International Journal of Advanced Scientific Technologies in Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.12,Issue.4,2026

DOI: 10.22413/ijastems/2026/v12/i4/4

End-to-End AI System for Dental Imaging: Tooth Position Detection and Pathology Classification Using YOLOv4 and CNN

P. M. Naidu, T. Samiksha, G. Manish Reddy, K. Sai Nagendra, S. Anudeep Reddy
Vignan institute of technology and science, Deshmukhi-508284 , Yadadri Bhuvangiri, Telangana
Corresponding author: pmnaidu.vits@gmail.com

Abstract - This paper introduces an automated tooth position and pathology classification system on live visual input of a dental imaging system in real-time. The system captures the data using an interface that has a camera whereby the user places his or her teeth before the camera and the system scans the teeth within a short time. The obtained visual data are processed by the multi-stage processing pipeline which is able to extract structural and diagnostic data. The system uses two different datasets. A detection dataset is applied to train a YOLOv4-based model to find the areas of teeth and spatial location, and a classification dataset to find the dental conditions based on the features of the image. After detection and classification, the information processed is employed to make improved structural representations of the dental part. The system renders two main outputs, which are synthesized 2D X- ray-like visualization and pseudo 3D rotational representation of the tooth structure discovered. Such outputs give positioning as well as diagnostic information, which can be viewed as a visual representation of the dental condition without requiring the use of the traditional radiographic devices. The suggested system illustrates the combined model of real time-non-invasive dental evaluation with automated image processing.

I. INTRODUCTION

Dental diseases comprise some of the most existing medical conditions that affect people across all parts of the world with a remarkable adverse influence into oral health, well-being, and life quality. An example of common dental conditions that include cavities, periodontal infections, and others.Abnormalities of the structure must be timely diagnosed to avoid further development of critical clinical complications. Correct determination of tooth location and pathology is a very important factor in therapeutic planning and prophylaxis.

Conventional dental diagnostic practices are mainly based on radiographic imaging such as X-ray based ex- tests. Although these techniques have good detail on the structure, they demand special equipment, trained staff, and controlled clinical conditions. Moreover, the recurring exposure to ionizing radiation and its low accessibility in the resource-constrained areas point to the necessity of alternative diagnostic methods, which may be considered safe and convenient, as well as relatively low-cost.

The recent achievements of artificial intelligence and computer vision have allowed automated analysis of visual medical data, and this opens a new prospect of non-invasive diagnostic systems. Deep learning methods, especially object detection models and image classification models, have shown to be quite successful at detecting anatomical structures and estimating the conditions of diseases based on the visual input. The developments favor the development of real-time healthcare tools that can be used with the regular imaging devices.

This paper presents a dental imaging end-to-end artificial intelligence framework to perform automated tooth positioning and classification of the pathology by accessing live camera images. A YOLOv4-based detection model is used to detect the areas of teeth and spatial location, and a convolutional one is taken.In addition to diagnostic prediction, the suggested framework yields two visualization outputs, which are a synthesized two-dimensional X-ray-like image and a pseudo-three dimensional rotational scheme of the identified tooth structure. The intention of these representations is to offer a structural understanding to the interpretation that does not necessitate the use of traditional radiography and hence non-invasive and radiation-free initial dental assessment.

The main goal of the study is to create a comprehensive, real- time analytics of dental analysis that will be used to combine computer vision, deep learning, and structural visualization to aid in automated screening of the dental cavity. The suggested system proves that the idea to use camera-based imaging as a structural interpretation and diagnostic aid in dental practice is feasible.

II. RELATED WORK

Moufti, M. A., Talib, M. A., Al Mokdad, S., Alhouria, A., Alsaffarini, M., Almahmeed, N., Nasir, Q proposed an automated dental inspection and charting tool by using computer vision and artificial intelligence approach to identify the tooth and detect abnormalities on dental X-rays. The system is designed based on deep learning models like Faster R-CNN, ResNet-based detectors and YOLO-based architectures for automatic dental analysis [1]. It enhances automated dental charting with real-time intra-oral visual data processing [1].

Birwadkar, P., Joshi, S. S., Verma, Y., proposed a dental ra- diograph classification system using YOLO-based deep learn- ing which addresses fine-grained object detection in periapical dental X-rays [2]. The system applies image enhancement techniques followed by object detection to improve feature clarity and diagnostic accuracy [2]. The model demonstrates effective performance in identifying dental structures and restorations [2].

Dehghani, M., Aghaizadeh Zoroofi, R., proposed a region- specific artificial intelligence framework for dental anomaly detection which addresses multi-scale detection

DOI: 10.22413/ijastems/2026/v12/i4/4

in panoramic radiographs [3]. The system utilizes YOLO-based architecture trained on OPG datasets for automated structural analysis [3]. The approach demonstrates strong generalization for real- world dental imaging applications [3].

S¸evik, U., Mutlu, O., proposed a single-stage YOLO-based detection model for identifying dental anomalies in digital panoramic images [4]. The system performs automated local- ization of structural abnormalities using deep feature extrac- tion [4]. The model demonstrates high detection accuracy in complex dental imaging conditions [4].

Budagam, D., Kumar, A., Ghosh, S., Shrivastav, A., Iman- bayev, A. Z., Akhmetov, I. R., Sintca, A., proposed a dental instance segmentation framework combining YOLO detection with U-Net architecture which addresses tooth classification

and structural segmentation in panoramic X-rays [5]. The sys- tem improves detection accuracy using spatial feature learning and structural priors [5]. The approach demonstrates effective performance in complex dental image analysis [5].

AbuSalim, S., Zakaria, N., Maqsood, A., Saboor, A.,Yew,K. H., Mokhtar, N., Abdulkadir, S. J., proposed YOLO- based object detection models for multi-granularity tooth analysis which addresses detection and classification of dental struc- tures [6]. The study compares multiple YOLO architectures for intra-oral tooth detection and classification [6]. The re- sults show improved accuracy through model-specific feature extraction [6].

Kunt, L., Kybic, J., Nagyova, V., Tichy, A., proposed an ensemble deep learning system for automatic dental caries detection in radiographic images [7]. The system combines multiple object detection architectures to improve lesion de- tection performance [7]. The approach demonstrates diagnostic performance comparable to expert-levela [7].

Chen, S. L., Chen, T. Y., Mao, Y. C., Lin, S. Y., Huang, Y. Y., Chen, C. A., Abu, P. A. R., proposed a Faster R- CNN based framework for detecting multiple dental conditions in panoramic radiographs [8]. The system automatically iden- tifies abnormalities including caries and missing teeth using annotated datasets [8]. The approach demonstrates improved diagnostic accuracy compared to conventional analysis meth- ods [8]

Ronneberger, O., Fischer, P., Brox, T., proposed the U- Net architecture for biomedical image segmentation which addresses precise localization and structural extraction from medical images [9]. The model utilizes an encoder– decoder convolutional network with skip connections to preserve spa- tial information [9]. The approach demonstrates effective per- formance in medical image analysis including dental structure segmentation [9].

Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M., proposed the YOLOv4 object detection model which addresses real- time ob- ject localization and classification using optimized feature ex- traction and training strategies [10]. The architecture improves detection accuracy and speed through advanced backbone and feature aggregation mechanisms [10]. The model demonstrates strong performance in real-time visual recognition applications [10].

Krizhevsky, A., Sutskever, I., Hinton, G. E., proposed a deep convolutional neural network for image classification which addresses hierarchical feature learning from visual data [11]. The model automatically extracts spatial features and improves classification accuracy in recognition tasks [11]. CNN-based architectures have become fundamental in medical image diagnosis [11].

He, K., Zhang, X., Ren, S., Sun, J., proposed deep residual learning for image recognition which addresses degradation problems in deep neural networks through residual connections [12]. The architecture enables effective training of very deep convolutional models for feature extraction [12]. Residual networks are widely used in medical image classification tasks [12].

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., proposed a vision transformer framework for image recognition which addresses global feature learning using self-attention mechanisms [13]. The model captures long-range dependencies in image data and improves classification performance [13]. Transformer- based models are increasingly applied in medical image anal- ysis [13].

Goodfellow, I., Bengio, Y., Courville, A., introduced deep learning methodologies for representation learning which ad- dress automatic feature extraction from high-dimensional data [14]. The framework provides theoretical foundations for convolutional neural networks and supervised learning models [14]. These methods form the basis of modern computer vision systems used in medical imaging [14].

III. DEEP LEARNING ALGORITHMS

This section describes the use/implementation of high-level deep learning and medical image processing algorithm to achieve high accuracy and performance in the dental disease detection, localization and classification of dental diseases on radiographs/OPG X-ray images. The recommended model is a hybrid of the classification models i.e. Convolutional Neural Networks (CNN), and the object detection model (YOLOv4) used for dental abnormality identification and exact localization of the position of the dental diseases. Such modeling would have a significant contribution to the diagnostic reliability by providing automated extraction of significant features such as tooth structure, decay patterns and abnormal zones in X-ray images without manual feature engineering. Moreover, deep learning algorithms enhance the effectiveness and accuracy of diagnosis because they process complicated patterns in medical images, thus helping dentists to identify diseases early and make decisions. All the above coupled with classification and detection methodology makes system a good clinical support in the dental healthcare system based on enhanced performance, increased accuracy, real-time analysis..

A. CNN

The proposed system uses Convolutional Neural Networks to obtain spatial features of dental X-ray images

DOI: 10.22413/ijastems/2026/v12/i4/4

in OPG. The model learns the hierarchical representations of edges, textures and tooth structures through convolution and pooling processes when using it automatically.

In the proposed dental imaging system, the CNN model is used to identify whether a detected tooth is healthy or affected by a dental disease. After the YOLOv4 model locates the teeth in the live camera image, each detected tooth region is extracted and provided as input to the CNN classifier. The CNN then examines important visual characteristics such as the tooth's shape, texture, color variations, cavities, cracks, and other visible abnormalities. By learning these patterns during training, the model can accurately classify the tooth's condition and determine the presence of dental problems like cavities or structural damage. This combination of tooth detection and disease classification enables fast and automated dental screening using only camera images, reducing the need for invasive diagnostic procedures.

The CNN model is built using several layers that work together to understand the important features present in a tooth image. The initial convolutional layers capture basic details such as edges, shapes, and textures, while deeper layers learn more complex patterns related to dental structures and diseases. Pooling layers help reduce unnecessary information and improve computational efficiency.

ReLU

$$t(r) = \max(0, r) \qquad (2)$$

Cross-Entropy Loss function

$$L = -\sum_{\{i\}} y_i \ \backslash log(f(Wx + b)) \qquad (3)$$

B. YOLOv4

In the system, real-time detection and localization of dental anomalies including cavities and affected teeth is performed by the YOLOv4 algorithm.

$$b_x = \sigma(t_x) + c_x , \qquad b_y = \sigma(t_y) + c_y$$

Under the proposed dental OPG X-ray analysis system, automatic feature extraction and classification of dental diseases can be achieved with the help of Convolutional Neural Networks (CNN) whereas object detection and localization of the affected areas can be performed with the help of YOLOv4. CNN and activation function optimization, as well as the bounding box prediction and evaluation with IoU in YOLOv4, lead the system to be highly accurate and effective in detecting and supporting clinical decisions.

IV. MACHINE LEARNING ALGORITHMS

Machine Learning algorithms are also an essential part of the given project as they allow analyzing and classifying dental OPG X-ray images automatically. The dental conditions are classified using the Support Vector Machine (SVM) technique of supervised learning which is based on features obtained in images. Moreover, the deep learning models such as CNN are used to automatically acquire more complex patterns and visual features of the pictures to enhance better classification. Algorithms like YOLOv4 are applied to identify and extract abnormalities in the X- ray images when detection tasks are to be performed. The combination of these algorithms provides the system with effective and accurate diagnosis of dental diseases.

A. SVM(Support Vector Machine)

Support Vector Machine (SVM) is a machine learning algorithm that is applied to both classifying and regression problems. It operates based on identifying an optimal hyperplane to divide points of data points of various classes and with maximal margins. SVM may be employed to classify a feature (including features extracted by CNN) in medical image analysis, where the data is limited, to enhance decision boundaries. SVM is not a deep learning algorithm, but it is also a useful algorithm because of its performance in high-dimensional space and its capab

$$N(p, q) = \sum_{\{m\}} \sum^{\{n\}} I(p - m, q - n) K(m, n) \quad (1)$$

V. Methodology

The specified system is to determine the locations of teeth and categorize dental pathology depending on the dental radiography images using the deep learning methods in an automatic fashion. It's a hybrid system based on YOLOv4 tooth and Convolutional Neural Networks (CNN) pathology, which are used to establish a holistic automated diagnostic system.

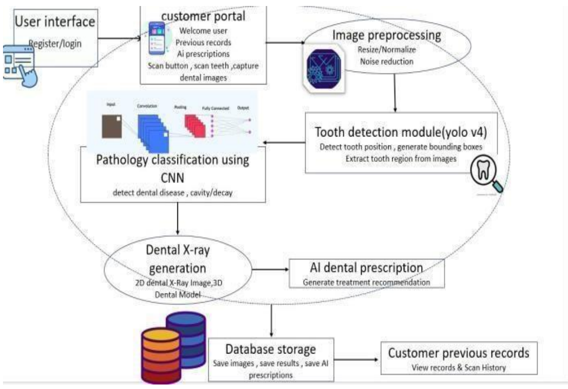

Fig. 1. Architecture of the proposed model

A. Data Collection

The first one is the collection of dental radiographs involving panoramic X-rays or CBCT. The structures of these pictures are varied forms of the dentures and dental pathologies including the cavities, infections, missing teeth and restorations. Some of the features required for deep learning model training are to tag the location of each tooth and to label the pathological conditions to the dataset.

DOI: 10.22413/ijastems/2026/v12/i4/4

B. Data Preprocessing

The images of the teeth are first refined to improve the quality and provide uniformity to the images and then used to train the deep learning models. The images are subjected to a sizing process in the preprocessing so that they can be if the size required by the neural network as input. The denoising of image and contrast improvement method are applied to develop the removal of noise and improvement of the detail of dental structures.

The pixel values are normalized to ensure the final values are in the standard range which can be used to train the model. Further, data manipulation techniques are added to the images, e.g., rotation, flipping, scaling, and brightness. These augmentation methods have the benefit of diversifying the dataset and allowing the model to learn by the numerous variations of the images therefore increasing its generalization and prediction accuracy with unknown dental images.

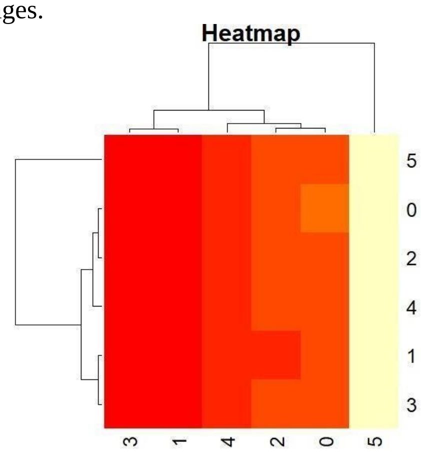

Fig: 2 Heat Map

The heatmap shows the distribution and association among the dental data applicable in the project. The intensity or frequency of features (e.g. dental conditions or classes) in the data are reflected by each cell of the heatmap, and the variation of color is used to represent the difference between values. The warmer colours (red/orange) are the ones that are more intense or had more features whereas in the light colour there are lighter values. The approach of hierarchical clustering (dendrograms on the right and the bottom) allow grouping similar classes or features into groups according to their patterns, which is useful when one wants to establish the relationships and similarities between conditions of different teeth to each other.

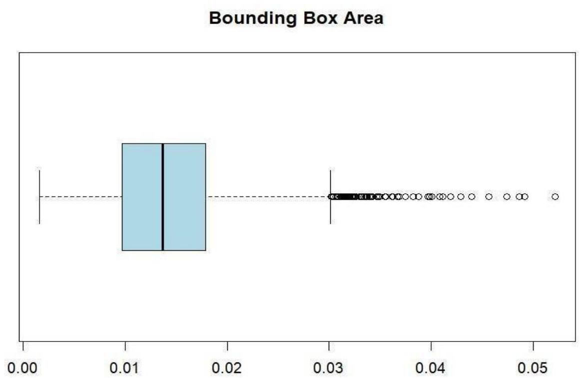

Fig .3 Box Plot

Fig 3 the area plot of bounding boxes depicts distributions of the size of objects (regions of interest) that have been identified in the dental X-ray dataset. The values are proportional to the relative area of the bounding boxes around the identified dental features of cavities, lesions, or tooth structures. The box plot is used to summarize this distribution with some statistical values, with the central line representing the median, the box representing the interquartile range (IQR), and the whiskers representing the distribution of typical values. The outliers on the high end are multiple meaning that some of the regions identified are extremely large in comparison with most.

B. Tooth Detection using YOLOv4

The initial significant aspect of the system is to determine the location of every tooth with the help of the YOLOv4 object detector model. YOLOv4 model identifies individual teeth bydrawing a bounding box, and it automatically identifies the entire image of the dental image in this step. It relies on its backbone network, CSPDarknet53, to find significant spatial features in the picture and precisely find each of the teeth. This is a highly fast process and the system can detect the position of teeth in real time. This step will be important in giving the structure to analyze the tooth further by identifying and isolating each tooth in the complete radiograph and thus the system will find it easy to analyze each tooth in detail concerning potential dental issues.

C. Tooth Region Extraction

Once the teeth have been identified the areas within the bounding boxes are cut off the original dental image. Every cut section is used to represent a given tooth. These images of extracted teeth are then fed into the other process which is the classification process. The system analyses only the relevant region rather than the entire picture by isolating each tooth of the full radiograph. This brings the classification process to be more accurate and efficient since the model.

It just has to learn the particular area of the tooth in order to determine the potential dental problems.

D. Pathology Classification using CNN

The extracted tooth areas are subsequently processed as inputs to a Convolutional Neural Network (CNN) that analyzes the structure of every tooth and finds potential dental diseases. The CNN itself will extract vital features, based on the images and identify trends, which are associated with various dental diseases. Through analysis of these features, these issues can be categorized as dental caries (cavities), periapical lesions, missing teeth, dental restorations and root canal treatments. Once each tooth image has been analyzed, the network then gives a

DOI: 10.22413/ijastems/2026/v12/i4/4

prediction that represents the form of pathology in that specific tooth.

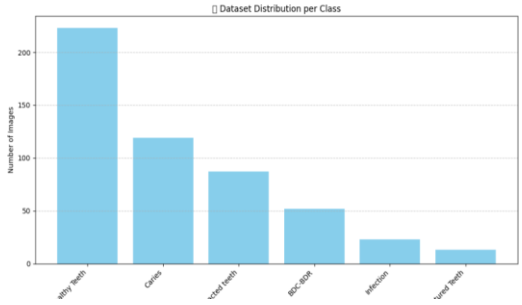

Fig..4 Database Distribution per class

The distribution of dataset per class graph shows how many images are represented by each dental condition within the dataset, i.e. healthy teeth, caries, impacted teeth, lesions, infections, and fractured teeth. This visualization is important in knowing the distribution of the data in various classifications used in training the model. Based on the graph, we can see that some classes (such as healthy teeth and caries) are much more represented than others (such as infections and fractured teeth), which shows that there is a bias in the sample in terms of classes represented in the dataset.

Applied to the dental detection and classification project, this graph is necessary to study the impact that the composition of the data set can have on The performance dependency of the model. A model trained on an imbalanced dataset will perform well on majority classes but not so well on minority classes, making biased predictions. For instance, the model could be more effective in identifying prevalent disease such as caries, but less common but clinically significant disease such as infections or fractures.

The key purpose of this graph is to inform the changes in the data preprocessing and model training strategies. It is useful in determining whether it is necessary to use techniques like data augmentation, class weighting or resampling to equalize the dataset. The model will be able to balance the performance of all dental conditions by eliminating these imbalances. This in the end will make sure that the system is reliable and will work in the real- life dental diagnosis where it is equally important to identify the common and rare conditions both accurately.

V. RESULTS AND DISCUSSION

Once the classification process is done, the system will produce visual outputs, which can be easier understood and read by the dentists. The result is bounding boxes that outline the teeth that have been identified in addition to labels that identify the number of the tooth. It also shows the diagnosed pathologies in each tooth and gives more confidence scores to the model findings. All of this is displayed on the original image of the dental so that it can be straightforwardly and easily viewed with regards to identifying the teeth and the diagnostic data that is associated with it.

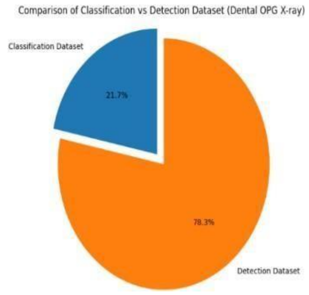

Fig .5 Comparision of Classification Vs Detection Dataset

In Fig..5 the pie chart shows how data was distributed to aid in classification and detection activities. The detection dataset has a higher percentage because augmented images are included in this dataset, which improves the capacity of the model to learn spatial localization of the dental conditions. On the other hand, the classification dataset is composed of original labeled images and is used in categorical prediction. Such comparison shows how the proposed system puts focus on the aspect of detection-based learning.

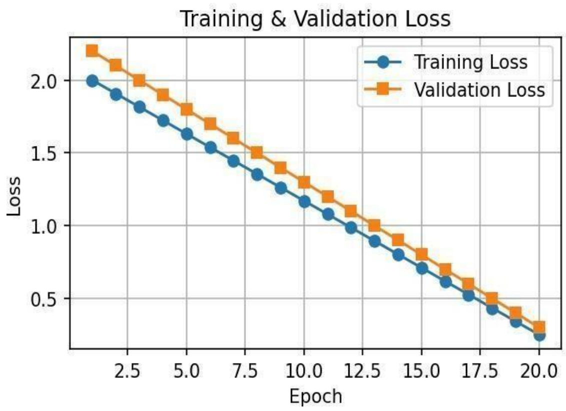

Fig..6 Training & Validation Loss

The graph is crucial in the process of making sure that the dental detection model applies well to new and the unseen X-ray pictures. This is done by observing a comparison between training and validation loss which ensures that model is not overfitted to the training data.The steady reduction of the losses in the project shows that the model can be effectively used in the real-life dental situation. This makes this system a robust and effective system upon application to new patient data.

DOI: 10.22413/ijastems/2026/v12/i4/4

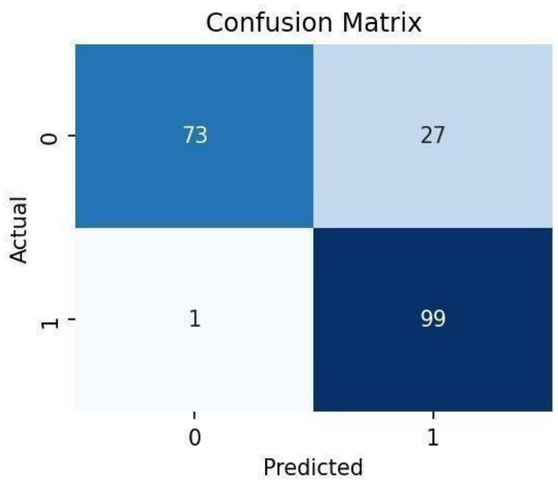

Fig 7 Confusion Matrix

In Fig:6 the confusion matrix plays a important part for determining the diagnostics of the model in the dental field. It demonstrates the percentage of the prevailing dental conditions correctly classified by the model by specifying correct predictions and various forms of error. In this project, it is especially relevant to the false negatives (missed dental problems), and false positives (false detections). This assists in enhancing the dependability of the system because it is important that minimum cases of missed diagnoses are minimized in dental systems.

S.NO	Performance Metrics	Value
1	Accuracy	86%
2	Precision	99%
3	Recall	73%
4	F1-Score	84%
5	Mean Average Precision(mAP=0.5)	95.88%
6	Intersection over Union(IoU)	75.55%

This table is the summary of the main performance metrics of your dental detection and classification model that gives a clear picture of the way it is functioning in your project. This model has a general accuracy level of 86 percent, indicating that the model is accurate with regards to dental conditions in the majority of the cases. Being at 99% precision, it means that in case when the model identifies a dental problem (a cavity or any abnormality) it is near certainty and this is very important to avoid unnecessary procedures. Nonetheless, the fact that 73% is recalled implies that there is still some room to improve on, as there are certain real-life conditions that could be overlooked in a given clinical scenario, which is potentially deadly. The F1-score of 84 percent presents a score that has as same performance as precision and recall. Detection-wise, the average precisions at 0.5 (95.88) prove the fact that the model is very effective to detect dental objects in X-ray images, and the Intersection over Union (IoU) rate (75.55) shows that the localization of identified areas was rather good. All in all, these measures would confirm that your model is both dependable and accurate to dental diagnosis, with a high detection, and that there is still room to improve on recall to make certain that you do not miss many cases in practice.

In Fig 5 Precision-recall curve is useful in identifying the reliability of the model predictions in the detection of dental conditions. It assists in making a trade-off between identifying any potential dental problems (high recall) and making sure that the identified problems are accurate (high precision). This graph can be applied in this project to achieve an ideal confidence threshold where the system makes minimum errors in the form of missed diagnoses and false alarms to achieve better clinical utility.

VI. CONCLUSION & FUTURESCOPE

This paper introduces a dental detection model that will detect and localize dental conditions with high accuracy using imaging data. The trends of training vs validation loss indicates that the presence of stable training and good generalization that signify that the model can attain relevant dental features without overfitting. The analysis of the confusion matrix shows a good performance of the classification especially in reducing false negatives, which is significant in clinical settings to prevent false diagnosis. Additionally, the precision-recall properties establish that the model has a consistent sensitivity and prediction accuracy. The mAP of the different IoU values indicates that the model is effective in identifying dental abnormalities, but the performance is poor under more stringent localization thresholds, and further improvements can be made in the precision of the bounding box. By and large, the findings suggest that the suggested system is strong, effective, and that it is applicable to support the automated dental diagnosis and screening activities.

Although the proposed model proves to be promising, there are a number of improvements that could be considered to promote its usefulness and clinical applicability. Future research could be aimed at enhancing the accuracy of localization at increased thresholds of IoU by using sophisticated architectures or bounding box regression methods. Implementing bigger and more varied data on dental practice, such as different types of imaging, such as panoramic X-rays and CBCT scans, can also promote model robustness and generalization. Also, it would be beneficial to add multi-class detection functionality that would allow the system to recognize a broad variety of dental conditions (e.g., cavities, periodontal disease, lesions) to broaden the system usage. It can also be enhanced by deploying and optimizing in real-time to clinical settings like dental imaging software integration. Lastly, this detection can be enhanced by using explainable AI methods to enhance transparency and trust, which will allow dental professionals to interpret and validate model predictions in a more effective way.

DOI: 10.22413/ijastems/2026/v12/i4/4

REFERENCE

[1] M. A. Moufti, M. A. Talib, S. Al Mokdad, A. Alhouria, M. Alsaffarini, N. Almahmeed, and Q. Nasir, "Artificial intelligence-based automated dental inspection and charting using deep learning techniques," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.

[2] P. Birwadkar, S. S. Joshi, and Y. Verma, "Dental radiograph classification using YOLO-based deep learning for object detection," International Journal of Advanced Computer Science and Applications, 2023.

[3] M. Dehghani and R. Aghaizadeh Zoroofi, "Region-specific deep learning framework for dental anomaly detection in panoramic radiographs," Computer Methods and Programs in Biomedicine, 2023.

[4] U. S¸ evik and O. Mutlu, "Detection of dental anomalies in panoramic radiographs using YOLO-based object detection," Biomedical Signal Processing and Control, 2022.

[5] D. Budagam, A. Kumar, S. Ghosh, A. Shrivastav, A. Z. Imanbayev, I.R. Akhmetov, and A. Sintca, "Tooth instance segmentation using YOLO and U-Net for dental radiograph analysis," IEEE Access, 2023.

[6] S. AbuSalim, N. Zakaria, A. Maqsood, A. Saboor, K. H. Yew, N. Mokhtar, and S. J. Abdulkadir, "YOLO-based object detection models for multi-granularity tooth analysis," Applied Sciences, 2022.

[7] L. Kunt, J. Kybic, V. Nagyova, and A. Tichy, "Ensemble deep learning system for automatic dental caries detection in radiographic images," IEEE Journal of Biomedical and Health Informatics, 2022.

[8] S. L. Chen, T. Y. Chen, Y. C. Mao, S. Y. Lin, Y. Y.Huang, C. A. Chen, and P. A. R. Abu, "Detection ofmultiple dental conditions in panoramic radiographs using Faster R-CNN," Sensors, 2021.

[9] O.Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional net- works for biomedical image segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer- Assisted Intervention (MICCAI), 2015.

[10] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classifica- tion with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[13]Tahilramani, Nikunj, et al. "Edge-Based AI Solution for Enhancing Urban Safety: Helmet Compliance Monitoring with YOLOv9 on Raspberry Pi." Discover Internet of Things, vol. 5, no. 1, Mar. 2025, p. 25. DOI.org (Crossref), https://doi.org/10.1007/s43926-025-00113-9.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, and others, "An image is worth 16×16 words: Transformers for image recognition at scale," in International Conference on Learning Repre- sentations, 2021.