Automated Detection of Oral Squamous Cell Carcinoma Using Deep Learning

Nayana G S
Assistant professor
Department of Computer Science
and Engineering, GM University
Davangere, India

Bhumika A
Assistant professor
Department of Computer Science
and Engineering, GM University
Davangere, India

Abstract - **Oral Squamous Cell Carcinoma (OSCC) accounts for a significant proportion of oral malignancies and is often diagnosed at advanced stages due to limitations in early screening methods. To support timely and objective diagnosis, this study presents an automated deep learning–based framework for classifying histopathological oral tissue images as cancerous or non-cancerous. A Convolutional Neural Network (CNN) was trained using a publicly available dataset comprising 1,224 Haematoxylins and Eosin–stained images collected from 230 patients at two magnification levels (100× and 400×). Image preprocessing and class-aware data augmentation were applied to improve learning stability and reduce bias arising from class imbalance. The trained model achieved a training accuracy of 84% and a test accuracy of 77.61%, with high recall for cancerous samples, indicating effective sensitivity to malignant tissue patterns. Although a performance gap between training and validation results suggests moderate overfitting, the findings demonstrate that CNN-based analysis of histopathological images can serve as a reliable assistive tool for oral cancer screening. Future work will focus on enhancing generalization through larger, multi-center datasets and the evaluation of more advanced network architectures.**

Keywords— Oral cancer detection, Convolutional Neural Network, deep learning, histopathological images, medical image classification

I. INTRODUCTION

Oral Squamous Cell Carcinoma (OSCC) represents one of the most frequently diagnosed malignancies affecting the oral cavity and remains a major public health concern worldwide. The incidence of OSCC is particularly high in regions where tobacco use, alcohol consumption, and betel nut chewing are prevalent, and where access to routine oral screening and specialized oncology services is limited. In such settings, patients are often diagnosed only after the disease has progressed to advanced stages, which substantially reduces survival rates and limits treatment options.

OSCC develops from the epithelial lining of the oral mucosa and is associated with gradual cellular and tissue-level alterations, including changes in nuclear morphology, epithelial thickness, and tissue organization. During the early phases of disease progression, these pathological changes may not be easily distinguishable through visual inspection alone. Consequently, early-stage lesions are frequently overlooked in busy clinical environments or in primary healthcare settings with limited diagnostic expertise. Histopathological examination following biopsy remains the gold standard for diagnosis; however, this process is invasive, time-consuming, and dependent on expert interpretation, making it impractical for large-scale or repeated screening.

Advances in artificial intelligence, particularly in the field of deep learning, have enabled the development of automated systems capable of analysing complex medical image data with minimal manual intervention. Unlike traditional machine learning methods that rely on handcrafted features, deep learning models can directly learn discriminative representations from raw image inputs. Convolutional Neural Networks (CNNs) are especially well suited for histopathological image analysis due to their ability to model spatial hierarchies and capture subtle textural and structural patterns associated with malignant transformation.

Although several CNN-based approaches have been reported for oral cancer detection, many existing studies focus on limited datasets, employ computationally intensive architectures, or demonstrate reduced robustness when evaluated across images acquired at different magnification levels. In addition, some models prioritize overall accuracy without adequately addressing clinically important metrics such as sensitivity to cancerous samples. These limitations restrict the practical applicability of such systems in real-world diagnostic settings.

Motivated by these challenges, the present study introduces a CNN-based classification framework designed to distinguish between cancerous and non-cancerous oral tissue using histopathological images captured at multiple magnifications. Emphasis is placed on effective preprocessing, data augmentation, and balanced model design to achieve reliable performance while maintaining computational efficiency. The overarching goal of this work is to contribute toward a non-invasive, computer-aided

diagnostic approach that can support clinicians by facilitating earlier detection of OSCC and reducing delays in clinical decision-making.

II. LITERATURE REVIEW

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have become central to recent advances in oral cancer image analysis due to their ability to learn discriminative representations directly from medical imaging data. Unlike conventional machine learning approaches that depend on manually engineered features, CNN-based models exploit spatial hierarchies and contextual information within images, making them well suited for identifying pathological patterns in oral tissue.

Earlier studies primarily relied on traditional machine learning classifiers to predict oral cancer stages or tissue abnormalities. For example, Fatihah Mohd et al. [1] explored algorithms such as Naïve Bayes, Multilayer Perceptron, K-Nearest Neighbours, and Support Vector Machines using a reduced feature set. While these methods demonstrated improved predictive capability, their performance was strongly influenced by feature selection and limited scalability to complex image data. Subsequent research shifted toward neural network–based approaches, as illustrated by Shreyansh A et al. [2], who evaluated artificial neural networks and transfer learning models on dental radiographic images, highlighting the advantages of learned representations over handcrafted features.

The generalization ability of CNNs across different tissue types and imaging conditions has also been investigated. Halicek et al. [3] applied deep convolutional models to multiple head and neck cancer datasets, demonstrating that CNNs can capture shared morphological patterns across related malignancies. However, their work also emphasized the sensitivity of model performance to dataset size and image acquisition variability.

Several studies have focused on modality-specific oral cancer detection. Aubreville et al. [4] employed CNNs on Confocal Laser Endomicroscopy images using a patch-based classification strategy, enabling localized tissue assessment but increasing computational overhead. In contrast, Albalawi et al. [5] adopted a lightweight EfficientNet-based architecture for histopathological image classification and demonstrated that data augmentation and regularization play a critical role in improving robustness when training on limited datasets. Optimization-driven architectures have also been explored; Nagarajan et al. [6] integrated a swarm intelligence optimizer with MobileNetV3 to enhance feature learning efficiency, though such hybrid approaches may introduce additional computational complexity.

Comparative analyses of deep learning architectures further highlight trade-offs between accuracy and efficiency. Das et al. [7] evaluated multiple pretrained CNN models, including VGG16, ResNet50, and InceptionNet, reporting strong classification performance but at the cost of increased model complexity. Similarly, Panigrahi et al. [9] demonstrated the effectiveness of transfer learning for oral cancer detection, though reliance on large pretrained networks may limit deployment in resource-constrained clinical settings. Fusion-based methods, as investigated by Rahman et al. [8], have shown improved detection reliability but require careful integration of multiple classifiers.

Beyond oral cancer–specific applications, broader studies on histopathological image analysis, such as those by Komura et al. [10] and Fu et al. [11], reinforce the effectiveness of deep learning for tissue-level cancer detection while also emphasizing challenges related to dataset diversity and real-world variability. Related work in dental imaging by Ramzi Ben Ali et al. [12] further supports the applicability of deep neural networks in oral healthcare diagnostics.

Despite substantial progress, existing approaches often face limitations related to computational cost, dataset dependency, and reduced robustness across varying magnification levels and acquisition conditions. Motivated by these observations, the present study aims to develop a CNN-based oral cancer classification framework that balances detection performance with computational efficiency, while maintaining practical relevance for histopathological analysis in clinical environments.

III. MATERIALS AND DATASET

This study employs a publicly accessible histopathological image dataset obtained from an open Kaggle repository to develop and evaluate the proposed Convolutional Neural Network (CNN) for oral cancer classification. The dataset comprises Hematoxylin and Eosin (H&E)–stained oral tissue images representing both healthy epithelium and Oral Squamous Cell Carcinoma (OSCC), enabling supervised binary classification.

A total of 1,224 digitized histopathological images acquired from 230 patients are included in the dataset. Image acquisition was performed using a Leica ICC50 HD digital microscope at two distinct magnification levels, allowing the model to learn multi-scale morphological characteristics. Specifically, the dataset contains 528 images captured at 100× magnification, including 89 normal and 439 OSCC samples, and 696 images captured at 400× magnification, comprising 201 normal and 495 OSCC samples.

Prior to model training, all images were visually inspected and standardized to ensure consistency in resolution and color distribution. The dataset was then partitioned into training, validation, and testing subsets using a stratified split to preserve class balance across all subsets. Data augmentation techniques were applied exclusively to the training set to enhance sample diversity and mitigate overfitting, while validation and test sets were kept unaltered to ensure unbiased performance evaluation.

Representative examples of normal and cancerous histopathological images from the dataset are illustrated in Fig.1.demonstrates the visual differences captured at varying magnification levels.
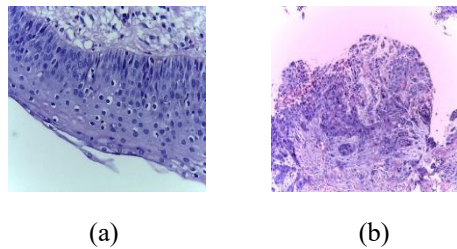
Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

(a)   (b)

Fig.1. Representative H&E-stained histopathological images from the OSCC dataset: (a) normal oral epithelium and (b) oral squamous cell carcinoma (OSCC).

IV. PROPOSED METHODOLOGY

This section describes the proposed deep learning framework for automated oral cancer detection from histopathological images. The methodology is designed to balance classification accuracy and computational efficiency for practical clinical deployment.

A. Overview of the Proposed System

The proposed system follows a sequential processing pipeline as illustrated in Fig. 3.2 The block diagram of the proposed oral cancer detection system illustrates the sequential flow of operations involved in automated classification. The process begins with the acquisition of input medical images, specifically histopathological images of oral tissue. These images are subjected to an image preprocessing stage that includes resizing, normalization, and enhancement to ensure uniform input quality. An optional image segmentation step may be employed to isolate regions of interest and focus the analysis on lesion-affected areas. The processed images are then forwarded to the Convolutional Neural Network (CNN) for feature extraction, where relevant spatial and textural characteristics are learned automatically. During the model training phase, these features are used to optimize the network parameters using labeled data. Performance evaluation is subsequently performed using standard metrics to assess classification reliability, and finally, the trained system generates a prediction indicating whether the input image corresponds to cancerous or non-cancerous oral tissue. In addition, the modular structure of the system ensures that each processing stage can be independently analyzed and optimized. This design flexibility allows improvements in preprocessing or feature extraction without affecting the overall framework. The CNN-based approach reduces reliance on manual feature engineering and subjective interpretation. Furthermore, the automated workflow minimizes diagnostic variability and supports consistent decision-making. Overall, the proposed system is designed to function as an effective computer-aided diagnostic tool for assisting clinicians in oral cancer detection.
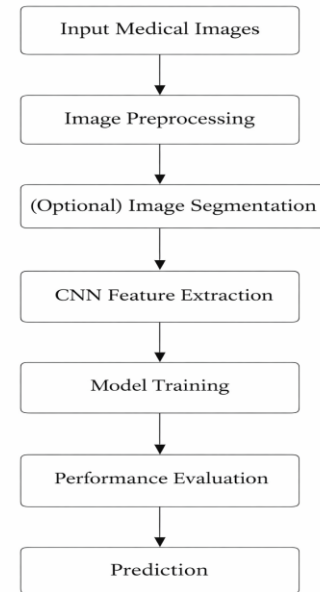

Fig. 3.2. Block diagram of the proposed CNN-based oral cancer detection system.

B. Image Preprocessing

Histopathological images often exhibit variations in resolution, staining intensity, and illumination, which can adversely affect model learning. To address this, all images are resized to a fixed spatial resolution compatible with the CNN input layer. Pixel intensity normalization is applied to scale image values into a consistent range, thereby improving numerical stability during training.

To enhance generalization and reduce sensitivity to spatial orientation, data augmentation techniques including random rotations and horizontal/vertical flipping are applied during training. These transformations increase effective sample diversity without altering the underlying tissue structure. Augmentation is strictly limited to the training set to prevent information leakage into validation and test data. A representative example illustrating the effect of preprocessing is shown in Fig. 3.3.
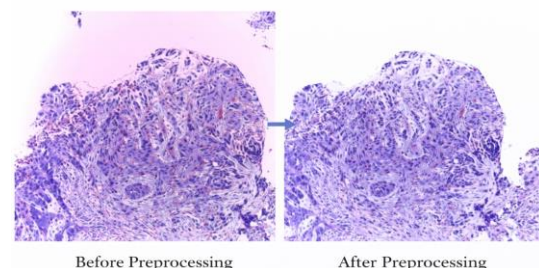

Fig. 3.3. Representative histopathological image before and after preprocessing

C. CNN Architecture Design

The proposed CNN architecture is specifically tailored for binary classification of histopathological oral

tissue images and is illustrated in Fig. 3.4. The network consists of multiple convolutional blocks followed by fully connected layers.

Each convolutional block comprises a convolutional layer with small-sized kernels (e.g., 3×3) to capture fine-grained cellular and textural patterns, followed by a Rectified Linear Unit (ReLU) activation to enable efficient gradient propagation. Max-pooling layers are introduced after selected convolutional blocks to reduce spatial resolution while retaining salient features, thereby lowering computational complexity and improving robustness to minor spatial variations.

As the network depth increases, the number of feature maps is progressively expanded to allow learning of higher-level representations, such as nuclei arrangement and tissue irregularities associated with malignancy. The final convolutional output is flattened into a one-dimensional feature vector, which is passed through fully connected layers for high-level feature integration. Dropout regularization is incorporated in these layers to reduce co-adaptation of neurons and mitigate overfitting.

The output layer consists of a single neuron with a sigmoid activation function, enabling probabilistic binary classification between cancerous and non-cancerous tissue classes.
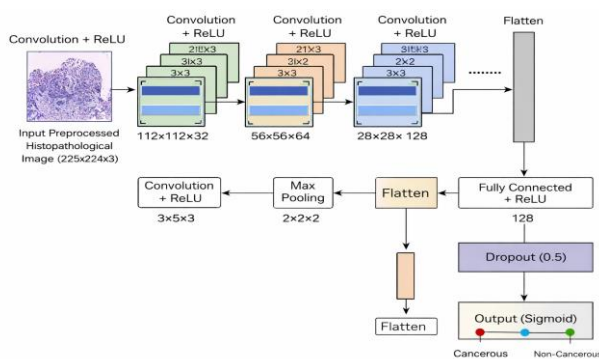

Fig. 3.4 Architecture of the proposed Convolutional Neural Network (CNN) for histopathological oral cancer classification.

D. Model Training Strategy

The CNN model is trained using a supervised learning approach with labeled histopathological images. The dataset is divided into training, validation, and testing subsets using a stratified splitting strategy to preserve class distribution across all subsets.

Model optimization is performed using the Adam optimizer due to its adaptive learning rate capability and stable convergence behavior in deep networks. Binary cross-entropy is employed as the loss function, as it is well suited for two-class classification problems. Training is conducted over a fixed number of epochs with an empirically selected batch size. Validation loss and accuracy are monitored during training, and the model exhibiting the best validation performance is retained for final testing.

E. Performance Evaluation Metrics

The trained CNN model is evaluated on an independent test dataset to assess its generalization capability. Multiple performance metrics are used to provide a comprehensive evaluation, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC–ROC).

While accuracy reflects overall classification correctness, recall is emphasized due to its clinical importance in minimizing false negatives during cancer detection. Precision evaluates the reliability of positive predictions, and the F1-score provides a balanced measure of precision and recall. The AUC–ROC metric is used to analyze the model's discriminative ability across varying decision thresholds.

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results obtained from training and testing the proposed CNN model for oral cancer detection using histopathological images.

A. Experimental Results

The proposed CNN model was trained for multiple epochs using labeled cancerous and non-cancerous oral images. The model performance was evaluated on training, validation, and test datasets. Fig. 4.1and Fig. 4.2 illustrates the training and validation accuracy and loss curves.

As shown in Fig. 4.1 and Fig. 4.2, the training accuracy increases steadily and reaches 83.47%, while the training loss decreases consistently, indicating effective learning. In contrast, the validation accuracy peaks at approximately 62% and shows fluctuations in later epochs, suggesting potential overfitting and the need for improved regularization or increased dataset diversity.
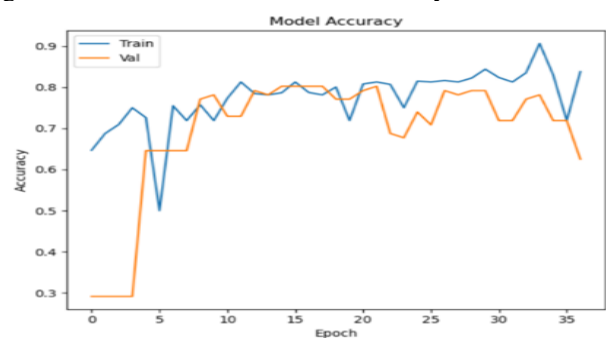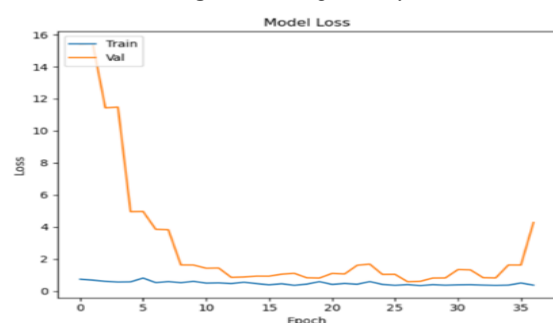

Fig. 4.1. Training accuracy

Fig. 4.2. loss accuracy

To further analyze the classification performance, a confusion matrix is generated on the test dataset, as shown in Fig. 4.2. The confusion matrix indicates that the proposed model correctly classifies a large number of cancerous and non-cancerous samples, demonstrating its effectiveness in oral
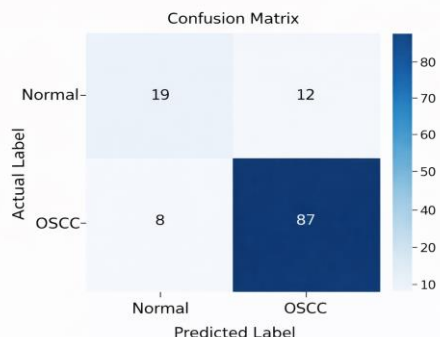

Fig. 4.2. Confusion matrix of the proposed CNN model.

A quantitative evaluation of the proposed CNN model is conducted using standard classification metrics, including precision, recall, F1-score, and accuracy. The classification report indicates that the model achieves a precision of 0.88 and a recall of 0.92 for cancerous samples, demonstrating high sensitivity, which is essential in medical diagnosis. For normal tissue samples, the precision and recall values are 0.70 and 0.61, respectively. The overall accuracy of the model is 84%, with macro and weighted average F1-scores of approximately 0.78, indicating balanced performance across both classes.

The final evaluation on the test dataset yields a testing accuracy of 77.61%. Although the testing performance is slightly lower than the training accuracy, the results demonstrate that the proposed CNN model is capable of effectively distinguishing between cancerous and non-cancerous oral tissues. Sample prediction results for normal and cancerous tissue images are shown in Fig. 4.3 and Fig. 4.4, respectively.
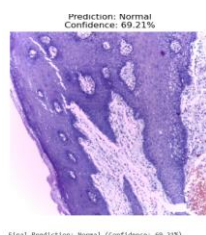

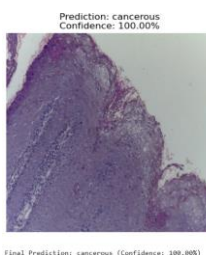**Fig. 4.3.** Sample histopathological image of normal oral tissue.

Fig. 4.4. Sample histopathological image of cancerous oral tissue.

B. Discussion

Overall, the experimental results demonstrate that the proposed CNN model is effective in distinguishing cancerous and non-cancerous oral tissues from histopathological images. The high recall achieved for cancerous samples highlights the model's ability to minimize false negatives, which is critical in medical diagnosis. Although a performance gap between training and validation accuracy indicates moderate overfitting, the obtained results remain promising. Future improvements can be achieved by incorporating larger datasets, enhanced data augmentation, regularization, and transfer learning techniques to further improve generalization performance.

VI. CONCLUSION

This study investigated the application of a Convolutional Neural Network (CNN) for oral cancer detection using histopathological image data. The experimental results demonstrate that the proposed CNN model effectively learns discriminative features from oral tissue images and achieves reliable classification performance. The observed trends in training and validation accuracy and loss indicate stable convergence and satisfactory generalization capability. Confusion matrix analysis further confirms that the model correctly identifies a substantial proportion of both cancerous and non-cancerous samples. Moreover, the obtained precision, recall, and F1-score values highlight the robustness of the proposed approach, particularly in detecting cancerous tissues, which is a critical requirement in clinical diagnosis. Overall, the results validate the suitability of CNN-based methods for automated oral cancer detection and suggest that the proposed model can serve as a supportive decision-making tool for healthcare professionals. Future work will focus on improving generalization through larger multi-center datasets and the integration of advanced deep learning architectures.

REFERENCES

[1] E. Albalawi *et al.*, "Oral squamous cell carcinoma detection using EfficientNet on histopathological images," *Frontiers in Medicine*, vol. 10, Art. no. 1349336, 2024.

[2] M. Aubreville *et al.*, "Automatic classification of cancerous tissue in laser endomicroscopy images of the oral cavity using deep learning," *Scientific Reports*, vol. 7, Art. no. 11979, 2017.

[3] N. F. Mohd, N. M. M. Noor, Z. A. Bakar, and Z. A. Rajion, "Analysis of oral cancer prediction using feature selection with machine learning," in *Proc. 7th Int. Conf. on Information Technology (ICIT)*, 2015, pp. xx–xx.

[4] Q. Fu *et al.*, "A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study," *EClinicalMedicine*, vol. 27, Art. no. 100558, 2020.

[5] M. Halicek *et al.*, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *J. Biomed. Opt.*, vol. 22, no. 6, Art. no. 060503, 2017.

[6] B. Nagarajan *et al.*, "A deep learning framework with an intermediate layer using the swarm intelligence optimizer for diagnosing oral squamous cell carcinoma," *Diagnostics*, vol. 13, no. 22, Art. no. 3461, 2023.

[7] S. Panigrahi *et al.*, "Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning," *Heliyon*, vol. 9, Art. no. e13444, 2023.

[8] S. A. Prajapati, R. Nagaraj, and S. Mitra, "Classification of dental diseases using CNN and transfer learning," in *Proc. 5th Int. Symp. on Computational and Business Intelligence*, 2017, pp. xx–xx.

[9] T. Y. Rahman, L. B. Mahanta, A. K. Das, and J. D. Sarma, "Study of morphological and textural features for classification of oral squamous cell carcinoma using traditional machine learning techniques," Cancer Medicine, vol. 9, no. 3, pp. 2020.

[10] D. K. Shumpei, "Machine learning methods for histopathological image analysis," Computational and Structural Biotechnology Journal, vol. 16, pp. 34–42, 2020.