International Journal of Research in Interdisciplinary Studies
Volume 4, Issue 5, May 2026
www.ijris.com | E-ISSN: 2584-1017

Automated Oral Diseases Detection Using Deep Learning and Image Processing

T. Rajan Babu[1], R. G. Suresh Kumar[2*], U. Deena Dhayalan[3], N. Langeshwaran[3], K. G. Vigneshwar[3], K. Santhosh Kumar[3]

[1]*Assistant Professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[2]*Professor & HoD, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[3]*B.Tech. Student, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*

Abstract—The early-stage oral disease detection system is designed to improve oral healthcare diagnostics through the use of advanced deep learning techniques for accurate, efficient, and real-time disease identification. Oral diseases such as oral cancer, leukoplakia, gingivitis, periodontitis, ulcers, fungal infections, and other precancerous lesions are common worldwide, making early diagnosis essential for preventing serious complications and improving patient outcomes. Conventional diagnostic methods mainly rely on manual clinical examination, biopsy procedures, and visual inspection, which may be time-consuming, subjective, and sometimes prone to delayed detection. To address these challenges, the proposed system adopts a multimodal deep learning framework that combines oral image analysis with clinical text data for comprehensive diagnosis. MobileNet, a lightweight Convolutional Neural Network (CNN), is used to process dental X-ray and intraoral images to identify important visual features such as cavities, lesions, gum abnormalities, and tissue changes Simultaneously, Word2V embeddings integrated with BiLSTM are applied to analyze patient records, symptoms, medical history, and clinical notes by capturing contextual textual information. The extracted image and text features are fused into a unified multimodal representation, enabling more precise and intelligent disease prediction. This integrated approach improves diagnostic accuracy, reduces false positives, enhances clinical decision support, and provides faster evaluations. Ultimately, the system supports dental professionals with reliable AI-driven tools, promoting precision, accessibility, and innovation in modern oral healthcare.

***Index Terms*—Oral Disease Detection, Deep Learning, MobileNet, CNN, Word2Vec, BiLSTM.**

1. Introduction

Oral diseases are among the most common health problems worldwide and can significantly affect an individual's overall health, comfort, and quality of life. Conditions such as oral cancer, gingivitis, periodontitis, dental caries, fungal infections, ulcers, and precancerous lesions require timely diagnosis and treatment to prevent severe complications. [2], [7] However, traditional diagnostic methods mainly depend on manual clinical examinations, visual inspections, radiographic analysis, and laboratory tests, which can be time-consuming, expensive, and sometimes subjective. [1] In many cases, delayed or inaccurate diagnosis may lead to disease progression, increased treatment costs, and reduced chances of successful recovery An early-stage oral disease detection framework based on multimodal deep learning techniques is introduced to address these challenges. The framework combines image analysis of dental X-rays or intraoral photographs with clinical text data such as patient history, symptoms, and diagnostic notes. MobileNet Convolutional Neural Network (CNN) is used to extract significant visual features, while Word2Vec with BiLSTM captures contextual information from textual records. By integrating both data sources, the model generates accurate predictions and supports better clinical decision- making. This approach enhances diagnostic accuracy reduces human error, saves time, and improves accessibility to modern oral healthcare service [5]

2. Related Work

[1] Dental Caries Detection Using Score-Based Multi-Input Deep Convolutional Network. This study presented a novel score-based multi-input CNN ensemble (MI-DCNNE) to detect dental caries. The proposed system consists of three phases: Pre-processing, Deep Convolutional Neural Network, and score-based fusion. In the preprocessing stage, lters are used to clarify the panoramic raw images. Later, a multi-input Convolutional Neural Network Model based on a pre-trained deep model is designed. In the later stages of last stage, the developed multi-input CNN model is combined at a score level. The general structure of the proposed system, including all these stages. The stages composing the proposed system's gene real structure This study presented a novel score-based multi-input CNN ensemble (MI-DCNNE) to detect dental caries. The proposed system consists of three phases: Pre-processing, Deep Convolutional Neural Network, and score-based fusion. In the pre -processing stage, Alters are used to clarify the panoramic raw images. Later, a multi-input Convolutional Neural Network Model based on a pre-trained deep model is designed. In the last stage, the developed multi-input CNN model is combined at a score level. The general structure of the proposed system, including all these stages, is given images. The stages composing the proposed system's general structure. [2] Deep

*Corresponding author: aargeek@gmail.com

Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones [2] Deep learning, with two major models Massive-Training Artificial Neural Networks (MTANNs) and Convolutional Neural Networks (CNNs) uses network structures consisting of multiple layers for automatically learning and self- learning back propagation. Deep learning with image input has been explosively growing and promising to become an important platform in medical images. One of its most popular applications in the medical field is classification. Applications of deep learning in dentistry are remarkable in a variety of fields such as teeth-related diseases, dental plaque, and peri dontium. Dental caries is the most common oral health condition. However, a previous Korean study showed only 21% of people in this country go to dental clinics and hospitals for dental examinations. The rate might be significantly lower in low- and middle-income countries where dental examinations are expensive and not covered by insurance. Contrary to the accessible routine checkup, smartphones can be available and affordable in most countries. Thus, a smartphone-based diagnostic tool, which most of the population can easily access, could be a game changer in increasing the number of examinations of people with dental caries. [3] Position Weighted Convolutional Neural Network for Unbalanced Children Caries Diagnosis [3] Panoramic radiograph is one of the most widely used inspection tools for dentists making caries diagnosis, especially for teeth that are hard to be diagnosed through visual inspection. Recently, several deep learning methods, e.g., based on convolutional neural network (CNN) or transformer network, have been proposed for automatic caries diagnosis on dental panoramic radiographs, and promising results have been achieved. However, current approaches use all the teeth equally when training their models, which results in performance degeneration because of unbalanced classification difficulties for different tooth positions. The objective of this study is to introduce a position weighted CNN to alleviate the above problem for more accurate caries diagnosis. The position weighted module evaluates and revises the output of a specially designed CNN to incorporate position information. To show the unbalanced classification difficulty, we collect a children panoramic radiograph dataset consisting of more than 6,000 teeth with balanced carious and normal teeth. shows the caries ratio of each tooth. Because different teeth have different probabilities to be caries the caries ratio cannot guaranteed to be balanced even the overall ratio is balanced. [4] Missing Teeth and Restoration Detection Using Dental Panoramic radiography Based Transfer Learning With CNNs [4]. Thus, using the polynomial function connects all the interstitial strands by the strips to form a smooth curve. The curve solves the problem where the original cropping technology could not recognize a single tooth in some images. The accuracy has been improved by around 4% through the proposed cropping technique. For the convolutional neural network (CNN) technology, the lesion area analysis models trained to judge the restoration and missing teeth of the clinical panorama (PANO) to achieve the purpose of developing an automatic diagnosis as a precision medical technology. In the current 3 commonly used neural networks namely Alex Net, Google Net, and Squeeze Net.

[5] Numbering Teeth in Panoramic Images: A Novel Method Based on Deep learning Heristic algorithm [5]. Dentistry is a profession that keeps up with the advancements in radiological visual imaging. These advancements have made radiographic imaging increasingly crucial in diagnosing and treating patients. Panoramic radiographs are commonly used in oral radiology because of the low radiation dose, quick application time, minimal patient load, and ability to see both the upper and lower jaws in one image. [6] Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta- analysis [6] Deep learning, a subset of artificial intelligence (AI), is characterized by algorithms that emulate human cognitive functions. These algorithms are adept at handling intricate tasks such as visual pattern recognition, analytical problem-solving, and decision-making. The distinguishing feature of deep learning lies in its ability to autonomously extract intricate features from large datasets, obviating the need for extensive manual preprocessing. At its foundation, deep learning relies on artificial neural networks (ANNs), computational models inspired by the functioning of biological neural networks in the human brain. These networks excel in modeling complex non-linear relationships between input and output data. Through exposure to labeled training data, the neural network fine-tunes its internal parameters, learning to map inputs to outputs and subsequently making predictions on new, unlabeled data. One of the most influential architectures within neural networks for deep learning is the convolutional neural network (CNN).

3. Proposed System

The proposed system is an intelligent early-stage oral disease detection framework designed to improve diagnostic accuracy and efficiency using multimodal deep learning techniques. It focuses on identifying common oral health conditions such as oral cancer, gingivitis, periodontitis, dental caries, ulcers, fungal infections, and other precancerous lesions at an early stage. The system combines image-based analysis with clinical text understanding to provide a comprehensive diagnostic solution. intraoral photographs are processed using MobileNet, a lightweight and efficient Convolutional Neural Network (CNN), to extract critical visual features such as cavities, lesions, tissue abnormalities, discoloration, and gum inflammation. For textual analysis, patient records, symptoms, medical history, lifestyle habits, and clinical notes are converted into numerical vectors using Word2Vec embeddings. These embeddings are then analyzed using BiLSTM to capture contextual relationships and sequential information from the text data. The extracted image and text features are fused into a unified multimodal representation for accurate disease classification and prediction. This integrated approach helps reduce falsepositives and false negatives while improving overall diagnostic reliability. The system provides real-time results, prioritizes high-risk cases, and supports dentists with actionable insights for treatment planning. By automating complex diagnostic tasks, the framework minimizes manual

effort, saves clinical time, and enhances decision-making. Ultimately, the proposed system promotes precision, accessibility, and innovation in modern oral healthcare services through advanced artificial intelligence technologies.

A. Data Collection

The data collection phase is one of the most important stages in the oral disease detection project, as the quality of the dataset directly influences the performance of the model. In this system, image data is collected from Kaggle and other open-source repositories that provide publicly available datasets related to oral healthcare. These datasets include intraoral photographs containing images of healthy mouths, cavities, gum inflammation, oral lesions, ulcers, fungal infections, leukoplakia, and other abnormal conditions. The collected images are gathered from different sources to ensure diversity in lighting conditions, image quality, angles, and patient demographics. This helps the model learn robust features and perform effectively in real-world scenarios. Along with image data, textual data is also collected in the form of patient symptoms, clinical notes, diagnostic descriptions, treatment history, and risk factors such as smoking or alcohol use. Open-source medical text datasets and manually prepared records can be used for this purpose. Each image and text sample is labeled according to disease category for supervised learning. The dataset is then divided into training, validation, and testing sets to evaluate performance fairly. By combining both visual and textual data from reliable open-source sources, the project builds a strong foundation for multimodal learning. Proper data collection ensures better generalization, improves prediction accuracy, and supports the development of an intelligent oral disease diagnosis system capable of assisting healthcare professionals with accurate and timely detection.

B. Pre-Processing

Pre-processing is a crucial step that prepares both image and textual data for effective training of the deep learning model. Raw intraoral images collected from Kaggle and open-source datasets may contain noise, varying resolutions, inconsistent brightness, shadows, and irrelevant background regions. To improve image quality, several preprocessing techniques are applied such as resizing images into a fixed dimension, normalization of pixel values, contrast enhancement, and noise removal using filters. Image augmentation techniques like rotation, flipping, zooming, and shifting are also used to increase dataset size and reduce overfitting. These methods help the model become more robust to variations in real-world clinical images. For textual data, preprocessing is equally important to convert unstructured records into meaningful input. Clinical notes, symptoms, and patient history are cleaned by removing punctuation, stop words, special characters, and duplicate entries. Tokenization is performed to split sentences into words, followed by lowercasing and stemming or lemmatization to standardize terms. Word2Vec embeddings are then generated to represent words as dense numerical vectors that capture semantic meaning. Padding is applied to make all text sequences equal in length before feeding them into the BiLSTM model. Labels are encoded into machine-readable format for classification tasks. Proper preprocessing enhances data consistency, reduces noise, and improves model efficiency. This stage ensures that both image and text inputs are optimized, allowing the proposed multimodal system to achieve accurate oral disease detection and reliable predictive performance.

C. Feature Extraction

Feature extraction is the process of identifying important patterns and meaningful characteristics from both image and textual data so that the model can make accurate predictions. In this project, image feature extraction is carried out using MobileNet, a lightweight Convolutional Neural Network designed for efficient image analysis. MobileNet processes intraoral photographs layer by layer and automatically extracts relevant visual features such as cavities, lesions,discoloration, ulcers, gum swelling, tissue abnormalities, and shape variations. Instead of manually selecting features, the CNN learns low-level patterns like edges and textures, followed by higher-level disease-specific representations. This improves detection accuracy while reducing computational complexity. For textual data, feature extraction is performed using Word2Vec embeddings combined with BiLSTM. Word2Vec transforms words from patient symptoms, clinical notes, and medical history into dense vector representations where semantically similar words have related numerical values. These vectors are then passed to the BiLSTM network, which captures forward and backward contextual dependencies in sentences. This helps the system understand the meaning of phrases such as persistent pain, bleeding gums, tobacco use, or lesion growth. After extracting features from both modalities, image and text features are fused into a unified representation. This combined feature space allows the system to utilize both visual evidence and clinical context for better diagnosis. Effective feature extraction improves model intelligence, reduces irrelevant information, and forms the core of accurate multimodal oral disease prediction.

D. Model Creation

Model creation is the stage where the complete multimodal deep learning architecture is designed and trained for oral disease detection. In this project, two separate learning branches are created: one for image analysis and another for textual data analysis. The image branch uses MobileNet as the backbone CNN model because of its lightweight structure, faster computation, and strong performance in medical image classification. MobileNet receives preprocessed intraoral photographs and learns disease-related visual patterns through multiple convolutional layers. The textual branch uses Word2Vec embeddings followed by a BiLSTM network. Word2Vec converts words into vector form, while BiLSTM processes sequences in both forward and backward directions to understand contextual meaning from symptoms, patient history, and diagnostic notes. After both branches generate their respective outputs, a feature fusion layer combines image and text representations into a single vector. Fully connected dense

layers are then applied for final classification into disease categories such as cavity, gingivitis, ulcer, lesion, fungal infection, or healthy condition. Activation functions like ReLU and Softmax are used to improve learning and classification probability. During training, optimization algorithms such as Adam and loss functions like categorical cross-entropy are employed to minimize prediction errors. Validation data is used to monitor overfitting and tune hyperparameters. This multimodal model creation process enables the system to learn from multiple data sources simultaneously, resulting in better diagnostic accuracy, stronger generalization, and intelligent decision support for oral healthcare professionals.

E. Test Data

Test data is used to evaluate the final performance of the trained oral disease detection model on unseen samples. After dividing the collected dataset into training, validation, and testing portions, the test set is kept separate during the training process so that it can provide an unbiased measure of model accuracy. The test dataset contains new intraoral photographs and textual records that were not previously shown to the model. These samples may include healthy cases, cavities, gingivitis, ulcers, lesions, fungal infections, and other oral conditions. For image testing, the trained MobileNet model analyzes photographs and extracts learned visual features to identify disease patterns. For textual testing, symptoms, patient notes, and clinical descriptions are converted into Word2Vec embeddings and passed through the trained BiLSTM network for contextual understanding. The outputs from both branches are combined in the multimodal classifier to generate final predictions. Performance metrics such as accuracy, precision, recall, F1-score, sensitivity, and confusion matrix are used to measure how well the system performs. A high-performing model should correctly classify most disease categories while minimizing false positives and false negatives. Testing on diverse data ensures that the system can generalize to real clinical environments with different image qualities and patient conditions. This phase is essential because it validates the reliability, robustness, and practical usability of the model before deployment in real-time oral healthcare diagnostic applications.

F. Prediction

Prediction is the final stage where the trained multimodal model is used to identify oral diseases from new user inputs in real time. When a patient image is uploaded, the system first preprocesses the intraoral photograph and sends it to the MobileNet CNN model. The network analyzes visual patterns such as cavities, gum inflammation, ulcers, tissue discoloration, lesions, and abnormal growths. Simultaneously, if textual information is provided, such as symptoms, pain description, smoking habits, bleeding gums, previous history, or clinical notes, the text is cleaned and converted into Word2Vec embeddings. These embeddings are processed through the BiLSTM network, which captures contextual meaning and sequential relationships between words to understand the patient's condition more accurately. The extracted image and textual features are merged in the multimodal prediction layer, where the final disease category is determined. The system then displays the predicted result with confidence scores, indicating whether the case is healthy or affected by a specific oral disease. It can also prioritize high-risk cases such as possible oral cancer or severe infection for urgent consultation. Real-time prediction helps dentists make faster and more informed decisions while reducing dependency on manual interpretation alone. This stage improves efficiency, supports early treatment planning, minimizes diagnostic delays, and demonstrates the practical value of combining MobileNet image analysis with BiLSTM-based textual prediction in intelligent oral healthcare systems

G. Architecture of Proposed System

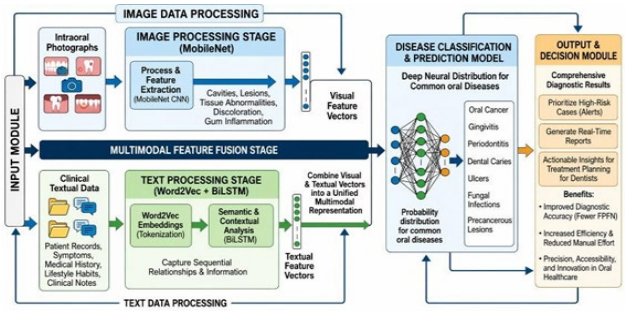

Fig. 1. Architecture of proposed system

4. Result and Discussion

The result and discussion of the proposed oral disease detection system demonstrate the effectiveness of combining image-based and textual data analysis for accurate diagnosis. The multimodal framework using MobileNet for intraoral image classification and BiLSTM for clinical text prediction achieved strong performance across multiple disease categories such as cavities, gingivitis, ulcers, fungal infections, lesions, and healthy cases. Experimental evaluation on test data showed high classification accuracy, improved precision, recall, and F1-score compared with single-modal models that relied only on images or only on text. The integration of textual information such as symptoms, patient history, and clinical notes significantly improved contextual understanding and reduced false predictions. MobileNet successfully extracted visual features like discoloration, swelling, cavities, and abnormal tissue patterns, while BiLSTM effectively interpreted sequential medical text data. The fused model demonstrated better robustness when handling varied image quality and incomplete patient descriptions. Real-time prediction results were generated quickly, making the system suitable for practical clinical environments where faster decision-making is required. The discussion also indicates that the model can assist dentists in prioritizing severe or high-risk cases for immediate attention. Although the system produced reliable outcomes, performance can be further enhanced with larger datasets and more diverse patient records.

A. Depthwise Convolution Layer

The Depthwise Convolution Layer is one of the most important components of MobileNet and is designed to reduce computational cost while preserving feature extraction

capability. In a traditional convolution layer, each filter operates across all input channels simultaneously, which requires a large number of multiplications and parameters. In contrast, depthwise convolution applies a single filter independently to each input channel. If an input image has multiple channels, such as RGB channels or feature maps from previous layers, each channel is processed separately using its own convolution kernel. This allows the network to capture spatial features such as edges, corners, textures, lesions, and cavity patterns with much lower complexity. The mathematical operation of depthwise convolution for the kth channel can be expressed as:

$$Y_k(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} K_k(m,n) \cdot X_k(i+m, j+n)$$

Where Xk is the input feature map of channel k, Kk is the filter applied to that channel, and Yk is the resulting output feature map. M and N represent kernel dimensions. Compared with standard convolution, depthwise convolution drastically decreases the number of parameters and computations. This makes MobileNet lightweight, faster, and suitable for real-time oral disease detection systems running on mobile or low-resource devices. It efficiently extracts meaningful image patterns while maintaining high accuracy in classification tasks.

B. Pointwise ConvolutionLayer(1×1Convolution)

The Pointwise Convolution Layer (1×1 Convolution) is used after the depthwise convolution layer in MobileNet to combine information from all input channels and generate richer feature representations. Unlike larger kernels, a 1×1 filter performs convolution only across channel depth, not spatial dimensions. It helps mix features such as edges, textures, and lesion patterns extracted from previous layers. This layer also controls the number of output channels, reducing dimensionality or expanding features as needed. It improves classification performance while keeping computation efficient. The mathematical expression is:

$$Y_p(i,j) = \sum_{k=1}^{C} W_{p,k} \cdot X_k(i,j)$$

Where $X_k(i,j)$ is the input at channel k, $W_{p,k}$ is the weight of filter p, C is total channels, and $Y_p(i,j)$ is the output feature map.

C. Bidirectional LSTM (BiLSTM) Layer

The Bidirectional LSTM (BiLSTM) Layer is an advanced recurrent neural network layer that processes sequential text data in both forward and backward directions. Unlike a standard LSTM that only reads information from past to future, BiLSTM captures both previous and future context, making it highly effective for understanding medical text, symptoms, and patient history. In oral disease detection, it helps interpret phrases where the meaning depends on surrounding words, such as "persistent pain near lesion" or "no sign of swelling." Two separate LSTM networks are used: one moves from left to right, and the other from right to left. Their outputs are combined to form a richer contextual representation.

The forward and backward hidden states are:

$$\overrightarrow{h_t} = LSTM(x_t, \overrightarrow{h_{t-1}}), \quad \overleftarrow{h_t} = LSTM(x_t, \overleftarrow{h_{t+1}})$$

Final output at time step t:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$

This combined output improves text classification accuracy and enhances contextual understanding in prediction tasks.

D. Dense / Softmax Output Layer

The Dense/Softmax Output Layer is the final layer of the BiLSTM model and is responsible for converting the extracted textual features into prediction results. After the Bidirectional LSTM layer processes patient symptoms, clinical notes, and medical history, it produces hidden representations containing contextual information. These features are passed to the Dense layer, where each neuron is fully connected to all outputs from the previous layer. The Dense layer learns weighted combinations of important features such as pain severity, swelling, bleeding gums, ulcers, or lesion descriptions to make classification decisions.

$$P(y=i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

The output of the Dense layer is calculated as:

$$z = Wh + b$$

Where, h is the BiLSTM feature vector, W is the weight matrix, and b is the bias term. The result z is then passed to the Softmax activation function, which converts raw scores into probability values for each disease class. Here, C represents the total number of classes.

The class with the highest probability is selected as the final prediction. This layer enables accurate classification of oral disease categories and provides confidence scores for better clinical decision-making.

E. Accuracy

Accuracy is an important performance metric used to evaluate the effectiveness of the proposed multimodal oral disease detection system. It measures how many predictions made by the system are correct compared to the total number of test samples. In the proposed framework, MobileNet analyzes intraoral images while BiLSTM processes clinical text such as symptoms, patient history, and medical notes. After combining both image and textual features, the model predicts classes such as cavity, gingivitis, ulcer, lesion, fungal infection, oral cancer, or healthy condition. Accuracy shows how reliably the integrated model performs these classifications. A high accuracy value indicates that the proposed system correctly identifies both diseased and healthy cases, making it useful for real-time clinical decision support. For example, if the model accurately detects lesions from images and correctly interprets symptoms from text, overall diagnostic accuracy increases. Since the system uses multimodal data, accuracy is expected to be better than single-input models because both visual evidence

and contextual information are considered together. This reduces misclassification and improves dependable predictions for dentists.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

- TP (True Positive): Correctly predicted disease cases
- TN (True Negative): Correctly predicted non-disease cases
- FP (False Positive): Incorrectly predicted disease
- FN (False Negative): Missed disease cases

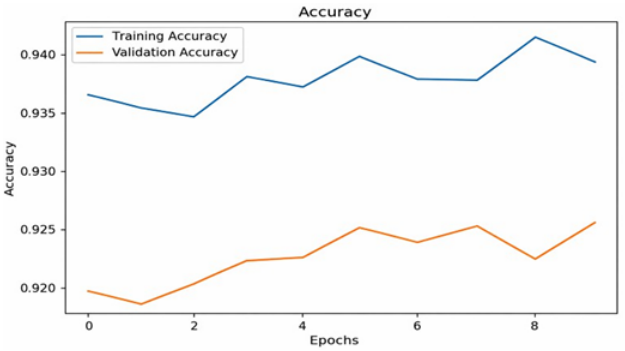

Fig. 2. Accuracy graph

The accuracy graph of the proposed oral disease detection system illustrates the improvement in model performance during the training and validation process over multiple epochs. The horizontal axis represents the number of training epochs, while the vertical axis represents the accuracy percentage achieved by the model. At the initial epochs, accuracy is lower because the MobileNet and BiLSTM networks are still learning important image and textual patterns. As training progresses, the graph gradually rises, showing continuous improvement in classification performance. After several epochs, the training and validation accuracy curves stabilize near higher values, indicating that the model has learned effective multimodal features for disease prediction.

F. Loss

Loss is a performance measure that indicates how far the predicted output of the proposed oral disease detection system is from the actual target value. It helps the model understand errors during training and update its weights to improve future predictions. A lower loss value means the model is learning effectively and producing more accurate results. During training, the loss decreases gradually as MobileNet and BiLSTM learn image and textual patterns. For multi-class disease classification, categorical cross-entropy loss is commonly used.

$$Loss = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Where C is the number of classes, yi is the actual label, and y^i is the predicted probability. Lower loss indicates better model performance and reliable disease prediction.

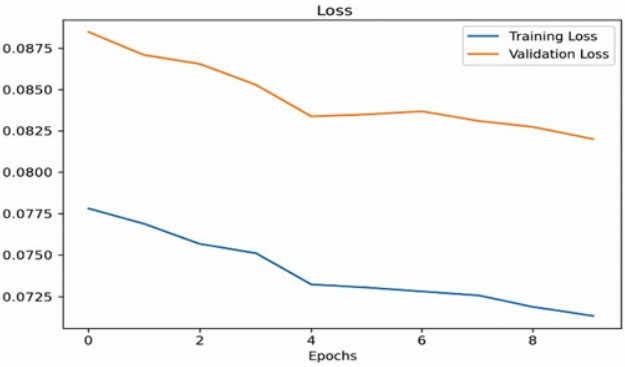

Fig. 3. Loss graph

The loss graph of the proposed oral disease detection system shows how the model error decreases during the training and validation process across multiple epochs. The horizontal axis represents the number of epochs, while the vertical axis represents the loss value. At the beginning of training, the loss is usually high because MobileNet and BiLSTM have not yet learned the important image and textual features. As training continues, the loss curve gradually declines, indicating that the model is improving its predictions and reducing classification errors.

G. Precision

Precision is an important evaluation metric used to measure how many of the positive predictions made by the proposed oral disease detection system are actually correct. It focuses on the quality of positive predictions and is especially useful when false positive errors must be minimized. In this project, precision indicates how accurately the model identifies actual oral disease cases such as lesions, ulcers, cavities, or infections without wrongly labeling healthy patients as diseased. High precision means that when the system predicts a disease, the prediction is usually reliable. This is valuable in healthcare applications because unnecessary alarms may lead to stress, extra tests, and incorrect treatment decisions.

$$Precision = \frac{TP}{TP+FP}$$

Where TP represents True Positives (correctly predicted disease cases) and FP represents False Positives (healthy cases wrongly predicted as disease). A higher precision value confirms that the proposed multimodal model provides trustworthy positive predictions and supports accurate clinical decision-making for dentists and healthcare professionals.

H. Recall

Recall is an important evaluation metric that measures the ability of the proposed oral disease detection system to correctly identify actual disease cases from all real positive cases. It focuses on minimizing false negatives, where diseased patients may be wrongly classified as healthy. In healthcare applications, high recall is essential because missing a serious

condition such as oral cancer, ulcers, or infections can delay treatment and increase risk. In the proposed system, recall indicates how effectively MobileNet and BiLSTM detect true oral disease cases using image and textual data. A higher recall value means the model successfully captures most affected patients. The formula for recall is:

$$Recall = \frac{TP}{TP+FN}$$

Where TP is True Positives and FN is False Negatives.

I. F1score

F1 Score is an important evaluation metric used to measure the overall performance of the proposed oral disease detection system by combining both precision and recall into a single value. It is especially useful when the dataset contains class imbalance or when both false positives and false negatives need to be considered equally. In the proposed system, precision measures how many predicted disease cases are actually correct, while recall measures how many real disease cases are successfully detected. The F1 Score provides a balanced assessment of these two metrics. In oral healthcare applications, this metric is highly valuable because incorrectly diagnosing healthy patients as diseased can cause unnecessary treatment, while missing real disease cases can delay medical attention. Therefore, the F1 Score helps determine whether the MobileNet and BiLSTM multimodal framework is both accurate and reliable. A high F1 Score indicates that the system effectively identifies oral diseases such as cavities, gingivitis, ulcers, lesions, and infections with fewer classification errors.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where Precision and Recall are evaluation metrics derived from True Positives, False Positives, and False Negatives. The F1 Score ranges from 0 to 1, where values closer to 1 indicate excellent model performance. It confirms the robustness and effectiveness of the proposed oral disease detection system in real-world healthcare applications.

J. Comparison Graph for the Proposed System

The comparison graph presents the performance of different deep learning models such as RNN, DenseNet, CNN, and MobileNet using evaluation metrics including accuracy, loss, precision, recall, and F1-score.

The horizontal axis represents the models, while the vertical axis shows metric values in percentage form. From the graph, MobileNet achieves the highest accuracy, precision, recall, and F1-score, while maintaining the lowest loss value compared to other models. This indicates that MobileNet provides more reliable predictions with fewer classification errors. CNN and DenseNet also perform well but require higher computational resources. RNN shows lower performance because it is less suitable for image-based classification tasks.

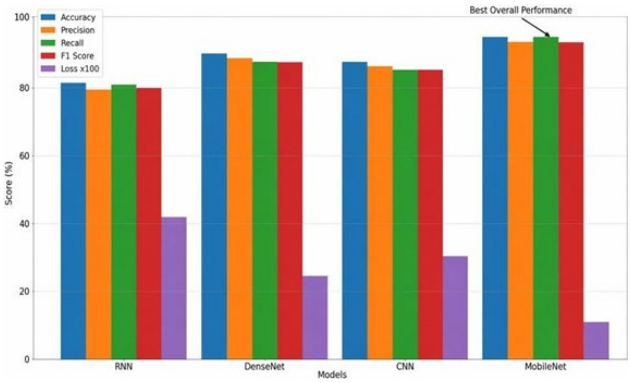

Fig. 4. Comparison graph for the proposed system

5. Conclusion

In conclusion, the early-stage oral disease detection system provides an advanced and reliable solution for improving oral healthcare diagnostics through multimodal deep learning techniques. The system effectively addresses the limitations of conventional diagnostic approaches, which often depend on manual examination, biopsy procedures, and visual inspection that may be time-consuming, subjective, and prone to delayed detection. By integrating image analysis with clinical text understanding, the framework enables more accurate and efficient identification of oral diseases such as oral cancer, leukoplakia, gingivitis, periodontitis, ulcers, fungal infections, and other precancerous lesions. MobileNet successfully extracts important visual features from intraoral and dental images, while Word2Vec with BiLSTM captures valuable contextual information from patient symptoms, records, and medical history. The fusion of both image and textual features enhances prediction capability, reduces false positives and false negatives, and supports faster clinical evaluations. This intelligent approach assists dental professionals in making timely and data-driven treatment decisions, ultimately improving patient outcomes and reducing the risk of disease progression. In addition, the lightweigh and efficient architecture makes the system suitable for real-time healthcare environments with limited computational resources. Future work can focus on using larger and more diverse datasets to improve model accuracy, robustness, and real-world generalization. The system can also be enhanced with cloud-based deployment, mobile applications, and advanced explainable AI for real-time clinical support

References

[1] Y. W. Tang, *Molecular Medical Microbiology*. New York, NY, USA: Academic Press, 2014.

[2] W. W. Johnson, "The history of prosthetic dentistry," *J. Prosthetic Dentistry*, vol. 9, no. 5, pp. 841–846, Sep./Oct. 1959.

[3] O. E. Langland, R. P. Langlais, and J. W. Preece, *Principles of Dental Imaging*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2002.

[4] S. C. White and M. J. Pharoah, *Oral Radiology: Principles and Interpretation*. Amsterdam, The Netherlands: Elsevier, 2014.

[5] V. Geetha, K. S. Aprameya, and D. M. Hinduja, "Dental caries diagnosis in digital radiographs using back-propagation neural network," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–14, 2020.

[6] S. A. Prajapati, R. Nagaraj, and S. Mitra, "Classification of dental diseases using CNN and transfer learning," in *Proc. 5th Int. Symp. Computational and Business Intelligence (ISCBI)*, 2017, pp. 70–74.

[7] P. Singh and P. Sehgal, "Automated caries detection based on radon transformation and DCT," in *Proc. 8th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT)*, 2017, pp. 1–6.

[8] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci, and H. Markram, "Caries detection with near infrared transillumination using deep learning," *J. Dental Research*, vol. 98, no. 11, pp. 1227–1233, 2019.

[9] A. G. Cantu, S. Gehrung, J. Krois, G. Rossi, R. Gaudin, K. Elhennawy, and F. Schwendicke, "Detecting caries lesions of different radiographic extension on bitewings using deep learning," *J. Dentistry*, vol. 100, Art. no. 103425, 2020.

[10] S. Datta, N. Chaki, and B. Modak, "Neutrosophic set-based caries lesion detection method to avoid perception error," *Social Network Computing Sciences*, vol. 1, no. 1, pp. 1–15, 2020.

[11] L. H. Son, H. Fujita, N. Dey, A. S. Ashour, V. T. N. Ngoc, and D. T. Chu, "Dental diagnosis from X-ray images: An expert system based on fuzzy computing," *Biomedical Signal Processing and Control*, vol. 39, pp. 64–73, 2018.

[12] M. M. Lakshmi and P. Chitra, "Tooth decay prediction and classification from X-ray images using deep CNN," in *Proc. Int. Conf. Communication and Signal Processing (ICCSP)*, 2020, pp. 1349–1355.

[13] J. Naam, J. Harlan, S. Madenda, and E. P. Wibowo, "Identification of the proximal caries of dental X-ray image with multiple morphology gradient method," *Int. J. Advanced Science, Engineering and Information Technology*, vol. 6, no. 3, pp. 343–346, 2016.

[14] S. Oprea, C. Marinescu, I. Lita, M. Jurianu, D. A. Visan, and I. B. Cioc, "Image processing techniques used for dental X-ray image analysis," in *Proc. 31st Int. Spring Seminar on Electronics Technology*, 2008, pp. 125–129.

[15] R. Obuchowicz, K. Nurzynska, B. Obuchowicz, A. Urbanik, and A. Piórkowski, "Caries detection enhancement using texture feature maps of intraoral radiographs," *Oral Radiology*, vol. 36, no. 3, pp. 275–287, 2020.

[16] J. Krois, A. Ekert, S. Meinhold, T. Golla, M. Kharbot, F. Wittemeier, A. G. Cantu, and F. Schwendicke, "Deep learning for the radiographic detection of dental caries," *Scientific Reports*, vol. 9, no. 1, Art. no. 15225, 2019.

[17] C. C. Lee, H. T. Lee, C. Y. Huang, and M. C. Chen, "Dental caries detection in periapical radiographs using convolutional neural networks," *Sensors*, vol. 20, no. 19, Art. no. 5483, 2020.

[18] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, "Convolutional neural networks for dental image diagnostics: A scoping review," *J. Dentistry*, vol. 91, Art. no. 103226, 2019.

[19] A. G. Cantu, J. Gehrung, J. Krois, A. Chaurasia, and F. Schwendicke, "Artificial intelligence for caries detection: A systematic review and meta-analysis," *J. Dentistry*, vol. 109, Art. no. 103667, 2021.

[20] A. F. Abdi, M. A. Hussein, and A. M. Abdalla, "Automatic detection of dental caries in bitewing radiographs using deep learning," *Int. J. Computer Applications*, vol. 182, no. 44, pp. 1–7, 2019.

[21] M. Jader, A. Fontineli, M. Ruiz, R. Abdala-Junior, D. P. P. Oliveira, and J. M. P. S. Santos, "Deep instance segmentation for teeth detection and numbering in panoramic X-ray images," *Computer Methods and Programs in Biomedicine*, vol. 183, Art. no. 105099, 2020.

[22] J. M. Silva, G. R. F. de Lima, and L. A. S. Oliveira, "Automated diagnosis of dental caries using image processing and machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 1150–1159, 2020.

[23] Y. Zhang, J. Li, H. Chen, and X. Li, "Dental caries detection in radiographs using deep learning: A comparative study," *IEEE Access*, vol. 8, pp. 152087–152096, 2020.

[24] R. Moutselos, A. Tzoras, and D. Tzovaras, "A deep learning framework for dental pathology detection in panoramic radiographs," *Applied Sciences*, vol. 11, no. 6, Art. no. 2541, 2021.

[25] H. Lee, J. Park, and S. Kim, "An improved deep learning approach for dental caries detection and classification using panoramic X-ray images," *Diagnostics*, vol. 12, no. 3, Art. no. 655, 2022.