Expert Systems With Applications 296 (2026) 129056

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Advanced hybrid deep learning approach for automated OSCC classification using improved pattern-based features

Anuradha Suresh Pandit [a,*], Vaibhav Vitthalrao Dixit [b]

[a] Department of Electronics & Telecommunication, G. H. Raisoni College of Engineering and Management, Pune 412207 Maharashtra, India
[b] Department of Electronics & Telecommunication, RMD Sinhgad Technical Institutes Campus, Warje, Pune, Maharashtra, India

ARTICLE INFO

Keywords:
Oral squamous cell carcinoma
Gaussian filter
MCD-BIRCH
MMBP
CAA-Bi-LSTM

ABSTRACT

Oral Squamous Cell Carcinoma (OSCC) is a prevalent and aggressive form of oral cancer, characterized by high morbidity and mortality rates. For better patient results and survival, an early and precise diagnosis is essential. Traditional diagnostic methods rely heavily on clinical expertise, which can introduce variability and potential errors. In this paper, a novel Deep Learning model-based Automated Classification method for Oral Squamous Cell Carcinoma (AC-OSCC-DL) is proposed. This study presents an advanced automated classification system for OSCC using the following comprehensive pipeline. The preprocessing step involves noise reduction using a Gaussian filter, while segmentation is done through Modified Cluster distance-based Balanced Iterative Reducing and Clustering using the Hierarchies (MCD-BIRCH) segmentation technique, which enhances the accuracy of identifying relevant regions within histopathological images. Feature extraction is performed by Visual Geometry Group Network with 16 layers (VGG-16) and Residual Network (ResNet) architectures, alongside statistical features and Modified Median Binary Patterns (MMBP), to capture a wide range of image characteristics. For classification, a hybrid model combining Collaborative Attention layer Assisted Bi-directional Long Short-Term Memory (CAA-Bi-LSTM) and Link Net models are employed to accurately categorize OSCC into different grades. The proposed system demonstrates enhanced performance by mitigating traditional limitations and offers a robust, automated approach to OSCC diagnosis, facilitating timely and precise treatment decisions. The CAA-Bi-LSTM + Link Net strategy achieved the highest accuracy of 0.986, an F-measure of 0.979, and a Precision of 0.989.

1. Introduction

Oral cancer accounts for 30–40 % of all cancers and ranks as the sixth most prevalent cancer globally. In India, 15.62 % of the population is affected, with more than 33 % of these cases are diagnosed at advanced stages. The risk factors for oral squamous cell carcinoma (OSCC; Eggermont et al., 2023; Sievert et al., 2022) include smoking, smokeless tobacco use, alcohol consumption, poor oral hygiene, and genetic alterations. Disruptions in the epigenetic network caused by external factors can potentially lead to malignancy (Greeshma et al., 2023; Kumar et al., 2024).

Surgery for oral tumors is said to have a 5-year survival rate of less than 50 %.Surgeons often rely on visual inspection and palpation for intraoperative testing, but up to 87 % of tumour-positive margins may be found in deeper soft tissue layers (Ali et al., 2024; Sordi et al., 2023; de Lima et al., 2023). Due to the time-consuming nature of the frozen section technique for pathological evaluation, effectively excising oral tumor cells during surgery is challenging when based only on the surgeon's expertise and the pathological findings. Therefore, a significant number of recent studies are dedicated to advanced non-invasive detection technologies assisting surgeons in determining the adequacy of surgical margins. Optical technology (Tsai et al., 2023; Öhman et al., 2023) offers advantages such as accuracy, robustness, portability, low cost, and ease of use, making it a valuable tool for clinical applications (Steybe et al., 2023; Liang et al., 2023).

Due to delays in detection, the likelihood of successful treatment is very low, resulting in a high mortality rate (Muthupalani et al., 2023; Meng et al., 2023). The survival rate could be increased to 90 % with an early diagnosis and the right therapies. Thus, there is an urgent need for oral cancer (OC) diagnosis techniques that are accurate, rapid, user-friendly, and non-invasive (Tarrad et al., 2023; de Lanna et al., 2022). The dentist begins by evaluating the primary indicators of abnormal tissue in the oral cavity. If potential malignancy is suspected, the patient

* Corresponding author.
E-mail address: anuradha.pandit.phdetc@ghrcem.raisoni.net (A.S. Pandit).

https://doi.org/10.1016/j.eswa.2025.129056
Received 2 January 2025; Received in revised form 4 July 2025; Accepted 16 July 2025
Available online 19 July 2025

Nomenclature

Abbreviation	Description
AC-OSCC-DL	Automated Classification Approach for Oral Squamous Cell Carcinoma Using Deep Learning
AI	Artificial Intelligence
OSCC	Oral Squamous Cell Carcinoma
Bi-GRU	Bidirectional Gated Recurrent Unit
CAA-Bi-LSTM	Collaborative Attention Layer Assisted Bi-Directional Long Short-Term Memory
CNN	Convolutional Neural Networks
DCNNs	Deep Convolutional Neural Networks
DL	Deep Learning
ECIS	Electric Cell-Substrate Impedance Sensing
DNA	Deoxyribonucleic Acid
Bi-LSTM	Bi-Directional Long Short-Term Memory
GAP	Global Average Pooling
Light GBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
ResNet	Residual Network
MCD-BIRCH	Modified Cluster Distance-Based Balanced Iterative Reducing and Clustering Using Hierarchies
MMBP	Modified Median Binary Patterns
OC	Oral Cancer
MBP	Median Binary Pattern
PCA	Principal Component Analysis
RNN	Recurrent Neural Network
TSCC	Tongue Squamous Cell Carcinoma
VGG-16	Visual Geometry Group Network With 16 Layers

is referred to an oral or maxillofacial surgeon for further testing (do Valle et al., 2023; Sukegawa et al., 2023). However, existing diagnostic methods rely heavily on the clinician's knowledge and expertise, making them subjective and prone to errors. The challenge of setting thresholds for node parameters further complicates the classification of oral conditions. Therefore, an automated system would significantly benefit early OC detection and reduce false-positive rates. This research proposes a new Automated Classification approach for Oral Squamous Cell Carcinoma (AC-OSCC-DL) based on Deep Learning models.

The major contributions are:

- Adopting the Gaussian filtering technique to preprocess the input image. Using the MCD-BIRCH-based segmentation approach, the preprocessed image is segmented to obtain the region of interest. Here, proposed Jensen-Shannon distance instead of Euclidean distance to compute the cluster distance.
- Contributing MMBP for feature extraction step that extracts texture patterns from the segmented image; along with this, VGG-16, ResNet, and statistical features are extracted.
- Proposing CAA-Bi-LSTM for classifying OSCC, where it utilizes an improved collaborative attention layer. This layer adopts the G-softmax function to compute the attention weights in the attention mechanism.

The format of the paper is: Section 2 discusses the related works on OSCC. Section 3 explains the developed method. Section 4 illustrates the results, comparing them with previous studies and outlining implications. Section 5 includes the conclusion.

2. Literature review

Haq et al. (2023) has explored the transformative capabilities of AI in OSCC diagnosis. Their approach classified extracted features using CatBoost individually, then concatenated them for image classification. The most effective strategy was the third one, which integrated Gabor filtering with CatBoost classification and ResNet50 feature extraction.

Wang et al. (2024) have evaluated the anti-cancer potential of the Lobophytum crassum extract in various cancer cell types. Although its effects on OSCC cells have not been explored, this study seeks to determine how the extract influences OSCC cells. To ensure accurate results, ECIS was used in parallel with SAS cells.

Raja et al. (2024) has aimed to evaluate and compare the DNA methylation of the MAP1LC3Av1 and ATG5 genes in oral leukoplakia and OSCC. The study, involving 48 tissue samples diagnosed as "OL, OSCC, or normal tissue", revealed statistically significant differences among the groups. These findings shed light on the genes' important roles in tumor progression.

Yan et al. (2020) has developed an ensemble CNN framework utilizing fiberoptic Raman spectroscopy alongside DL techniques to distinguish TSCC from non-tumor tissues. This data was then processed by the ensemble CNN strategy. The final step involved generating a feature vector through a fusion layer, which was subsequently used by the FC layer for TSCC classification.

Goswami et al. (2024) has introduced a technique capable of differentiating benign from malignant oral cavity lesions and identifying pre-cancerous stages. Their method explores five unique color spaces to extract texture as well as color features, which are then classified using the LightGBM. By utilizing hand-crafted features and an ML classifier, the proposed approach is both resource-efficient and relatively quick.

Panigrahi et al. (2022) have proposed an innovative approach for oral cancer classification using the DL technique known as capsule networks. The network's capsule dynamic routing as well as routing through agreement features offer increased resilience to rotation and affine transformations in augmented oral datasets. Its capacity to handle different views, orientations and poses makes it particularly appropriate for analyzing oral cancer images in their early stages.

Panigrahi et al. (2020) has proposed two strategies for the classification of OSCC histopathology images. The approach began by implementing transfer learning with DCNNs to establish the most effective model for differentiating malignant tumors from benign ones. This was followed by a comparative assessment of the models based on their classification accuracy and other performance criteria.

Das et al. (2020) has investigated oral biopsy images using two methods: "i) transfer learning with pre-trained DCNNs, testing four models like AlexNet, VGG-16, VGG-19, and ResNet-50, to determine the most effective classification model, and ii) a newly proposed CNN model". Due to the study, diagnosing patients with OSCC may be greatly aided by the CNN-based multi-class grading system.

Sukegawa et al. (2023) has introduced to advance OSCC diagnosis through the use of CNN deep-learning models for classifying histopathological images. Prepared and labelled by pathologists, images were tiled and analyzed with ResNet50 and VGG16 models, employing optimizers both without and with a learning rate scheduler, SAM and stochastic gradient descent with momentum. The study identified key conditions for optimal CNN performance through performance metrics analysis.

Soni et al. (2024) has proposed a deep-learning approach for automating the early diagnosis of oral cancer from histopathology images. Their CNN model categorizes oral biopsy images as benign or malignant. Among 17 pre-trained models, EfficientNetB0 was determined through a two-step analysis, to be the most effective. Moreover, the effectiveness of the model was further improved by integrating a dual attention network (DAN). Table 1 shows the summary of related works.

Table 1
Review on extant works.

Author [Citation]	Methodology	Features	Challenges	Results obtained
(Haq et al., 2023)	ResNet50 and PCA	It reduces the issue of overfitting.	Additional research is needed to enhance AI-based tools for cancer therapy and detection, improving their effectiveness and accessibility.	Accuracy of 94.92 %
(Wang et al., 2024)	ECIS	The groups treated with elevated concentrations of C127 showed uniform results.	More research is necessary to understand how the Keap1/Nrf2 pathway affects apoptosis induction and cell migration inhibition by C127 in oral cancers.	C127 inhibited cell viability in a dose dependent manner ($P < 0.001$)
(Raja et al., 2024)	MAP1LC3Av1 gene and DNA methylation of ATG 5	According to the approach, methylation silencing of MAP1LC3Av1 induced by Helicobacter pylori may disrupt advanced gastric carcinogenesis and autophagy.	To refine future studies, consider increasing sample size, using a multicentric approach, stratifying by clinical grade, and evaluating the tumor invasion index.	The statistical significance of p value <0.05
(Yan et al., 2020)	CNN	The approach provides two advantages it reduces noise creation while simultaneously increasing its distinguishing power.	The TSCC detection approach, integrating Raman spectroscopy with an ensemble CNN model, is expected to enhance surgeries targeting TSCC edges.	Accuracy is 99.2 %
(Goswami et al., 2024)	LightGBM	The analysis revealed that 126 features had the highest information gain.	The approach uses handcrafted feature extraction with an ML classifier, making it significantly more resource- and time-efficient than DL methods.	Accuracy of 99.25 %,
(Panigrahi et al., 2022)	Capsule network	The network's ability to manage different poses, orientations, and views makes it appropriate for validating oral cancer images.	The proposed system has the potential to be extended for staging oral cancer at different levels.	97.35 % Accuracy
(Panigrahi et al., 2020)	DCNNs	In transfer learning, fine-tuning adapts a model trained on one task to improve its performance on a new task.	The scalability of this work can be enhanced by using a larger dataset, leveraging high-performance GPU systems and adding more deep dense layers.	Accuracy of 96.6 %
(Das et al., 2020)	CNN	Achieving 97.5 % accuracy and offering the highest performance in the comparison.	To guarantee precise treatment, it is crucial to establish a real-time decision support system.	Accuracy of 96.94 %
(Sukegawa et al., 2023)	CNN	Although the number of epochs was limited, the learning rate scheduler successfully improved performance.	Sufficient resources, capable of managing the computational demands, are essential for validating the use of more sophisticated CNN models.	Accuracy is 0.8622
(Soni et al., 2024)	DL-CNN	The EfficientNetB0 model achieved the highest effectiveness among the tested methods, with a mean accuracy of 86.66 %.	Using attention mechanisms and saliency maps to improve the interpretability of DL-CNN model predictions is critical for establishing clinician trust and integrating the model into clinical settings.	Accuracy of 91.1 %

2.1. Problem statement

Oral Squamous Cell Carcinoma (OSCC) represents a major global health challenge due to its high incidence and the fact that it is often diagnosed at advanced stages, significantly impacting patient survival rates. Despite advances in medical imaging and diagnostic techniques,

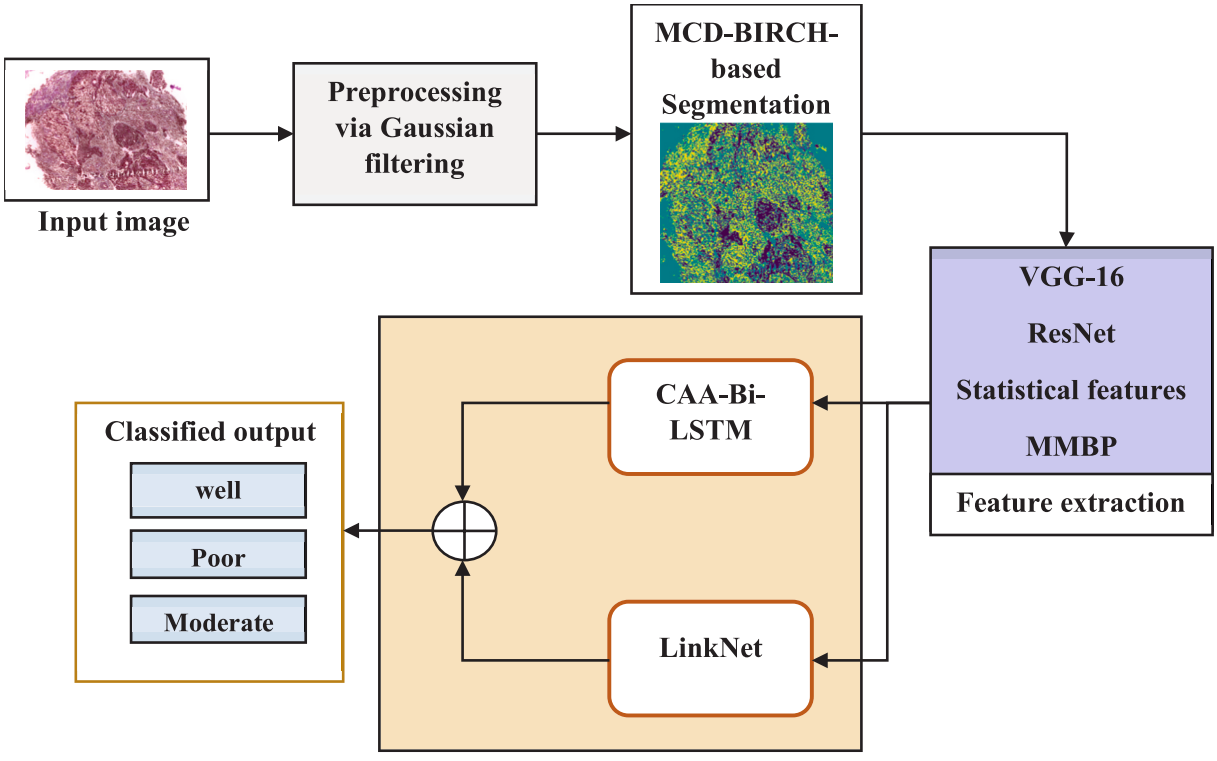

Fig. 1. Structure of the suggested AC-OSCC-DL model.

early detection remains difficult, and traditional methods are hindered by limitations in accuracy, scalability, and clinical application. Current diagnostic approaches often rely heavily on clinical expertise, introducing variability and increasing the risk of errors in interpretation, especially in complex or subtle cases. Furthermore, there is a critical need for more robust, automated systems capable of handling large datasets and providing consistent, accurate results. These drawbacks underscore the pressing need for creative solutions that might improve OSCC diagnosis accuracy and dependability while also being simple to incorporate into clinical procedures. This paper proposes an advanced deep learning-based approach that addresses these gaps, offering a more accurate, scalable, and efficient solution for early OSCC detection, thereby improving patient outcomes and facilitating timely treatment decisions.

3. An overview of proposed automated classification oral squamous cell carcinoma (OSCC)

Oral Squamous Cell Carcinoma (OSCC) is a highly common and aggressive type of oral cancer, frequently linked to significant mortality and morbidity. To improve patient outcomes and raise survival rates, prompt detection and accurate diagnosis are crucial. However, conventional diagnostic approaches depend heavily on clinician expertise, making them susceptible to variability and human error. As a result, this study suggests a new Deep Learning model-based Automated Classification method for OSCC (AC-OSCC-DL). As seen in Fig. 1, the suggested procedure for OSCC classification takes a methodical approach, starting with input picture preparation. In order to smooth the images by reducing noise and highlighting important characteristics, the images are first preprocessed using a Gaussian filter. Following this, the images are segmented using the MCD-BIRCH segmentation technique. This method efficiently groups pixels, distinguishing tumor regions from the surrounding tissue.

After segmentation, feature extraction is performed using a combination of deep learning and statistical techniques. The VGG-16 and ResNet architectures extract deep features, capturing complex patterns from the images. To provide an extensive analysis of the visual data, statistical features are also extracted. The feature set is further detailed by applying the MMBP to the extraction of texture-based features. After that, the features are sent into the classification stage, where a hybrid model that combines the LinkNet architecture and the CAA-Bi-LSTM network is used. The final classified output categorizes the OSCC images into three levels well differentiated, moderately differentiated, and poorly differentiated, which are critical for determining the appropriate clinical treatment and prognosis for patients.

3.1. Preprocessing via Gaussian filter

The preprocessing step in this workflow is essential for preparing input images for further analysis, especially for classifying Oral Squamous Cell Carcinoma (OSCC). The primary objective of preprocessing is to improve image quality, reduce noise, and ensure that critical features are more visible for the subsequent processing stages. This process starts with acquiring raw images (considered $\aleph^{Inp}$ to be the input image), which may have noise, artefacts, or inconsistencies due to variations in lighting or staining methods. To mitigate these issues, a Gaussian filtering technique is applied that is discussed as follows:

Gaussian filter: A Gaussian filter (Kalaivani and Asnath Victy Phamila, 2018) is a linear smoothing filter where weights are determined by a Gaussian function. Constructing this filter with Pascal's triangle and applying it recursively helps form a Gaussian pyramid, thereby improving its performance. Gaussian filtering is particularly well-suited for preprocessing OSCC images due to its ability to reduce noise while preserving essential features. It outperforms other methods, such as mean or median filters, by maintaining structural integrity and improving the SNR. The use of Gaussian filtering ensures that subsequent steps, such as segmentation, feature extraction, and classification, are performed on cleaner, more informative images, ultimately enhancing the overall performance of the OSCC detection system. Quantitative metrics like SNR improvements, combined with visual comparisons, help demonstrate the significant impact of this preprocessing step on image quality and subsequent processing stages. A 3 $\times$ 3 kernel, based on the third row of Pascal's triangle, functions as a particular kind of averaging filter. The filtering process involves several steps: "First, the outer boundary pixels of the image are padded. Next, a Gaussian mask is generated and applied to each pixel in the noisy image. This is followed by element-wise multiplication of the mask with the pixel's neighborhood, and summing the results. The computed sum is then assigned to the central pixel of the kernel. These steps are repeated for every pixel in the image to achieve a smooth, noise-reduced result". Then, the noise-free or preprocessed image is denoted by $\aleph^{Pre}$.

3.2. Augmentation process

After preprocessing, the noise-reduced images undergo augmentation to enhance the robustness of the dataset. The augmented images are created by applying various transformation techniques, such as:

- Shear
- Rotation
- Shifting
- Scaling
- Translation

These augmented images, denoted by $\aleph^{Aug}$, provide diversity in the dataset, making the model more resilient to variations. Once augmented, the images are prepared for the next stage of segmentation.

3.3. Segmentation via MCD-BIRCH approach

The segmentation step is a critical phase in image analysis that involves isolating an image, $\aleph^{Aug}$ into different segments or regions to facilitate more detailed examination and classification. In the context of classifying OSCC, segmentation focuses on isolating relevant regions of the image, such as tumor areas, from the surrounding tissue. The primary goal of segmentation is to separate and identify Regions of Interest (RoI) within the image, such as tumor tissues or abnormal areas, from the rest of the image. This allows for a more focused analysis of specific areas relevant to OSCC diagnosis. In this workflow, the Modified Cluster distance-based Balanced Iterative Reducing and Clustering using Hierarchies (MCD-BIRCH) segmentation technique is employed and it is described as follows.

MCD-BIRCH segmentation: The BIRCH algorithm (Ramadhani et al., 2020) is a hierarchical clustering technique that uses Clustering Features (CF) and a CF Tree to efficiently manage and store clustering information. This approach significantly reduces memory usage and improves the algorithm's scalability and speed, making it effective for clustering large datasets with both discrete and continuous attributes.

In the BIRCH algorithm, objects are organized into sub-clusters as CFs. Each CF is a triple $CF = (\eta, lS, sS)$, where lS is the aggregation of attribute values, η is the count of data points, and sS is the aggregation of squared attribute values. When merging CFs, specific update rules are applied as in Eq. (1).

$$CF_{12} = (\eta_1 + \eta_2, \overline{lS_1} + \overline{lS_2}, sS_1 + sS_2) \tag{1}$$

In BIRCH, the CF sub-clusters are summarized incrementally, with only the Vector CF stored in memory. This Vector CF is adequate for calculating essential sub-cluster characteristics, including radius, centroid, and diameter, providing a space-efficient method by summarizing the sub-cluster rather than retaining all data points. To determine which cluster features should be merged, the Euclidean distance formula

is employed as in Eq. (2).

$$ED(u,v) = \sqrt{(lS_1 - sS_1)^2 + (lS_2 - sS_2)^2} \tag{2}$$

The Euclidean distance used in BIRCH can be quite sensitive to the order in which data is inputted. This sensitivity may result in suboptimal segmentation performance, especially when clusters exhibit irregular shapes and varying densities. To address this, a new MCD-BIRCH algorithm is used for better segmentation. Here, it adopts improved Jensen Shannon distance (Zunino et al., 2022). The standard Jensen-Shannon distance is formulated as in Eq. (3)

$$JS_{dis}(lS, sS) = \frac{1}{2}\left[D_{kl}\left[lS, \frac{lS+sS}{2}\right] + D_{kl}\left[sS, \frac{lS+sS}{2}\right]\right] \tag{3}$$

where, $D_{kl} = \sum_{i=1}^{n} lS_i \log\left(lS_i / sS_i\right)$ and these metrics are sensitive to outliers. To overcome this issue, improved Jensen-Shannon distance is proposed which is defined as Eq. (4).

$$^{proposed}JS_{dis}(lS, sS) = \frac{1}{2}\left[{}^{proposed}D_{kl}\left(lS, \frac{lS+sS}{2}\right) * Df + {}^{proposed}D_{kl}\left(sS, \frac{lS+sS}{2}\right) * Df\right] \tag{4}$$

where, $^{proposed}D_{kl} = \sum_{i=1}^{n} Cd(lS_i) \times \log[Cd(lS_i/sS_i)]$; Cauchy distribution value $Cd = \frac{1}{\pi(1+x^2)}$; and distance factor $Df = \frac{lS_i * sS_i}{\sum_{i=1}^{n} lS_i * \sum_{i=1}^{n} sS_i}$.

Therefore, using the improved Jensen-Shannon distance can mitigate the impact of sensitive outliers. Additionally, the Jensen-Shannon distance enhances the stability and consistency of CF-tree construction. This approach ultimately leads to better cluster quality and more accurate segmentation results.

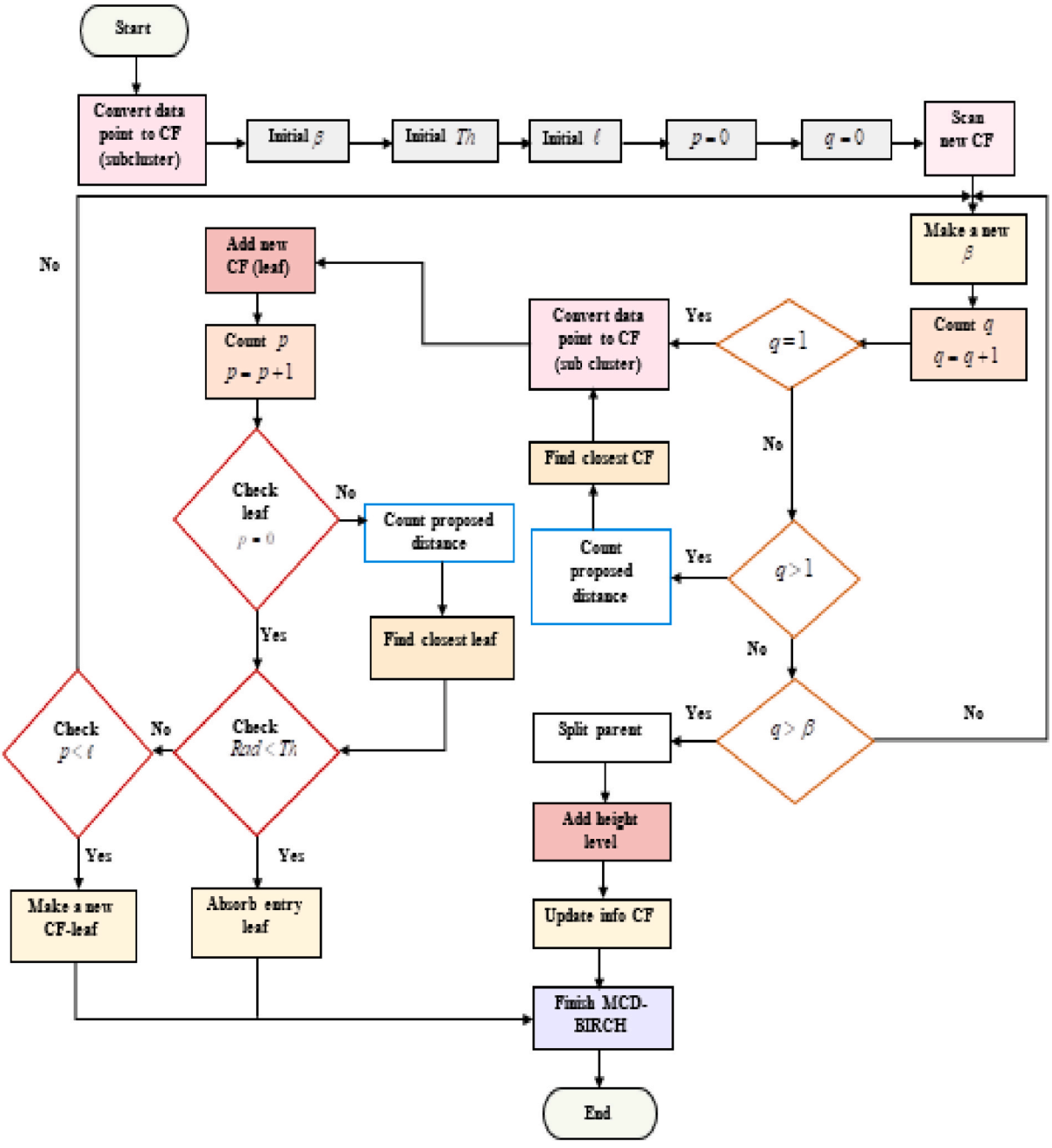

Fig. 2. Flowchart of MCD-BIRCH algorithm.

Further, Eq. (5) shows the computation of the radius of CF-leaf.

$$Rad = \frac{\sqrt{sS - (lS)^2}\Big/n}{n} \tag{5}$$

The flowchart of the MCD-BIRCH algorithm is depicted in Fig. 2 and the algorithm of MCD-BIRCH is given in the following steps:

Step 1: Convert data points into CF form $CF = (\eta, lS, sS)$. Where sS indicates the sum of squared distances, η represents the number of data points, lS is the aggregation of attribute values.

Step 2: Use CF-Tree for clustering CFs, specifying the number of branches, β.

Step 3: Set a static CF-tree threshold for new CF entries.

Step 4: Initialize branches on non-leaf CFs (p), number of leaves (ℓ), and leaf branches on CF-leaves (q).

Step 5: Assign each record to the closest clustering feature (CF) at the root node.

Step 6: The process continues by descending to non-leaf nodes, where the record is compared to non-leaf CFs and assigned to the nearest one.

Step 7: Move to non-leaf and then leaf nodes, assigning records to the nearest CF.

Step 8: If the leaf radius, *Rad* is within the threshold *Th*, update the leaf; otherwise, create a new leaf and update CFs.

Step 9: If the number of leaves q exceeds ℓ, add an extra branch β.

Step 10: If β exceeds its limit, split and reorganize CFs, updating the CF structure.

Thereby, the segmented image is indicated as $\aleph^{Seg}$.

3.4. Feature extraction: VGG-16, ResNet, statistical features and MMBP

The feature extraction step is a critical phase in the image analysis workflow, especially in the context of classifying OSCC. This step involves deriving meaningful and quantifiable attributes from the segmented image, $\aleph^{Seg}$ to facilitate accurate classification. The goal of feature extraction is to convert segmented image data $\aleph^{Seg}$ into a set of descriptive features that capture essential features of the images. In this work, the extracted features are elucidated as follows:

3.4.1. VGG-16

VGG-16 (Tammina, 2019) is a well-regarded CNN known for its simplicity and effectiveness in image classification. It features 16 trainable layers, including 13 convolutional layers with 3x3 filters for capturing detailed features and 3 fully connected layers for classification. Max-pooling layers with a 2x2 window reduce spatial dimensions while preserving key information. The VGG-16 network is effective at extracting fine-grained patterns and features from the image. The advantages of VGG-16, it captures fine-grained, low-to-mid level features crucial in distinguishing between different OSCC grades, especially in histopathological image textures. The network's design enables it to efficiently process and classify images by learning both fine-grained and high-level features and the obtained feature is denoted by Vgg^F.

3.4.2. ResNet

Residual Network (ResNet; Liang, 2020) is a deep convolutional neural network architecture designed to train very deep models effectively. Its key feature is residual connections, which skip one or more layers, helping to address vanishing gradients and simplify learning. ResNet uses residual blocks, batch normalization, and ReLU activations, and employs Global Average Pooling (GAP) instead of fully connected layers to reduce parameters and prevent overfitting. ResNet is particularly justified for OSCC classification because the disease's histopathological images often have intricate structures that benefit from deeper networks that can learn complex features. The advantages of ResNet captures high-level abstract features and complex visual patterns that may represent subtle pathological differences (e.g., tumor cell distribution, boundary irregularities). This design allows ResNet to learn complex features and maintain high performance in deep networks and the extracted feature is represented as Res^F.

3.4.3. Statistical features

Statistical features, Sta^F(Esmael et al., 2012) are quantitative measures derived from data that summarize and describe key aspects of a dataset. The use of statistical features is well justified because they complement the deep learning-based feature extraction by providing additional summary information that could help differentiate between benign and malignant lesions.

Min and Max: Represent the smallest and largest pixel values, respectively, which can indicate the range of intensities within the image.

Mean and Median: Offer insights into the average and central pixel values, providing information about the overall brightness and contrast of the image.

Skewness: Measures the asymmetry of the pixel value distribution.

Kurtosis: Indicates the peakedness or flatness of the distribution. Both features help in understanding the texture and contrast of the image.

Moment: Statistical moments (e.g., first moment, second moment) describe the shape and distribution of pixel intensities, providing additional texture and pattern information.

3.4.4. MMBP

The Median Binary Pattern (MBP; Hafiane et al., 2007) derives binary patterns by comparing each pixel's value with the median of its 3×3 neighborhood. With a median value of 120, and including the central pixel in this process, the method produces 29 distinct patterns, as shown in Fig. 3. Then, the MBP descriptor is formulated as in Eq. (6).

$$Mbp = \sum_{i=1}^{M} f(It_i) \times 2^i, \quad where \;\; f(It_i) = \begin{cases} 1, & if\ It_i \geqslant Median \\ 0, & Otherwise \end{cases} \tag{6}$$

where, It_i represents intensity value and M indicates neighbors count.

This approach, however, is vulnerable to changes in pixel arrangement or intensity, particularly in regions where the median value fluctuates. Consequently, it may produce inconsistent patterns, especially in noisy images or those with minor distortions. To tackle this issue, the MBP is modified; then, the modified form of the MMBP descriptor is defined as in Eq. (7).

$$\begin{aligned} {}_{proposed}Mbp &= \sum_{i=1}^{M} f(It_i) \times 2^i \times \tau, \quad where \;\; f(It_i) \\ &= \begin{cases} 1 & if\ Q_{\min} < Q_{median} < Q_{\max} \\ \aleph^{Seg}(i,j) & else\ Q_{\min} < \aleph^{Seg}(i,j) < Q_{\max} \\ 0 & Otherwise \end{cases} \end{aligned} \tag{7}$$

where, $Q_{\max}$ indicates the maximum value of $\aleph^{Seg}(i,j)$; $Q_{\min}$ represents the minimum value of $\aleph^{Seg}(i,j)$; and Q_{median} denotes the median value of $\aleph^{Seg}(i,j)$. Here, τ signifies a small constant value according to Kapur entropy (Ji & He, 2021) and it is defined as in Eq. (8).

$$\tau = \frac{\hbar_i}{\sum_{i=0}^{G-1} \hbar(i)} \tag{8}$$

where, i indicates grey level; $\hbar_i$ refers to pixels count and G signifies some levels. Utilizing Kapur entropy with MMBP leads to better texture differentiation by capturing finer details, thereby improving the capability to distinguish between similar textures.

The MMBP operator is unaffected by monotonic changes in grayscale intensity, as the threshold does not vary with intensity levels. The patterns detected are determined by spatial interactions within the local area. A lack of contrast in a neighborhood results in it being classified as a spot.

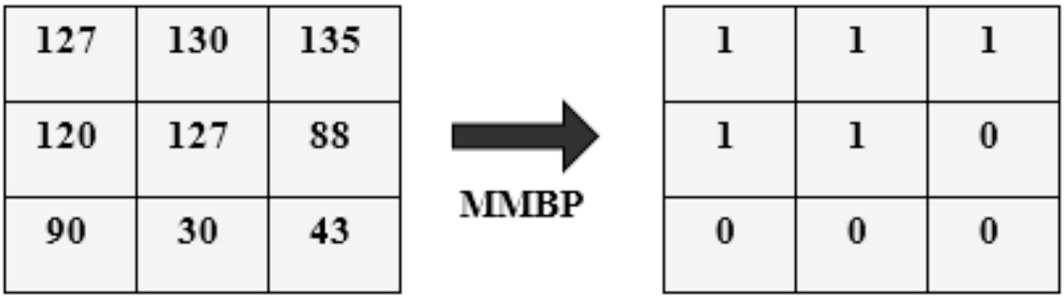

Fig. 3. Exemplary illustration of MMBP with a median value of 120.

The MMBP operator segments the image into two groups of pixels according to the median value, resulting in a defined pattern. Below each image, the output is displayed with hashed regions indicating the value 1. This approach captures contrasts between intensity ranges, impacting local structures and forming the core of texture definition, Mbp^F. Thus, the combined features are represented as $Ext^F = \left[Vgg^F, Res^F, Sta^F, Mbp^F\right]$.

3.5. Classification via a hybrid model

In the classification step for OSCC, the goal is to categorize histopathological images or other relevant data into specific classes. The hybrid model leverages both segmentation and classification, ensuring that the model focuses on relevant regions of the image and processes them effectively using attention mechanisms and sequential dependencies. This combination enhances both the accuracy of segmentation and the precision of classification, making it superior to individual models. This stage comprises Collaborative Attention layer Assisted Bi-LSTM (CAA-Bi-LSTM) and LinkNet model for classification by using extracted feature set Ext^F. The classifiers are trained using a feature set Ext^F as well as offering the intermediate scores. Then, these scores of both the CAA-Bi-LSTM and LinkNet models are averaged to offer the classified output as "well-differentiated, moderately differentiated, or poorly differentiated". During training, the model learns to associate the extracted features with the correct classes by adjusting its parameters to minimize classification errors.

3.5.1. CAA-Bi-LSTM

Hochreiter and Schmidhuber designed the Long Short-Term Memory (LSTM) network, which assists "three gates and two conveyor belts" within a single unit to manage control information flow and the state of each neuron. By using this gating mechanism, precise information transfer is achieved across the four interconnected layers of an LSTM network. To understand the structure and functionality of a Bidirectional LSTM (Bi-LSTM; Abualuroug et al., 2024), it is helpful to first explore unidirectional LSTM networks, which are a type of RNN. The cyclic connections within the hidden layers of traditional RNNs are used to handle short-term memory and sequential data, yet these networks do not feature a clearly defined layer structure. CAA-Bi-LSTM brings sequential processing and attention mechanisms, which help the model capture long-range dependencies and focus on the most relevant features in the image. The Bi-LSTM component, enhanced with a Collaborative Attention Layer (CAA), enables the model to capture long-range dependencies and temporal relationships among sequential feature vectors extracted from different image patches or layers. LSTM networks address the challenge of long-term dependencies by utilizing a memory cell to store important information throughout the network, functioning similarly to a conveyor belt. An LSTM unit comprises a memory cell along with three gates: the forget gate, input gate, and output gate, which regulate the flow of information by determining what to remember and what to ignore. This architecture enables LSTMs to capture enduring relationships and overcome challenges associated with long-term memory. Bi-LSTM layers extend this capability by processing input in both "forward and backward directions", allowing for a more comprehensive understanding of sequential data as in Eq. (9) and Eq. (10), respectively.

$$\lambda_{j,t}^{fwd} = LSTM\left(Ext^F, \lambda_{j,t-1}^{fwd}\right) \tag{9}$$

$$\lambda_{j,t}^{bwd} = LSTM\left(Ext^F, \lambda_{j,t-1}^{bwd}\right) \tag{10}$$

here, the input at the time step t is referred to as Ext^F, while the LSTM cell is represented as $LSTM$. The resultant of the j^{th} LSTM cell, labeled H_{out}, falls within the interval of 1 to 2. The combined forward and backward states generate the final output as in Eq. (11).

$$H_{out} = \left[\lambda_{j,t}^{fwd}, \lambda_{j,t}^{bwd}\right] \tag{11}$$

However, while the current model emphasizes relevant portions of the input, a Bi-LSTM may become prone to overfitting irrelevant patterns in the data. This issue is especially problematic in noisy datasets or when there is significant redundancy within the input sequence. To avoid this problem, a Collaborative Attention layer-assisted Bi-LSTM (CAA-Bi-LSTM) model is proposed, as shown in Fig. 4. This layer is designed to identify complementary features by calculating the attention scores for each feature. The attention mechanism encompasses three stages such as score, attention weights and context vector.

The attention mechanism (Ahmed et al., 2021) helps the decoder concentrate on the most relevant features of the input sequence by calculating a weighted sum of all preceding hidden states. At every time step, the hidden state of the CAA-Bi-LSTM layer, represented as $H_{out} = \left[\lambda_{j,t}^{fwd}, \lambda_{j,t}^{bwd}\right]$, H_{out} is fed into the attention mechanism layer. This process unfolds in three phases: aligning scores, computing weights, and generating the context vector. Eq. (12) describes the score alignment. Here, w represents weight and b signifies bias term.

$$Scr_t = \tanh(H_{out}.w + b) \tag{12}$$

Conventionally, the attention weights Atw_t are derived by applying the Softmax function to the scores Scr_t as in Eq. (13).

$$Atw_t = soft\max(Scr_t) \tag{13}$$

The main limitation of adopting the softmax function is that the network faces a vanishing gradient issue. To overcome this issue, an improved G-Softmax function is proposed in the attention weight function as in Eq. (14).

$$_{proposed}Atw_t = G - soft\max(Scr_t) \tag{14}$$

Traditionally, the G-softmax function is defined as in Eq. (15). The proposed G-Softmax function is formulated as in Eq. (16). Here, σ_i represents standard deviation; Dt refers to distributed team; and μ_i denotes mean.

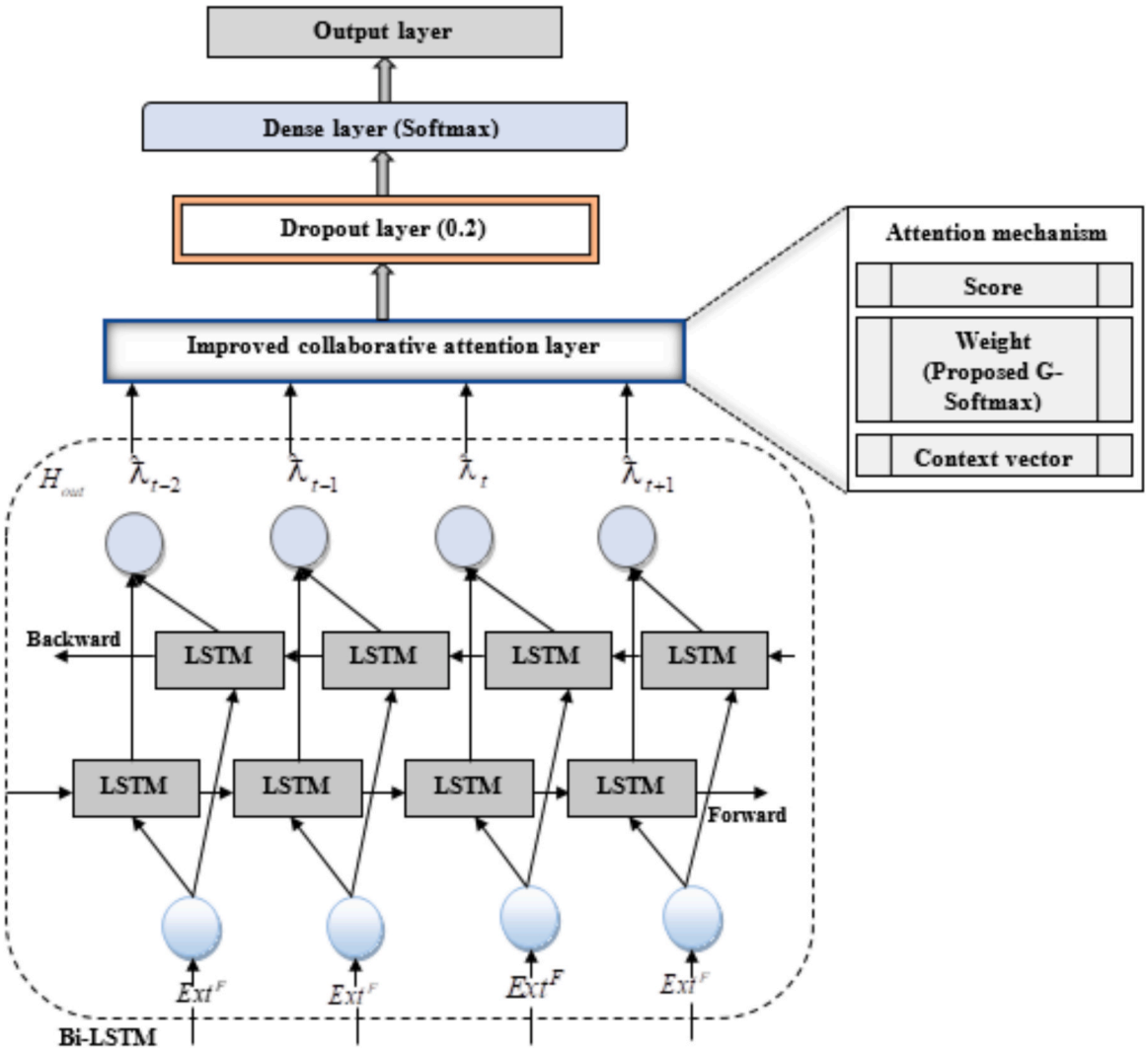

Fig. 4. Diagrammatic representation of CAA-Bi-LSTM.

$$S(x) = \frac{\exp(x_i)}{\sum_{j=1}\exp(x_j)} \tag{15}$$

$$G - Softmax(x) = \frac{\exp(x_i + Dt(x : \mu_i, \sigma_i))}{\sum_{j=1}\exp(x_j + Dt(x : \mu_i, \sigma_i))} \tag{16}$$

where, $Dt(x : \mu_i, \sigma_i) = \lambda * \frac{\sqrt{2}\exp\left(-\frac{(\mu_i - x_i)^2}{2\sigma_i^2}\right)}{2\sqrt{\pi}\sigma_i}$; λ indicates the control parameter which is in the range (0, 1).

After obtaining the attention weights, the context vector referred to as the attention vector is computed as a weighted sum of the J hidden states, according to Eq. (17).

$$Atw_t = \sum_{i=1}^{J} Atw_t \hat{\lambda}_t \tag{17}$$

In the collaborative attention layer, various entities may need to contribute differently based on the context. The proposed G-Softmax facilitates stable training by addressing issues such as vanishing or exploding gradients, resulting in more effective learning of attention weights.

Further, the resultant from the attention mechanism is passed into the dropout layer with a dropout rate of 0.2 and subjected to a dense layer with an activation function, softmax and provides the classified intermediate output.

3.5.2. LinkNet

LinkNet (Chaurasia & Culurciello, 2017) is an advanced deep-learning architecture crafted for semantic segmentation, effectively balancing efficiency and precision in pixel-level classification tasks. It employs an encoder-decoder framework, as shown in Fig. 5: the encoder captures hierarchical features, Ext^F through a sequence of convolutional and pooling layers, progressively reducing the image's spatial dimensions. LinkNet, a lightweight encoder-decoder architecture, excels in semantic segmentation tasks, making it highly effective in preserving spatial resolution and local structures in histopathological images is critical for identifying cancerous regions. The decoder then employs transposed convolutions to upsample these features, restoring the segmentation map to the original image resolution. The inclusion of residual blocks improves training stability by supporting gradient flow while skipping connections between the encoder and decoder preserves detailed spatial information. This streamlined and resource-efficient design enables LinkNet to achieve high performance in real-time scenarios without excessive computational overhead. The network's final classification head generates a per-pixel segmentation map, making LinkNet highly effective for applications like medical image analysis and autonomous driving. The high-resolution outputs from LinkNet allow the system to focus on specific lesion regions in the image, making the classification task more accurate by using the most relevant data.

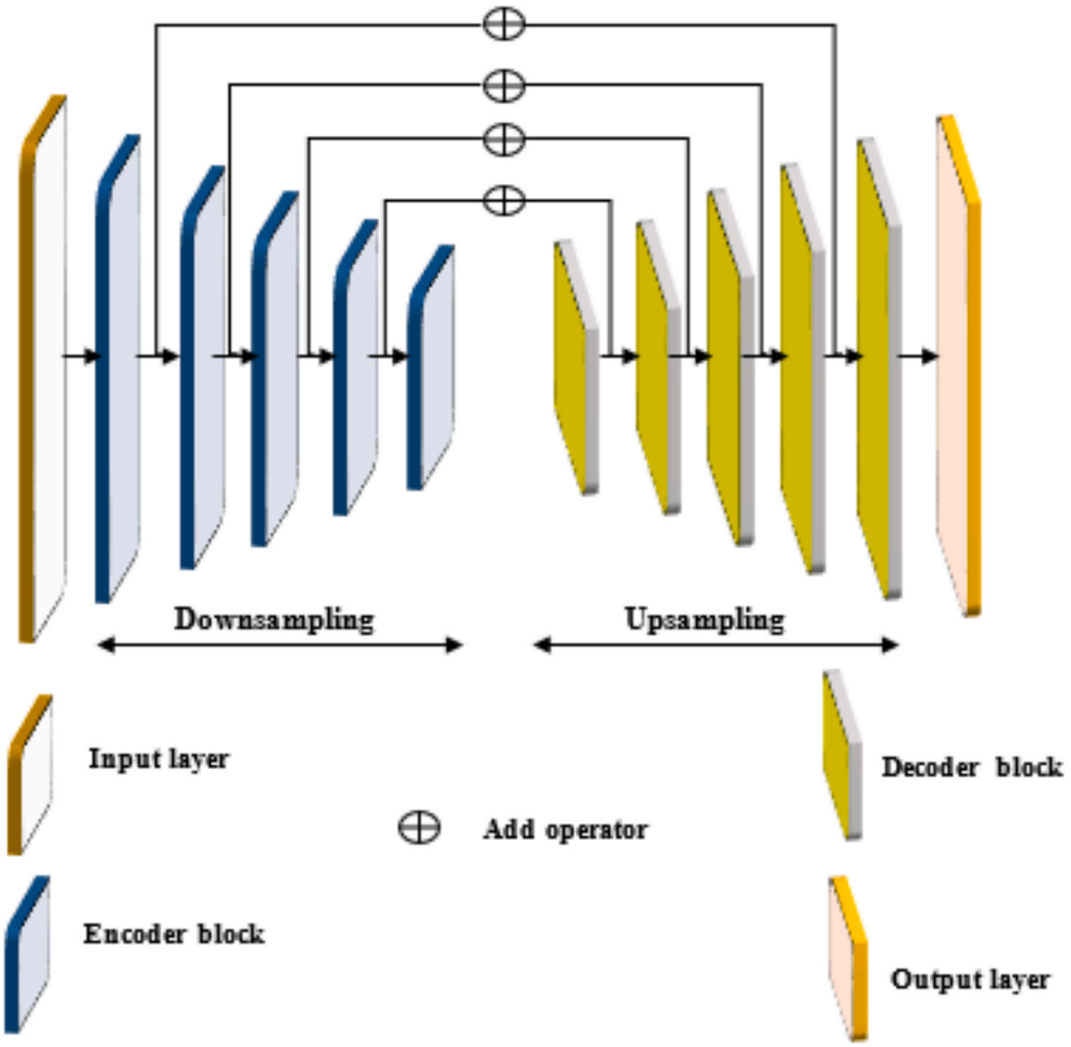

Fig. 5. Structure of LinkNet.

The combination of CAA-Bi-LSTM and LinkNet was purposefully selected to address the dual challenges in OSCC classification: accurately identifying spatially relevant tissue regions and effectively interpreting the complex, contextual relationships between extracted features.

4. Results and discussion

4.1. Simulation procedure

The proposed automated OSCC classification system was implemented in Python 3.7, utilizing an "11th Gen Intel® Core™ i5-1135G7 processor running at 2.40 GHz and equipped with 16.0 GB of RAM.".

4.2. Dataset description: histopathological imaging database for oral cancer analysis

The Histopathological Imaging Database for Oral Cancer Analysis, publicly available on Mendeley Data(https://data.mendeley.com/datasets/ftmp4cvtmb/2),serves as a benchmark dataset for the development and evaluation of automated classification systems. This dataset was clinically validated by Dr. Amit Nisal from Bharati Vidyapeeth Medical College, Pune. It includes histopathological images captured at 100× magnification, primarily used for diagnosing and grading Oral Squamous Cell Carcinoma (OSCC) based on tissue differentiation levels. Initially, the dataset comprised 155 images, with 12 labeled as Well Differentiated (Class 0), 83 as Moderately Differentiated (Class 1), and 60 as Poorly Differentiated (Class 2). To address class imbalance, data augmentation techniques were applied, expanding the dataset to 930 images, with 310 images per class, resulting in a balanced distribution of Well Differentiated (Class 0), Moderately Differentiated (Class 1), and Poorly Differentiated (Class 2) images. This dataset serves as a valuable benchmark for the development and evaluation of automated OSCC classification systems. All images have an input shape of (1536, 2048, 3), indicating the resolution is 1536 x 2048 pixels with three color channels (RGB). This dataset serves as a valuable resource for the automated classification and analysis of OSCC using deep learning techniques. The details of testing and training data samples are tabulated in Table 2.

4.3. Performance analysis

An extensive evaluation was carried out comparing the proposed

Table 2
Training and testing data.

Training percentage	training data	Testing data
60 %	558	372
70 %	651	279
80 %	744	186
90 %	837	93

OSCC classification method with established techniques. The assessment included various metrics such as "Sensitivity, NPV, Specificity, F-measure, FNR, Precision, FPR, MCC, and Accuracy." The analysis included ablation tests and statistical analyses, comparing the CAA-Bi-LSTM + LinkNet approach with state-of-the-art methods like Improved EfficientNetB0 (Soni et al., 2024) and CNN (Sukegawa et al., 2023), and traditional classifiers including Dense Net, Bi-GRU, Link Net, ResNet, and LSTM. The preprocessing analysis of image results is shown in Fig. 6.

4.4. Preprocessing analysis

4.4.1. Segmentation analysis

A comparison of sample photos and their segmented results utilizing the Automated OSCC Classification approaches of Conventional BIRCH, FCM, K-Means, and MCD-BIRCH is shown in Fig. 7.The analysis presented visually underscores how MCD-BIRCH achieves better segmentation results than conventional methods The enhanced results of MCD-BIRCH are evident in its capability to more exactly and effectively delineate the regions of interest, demonstrating its advancement over the traditional methods in image segmentation.

4.4.2. Signal to noise ratio analysis

Table 3 presents a comparison of four filtering methods like Gaussian filter, median filter, adaptive histogram, and Wiener filter. The Gaussian filter outperforms the other methods, achieving the highest PSNR of 36.03 and SSIM of 0.938, indicating it is the most effective at preserving image quality while reducing noise. The median filter follows with a PSNR of 30.34 and SSIM of 0.906, also providing good results, though slightly less optimal than the Gaussian filter. Overall, the Gaussian filter is the most effective for noise reduction and image quality enhancement in this analysis.

Table 4 presents a detailed Segmentation analysis of the MCD-BIRCH strategy, comparing its performance with FCM, K-Means, and Conventional BIRCH methods for automated OSCC classification. The analysis focuses on three key metrics: Dice Coefficient, Jaccard Coefficient, and Segmentation Accuracy. The MCD-BIRCH establishes excellent performance, achieving a Jaccard Coefficient of 83.788, which is higher than FCM (74.608), Conventional BIRCH (71.509), and K-Means (60.988). It also excels in the Dice Coefficient with a score of 85.507, surpassing FCM (67.228), Conventional BIRCH (61.320), and K-Means (60.511). Additionally, it records the highest Segmentation Accuracy at 85.964, compared to FCM (63.953), Conventional BIRCH (74.677), and K-Means (69.128).

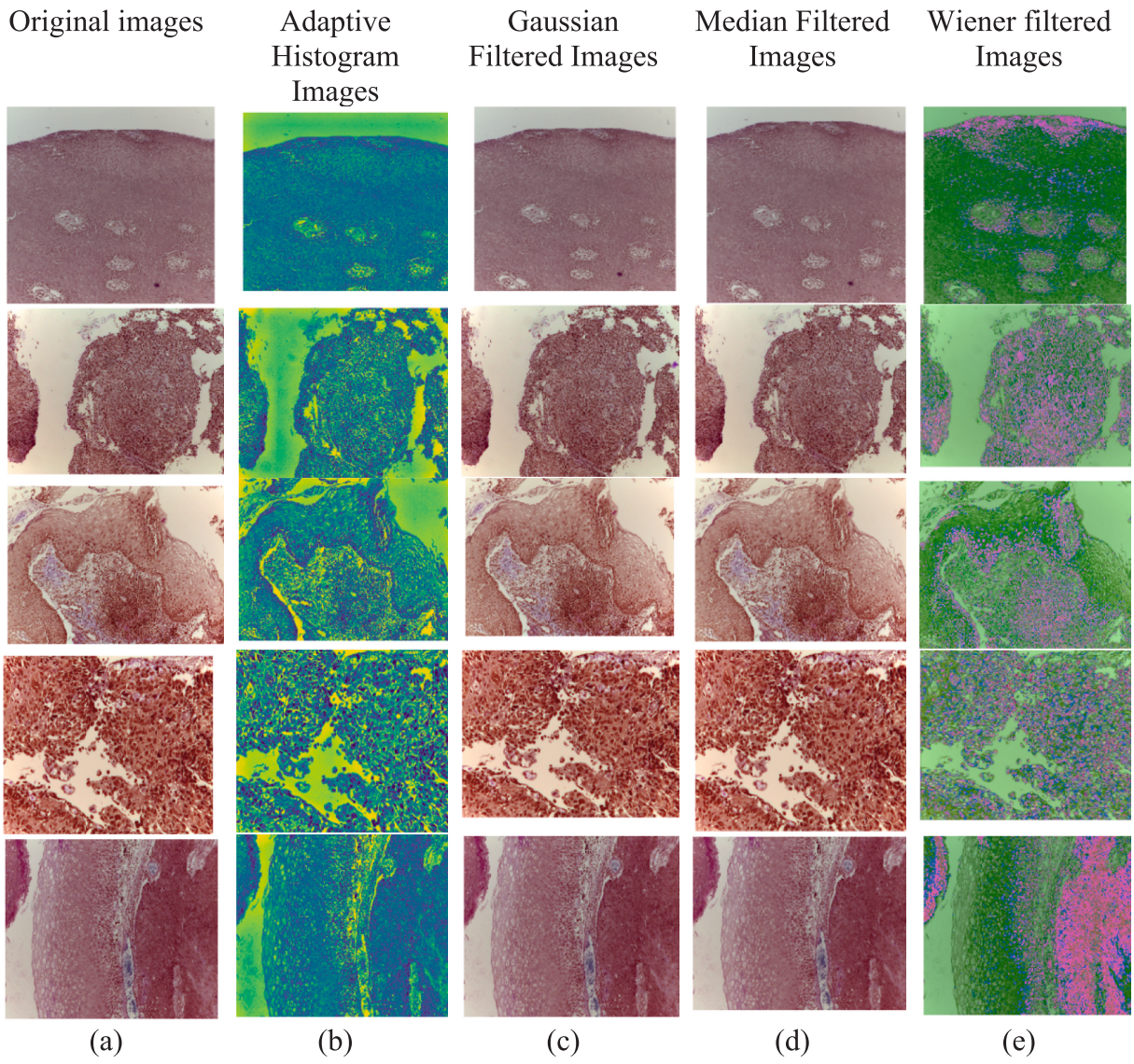

Fig. 6. Images for preprocessing (a) original images (b) adaptive histogram images (c) Gaussian filtered images (d) median filtered images (e) wiener-filtered images.

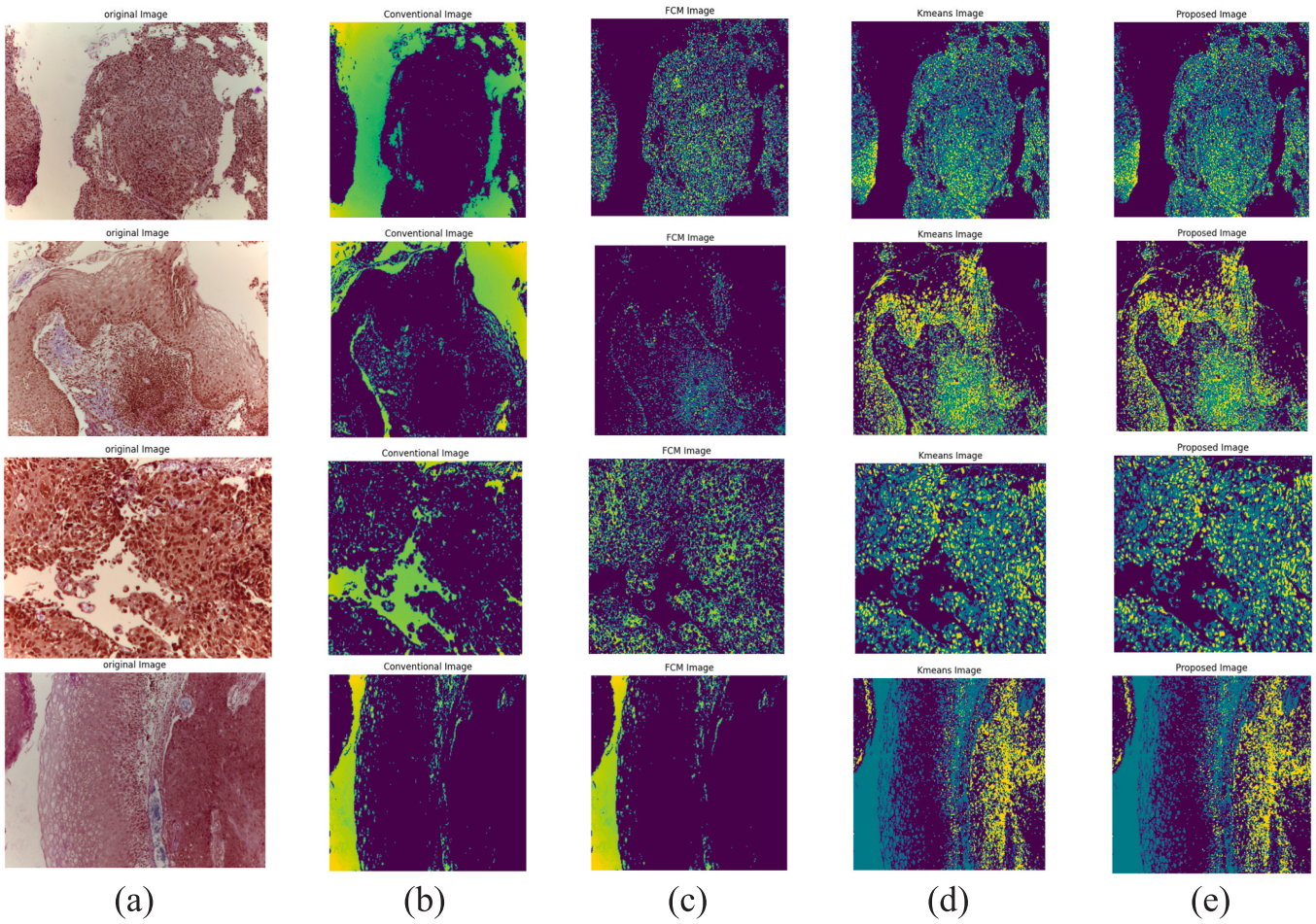

Fig. 7. Images for automated OSCC classification a) sample images b) conventional BIRCH c) FCM d) K-means and e) MCD-BIRCH.

Table 3
Analysis of SNR.

Filtering methods	PSNR	SSIM
Gaussian filter	36.03088	0.938155
median filter	30.33918	0.906156
adaptive histogram	23.79764	0.762441
Wiener	19.31359	0.538652

Table 4
Evaluation of segmentation methods: MCD-BIRCH vs. Conventional Methods, using Dice Coefficient, Jaccard Coefficient, and Segmentation Accuracy.

Measures	FCM	K-Means	Conventional BIRCH	MCD-BIRCH
Jaccard	74.608	60.988	71.509	83.788
Dice	67.228	60.511	61.320	85.507
Segmentation accuracy	63.953	69.128	74.677	85.964

4.4.3. Segmentation output images

Features indexed from 0 to 99 are derived from the ResNet architecture, capturing complex, high-level visual patterns in the image. Features 100 to 149 are extracted using the VGG16 model, emphasizing structural and textural details through a deep learning-based approach. Features 150 to 249 correspond to the Improved Modified Binary Patterns (MBP), which focus on texture analysis by capturing intricate patterns in the image, offering enhanced robustness to noise. Finally, features 250 to 258 are traditional statistical features, which include global image properties like mean, variance, and entropy, providing complementary information to the deep learning-based features for a more comprehensive image representation. These feature classes collectively contribute to a more accurate and detailed classification of Oral Squamous Cell Carcinoma (OSCC) images. Fig. 8 depicts the segmentation output images of different feature classes.

4.5. Classification analysis

4.5.1. Assessment on positive metrics

The effectiveness of the CAA-Bi-LSTM + LinkNet strategy for Automated Classification of OSCC is assessed by comparing its positive metric evaluation against established models, including DenseNet, Bi-GRU, LinkNet, ResNet, 10-layer CNN, ViT, CNN (Sukegawa et al., 2023), LSTM, and Improved EfficientNetB0 (Soni et al., 2024), as illustrated in Fig. 9. The model must have high positive metric values, which indicate its efficacy in precisely identifying and categorizing OSCC cases, in order to achieve robust classification performance. With 60 % training data, it achieves 0.925, leading over conventional methods like ResNet (0.801) and DenseNet (0.776). This trend continues with 70 % data (0.945), 80 % data (0.946), and peaks at 90 % data with 0.986 accuracy. In evaluating precision for automated OSCC classification, the CAA-Bi-LSTM + LinkNet strategy achieves a notable precision score of 0.989 at 90 % of training data. This is significantly higher compared to DenseNet, which scores 0.699, Bi-GRU at 0.968, and ResNet at 0.710. The CAA-Bi-LSTM + LinkNet model also outperforms LinkNet, which has a precision of 0.817, CNN (Sukegawa et al., 2023) at 0.839, LSTM with 0.849, and Improved EfficientNetB0 (Soni et al., 2024), which scores 0.570.

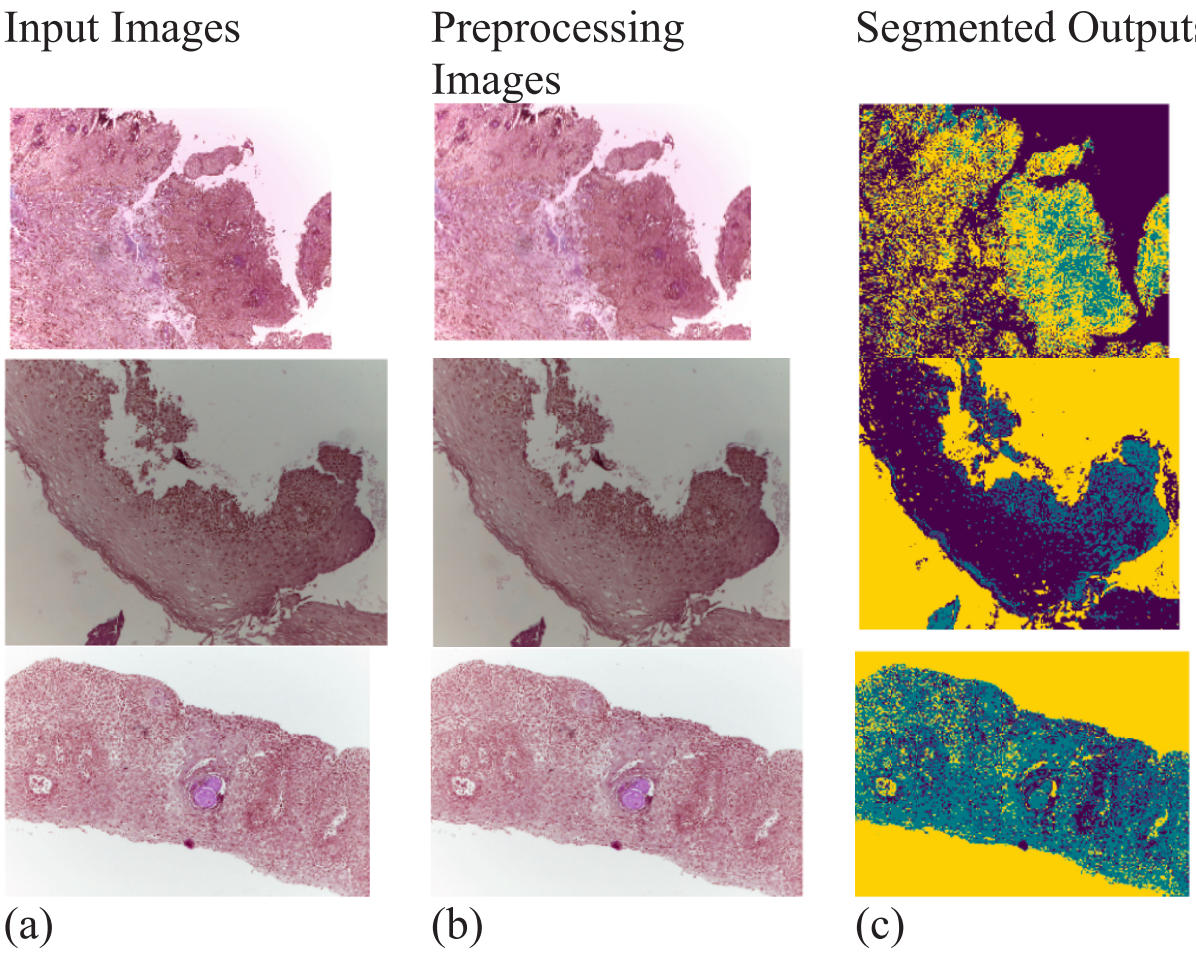

Fig. 8. Different feature of classes (a) input images (b) preprocessing images (c) segmented outputs.

The CAA-Bi-LSTM + LinkNet model demonstrates exceptional specificity across all training data levels. With 60 % of training data, the CAA-Bi-LSTM + LinkNet obtains a specificity of 0.945; as more training data is used, the performance gets better. At 70 %, specificity rises to 0.960, reaching 0.962 with 80 % data. The model peaks at 0.995 specificity with 90 % training data. According to the Positive Metric analysis, the CAA-Bi-LSTM + LinkNet model demonstrates a clear advantage over conventional methods, with higher values across all measured metrics. The use of MCD-BIRCH and MMMBP approaches, which work together to produce more accurate and effective categorization, is associated with this performance gain. The use of a hybrid model further refines the accuracy of the classification process, demonstrating its superiority in classifying OSCC compared to conventional approaches.

4.5.2. Assessment on negative metrics

Fig. 10 provides a comparative analysis of negative metric evaluation for CAA-Bi-LSTM + LinkNet methodology against conventional models, including DenseNet, Bi-GRU, LinkNet, ResNet, 10-layer CNN, ViT, CNN (Sukegawa et al., 2023), LSTM, and Improved EfficientNetB0 (Soni et al., 2024), in the context of Automated Classification of OSCC. To ensure effective classification, the model must accomplish diminished negative measure ratings, which signify fewer errors in OSCC identification. The CAA-Bi-LSTM + LinkNet model, when trained on 60 % of the data, achieves an FNR of 0.115, outperforming DenseNet (0.337), Bi-GRU (0.142), and ResNet (0.299). With 70 % training data, the model's FNR improves to 0.085, remaining lower than DenseNet (0.342), but slightly higher than Bi-GRU (0.078) and still better than ResNet (0.313). At 80 % of training data, the FNR stays at 0.085, continuing to outperform DenseNet (0.314) and ResNet (0.346). In the end, the CAA-Bi-LSTM + LinkNet model achieves its lowest FNR of 0.032 with 90 % training data, nevertheless lagging behind DenseNet (0.316) and Bi-GRU (0.053).

When analyzing FPR for automated OSCC classification with 90 % of the training data, the CAA-Bi-LSTM + LinkNet model exhibits a markedly lower FPR than the other models. The CAA-Bi-LSTM + LinkNet model achieves an FPR of 0.005, significantly lower than DenseNet's 0.152, Bi-GRU's 0.016, and ResNet's 0.147. Other models such as LinkNet have an FPR of 0.092, CNN (Sukegawa et al., 2023) at 0.082, LSTM at 0.076, and Improved EfficientNetB0 (Soni et al., 2024) with 0.217, all of which are higher than the CAA-Bi-LSTM + Link Net model's FPR. Consequently, the Negative Metric analysis reveals that the CAA-Bi-LSTM + Link Net model records lower values compared to traditional methods, reflecting a decrease in error values. This enhancement is largely due to the combination of MCD-BIRCH and MMMBP techniques. Further, the hybrid model approach optimizes classification accuracy, leading to more consistent and effective results.

4.5.3. Assessment on other metrics

Fig. 11 presents a comparison of other metric evaluations for CAA-Bi-LSTM + LinkNet strategy against DenseNet, Bi-GRU, LinkNet, ResNet, 10 layer CNN, ViT, CNN (Sukegawa et al., 2023), LSTM, and Improved EfficientNetB0 (Soni et al., 2024) within the context of Automated Classification of OSCC. This analysis highlights the developed model performs on various other metrics, which are crucial for comprehensive classification performance. For the model to be deemed effective in OSCC classification, it must accomplish higher ratings across these metrics, indicating a more robust and comprehensive capability in precisely recognizing and classifying OSCC cases. This advantage

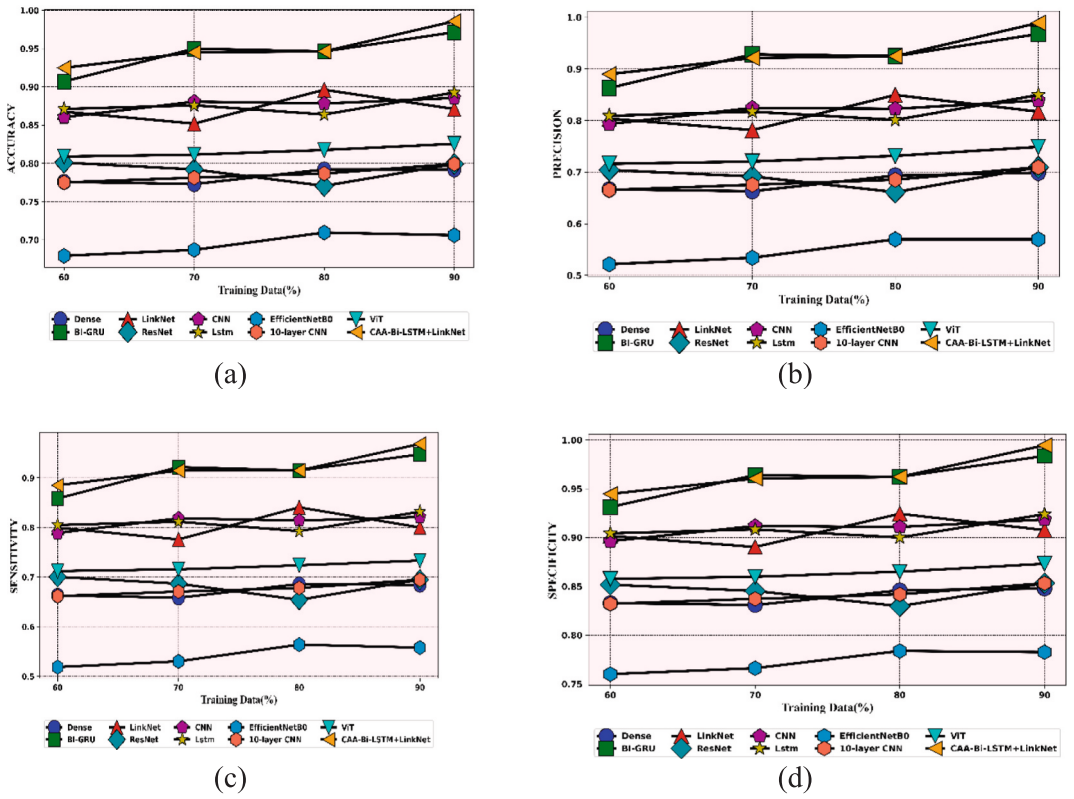

Fig. 9. Assessing positive metrics: a comparative study of CAA-Bi-LSTM + LinkNet and conventional methods a) accuracy b) precision c) sensitivity and d) specificity.

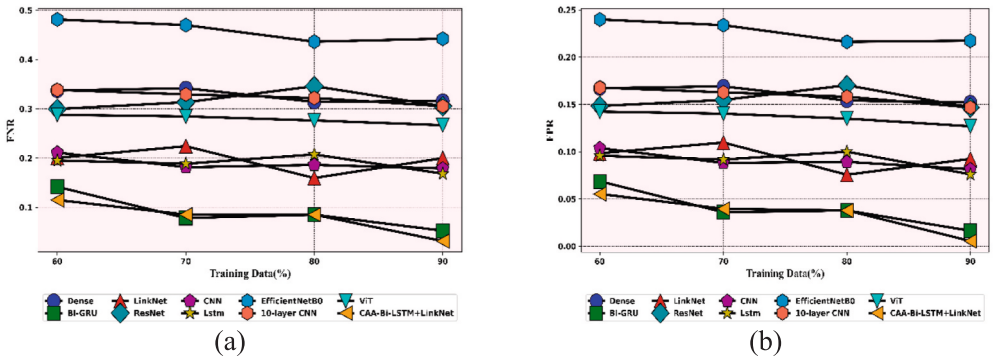

Fig. 10. Assessing negative metrics: a comparative study of CAA-Bi-LSTM + LinkNet and conventional methods a) FNR b) FPR.

continues with 70 % data, where the CAA-Bi-LSTM + LinkNet model reaches 0.918 surpassing ResNet and DenseNet. With 80 % training data, the F-Measure reaches 0.920, leading over ResNet and CNN (Sukegawa et al., 2023). By 90 % of training data, the CAA-Bi-LSTM + LinkNet model achieved an impressive F-Measure of 0.979, which is significantly higher than the values of DenseNet (0.691), Bi-GRU (0.957), LinkNet (0.809), ResNet (0.702), CNN (Sukegawa et al., 2023; 0.830), LSTM (0.840), and Improved EfficientNetB0 (Soni et al., 2024; 0.564).

With 80 % training data, the suggested model exhibits a notably higher MCC of 0.879 compared to conventional methods. DenseNet achieves an MCC of 0.534, Bi-GRU at 0.879, LinkNet at 0.767, ResNet at 0.485, CNN (Sukegawa et al., 2023) at 0.727, LSTM at 0.694, and Improved EfficientNetB0 (Soni et al., 2024) at 0.349, respectively. Hence, the Other Metric analysis shows that the CAA-Bi-LSTM + LinkNet model outperforms conventional models across various measures. This enhancement is attributed to the application of MCD-BIRCH, BMP, and the precise classification capabilities of the hybrid model. These advanced methodologies enhance the model's overall performance, leading to superior results in automated OSCC classification.

4.5.4. Statistical analysis on accuracy

Table 5 compares the performance of several classifiers in the context of OSCC classification using important statistical parameters. Among the models evaluated, the CAA-Bi-LSTM + LinkNet hybrid classifier outperformed all others, achieving the greatest mean accuracy of 0.950, a strong median of 0.946, and a maximum accuracy of 0.986, demonstrating both high effectiveness and consistency. Bi-GRU also performed well, with a mean accuracy of 0.944 and relatively higher variability (Std = 0.023), indicating some fluctuation across runs. Other classifiers such as CNN and LSTM showed competitive performance with mean accuracies of 0.876 and 0.876, respectively, but with slightly higher

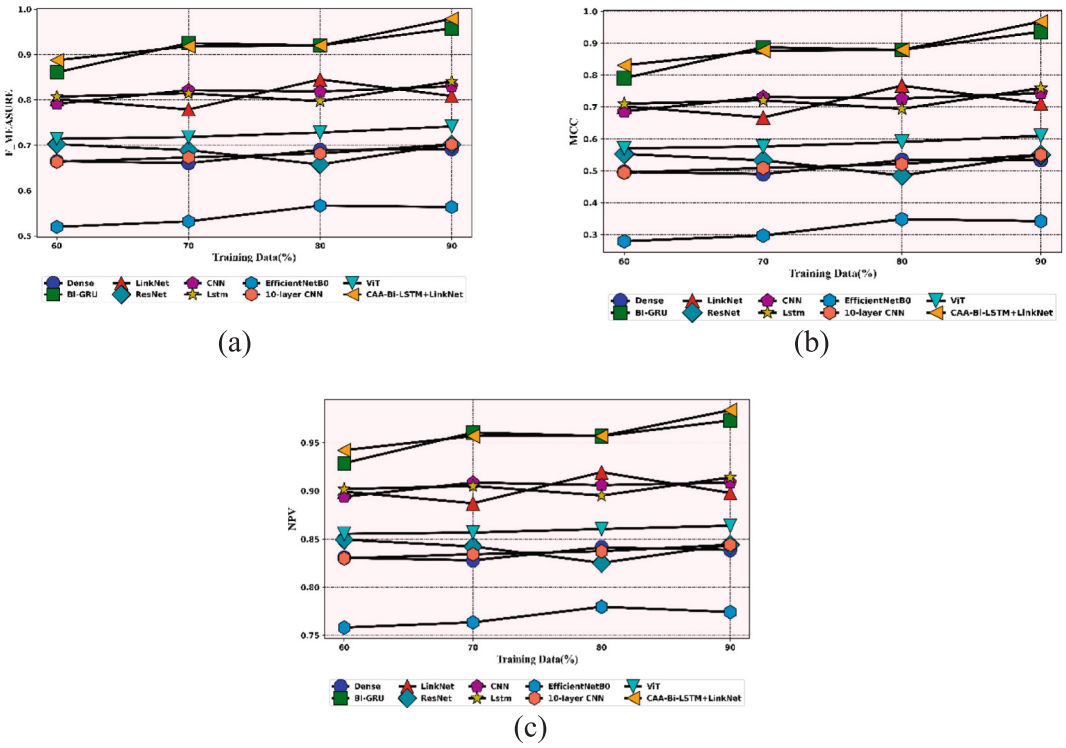

Fig. 11. Assessing other metrics: a comparative study of CAA-Bi-LSTM + LinkNet and conventional methods a) F-measure b) MCC and c) NPV.

Table 5
Statistical evaluation on accuracy.

Classifiers	Mean	Median	Std	Min	Max
Dense	0.783303	0.78405	0.008874	0.772999	0.792115
BI-GRU	0.943548	0.948029	0.02328	0.90681	0.971326
LinkNet	0.871565	0.869176	0.015861	0.851852	0.896057
ResNet	0.790771	0.795699	0.012113	0.770609	0.801075
CNN	0.876045	0.879331	0.009497	0.860215	0.885305
Lstm	0.875747	0.873357	0.010552	0.863799	0.892473
EfficientNetB0	0.69549	0.696535	0.012759	0.679211	0.709677
10-layer CNN	0.785618	0.78405	0.008902	0.77509	0.799283
ViT	0.815873	0.814466	0.006398	0.80899	0.825568
CAA-Bi-LSTM + LinkNet	0.950418	0.945639	0.02207	0.924731	0.985663

standard deviations than top performers. Traditional deep learning models like ResNet, Dense, and 10-layer CNN achieved moderate accuracies (around 0.78–0.79), while EfficientNetB0 exhibited the lowest mean accuracy (0.695), suggesting its relative unsuitability for this task. Overall, the hybrid model emerged as the most robust and accurate classifier for automated OSCC detection.

4.6. Ablation study on CAA-Bi-LSTM + LinkNet

The ablation study of the CAA-Bi-LSTM + LinkNet approach is detailed in Table 6. This analysis aims to evaluate the contributions of different components and methodologies in the Automated OSCC Classification process. The study demonstrates the impact of each component on the effectiveness and accuracy of OSCC classification by reviewing the performance of each configuration. The CAA-Bi-LSTM + LinkNet Approach has a specificity of 0.995, which exceeds that of the Model with Conventional MBP (0.834), the Model with Conventional Segmentation (0.824), and the Model without Segmentation (0.831). Furthermore, the CAA-Bi-LSTM + LinkNet shows an NPV of 0.984, which is significantly higher compared to 0.833 for the Model employing Conventional MBP, 0.823 for the Model utilizing Conventional Segmentation, and 0.829 for the Model without Segmentation.

Table 6
Ablation evaluation CAA-Bi-LSTM + LinkNet, model with conventional MBP, model with conventional segmentation and model without segmentation.

Metrics	A model with conventional MBP	CAA-Bi-LSTM + LinkNet	Model with conventional segmentation	Model without segmentation
Precision	0.668	0.989	0.649	0.662
NPV	0.833	0.984	0.823	0.829
Specificity	0.834	0.995	0.824	0.831
Sensitivity	0.666	0.968	0.647	0.660
FNR	0.334	0.032	0.353	0.340
F-Measure	0.667	0.979	0.648	0.661
Accuracy	0.778	0.986	0.765	0.774
MCC	0.500	0.968	0.472	0.491
FPR	0.166	0.005	0.176	0.169

Table 7 evaluates the individual and combined contributions of key components within the proposed OSCC classification pipeline. By systematically disabling or replacing specific modules—such as pattern-based features (pwc_MBP), segmentation variants (pwc_seg, pwo_seg), deep feature extractors (VGG-16 + ResNet), classification modules (CAA-BiLSTM + Link Net), and segmentation strategy (MCD-BIRCH)—the study reveals their impact on classification performance across multiple metrics. Among all configurations, the proposed full model significantly outperforms others, achieving the highest accuracy (98.57 %), F-measure (97.87 %), precision (98.92 %), and MCC (96.80 %). Similarly, the MCD-BIRCH segmentation method improves overall performance metrics, highlighting the importance of accurate region-of-interest extraction. These results validate that each module contributes meaningfully to the final model's robustness and high diagnostic accuracy.

Table 7
Ablation study of VGG-16 and ResNet), and the hybrid classifier (CAA-BiLSTM + Link Net).

Metrics	pwc_MBP	pwc_seg	pwo_seg	VGG-16 + ResNet	CAA-BiLSTM + Link Net	MCD-BIRCH	proposed
Accuracy	0.777778	0.764957	0.773504	0.769231	0.857412	0.813321	0.985663
Sensitivity	0.666134	0.646965	0.659744	0.653355	0.68541	0.669382	0.968421
Specificity	0.833868	0.824238	0.830658	0.827448	0.84512	0.836284	0.994565
Precision	0.668269	0.649038	0.661859	0.655449	0.701245	0.678347	0.989247
F_measure	0.6672	0.648	0.6608	0.6544	0.71452	0.68446	0.978723
MCC	0.500402	0.471579	0.490794	0.481186	0.58741	0.534298	0.968039
NPV	0.832532	0.822917	0.829327	0.826122	0.832541	0.829331	0.983871
FPR	0.166132	0.175762	0.169342	0.172552	0.1684	0.170476	0.005435
FNR	0.333866	0.353035	0.340256	0.346645	0.2145	0.280573	0.031579

4.7. K fold analysis

Table 8 evaluates the performance of various deep learning models across five different folds, providing an understanding of their consistency and reliability in OSCC classification. The reliability and efficacy of the CAA-Bi-LSTM + LinkNet hybrid model were demonstrated by its consistent top performance across all folds, with accuracy scores ranging from 0.956 to 0.978. With accuracy ratings ranging from 0.872 to 0.899 and 0.862 to 0.888, respectively, CNN and LSTM models also demonstrated strong performance. While BI-GRU and Dense models demonstrated reasonable accuracy, ranging from 0.798 to 0.837 and 0.798 to 0.822, accordingly, LinkNet produced strong results, with scores between 0.855 and 0.881. These models may be less effective for this specific task, as demonstrated by the significantly lower accuracy scores of ResNet and EfficientNetB0, which ranged from 0.791 to 0.815 for ResNet and 0.771 to 0.795 for EfficientNetB0. Overall, the analysis shows that the CAA-Bi-LSTM + LinkNet model performs better than the others, making it the most accurate and dependable option for OSCC classification in this dataset.

4.8. Computational time analysis

Table 9 compares the evaluation of time in OSCC classification. The CAA-Bi-LSTM + LinkNet hybrid model achieved the fastest processing time at 53.33 s, followed by BI-GRU (57.86 s) and LinkNet (56.57 s), demonstrating efficient performance. LSTM and Dense models took 54.63 s and 60.93 s, respectively, while the more complex CNN model required 65.78 s. EfficientNetB0 and ResNet were slower, taking 64.30 s and 73.33 s, respectively, due to their deeper and more intricate architectures. This analysis shows that while more complex models like ResNet and EfficientNetB0 offer high accuracy, they come at the cost of longer computation times, whereas hybrid and simpler models such as CAA-Bi-LSTM + LinkNet strike a better balance between speed and performance.

4.9. AUC-ROC Graph analysis

Fig. 12 shows the ROC curve plots the true positive rate (sensitivity) against the false positive rate (1−specificity) at various classification thresholds, offering insight into the trade-off between sensitivity and specificity. A model with perfect discrimination would yield an AUC-ROC value of 1.0, while a value of 0.5 indicates no discriminative power. In this study, the proposed model achieves a near-perfect AUC-ROC, reflecting its exceptional ability to distinguish between OSCC and non-OSCC cases across all thresholds. This high AUC value confirms that the classifier maintains a strong balance between correctly identifying cancerous cases and minimizing false positives, which is especially critical in medical diagnosis.

Table 8
Analysis of K Fold validation.

Methods	fold-1	fold-2	fold-3	fold-4	fold-5
Dense	0.806452	0.814516	0.816935	0.798387	0.822581
BI-GRU	0.820789	0.828996	0.831459	0.812581	0.837204
LinkNet	0.863799	0.872437	0.875029	0.855161	0.881075
ResNet	0.799283	0.807276	0.809674	0.79129	0.815269
CNN	0.88172	0.890538	0.893183	0.872903	0.899355
Lstm	0.870968	0.879677	0.88229	0.862258	0.888387
EfficientNetB0	0.779534	0.787329	0.789668	0.771739	0.795125
CAA-Bi-LSTM + LinkNet	0.965663	0.97532	0.978217	0.956006	0.971255

Table 9
Time analysis of various models.

Methods	Computing time (sec)
Dense	60.93421
BI-GRU	57.86317
LinkNet	56.57134
ResNet	73.32998
CNN	65.77975
Lstm	54.63258
EfficientNetB0	64.2967
CAA-Bi-LSTM + LinkNet	53.32987

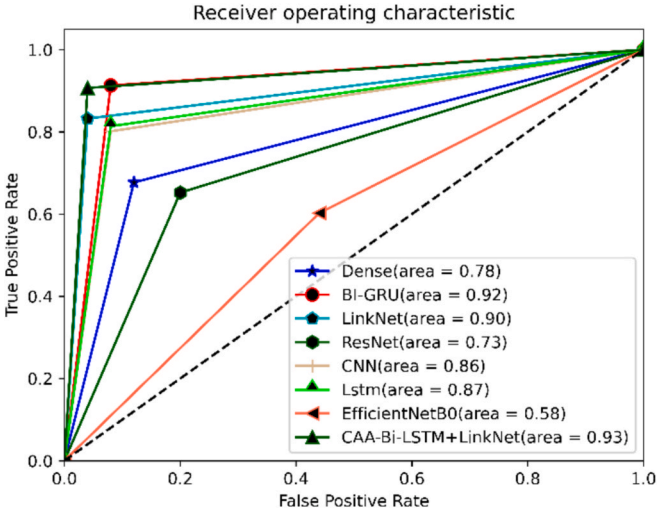

Fig. 12. ROC-AUC analysis.

5. Discussion

The proposed CAA-Bi-LSTM + LinkNet approach for OSCC classification demonstrates promising results, there are several areas where the system could be improved or further explored. One limitation of our work is the dependency on high-quality labeled datasets for training the model. The efficacy of the model remains significantly affected by the quality of the dataset, even though evaluation metrics like accuracy, precision, and F-measure indicate high efficiency. Another potential shortcoming lies in the computational complexity of the proposed

hybrid model. Although the CAA-Bi-LSTM + LinkNet architecture performs well, its heavy computational demand may pose challenges in real-time clinical applications, particularly in resource-constrained environments. The model's large number of parameters requires significant memory and processing power, which may not be feasible for deployment on low-cost or embedded systems commonly found in medical practice.

6. Conclusion

In this paper, a novel Deep Learning model for Oral Squamous Cell Carcinoma Automated Classification (AC-OSCC-DL) was proposed. The study presented an advanced automated classification system for OSCC using a comprehensive pipeline. The preprocessing step involved noise reduction using a Gaussian filter, while segmentation was achieved through the MCD-BIRCH segmentation technique, which enhanced the accuracy of identifying relevant regions within histopathological images. Feature extraction was performed using a combination of VGG-16 and ResNet architectures, alongside statistical features and MMBP, to capture a wide range of image characteristics. The proposed system demonstrated enhanced performance by mitigating traditional limitations and offered a robust, automated approach to OSCC diagnosis, facilitating timely and precise treatment decisions. In evaluating precision for automated OSCC classification, the CAA-Bi-LSTM + LinkNet strategy achieves a notable precision score of 0.989. This is significantly higher compared to DenseNet, which scores 0.699, Bi-GRU at 0.968, and ResNet at 0.710. The system is designed to handle large volumes of data efficiently. The automated system helps in the scalable processing and analysis of numerous samples as healthcare systems embrace digital pathology and electronic health records (EHRs). The MCD-BIRCH segmentation technique, coupled with the use of the Gaussian filter for noise reduction, enhances the accuracy of image segmentation. By identifying and isolating regions of interest (ROI) with higher precision, the system reduces the chances of misinterpreting unclear or ambiguous areas of the image. Furthermore, the model is able to concentrate on the most important features thanks to the utilization of enhanced collaborative attention layers, which further reduces problems caused by inadequate segmentation or fuzzy borders. Addressing integration with clinical workflows, guaranteeing data security, testing the system in various clinical settings, and obtaining regulatory permissions will all be crucial for a successful implementation. If these challenges are removed, the automated approach has the potential to revolutionize OSCC diagnosis and develop into a useful clinical tool, especially one that helps pathologists make better decisions and improve patient outcomes. Future research could focus on further training the proposed model with larger and more diverse datasets from various clinical settings.

- Future work will focus on incorporating advanced cryptographic techniques, including post-quantum cryptography, to ensure the security and privacy of patient data in AI-based healthcare systems, protecting against adversarial attacks and data breaches. In clinical deployment to integrate the model into clinical workflows and validate its performance across real-time diagnostic settings in hospitals and pathology labs.
- For real-time OSCC diagnosis, optimizing computational efficiency and enabling large-scale deployment in clinical settings, while ensuring that the system maintains high accuracy and robustness across diverse datasets. The multi-class pathology extension is expanding the model to support multi-class classification of various oral and head-and-neck pathologies to make it a more comprehensive diagnostic tool.
- An important direction will be the development of explainable AI techniques to improve the interpretability of the OSCC classification system, enabling healthcare professionals to better understand and trust AI-driven diagnosis results in clinical practice. In real-Time Optimization is to enhance computational efficiency to support real-time diagnosis on edge devices or within cloud-based platforms, ensuring accessibility and scalability in resource-constrained environments.

CRediT authorship contribution statement

Anuradha Suresh Pandit: Conceptualization, Methodology, Resources, Data curation. **Vaibhav Vitthalrao Dixit:** Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abualuroug, S., Alzubi, A., & Iyiola, K. (2024). Inventory prediction using a modified multi-dimensional collaborative wrapped bi-directional long short-term memory model. *Applied Sciences, 14*(13), 5817.

Ahmed, S., Saif, A. S., Hanif, M. I., Shakil, M. M., Jaman, M. M., Haque, M. M., Shawkat, S. B., Hasan, J., Sonok, B. S., Rahman, F., & Sabbir, H. M. (2021). Att-BiL-SL: Attention-based Bi-LSTM and sequential LSTM for describing video in the textual formation. *Applied Sciences, 12*(1), 317.

Ali, J. P., Mallick, B. A., Rashid, K., Malik, U. A., Hashmi, A. A., Zia, S., Irfan, M., Khan, A., & Faridi, N. (2024). Diagnostic accuracy of intraoperative frozen section for margin evaluation of oral cavity squamous cell carcinoma. *BMC Research Notes, 17*(1), 43.

Chaurasia, A., & Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE visual communications and image processing (VCIP)*.

Das, N., Hussain, E., & Mahanta, L. B. (2020). Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Networks, 128*, 47–60.

de Lanna, C. A., da Silva, B. N., de Melo, A. C., Bonamino, M. H., Alves, L. D., Pinto, L. F., Cardoso, A. S., Antunes, H. S., Boroni, M., & Cohen Goldemberg, D. (2022). Oral lichen planus and oral squamous cell carcinoma share key oncogenic signatures. *Scientific Reports, 12*(1), 20645.

de Lima, J. M., Macedo, C. C., Barbosa, G. V., Castellano, L. R., Hier, M. P., Alaoui-Jamali, M. A., & da Silva, S. D. (2023). E-liquid alters oral epithelial cell function to promote epithelial to mesenchymal transition and invasiveness in preclinical oral squamous cell carcinoma. *Scientific Reports, 13*(1), 3330.

do Valle, I. B., Oliveira, S. R., da Silva, J. M., Peterle, G. T., Có, A. C., Sousa-Neto, S. S., Mendonça, E. F., de Arruda, J. A., Gomes, N. A., da Silva, G., & Leopoldino, A. M. (2023). The participation of tumor residing pericytes in oral squamous cell carcinoma. *Scientific Reports, 13*(1), 5460.

Eggermont, C., Wakkee, M., Bruggink, A., Voorham, Q., Schreuder, K., Louwman, M., Mooyaart, A., & Hollestein, L. (2023). Development and validation of an algorithm to identify patients with advanced cutaneous squamous cell carcinoma from pathology reports. *Journal of Investigative Dermatology, 143*(1), 98–104.

Esmael, B., Arnaout, A., Fruhwirth, R., & Thonhauser, G. (2012). A statistical feature-based approach for operations recognition in drilling time series. *International Journal of Computer Information Systems and Industrial Management Applications, 4*(6), 100–108.

Goswami, B., Bhuyan, M. K., Alfarhood, S., & Safran, M. (2024). Classification of oral cancer into pre-cancerous stages from white light images using LightGBM algorithm. *IEEE Access, 12*, 31626–31639.

Greeshma, L. R., Joseph, A. P., Sivakumar, T. T., Raghavan Pillai, V., & Vijayakumar, G. (2023). Correlation of PD-1 and PD-L1 expression in oral leukoplakia and oral squamous cell carcinoma: An immunohistochemical study. *Scientific Reports, 13*(1), 21698.

Hafiane, A., Seetharaman, G., & Zavidovique, B. (2007). Median binary pattern for textures classification. *International Conference Image Analysis and Recognition*.

Haq, I. U., Ahmed, M., Assam, M., Ghadi, Y. Y., & Algarni, A. (2023). Unveiling the future of oral squamous cell carcinoma diagnosis: An innovative hybrid AI approach for accurate histopathological image analysis. *IEEE Access, 11*, 118281–118290.

Ji, W., & He, X. (2021). Kapur's entropy for multilevel thresholding image segmentation based on moth-flame optimization. *Mathematical Biosciences and Engineering, 18*(1), 7110–7142.

Kalaivani, K., & Asnath Victy Phamila, Y. (2018). Modified wiener filter for restoring landsat images in remote sensing applications. *Pertanika Journal of Science & Technology, 26*(3).

Kumar, N. A., Dikhit, P. S., Jose, A., Mehta, V., Pai, A., Kudva, A., & Rao, M. (2024). Oral metronomic chemotherapy in advanced and metastatic oral squamous cell carcinoma: A need of the hour. *Journal of Maxillofacial and Oral Surgery, 23*(4), 793–800.

Liang, J. (2020). Image classification based on RESNET. *Journal of Physics: Conference Series*.

Liang, L., Li, Y., Ying, B., Huang, X., Liao, S., Yang, J., & Liao, G. (2023). Mutation-associated transcripts reconstruct the prognostic features of oral tongue squamous cell carcinoma. *International Journal of Oral Science, 15*(1), 1.

Meng, L., Jiang, Y., You, J., Zhao, P., Liu, W., Zhao, N., Yu, Z., & Ma, J. (2023). IRF4 as a novel target involved in malignant transformation of oral submucous fibrosis into oral squamous cell carcinoma. *Scientific Reports, 13*(1), 2775.

Muthupalani, S., Annamalai, D., Feng, Y., Ganesan, S. M., Ge, Z., Whary, M. T., Nakagawa, H., Rustgi, A. K., Wang, T. C., & Fox, J. G. (2023). IL-1 β transgenic mouse model of inflammation driven esophageal and oral squamous cell carcinoma. *Scientific Reports, 13*(1), 12732.

Öhman, J., Zlotogorski-Hurvitz, A., Dobriyan, A., Reiter, S., Vered, M., Willberg, J., Lajolo, C., & Siponen, M. (2023). Oral erythroplakia and oral erythroplakia-like oral squamous cell carcinoma–what's the difference? *BMC Oral Health, 23*(1), 859.

Panigrahi, S., Nanda, B. S., Bhuyan, R., Kumar, K., Ghosh, S., & Swarnkar, T. (2020). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. *Heliyon, 9*(3), 2023.

Panigrahi, S., Das, J., & Swarnkar, T. (2022). Capsule network based analysis of histopathological images of oral squamous cell carcinoma. *Journal of King Saud University-Computer and Information Sciences, 34*(7), 4546–4553.

Raja, N., Ganesan, A., Lakshmi, K. C., & Aniyan, Y. (2024). Assessing DNA methylation of ATG 5 and MAP1LC3Av1 gene in oral squamous cell carcinoma and oral leukoplakia-a cross sectional study. *Journal of Oral Biology and Craniofacial Research, 14*(5), 534–539.

Ramadhani, F., Zarlis, M., & Suwilo, S. (2020). Improve BIRCH algorithm for big data clustering. *IOP conference series: materials science and engineering*.

Sievert, M., Mantsopoulos, K., Mueller, S. K., Rupp, R., Eckstein, M., Stelzle, F., Oetter, N., Maier, A., Aubreville, M., Iro, H., & Goncalves, M. (2022). Validation of a classification and scoring system for the diagnosis of laryngeal and pharyngeal squamous cell carcinomas by confocal laser endomicroscopy. *Brazilian Journal of Otorhinolaryngology, 88*(4), S26–S32.

Soni, A., Sethy, P. K., Dewangan, A. K., Nanthaamornphong, A., Behera, S. K., & Devi, B. (2024). Enhancing oral squamous cell carcinoma detection: A novel approach using improved efficient net architecture. *BMC Oral Health, 24*(1), 601.

Sordi, M. B., Panahipour, L., & Gruber, R. (2023). Oral squamous carcinoma cell lysates provoke exacerbated inflammatory response in gingival fibroblasts. *Clinical Oral Investigations, 27*(8), 4785–4794.

Steybe, D., Poxleitner, P., Metzger, M. C., Rothweiler, R., Beck, J., Straehle, J., Vach, K., Weber, A., Enderle-Ammour, K., Werner, M., & Schmelzeisen, R. (2023). Stimulated Raman histology for histological evaluation of oral squamous cell carcinoma. *Clinical Oral Investigations, 27*(8), 4705–4713.

Sukegawa, S., Ono, S., Tanaka, F., Inoue, Y., Hara, T., Yoshii, K., Nakano, K., Takabatake, K., Kawai, H., Katsumitsu, S., & Nakai, F. (2023). Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists. *Scientific Reports, 13*(1), 11676.

Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP), 9*(10), 143–150.

Tarrad, N. A., Hassan, S., Shaker, O. G., & AbdelKawy, M. (2023). Salivary LINC00657 and miRNA-106a as diagnostic biomarkers for oral squamous cell carcinoma, an observational diagnostic study. *BMC Oral Health, 23*(1), 994.

Tsai, Y. F., Chan, L. P., Chen, Y. K., Su, C. W., Hsu, C. W., Wang, Y. Y., & Yuan, S. S. (2023). RAD51 is a poor prognostic marker and a potential therapeutic target for oral squamous cell carcinoma. *Cancer Cell International, 23*(1), 231.

Wang, S. M., Lu, M. C., Hsu, Y. M., Huang, C. C., Fang, C. Y., Yu, C. H., Chueh, P. J., Yu, S., & Lee, P. (2024). Soft coral *Lobophytum crassum* extract inhibits migration and induces apoptosis capabilities in human oral squamous cell carcinoma cells. *Journal of Dental Sciences*.

Yan, H., Yu, M., Xia, J., Zhu, L., Zhang, T., Zhu, Z., & Sun, G. (2020). Diverse region-based CNN for tongue squamous cell carcinoma classification with Raman spectroscopy. *IEEE Access, 8*, 127313–127328.

Zunino, L., Olivares, F., Ribeiro, H. V., & Rosso, O. A. (2022). Permutation Jensen-Shannon distance: A versatile and fast symbolic tool for complex time-series analysis. *Physical Review E, 105*(4), Article 045310.