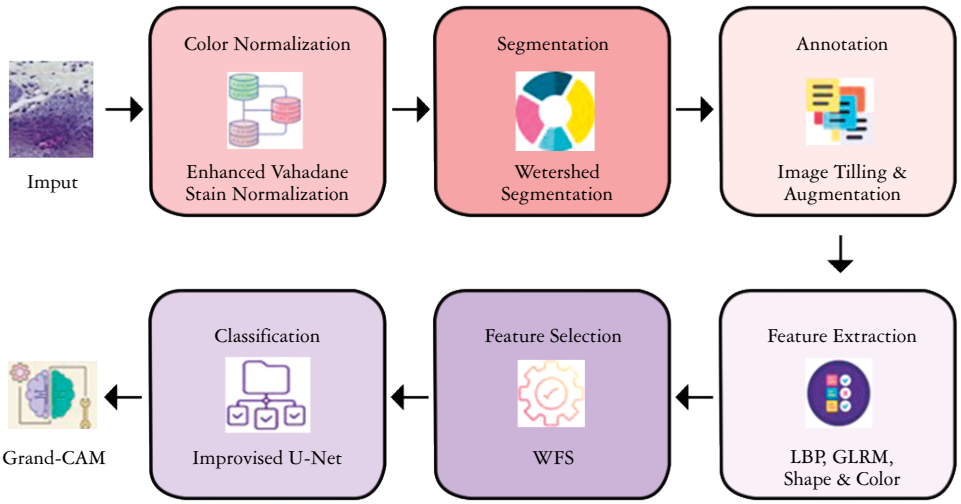

Fig. 1. Diagrammatic representation of the proposed model

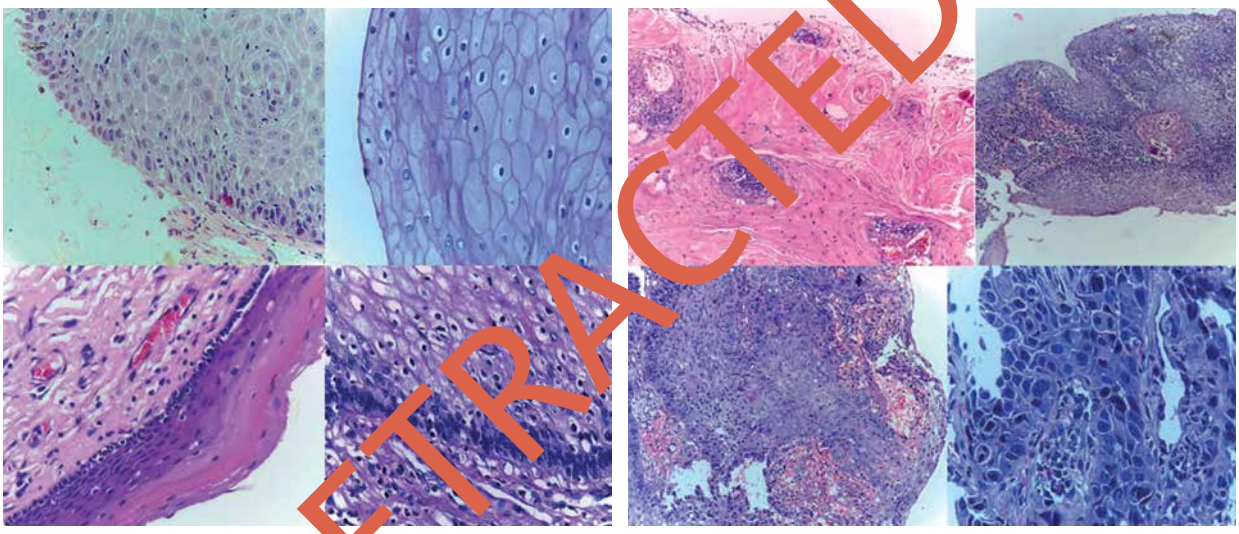

Fig. 2. Histopathological image samples from the dataset for normal images (LHS) and cancerous images – oral squamous cell carcinoma (RHS)

developed a DL-based model for OC classification and enhanced diagnosis decision-making and interpretability.

Proposed methodology

The proposed methodology integrates advanced image processing techniques with DL methods to enhance the accuracy and reliability of OC detection using histopathological images. Initially, the images are pre-processed to standardize color variation across samples. Watershed Segmentation algorithm is then employed to effectively delineate regions with unclear boundaries. The segmented images are further processed through image tiling and image augmentation techniques. After that, the shape and textual features are extracted from the images using local binary patterns (LBP) and gray level run length matrix (GLRLM). To address the class imbalance, the WFS method incorporates class weights to prioritize features that differentiate minority classes, thereby improving the model's performance. For classification, an improvised U-Net model is developed that utilizes features instead of entire images as input, reducing the duration of training. Finally, the XAI technique Grad-CAM is employed for image interpretability to identify critical image regions. The proposed approach, depicted in Fig. 1, helps differentiate non-cancerous and cancerous tissue.

Image source

This research utilizes histopathological images, which are microscopic images of tissues used for analysing diseases. Sections are stained with one or

more dyes to perceive various tissue components under a microscope. These images were sourced from datasets of histopathological images of OC on Kaggle. The images in the OC datasets are categorized into two groups: normal and cancerous, as shown in Fig. 2. Furthermore, we have also used the Multi Cancer Dataset for training and testing and the NDB-UFES dataset for validation purposes.

Color normalization

Color normalization is crucial for optimizing the performance of the CNN classifier. To achieve this, an intensity scaling factor is included to regulate intensity, and the proposed method adopts an improved Vahadane Three-Stain Parameter Normalization approach, utilizing three stain parameters to capture increased variability (Fig. 3). The approach is utilized to achieve accurate normalization results. The improved technique enhances performance by preserving nuclear patterns critical for distinguishing between cancerous and non-cancerous tissues. The incorporation of Sparse Non-Negative Matrix Factorization (SNMF) ensures precise stain separation, leading to consistent results across varying histopathological images. Initially, NMF separates stain components from the source image. This is enhanced through SNMF optimization, which ensures better separation by enforcing sparsity in the factorization. Additionally, a structure-preserving normalization technique is applied to maintain the structural integrity of the source images post-normalization. This technique preserves spatial information by minimizing alterations to structural features. The combination of SNMF and structure-preserving normalization yields exceedingly precise results while safeguarding the structural and spatial characteristics of the input image. SNMF is represented by the following Equations (1–3):

$$V_S \approx W_S * H_S \tag{1}$$

$$V_R \approx W_R * H_R \tag{2}$$

$$W,H \geq 0, \text{ and } \left|\left|V\text{-}WH\right|\right|_{Fro}^2 \leq \epsilon \tag{3}$$

Where ϵ is a small positive number, W is stain matrix, H is concentration matrix and V is OD matrix. Structure-Preserving Color Normalization (SPCN) is represented by Equation (4), the modified stain matrix is obtained by Equation (5), and the normalized concentration matrix is obtained by Equation (6).

$$Q = inv\ (W_R) * W_S \tag{4}$$

$$\omega = W_S * U * V_T \tag{5}$$

$$H_N = inv\ (\omega) * V_S \tag{6}$$

Q is the transformation matrix from which the left and right singular values (U and V_T) are obtained.

Achieving segmentation through watershed algorithm

The watershed segmentation algorithm employed in the proposed model uses mathematical morphology's topological principles. One notable advantage of watershed segmentation is its efficiency and robustness for delineating objects with indistinct boundaries. Fig. 4 shows the segmented outputs from the watershed segmentation process which are further used to classify the cancerous and non-cancerous images.

Image annotation

Multiple augmented versions of an input image are created using predefined transformations in the image annotation stage through image tiling and image augmentation. Image tiling divides the large segmented image into smaller patches, allowing detailed examination of specific sections without losing resolution. Image augmentation is a DL technique that enriches the training datasets by generating new examples from existing ones. This enhances model generalization and accuracy through diverse transformations such as rotation, flipping, blurring, translation, scaling, cropping, and padding.

Feature extraction

Feature extraction involves capturing pertinent information from each augmented image to facilitate indexing and retrieval. We have chosen LBP and GLRLM for feature extraction in our histopathological image analysis for their strong balance between computational efficiency and effectiveness in capturing critical texture and structural information. LBP excels in capturing local, fine-scale textural patterns, crucial in distinguishing subtle variations in tissue microstructures, such as histopathological cell arrangements. Meanwhile, GLRLM is adept at quantifying the spatial relationships of pixel intensities, which is valuable for identifying larger-scale tissue organization. Both methods are less computationally demanding than more complex techniques like Wavelet Transforms, Gabor Filters, or Gray-Level Co-occurrence Matrices (GLCM).

Local binary pattern

Local binary pattern (LBP), a texture descriptor, exhibits minimal computational complexity. It is relatively robust to variations in radiance due to its fo-

cus solely on the intensity association between sets of pixels.

The process involves comparing the intensity value of each pixel in an input image with the intensity values of P neighboring pixels sampled in a circular area with a radius R around it. If the central pixel has a lower intensity value than the neighboring pixels, a value of 1 is recorded; otherwise, 0 is recorded. This generates a binary array of length which represents a pattern index.

Gray level run length matrix

Gray level run length matrix (GLRLM) is based on counting the occurrences of consecutive picture points with similar gray levels, known as gray-level runs. A run-length matrix $D(s,t)$ is structured such that r is the number of runs in the grey level, s represents the number of consecutive runs of pixels with a specific gray level intensity, and t represents the length of each run in a particular direction. Using a displacement vector $p(x,y)$, orientations are defined where x and y denote displacements along the x-axis and y-axis. This approach utilizes four directions (0°, 45°, 90°, and 135°) to delineate texture runs, resulting in the generation of four run-length matrices. Through GLRLM, seven features namely Short Run Emphasis (SRE), Long Run Emphasis (LRE), Low Gray Level Run Emphasis (LGRE), High Gray Level Run Emphasis (HGRE), Gray Level Non-Uniformity (GLN), Run Length Non-Uniformity (RLN), and Run Percentage (RP) are procured. These are mathematically represented in Equation (7–13):

$$SRE = 1/r \sum_{s,t} (D(s,t))/t^2 \quad (7)$$

$$LRE = 1/r \sum_{s,t} t^2 D(s,t) \quad (8)$$

$$LGRE = 1/r \sum_{s,t} (D(s,t))/s^2 \quad (9)$$

$$HGRE = 1/r \sum_{s,t} s^2 D(s,t) \quad (10)$$

$$GLN = 1/r \sum_{s} (\sum_{t} D(s,t))^2 \quad (11)$$

$$RLN = 1/r \sum_{s} (\sum_{s} D(s,t))^2 \quad (12)$$

$$RP = \sum_{s,t} r/(D(s,t)t) \quad (13)$$

Shape features

Shape-based feature extraction is comprised of calculating the resemblance between shapes based on their typical features. The shape parameters, eccentricity, area, compactness, circularity ratio are assessed in the proposed system as expressed in Equations (14–17).

$$Eccentricity = \sqrt{(1 - w^2/y^2)} \quad (14)$$

In Equation (14), w is the minor axis and y is the major axis.

$$Area = \sum_{e,f=0} p(e,f) \quad (15)$$

where, $p(e,f)$ is the pixel intensity value at the (e,f) point and the image is of Z by B size.

$$Circularity = Area/A_0 \quad (16)$$

where A_0 specifies the area of the circle having the identical length as the shape.

$$Compactness = 4\varpi(Area)/Perimeter^2 \quad (17)$$

where *Perimeter* specifies the quantity of pixels along the boundary of the nucleus.

Color features

Hyperchromatism involves the proliferation of excess chromatin or unnecessary nuclear staining, typically indicative of pathological processes. In this study, numerous color features have been extracted, significantly contributing to identifying malignant nuclei.

Feature selection

The Weighted Between-Class Variance $SBw(f)$, Weighted Within-Class Variance $SWw(f)$, and WFS are computed through the following mathematical Equations (18–20):

$$SBw(f) = \sum_{(i=1)} K\, w_i\, E_i\, (\mu_i - \mu)^2 \quad (18)$$

$$SWw(f) = \sum_{(i=1)} K\, w_i\, V \in CS_i \sum (V_j - \mu_i)^2 \quad (19)$$

$$WFS = SBw(f) * SWs(f) \quad (20)$$

Where, E – total number of samples, K – total number of classes, F – total number of features, E_i – number of samples in class i, μ_i – mean for class i, μ – overall mean, w_i – weight for class i, V_j – value for the sample j in class i, and CS_i – set of samples in class i.

After WFS has been calculated for each feature, the features are ranked based on their Fisher scores in descending order. The features with the highest Fisher scores are selected for classification.

Classification

Predictions are obtained from the improvised classifier, U-Net, for each feature vector corresponding to an augmented image. The proposed U-Net uses feature-based inputs instead of full images, reducing complexity and training time. Its optimized feature selection and architecture enhance clinical application efficiency, generalization, and reliability. It fea-

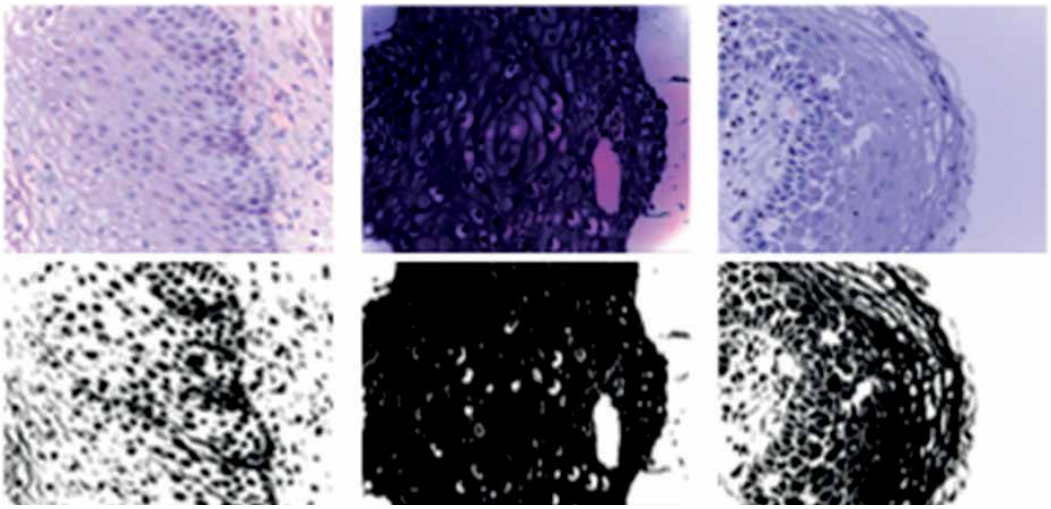

Fig. 3. Representative examples of color-normalized images obtained through the application of the Vahadane Three-Stain Parameter Normalization technique on two different histopathological images

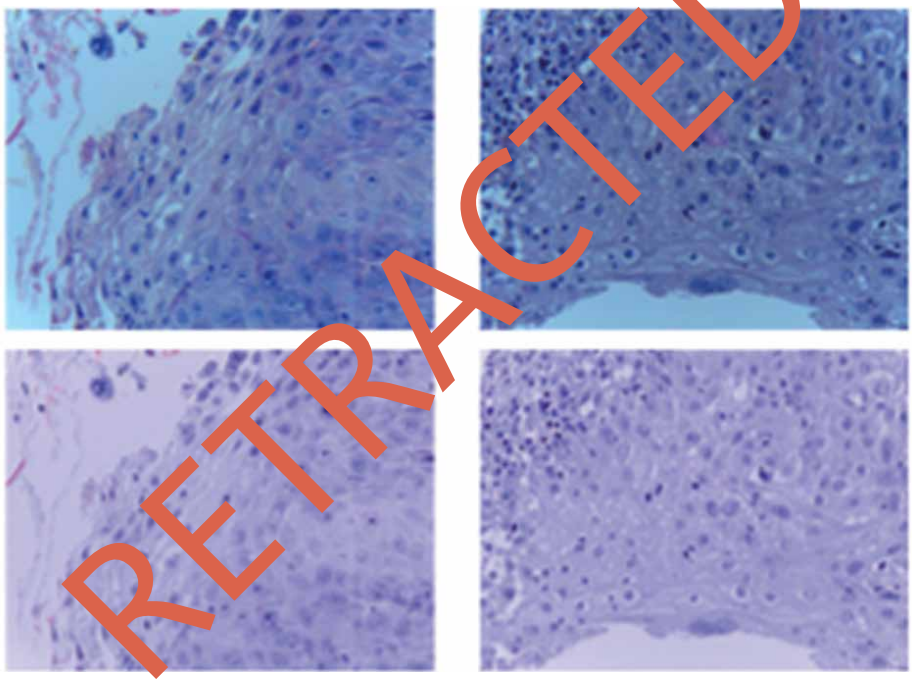

Fig. 4. Representative examples of segmented images obtained through the application of the Watershed Algorithm on four different color-normalized images

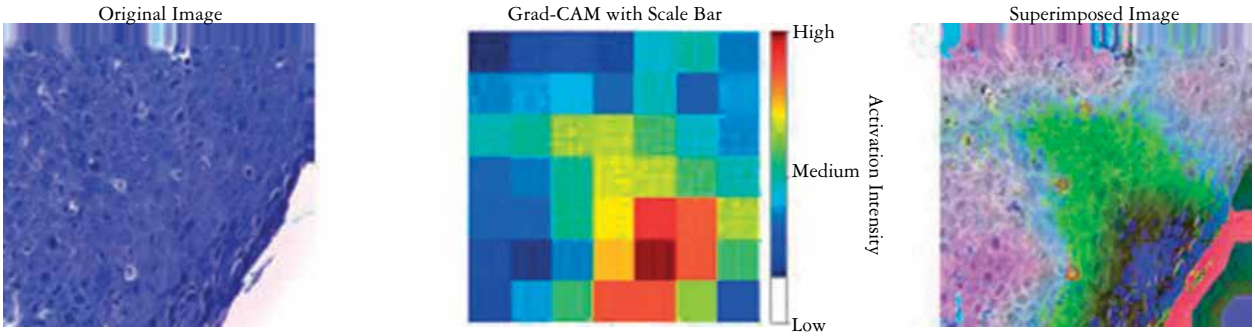

Fig. 5. Visual representation of the Grad-CAM explainable AI technique

Table I. Key performance metrics of the proposed model on datasets

Dataset / Metrics	Accuracy (%)	Precision (%)	F1 Score (%)	Recall (%)	Specificity (%)	NPV (%)	MCC (%)	FNR (%)	FPR (%)
Kaggle dataset	98.72	98.58	98.80	99	98.50	98.80	98.00	0.0078	0.0069
Multi Cancer Dataset	99.32	99.25	99.28	99.30	99.35	99.40	99	0.007	0.0065
NDB-UFES dataset	99.54	99.95	98.44	98.93	99.02	86.40	97.92	0.0254	0.0651

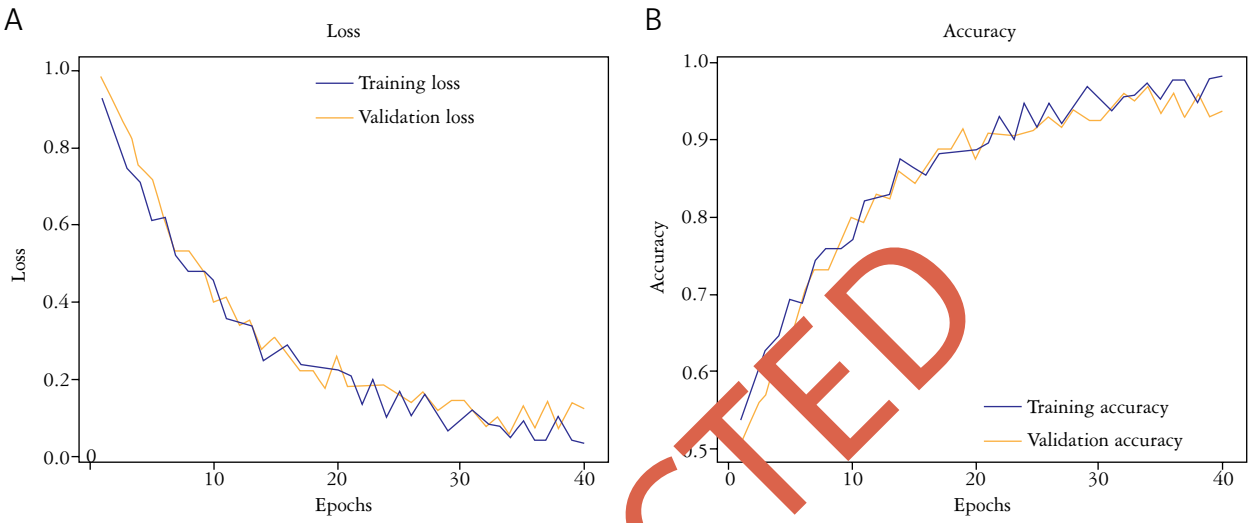

Fig. 6. A) Loss and **B)** accuracy curves of the proposed model on the Kaggle dataset

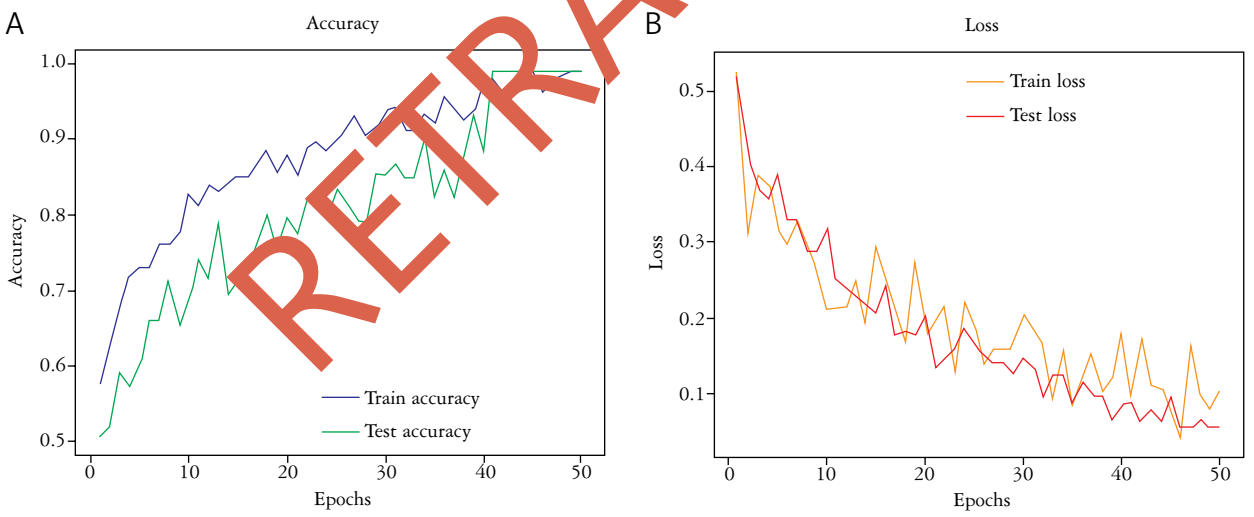

Fig. 7. A) Accuracy and **B)** loss curves of the proposed model on the NDB-UFES dataset

tures both an encoder and a decoder, offering advantages such as high accuracy and the absence of fully connected layers. The classifier comprises six convolutional sub-modules, each consisting of three convolution layers. Each convolution layer utilizes 32 filters with a window size of 3 × 3. The dimensionality of the features is found to be too high. To reduce the size, a pooling layer follows each convolution sub-module. This pooling layer incorporates both average and maximum pooling techniques. In average pooling, a 3 × 3 mask is applied to the feature set, computing the mean of feature values for the final response from the pooling layer. Conversely, maximum pooling applies a 3 × 3 mask to the feature set, selecting the maximum feature value for the final response. Following the concatenation of the

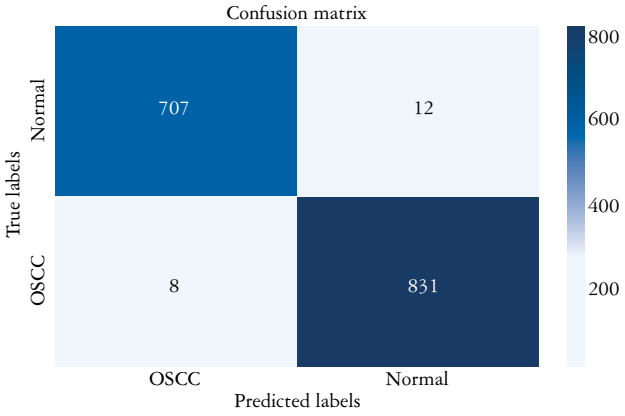

Fig. 8. Confusion matrix of the proposed model with the Kaggle dataset

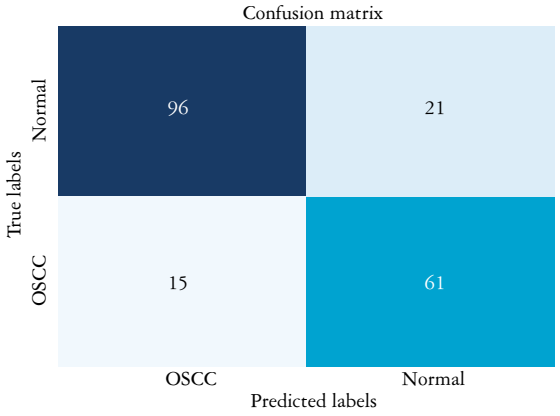

Fig. 9. Confusion matrix of the proposed model with the NDB-UFES dataset

Table II. Dividing Multi Cancer Dataset into training, testing, and validation phases

Datasets	Classes	Training (80%)	Validation (20%)	Testing (20%)
Multi Cancer Dataset	Normal oral cavity	3201	800	1000
	OSCC	3201	800	1000
Total		6402	1600	2000

Table III. Ten-fold cross-validation performance of the proposed model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Fold-1	98.2[illegible]	[illegible].50	98.30	98.40
Fold[illegible]	9[illegible].90	99.30	99.20	99.20
Fold-3	99.40	99.40	99	99.20
Fold-4	[illegible].10	98.70	98.40	98.50
Fold-5	99.50	99.30	99.40	99.40
[illegible]	99.40	99.60	99.30	99.50
[illegible]old-7	99.60	99.40	99.40	99.50
Fold-8	99.70	99.70	99.60	99.70
Fold-9	99.40	99.2	99.10	99.30
Fold-10	99.60	99.5	99.40	99.50
Average	99.20	99.2	99.0	99.20

Table IV. Comparison between the proposed approach and existing state-of-the-art approaches in OC classification

Classifiers / Metrics	Proposed Model	VGGNet	Dens[illegible]Net121	[illegible]tionV3
Accuracy (%)	99.54	94.0	[illegible].0	83
Precision (%)	99.95	9[illegible].2	84.6	77
Recall (%)	98.44	[illegible]8.0	95.0	88
F1-Score (%)	98.93	95.7	89.7	82

Table V. Comparative analysis of MSE and RMSE

Classifiers /Metrics	Proposed model	CNN	SVM	RESNET-152V2
RMSE	0.1133	0.269393	0.287861	0.2
MSE	0.0128	0.0726	0.082864	0.038

input features, a convolution operation is performed using two 3 × 3 kernels, which ReLU subsequently activates. Finally, a 1 × 1 convolution is applied to the final layer, assigning a vector of 56 components to the respective class. The features are input to the decoder of the U-Net architecture, which generates detailed predictions.

Feature prediction

XAI seeks to enhance the reliability of AI solutions by providing transparency into how CNN models make predictions. XAI technique GRAD-CAM has been used to provide comprehension of the model's predictions. The XAI methods identify regions of hyperchromatism as the key distinguishing feature in a subset of OSCC samples. The XAI technique produces a visual heat map to help clinicians verify if the highlighted regions align with pathological markers like tumors, non-tumors, or other disease-specific features. The XAI enables pathologists to detect cancerous regions earlier with greater precision, reducing diagnostic errors and supporting timely intervention. These insights align with clinical observations, underscoring the proposed model's potential to assist pathologists in decision-making.

Table VI. Comparison with recent works using the same datasets as the proposed model

Ref.	Models	Dataset	Accuracy (%)	Precision (%)	F1 score (%)
Ahmad *et al.* 2023 [19]	DenseNet201	Kaggle dataset	97	96.77	93.74
Redie *et al.* 2023 [20]	Modified VGG19	Kaggle dataset	98.64	98.59	98.7
Abbas *et al.* 2024 [21]	Adaptive Federated Learning	Multi Cancer Dataset	90	82	81
Rajadurai *et al.* 2024 [22]	Inceptionv3+Xception+CNN	Multi Cancer Dataset	99	99	99
Proposed model /2024	Improvised U-Net	Kaggle dataset / Multi Cancer Dataset / NDB-UFES dataset	98.72 / 99.32 / 99.54	98.58 / 99.25 / 99.95	98.8 / 99.28 / 98.44

Gradient-weighted Class Activation Mapping

Gradient-weighted Class Activation Mapping Grad-CAM identifies the specific areas the proposed model utilizes to formulate a prediction by aggregating activation maps from the final convolutional layer using linear weighting.

$$G\_C(a) = RELU(\Sigma_{(m=1)}^{M} X_m BN_m) \quad (21)$$

In equation (21), BN_m represents the m-th channel of the activation map from the final convolution layer, attainable during a forward pass, and X_m [illegible] weight allotted to each channel. Figure 5 represents the visualization of the Grad-CAM technique.

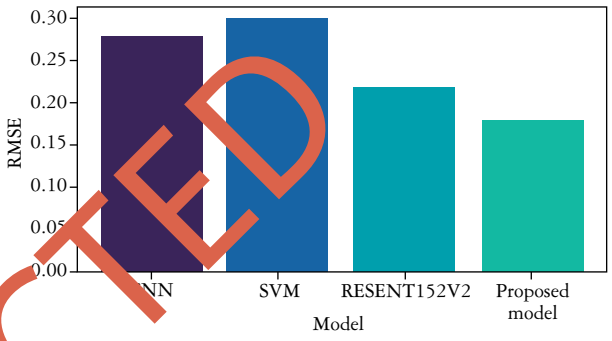

Fig. 10. RMSE assessment

Results and discussion

Dataset description

The proposed model utilized Oral Cancer Histopathological Image Dataset [16], Multi Cancer Dataset [17] and NDB-UFES dataset [18].

Performance analysis of proposed approach

The proposed model shows incomparable performance in OC detection on the Kaggle dataset, Multi Cancer Dataset, and NDB-UFES dataset across various metrics, including accuracy, precision, recall, specificity, F1 Score, Negative Predictive Value (NPV), False Positive Rate (FPR), False Negative Rate (FNR), and Matthews Correlation Coefficient (MCC), as shown in Table I.

We have used accuracy and loss curves to track the proposed model's performance on the Kaggle and NDB-UFES datasets during training. They provide insights into its learning effectiveness and adaptation over time. From Figs. 6 and 7, it is evident that the disparity between training and validation loss, as well as training and validation accuracy, is minimal.

Figures 8 and 9 represent the confusion matrix used in classification models for evaluating performance by comparing actual target values with model predictions.

K fold cross-validation

The Multi Cancer Dataset images are split into three subsets: 80 % for training, 20 % for validation, and 20 % for testing, as shown in Table II. To evaluate the proposed model's robustness, 10-fold cross-validation is used. The results in Table III results confirm the model's strong and stable performance, emphasizing its robustness and dependability.

Comparative evaluation of the proposed and existing models

In Table IV the accuracy of the proposed OC prediction approach is 98.716% whereas other existing state-of-the-art approaches namely VGGNet, DenseNet121, and InceptionV3 obtained accuracy values of 94%, 90%, and 83%, correspondingly. This was due to the combination of the enhanced Vahadane Three-Stain Parameter Normalization approach, watershed segmentation, WFS to address

class imbalance, and an improved U-Net. Table V compares the Root Mean Square Error (RMSE) and Mean Square Error (MSE) values across different models. Figure 10 illustrates the comparison of performance using RMSE between the proposed method and existing models.

The proposed OC prediction model exhibits superior performance compared to existing models, as evidenced by lower MSE and RMSE values. In Table VI, we have additionally compared our proposed work with recent works using the same datasets based on accuracy, precision, and F1-score. In this paper, early identification of OC cases utilizing image processing and DL techniques was investigated. The findings of the study demonstrate how the approaches effectively enable cost-effective early diagnosis of the disease when applied to histopathological images. Uncertainty quantification was employed to evaluate how TTA influenced the model. The classification report and confusion matrix offer valuable insights into the precision and effectiveness of the ensemble model in distinguishing between non-cancerous and cancerous tissue. Compared to individual models, the ensemble model's results highlight its superior overall accuracy. Moreover, visualizations from various XAI techniques enhance interpretability by explaining how the model makes predictions, revealing significant areas or patterns in images influencing the ensemble model's decision-making process.

Conclusions

While CNNs show promising results in detecting early-stage OC, validation through corresponding histopathological examination is necessary. This study introduces an automatic prediction model for OC incorporating an improved Vahadane Three-Stain Parameter Normalization technique to capture greater variability in three stain parameters, watershed segmentation for precise object delineation in ambiguous areas, WFS for addressing class imbalance, and an enhanced U-Net using input features to handle training time. The study demonstrates how DL and XAI methods enhance histopathological image analysis, achieving precise OC classification with computational efficiency via optimized feature selection and U-Net architecture. The XAI technique improves interpretability, supporting diagnostic accuracy in OC treatment. Future work could extend the model to predict OC stages 1–4 with explanatory insights.

Disclosures

1. Institutional review board statement: Not applicable.
2. Assistance with the article: None.
3. Financial support and sponsorship: None.
4. Conflicts of interest: None.

References

1. Al-Rawi N, Sultan A, Rajai B, et al. The Effectiveness of Artificial Intelligence in Detection of Oral Cancer. Int Dent J 2022; 72: 436-447.
2. Jubair F, Al-Karadsheh O, Malamos D, Al Mahdi S, Saad Y, Hassona Y. A novel lightweight deep convolutional neural network for early detection of oral cancer. Oral Dis 2022; 28: 1123-1130.
3. Warin K, Suebnukarn S. Deep learning in oral cancer – a systematic review. BMC Oral Health 2024; 24: 212.
4. Weber A, Enderle-Ammour K, Kurowski K, et al. AI-Based Detection of Oral Squamous Cell Carcinoma with Raman Histology. Cancers (Basel) 2024; 16: 689.
5. Yang SY, Li SH, Liu JL, et al. Histopathology-Based Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning. J Dent Res 2022; 101: 1321-1327.
6. Camalan S, Mahmood H, Binol H, et al. Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results. Cancers (Basel) 2021; 13 1291.
7. Musulin J, Štifanić D, Zulijani A, Ćabov T, Dekanić A, Car Z. An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue. Cancers (Basel) 2021; 13: 1784.
8. Tanriver G, Soluk Tekkesin M, Ergen O. Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders. Cancers (Basel) 2021; 13: 2766.
9. Deo BS, Pal M, Panigrahi PK, Pradhan A. An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. Int J Data Sci Anal 2024; 20: 1005-1022.
10. Marzouk R, Eatedal A, Sami D, et al. Deep Transfer Learning Driven Oral Cancer Detection and Classification Model. Computers, Materials and Continua 2022; 3: 3905-3920.
11. Das M, Dash R, Mishra SK. Automatic Detection of Oral Squamous Cell Carcinoma from Histopathological Images of Oral Mucosa Using Deep Convolutional Neural Network. Int J Environ Res Public Health 2023; 20: 2131.
12. Gomes RFT, Schmith J, Figueiredo RM, et al. Use of Artificial Intelligence in the Classification of Elementary Oral Lesions from Clinical Images. Int J Environ Res Public Health 2023; 20: 3894.
13. Myriam H, Abdelaziz AA, El-Sayed M, et al. Advanced meta-heuristic algorithm based on Particle Swarm and Al-biruni Earth Radius optimization methods for oral cancer detection. IEEE Access 2023; 11: 23681-23700.
14. Song B, Kc DR, Yang RY, Li S, Zhang C, Liang R. Classification of Mobile-Based Oral Cancer Images Using the Vision Transformer and the Swin Transformer. Cancers (Basel) 2024; 16: 987.
15. Confer MP, Falahkheirkhah K, Surendran S, et al. Rapid and Label-Free Histopathology of Oral Lesions Using Deep Learning Applied to Optical and Infrared Spectroscopic Imaging Data. J Pers Med 2024; 14: 304.
16. Oral Cancer Histopathological Image Dataset Link: https://www.kaggle.com/datasets/ashenafifasilkebede/dataset (Access: Jul 2023).
17. Multi Cancer Dataset Link: https://www.kaggle.com/datasets/obulisainaren/multi-cancer (Access: Jul 2023).
18. NDB-UFES dataset Link: https://data.mendeley.com/datasets/bbmmm4wgr8/4 (Access: Jul 2023).
19. Ahmad M, Irfan MA, Sadique U, et al. Multi-Method Analysis of Histopathological Image for Early Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning and Hybrid Techniques. Cancers (Basel) 2023; 15: 5247.

20. Redie DK, Bilgaiyan S, Sagnika S. Oral cancer detection using transfer learning-based framework from histopathology images. J Electronic Imaging 2023; 32: 053004-053004.
21. Abbas T, Fatima A, Shahzad T, Alharbi M, Khan MA, Ahmed A. Multidisciplinary cancer disease classification using adaptive FL in healthcare industry 5.0. Sci Rep 2024; 14: 18643.
22. Rajadurai S, Perumal K, Ijaz MF, Chowdhary CL. PrecisionLymphoNet: Advancing Malignant Lymphoma Diagnosis via Ensemble Transfer Learning with CNNs. Diagnostics (Basel) 2024; 14: 469.

Address for correspondence

V.P. Gladis Pushparathi
Department of Computer Science and Engineering,
Velammal Institute of Technology,
Panchetti, Chennai, Tamil Nadu,
601204, India
e-mail: gladisrathi@gmail.com