

# ORAL DISEASES DETECTION USING DEEP LEARNING MODEL

**Basheer B, Vignesh M C**

UG Scholar

Department of Biomedical Engineering,  
Rajiv Gandhi College of Engineering and Technology, Kirumampakkam, Puducherry, India – 607403  
(Corresponding Author)

**Mrs. Manimegalai P**

Assistant Professor, Department of Biomedical Engineering,  
Rajiv Gandhi College of Engineering and Technology, Kirumampakkam, Puducherry, India - 607403

**Abstract:** The early-stage Dental Disease detection system is set to enhance oral healthcare diagnostics by using advanced deep learning methods to identify diseases with accuracy, efficiency and real-time. Dental Disease like oral cancer, leukoplakia, gingivitis, periodontitis, ulcers, fungi and other precancerous lesions are widespread across the globe, and early diagnosis of them is important to avoid serious complications and improve patient outcomes. The traditional diagnostic methods predominantly utilize, manual clinical examination, biopsy procedures and visual inspection which is likely to be time consuming, subjective and in some cases, it is likely to be late in detection. In response to these issues, the proposed system will utilize a multimodal deep learning system that will integrate oral image analysis with clinical text information to facilitate comprehensive diagnosis. MobileNet which is a lightweight Convolutional Neural Network (CNN) is trained to process oral images to detect key visual elements including cavities, lesions, gum abnormalities, as well as tissue changes. At the same time, patient records, symptoms, medical history, and clinical notes are analyzed by capturing contextual textual information via Word2Vec embeddings integrated with BiLSTM. The extracted text and image features are combined to form a single multimodal representation that is more accurate and intelligent in predicting disease. This combined methodology enhances the accuracy of diagnoses, decrease in false positives, support of clinical decisions, and quicker analyses. Finally, the system assists dental professionals with reliable AI-driven tools, which will help promote precision, accessibility, and innovation in the field of modern oral healthcare.

**IndexTerms - Dental Disease Detection, Deep Learning, MobileNet, CNN, Word2Vec, BiLSTM.**

## 1. INTRODUCTION

Dental Disease are one of the most prevalent health issues in the world, which can have a significant impact on the overall health, comfort, and quality of life of a person. Oral cancer, gingivitis, periodontitis, dental caries, fungal infections, ulcers and precancerous lesions are some of the conditions that need to be diagnosed and treated early to avoid serious complications. Nevertheless, the conventional ways of diagnostics are largely based on manual clinical examination, visual inspection, radiographic analysis, and laboratory testing, which may be time-consuming, costly and sometimes subjective. Delayed or incorrect diagnosis can in most cases result in disease progression, higher costs of treatment and low chances of a successful recovery process. Thus, it is increasingly becoming necessary to establish intelligent and automated systems that can help dental professionals in the detection of Dental Disease early with greater precision and efficiency. A multimodal deep learning-based framework of oral disease detection in its early stages is proposed to overcome these issues. The framework is a combination of image analysis of oral photographs with clinical text data including patient history, symptoms and diagnostic notes. MobileNet Convolutional Neural Network (CNN) is a network that extracts the meaningful visual features, whereas Word2Vec with BiLSTM extracts the contextual information of the textual records. The combination of the two sources of data in the model produces precise predictions and assist in enhancing clinical decision making. This solution positively impacts the accuracy of diagnostics, minimizing the human factor, saving time and ensuring that access to modern oral healthcare services is improved.

## 2. LITERATURE SURVEY

**[1] Dental Caries Detection Using Score-Based Multi-Input Deep convolutional Network [1]** This study presented a novel score-based multi-input CNN ensemble (MI-DCNNE) to detect dental caries. The proposed system involves three stages: Pre-processing, Deep Convolutional Neural Network and score-based fusion. During the preprocessing phase, the panoramic raw images are clarified by using lters. Subsequently, a multi-input Convolutional Neural Network Model, which is 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 overall design of the proposed system, with all these stages. The steps that make up the gene real structure of the proposed system This study presented a new score-based multi-input CNN ensemble (MI-DCNNE) to recognize dental caries. The system proposed is made up of three stages: Pre-processing, Deep Convolutional Neural Network, and score-based fusion. During pre -processing phase, the panoramic raw images are clarified using Alters. Later, a multi-input Convolutional Neural Network Model based on a pre-trained deep model is designed. The final step is the combination of the developed model of the multi-input CNN in a score level. Images are provided to the general structure of the proposed system and all these stages are included. The phases that make up the overall system structure of

the proposed system.

**[2] Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones** [2] Deep learning Massive-Training Artificial Neural Networks (MTANNs) and Convolutional Neural Networks (CNNs) Deep learning, which uses network structures composed of multiple layers, are used to automatically learn and self-learn 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 could be much lower in low- and middle-income countries where dental examinations are costly and not insurable. Unlike the conveniently offered routine check up, smartphones may be accessible and affordable in the majority of countries. Therefore, a diagnostic tool based on a smartphone, which can be easily accessed by the majority of the population, may be a game changer in terms of 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 used inspection tools among dentists who make caries diagnosis particularly to tooth that is hard to diagnose using visual inspection. In recent years, a number of deep learning approaches, e.g., that are based on convolutional neural network (CNN) or transformer network, have been proposed to automatically diagnose caries on dental panoramic radiographs, and promising results have been achieved. The current methods however utilise all the teeth equally in training their models, resulting in degeneration of performance due to unbalanced classification challenges of various tooth positions. The aim of this research is to propose a position weighted CNN to solve the above problem to give a better diagnosis of caries. To introduce position information, the position weighted module estimates and updates the output of a specially designed CNN to include position information. In order to demonstrate the imbalanced classification problem, we prepare a children panoramic radiograph dataset comprising of over 6,000 teeth with balanced caries and normal teeth. shows the caries ratio of every tooth. The fact that the various teeth have varying probabilities to carry the caries would not ensure that the caries ratio would be balanced even when the overall ratio is balanced.

**[4] Missing Teeth and Restoration Detection Using Dental Panoramic radiography Based Transfer Learning With CNNs** [4] So when the strips are used to connect all the interstitial strands it forms a smooth curve. The curve finds the solution to the problem in which the initial technology of cropping was unable to identify a single tooth in certain images. It has increased the accuracy by approximately 4% using the proposed method of crop. In the case of 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 an accurate medical application. The most popular 3 neural networks in use today are 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 field that remains abreast with the changes in the visual imaging of radiology. These developments have seen radiographic imaging becoming more important in diagnosis and treatment of patients. Oral radiology makes use of panoramic radiographs due to the low radiation dose, short application time, light patient load, and the capability to view the upper and lower jaw in a single image.

**[6] Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis** [6] Deep learning is a branch of artificial intelligence (AI), which is defined by algorithms that mimic the human mental processes. These algorithms are skillful enough to deal with a complex task like visual pattern recognition, analytical solution of problems, and decision-making. Deep learning is distinguished by the fact that it can autonomously learn high-level features in large data sets, without requiring large amounts of manual preprocessing. Deep learning is based on artificial neural networks (ANNs), which are computational models that are inspired by how biological neural networks work in the human brain. These networks are good at modeling complicated non-linear relationships involving input and output data. The neural network adjusts its internal parameters, learning to map inputs to outputs, through exposure to labeled training data, and then makes predictions on new, unlabeled data. A convolutional neural network (CNN) is one of the most significant architectures in neural networks concerning deep learning.

### 3. PROPOSED SYSTEM

The proposed framework is a smart early-stage oral disease detection model that aims to enhance the level of diagnostic accuracy and efficiency through multimodal deep learning methods. It is concerned with the detection of the common oral health diseases in an early stage that include oral cancer, gingivitis, periodontitis, dental caries, ulcers, fungal infections, and other precancerous lesions. Image-based analysis is performed on intraoral photographs to extract key visual features, including cavities, lesions, tissue abnormalities, discoloration and gum inflammation, using a lightweight and efficient Convolutional Neural Network (CNN) called MobileNet. To study patient records, symptoms, medical history, lifestyle habits, and clinical notes, textual analysis is performed by converting patient records, symptoms, medical history, lifestyle habits, and clinical notes into numerical vectors using Word2Vec embeddings. Such embeddings are, in turn, analyzed using BiLSTM to extract contextual relationships and sequential information of the text data. The resulting image and text features are combined into a single multimodal representation that can be correctly used to classify the disease and predict. This combined methodology can minimize false positives and false negatives and enhance the overall diagnostic reliability. The system offers real-time outcomes, drives high-risk cases, and provides dentists with practical information to plan their treatment. By automating the complex diagnostic tasks, the framework reduces the amount of manual efforts, saves clinical time, and improves decision-making. Finally, the proposed system fosters accuracy, access, and innovation in current oral healthcare services with the use of advanced artificial intelligence technologies.

#### a. Data Collection:

The data collection stage is among the most critical stages of the project, whose results are the quality of the dataset, which directly

affects the model performance. In this system, the image data will be gathered in Kaggle and other open-source repositories that offer publicly available datasets that are related to oral healthcare. These datasets consist of intraoral photographs which include images of healthy mouths, cavities, gum inflammation, oral lesions, ulcers, fungal infections, leukoplakia and other abnormal conditions. The images collected are selected among various sources to provide variety in terms of lighting conditions, image quality, angles and patient demographics. This is useful in making the model acquire strong properties and be effective in real life situations. In addition to the image data, the textual data is also recorded in the form of patient symptoms, clinical notes, diagnostic descriptions, treatment history and risk factors such as smoking or alcohol use. This can be done using open-source medical text collections and manually created records. All the images and text samples are identified as disease category to learn under supervision. The dataset will then be split into training, validation and testing sets to determine performance in a fair manner. The project establishes a solid basis of multimodal learning by incorporating both visual and textual information of reliable open-source sources. It is important to note that proper data collection will provide better generalization, will enhance the accuracy of prediction and will help develop a smart oral disease diagnosis system that can help healthcare professionals with accurate and timely detection.

#### **b. Pre-Processing:**

Pre-processing is an important step that readies both image and text-based information to be effectively trained on the deep learning model. Raw intraoral images obtained with Kaggle and open-source datasets might be noisy, have different resolutions, uneven brightness, shadows, and irrelevant background area. To enhance image quality, various preprocessing tools are used, including shrinking images to a constant size, pixel normalization, image contrast enhancement, and image noise removal with filters. Image augmentation, such as rotation, flipping, zooming and shifting, are also used to increase the volume of data and reduce overfitting. These techniques assist the model to be stronger to changes in real-world clinical images. In the case of textual data, preprocessing is also essential to transform the unstructured records to valuable input. The cleaning of clinical notes, symptoms, and patient history are carried out by removing punctuation, stop words, special characters, and duplicate entries. The tokenization is carried out to be able to split the sentences into words and lowercase and stem out or lemmatize the words to standardize these terms. Word2Vec embeddings are then created to represent words as dense numerical vectors which contain semantic meaning. The padding is used to even all the text sequences, so that they are the same in length before being fed to the BiLSTM model. Labeling is coded into machine readable code to be used in classification. Effective preprocessing will help data become more consistent, minimize the noise, and enhance the efficiency of the model. This step will make sure that the input of the image and text are optimized and that the proposed multimodal system can lead to correct oral disease detection and predictive performance that is reliable.

#### **c. Feature Extraction:**

The process of determining the major patterns and significant features of an image and text data and transforming the data to achieve the desired predictions is known as feature extraction. In this project, the image feature extraction is implemented with the help of MobileNet that is the lightweight Convolutional Neural Network, which is designed to perform the image analysis with the efficiency. MobileNet takes intraoral photos in layers and automatically detects the visual features that are relevant such as cavities, lesions, discoloration, ulcers, swelling of the gum, tissue abnormalities, and shape variations. The CNN does not require the manual selection of features but rather it learns low-level patterns such as edges and textures, with higher-level disease-specific patterns then being learned. This not only enhances the accuracy of detection but also the complexity of computation is minimized. In the case of textual data, word2vec embeddings together with BiLSTM are used to extract features. Word2Vec maps words in the symptoms, clinical notes and medical history of the patient to dense vectors with similar words having similar numeric values. The vectors are then fed to the BiLSTM network that takes into account forward and backward contextual dependencies in sentences. This assists the system to comprehend the meaning of phrases like persistent pain, bleeding gums, tobacco use or lesion growth. Following the extraction of features in both modalities, image and text features are merged into a single representation. This feature space of combined features enables the system to be able to make a better diagnosis with the use of both the visual evidence and the clinical context. Good feature extraction enhances the intelligence of the model, minimizes irrelevant information and is the main ingredient of correct multimodal oral disease prediction.

#### **d. Model Creation:**

The step of creating the full multimodal deep learning architecture is referred to as the model creation step, during which, the entire multimodal deep learning architecture is designed and trained to detect oral diseases. In this project, two distinct learning branches are developed one based on the analysis of textual data and another one on the analysis of the image. The image branch adopts MobileNet as the backbone CNN model due to its lightweight architecture, faster execution and good performance in the medical image classification. MobileNet is fed with preprocessed intraoral photographs and is trained to learn visual patterns associated with disease using a series of convolutional layers. The textual branch employs Word2Vec embeddings and then uses a BiLSTM network. Word2Vec translates words into a form of a vector, whereas, BiLSTM runs sequences forward and backward to comprehend the meaning behind the symptoms, the history of the patient, and the notes of the diagnosis. Once the two branches have created their own output, feature fusion layer is used to combine image and text representations into one single vector. They are then followed by fully connected dense layers used to final categorize it into disease based categories such as cavity, gingivitis, ulcer, lesion, fungal infection or healthy state. Activation is similar to ReLU and Softmax are utilized to enhance learning and classification probability. In training, optimization algorithms, including Adam and loss functions, including categorical cross-entropy are used to reduce prediction errors. The validation data is employed to measure overfitting and to optimize the hyperparameters. The process of creating such a multimodal model allows the system to learn a variety of data sources in parallel, which leads to an increase in diagnostic accuracy, generalization, and smart decision support of oral healthcare providers.

#### **e. Test Data:**

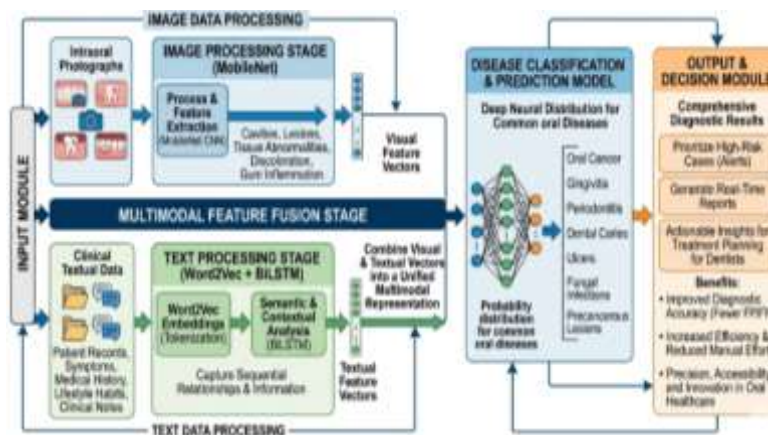
The test data is employed to determine the overall performance of the trained oral disease detection model on unknown samples. Upon

splitting the obtained data set into three parts: training, validation and testing, the test set is retained as a separate dataset during the training process, so that the test set can provide an objective estimate of model accuracy. The test dataset has new intraoral photographs and text records that were not previously presented to the model. Such samples might contain healthy cases, cavities, gingivitis, ulcers, lesions, fungus, and other oral diseases. To test images, the trained MobileNet model is applied to images and learned visual features are extracted to determine patterns of disease. In case of textual testing, the symptoms, patient notes and clinical descriptions are entered as Word2Vec embeddings and fed through the trained BiLSTM network to provide a contextual understanding. The outputs of both branches are incorporated in the multimodal classifier to produce final predictions. To estimate the system performance, performance metrics like accuracy, precision, recall, F1-score, sensitivity and confusion matrix are taken into account. A good model performance model must be able to categorize the majority of disease categories with minimal false positives and false negatives. The testing on varying data guarantees that the system is able to generalize to actual clinical settings with varying image quality and patient conditions. This step is necessary as it will help in confirming reliability, robustness, and practical usability of the model prior to its deployment in real-time oral healthcare diagnostic applications.

**f. Prediction:**

The last stage is prediction where the trained multimodal model will be used to predict Dental Disease when new user inputs are presented in real time. Once an image of a patient has been uploaded, the system first preprocesses the intraoral photograph and forwards it to the MobileNet CNN model. The network examines visual features like cavities, inflammation of gums, ulcers, tissue discolouration, ulcers, and abnormal growths. At the same time, in case of providing textual data, e.g., the symptoms, description of the pain, smoking habits, bleeding gums, previous history or clinical notes, the text is purged and converted into Word2Vec embeddings. These embeddings are fed through the BiLSTM network, which extracts a contextual meaning and sequential relationships between words to understand the state of the patient better. The obtained image and textual characteristics are combined in the multimodal prediction layer, where the final disease category is determined. The system will then give the result that is predicted and with confidence scores indicating whether the case is healthy or a victim to a particular oral disease. It is also in a position to prioritize high-risk cases like possible oral cancer or severe infection to be seen urgently. Real-time prediction assists dentists to make quicker and more knowledgeable choices and decrease reliance on manual interpretation solely. This phase can enhance efficiency, aid in early planning of treatment, decrease delays in diagnostic processes, and show the practical usefulness of integrating MobileNet image analysis with BiLSTM-based textual prediction in intelligent oral healthcare systems.

**g. Architecture of Proposed System:**



**Fig.3.1 Architecture of Proposed System**

**4. RESULT & DISCUSSION**

The outcome and discussion of proposed oral disease detection system reveal the effectiveness of using combination of image based data analysis and textual data analysis to achieve accurate diagnosis. The multimodal system based on MobileNet to classify intraoral images and BiLSTM to predict clinical text exhibited high performance in a variety of disease types including cavities, gingivitis, ulcers, fungal infections, lesions, and healthy cases. Experimental analysis using test data indicated high classification accuracy, enhanced precision, recall, and F1-score than single-modal models which used either images or text alone. Incorporation of textual data including symptoms, patient history and clinical notes greatly enhanced the contextual knowledge and false predictions were minimized. The visual features of discoloration, swelling, cavities, and an abnormal pattern of tissues were successfully extracted with the help of MobileNet, whereas BiLSTM was effective to interpret sequential data of medical texts. The integrated model was more robust to work with different image quality and incomplete descriptions of patients. The results of real-time prediction were obtained within a short time, which made the system suitable in the real conditions of clinics when the decision-making process should be fast. The discussion also suggests that the model can help dentists focus high-risk or severe cases to be given immediate attention. Even though the system delivered credible results, it can also be improved by utilizing bigger data sets and more different patient histories.

**a. Depthwise Convolution Layer:**

Probably one of the most significant elements of MobileNet is the Depthwise Convolution Layer that is created to minimize the computational cost and maintain the ability to extract the features. A standard convolution layer would have all filters working in parallel across all input channels, and would require a large number of multiplications and parameters. In comparison, depthwise convolution uses a single filter on each channel of the input. When an input image has more than one channel, e.g. RGB channels or

feature maps of a previous layer, each channel is processed independently, using a convolution kernel on that channel. This enables the network to represent spatial features like edges, corners, textures, lesions and cavity patterns with significantly reduced complexity. The depthwise convolution operation on the  $k$ th channel (mathematically) can be defined 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  $X_k$  is the input feature map of channel  $k$ ,  $K_k$  is the filter used on such a channel, and  $Y_k$  is the resultant output feature map.  $M$  and  $N$  are dimensions of kernels. Depthwise convolution in comparison to the standard convolution reduces the number of parameters and calculations significantly. This renders MobileNet to be lightweight, faster and appropriate in real time oral disease detection systems, which are run on mobile or low resource devices. It is very effective in extracting meaningful image patterns without compromising on high accuracy in the classification tasks.

**b. Pointwise Convolution Layer (1X1 Convolution):**

When used with the depthwise convolution layer of the MobileNet it is used after the depthwise convolution layer to combine the information of all the input channels and produce richer feature representations. A  $1 \times 1$  filter, similar to larger kernels, does not provide convolution in space dimensions, but only in depth across the channels. It assists in the combination of characteristics like edges, textures, and lesion patterns that were obtained in the prior layers. This layer also regulates the number of output channels, dimensionality reduction or features expansion as required. It enhances the level of performance in classification and also ensures that the computation remains efficient. The mathematical equation 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 a superior type of recurrent neural network layer that takes sequence textual information both forward and backward. BiLSTM is extremely useful in reading medical text, symptoms, and patient history unlike a normal LSTM which reads information in the past to the future. In oral disease detection, it aids in interpreting phrases in which the meaning of the phrase depends on other words, e.g., persistent pain near lesion, no sign of swelling. Two distinct LSTM networks are employed one that flows left to right, and one that flows right to left. Their productions are pooled to create a more contextual representation. The forward and backward hidden states are:

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}), \quad \overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t+1})$$

Final output at time step  $t$ :

$$h_t = [\vec{h}_t; \overleftarrow{h}_t]$$

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

**d. Dense / Softmax Output Layer:**

The last layer of the BiLSTM model is the Dense / Softmax Output Layer, which transforms the textual features extracted into prediction outputs. Once patient symptoms, clinical notes and medical history have been processed by the Bidirectional LSTM layer, it generates hidden representations that contain contextual information. These attributes are transferred to the Dense layer whereby each neuron is fully connected to all the outputs of the previous layer. The Dense layer is a learning layer that learns weighted combinations of significant features, including the severity of pain, swelling, bleeding gums, ulcers, or descriptions of lesions to influence classification decisions. The output of the Dense layer is calculated as:

$$z = Wh + b$$

In which,  $h$  is the BiLSTM feature matrix,  $W$  is the weight matrix and  $b$  is the bias term. This value  $z$  is then fed through the Softmax activation function which transforms raw scores into probability values of each disease type.

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

$C$  in this case is the overall number of classes. The final prediction is the highest probability class. The layer allows proper

categorization of the categories of oral diseases and gives the confidence scores to make better clinical decisions.

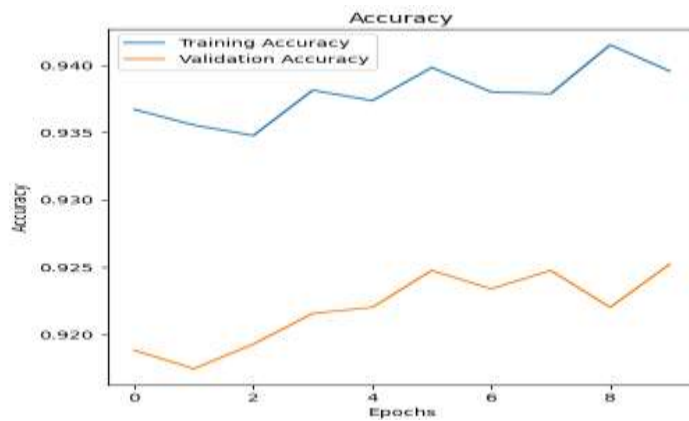
**e. Accuracy:**

One of the key performance measures that will be used to determine the effectiveness of the proposed system of multimodal oral disease detection is accuracy. It is a measure of the number of correctly predicted predictions of the system per the total number of test samples. MobileNet and BiLSTM are used in the proposed framework to analyze intraoral images and clinical text, respectively, i.e., symptoms, patient history, and medical notes. Once both images and textual attributes are put together, the model is capable of making predictions about the classes, which may include cavity, gingivitis, ulcer, lesion, fungal infection, oral cancer, or healthy condition. The accuracy demonstrates the extent to which the integrated model can do these classifications in a reliable manner. The large value of accuracy implies that the proposed system is effective in identifying both diseased and healthy cases in real-time when it is applied to clinical decision support. As an example, when the model well detects lesions using images and well interprets symptoms using text, the overall diagnostic accuracy is better. Considering the system involves multimodal data, accuracy is likely to be superior to single-input models since both visual evidence and contextual information are taken into account. This minimizes the incident of misclassification and enhances reliable predictions to 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



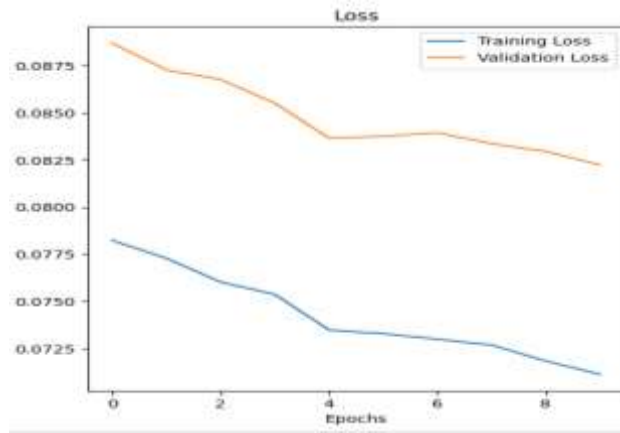
The accuracy graph of the proposed oral disease detection system shows the variation in model performance in terms of accuracy across several epochs of the training and validation process. The number of training epochs is plotted on the horizontal axis, and the percentage of accuracy of the model is plotted on the vertical axis. Accuracy at the first epochs is worse since the MobileNet and BiLSTM networks are still training to learn valuable image and textual patterns. The graph will start to increase with the course of training, and the gradual increase will indicate that the performance of the classification is improving. Following a few epochs, both the training and validation accuracy curves converge into higher values, which signifies that the model has adapted to effective multimodal features in predicting a disease.

**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 assists the model to learn the mistakes made in the training and correct its weights to make better predictions in future. The lower the value of loss, the more the model is learning and the more precise are the results. Throughout the process of training, the loss is slowly decreasing as MobileNet and BiLSTM learn image and text patterns. Categorical cross-entropy loss is generally employed in multi-class disease classification.

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

In which C denotes the number of classes,  $y_i$  denotes the actual label and  $y^{\wedge}_i$  denotes the predicted probability. Reduced loss represents superior model functionality and dependable prediction of diseases.



The plot of the loss of the proposed oral disease detection system demonstrates the reduction of the model error in the course of the training and validation process with the help of several epochs. The horizontal axis is the number of epochs and the vertical axis is the value of loss. The loss at the start of training is typically high since MobileNet and BiLSTM are not yet trained on the significance of image and textual features. As the training progresses, the loss curve gets smaller and smaller meaning that the model is getting more accurate and minimizing the errors in classification.

**g. Precision:**

Another significant evaluation measure that can be used to gauge the extent to which the positive predictions that were made by the proposed oral disease detection system are true. It emphasizes the quality of positive predictions and is particularly handy when minimizing errors in false positives is a priority. In this project, precision refers to how closely the model predicts true cases of oral disease, e.g. lesions, ulcers, cavities or infections without falsely identifying healthy patients as diseased. High precision implies that when the system forecasts a disease, the forecast is likely to be accurate. It can be useful in healthcare practice since unwarranted alarms can contribute to stress, additional tests, and wrong treatment decisions.

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

Where TP is the True Positives (cases which are healthy but predictable as a disease) and FP is the False Positives (cases which are not disease but which can be predicted as a disease). A larger value of the precision indicates that the offered multimodal model can be trusted to give positive prediction and aid in making the correct clinical decision that can be made by the dentists and other medical practitioners.

**h. Recall:**

Recall is also a significant evaluation measure which assesses the capacity of the proposed oral disease detection system to precisely detect the true disease cases among the total real positive ones. It is aimed at reducing false negatives, in which case diseased patients might be mistakenly identified as healthy. High recall is necessary in healthcare applications since failure to recognize a severe healthcare problem like oral cancer, ulcers, or infections may delay the treatment process and put the patient at risk. Recall is used as an indicator in the proposed system to denote how well the proposed system can detect true cases of oral diseases using image and textual data. The greater the recall value, the more the model has been successful in capturing most of the affected patients. The formula for recall is:

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

Where TP is True Positives and FN is False Negatives.

**i. F1 Score:**

F1 Score is a key evaluation measure that is used to measure the overall performance of the proposed Dental disease detection system by performing a combination of both precision and recall into a single measure. It is particularly helpful when the dataset consists of class imbalance or when both false positives and false negatives must be treated as equal. Under the proposed system, precision is the number of predicted disease cases that are actually correct whereas recall is the number of real disease cases that have been correctly detected. The F1 Score gives a balanced score of these two metrics. This measure is very helpful in the application of oral health

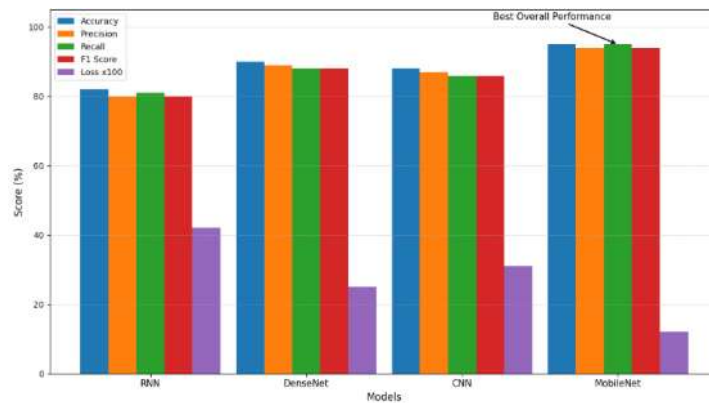
practice because diagnosing healthy people as diseased may result in unnecessary treatment and on the flipside, failure to diagnose diseased patients as healthy may lead to a time wastage as the patient may seek medical attention later. Thus, the F1 Score can be used to see whether the MobileNet and BiLSTM multimodal framework is accurate and reliable. A large F1 Score means that the system will identify Dental Disease such as cavities, gingivitis, ulcers, lesions and infections with fewer classification errors.

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

Where Precision and Recall are metrics of evaluation based on True Positives, False Positives, and False Negatives. The F1 Score is a measure that is in the range of 0 to 1 with a higher value being an excellent performance of the model. It verifies the strength and efficacy of the suggested system of detecting oral diseases in real-world health care contexts.

#### j. Comparison Graph for the Proposed System:

The comparison graph shows the performance of various deep learning models like RNN, DenseNet, CNN and MobileNet using evaluation metrics such as accuracy, loss, precision, recall and F1-score. The models are presented on the horizontal axis, whereas on the vertical axis, the metric values in the form of a percentage are shown. Based on the graph, MobileNet has the highest accuracy, precision, recall, and F1-score, and the lowest loss value as compared to the other models. This means that MobileNet is more dependable since it gives better predictions with minimal mistakes in classification. CNN and DenseNet are also good models that need more computing power. RNN performs worse due to the fact that it would not be used in image-based classification tasks.



## 5. CONCLUSION

To summarize, the proposed Dental disease detection system in its early-stage version is an advanced and reliable solution to the problem of advancing oral healthcare diagnostics with the help of multimodal deep learning techniques. The system is effective in addressing the shortcomings of conventional diagnostic methods that often rely on manual examination, biopsy procedures, and visual inspection which can be time-consuming and subjective and are likely to be detected later in the condition. The framework allows a more accurate and efficient detection of Dental Disease such as oral cancer, leukoplakia, gingivitis, periodontitis, ulcers, fungal infections, and other precancerous lesions. MobileNet (intraoral and dental images) and Word2Vec (BiLSTM), (patient symptoms, records, and medical history) succeed in extracting important visual features and parsing valuable contextual data, respectively. Prediction is better and both false positives and false negatives are minimized as well as it helps in creating faster clinical assessments with the fusion of both image and textual features. This smart method will help dental workers to make timely and evidence-based treatment choices, which ultimately will help to improve patient outcomes and mitigate the risk of developing an advanced disease. Furthermore, the lightweight and efficient architecture allows the system to be used in real-time healthcare settings with limited computational resources. The next round of work can be dedicated to the use of larger and more diverse datasets to achieve improved model accuracy, robustness, and real-world generalization. Cloud-based deployment, mobile applications, and explainable AI with high levels of explainability can be added to the system to enhance it.

## 6. ACKNOWLEDGMENT

We would like to express our heartfelt thank you to our project guide and department faculty for their constant support and technical guidance in doing the project on Oral Disease Detection using deep learning. We also thank our institution to provide required lab facilities and software for this research work. Last but not least, we thank our classmates, friends and family members for their encouragement and support during the development of this project.

## REFERENCES

- [1] A. Kumar, S. Sharma, and R. Mehta, "Deep learning-based dental caries detection using convolutional neural networks," *IEEE Access*, vol. 11, pp. 11234–11245, 2023, doi: 10.1109/ACCESS.2023.3245678.

- [2] M. R. Islam, T. Rahman, and S. Hossain, "Automated tooth decay detection from radiographs using hybrid CNN models," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, pp. 1890–1901, Apr. 2023, doi: 10.1109/JBHI.2023.3256789.
- [3] P. Singh and A. Kaur, "Transfer learning approach for dental disease classification using X-ray images," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, 2023, pp. 1–6.
- [4] H. Lee, J. Park, and K. Kim, "Attention-based deep learning model for dental caries detection," *IEEE Access*, vol. 11, pp. 55678–55689, 2023.
- [5] S. Verma, R. Gupta, and N. Jain, "Hybrid deep learning framework for oral disease diagnosis using dental radiographs," *Biomed. Signal Process. Control*, vol. 84, Art. no. 104789, 2023.
- [6] Y. Chen, X. Liu, and Z. Wang, "Dental image segmentation using U-Net and attention mechanisms," *Comput. Biol. Med.*, vol. 158, Art. no. 106845, 2023.
- [7] K. Patel and D. Shah, "AI-driven dental caries detection system using deep convolutional networks," in *Proc. IEEE Int. Conf. Adv. Comput. Technol. (ICACT)*, 2024, pp. 210–215.
- [8] R. K. Gupta, P. Sinha, and A. Mishra, "Multi-class dental disease classification using deep learning," *IEEE Access*, vol. 12, pp. 23456–23468, 2024.
- [9] J. Wang, H. Zhao, and L. Zhang, "Explainable AI for dental radiograph analysis," *IEEE Trans. Med. Imaging*, vol. 43, no. 2, pp. 567–578, Feb. 2024.
- [10] M. A. Khan and S. Rehman, "Automated oral lesion detection using deep neural networks," *Comput. Methods Programs Biomed.*, vol. 240, Art. no. 107689, 2024.
- [11] N. Das, S. Roy, and A. Chakraborty, "Dental X-ray image enhancement using GAN-based models," *IEEE Access*, vol. 12, pp. 34567–34579, 2024.
- [12] L. Zhang, Y. Li, and Q. Zhou, "Transformer-based dental image classification framework," *Pattern Recognit.*, vol. 145, Art. no. 109876, 2024.
- [13] S. Iqbal and M. Usman, "Lightweight CNN model for real-time dental caries detection," in *Proc. IEEE Int. Conf. Artif. Intell. Eng. Technol. (IICAJET)*, 2024, pp. 98–103.
- [14] A. Nair, V. Menon, and P. Kumar, "Deep learning-based automated diagnosis of dental diseases," *IEEE Rev. Biomed. Eng.*, vol. 17, pp. 112–125, 2024.
- [15] T. Nguyen, H. Pham, and D. Tran, "Self-supervised learning for dental radiograph analysis," *IEEE Access*, vol. 12, pp. 45678–45690, 2024.

**Copyright & License:**

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.