

Optimized Deep Learning for Feature Extraction and Detection of Oral Cancer from Histopathological Images

Tamara Afif Anai^{1,*} , Mays Afif Anaee² , Samah Mohamed³

¹ Department of Basic Science (Computer Science), College of Dentistry, Tikrit University, Tikrit, Iraq.

² Department of Network and Cybersecurity Engineering at Al-Iraqia University, Baghdad, Iraq.

³ Wireless Intelligent Networks Center (WINC), Nile University, Giza 12677, Egypt.

*Corresponding Author: Tamara Afif Anai

DOI: <https://doi.org/10.31185/wjcms.542>

Received 25 June 2026; Accepted 07 September 2026; Available online 30 September 2026

ABSTRACT: This study aims to develop an efficient, powerful deep learning model that accurately classifies oral cancer from high-resolution histopathology images, especially in cases with high intra-class variability and extreme class imbalance. We propose a novel Ensemble Deep Learning framework combined with two streams: CNN1 (EfficientNet-B5), CNN2 (DenseNet-201), and a new Hybrid -Spatial-Channel Attention module to extract and fuse the cytological and structural features. We addressed class imbalance using a cost-sensitive learning approach and data augmentation techniques. The Whale Optimization Algorithm (WOA) was used for optimizing hyperparameters. The model is based on late-fusion of CNN1 (fine-grained features) and CNN2 (architectural features). The procedure that was followed was to utilize seven classes of 12425 high-resolution images from (ORCHID and UFES-NDB) and diagnose the seven specific histopathological classes into three primary diagnostic groups: normal, pre-cancer, and cancer. The dataset was divided into Training (80%), Validation (10%), and Testing (10%) to prevent data leakage. It is found that the overall accuracy of the proposed architecture is very robust at 90.35% on the test set. Using five-fold cross-validation. The model achieved the highest Matthews Correlation Coefficient (MCC), particularly in highly imbalanced classes. The proposed method can effectively improve the representation of features and the precision of classification for complex medical image analysis by achieving high metrics in Accuracy, Precision, Recall, Specificity, and F1-Score, which will be a powerful computational tool for automated histopathology analysis. Optimized ensemble learning enhances the detection and classification of Oral Squamous Cell Carcinoma (OSCC) and precancerous lesions by merging fine-grained features and architectural features. Future models should be validated in independent patient populations on external datasets to make it ready for full clinical use and investigated regarding the incorporation of multimodal patient information.

Keywords: Deep learning, Feature Extraction, EfficientNet-B5, DenseNet-201, Whale Optimization Algorithm, Oral Cancer, Histopathological Images.



1. INTRODUCTION

Oral cancer is one of the most prevalent cancers worldwide, particularly in areas where the use of tobacco and alcoholic beverages is common. Oral cancer may be influenced by genetic factors, vegetables, fruits, and persistent irritation from badly fitting dental profited prostheses. Oral cancer can be difficult to detect, as many early signs of oral cancer are subtle. Oral Squamous Cell Carcinoma (OSCC) is a cancer cells that develop in the squamous cells (the cells that line mucous membranes) of the inside of the mouth (tongue, gums, buccal or floor of the mouth). A type of oral cancer that is very common is squamous cell carcinoma [1].

Common methods of diagnosing oral cancer include visual examinations, biopsies, and tests such as Computed Tomography scans (CT) and Magnetic Resonance Imaging scans (MRI). The definitive diagnosis is based on histopathology, which, although a gold standard, is time-consuming. The Artificial Intelligence (AI) holds great promise for improving the way oral cancer is diagnosed, especially in light of the difficulties faced by conventional methods. AI algorithms, especially Convolutional Neural Networks (CNNs), are capable of accurately interpreting medical images that can help detect and classify any potential malignant lesions at an early stage. These automated systems improve the efficiency of diagnostics, minimize human error, and enable quick decision-making, which will lead to better survival rates for patients [2].

Deep Learning (DL) and particularly CNNs are able to extract the hierarchical feature representation of images and to accurately classify complex patterns with pathological conditions. While CNN single architectures have proved to be useful, they have pitfalls in generalizing across a wide range of datasets, overfitting, and changes in image capture techniques. Such restrictions call for an ensemble-based DL system that combines various CNN structures to harness their complementary characteristics. One of the most popular methods in the field of Machine Learning is Ensemble learning which consists of the aggregation of several classifiers. Ensemble CNNs have been demonstrated to improve the diagnostic performance in medical image classification tasks by utilizing a large number of different feature representations extracted from different network architectures. Here, CNN1 is given by the EfficientNet-B5 model while CNN2 is provided by DenseNet-201 to enhance the classification accuracy and computational efficiency for cancer detection in advanced ensemble CNN [2]-[5].

To improve localization of cancerous areas, the EfficientNet-B5 model uses a Squeeze-and-Excitation (SE) attention module with a Hybrid spatial-channel attention module. DenseNet-201 can extract the complex textural and hierarchical information which is crucial for good classification. In the fusion of two models, we keep the feature maps of the two models in the penultimate layer and fuse them, then pass them through a global attention pooling layer, enhancing the representation even more before final classification. The selection of *LearningRate* (*LR*), *BatchSize* (*BS*), *DropoutRate* (*DR*), and *WeightDecay* (*WD*) in training DL models can greatly affect the convergence and generalization of the models [3]-[5].

To overcome this problem, the (WOA) is a nature-inspired metaheuristic optimization algorithm that is inspired by the whale hunting behavior of humpback whales. Adaptive mechanisms are considered to examine the solution space and this renders WOA highly suitable for hyper parameter optimization in DL. The effectiveness of WOA is analyzed by investigating the optimization of CNN1 and CNN2 hyperparameters to be used in the ensemble CNN model, which can converge faster during training and improve the classification accuracy [6]. The study includes the ORCHID and UFES-NDB datasets, which are large in size histological images.

Research motivation: Three primary challenges are encountered by present automatic systems for oral cancer diagnosis:

- (1) intra-class morphological variations at different stages of the disease.
- (2) high class imbalance which affects the prediction.
- (3) domain shift due to a different staining protocol used at different institutions.

CNNs have the tendency to simply overfit or not simultaneously capture both local cellular details and global tissue architecture when used in a single architecture. Previous studies have used single CNN models, and the common problem is extreme class imbalance and domain shift between datasets. This study aims to fill this gap by proposing a powerful ensemble model with a hybrid-attention module that adapts features dynamically. In addition, this work, unlike most of the literature, strictly adheres to patient-level data splitting to prevent any possibility of data leakage. To obtain a more reliable diagnostic tool, (WOA) is used to optimally select the hyperparameters.

The following is a possible conclusion to this study:

- Feature fusion approach: EfficientNet-B5 and DenseNet-201 are proposed to combine the efficient feature extraction capability of EfficientNet-B5 with the deep hierarchical learning capability of DenseNet-201 to improve the efficiency of the feature extraction process and enhance the classification efficiency.
- Implementation of the (WOA) hyperparameter optimization technique to enhance the training parameters of the CNN and improve the model performance.
- Evaluating the Performance: by using performance metrics to prove its effectiveness and generalizability.

This paper is organized as follows: In Section 2, related works on deep learning-based cancer detection. The methodology is described in Section 3, including dataset, CNN models, and the WOA algorithm. The experimental results are presented in Section 4. Finally, the conclusion and recommendations are presented in Section 5.

2. RELATED WORKS

Deep Learning has proven successful in medical image analysis and has become a major focus, especially for cancer detection [7]-[9]. The model being proposed for oral cancer detection carefully preprocesses the ORCHID and UFES-NDB datasets to provide a high-quality representation of features. Feature extraction enhancement of ensemble models using Mask-Mean-Shift CNN and SV-Onion-Net is a novel DL architecture that improves the accuracy and performance for detecting (OSCC) in histopathological images [10].

Kumar and Harikumar [11] proposed a neural network to identify and classify oral cancer. Determined the types of oral lesions; some oral lesions can develop into oral cancer. These lesions can help in early detection of benign or malignant lesions in the mouth. A linear network was used for this study to identify the type of lesion. A diagnostic method was proposed and tested on 75 patients. The accuracy of categorization was about 85.19%, 100%, 94.12%, and 84.21%.

Fahed et al. [12] detected oral cancer using CNN. CNN method is applied to the classification of cancerous and non-cancerous by using EfficientNet-B0 to train the model. The CNN has been trained and evaluated to achieve its performance with 715 clinical images. The accuracy was ~85.0%, the specificity was 84.5% and the sensitivity was 87%.

According to the results, CNN technique could be a good, efficient and affordable technique for the diagnosis of mouth cancer.

Camalan et al. [13] proposed a CNN model for the diagnosis of Multilayered (SCC) of the oral cavity using the Inception-ResNet-V2. Images of 54 patients with oral lesions of multiple layers were used to assess the performance of the network for SCC detection. The results of the classification of malignant lesions showed that the accuracy of 74%. The results showed that the proposed network was effectively applied in diagnosing mouth diseases.

Rajan and Samuel [14] Using a deep CNN for the diagnosis of cancer in the mouth, which occurred in 287 cases, suggests that a correct and timely diagnosis of the presence of cancer in the mouth could significantly help the early treatment of the disease. There are many methods to diagnose this disease. The accuracy, sensitivity, and specificity of the CNN's performance in the assessment of malignant lesions were used to evaluate the CNN. The precision was found 91.4%, sensitivity 94% and specificity 91%. In addition, the accuracy of categorization of mouth cancer was 95%.

Hu et al. [15] used anisotropic preprocessing and fuzzy C-means categorization to identify the malignant mouth lesions. In the present work two methods have been proposed, namely, anisotropic diffusion method and fuzzy C-means, both of which are used to enhance the image resolution for the better identification. The results showed an accuracy of 90.11%, 87.5% specificity value, and 92.16% sensitivity value. The results suggested that the proposed technique was effective for the early diagnosis of malignant oral lesions that are malignant.

To improve the generalization capability and decrease the variance, Ensemble-based methods have been shown to outperform individual CNN designs by combining complementary information from several models [16]. Hyperparameter optimization using nature-inspired algorithms has proven to be effective in improving convergence and performance [17].

The feature selection technique was improved by Ali and Mohammed [18] showed the significant and accurate classification of cancer subtypes. The model achieved a result rate of 81%. After feature selection, the accuracy turned out to be quite outstanding at 100% and better than existing models. That research provided a solid foundation for further biological studies, which would lead to new information and advances in cancer research. Also, Mohammed and Ali [19] used the Quantum Cat Swarm Optimization (QCSO) algorithm to select features along with K-means clustering and Support Vector Machine (SVM) for classification. Using QCSO allowed for the most relevant features to be identified, which resulted in a higher accuracy.

The proposed oral cancer detection model could be extended by using similar methods for improving the applicability of the model in clinical use. The incorporation and clinical implementation of AI technologies in healthcare environments have been recently examined, with particular focus on the need for explainable AI (XAI) to increase the transparency and clinical trust of these systems, especially in diagnostics [20].

The literature review showed a few limitations in the methods used for detecting oral cancer. The early works focused on shallow classifiers or single-stream CNNs (such as Inception-ResNet-V2 or EfficientNet-B0) with moderate accuracy (74%–85%) but lacked the capacity for feature extraction. The previous works used shallow classifiers or single-stream CNNs (such as Inception-ResNet-V2 or EfficientNet-B0), and the accuracy was around 74%–85%, but they didn't have enough capacity to differentiate pathohistological images in multiple classes. Recent studies employed feature selection algorithms, such as QCSO and fuzzy logic, to improve the diagnostic performance, and these algorithms were primarily used for binary classification (normal versus malignant) in small and single-center cohorts.

In addition, another common weakness in the literature is the absence of strict splitting of patient-level data. When using a random split (as opposed to patient slides), data leakage can produce artificially higher performance metrics. Furthermore, most of the previous research publications do not consider the variability of stains between multi-source datasets or the problem of a high-class imbalance in multi-grade carcinomas. This work directly tackles these limitations by proposing a 7-class patient-level split framework with stain normalization, dual-stream feature fusion, attention recalibration, and hyperparameter optimization by WOA.

3. THE METHOD

3.1 DATASETS

The Histopathological images using in this study for the oral cancer database (ORCHID) [21] and (UFES-NDB) [22], which are large-scale datasets created to facilitate research on AI applications for oral cancer detection from histological images. It includes 12425 high-quality images of a range of oral cancer and precancerous lesions. 300 x 300-pixel resolution were created by sequentially cropping the images (150-pixel) to ensures complete coverage of the tissue sample and quality for images. The dataset has been preprocessed to make the processing of oral cancer diagnosis more efficient. The training set was extensively preprocessed and augmented (Random Resized Crop, Rotation, and Flipping). The samples were obtained from rigorous stratified sampling and divided into 80% Training, 10% Validation, and 10% Testing, which was done at the patient/original slide level prior to any patch cropping. This technique was used to make sure that there is no data leakage between the splits. Cross-validation technique using the k-fold ($k = 5$) to test the ability for generalize. This approach will ensure that all data is utilized for training as well as validation and will

minimize the risk of overfitting. According to the proposed model, the images are classified into seven different groups, each of which is clinically categorized into one of three categories: normal, pre-cancerous, and cancerous. Figure 1 displays the dataset, and distribution the dataset is showed in Table 1.

The details of the dataset are as follows:

1. Normal: Normal oral mucosa, with no inflammatory or malignant changes in the cell structure. Figure 1(a) [23].
2. Oral Submucous Fibrosis (OSMF): A serious, high-risk pre-cancerous disease which affects the mouth and is strongly associated with betel quid chewing and results in scarring of tissue and significantly restricted mouth opening. Figure 1(b) [24].
3. Leukoplakia without Dysplasia: A low-risk white lesion that has thickened tissue, but no cancerous cells. Figure 1(c) [25].
4. Leukoplakia with Dysplasia: A white patch that is pre-cancerous with abnormal structural changes in the cells that may develop into cancer. Figure 1(d) [26].
5. Well-Differentiated OSCC (WDOSCC): A less aggressive, slower-growing oral cancer. Unfortunately, the malignant cells still closely resemble normal cells, and generally have a better prognosis. Figure 1(e) [27].
6. Moderately Differentiated OSCC (MDOSCC): The cells become partially distorted and grow and spread at a rate greater than WDOSCC. Figure 1(e) [27].
7. Poorly Differentiated OSCC (PDOSCC): The most aggressive and fast-growing oral cancer. Cells are very abnormal and are not recognized, have a high chance of quickly spreading to other parts of the body, and are the hardest to treat. Figure 1(e) [27].

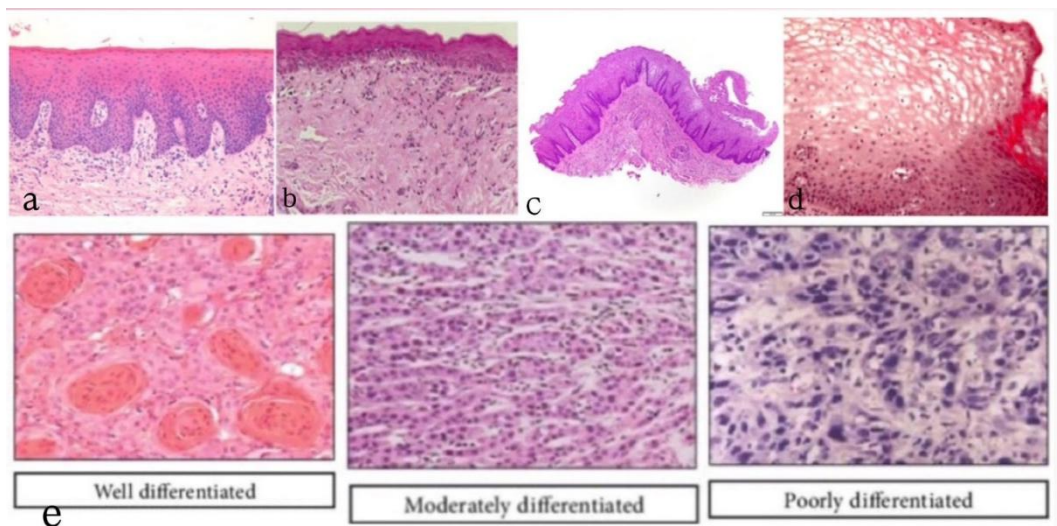


Figure 1: Histopathological Images a) The healthy oral mucosa, b) Image of OSMF, c) The Leukoplakia without dysplasia, d) The Leukoplakia with dysplasia, and e) The Grades of OSCC.

Table 1: Distribution of the Dataset

Classes	Description	No. of Images
Normal	Healthy Tissue	1,045
Leukoplakia without Dysplasia	Benign/Reactive Tissue	589
Leukoplakia with Dysplasia	Pre-cancer / OPMD	1,608
OSMF	Pre-cancer / OPMD	2,095
WDOSCC	Well-Differentiated Carcinoma	2,790
MDOSCC	Moderately Differentiated Carcinoma	2,699
PDOSCC	Poorly Differentiated Carcinoma	1,599
Total of Images		12,425

3.1.1 COMPUTATIONAL FRAMEWORK AND TRAINING DURATION

The training of model and evaluation process was performed locally, and an NVIDIA GeForce RTX 5060 OC (8GB GDDR7) GPU was used to boost the computation time. The initial 50 epochs of training on the entire set of 12,425 high-resolution images took about 10 hours of continuous compute time. In order to train the complex Ensemble network under the constraint of 8GB VRAM on the GPU, memory-management techniques like mixed precision training and gradient accumulation were used. This enabled the system to sustain an effective batch size that is large enough to guarantee stable convergence, while avoiding memory constraints.

3.2 PROPOSED METHOD

In this paper, we propose a deep learning-based classification system using an optimized ensemble of EfficientNet-B5 and DenseNet-201. The model is optimized through hyperparameter optimization using the Whale Optimization Algorithm (WOA) to achieve optimal training efficiency and system reliability. In addition, data pre-processing, data augmentation, and Region of Interest (ROI) extraction improve the system's performance, enabling the precise mapping of seven specific histopathological classes into three primary diagnostic groups: normal (Normal oral mucosa), pre-cancer ((OSMF), (Leukoplakia without Dysplasia), and (Leukoplakia with Dysplasia)), and cancer ((WDOSCC), (MDOSCC), and (PDOSCC)). The pipeline starts with an organized data assembly process to ensure the collection of a robust and representative set of data. Images are first resized (224×224 pixels) to align with the fixed input constraints set by the Enhanced-EfficientNet-B5 and DenseNet-201 ensemble. After resizing the pixels, the intensity values are scaled to (0 & 1 or -1 & 1). The normalization process ensures a more stable training process, even when there are discrepancies in illumination and contrast, while maintaining a good learning curve. There is a normalization of the color space, which is especially implemented for histopathological images, in order to counteract the variation of stains and improve the visibility of tissue. Data augmentation adds deliberately modified versions of the data to increase the size of the dataset, reduce overfitting, and improve generalization. Finally, image segmentation and ROI extraction (mainly on histopathological) are carried out to exclude unwanted noise from the image, thus directing the attention of the model to the important structures of the cancerous tissue. The non-tissue background noise was removed from the patches by converting them to grayscale, applying Otsu thresholding to produce a binary tissue mask, and cropping bounding boxes around the tissue structures. These normalization and ROI extraction processes are carried out comprehensively to adjust the imaging conditions in the merged image datasets. This will help the model to master pathological features and not dataset-specific artifacts or illumination and staining variations, and Class imbalance was addressed using Weighted Cross-Entropy Loss ($\mathcal{L}_{WCE}$) using Equation (1).

$$\mathcal{L}_{WCE} = - \sum_{i=1}^N \sum_{c=1}^K w_c \cdot y_{i,c} \log(\hat{y}_{i,c}), \quad w_c = \frac{N_{total}}{K \cdot N_c} \quad (1)$$

Where $K = 7$, N_{total} is the total sample count, and N_c is the sample count for class c .

For feature extraction, the EfficientNet-B5 model employs a hybrid spatial-channel attention mechanism and (SE) blocks to optimize the extraction of features. The SE block normalizes the feature responses with channel-wise attention to emphasize informative channels, and the Hybrid Spatial-Channel Attention strategically tunes the spatial and channel weights to improve target localization. These mechanisms of attention work together to improve the representations of features before they are fed into classification layers. In the end, cross-validation is used to guarantee strong training and model optimization.

3.2.1. CNN1: IMPROVED EFFICIENTNET-B5 ARCHITECTURE

The first DL architecture is CNN1, which implements compound scaling methodology for the depth, width, and resolution of the network, similar to the popular EfficientNet-B5 framework [28]. It is very useful for distinguishing the subtle textures, color differences, and structural patterns of lesions, which are essential to early malignancy diagnosis. In this model, a Squeeze Excitation (SE) block is used to improve feature extraction, and the hybrid spatial-channel attention mechanism is used to accurately perform the localization of the corresponding regions in the oral cancer images. The structural changes that are added to Enhanced EfficientNet-B5 include:

- Convolutional the Layer: In the beginning input (224×224 pixels) image goes through a 3×3 convolutional layers with stride 2. This is followed by a Swish activation function and batch normalization in sequence to enhance feature propagation across the network.
- MBConv Blocks: It consists of multiple Mobile Inverted Bottleneck Convolution (MBConv) units, along with feature expansion ratios that enable efficient computation for extracting features.
- (SE) Attention Mechanism: This mechanism dynamically calculates channel-wise dependencies to recalibrate feature maps, focusing on informative regions across the image. Hybrid-Spatial-Channel

Attention Module: It is designed to enhance both the spatial and channel-wise dependencies, thereby improving the network's localization of cancerous anomalies.

- Global Average Pooling Layer: This layer is used to extract the feature maps and shrink the dimensional space into a single dimension $1 \times 1 \times 2048$, to be passed on to the classification stages.
- Fully Connected Layer: This last layer takes the final vector and calculates class probabilities through the use of the Softmax activation function.

Feature maps are calibrated sequentially by the Hybrid Spatial-Channel Attention Module by:

1. Channel Attention, see Equation (2):

$$M_c(F) = \sigma \left(W_1 \left(W_0(\text{AvgPool}(F)) \right) + W_1 \left(W_0(\text{MaxPool}(F)) \right) \right) \quad (2)$$

Where $F \in \mathbb{R}^{C \times H \times W} \rightarrow$ input feature map, σ is the Sigmoid, and $r = 16$ reduction ratio. The channel-refined map is $F' = M_c(F) \otimes F$

2. Spatial Attention, see Equation (3):

$$M_s(F') = \sigma(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (3)$$

Where $f^{7 \times 7}$ is a 7×7 convolution layer. The final feature map is $F'' = M_s(F') \otimes F'$.

In order to enhance the feature refinement step further, the basic EfficientNet-B5 has been enhanced with the above-discussed SE attention mechanism. The spatial attention part selectively extracts spatial regions of interest, and the channel attention part selectively extracts the most discriminative channels to enhance the accuracy of the overall diagnosis [28]. Figure 2 shows the architecture of EfficientNet-B5 [4].

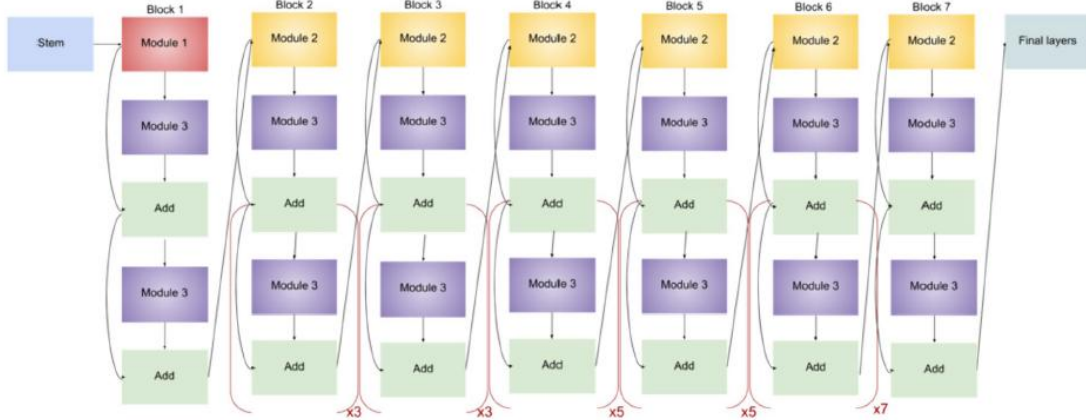


Figure 2: The Architecture of EfficientNet-B5.

3.2.2. CNN2: DENSENET-201 ARCHITECTURE

The model employs dense connectivity, where each layer is connected to all the previous layers, enabling maximum feature reuse and more direct gradient flow, which is essential for a finer level of detail extraction at multiple layers for accurate lesion analysis.

The output of each layer is fed into all the outputs of the previous layer and is processed with a given "growth rate" parameter specific to that layer. The number of parameters hence increases with every layer, and fewer parameters will give better performance [29]. The DenseNet Family is a series of dense networks; DenseNet-201 is one of them with 201 layers, which are condensed networks for constructing simple and efficient models. It is highly performing because it is possible to directly access the feature maps of all previously used layers [30]. It is architecturally distinct, incorporating "skip connections" that enable gradients (measured errors) to pass through the network readily, endowing it with a very strong ability to learn deep structural hierarchies. DenseNet-201 is a stabilizer. This way, even if the EfficientNet-B5 is

looking at a small detail, the DenseNet-201 is still able to provide the global context of the tissue sample. Figure 3 shows the architecture for DenseNet-201[5].

The workflow starts with the extraction of features, where **CNN1andCNN2** independently produce 2048×1 feature vectors that are H-dimensional embed for the input image. These attributes are then combined into a single 4096×1 representation to take advantage of complementary features of both architectures. This optimization step improves model accuracy by identifying important patterns in the oral cancer lesions and discarding the irrelevant background information. The final architecture stage consists of a fully connected layer that maps the refined feature vector to 3×1 for the three categories of prediction: normal, pre-cancer, and cancer(malignant). The output probabilities are distributed over the ranges of the output neurons so that their sum is 1, to enable clear decision-making. The Softmax activation function is used to get deterministic classification by spreading the output probabilities between the neurons, which combines feature fusion and classification, significantly improving the accuracy of diagnosis. The simultaneous extraction of both CNN networks' spatial, textural, and hierarchical features enables the system to achieve highly robust and accurate identification of oral cancer. However, hyperparameter tuning is critical to get the best training performance out of these types of ensemble deep learning.

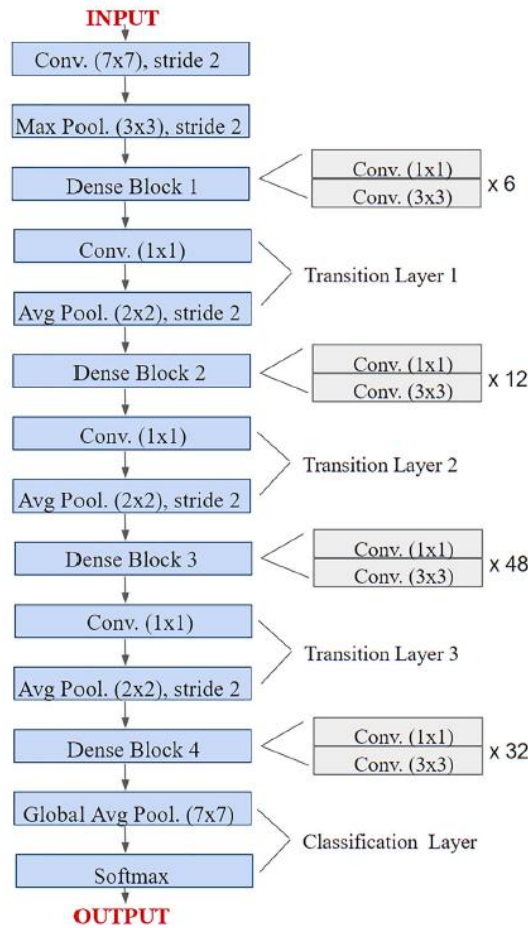


Figure 3: DenseNet-201 architecture.

3.2.3 HYPERPARAMETER OPTIMIZATION (WOA)

The (WOA) is a nature-inspired Metaheuristic optimization algorithm based on an interesting hunting method of the humpback whale. The feeding technique, called "bubble-net," involves swimming in unique spirals or circles to surround prey. Exploiting the current knowledge and searching for new knowledge in the solution space is achieved through adaptive mechanisms, and this makes WOA highly apt for the hyperparameter optimization in DL. The mathematical modelling of WOA consists of three parts: Encircling Prey, Bubble-Net Attack (exploitation), and Searching for Prey (exploration) [6], [31].

1. Encircling Prey

Humpback whales are able to locate their feed and surround them. WOA assumes that the current best candidate solution is the target prey [31]. The other search agents (whales) update their positions towards the best search agent using Equations (4) & (5):

$$D = |C \cdot X^*(t) - X(t)| \quad (4)$$

$$X(t+1) = X^*(t) - A \cdot D \quad (5)$$

Where:

- (t) refers to the current iteration.
- $X^*(t)$ is the position vector of the best solution obtained so far.
- $X(t)$ is the position vector of the current solution.
- A & D are coefficient vectors calculated as follows:

$$A = 2a \cdot r - a$$

$$C = 2 \cdot r$$

Components of a decrease linearly from **2 to 0** over a sequence of iterations, and $\vec{r}$ is a random vector in $[0, 1]$.

2. Bubble-Net Attacking Mechanism (Exploitation Phase)

Two approaches are designed to model mathematically the bubble-net behavior:

- Shrinking Encircling Mechanism: Set by making a smaller in the equation for A .
- Spiral Position Updating: Implements the “helix” motion of humpback whales. Equation (6) represented the distance between the whale and the prey:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (6)$$

Where $D' = |X^*(t) - X(t)|$ indicates the distance of the i -th whale to the prey (best solution), b is a constant for defining the shape of the logarithmic spiral, and l is a random number in $[-1, 1]$.

At the same time whales circle and move along a spiral path to the prey, which is moving in a shrinking circle. Whales swim around prey in a shrinking circle and along a spiral path at the same time [6], [31]. Assuming a probability of 50% ($p \leq 0.5$) to select which mechanism to model, either choice is made using Equation (7).

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & \text{if } p < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (7)$$

3. Search for Prey (Exploration Phase)

A similar approach based on the variation of the A vector can be used to search for prey (exploration). $|A| \geq 1$, the search agents will do a random search instead of the best whale found so far. This mechanism ensures that the algorithm will not get stuck at the local optimum and will do the global search by Equations (8) & (9):

$$D = |C \cdot X_{\text{rand}} - X| \quad (8)$$

$$X(t+1) = X_{\text{rand}} - A \cdot D \quad (9)$$

Where X_{rand} is a random position vector from a random whale from the population

3.2.4 OPTIMIZING THE MODEL USING WOA

The hyperparameters in the optimized hybrid EfficientNet-B5+ DenseNet-201 model are crucial in controlling the model's behavior and stability. The Learning Rate (LR) determines the update speed of the weight and thus also the ability to catch the optimal solutions, whereas the Batch Size (BS) determines the speed of the data and thus also the stability of the gradient and the amount of required memory, the Dropout Rate (DR) and Weight Decay (WD) are used as regularization mechanisms to stop overfitting and the model will be able to generalize well to classify the unseen oral cancer images. In this iterative process, configurations with lower accuracy and higher losses are discarded, whereas the ones with higher accuracy and lower losses are fine-tuned. The optimization is iterated until a stopping criterion is satisfied, e.g. the maximum number of iterations was reached. The best set of hyperparameters is then used to train the final ensemble model.

The algorithm starts with a population of diverse candidate solutions to cover the search space. The fitness score using an objective function (Entropy Loss ($\mathcal{L}_{\text{WCE}}$)). Fitness better configurations are kept as reference points (X^*) for the next rounds. Table 2 displays the parameters that were set in WOA.

Table 2: Optimized Hyperparameters for WOA

Hyperparameter	Search Range	Optimized value
Population Size	30 Whales	N/A
Maximum Iteration	50	N/A
Fitness function	validation Weighted Cross – Entropy Loss ($\mathcal{L}_{WCE}$)	N/A
LR	$[10^{-5}, 10^{-3}]$	0.0002
BS	$\{16, 32, 64\}$	32
DR	$[0.2, 0.5]$	0.35
WD	$[10^{-5}, 10^{-2}]$	0.0001

3.2.5 OVERFITTING MITIGATION STRATEGIES

The following mechanisms were incorporated in the training pipeline in order to strictly avoid overfitting and make sure the model is generalizable for unseen clinical data:

- Stochastic Data Augmentation: Heavy augmentation pipeline used at train time, including Random Resized Cropping between 80% to 100% scaling, Random Horizontal/Vertical Flipping, and Random Rotations (up to 30 degrees). This makes the model detect anomalies in cells, irrespective of their direction or smallness.
- To reduce overfitting, the dropout rate (0.3) was used in the ensemble's fusion classifier. The model learns to randomly deactivate 30% of the neurons in the network during each training iteration, to make the model learn a more robust, distributed representation of the pathology that can't be captured by any single feature map.
- Weight Decay (L2 Regularization): This treats large values of network weights as a penalty, smoothing decision boundaries and avoiding becoming too sure on spurious, sample-specific information.
-

3.2.6 THE PROCESS FOR WOA OPTIMIZATION

To perform this process, see Algorithm 1.

Step 1: Initialize the parameters of WOA (population size, max iterations number, constants b , etc.).

Step 2: Randomly set the parameters of the whale position, (LR, BS, DR, WD).

Step 3: Train an EfficientNet-B5 + DenseNet-201 for each of the candidate hyperparameters.

Step 4: Calculate the classification loss (fitness value) for each search and identify the best solution (X^*).

Step 5: Modify $\rightarrow$ parameters a, A, C, l , and p , Equations (4 – 9).

Step 6: Update $\rightarrow$ position of each search agent:

- **If ($p < 0.5$):**
 - If ($|A| < 1$), move to the new position to the Encircling Prey mechanism.
 - If ($|A| \geq 1$), randomly choose one search agent (X_{rand}) and update the position using the Search for Prey (Exploration) mechanism.
- **If ($p \geq 0.5$):** Update the position using the Spiral Position Updating (Exploitation) mechanism.

Step 7: Train the ensemble model on the new hyperparameter

Step 8: Compute the validation loss, check the fitness, and update the best-known solution (X^*) if there is improvement.

Step 9: Repeat steps 5-8 until the stopping condition (max-iter / min-loss-change).

Step 10: Find the best hyperparameters from WOA are used to train the final ensemble model.

Step 11: Calculate the model's performance using standard *metrics* (Accuracy, Precision, Recall, and $F1 - score$).

3.2.7 EVALUATION THE MODEL AND FRAMEWORK ARCHITECTURE

To test the proposed method, we employed a five-fold cross-validation. The data was randomly divided into five equal parts. In each iteration, four folds were used for training, and one for validation. Five folding were done to achieve a one-time validation set for each folding. All five folds were used in the performance and an average value was used to get a stable and strong measure in which the fluctuation due to data partitioning is minimized. A two different CNN parallel feature extraction is used:

1. CNN 1 (EfficientNet-B5): Combines specialized Conv2D layers, MBConv blocks, Squeeze Excitation (SE) modules and hybrid attention mechanisms to learn fine-grained patterns. CNN1 is unique in the identification of fine-grained cellular chaos as in differentiation between dysplastic and non-dysplastic leukoplakia, nuclear pleomorphism.

2. CNN 2 (DenseNet-201): Utilizes dense connectivity patterns where each layer receives direct inputs from all preceding layers, optimizing feature reuse and gradient flow. CNN2 is utilized to capture broader structural patterns, such as the depth of invasion and keratin pearl formation.

Both architectures learn feature maps that are then fused to make use of complementary spatial and contextual information. Features Fusion is fed into a Softmax classifier which classifies the images as (normal, pre-cancer, and cancer(malignant)). Figure 4 shows the proposed approach.

4.RESULTS AND STATISTICAL VALIDATION

4.1 CROSS-VALIDATION

To ensure robustness and avoid overfitting on a particular train-test split, a 5-fold stratified cross-validation protocol was used to rigorously validate the model. The dataset was dynamically split into 5 different sub-sets, where all the 7 classes were present in the same ratio in each sub-set. Each time the model was initialized and trained from scratch.

K-Fold Analysis: The standard deviation after the analysis of the five folds was found to be very low ($\pm 1.32\%$), which indicates that the numbers are consistent and there is procedural consistency in data partitions.

4.2 STATISTICAL PERFORMANCE

The experimental evaluation was conducted with a two-stage protocol: (1) 5-fold cross validation of the combined training/validation dataset was performed to assess model stability, and (2) evaluation of the model on a held-out test set (10% of the total samples). The finalized 7-class ensemble model was tested on a completely unseen test partition to prove that this model generalizes to real-world data. The Overall Test Accuracy of the model was 90.35%. The same test-set accuracy of 91.71% is close to the best validation accuracy that was obtained during the 50-epoch training cycle (Epoch 42). Results from minimal degradation between validation and test accuracy empirically validate anti-overfitting techniques (stochastic augmentation, dropout, and weight decay) and validate model high diagnostic reliability in a simulated clinical setting. The performance metrics of the model suggest a very stable and ready-to-use system.

4.2.1 CONFIDENCE INTERVAL ANALYSIS

95% Confidence Interval (CI): The 95% CI for the model's accuracy is [88.71%, 91.99%]. The tight interval ensures a high level of reliability in actual clinical settings [32] by Equation (10).

$$CI = \hat{p} \pm Z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (10)$$

Where:

- $\hat{p}$ (Sample Proportion / Accuracy) = 0.9035
- (Z – score for 95% confidence) = 1.96
- n (sample size) = 1244

1- Compute the Standard-Error (SErr)

$$SErr = \sqrt{\frac{0.9035 \times (1 - 0.9035)}{1244}}$$

$$SErr = \sqrt{0.00007008}$$

$$SErr \approx 0.00837$$

2- Compute the Margin of Error (ME)

$$ME = 1.96 \times 0.00837$$

$$ME \approx 0.0164 \text{ (or } 1.64\%)$$

3- Use the (ME) in the Accuracy

$$CI = 90.35\% \pm 1.64\%$$

The 95% CI for the model's [88.71%, 91.99%], so narrow ($\pm 1.64\%$) that it provides a high level of statistical reliability in real-life clinical use.

4.2.2 DETAILS OF ANALYSIS AND CONFUSION MATRIX

The following is a detailed Class-wise analysis and confusion matrix, as follows [32].

Accuracy: Measures the overall level of correct predictions, as a percentage of the total number of cases studied using Equation (11).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

Where;

True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN)

Recall (also known as Sensitivity or True Positive Rate - TPR): Measures the percentage of the actual positive cases (e.g., actual cancer/OSMF) that was correctly identified by the model. This is an essential part of medical screening because no disease will be missed using Equation (12).

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

Specificity (also known as True Negative Rate - TNR) Measures the proportion of actual negative cases (e.g., healthy tissue) that were correctly identified as negative, Equation (13).

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

Precision (also called Positive Predictive Value - PPV) is the percentage of positive predictions that were accurate using Equation (14).

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

F1-Score: Mean of the harmonic of Precision and Recall. It offers a single score that takes into account both aspects, particularly when there are imbalanced datasets by Equation (15).

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (15)$$

Negative Predictive Value (NPV) is the percentage of negative predictions which are truly negative using Equation (16).

$$NPV = \frac{TN}{TN+FN} \quad (16)$$

False Positive Rate (FPR): Percent of negative cases that were falsely identified as positive (False Alarms) by Equation (17).

$$FPR = \frac{FP}{FP+TN} = 1 - Specificity \quad (17)$$

False Negative Rate (FNR): The percentage of those who do have a positive test result who have it labeled as negative (Missed Diagnoses) using Equation (18).

$$FNR = \frac{FN}{TP+FN} = 1 - Recall \quad (18)$$

False Discovery Rate (FDR): The percentage of true positives among the number of positive predictions by Equation (19).

$$FDR = \frac{FP}{TP+FP} = 1 - Precision \quad (19)$$

Matthews Correlation Coefficient (MCC): the most robust classification assessment functions particularly for imbalanced data see Equation (20). It considers all four elements of the confusion matrix and returns a number between -1 and +1, with +1 meaning perfectly, 0 meaning not predict better than random guess and -1 meaning totally in disagreement. Figure 5 shows the Confusion-Matrix.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (20)$$

In Table 3 shows the performance of the Diagnostic Models in different classes. Tables 4 and 5 display the performance metrics for the proposed method with different measures.

The proposed approach showed good diagnostic performance for OSMF (Accuracy: 98.39%, MCC: 0.9453) and Normal tissue (Accuracy: 99.12%, MCC: 0.9417).

The very high sensitivity of 99.82% of the Normal class is noteworthy from a clinical perspective, as it means that the system makes few false positive diagnoses, which can reduce the risk of patient anxiety or workload. This pre-cancerous condition mustn't be missed when performed in screening, as it is very sensitive (99.05%).

From Tables 3-5, we can see the following: Leukoplakia without Dysplasia (F1-score down to 70.87%, and the MCC down to 0.695, both the lowest among all classes. It is immediately clear from these figures how difficult it can be to differentiate between a benign leukoplakic lesion and healthy mucosa or early dysplastic changes in tissue. There is considerable overlap in the appearance of these under the microscope.

The accuracy of this category is 97.03%, while the specificity is 98.06%. But the precision is 66.18%. So, it was the model that pointed out several examples, including Leukoplakia without Dysplasia, which was not an appropriate classification. The sensitivity is 76.27%, which still means that a number of true cases of benign leukoplakia were missed. This lack of clarity is understandable clinically. Leukoplakia without dysplasia may exhibit only slight changes – no significant changes – and can be confused with normal mucosa or early dysplastic changes. Hence, the most difficult diagnostic challenge for the model is with the class.

The study could not be compared with previous ones due to different datasets. Furthermore, the datasets used by previous studies had fewer images than the dataset used in this research, as shown in Section 2, which could make comparisons of performance results difficult.

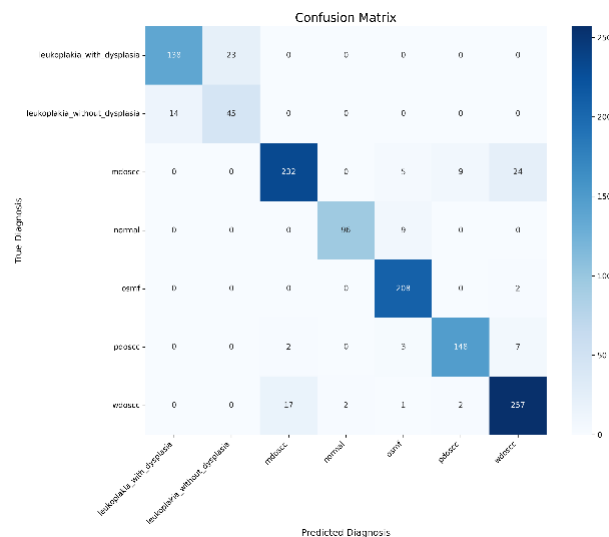


Figure 4: Confusion-Matrix

Table 3: Performance for the Diagnosis classes

Class (Diagnosis)	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)
Leukoplakia_With_Dysplasia	138	14	23	1069
Leukoplakia_Without_Dysplasia	45	23	14	1162
MDOSCC	232	19	38	955
NORMAL	96	2	9	1137
OSMF	208	18	2	1016
PDOSCC	148	11	12	1073
WDOSCC	257	33	22	932

Table 4: The performance metrics for the proposed method for (Accuracy, PPV, Recall, TNR, and F1-Score)

Class (Diagnosis)	Accuracy	Precision (PPV)	Recall (Sensitivity)	Specificity (TNR)	F1-Score
Leukoplakia_with_Dysplasia	97.03%	90.79%	85.71%	98.71%	88.18%
Leukoplakia_without_Dysplasia	97.03%	66.18%	76.27%	98.06%	70.87%
MDOSCC	95.42%	92.43%	85.93%	98.05%	89.06%
NORMAL	99.12%	97.96%	91.43%	99.82%	94.58%
OSMF	98.39%	92.04%	99.05%	98.26%	95.41%
PDOSCC	98.15%	93.08%	92.50%	98.99%	92.79%
WDOSCC	95.58%	88.62%	92.11%	96.58%	90.33%

Table 5: The performance of the proposed method for (VPV, FPR, FNR, FDR, and MCC)

Class (Diagnosis)	Negative Predictive Value (NPV)	False Positive Rate (FPR)	False Negative Rate (FNR)	False Discovery Rate (FDR)	Matthews Correlation Coefficient (MCC)
Leukoplakia_with_Dysplasia	97.89%	1.29%	14.29%	9.21%	0.8653
Leukoplakia_without_Dysplasia	98.81%	1.94%	23.73%	33.82%	0.695
MDOSCC	96.17%	1.95%	14.07%	7.57%	0.8626
NORMAL	99.21%	0.18%	8.57%	2.04%	0.9417
OSMF	99.80%	1.74%	0.95%	7.96%	0.9453
PDOSCC	98.89%	1.01%	7.50%	6.92%	0.9173
WDOSCC	97.69%	3.42%	7.89%	11.38%	0.875

4.2.3 CARCINOMA CLASSIFICATION

The carcinoma subclasses—WDOSCC, MDOSCC, and PDOSCC—the model achieved good results throughout. The highest accuracy was seen in the PDOSCC (98.15%) and the MCC in the PDOSCC was (0.9173). It is a good predictor that's a good sign. The accuracy of WDOSCC and MDOSCC were also good, with 95.58% and 95.42% respectively. MCC values (0.875 and 0.8626) followed closely.

The specificity was outstanding in each of the three carcinoma types with values always exceeding 96%. That implies that FPs were uncommon, and exactly what is necessary to prevent that benign or pre-malignant lesions are incorrectly diagnosed as cancer.

High sensitivity scores were also recorded with the carcinoma classes ranging, from 85.93% to 92.50%. The model thus often missed identifying cases with malignancy; a critical condition in histopathology, where all cases of cancer need to be identified.

4.2.4 MEAN AND STANDARD DEVIATION

Table 6 displays the mean and Standard Deviation (SD). The SD ($\pm 1.32\%$) across the five folds mathematically proves the stability of the proposed model. It shows that no matter what the disease is, features of oral pathology are not memorizing the training data. The consistent mean accuracy of 87.37% on a very complex 7-class medical diagnostic task shows good production-ready generalization.

Table 6: Mean and SD

Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean	($\pm$ SD)
88.33	86.87	85.01	88.53	88.12	87.37	($\pm 1.32\%$)

4.2.5 TRAINING AND VALIDATION LOSS

Figure 5 Shows the loss curves for Training and validation with stable convergence and no divergence. This was partly the reason there was no significant separation between curves, thus diminishing risk and consequently enhancing the system.

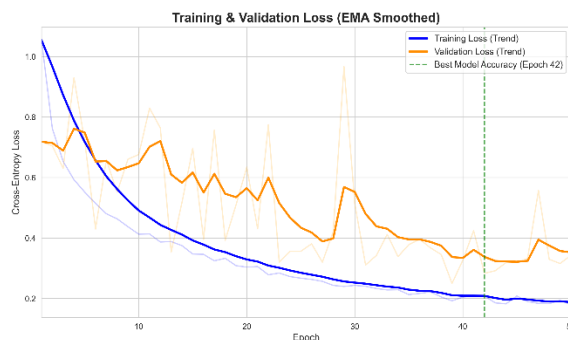


Figure 5: The loss curve for the proposed method.

4.3 ABLATION STUDY

This study quantified each proposed component's contribution. Despite the simplicity of fusion, the fully optimized model yielded the highest accuracy (90.35%), proving the necessity of the proposed modules. Table 7 compares the performance of each component on the test set.

Table 7: The performance of comparison of individual components on the test set

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNet-B5 (Baseline)	86.40	85.12	84.80	84.96
DenseNet-201 (Baseline)	85.10	84.30	83.90	84.10
(Proposed + WOA)	90.35	89.80	89.50	89.65

5. CONCLUSION AND RECOMMENDATIONS.

This study presents an optimized DL model for oral cancer detection that consists of two models, Efficient-Net-B5 and DenseNet-201, optimized using the (WOA). The proposed method is effective in addressing the drawbacks of CNN architectures by combining the feature extraction of two CNNs. The model functioned successfully, achieving a mean cross-validation accuracy equal to 87.37% ($\pm 1.32\%$) and achieving 90.35% accuracy on the test set. Matthews Correlation Coefficient (MCC) values demonstrate that the model effectively distinguishes these categories—despite difficult circumstances, such as differentiating carcinoma subtypes or simply normal tissue on the ORCHID and UFES-NDB datasets, which is a good performance in diagnosing normal, pre-cancer, and cancer. The proposed method achieved high precision, recall, and F1-score for all categories. The 95% Confidence Interval for the model's accuracy is [88.71%, 91.99%]. A major advantage of the methodology is that it can overcome the drawback of using individual CNN models, combining them to get a holistic high decision in processing. The hyperparameter optimization based on WOA further enhances the generalization power, effectively improves the performance of the model, and reduces the phenomenon of overfitting by adjusting the key parameters like LR, BS, DR, and WD. Although the proposed framework is validated to give high diagnostic parameters, it must be validated in clinical deployment for multi-center clinical cohorts.

In recommendations, integrate more than one model and metaheuristic algorithm to come up with radical solutions in cancer treatment. Further extensions to the clinical use of this approach would involve investigating the possibilities of using other CNN models and multi-modal imaging modalities, as well as real-time diagnostic applications.

REFERENCES

- [1] B. Ilhan, K. Lin, P. Guneri, and P. Wilder-Smith, "Improving Oral Cancer Outcomes with Imaging and Artificial Intelligence," *J. Dent. Res.*, vol. 99, no. 3, pp. 241–248, 2020, doi: 10.1177/0022034520902128.
- [2] N. Hoda, M. Aastha, B. Akshay, A., and S. K., S., "Artificial Intelligence Based Assessment and Application of Imaging Techniques for Early Diagnosis in Oral Cancers," *Int. Surg. J.*, vol. 11, no. 2, pp. 318–22, 2024, doi: 10.18203/2349-2902.isj20.
- [3] R. Dharani and K. Danesh, "Optimized deep learning ensemble for accurate oral cancer detection using CNNs and metaheuristic tuning," *Intell. Med.*, vol. 11, p. 100258, 2025, doi: 10.1016/j.ibmed.2025.100258.
- [4] D. Sezer, Y. Esra, Kavalcı, A. Kemal, and A. Erdinç, "Performance of vision transformer and swin transformer models for lemon quality classification in fruit juice factories," *Eur. Food Res. Technol.*, vol. 250, no. 9, pp. 2291–2302, 2024, doi: 10.1007/s00217-024-04537-5.
- [5] J. Aayush, G. Neha, S. Dilbag, K. Vijay, and K. Manjit, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *J. Biomol. Struct. Dyn.*, vol. 39, no. 15, pp. 5682–

5689, 2021, doi: 10.1080/07391102.2020.1788642.

- [6] Z. Liang, S. Ting, and D. Zuohua, "A Novel Improved Whale Optimization Algorithm for Global Optimization and Engineering Applications," *Mathematics*, vol. 12, no. 5, p. 636, 2024, doi: 10.3390/math12050636.
- [7] N. Ali, Fahem, A. Ahmad, Shaker, M. Mohammed, Ibrahim, and R. Khan, Ihtiram, "Feature Fusion for Improved Skin Cancer Diagnosis Using Support Vector Machines," *Wasit J. Comput. Math. Sci.*, vol. 4, no. 1, pp. 53–58, 2025, doi: 10.31185/wjcms.382.
- [8] N. Ahmed, Adil, I. Mohammed, Salah, S. Mustafa, Muslih, A.-K. Kibriya, A. Hiba, Rashid, and A.-A. Mohammed, "Deep Learning Algorithm for Lung Cancer Detection Using EfficientNet-B3," *Wasit J. Comput. Math. Sci.*, vol. 2, no. 4, pp. 68–76, 2023, doi: 10.31185/wjcms.209.
- [9] G. Akilandasowmya, G. Nirmaladevi, S. Suganthi, and A. Aishwariya, "Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification," *Biomed. Signal Process. Control*, vol. 88, no. part c, p. 105306, 2024, doi: 10.1016/j.bspc.2023.105306.
- [10] R. DharaniDanesh and K. Danesh, "Oral Cancer Segmentation and Identification System Based on Histopathological Images using MaskMeanShiftCNN and SV-OnionNet," *Intell. Med.*, vol. 10, no. 2, p. 100185, 2024, doi: 10.1016/j.ibmed.
- [11] P. Sunil, Kumar and R. Harikumar, "Performance analysis of linear layer neural networks for oral cancer classification," in *2017 6th ICT International Student Project Conference (ICT-ISPC)*, Johor, Malaysia, 2017, pp. 1–4. doi: 10.1109/ICT-ISPC.2017.8075357.
- [12] J. Fahed, A. Omar, M. Dimitrios, A. M. Samara, S. Yusser, and H. Yazan, "A Novel Lightweight Deep Convolutional Neural Network for early detection of oral cancer," *Oral Dis.*, vol. 28, no. 4, pp. 1123–1130, 2022, doi: 10.1111/odi.13825.
- [13] S. Camalan et al., "Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results," *Cancers (Basel)*, vol. 13, no. 6, p. 1291, 2021, doi: 10.3390/cancers13061291.
- [14] J. Pandia, Rajan and N. Edward, Rajan, Samuel, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *J. Cancer Res. Clin. Oncol.*, vol. 145, pp. 829–837, 2019, doi: 10.1007/s00432-018-02834-7.
- [15] Z. Hu, A. Abeer, M. Paul, P. W. C. Prasad, A. Salih, and A. Elchouemic, "Early stage oral cavity cancer detection: Anisotropic pre-processing and fuzzy C-means segmentation," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 2018, pp. 714–719. doi: 10.1109/CCWC.2018.8301673.
- [16] G. Yunus, Emre, "Ensemble deep learning framework integrating deep image features and statistical descriptors for robust tumor diagnosis," *Biomed. Signal Process. Control*, vol. 118, p. 109645, 2026, doi: 10.1016/j.bspc.2026.109645.
- [17] M. Rajakani, R. Kavitha, and B. Kannan, "Metaheuristic Optimization Based Deep Learning Model for Multispectral Image Classification," *Math Oper Res*, vol. 3, no. 2, pp. 45–58, 2023, doi: 10.21203/rs.3.rs-2731247/v1.
- [18] M. Ali, Ali and M. Mazin, Abed, "Optimized Cancer Subtype Classification and Clustering Using Cat Swarm Optimization and Support Vector Machine Approach for Multi-Omics Data," *J. Soft Comput. Data Min.*, vol. 5, no. 2, pp. 223–44, 2024, doi: Ali, Ali Mahmoud, and Mazin Abed Mohammed.
- [19] A. Mohammed, Mazin and M. Ali, Ali, "Enhanced cancer subclassification using multi-omics clustering and quantum Cat swarm optimization," *Iraqi J. Comput. Sci. Math.*, vol. 5, no. 3, pp. 552–582, 2024, doi: 10.52866/ijcsm.2024.05.03.035.
- [20] T. Zixuan et al., "A literature review of artificial intelligence (AI) for medical image segmentation: from AI and explainable AI to trustworthy AI," *Quant Imaging Med Surg.*, vol. 14, no. 12, pp. 9620–9652, 2024, doi: 10.21037/qims-24-723.
- [21] N. Chaudhary et al., "High-resolution AI image dataset for diagnosing oral submucous fibrosis and squamous cell carcinoma," *Zenodo Data*, V1, 2024. https://zenodo.org/records/12636426?utm_
- [22] R. Maria, Clara, Falcão et al., "NDB-UFES: An oral cancer and leukoplakia dataset composed of histopathological images and patient data," *Mendeley Data*, V4, 2023. <https://data.mendeley.com/datasets/bbmmm4wgr8/4>
- [23] U. Themes, "Oral mucosa: Physiological and physicochemical aspects," *Internet*, 2026. <https://basicmedicalkey.com/oral-mucosa-physiological-and-physicochemical-aspects/>
- [24] V. Thakur, S. Jassal, A. Kumar, S. Malik, P. Sharma, and S. Sahi, "A short review on OSMF: oral submucous fibrosis," *J Curr Med Res Opin*, vol. 3, no. 9, pp. 619–624, 2020, doi: 10.15520/jcmro.v3i09.337.
- [25] S. Ivan, J., "Oral cavity & oropharynx Potentially malignant & dysplasia/ Leukoplakia," *Internet*, 2021. <https://share.google/ghoCGIhyy9NOEZY2M>
- [26] D. Olinici et al., "The ultrastructural features of the premalignant oral lesions," *Rom J Morphol Embryol*, vol. 59, no. 1, pp. 243–248, 2018, [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29940634/>

- [27] M. Pereira, D. Oliveira, G. Landman, and L. Kowalski, "Histologic subtypes of oral squamous cell carcinoma: prognostic relevance," *J Can Dent Assoc*, vol. 73, no. 4, pp. 339–44, 2007, [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17484800/>
- [28] T. Mingxing and L. Quoc, V., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114. doi: 10.48550/arXiv.1905.11946.
- [29] P. Padhi and M. M. Das, "Hand gesture recognition using denseNet201- mediapipe hybrid modelling," in *International Conference on automation, computing and renewable systems (ICACRS)*, Pudukkottai, India: IEEE, 2022, pp. 995–999. doi: 10.1109/ICACRS55517.2022.10029038.
- [30] A. Abadicio et al., "Ground-level Post-disaster image classification using DenseNet201 for disaster damage assessment," in *2023 International Conference On Cyber Management And Engineering (CyMaEn)*, 2023, pp. 132–137. doi: 10.1109/CyMaEn57228.2023.10050981.
- [31] L. Maodong, X. Guanghui, L. Qiang, and C. Jie, "A chaotic strategy-based quadratic Opposition-Based Learning adaptive variable-speed whale optimization algorithm," *Math. Comput. Simul.*, vol. 193, pp. 71–99, 2022, doi: 10.1016/j.matcom.2021.10.003.
- [32] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020, doi: 10.1186/s12864-019-6413-7.