Enhanced Early Detection of Oral Squamous Cell Carcinoma via Transfer Learning and Ensemble Deep Learning on Histopathological Images

Gurjot Kaur[1], Sheifali Gupta[2], Ashraf Osman Ibrahim[3]*, Salil bharany[4]*,
Marwa Anwar Ibrahim Elghazawy[5], Hadia Abdelgader Osman[6], Ali Ahmed[7]
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India[1, 2, 4]
Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Malaysia[3]
Positive Computing Research Center, Emerging and Digital Technologies Institute[3]
Computer Department, Applied College, Northern Border University, Arar, Saudi Arabia[5, 6]
Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, 21589, Saudi Arabia[7]

***Abstract*—Oral Squamous Cell Carcinoma (OSCC) is one main kind of oral cancer; early diagnosis is rather important to increase patient survival chances. This study investigates the application of advanced deep learning techniques including transfer learning and ensemble learning to increase the accuracy of oral squamous cell cancer (OSCC) diagnosis using histopathological image analysis. Two transfer learning models, EfficientNetB3 and ResNet50, support the suggested method to extract suitable features from the histopathological images. Both models permit fine-tuning to improve their classification accuracy. On tests taken after the initial training, the EfficientNetB3 model scored 96.15%. Later on, training ResNet50 yielded a test accuracy of 91.40%. Weighted voting merged several models into an ensemble model designed to maximize the strengths of each network. With a test accuracy of 98.59% and a training accuracy of 99.34%, the ensemble model showed notably higher performance than the values obtained by the individual models. Divided into OSCC and standard categories, the collection has 5,192 extremely well-resolved images. The images were used to create training, validation, and testing sets. We used this method to consistently evaluate the model's performance and reduce overfitting. Furthermore, the ensemble model proved to be quite accurate with recall and F1 scoring, thereby proving its capacity to routinely identify OSCC images. Both groups produced ROC curves, and the area under the curve (AUC) demonstrated excellent model performance. Transfer learning and ensemble learning are used together in this study to show that OSCC can be found early and consistently in histopathology images. The findings reveal that the recommended strategy could be a consistent tool to assist pathologists in the precise and timely detection of OSCC, thereby improving patient treatment and outcomes.**

Keywords*—Oral Squamous Cell Carcinoma (OSCC); histopathology images; transfer learning; ensemble learning; EfficientNetB3; ResNet50; deep learning; cancer detection; medical image analysis

I. Introduction

Oral Squamous Cell Carcinoma (OSCC) [1] is one of the most prevalent forms of cancer worldwide, accounting for over 90% of oral cancers. Usually affecting the squamous cells guarding the oral cavity, the disorder arises in malignant tumors that, if not found and treated quickly, can spread to other parts of the body. According to the World Health Organization (WHO), OSCC ranks in the top ten most common malignancies globally, with especially high prevalence in South Asia, Southeast Asia, and parts of Europe. OSCC has a significant global influence considering an estimated 300,000 new cases and around 145,000 deaths annually. The OSCC diagnosis primarily dictates its prognosis; early-stage diagnosis considerably increases the chances of successful therapy and survival [2]. Among the most often occurring risk factors of OSCC, a multifactorial etiopathogenesis including tobacco usage, alcohol intake, betel quid chewing, and human papillomavirus (HPV) infection stands. Growing emphasis on the importance of molecular changes and genetic predispositions in the evolution of OSCC has been observed in recent years. Although OSCC's molecular biology is now well known, the main diagnostic tool that remains is histopathological investigation of tissue biopsies. This method is formed by a pathologist's microscopic examination of tissue samples in search of malignant cells. Nevertheless, this standard diagnostic approach has important disadvantages including inter-observer variability, the time-consuming character of the operation, and the likelihood of misinterpretation coming from the subjective interpretation of histological features [3]. Early and correct OSCC detection is highly significant, thus the scientific and medical industries have worked hard to develop automated diagnostic tools that can help pathologists and minimize diagnosis mistakes. In this sense, applying machine learning (ML) and artificial intelligence (AI) in medical imaging has become an interesting road for innovation. The development of digital pathology and advances in image analysis methods have presented opportunities for the design of AI-driven diagnostic systems capable of very accurate and fast histological image interpretation [4]. OSCC presents varied clinical signs based on tumor sizes and location, such as erythroplakia, leukoplakia, non-healing ulcers, lumps, or persistent sores. Other symptoms include voice changes, dysphagia, pain, numbness, bleeding, loose teeth, and, in advanced stages, neck swelling due to lymph node metastasis. The etiology of OSCC is unclear, though major risk factors include tobacco use and alcohol consumption, which together significantly increase cancer risk [5]. High-risk human

papillomavirus (HPV), particularly HPV-16, is also a major factor, especially for oropharyngeal cancers [6]. Additional risks include poor oral hygiene, environmental toxins, chronic trauma, a weakened immune system, and genetic predispositions. Understanding these factors is critical for early detection and prevention programs aimed at reducing OSCC incidence and mortality [7]. Treatments for OSCC combine a multimodal approach tailored to the tumor's stage, location, and patient health. Major treatments include surgery, radiation, and chemotherapy, often used together for optimal results [8]. Surgery, the primary treatment for localized OSCC, aims to remove the tumor with clear margins to minimize recurrence. Reconstructive surgery may follow to restore function and appearance after partial or full removal of affected parts, such as the tongue or jawbone. Radiation therapy, typically external beam, is used as a supplement to surgery or as a main treatment when surgery isn't viable [9]. Chemotherapy, often paired with radiation (chemoradiation), is reserved for advanced or metastatic cases and patients with recurrent OSCC. Targeted therapies and immunotherapies offer appealing adjunct treatments by focusing on molecules linked to tumor growth [10]. Supportive care, including nutrition, pain management, and speech therapy, helps maintain quality of life. Multidisciplinary tumor boards optimize treatment by balancing patient-specific factors with evidence-based guidelines and chosen treatment methods. In oncology, where early detection significantly influences patient prognosis, the diagnosis of Oral Squamous Cell Carcinoma (OSCC) provides a huge challenge. Sometimes subjectivity, variability in interpretation, and the need for specialist knowledge limit conventional diagnosis approaches such as histological study and clinical examination, therefore causing inconsistent and delayed diagnosis [11]. Although machine learning has advanced, present models for OSCC detection from histopathology images often depend on single-model architectures, which might not adequately reflect the complex and heterogeneous character of OSCC. Furthermore, confusing accurate detection results in variations in image quality and staining methods. More strong, accurate, and dependable diagnostics tools that can combine several models to improve detection possibilities are desperately needed. This work fills in this need by looking at the application of transfer learning and ensemble learning approaches to create a better, ensemble-based model for early and accurate OSCC identification [12]. The main goal of this work is to provide a strong and accurate strategy based on advanced deep learning [13] approaches more especially, transfer learning [14] and ensemble learning [15] for the detection of Oral Squamous Cell Carcinoma (OSCC). The work intends to extract discriminative characteristics from histopathology images using the strengths of two state-of-the-art pre-trained models, EfficientNetB3 and ResNet50. The study aims to raise the general classification accuracy and dependability of OSCC detection by optimizing these models and combining their outputs using an ensemble learning method. With an eye on measurements including accuracy, precision, recall, F1 score, and AUC-ROC, the study also seeks to evaluate the ensemble model against individual models. The ultimate aim is to develop an ensemble model significantly above present methods, thus providing a more consistent diagnostic tool for early OSCC diagnosis and aiding in improving patient outcomes.

The primary contributions of this research are:

- This work presents a new ensemble learning method based on two state-of-the-art pre-trained models, EfficientNetB3 and ResNet50, for Oral Squamous Cell Carcinoma (OSCC) identification from histopathology images.
- Specifically, with OSCC identification, the work shows medical image analysis transfer learning performance. The work uses transfer learning to overcome the difficulties provided by insufficient labeled data thus improving the generalizing power of the model by fine-tuning pre-trained models on histomorphology images.
- The study evaluates the proposed ensemble model exhaustively using significant performance measures including accuracy, precision, recall, F1 score, and AUC-ROC. Since the ensemble model trumps individual models, the results demonstrate that it offers a more consistent diagnosis tool for early OSCC identification. Development of the field of medical image analysis and improvement of oncology clinical results depend on this study.

II. Literature Review

The detection of OSCC has become even more critical considering the substantial fatality rate connected with late-stage diagnosis. Deep learning, notably in CNNs and Transfer Learning, has lately shown promise in boosting the accuracy of OSCC detection from histopathological images.

Fanizzi et al. (2024) [16] used an explainable CNN model to examine oropharyngeal squamous cell cancer, training on CT images of 499 OPSCC patients with an independent test set of 92. Using an Inception-V3 architecture, they achieved a 73.50% AUC, highlighting tumor locations via Grad-CAM and emphasizing CNN relevance in therapy.

Paramasivam et al. (2024) [17] used deep learning with three modified CNN architectures, including DENSENET-121 variants, to diagnose OSCC in 5,492 histopathological images, achieving 97.03% accuracy. The study highlighted AI's potential to reduce human errors and improve diagnostic accuracy, addressing the limitations of traditional mouth cancer diagnosis methods.

Weber et al. (2024) [18] studied Stimulated Raman Histology (SRH) with deep learning for Oral Squamous Cell Carcinoma categorization. Using a VGG19 CNN, the model achieved balanced accuracies of 0.90 (SRS) and 0.87 (SRH) from SRH images transformed to resemble H&E sections, highlighting AI's efficiency in intraoperative OSCC detection.

Albalowitz et al. (2024) [19] developed a deep learning model using EfficientNetB3 for OSCC diagnosis, achieving 99% accuracy, precision, recall, and F1-score. The study utilized 1,224 histograms from 230 individuals, leveraging data augmentation, regularization, and optimization. This work demonstrates the potential of deep learning in improving OSCC diagnostic accuracy.

Kumar et al. (2024) [20] examined deep learning for early OSCC identification from histopathology images. Using an

enhanced Inception-Resnet-V2 CNN model, the study achieved 91.78% accuracy in distinguishing benign from malignant biopsy images. The paper highlights deep learning's potential in automating OSCC diagnosis and improving early clinical interventions.

Mishra et al. (2024) [21] explored CNNs for early OSCC identification, achieving 98.49% training accuracy, 86.89% validation accuracy, and 89.37% testing accuracy. The model's F1-score of 0.89 for both classes highlights CNNs' potential for improving OSCC screening protocols and patient outcomes, thus reducing fatality rates linked to late diagnosis.

Siddique et al. (2024) [22] used biopsy-derived histological images to explore oral cancer detection techniques, emphasizing image pre-processing's role in enhancing CNN model performance. Models including VGG16, VGG19, InceptionV3, AlexNet, and ResNet50 achieved respective accuracies of 84%, 82%, 67%, 76%, and 42%, showcasing deep learning's potential in cancer diagnosis.

SMira et al. (2024) [23] investigated early oral cancer detection using deep learning and smartphone-based imagery. They developed a resampling method and "center positioning" image-capturing technique to manage variability. The deep learning network achieved 83.0% sensitivity, 96.6% specificity, 84.3% accuracy, and 83.6% F1 score on 455 test images.

Deo et al. (2024) [24] studied the categorization of Oral Squamous Cell Carcinoma (OSCC) using a Vision Transformer (ViT) framework. The updated ViT model outperformed eight pre-trained deep learning models on a dataset of 4,946 images, achieving 97.78% accuracy, 96.72% specificity, and 98.80% sensitivity, proving ViT's efficacy on smaller datasets.

Soni et al. (2024) [25] studied early OSCC diagnosis using an improved EfficientNetB0 model and Dual Attention Network (DAN). Achieving 91.1% accuracy, 92.2% sensitivity, 91.0% specificity, and a 92.3% F1 score, EfficientNetB0 outperformed 17 pre-trained models, highlighting deep learning's clinical potential for early OSCC identification and therapy enhancement.

Deo et al. (2024) [26] developed a DL model for OSCC detection from histopathological images. Using 2D empirical wavelet transform and a ResNet50-DenseNet201 ensemble, binary classification achieved 92% accuracy. The method enhances diagnostic accuracy, reduces human error, and accelerates classification, demonstrating deep learning's potential to aid clinical decision-making.

Saraswathi et al. (2023) [27] examined AlexNet's classification of Oral Squamous Cell Cancer (OSCC) using 5,192 histopathological images. AlexNet outperformed ResNet, achieving 89% accuracy and 60% loss compared to ResNet's 78% accuracy and 82% loss. The study suggested future improvements through larger datasets, software applications, and modified AlexNet models.

Ahmad et al. (2024) [28] studied AI approaches for OSCC detection via histopathological images. Three methods were compared: Gabor+CatBoost, ResNet50+CatBoost, and a hybrid Gabor+ResNet50+CatBoost. The hybrid achieved 94.92% accuracy, 95.51% precision, and 94.9% F1 score. PCA reduced feature dimensionality, improving performance for accurate OSCC diagnosis.

Nagarajan et al. (2023) [29] developed a deep learning system to diagnose Oral Squamous Cell Carcinoma from histopathology images, using a Modified Gorilla Troops Optimizer. MobileNetV3 achieved the highest accuracy (0.89), which increased to 0.95 after optimization. This method demonstrates swarm intelligence's potential to improve OSCC detection accuracy.

Begum et al. (2023) [30] use deep learning models to automatically detect OSCC from histological images, employing transfer learning and layer modification. Among four pre-trained CNN models (NASNet Large, InceptionNet, Xception, DenseNet 201), DenseNet 201 achieved the highest accuracy of 91.25%, showcasing DL-based advancements in OSCC diagnosis and treatment.

III. Methodology

A. Dataset Distribution

This work uses a dataset [31] consisting of 5,192 histopathology images, carefully selected to identify Normal tissues and Oral Squamous Cell Carcinoma (OSCC). As Fig. 1 shows, the dataset is almost balanced with 2,698 OSCC images (52%) and 2,494 Normal images (48%). The right pie chart shows the percentage distribution and the left bar chart shows the image count in every class. This balanced set guarantees a fair training approach, therefore reducing possible categorization bias. Three subsets comprising 3,634 images for training, 779 images for validation, and 779 images for testing comprise the dataset. Using a strong and well-organized dataset, this section helps deep learning models to be developed, refined, and evaluated for effective OSCC detection.

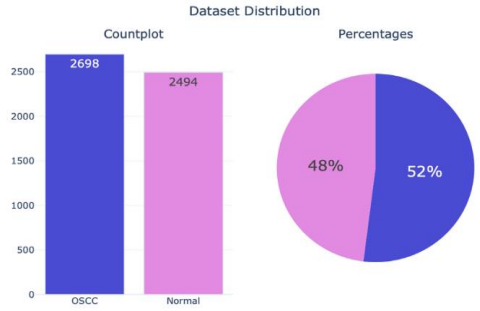

Fig. 1. Dataset distribution.

B. Input Dataset

The normal tissue samples in Fig. 2 show a well-organized structure typical of healthy oral epithelium. The cells have exactly defined borders and firmly packed layers that match their consistent size and form. The way the compartments are arranged is logical there are no obvious anomalies or deviations. There is no hyperchromatism, which would indicate non-malignant tissue; the nuclei are constant in size. Conversely, the Oral Squamous Cell Carcinoma (OSCC) samples in Fig. 3 reveal a disturbance in cellular architecture. OSCC sample cells are irregular, bigger, and show pleomorphism that is, a

great range in size and form. Commonly an indication of cancer, the hyperchromatic nature of the nuclei results from the higher DNA content, which seems darker. Furthermore, the OSCC images show abnormal keratinization and uneven stratification of the epithelium with layers that are no longer clearly defined or ordered. These pathological abnormalities are important in histopathological diagnosis since they mirror the aggressive character of OSCC, in which disorganized, malignant cells replace the normal, ordered tissue.

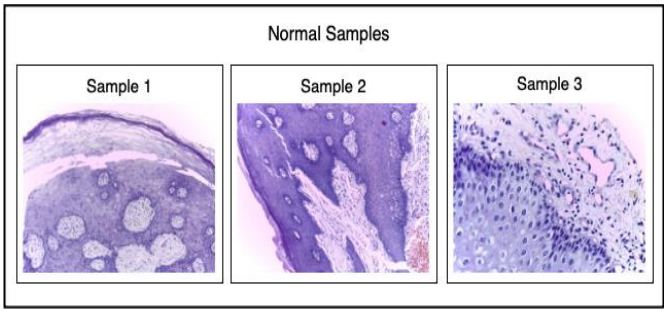

Fig. 2. Dataset samples of normal images.

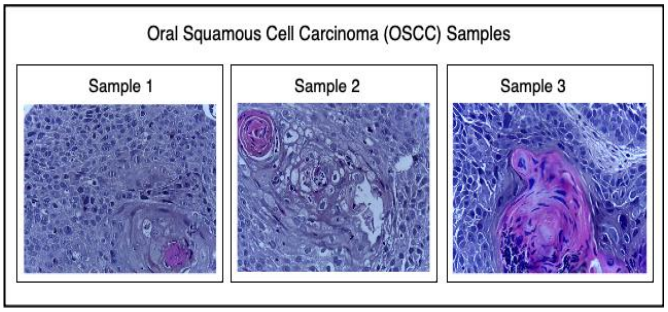

Fig. 3. Dataset samples of OSCC images.

C. Data Preprocessing and Augmentation

Preprocessing is a crucial step that greatly influences the performance of a deep-learning model developed to classify OSCC from histopathological images. Essential for getting the dataset ready for efficient training, validation, and testing are image resizing, normalizing, and augmenting preprocessing techniques [32].

1) Image resizing: Every image is reduced to a consistent dimension suitable for input into the deep learning models thereby guaranteeing consistency over the dataset. This scaling method converts every image into such size since many CNN designs demand each image to be a specific size usually (224×224). This phase guarantees that the model manages images of the same resolution, therefore facilitating efficient training and reducing computer complexity.

2) Normalization: Normalization is carried out to rescale the pixel values of the images to a standardized range, usually ranging from 0 to 1. The process of normalizing can be quantitatively expressed in (1):

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Where, X represents the original pixel values, min(X) is the minimum pixel value, and max(X) is the maximum pixel value. By normalizing the input values to a similar range, this step enhances the convergence speed and stability of the training process.

3) Augmentation: Data augmentation provides artificial unpredictability and increases the training dataset size. Another uses horizontal flipping, rotation, zooming, and shifting among other techniques. Mathematically, augmentation can be stated in (2):

$$X' = T(X) \quad (2)$$

Where X' is the augmented image and T is the transformation carried out, say rotation by a certain angle θ or scaling by a factor s. Exposing the model to several image transformations during training, this stage is vital for improving its robustness and lowering overfitting. Using their integration into the model pipeline, these preprocessing actions guarantee the well-preparedness of the dataset, so facilitating the deep learning model to achieve high accuracy in OSCC detection and generalization capability.

IV. Proposed Methodology

Fig. 4 demonstrates the proposed methodology for the detection of Oral Squamous Cell Carcinoma (OSCC). The present study is based on a combination of two pre-trained deep learning models: EfficientNetB3 and ResNet50. These models utilize transfer learning techniques and are fine-tuned to increase their performance on the OSCC dataset. The outputs of various models are then integrated using ensemble learning, leading to an enhanced classification performance. Below is a thorough description of the proposed methodology.

A. EfficientNetB3 Architecture

EfficientNetB3 is a deep learning model known for balancing accuracy and efficiency by scaling network depth, width, and resolution, as shown in Fig. 5. It uses a 3x3 convolutional layer followed by MBConv layers, depthwise separable convolutions, and squeeze-and-excitation (SE) blocks for feature extraction. Larger kernel sizes (5x5) in deeper layers help capture contextual information for OSCC differentiation. After feature extraction, the model classifies Normal and OSCC tissues. Its computational efficiency and performance make it ideal for medical imaging, especially when combined with ResNet50 in an ensemble model for OSCC detection.

EfficientNet-B3 is a scalable model that balances network depth, width, and resolution for optimal performance. The scaling is governed by a compound coefficient ϕ, with the scaling laws given in (3):

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi} \quad (3)$$

where d, w, and r represent the depth, width, and resolution of the network, respectively, and α, β, γ are constants determined through grid search. The output of EfficientNet-B3 is passed through a global average pooling layer, followed by a dense layer to produce the final classification output shown in (4):

$$Z_{efficientnet} = softmax\left(W_{efficientnet} \cdot f_{global} + b_{efficientnet}\right) \quad (4)$$

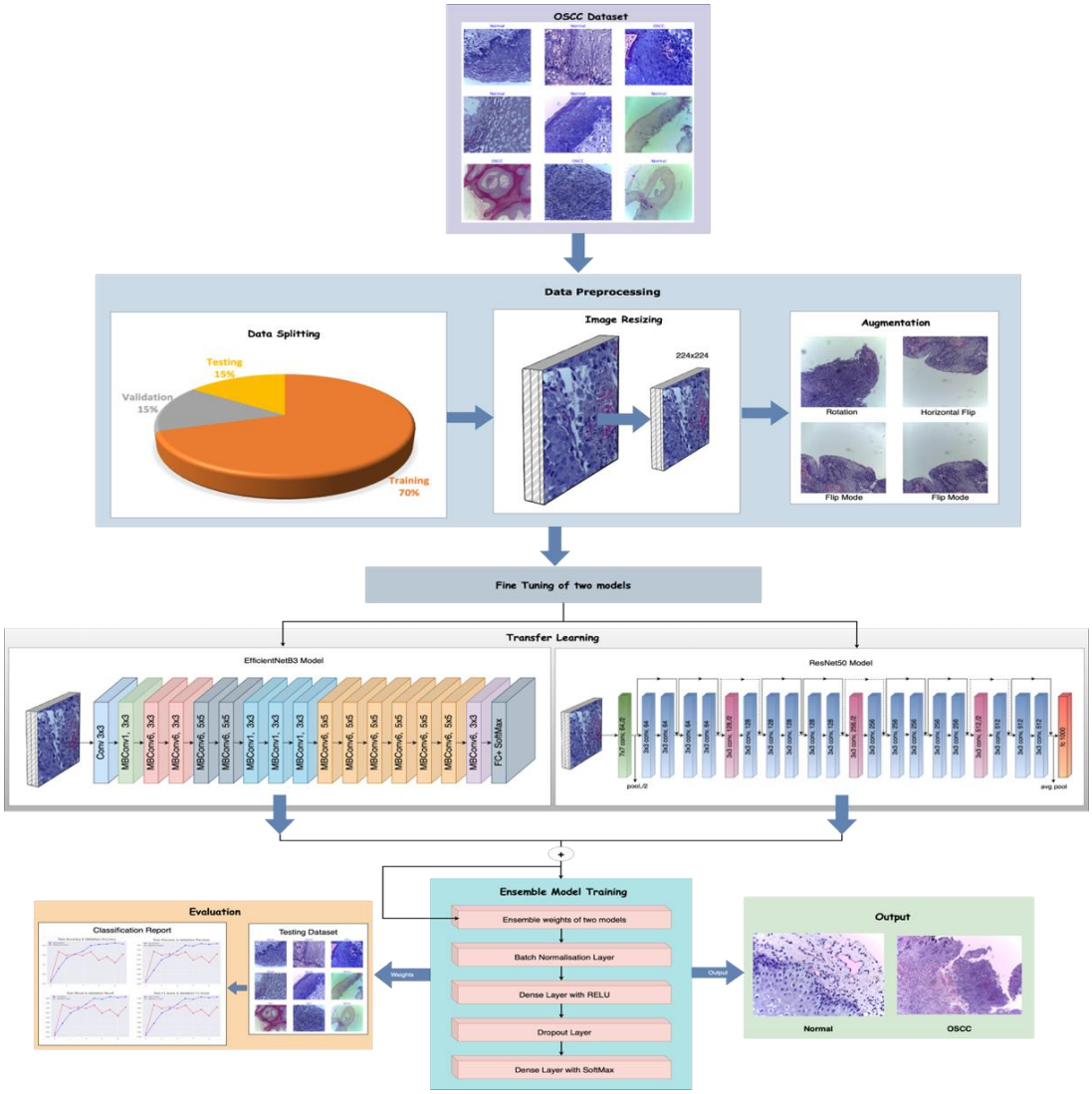

Fig. 4. Proposed methodology.

Where, $Z_{efficientnet}$ represents the predicted probabilities, f_{global} is the feature vector after global pooling, and $W_{efficientnet}$ and $b_{efficientnet}$ are the weights and biases of the final classification layer.

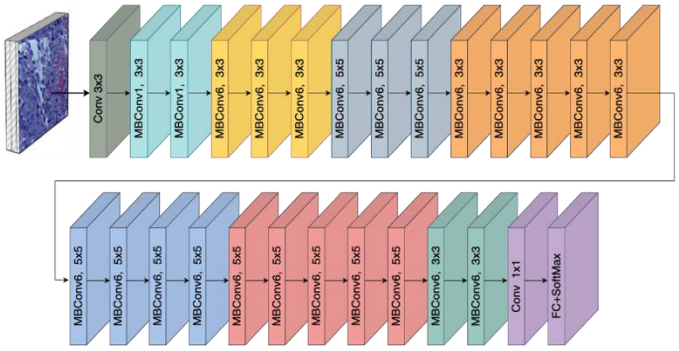

Fig. 5. EfficientNetB3 architecture.

B. *Resnet50 Architecture*

Fig. 6 shows the ResNet50 architecture, a deep CNN designed to overcome challenges in deep network training with 50 convolutional layers, batch normalization, and ReLU activations. Residual connections prevent vanishing gradients, allowing deeper networks. The architecture starts with a 7x7 convolution and max-pooling, followed by four stages with residual blocks and filters ranging from 64 to 512. Spatial dimensions are reduced via stride-2 convolutions for efficiency. The final section includes global average pooling and a fully connected layer. For OSCC detection, the last layer classifies images into Normal and OSCC, helping detect subtle variations in oral cancer.

ResNet50 is a residual network that employs skip connections to address the problem of vanishing gradients in deep networks. The fundamental building block of ResNet50 is the residual block, defined mathematically as in (5):

$$y = F(x, \{W_i\}) + x \tag{5}$$

Where, x is the input, $F(x,\{W_i\})$ represents the convolutional operation with weight parameters $\{W_i\}$, and y is the output of the residual block. The network consists of 50 layers, including convolutional, batch normalization, and ReLU activation layers. The final output of the ResNet50 model, after passing through a global average pooling layer, is given by (6):

$$Z_{resnet} = softmax\left(W_{resnet}.Y_{global} + b_{resnet}\right) \quad (6)$$

Where, Z_{resnet} is the predicted class probabilities, W_{resnet} and b_{resnet} are the weights and biases of the final dense layer, and Y_{global} is the output from the global pooling layer.

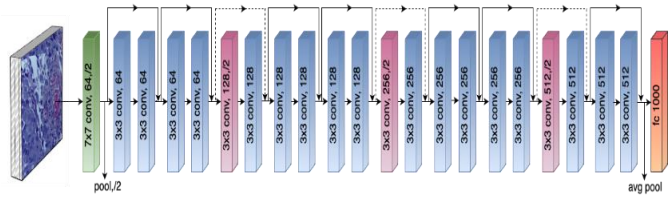

Fig. 6. ResNet50 architecture.

Transfer learning significantly enhanced OSCC detection accuracy by utilizing pre-trained models like EfficientNetB3 and ResNet50, which were fine-tuned on the OSCC dataset. These models transfer learned features, such as texture and shape, from large datasets, reducing the need for extensive labeled data and improving feature generalization. Fine-tuning helps the models capture OSCC-specific patterns, preventing overfitting on small datasets and accelerating training convergence, allowing for efficient and precise differentiation between normal and malignant tissues. This technique proved essential for optimizing accuracy and training speed in medical image analysis.

C. Ensemble Learning

The ensemble model with the architecture seen in Fig. 7 combines the predictions of ResNet50 and EfficientNet-B7 to leverage the strengths of both architectures, thereby enhancing overall classification accuracy.

1) Ensemble model design: The ensemble approach is implemented using a weighted average method, where the final prediction is a weighted combination of the individual models' outputs. The output predictions from ResNet50 and EfficientNet-B7 are combined using a weighted average approach shown in (7):

$$Z_{ensemble} = W_{resnet}.Z_{resnet} + W_{efficientnet}.Z_{efficientnet} \quad (7)$$

where $Z_{ensemble}$ is the final ensemble prediction, and $W_{resnet}, W_{efficientnet}$ are the weights assigned to the ResNet50 and EfficientNet-B3 outputs, respectively. These weights are optimized based on the validation performance of each model.

2) Ensemble model algorithm: This algorithm outlines the process of training individual models, obtaining their predictions, combining those using weighted averages or voting, and optionally training a meta-learner to refine the ensemble's predictions. The Algorithm of the ensemble model is shown in Table I.

TABLE I. ALGORITHM OF ENSEMBLE MODEL

Algorithm 1: Ensemble Model for Oral Squamous Cell Carcinoma Detection

Input: Histopathological images $D_{train}, D_{val}, D_{test}$, number of epochs E, batch size B, learning rate η.
Output: Trained ensemble model for Oral Squamous Cell Carcinoma detection.
Step 1: Model Initialization
1.1 Initialize EfficientNetB3 and ResNet50 models with pre-trained ImageNet weights.
1.2 Configure the final dense layers for detection of two classes (Normal and OSCC).
Step 2: Model Training
2.1 For each epoch e from 1 to E:
2.2 For each batch b in D_{train}:
2.2.1 Perform forward propagation through EfficientNetB3 to obtain $Z_{efficientnet.}$
2.2.2 Perform forward propagation through ResNet50 to obtain Z_{resnet}.
2.2.3 Combine the outputs using the weighted average:
$$Z_{ensemble} = W_{resnet}.Z_{resnet} + W_{efficientnet}.Z_{efficientnet}$$
2.2.4 Compute the cross-entropy loss $\mathcal{L}(Z_{ensemble}, Y)$ where Y is the true label.
2.2.5 Backpropagate the loss and update the weights using the Adamax optimizer with learning rate η.
Step 3: Model Validation
3.1 Evaluate the ensemble model on D_{val} after each epoch.
3.2 Adjust the weights $W_{efficientnet}$ and W_{resnet} based on validation performance.
Step 4: Model Evaluation
4.1 After completing training, evaluate the model on D_{test}.
4.2 Compute the final accuracy, precision, recall, and F1-score.
Step 5: Model Deployment
5.1 Save the trained ensemble model for future use.

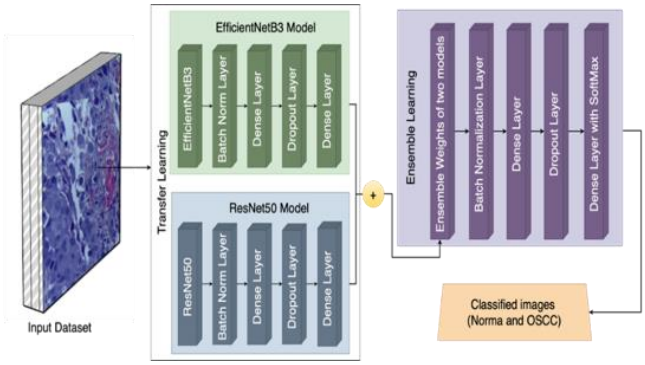

Fig. 7. Ensemble learning architecture.

V. RESULTS

A. Model Evaluation

The work aimed at the identification of Oral Squamous Cell Carcinoma (OSCC) using a deep learning approach combining the strengths of two strong models EfficientNetB3 and ResNet50 by constructing the Ensemble model for improved results. Divided into training, validation, and testing, the models were developed using a set including 5,192 histopathology images. The results reveal the degree of increase in OSCC detection classification accuracy resulting from the combined approach. The Confusion Matrix, Performance criteria, and State of the art Comparison are fully described below.

1) EfficientNet-B3's accuracy and loss analysis: The examination of the EfficientNetB3 model's training and validation accuracy, along with its loss metrics across numerous epochs, is shown in Fig. 8 and Table II. The model demonstrates strong generalization and learning capacity, with an early drop in training loss and gradual reduction in both training and validation loss, ultimately converging to low values. The model quickly reaches 90% accuracy early in training, with validation accuracy following suit. Both accuracy measures level off in later stages, indicating the model's peak performance and suitability for reliable predictions, without overfitting.

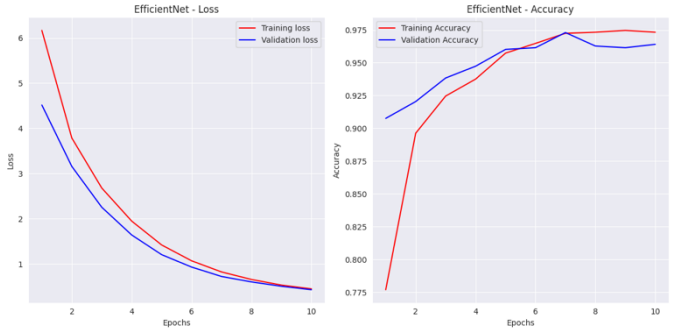

Fig. 8. Loss and accuracy analysis of EfficientNetB3 architecture.

The precise performance per epoch is shown in Table II. It shows that with every epoch the model reduces loss and steadily gains accuracy. The strong learning capacity and adaptability of the EfficientNetB3 model to fresh data are shown by this constant improvement and convergence in both accuracy and loss values [33-36].

TABLE II. EFFICIENTNETB3 MODELS PERFORMANCE PER EPOCH

Epoch	Time(sec)	Loss	Accuracy	Validation Loss	Validation Accuracy
1	131	6.1664	0.7768	4.5153	0.9076
2	54	3.7852	0.8963	3.1570	0.9204
3	54	2.6791	0.9246	2.2538	0.9384
4	53	1.9470	0.9375	1.6400	0.9474
5	54	1.4193	0.9573	1.2048	0.9602
6	54	1.0688	0.9648	0.9303	0.9615
7	54	0.8223	0.9725	0.7222	0.9730
8	54	0.6560	0.9733	0.6015	0.9628
9	54	0.5315	0.9747	0.5054	0.9615
10	54	0.4487	0.9733	0.4286	0.9641

2) ResNet-50's accuracy and loss analysis: The ResNet50 model was trained for 10 epochs, with performance metrics in Table III and accuracy and loss curves in Fig. 9. The training and validation losses steadily decreased, showing effective learning, with a slight generalization gap around the 7th epoch. Training loss stabilized after the 4th epoch, while training accuracy approached 95% by the 10th epoch. Validation accuracy improved gradually, staying slightly below training accuracy. The close alignment of both accuracy curves indicates minimal overfitting and strong generalization. This demonstrates the ResNet50 model's suitability for the current classification task.

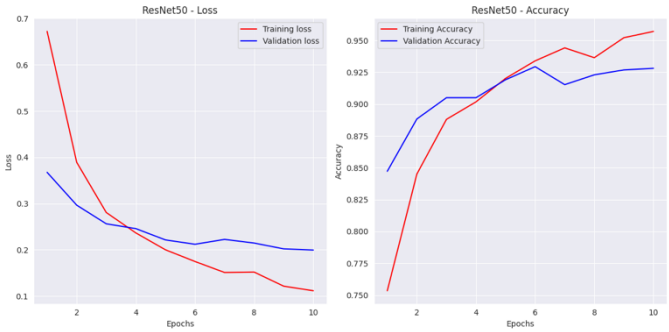

Fig. 9. Loss and accuracy analysis of ResNet50 architecture.

Table III, which illustrates the performance measures for each epoch, shows a constant increase in both training and validation accuracy together with related declines in loss. This development implies that in every epoch the parameters of the model are being tuned somewhat successfully.

TABLE III. RESNET50 MODELS PERFORMANCE PER EPOCH

Epoch	Time(sec)	Loss	Accuracy	Validation Loss	Validation Accuracy
1	43	0.6721	0.7534	0.3674	0.8472
2	36	0.3891	0.8451	0.2965	0.8883
3	36	0.2806	0.8880	0.2562	0.9050
4	36	0.2365	0.9018	0.2457	0.9050
5	37	0.2000	0.9202	0.2214	0.9191
6	41	0.1747	0.9340	0.2120	0.9294
7	36	0.1509	0.9441	0.2225	0.9153
8	37	0.1518	0.9364	0.2146	0.9230
9	35	0.1213	0.9521	0.2020	0.9268
10	36	0.1115	0.9571	0.1994	0.9281

3) Ensemble model's performance analysis: Table IV summarizes the Ensemble model's performance over ten epochs, with corresponding accuracy and loss graphs shown in Fig. 10. The model effectively aggregates multiple classifiers to boost prediction accuracy and reduce generalization error. Training and validation losses consistently decrease, though validation loss shows minor swings, suggesting potential overfitting. Accuracy reaches 99% for training, but validation accuracy varies slightly, indicating room for improvement. Precision and recall graphs in Fig. 11 reveal high precision and steadily improving recall, with validation trends reflecting minor fluctuations. The F1-score and ROC curve in Fig. 12 offer a comprehensive performance view, showing the model's balance between precision and recall and strong discriminative ability. The ROC curve achieves near-perfect AUC scores of 1.00 for both Normal and OSCC classes, highlighting the model's effectiveness in classification with minimal errors, and demonstrating strong reliability.

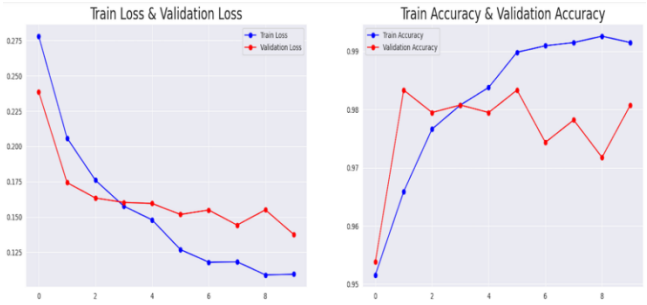

Fig. 10. Loss and accuracy analysis of ensemble model.

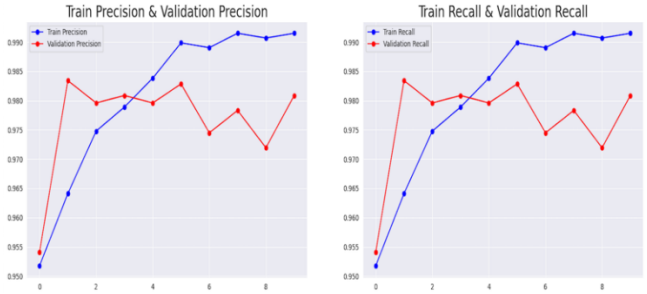

Fig. 11. Precision and recall analysis of ensemble model.

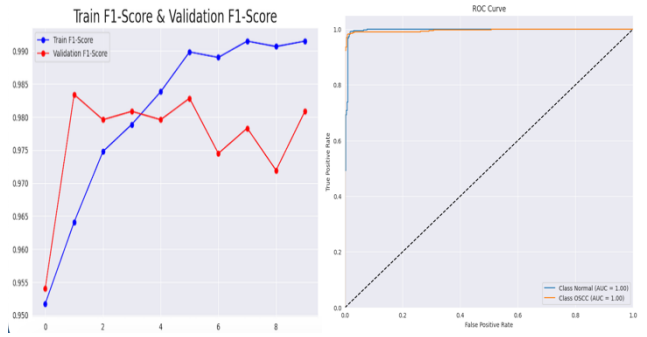

Fig. 12. F1-score and ROC analysis of ensemble model.

Table IV shows the performance measures per epoch, demonstrating constant loss and accuracy improvement. The Ensemble model's efficiency in this classification problem is shown by its ability to preserve great accuracy while reducing loss over epochs.

TABLE IV. ENSEMBLE MODELS PERFORMANCE PER EPOCH

Epoch	Time(sec)	Loss	Accuracy	Validation Loss	Validation Accuracy
1	138	0.1015	0.9854	0.1266	0.9718
2	62	0.0768	0.9907	0.1210	0.9769
3	62	0.0721	0.9923	0.1068	0.9807
4	62	0.0664	0.9934	0.1081	0.9820
5	62	0.0646	0.9951	0.1081	0.9820
6	62	0.0634	0.9942	0.1085	0.9820
7	62	0.0592	0.9940	0.1113	0.9795
8	62	0.0620	0.9926	0.1086	0.9782
9	62	0.0620	0.9926	0.1086	0.9782
10	62	0.0545	0.9934	0.0936	0.9835

B. Confusion Matrix

The three models' performance EfficientNetB3, ResNet50, and the Ensemble Model is evaluated here based on the confusion matrices in Fig. 13, 14, and 15. In Fig. 13 the EfficientNetB3 model displays excellent detection performance, properly classifying 385 Normal instances and 384 OSCC cases. However, it misclassifies 9 Normal cases as OSCC and 21 OSCC cases as Normal, demonstrating a minor bias toward misidentifying OSCC as Normal. Despite these flaws, the model exhibits a high level of accuracy overall.

In Fig. 14, the ResNet50 model accurately detects 349 Normal cases and 363 OSCC cases. With 25 Normal instances mistakenly categorized as OSCC and 42 OSCC cases incorrectly identified as Normal, it reveals a higher number of misclassifications than EfficientNetB3. This suggests that although ResNet50 performs well, it finds greater difficulty differentiating the two groups than EfficientNetB3.

Among the three, the Ensemble model in Fig. 15, which integrates the capabilities of EfficientNetB3 and ResNet50 showcases the best performance. With only three Normal cases misclassified as OSCC and eight OSCC cases misclassified as Normal, it correctly classifies 307 Normal cases and 369 Normal cases. The much smaller number of misclassifications implies that the Ensemble model efficiently lowers the mistakes observed in the individual models, hence producing better general accuracy.

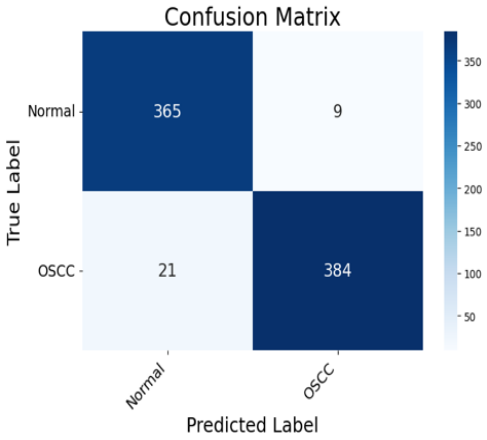

Fig. 13. Confusion matrix of EfficientNetB3.

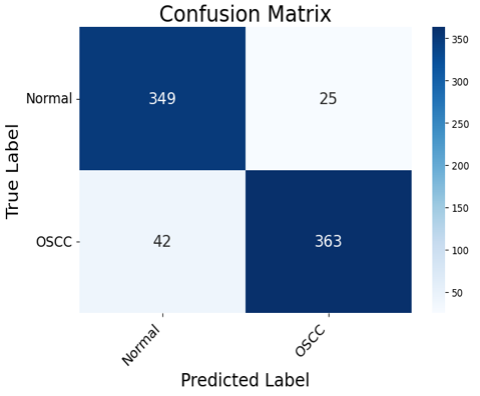

Fig. 14. Confusion matrix of ResNet50.

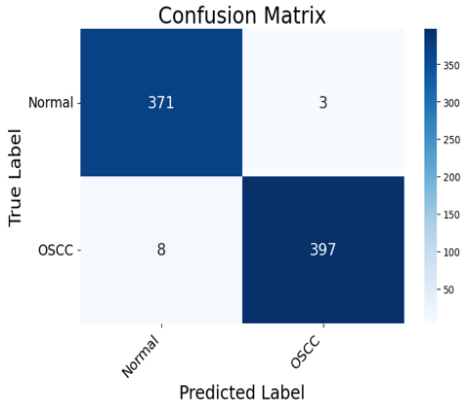

Fig. 15. Confusion matrix of ensemble model.

C. Performance Parameters

Table V presents an extensive three-model performance parameter comparison- precision, recall, F1-score, support, and accuracy across EfficientNetB3, ResNet50, and the Ensemble Model. EfficientNetB3 shows great precision and recall reflecting its great capacity to properly identify positive instances with low false positives and missed real positives. ResNet50 struggles more in differentiating between classes and demonstrates somewhat lower precision and recall even if it is still efficient. Though ResNet50 trails somewhat behind EfficientNetB3, the F1-scores for both models show a reasonable mix of precision and recall. Reflecting its better capacity to balance precision and recall while minimizing classification mistakes, the Ensemble model beats both individual models by obtaining the highest precision, recall, and F1 score. Furthermore, the Ensemble model has the best accuracy, which emphasizes the fact that it can appropriately classify most of the cases. All models show consistent support that guarantees these measures fairly represent performance over a balanced dataset. Combining the strengths of EfficientNetB3 and ResNet50 to provide excellent classification performance, the Ensemble model shows overall to be the most dependable and efficient classifier.

D. State-of-the-Art Comparison

Table VI presents a state-of-the-art comparison of the proposed Ensemble model with EfficientNetB3, ResNet50, and other top approaches in OSCC detection. The Ensemble model outperforms in F1 scores and general accuracy by aggregating model strengths, setting a new benchmark in medical imaging classification tasks.

TABLE V. PERFORMANCE PARAMETERS OF EFFICIENTNETB3, RESNET50 AND ENSEMBLE MODEL

Model	Classes	Precision	Recall	F1-Score	Support	Accuracy
EfficientNetB3	Normal	0.95%	0.98%	0.96%	374	0.96%
	OSCC	0.98%	0.95%	0.96%	405	
ResNet 50	Normal	0.89%	0.93%	0.91%	374	0.91%
	OSCC	0.94%	0.90%	0.92%	405	
Ensemble Model	Normal	0.98%	0.99%	0.99%	374	0.99%
	OSCC	0.99%	0.98%	0.99%	405	

TABLE VI. STATE-OF-THE-ART COMPARISON

Reference No.	Image Type	Technique	Images Count	Accuracy
(2024) [16]	CT images	Inception-V3	499	73.50%
(2024) [17]	Histopathology images	DenseNet121	5192	97.02%
(2024) [18]	Raman Histology	VGG-19	21,703	0.90%
(2024) [19]	Histopathological images	EfficientNetB3	1224	99%
(2024) [20]	Histological images	InceptionV3,Xception,InceptionResNetV2, NASAnet	5685	89.3%,89.5%,91.78%,90.8%
(2024) [21]	Histological images	Convolutional Neural Network	5192	98.49%, 86.89%, 89.37%
(2024) [22]	histopathological images	VGG16,VGG19,InceptionV3, AlexNet, ResNet50	1224	84%,82%,67%,76%,42%
(2024) [23]	Smartphone based images	Deep Learning	760	84.3%
(2024) [24]	Histological images	Deep Learning Models	4946	97.78%
(2024) [25]	histopathological images	DL-CNN	1224	91.1%
(2024) [26]	histopathological images	Deep Learning Models	696	0.92%
(2023) [27]	Histological images	AlexNet	5192	89%
(2024) [28]	Histological images	AI based approaches	5192	94.92%
(2023) [29]	histopathological images	MobileNetV3, InceptionV2, EfficientNetB3	5192	0.89%,0.88%,0.52%
(2023) [30]	Histopathological images	DL-CNN models	1224	91.25%
Proposed Model	Histopathological images	Ensemble Learning	5192	99.34%

VI. Conclusion

This work efficiently applied ensemble learning and transfer learning approaches for the identification of Oral Squamous Cell Carcinoma (OSCC) from histopathology photos. By combining the best features of pre-trained models such as EfficientNetB3 and ResNet50 via ensemble learning, the proposed approach shows astonishing accuracy and robustness. Outliving individual models, the last ensemble model unequivocally displayed increasing generality and accuracy. The detailed study of the performance of the deep learning models in medical image processing highlights its promise since it reveals how to routinely raise accuracy and reduce loss during training and validation datasets. Especially the EfficientNetB3 model has shown remarkable performance measures, thereby addressing the problems of early OSCC detection as a required need for timely and effective treatment. The requirement of incorporating modern deep-learning architectures with ensemble learning techniques to provide reliable and efficient diagnostic tools is underlined in this work. Apart from improving detection accuracy, the proposed method provides a structure suitable for other kinds of cancer and medical image analysis uses. Using more varied datasets and investigating other deep-learning approaches will help to improve the performance of the model, hence producing strong, real-time diagnostic systems for clinical use.

VII. Limitations and Future Scope

Although the suggested transfer learning and ensemble learning method greatly increased the accuracy of OSCC detection, some constraints have to be mentioned. First of all, compared to larger-scale medical imaging datasets, the dataset size is somewhat tiny even if it is enough for the present work. This may restrict the model's generalizability to unprocessed data from many populations. Second, depending on pre-trained models such as EfficientNetB3 and ResNet50 implies that more specific medical datasets for OSCC could help to improve the model's performance even further. Including a bigger, more varied dataset in the next projects could help the model to be more resilient. A deeper understanding of model predictions may also come from investigating more sophisticated ensemble methodologies and including explainable artificial intelligence approaches. Another important development of this work would be implementing the model in actual clinical environments and verifying its efficacy among several institutions.

References

[1] S. Y. Yang et al., "Histopathology-based diagnosis of oral squamous cell carcinoma using deep learning," J. Dent. Res., vol. 101, no. 11, pp. 1321–1327, 2022.

[2] C. Carreras-Torras and C. Gay-Escoda, "Techniques for early diagnosis of oral squamous cell carcinoma: Systematic review," Med. Oral Patol. Oral Cir. Bucal, vol. 20, no. 3, pp. e305-15, 2015.

[3] S. R. Larsen, J. Johansen, J. A. Sørensen, and A. Krogdahl, "The prognostic significance of histological features in oral squamous cell carcinoma," J. Oral Pathol. Med., vol. 38, no. 8, pp. 657–662, 2009.

[4] A. K. Shaw et al., "Diagnostic accuracy of salivary biomarkers in detecting early oral squamous cell carcinoma: A systematic review and meta-analysis," Asian Pac. J. Cancer Prev., vol. 23, no. 5, pp. 1483–1495, 2022.

[5] B. F. Adeyemi and B. Kolude, "Clinical presentation of oral squamous cell carcinoma," Niger. Postgrad. Med. J., vol. 20, no. 2, pp. 108–110, 2013.

[6] L. Nokovitch et al., "Oral cavity squamous cell carcinoma risk factors: State of the art," J. Clin. Med., vol. 12, no. 9, 2023.

[7] C. Seethalakshmi, "Early detection of oral squamous cell carcinoma (OSCC)-Role of genetics: A literature review," Journal of clinical and diagnostic research, vol. 7, no. 8, 2013.

[8] P. Hollows, P. G. McAndrew, and M. G. Perini, "Delays in the referral and treatment of oral squamous cell carcinoma," Br. Dent. J., vol. 188, no. 5, pp. 262–265, 2000.

[9] J. J. Sciubba, "Oral cancer: the importance of early diagnosis and treatment," American journal of clinical dermatology, vol. 2, pp. 239–251, 2001.

[10] C. E. Palme, P. J. Gullane, and R. W. Gilbert, "Current treatment options in squamous cell carcinoma of the oral cavity," Surg. Oncol. Clin. N. Am., vol. 13, no. 1, pp. 47–70, 2004.

[11] D. K. Das, S. Bose, A. K. Maiti, B. Mitra, G. Mukherjee, and P. K. Dutta, "Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis," Tissue Cell, vol. 53, pp. 111–119, 2018.

[12] S. Arora, A. Matta, N. K. Shukla, S. V. S. Deo, and R. Ralhan, "Identification of differentially expressed genes in oral squamous cell carcinoma. Molecular Carcinogenesis," vol. 42, pp. 97–108, 2005.

[13] M. Das, R. Dash, and S. K. Mishra, "Automatic detection of oral squamous cell carcinoma from histopathological images of oral mucosa using deep convolutional neural network," Int. J. Environ. Res. Public Health, vol. 20, no. 3, 2023.

[14] R. Marzouk et al., "Deep transfer learning driven oral cancer detection and classification model," Comput. Mater. Contin., vol. 73, no. 2, pp. 3905–3920, 2022.

[15] S. Krishna, J. Lavanya, G. Kavya, N. Prasamya, and Swapna, "Oral cancer diagnosis using deep learning for early detection," in 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022.

[16] A. Fanizzi et al., "Explainable prediction model for the human papillomavirus status in patients with oropharyngeal squamous cell carcinoma using CNN on CT images," Sci. Rep., vol. 14, no. 1, p. 14276, 2024.

[17] M. E. Paramasivam, B. S. Sriganesh, and S. Sureshkrishna, "Oral cancer detection using convolutional neural network," in 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), 2024.

[18] A. Weber et al., "AI-based detection of oral squamous cell carcinoma with Raman Histology," Cancers (Basel), vol. 16, no. 4, p. 689, 2024.

[19] E. Albalawi et al., "Oral squamous cell carcinoma detection using EfficientNet on histopathological images," Front. Med. (Lausanne), vol. 10, p. 1349336, 2023.

[20] K. V. Kumar, S. Palakurthy, S. H. Balijadaddanala, and S. Reddy, "Early Detection and Diagnosis of Oral Cancer Using Deep Neural Network," Journal of Computer Allied Intelligence, vol. 2, no. 02, pp. 22–34, 2024.

[21] A. Mishra, K. Srinivas, and A. Charan Kumari, "Empowering Oral Squamous Cell Carcinoma detection with deep learning: Insights from convolutional neural network analysis of histopathological images," J. Tr., Chal. Art. Intell., vol. 1, no. 2, pp. 45–50, 2024.

[22] M. U. A. Siddique, S. M. Rabha, J. Periwal, N. Choudhury, and R. Mandal, "Early detection of oral cancer using image processing and computational techniques," in Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology, Singapore: Springer Nature Singapore, 2024, pp. 37–54.

[23] E. S. Mira et al., "Early diagnosis of oral cancer using image processing and artificial intelligence," Fusion: Practice and Applications, vol. 14, no. 1, pp. 293–308, 2024.

[24] B. S. Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "Supremacy of attention-based transformer in oral cancer classification using histopathology images," Int. J. Data Sci. Anal., 2024.

[25] A. Soni, P. K. Sethy, A. K. Dewangan, A. Nanthaamornphong, S. K. Behera, and B. Devi, "Enhancing oral squamous cell carcinoma detection:

a novel approach using improved EfficientNet architecture," BMC Oral Health, vol. 24, no. 1, p. 601, 2024.

[26] B. S. Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification," Int. J. Data Sci. Anal., 2024.

[27] T. Saraswathi and V. M. Bhaskaran, "Classification of Oral Squamous Carcinoma Histopathological images using Alex Net," in 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), IEEE, 2023, pp. 637–643.

[28] M. Ahmad, M. I. Khattak, A. Jan, and I. U. Haq, A Novel Hybrid AI-Based System for Early Detection of Oral Squamous Cell Carcinoma via Histopathological Images.

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

[30] S. H. Begum and P. Vidyullatha, "Deep Learning Model for Automatic Detection of Oral squamous cell carcinoma (OSCC) using Histopathological Images," International Journal of Computing and Digital Systems, 2023.

[31] A. F. Kebede, "Histopathologic Oral Cancer Detection using CNNs." 21-Jul-2021.

[32] B. Ananthakrishnan, A. Shaik, S. Kumar, S. O. Narendran, K. Mattu, and M. S. Kavitha, "Automated detection and classification of oral squamous cell carcinoma using deep neural networks," Diagnostics (Basel), vol. 13, no. 5, p. 918, 2023.

[33] S. Bharany and S. Sharma, "Intelligent green internet of things: An investigation," in Machine Learning, Blockchain, and Cyber Security in Smart Environments. Chapman and Hall/CRC, 2022, pp. 1–15.

[34] S. Badotra et al., "A DDoS Vulnerability Analysis System against Distributed SDN Controllers in a Cloud Computing Environment," Electronics, vol. 11, no. 19. MDPI AG, p. 3120, Sep. 29, 2022. doi: 10.3390/electronics11193120.

[35] K. Kaushik et al., "Multinomial Naive Bayesian Classifier Framework for Systematic Analysis of Smart IoT Devices," Sensors, vol. 22, no. 19. MDPI AG, p. 7318, Sep. 27, 2022. doi: 10.3390/s22197318.

[36] S. Bharany, K. Kaur, S. E. M. Eltaher, A. O. Ibrahim, S. Sharma, and M. M. M. A. Elsalam, "A Comparative Study of Cloud Data Portability Frameworks for Analyzing Object to NoSQL Database Mapping from ONDM's Perspective," International Journal of Advanced Computer Science and Applications, vol. 14, no. 10. The Science and Information Organization, 2023. doi: 10.14569/ijacsa.2023.0141086.