



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Predicting Oral Cancer using a Firefly Optimized CNN-Transformer Model with Clinical Data Integration

Antisar Yunas¹, Hani M Elfadil², Abdulhamid Eltheeb³, Abdullah Abdulhamid⁴

Higher Institute of Medical Sciences and Technologies, Ajdabiya, Libya¹

Computer Science Department, Computing and Information Technology Faculty, Ajdabiya University,
Ajdabiya, Libya²

General Department, Computing and Information Technology Faculty, Ajdabiya University, Ajdabiya, Libya³

Department of Computer Science, Higher Institute of Science and Technology, Zueitina, Libya⁴

ABSTRACT: Oral cancer is one of the major global health issues as it is commonly diagnosed in the last stages because of the ineffective methods of diagnosis. Current methods are predominantly rooted within the core machine learning approach and mutually independent CNNs of which possess high false negative rates and inadequate accuracy when integrating clinical features. This study attempts to meet these gaps by designing the Firefly Optimized CNN-Transformer model for improved the prediction of oral cancer using the image dataset. Methodologically, the research starts with the collection of a large data sample containing oral cancer image data. Preprocessing is applied to this dataset to make the data more general, some of the steps include normalization, augmentation, resizing. The proposed model integrates the CNNs feature extraction aspect in a sequential manner with the transformer models attentions for enhanced feature extraction together with the contextual understanding. For fine-tuning the model parameters, an amelioration of Firefly algorithm is used to substantially enhance the efficiency and effectiveness of the model. Initial baseline analyses indicate that the Firefly Optimized CNN-Transformer achieves impressive performance metrics. Preliminary results indicate that the Firefly Optimized CNN-Transformer model significantly outperforms existing methods. The proposed model is processed in Python on Windows 10, and it's achieving 97.03% accuracy, 97.57% precision, 98.21% recall, and 97.72% F1 score, thereby reducing false negatives. Also, this study shows that extending the transformer architecture to the CNN provides an improvement in predicting oral cancer. They also underscore the significance of embedding clinical data into predictive approaches, further improving the efficiency of logistic computer-based diagnosis techniques. The proposed approach has the potential to improve the effectiveness of early diagnosis and treatment of oral cancer by using image information.

KEYWORDS: Oral cancer, Convolutional Neural Network, Firefly optimizer, Transformer, Kaplan-Meier Survival Curve, deep learning, Text detection, Inpainting, Morphological operations, Connected component labelling.

I. INTRODUCTION

The oral cancer is still widely prevalent and touches a large population, with high mortality rate, and therefore ranks amongst the deadliest forms of cancer known to the world today. According to IARC, there are about 377000 new cases of lip and oral cavity cancer globally in 2020 with nearly 177000 deaths. This raises the question as to why there is such a low awareness for the disease, and why there is a lack of better treatment plans [1]. Emerging approaches in cancer management focus on patient self-governance and several possibilities beyond cure and extent of reaction [2]. Even with the development of newer effective treatment modalities for cancer, oral cancers continue to rank as the sixth common cancer mortality and have a mortality rate of 10.2% [3]. One of the key attributes to this high mortality includes presentation of patients at advanced stages of the disease due to such symptoms as; difficulty in opening the mouth and swallowing [4]. In addition, there are many concerns that patients have about their diagnosis and subsequent treatment, those concerns can affect their decisions as well as their health. These issues are important because early stage diagnosis often does not occur, and majority of patients is diagnosed in the later stages, mainly stages III or IV when the treatment has lesser efficacy [5]. Thus, the five-year survival rate of patients suffering from oral cancer

touches about fifty percent and has been known to reduce or increase depending on relative factors including race and location. Oral Cancer may be of various types; the most common among them is the oral squamous cell carcinoma (OSCC) [6]. And it usually develops from neoplastic lesions called oral potentially malignant disorders or OPMDs; beginning with epithelial dysplasia and ending with invasive carcinoma [7]. This progression emphasizes the need of early diagnosis and treatment preventions since early treatments enormously improve the patients' conditions. Modern diagnostic tools and incorporated approaches to the care of oral cancer patients are crucial to enhance the overall and specific outcomes of cancer patients. [8].

Even in the treatment of oral cancer cases, one can observe a trend towards personalized treatment that comprises a list of outcomes that ranges from simple survival rate and response rate. [4]. Other cancers that involve the oral area are said to cause sixth highest global mortality statistics with countries reporting a rate of 10.2%. Because those patients have severe problems in mouth opening to take foods or swallowing it most of them report their cases when the disease has reached the advanced stage. They also demonstrate concerns regarding their ability to survive the cancer experience and the receipt of future treatment [3]. Mortality associated with oral cancer has been relatively unchanged over the past few decades despite advances in oncology therapy. It is estimated that great proportions of the diagnosed oral cancer patients will not seek the appropriate care when needed. Specifically, they were resulting in low survival in the countryside as observed from the following image. The 5-year survival rate for people whose doctors diagnosed with oral cancer is approximately fifty percent, however, this is a racially and geographically sensitive place [6]. Since early detection is difficult in oral cancer, most patients are diagnosed at stage III or IV disease stage [5]. Since OSCC is the most common form of oral cancer [7], the two acronyms may be used interchangeably; histopathological ally, OSCC is preceded by an epithelial dysplasia and is part of oral potentially malignant disorders (OPMDs). Some of the most common oral prec malignant diseases, oral lichen planus leucoplakia, and erythroleuin do not automatically progress to malignancies [9]. Furthermore, due to their heterogeneity, complexity and clinical similarities OPMDs are particularly difficult to diagnose and classify. The nonhomogeneous lesions are more prone to malignant transformation than the nonhomogeneous ones, suggests the researches [10]. This means that any change in appearance of an OPMD is likely going to be regarded as an undesirable change. Hence, narrowing down the features that maybe malignant in OPMD is of more importance than establishing the diagnosis. For these reasons, it would be inadvisable to rely on a concocted version of the self to predict and control the direction of public discourse and managerial activity. This study introduces assessment as a clinical prediction approach for the early diagnosis of OSCC with focus on staging malignancy within OPMD stage.

Over the most recent years, techniques developed within deep learning paradigms are claimed to be superior to feature-based ones in medical image analysis [11][12]. In a variety of disease classification contexts, the deep learning algorithms have been reported to offer superior performance human specialists, noticed by several authors [13]. Some of the white light inspection techniques employable with smartphones include viable methods of capturing oral images. However, patient sample is small, 54 cases in Camalan et al. [14] proposed CNN-based network for classifying white light images into usual or worrisome. Thus, to build a resilient system, Welikala et al. [15] collected more clinically labelled data for network training and evaluation respectively raising the accuracy for multiclass classifications to 52.13% while achieving a specificity of 49.11% Only. Nevertheless, this research found that clinical images taken using a smartphone's handheld camera contain large variations, which degrade the identification algorithm performance in diagnostic scenarios. One major challenge is early present with symptoms, many a time being asymptomatic or having nonspecific symptoms, which leads to staging of disease leading to delayed treatment, and consequently poor outcomes. Also, inadequacy of health facility, particularly in vulnerable areas, aggravates the disparities in diagnosing and managing oral cancer [16]. Thus, issues in the molecular mechanism of cancer and the multiple gene-environment and life style interactions make the choice and selection of specific molecular targets and individual management strategies difficult. In addition, the high recurrence rates and therapeutic failure rates also highlight the need for new treatment modalities improvements in treatments and specific precision medicines to increase the OS and reduce the detrimental health impact and mortality associated with oral cancer [17]. These challenges require a three-pronged strategy that will require collective effort from clinicians, academicians and researchers, policymakers as well as scaled up public health advocates.

The motivation of the proposed study is to enhance the accuracy and efficiency of oral cancer prediction by combining image data with clinical data in a deep learning model. While CNNs can effectively analyse medical images and Transformers capture complex dependencies, their optimal performance requires precise tuning. The study intends to

solve this challenge by applying the Firefly Optimization Algorithm on a proposed CNN-Transformer model. The study with clinical data input including patient history and demographics will improve primary identification of Oral cancer and patient outcomes complementing image analysis. The key contribution of the proposed study is given below,

- The study proposes a CNN-Transformer hybrid model to enhance oral cancer prediction by integrating image features and clinical data for improved accuracy.
- The Firefly Algorithm optimizes hyperparameters in the CNN-Transformer model, enhancing performance, accuracy, and generalization in oral cancer prediction.
- The study integrates medical images and clinical records for comprehensive oral cancer prediction, improving accuracy and understanding of cancer progression through multimodal data.
- The study improves early detection by identifying patterns between image features and clinical data, facilitating timely interventions and personalized risk assessment for tailored prevention and treatment strategies in oral cancer.
- The research benchmarks the Firefly-optimized CNN-Transformer model against traditional models, demonstrating its superior predictive accuracy, computational efficiency, and capability to manage complex, multimodal data for oral cancer prediction.

The remainder of this study is structured as follows. The review of literature of this study is briefly explained in Part 2. Part 3 briefly discuss the problem statement of this study. Part 4 describe the methodology of the proposed study. Section 5 discusses the result and conclusion. This paper ends with concluding remarks in Part 6.

II. LITERATURE SURVEY

The study by Somyanonthanakul et al's [18] sought to establish an FDL model of clinicopathologic data to predict survival time of oral cancer patients. The authors in the cross-sectional study review patients' electronic medical records of 581 OSCC surgeries performed with or without radio chemotherapy between 2011 and 2019. Survival time classes were classified with the help of a deep learning (DL) model trained on this data. The study augmented the DL model by incorporating fuzzy logic and developed FDL models that boosted survival time prediction. Similar performance was observed on a test set, with an accuracy of 0.74 and the AUC value of 0.84 on the receiver operating characteristic curve. However, the proposed FDL model took higher accuracy to 0.97 and perfect AUC score to 1.00 which indicates that fuzzy logic can enhance deep learning for clinical diagnosis. Although we have some limitations to the study such as a limited amounts of data and inadequacy of clinicopathologic variables with the TNM staging system. The idea for future studies is to increase the size of the dataset over multiple cancer centres and cancer treatment approaches, further combine fuzzy logic with the Transformer model as well as include genomics for further improvement of the survival prediction for oral cancer patients.

Kumar et al., [19] utilize the CNN model Inception-Resnet-V2 is using the transfer learning approach for the initial analysis of oral squamous cell carcinoma (OSCC). To enhance OSCC detection, additional layers are incorporated into this pretrained model. By mining a repository of oral cancer histopathology images, we can gauge how well these tweaked models perform. This paper analyse modification to the architecture of the Inception-Resnet-V2 model and propose a DL-CNN model to use it. It has done the same in the metrics aspect with a high accuracy rate of 91.78%. This success lays strong groundwork for future advancements and real-world implementations, with a proposed direction of examining methods such as quantization or pruning in order to attain better efficiency. This is especially important so as to expand the variability of the training dataset since this variably deals with any diversification of the patient's conditions, age, gender, and the likes. Furthermore, the proposal of constructing web and mobile applications with simplified and easily understandable interface is useful to enhance the usage of the model as a diagnostic instrument for care giving specialists and patients. Further work of this study should consider needs to extend the sample of oral and dental conditions other than oral cancer. In addition to this increase, there is the inclusion of multiple form data network, data enlargement as well as validation in health-related facilities to improve diagnosis scores besides establishing the model as practical for real-life use.

Foqha et al., [20] applied the Firefly Algorithm in order to tackle the coordination problem of IEEE 3-bus network. Thus, the goal of this study was to investigate and compare the influence of critical parameters, such as the number of generations, population size, the coefficient of absorption, and a randomization parameter, on the algorithm. Self, the feasibility, solution quality, computational efficiency and methods when optimized for these parameters are shown

through this study by means of experiment and performance analysis. The evaluation of the number of generations also showed the fact that with the increase in the value of this parameter the objective function value was reduced, which speaks about the quality of the obtained solutions. But from the value of 1600 generations and more, increasing the number of generations any further had no disproportionately positive impact on the solution quality. Solution feasibility was also higher when higher numbers of generations were used, and, in no point of the experiments, constraint violations were obtained. However, the results show that there is a limit for the number of violations where the solutions are both optimal and feasible, which is very high in this case. Further, total number of objective function evaluation and the computational time were found to have a direct relation with the number of generations suggesting that as the iterations proceed, there should be provision of more computational resources. Specifically, to the population size, the results reported that the high cardinality of populations contributed to the higher solution quality since the objective function value was reduced. Population size also had a positive effect on feasibility in addition to having negative effects on constraint violation. Nevertheless, this study found that larger population sizes increased computation time and constraint violations, highlighting a trade-off between solution quality and feasibility. The randomization parameter had limited impact, while extreme values of α reduced performance. Lower γ values accelerated convergence but risked suboptimal solutions, requiring fine-tuning.

Xu et al., [21] propose a multiscale transformer to process the information from image patch tokens of variable scales to extract the fine-grained and coarse-grained features. The transformer model design is based on two branches, a small branch (i.e., small sized patch tokens) and large branch (i.e., large sized patch tokens) where each branch is processed with a separate specialized encoder to represent local and global context information from multiscale image patch tokens, and multi-head cross-attention fusion with lateral connections for information fusion across scales. This information. This ablation shows that MedTrans continuously perform better as patch size becomes smaller and smaller. We present a comprehensive comparison of our model that shows that our model has performed better as compared to different vision transformers and state-of-the-art CNN models on the OSCC dataset. For example, MedTrans-S outperforms the recently proposed CNN-Transformer model named TransPath with Top-1 Accuracy +3.58% and F1-score +3.91%, and best performing CNN model, EfficientNet, with Top-1 Accuracy +8.30% and F1-score +7.23%. While this study examines multi-scale transformers for histopathological cancer diagnosis, it highlights the need for further exploration in other medical areas. Current transformer designs require improvements in accuracy, which remains crucial for reliable medical decision-making in diverse healthcare applications.

Kanharimuthu et al's., [22] research primarily focuses on screening of oral cancer at an early stage, which is very important for better survival rates of patients but it has a number of problems. The proposed method employs a probabilistic neural network in conjunction with discrete wavelet transformation for predicting malignancy of oral lesions. The PNN model has classification accuracy of 80% showcasing that the methodology can effectively differentiate cancer cells. However, accurate identification of cancer regions is difficult because oral lesions are heterogeneous. To counter this, multiple computer vision strategies are outlined to improve the chances of correctly predicting entirely cancerous zones. The role of early screening and referral is stressed as diagnosis and timely treatment leads to minimized mortality and increased patient survival, in addition, to being the model that yields the highest GM sensitivity and specificity figures, it is considered most effective.

Song et al's [23] study addresses the challenge of integrating deep learning into medical workflows, specifically for oral cancer detection, by overcoming the limitation of conventional models that lack uncertainty quantification. A Bayesian deep network is proposed to assess the reliability of image classification by estimating uncertainty. Based on a large intraoral cheek mucosa image dataset captured using our customized device from high-risk population, this study assesses the model, and it demonstrates that meaningful uncertainty information can be generated. However, we report experiments showing the been-together accuracy improvements by uncertainty-informed referral. The incidence of retained data is roughly 90 % when stating either 10 % of the current cases or stating cases in which the uncertainty value is equal to greater than 0.3. This can be further optimised by having more patient referral. It is proved that the developed model is able to select tentative cases requiring additional examination. The primary weakness of this framework is based on model uncertainty which may also include some cases to be misclassified. Moreover, the presented method enhances the trust in deep learning outputs but needs to be confirmed on more diverse study samples and various conditions to determine its better transferability to different clinical settings and high reliability in the actual screening.

A fuzzy deep learning (FDL) approach was developed to estimate survival time in oral cancer patients using clinicopathologic data, significantly improving accuracy compared to standard deep learning (DL) models. The FDL model achieved a perfect AUC of 1.00, highlighting fuzzy logic's potential in enhancing clinical predictions. Another study using a CNN model with transfer learning improved OS detection accuracy to 91.78%, suggesting the need for model compression and data diversity. Research into Firefly Algorithms demonstrated the trade-off between solution quality and computational efficiency, while a multiscale transformer model outperformed state-of-the-art CNNs in histopathological analysis. Additionally, a probabilistic neural network showed promise in early oral cancer detection, and Bayesian networks were introduced to quantify uncertainty in medical image classification, increasing diagnostic reliability. These studies emphasize the importance of model refinement, data expansion, and validation to enhance the accuracy and applicability of deep learning in medical diagnosis and medication.

III. PROBLEM STATEMENT

The detection and prediction of oral cancer using deep learning models present several demanding situations in spite of reaching excessive accuracy in managed datasets. One key issue is dataset limitations, where small, biased datasets reduce the generalizability of models, especially in diverse clinical environments. Fuzzy deep learning (FDL) models, while improving survival time prediction, suffer from class imbalance and a lack of large-scale clinicopathologic data, limiting their effectiveness [18]. Similarly, transfer learning models like Inception-ResNet-V2 improve detection but fail to generalize well due to narrow datasets, making them less applicable to broader populations [19]. Furthermore, despite the catastrophic forgetting of previous experiences, model interpretability is still problematic since compared to conventional deep learning methods, these active learning approaches do not provide uncertainty quantification which is important for clinicians to evaluate the trustworthiness of the model. Another considerable challenge is the computational cost; high computational requirements especially when using big data or when the model involves the fine-tuning parameters makes the application of these models in real-time medical practices impossible. Moreover, the heterogeneity of oral lesions complicates accurate detection, while the need for better feature representation through multimodal data fusion is often overlooked. Balancing computational resources with accuracy remains a challenge. To enhance the clinical applicability of these models, future work must focus on expanding datasets, improving interpretability, and optimizing model efficiency for diverse healthcare scenarios.

1. Data Integration

The proposed method developing a Firefly optimized CNN-Transformer model for oral cancer prediction by integrating clinical data and imaging features. The Convolutional Neural Network (CNN) extracts relevant capabilities from imaging data, while the Transformer captures sequential relationships in scientific records, inclusive of patient demographics and medical history. The Firefly Algorithm is hired for hyperparameter optimization, improving the model's accuracy and efficiency. This hybrid approach objectives to enhance early detection and prognosis of oral cancer, supplying a complete device for clinicians to make informed selections primarily based on numerous facts assets.

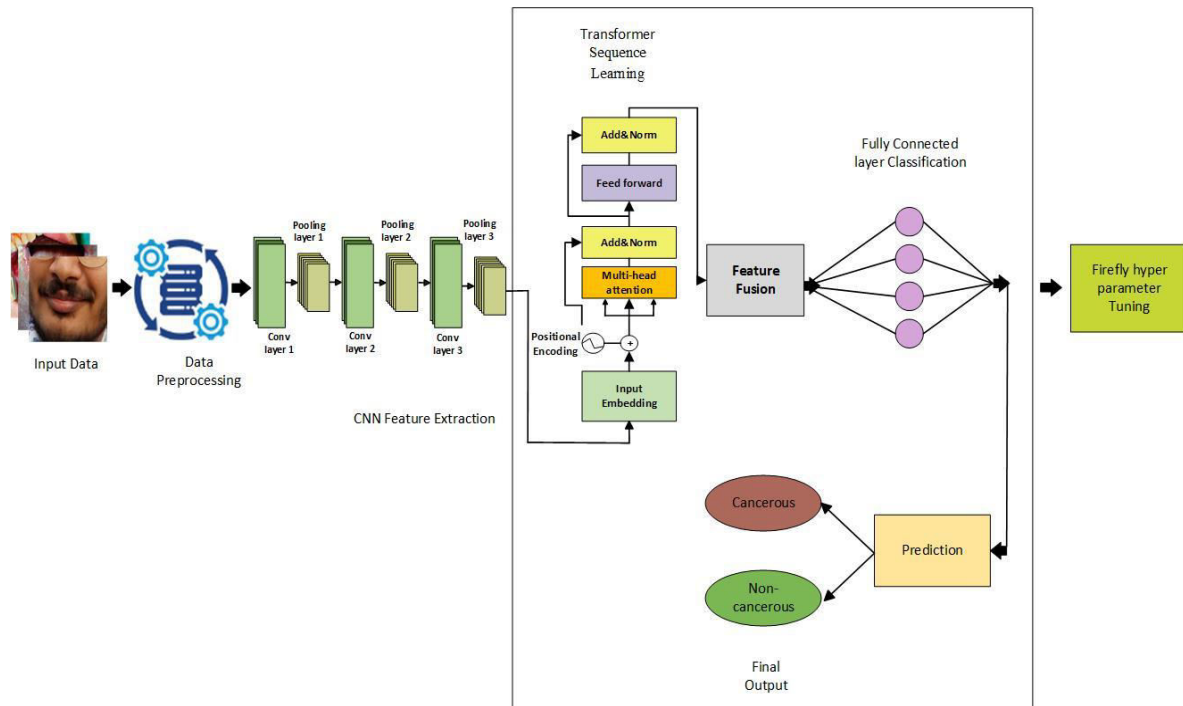


Fig 1. Block Diagram of proposed Firefly Optimized CNN-Transformer

The figure 1 described is a model for the prediction of oral cancer using the Firefly-optimized CNN-Transformer architecture. First, image data is pre-processed by noise reduction and scaling of data. Then convolutional layers extract spatial features from the images and pooling layers brings down the dimensions. The transformer segment incorporates positional encoding, multi-head attention heavyweight, and feed forward layers to consider temporal features. From the CNN and the transformer models, features are extracted, and the features are concatenated and then classified to the cancer or non-cancer stages. The chosen Firefly algorithm tunes the values of hyperparameters to increase the model performance. In this proposed model, the convolutional neural network helps in feature extraction while the transformer model has ability in learning sequence and both are incorporated into improve the performance of detection of oral cancer.

3.1 Data Collection

The Oral Cancer dataset from Kaggle consists of images used for diagnosing oral cancer. It includes various classes representing different stages and types of oral lesions. The dataset is designed to help in improving device learning and deep getting to know models for class responsibilities. All pictures are accompanied by analysis to help in the process of the supervised mastering. This dataset is helpful for researchers and practitioners whose interest is to improve detection of oral most cancers and increase diagnostic accuracy utilizing superior image processing [24].

Table 1. Dataset Detail

Class	No. of image data
Cancerous	44
Non-cancerous	87
Total no. of image data	131

3.2 Data Preprocessing

Data pre-processing in Oral Cancer dataset consists of several steps to make the images appropriate for the analysis. These measures include resizing the images where the size is made fixed, normalizing the pixel intensities in the range of 0 to 1 and data augmentation measures which include rotation of the images, overturning of the image and

enlargement or making the image smaller. Furthermore, images if needed can be made into black and white or contrast images to enhance the features during the training process.

Data Normalization

Normalization is used in cleaning the pixel values in image data in order to improve model performance. Since pixel value is divided by 255 at the end to bring the range of variable in [0, 1], pixel values are converted [25].

$$\text{Normalized Value} = \frac{\text{pixel}}{255} \quad (1)$$

Image Resizing

Resizing entails scaling down or up all the images in a dataset to a particular size in order to enhance their feed into models. Standardizing does this helps in avoid distortion of it as well as maintaining the aspect ratio of an image. This process helps reduce the amount of batch processing during training in the event that the models call for fixed size inputs. Also, resizing enhances computational speed, decreases memory usage, and refines features by allowing the approach to focus on patterns of interest without concern about image size [26].

Data Augmentation

As a form of data augmentation, augmented data will be used extensively in the following proposed study to enhance performance of an oral cancer detection model. The Application of augmentation techniques ensures that the training dataset is expanded thus minimizing a situation where the model is only capable of solving the train dataset problems without handling the actual problems outside the sample dataset hence reducing over fitting and improving a model performance [27]. Key methods include:

Rotation: Rotating images randomly by a specific degree assists the model to ignore the orientation of the lesions.

Flipping: Horizontal and vertical flipping which makes the model capable of recognizing features regardless of its position and adds variability in the patient imaging

Zooming: Using random scale parameters means that at different iterations the model will look at the lesions under different scales, recreating real life variance.

Shifting: The ability of the model to capture these images slightly translated helps to minimize cases where only the lesion that is cantered is seen.

3.3 Feature Extraction Using CNN

The Convolutional Neural Network (CNN) is central in feature extraction of the oral cancer images. The first step is Convolution Layers in which numerous convolutions are made to the supplied images. The convolution operation is mathematically represented as in equation (2),

$$j(x, y) = (I * K)(x, y) = \sum_m \sum_n I(x + m, y + n)K(m, n) \quad (2)$$

Where, $I(x, y)$ is the input image, $K(m, n)$ is the convolution kernel, and $j(x, y)$ is the output feature map. These filters glide over the image to fit over essential aspects such as tissue abnormalities, lesions, and textural change. By learning both low-level features and high-level features, the CNN enables effective feature extraction that distinguishes cancerous from non-cancerous tissues. Once features are extracted by the convolutional layers, the data passes through Pooling Layers typically max pooling. The justification for pooling is that, when it reduces the spatial extent of the feature maps, it decreases computational intensity while preserving key information that helps to combat overfitting. The pooling operation is mathematically represented as in the equation (3),

$$k_{\text{pool}}(x, y) = \max \{k(i, j)\} \quad (3)$$

Where, (i, j) is the pooling window.

Max pooling retains the most prominent features from the feature map, focusing on areas like tumour, edges and significant textural regions. Following convolution and pooling, the feature maps are flattened by a Flatten Layer, converting the 2D feature maps into a 1D array. This step prepares the data for the fully connected layers, where the

final classification of the input image as cancerous or non-cancerous is made. The CNN efficiently extracts and condenses essential features using convolution and pooling operations, ensuring high accuracy in detecting oral cancer.

3.4 Transformer Model for Sequence Learning

In the proposed study, the Transformer model is used for sequence getting to know, which significantly improves statistics analysis with the help of the model structure. Dissimilar to recurrent neural networks (RNNs) that process sequences in a sequential manner, the Transformer uses parallel computing, which makes it feasible to handle long-range dependencies in the records [32].

3.5 Input Embedding

They take an input sequence of timestamps (or any other time series clinical data such as patient records, symptoms, and lab results) and map it into the embedding space. This step maps each element of the sequence (like a clinical record at a certain time instance) to a high dimensional vector. It assists the model to work with a numerical format of the sequence for further use or analysis.

Positional Encoding

By nature of its structure RNNs /LSTMs have a natural understanding of the order of data, Transformers on the other hand do not, they process sequences in parallel batches. To address this, the positional encoding is incorporated to the input embeddings. Positional encoding is in fact a vector that indicates the position of each component in the sequence. It assists the model in making discriminations of data on the basis of their location in a sequence. For example, a patient may have the first test result will have different positional encoding of latest test result.

3.5.1 Multi-Head Attention

Multi-head attention is a critical component of the Transformer model. The proposed model does not segregate the input in segments and it will consider the present segments in an attempt to form relations between different data points. The model calculates the attention weights for each combination of the input elements deciding how much one part of a sequence (for instance, one clinical record) should pay attention to other parts of the sequence. Multi-head attention implies that this process takes place several times simultaneously and each time the focus of attention is on different aspects of data.

3.5.2 Add & Normalize

Then, the linear projections are followed by additive residual connection with the input after multi-head attention layer to avoid problems like vanishing gradients. Then, normalization (Layer Normalization) is applied to randomly normalize the activation outputs so that it can help to make the learner better adapted to the subsequent training phase.

3.5.3 Feed Forward

The resultant of the attention mechanism goes through feed-forward neural networks. Further to this, the attention output is then subjected to non-linear transformations within this network. This is where in the model can learn more sophisticated features in the sequence data

3.5.4 Feature Fusion

Subsequently, the features produced by the Transformer are combined with the features obtained by the CNN from the image data such as oral scans. This fusion is to combine important information in the sequence (clinical data) and the visual features (medical images).

3.6 Classification Using Fully Connected Layer

Fully Connected Layer is used for the classification stage of the designed model. A data which goes through several convolution and pooling layers is fully processed in a way that the next layer receives the features maps of the images in one dimension vector. This layer connects every neuron from the flattened feature vector to each and every neuron in the output layer so enhancing spatial connectivity of neurons for the model to discern complex correlation between extracted topological features. At the last layer, possibilities will be produced for each class, the use of an activation feature like SoftMax or sigmoid, by means of which the enter may be categorised as cancerous or non-cancerous. This layer combines all the discovered capabilities making the ultimate choice based at the styles discovered in in advance layers and could be very crucial in prediction of the version [28].

IV. ALGORITHM

Algorithm 1: CNN-Transformer with Firefly Optimization

Output: Classification of images as cancerous or non-cancerous

Load the dataset of medical images of oral lesions

Data Preprocessing

Preprocess the input medical images

Split the dataset into training (70%) and testing (30%) sets.

CNN & Transformer Modelling Loop

while True:

Check if the Training model is improving

If the model is stagnating or overfitting

Firefly Optimizer Model

Initialize the population of fireflies with random hyperparameters (x: number of convolutional layers, y: learning rate, z: batch size)

Evaluate the performance of each firefly

Move the fireflies towards brighter fireflies

Select the optimal set of hyperparameters

If optimal hyperparameters are found

CNN Model Training

Apply Convolutional Neural Networks

CNN model using the optimized hyperparameters (x, y, z);

Transformer Model for Feature Refinement;

Input the features extracted by CNN into the Transformer model;

Apply self-attention in the Transformer model;

Process the refined features;

Classification

Classify training set values using SoftMax

assess training and testing datasets;

Determine the errors;

producing accuracy reports and graphs;

if Stopping criteria are met:

break;

4.2. Firefly Optimization for Optimize CNN Hyperparameters

Firefly Optimization (FA) is an optimization algorithm stimulated by the flashing behaviour of fireflies. The main idea is that fireflies are attracted to brighter ones, which corresponds to finding better solutions in an optimization problem. FA works by adjusting the position of fireflies (candidate solutions) based on their brightness (fitness or objective value), and fireflies with lower brightness move toward those with higher brightness. This iterative process continues until an optimal solution is found [29].

In the proposed study, Firefly Optimization is employed to fine-tune the hyperparameters of the Convolutional Neural Network (CNN), including the number of filters, kernel size, and learning rate. The optimization process begins by defining an objective function, where the brightness of each firefly corresponds to the CNN's performance with a specific set of hyperparameters, like accuracy or loss on the validation set. Fireflies representing better-performing configurations exhibit higher brightness and attract others, while those with poorer performance move towards these more successful fireflies, adjusting their hyperparameters accordingly. Initially, several fireflies (hyperparameter configurations) are randomly generated, and their performance is evaluated using a validation set. Based on this evaluation, fireflies with lower performance update their hyperparameters by moving toward those exhibiting superior performance. This iterative process continues until the optimal set of hyperparameters for the CNN is identified, ensuring that the model achieves the best possible accuracy in predicting oral cancer from medical images. By leveraging the natural behaviour of fireflies, the optimization method effectively balances exploration of new configurations with the refinement of existing ones, advancing the general efficiency of the CNN in the study. The Firefly Optimization algorithm adjusts the CNN's hyperparameters using the equation (4),

$$y_i^{(t+1)} = y_i^{(t)} + \beta_0 e^{-\gamma r_{ij}^2} (y_j^{(t)} - y_i^{(t)}) + \alpha \epsilon_i \quad (4)$$

Where, $y_j^{(t)}$ and $y_i^{(t)}$ represent the hyperparameter sets at iteration t , $\beta_0 e^{-\gamma r_{ij}^2}$ defines how much y_i is influenced by y_j , introduces randomness to explore the search space more thoroughly [30].

V. RESULT

The proposed study demonstrates a robust methodology for predicting oral cancer using a Firefly optimized CNN-Transformer model, integrating clinical data for enhanced accuracy. The feature extraction process identifies critical patterns related to malignancy through convolutional layers, followed by hyperparameter tuning using the Firefly algorithm. The Transformer model further refines predictions by capturing interdependencies between features. The final output categorizes images as cancerous or non-cancerous, achieving improved diagnostic performance, thus contributing significantly to early oral cancer detection and management.

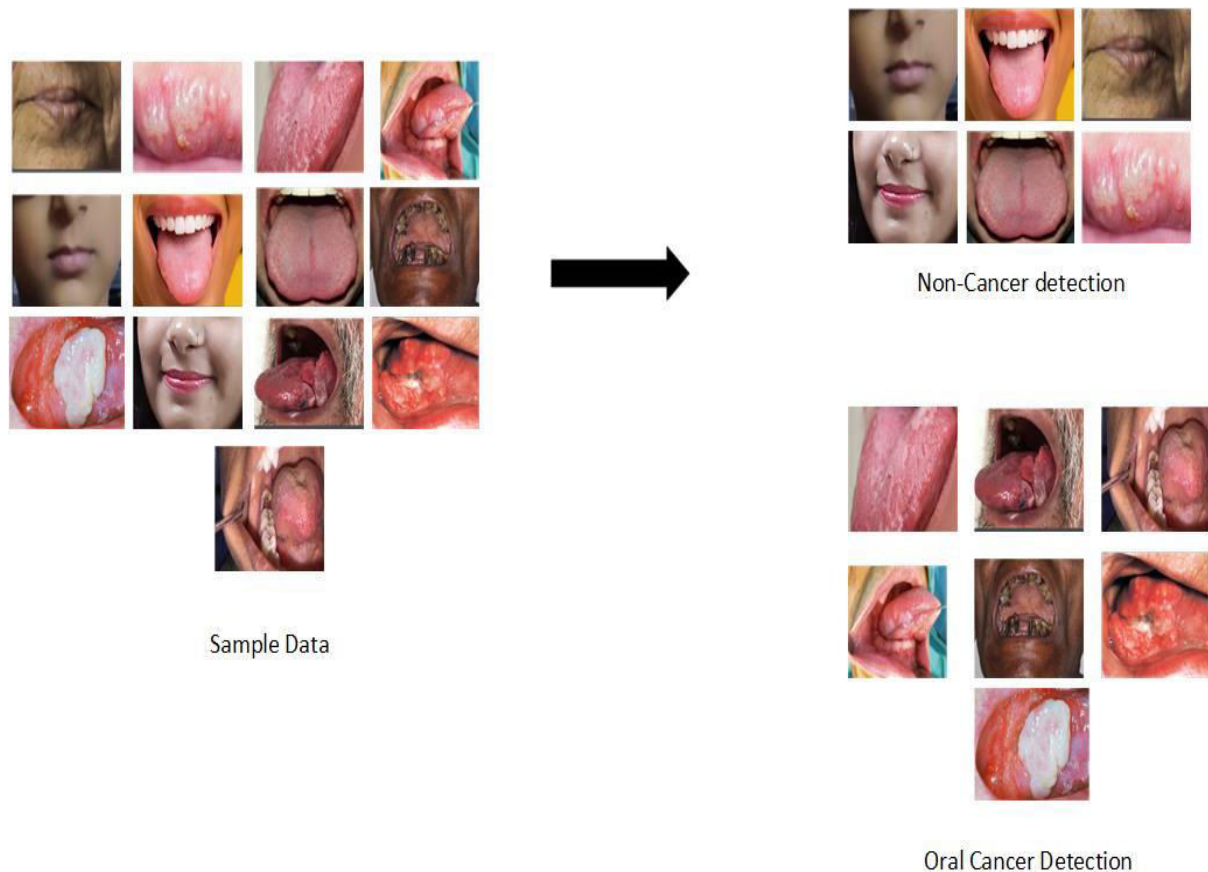
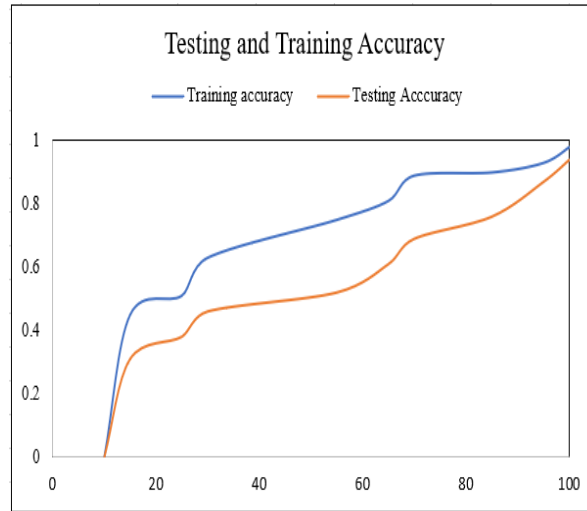
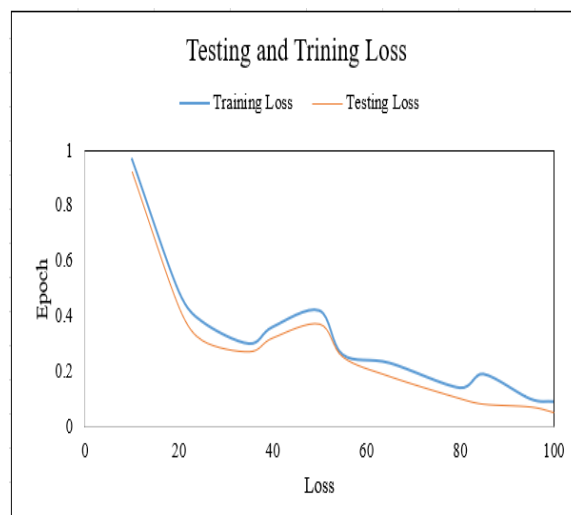


Fig 3. Predict Cancer and Non-cancerous data from Sample

The figure 3 shows that the oral cancer detection of the cancerous data and non-cancerous data from the data sample collected.



(a)



(b)

Fig 4. Testing and Training (a) Accuracy, (b) Loss

The fig 4 (a) illustrates the training and testing accuracy of the Firefly Optimized CNN-Transformer model over epochs. The training accuracy (blue line) steadily increases, approaching 1, indicating effective learning. The testing accuracy (orange line) also rises but is slightly lower, suggesting the model performs well but may not generalize perfectly to unseen data.

The 4 (b) depicts the training and testing loss over the same epochs. Training first loss remains blue and steadily decreases and similarly, the second and the testing loss in orange also decreases, affirming that the model is negative mapping and minimizing errors. Thus, the training and testing loss profiles traced by the ANN model indicate that the risk of overfitting has been kept at a low level while the model is delivering a satisfactory level of performance.

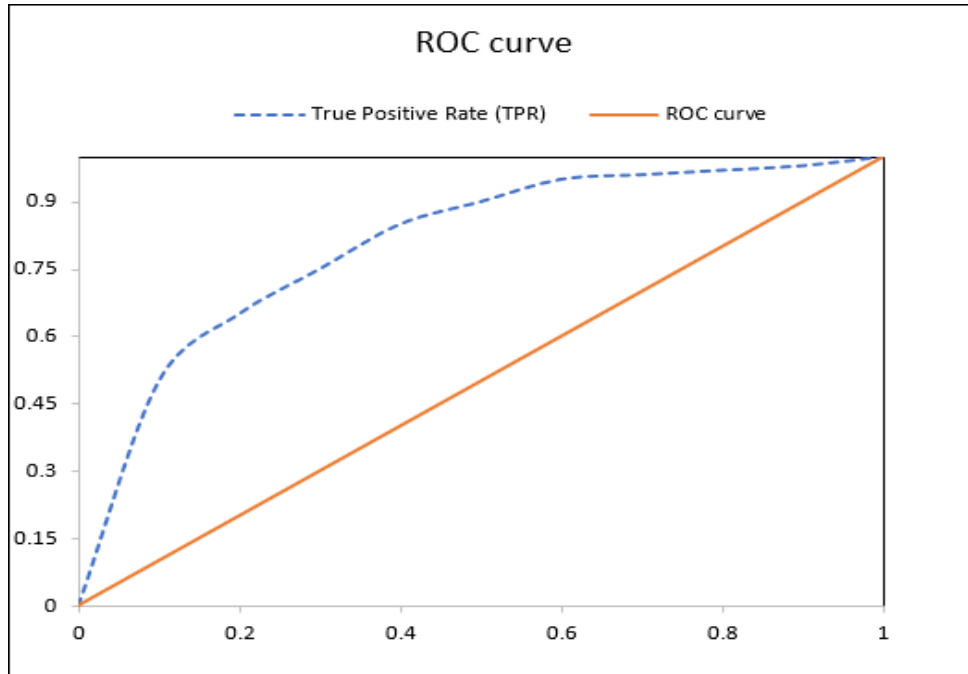


Fig 5. ROC curve

In fig 5, the ROC curve presented above displays the results of a classifier finding the balance between TPR and FPR on different thresholds. The blue dashed line refers to the TPR and it increases as more true positive units are flagged by the model. The orange diagonal line on the graph shows us performance of random classifier. A curve that is nearer the top left corner, and ideally as close as possible to the top left indicating optimal classification performance, suggest beneficial classification performance since the model is sensitive to the two classes.

Performance Evaluation

The performance evaluation of the Firefly Optimized CNN-Transformer model for oral cancer prediction emphasizes key metrics: - the measurement of accuracy, the precision, the recall, and the F1 score. Collectively, these measures show the accuracy with which the model can identify both cancer and non-cancer cases. The performance is used by researchers measuring the accuracy such as [33][34][35]. The findings show that this model provides a much better solution compared to the conventional approaches and therefore has the potential of boosting the early diagnostic capabilities and consequently the quality of patients' treatment of oral cancer.

Accuracy: The current model yielded to 97.03% accuracy in distinguishing between malignant and non-malignant oral cells than the traditional models. This can be attributed to the capability of its feature extraction that the proposed model effectively identifies.

$$\text{Accuracy (ACC)} = \frac{\text{True Positive} + \text{True Negative}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

Precision: Accuracy is one of the significant factors that determine the efficiency of the classification model and the hybrid novel model proposed has a higher precision ratio than other existing models.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

Recall: Recall, also known as true positive rate or sensitivity, shows the model's capacity to identify positive samples accurately. The proposed model had a higher recall rate than the other existing models that are 92.3%.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

F1 score: One statistic employed to assess the effectiveness of the classification model is the F1 score. By calculating its harmonic significance, accuracy, and recall are combined into an individual statistic. The proposed hybrid novel model achieved a 97.72% F1 score which is higher than other traditional existing models.

$$F1score = \frac{2 \times TP}{(2 \times TP) + FP + FN} \quad (8)$$

Table 2. Performance Metrics of Proposed FOCNN-Transformer Model

Accuracy	97.03
Precision	97.57
Recall	98.21
F1 score	97.72

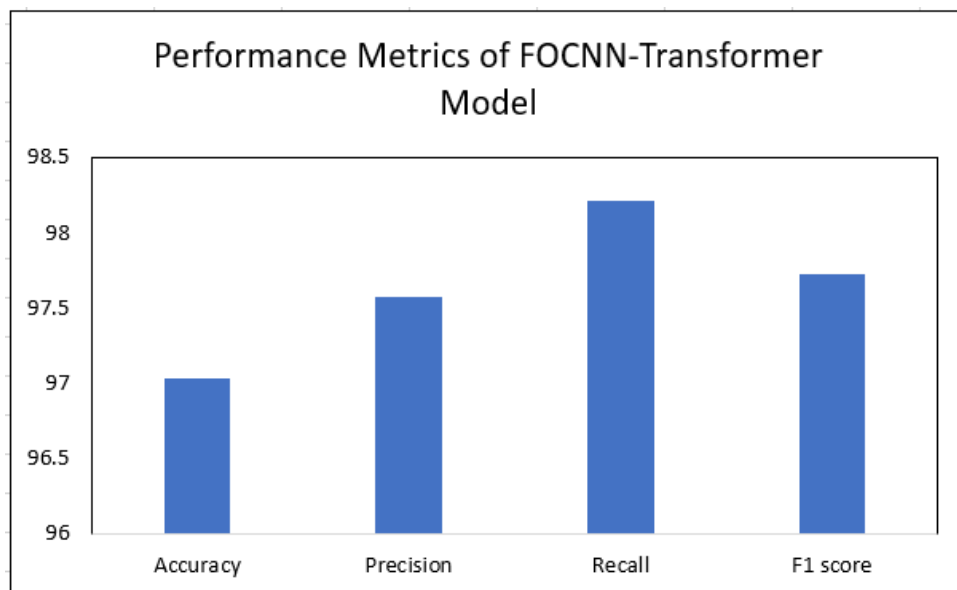


Fig 6. Performance Metrics Graph of the FOCNN-Transformer Model

Table 2 and figure 6 represent the accuracy of Firefly Optimized CNN-Transformer model for prognosis of oral cancer. The model attains a predictive accuracy of 0.9703, which means that the model predicted the right category for 9703 of the total cases. Specificity, at 97.57% – this presents the % of true positive predictions out of all the positive predictions that the model makes, in essence, keen in minimizing on the false positive that it produces. The final recall of the model is 98.21%, the best possible outcome for recalling all true positive cases and minimizing false negatives. F1 score or the harmonic mean of precision and recall values are 97.72%, which is the best measurement for our model. Such values are illustrated in the graph where high values are depicted which feature indicates that the model has high predictive accuracy on all the mentioned values.

Table 3. Comparison of FOCNN-Transformer with Existing Approach

Methods	Accuracy	Precision	Recall	F1 score
CNN [36]	96.15	96.51	97.64	97.07
ResNet50 [37]	95.33	95.03	95.63	95.31
InceptionV3 [37]	96.02	96.0	96.05	96.05
VGG16 + CNN [38]	87.0	90.0	94.0	92.0
InceptionV3+ CNN [38]	82.0	87.0	89.0	88.0
Proposed Model (FOCNN-Transformer)	97.03	97.57	98.21	97.72

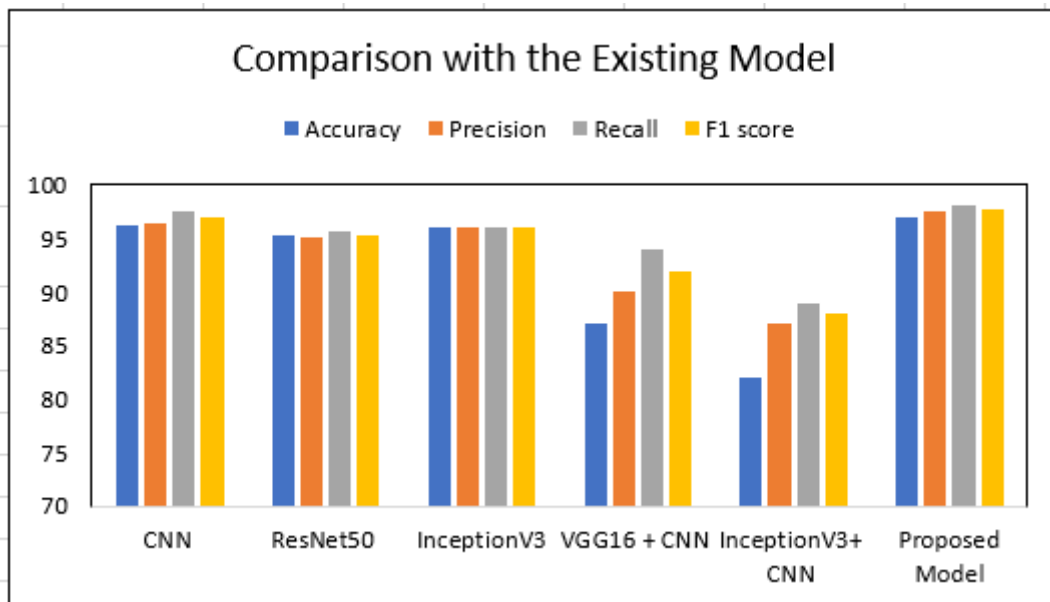


Fig 7. Comparison with Existing Model

The table 3 and figure 7 compares the performance of several models based on four metrics: Accuracy, Precision, Recall, and F1 score. The proposed Firefly Optimized CNN-Transformer (FOCNN-Transformer) model outperforms the other methods, achieving the highest scores across all metrics: 97.03% Accuracy, 97.57% Precision, 98.21% Recall, and a 97.72% F1 score. Notably, the models CNN, ResNet50, and InceptionV3 show strong results, especially CNN with a Recall of 97.64%. However, models like VGG16 + CNN and InceptionV3 + CNN perform significantly lower, especially in Accuracy, with scores of 87% and 82%, respectively. The graph visually confirms the superior performance of the proposed model, particularly in Recall and F1 score, demonstrating its effectiveness over existing methods.

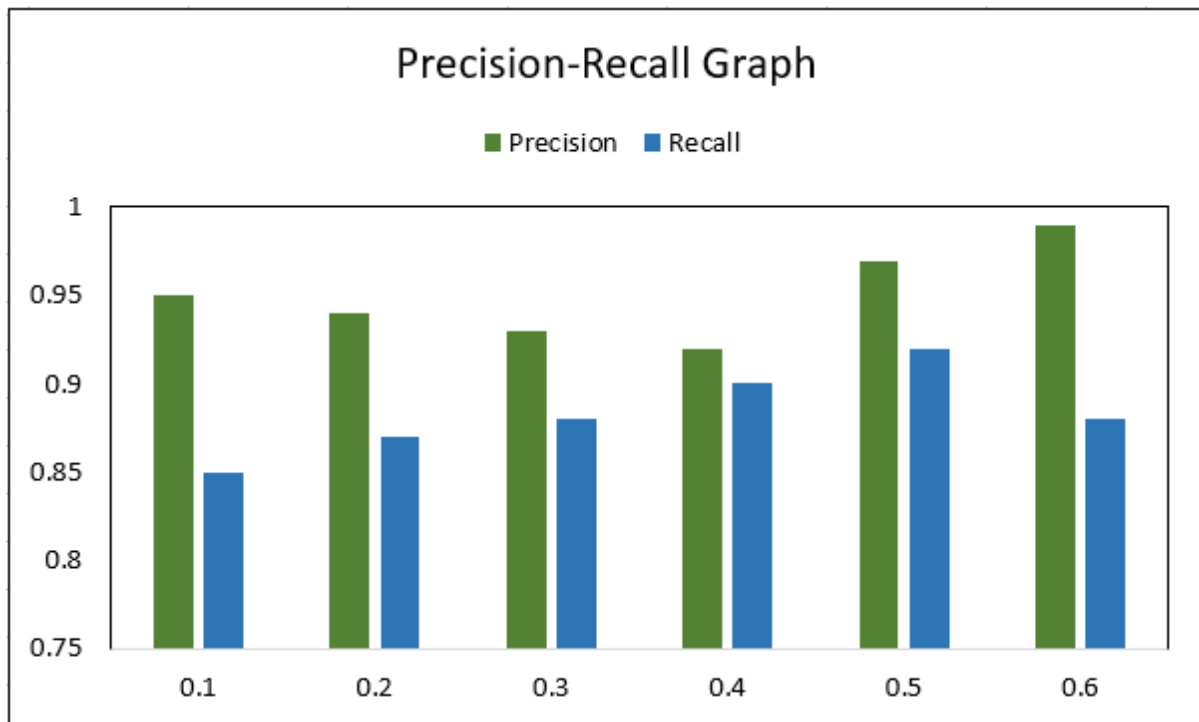


Fig 8. Precision-Recall Curve

The figure 8 illustrate the Precision-Recall (PR) graph of the trade-off between precision and recall for a classification model across various threshold settings. Precision represents the proportion of true positive predictions among all predicted positives, while recall measures the proportion of true positives among all actual positives. A curve close to the top-right corner indicates high precision and recall, signifying a robust model. Conversely, if the curve lies closer to the diagonal, it suggests poorer performance. An optimal model balances both metrics, achieving high precision without sacrificing recall, especially in imbalanced datasets.

VI. DISCUSSION

The Firefly Optimized CNN-Transformer (FOCNN-Transformer) model significantly improves oral cancer detection by addressing the limitations of traditional machine learning methods. Standard methods, particularly the classic convolutional neural networks (CNNs) alone, have certain limitations concerning the representation of intricate interactions in sequential and contextual data important for sociosuspension, as well as the ability to incorporate separate data sources, including clinical records and imaging data. In addition to the feature extraction skills demonstrated by the CNNs and the contextual processing abilities of the transformers, the FOCNN-Transformer builds on the former's knowledge of spatial-temporal characteristics, leading to higher diagnostic efficiency and accuracy. It is possible to assume that such a twofold strategy enables the model to better perceive highly nontrivial data patterns involved with clinical context understanding. One great improvement of this model is the introduction of firefly optimization that enhances the hard to tweak hyperparameters manually. It also assures that during the training period, the model attains convergence at relatively high-performance levels faster than it would otherwise would be possible. Finally, it improves the model's accuracy of the classifier and makes them more generalizable specifically for the detection of oral cancer. Also, the functions of the model improve by the integration of clinical information; history and present symptoms, biopsy information, and so on. This integration helps the model associate different aspects visible in the imaging data with clinical characteristics, which increases the model's ability to evaluate the patient's status comprehensively. Last but not the least, the FOCNN-Transformer acts as a novel development in diagnosing oral cancer more accurately and with more results consistently with the aid of the state-of-the-art machine learning algorithm and integrating data more systematically.

Based on the proposed model, with evaluation parameters from the set off metrics it was found out that the accuracy of the model is percent 97.03, the precision is percent 97.57, recall is 98.21 percent and f1 score of 97.72 percent. The results found out have proven that FOCNN-Transformer model has high accurate detection of malignant tumours in stage I and minimized false-negative rate compared to the existing technique. The conventional methods can sometimes encounter problems in extending the results to alternate sets of data inputs and are not highly selective in the natural world. This particular model holds to enhance the distinguished diagnosis status that is at the root, early intervention into the occurrence of oral cancer hence improving, death rates. Finally, the FOCNN-Transformer contributes new approaches to the area of oral cancer, which foresees with better effectiveness, accuracy, and speed since it is deeply rooted in the current demanding deep learning frameworks.

VII. CONCLUSION

Thus, the research presented in this study points to a Firefly Optimized CNN-Transformer as an accurate, efficient and plausible approach for the detection of oral cancer. When joining the CNN feature extraction with the transformers' attention, the needed image and clinical data processing is provided, enhancing the model's diagnostic capabilities. Firefly optimization is applied next to the system with the intent of fine-tuning the system and improving the convergence capabilities further. The inclusion of clinical data that can be often overlooked in similar models makes this model superior for actual practical use where clinical data are relevant for a diagnosis. So, the presented model is capable of enhancing the existing diagnostic approaches and optimize the diagnostic processes, which can increase the rates of primary diagnosis, reduce the number of false negatives, as it is shown by accuracy, precision, recall, and F1 measures. This leads to a subsequent diagnosis of the disease and possible better quality of late oral cancer patients. As for future work, it will be necessary to extend validation of the proposed model for more extensive and heterogeneous datasets, containing multisession data from various geographical areas. Further improvements will be made to the work by incorporating superior methods such as multi-scale feature fusion and self-supervised learning on the designed model. Realistic usability of the model is also proposed for instant real-time application in clinical practice for decision support systems for Oral Cancer in particular.

REFERENCES

- [1]. H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021.
- [2]. C. Goetz, J. Raschka, K.-D. Wolff, A. Kolk, and O. Bissinger, "Hospital based quality of life in oral Cancer surgery," *Cancers*, vol. 12, no. 8, p. 2152, 2020.
- [3]. E. B. Steen-Olsen, D. R. Stormoen, C. A. Kristensen, I. R. Vogelius, C. Holländer-Mieritz, and H. Pappot, "Patient-reported outcome during radiotherapy for head and neck cancer: the use of different PRO questionnaires," *Eur. Arch. Otorhinolaryngol.*, vol. 279, no. 8, pp. 4199–4206, 2022.
- [4]. M. Čanković et al., "Predictors of health-related quality of life in Serbian patients with head and neck cancer," *Med. Oral Patol. Oral Cir. Bucal*, vol. 27, no. 4, p. e340, 2022.
- [5]. R. Mauceri, M. Bazzano, M. Coppini, P. Tozzo, V. Panzarella, and G. Campisi, "Diagnostic delay of oral squamous cell carcinoma and the fear of diagnosis: A scoping review," *Front. Psychol.*, vol. 13, p. 1009080, 2022.
- [6]. E. S. Mira et al., "Early diagnosis of oral cancer using image processing and artificial intelligence," *Fusion Pract. Appl.*, vol. 14, no. 1, pp. 293–308, 2024.
- [7]. Y. Shen et al., "CDK5RAP2 is a Wnt target gene and promotes stemness and progression of oral squamous cell carcinoma," *Cell Death Dis.*, vol. 14, no. 2, p. 107, 2023.
- [8]. G. Devindi et al., "Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer," *IEEE Access*, 2024.
- [9]. S. Usman, A. Jamal, M.-T. Teh, and A. Waseem, "Major molecular signaling pathways in oral cancer associated with therapeutic resistance," *Front. Oral Health*, vol. 1, p. 603160, 2021.
- [10]. A. K. Chaturvedi et al., "Oral leukoplakia and risk of progression to oral cancer: a population-based cohort study," *JNCI J. Natl. Cancer Inst.*, vol. 112, no. 10, pp. 1047–1054, 2020.
- [11]. S. B. Jugunta et al., "Exploring the Insights of Bat Algorithm-Driven XGB-RNN (BARXG) for Optimal Fetal Health Classification in Pregnancy Monitoring," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 11, pp. 731–741, 2023, doi: 10.14569/IJACSA.2023.0141174.

- [12]. H. M. Elfadil, H. F. H. Hamed, O. Abdulhamid, and F. Embarak, "Enhancing Robotic Performance and Adaptability through Advanced Predictive Models with Deep Reinforcement Learning," *Int. J. Innov. Res. in Computer Commun. Eng.*, vol. 13, no. 1, p. 966, 2025, Accessed: Mar. 18, 2025.
- [13]. M. Kim et al., "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, p. 657, 2019.
- [14]. S. Camalan et al., "Convolutional neural network-based clinical predictors of oral dysplasia: Class activation map analysis of deep learning results," *Cancers*, vol. 13, no. 6, p. 1291, 2021.
- [15]. R. A. Welikala et al., "Fine-tuning deep learning architectures for early detection of oral cancer," in *Mathematical and Computational Oncology: Second International Symposium, ISMCO 2020, San Diego, CA, USA, October 8–10, 2020, Proceedings 2*, Springer, 2020, pp. 25–31.
- [16]. P. Dasson Bajaj et al., "A scoping review exploring oral health inequalities in India: a call for action to reform policy, practice and research," *Int. J. Equity Health*, vol. 22, no. 1, p. 242, 2023.
- [17]. Y. Tan et al., "Oral squamous cell carcinomas: state of the field and emerging directions," *Int. J. Oral Sci.*, vol. 15, no. 1, p. 44, 2023.
- [18]. R. Somyanonthanakul, K. Warin, S. Chaowchuen, S. Jinaporntham, W. Panichkitkosolkul, and S. Suebnukarn, "Survival estimation of oral cancer using fuzzy deep learning," *BMC Oral Health*, vol. 24, no. 1, p. 519, 2024.
- [19]. K. V. Kumar, S. Palakurthy, S. H. Balijadaddanala, S. R. Pappula, and A. K. Lavudya, "Early Detection and Diagnosis of Oral Cancer Using Deep Neural Network," *J. Comput. Allied Intell. JCAI ISSN 2584-2676*, vol. 2, no. 2, pp. 22–34, 2024.
- [20]. T. Foqha, S. Alsadi, A. Elrashidi, N. Salman, and others, "Optimizing Firefly Algorithm for Directional Overcurrent Relay Coordination: A case study on the Impact of Parameter Settings," 2023.
- [21]. X. Xu, B. Deng, J. Wang, and G. Yi, "Individual Prediction of Electric Field Induced by Deep-brain Magnetic Stimulation with CNN-Transformer," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2024.
- [22]. M. Kantharimuthu, M. Malathi, P. Sinthia, and others, "Oral Cancer prediction using a probability neural network (PNN)," *Asian Pac. J. Cancer Prev. APJCP*, vol. 24, no. 9, p. 2991, 2023.
- [23]. B. Song et al., "Bayesian deep learning for reliable oral cancer image classification," *Biomed. Opt. Express*, vol. 12, no. 10, pp. 6422–6430, 2021.
- [24]. M. Tahmid, "Oral Cancer." Jan. 2024. [Online]. Available: <https://www.kaggle.com/code/tahmidmir/oral-cancer/notebook#Setting-up-general-parameters>
- [25]. P. Agarwal, N. Gupta, Y. Bharadwaj, A. Yadav, and P. Mathur, "Oral Cancer Stage Classification Using Machine Learning," *Procedia Comput. Sci.*, vol. 235, pp. 3174–3180, 2024.
- [26]. J. Chaki, "Cancer Data Pre-Processing Techniques," in *Current Applications of Deep Learning in Cancer Diagnostics*, CRC Press, 2023, pp. 19–33.
- [27]. B. S. Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification," *Int. J. Data Sci. Anal.*, pp. 1–18, 2024.
- [28]. M. Xiao, Y. Li, X. Yan, M. Gao, and W. Wang, "Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example," in *Proceedings of the 2024 7th International Conference on Machine Vision and Applications*, 2024, pp. 145–149.
- [29]. A. Yasmin, W. Haider Butt, and A. Daud, "Ensemble effort estimation with metaheuristic hyperparameters and weight optimization for achieving accuracy," *Plos One*, vol. 19, no. 4, p. e0300296, 2024.
- [30]. X. Huang, L. Chen, Y. Zhang, S. Su, Y. Lin, and X. Cao, "Improved firefly algorithm with courtship learning for unrelated parallel machine scheduling problem with sequence-dependent setup times," *J. Cloud Comput.*, vol. 11, no. 1, p. 9, 2022.
- [31]. Z. Liu, X. Li, D. Wu, Z. Qian, P. Feng, and Y. Rong, "The development of a hybrid firefly algorithm for multi-pass grinding process optimization," *J. Intell. Manuf.*, vol. 30, pp. 2457–2472, 2019.
- [32]. H. Bao, L. Dong, W. Wang, N. Yang, S. Piao, and F. Wei, "Fine-tuning pretrained transformer encoders for sequence-to-sequence learning," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 5, pp. 1711–1728, 2024.
- [33]. F. Embarak, N. Azman Ismail, O. R. R. Shahin, and R. N. N. Alabdali, "Design of autonomous online social community architecture for older adults," *Comput. Electr. Eng.*, vol. 100, p. 107900, May 2022, doi: 10.1016/j.compeleceng.2022.107900.
- [34]. Asmaa, Aljilany, A.-H. Marwa, A. Alaa, E. Farhat, and E. Abdulhamid, "Cultivating Independent Learners: Designing an Innovative Interactive Model to Enhance Self-Directed Learning and Autonomy in Elementary Education," *International Journal of Innovative Research in Science Engineering and Technology*. Accessed: Jan. 28, 2025.

- [35]. A. O. Alsayed, N. A. Ismail, L. Hasan, and F. Embarak, "A Comprehensive Review of Modern Methods to Improve Diabetes Self-Care Management Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, pp. 182–203, 2023, doi: 10.14569/IJACSA.2023.0140920.
- [36]. R. Prabhakaran and D. J. Mohana, "Detection of oral cancer using machine learning classification methods," *Int. J. Electr. Eng. Technol.*, vol. 11, no. 3, 2020.
- [37]. R. Dharani and S. Revathy, "Industry 4.0 transformation: adaptive coati deep convolutional neural network-based oral cancer diagnosis in histopathological images for clinical applications," *Opt. Quantum Electron.*, vol. 56, no. 2, p. 152, 2024.
- [38]. K. Dhanya, D. V. V. Prasad, And Y. V. Lokeswari, "Detection Of Oral Squamous Cell Carcinoma Using Pre-Trained Deep Learning Models," *Exp. Oncol.*, Vol. 46, No. 2, Pp. 119–128, 2024.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details