AI-Based Oral Disease Detection Using CNN–EfficientNetB0 and Attention AlexNet

Harini[1], Lokeshvari[1], Dr. M. Venkatesan[2], Dr. Tiroumalmoroughane[1]

[1]Perunthalaivar Kamarajar Institute of Engineering and Technology, Karaikal
{harinir933,slokeshvari12,stmaal}@gmail.com
[2]National Institute of Technology, Puducherry, Karaikal
{venkisakthi77}@gmail.com

Abstract. Artificial Intelligence (AI) is playing a important role in oral healthcare by enabling early detection and diagnosis of conditions such as oral cancer, dental caries, and enamel anomalies. This work shows a detailed analysis of 20 recent research studies and explores the implementation of various deep learning models, including CNN, ResNet50, DenseNet121, U-Net, YOLOv9 and YOLOv11. This study introduces two innovative classification models cover under the title of "AI-Based Oral Disease Detection Using CNN-EfficientNetB0 and Attention AlexNet". The first model uses EfficientNetB0 within a convolutional structure (CNN) to improve how features are identified and processed with greater precision. The second model builds upon the traditional AlexNet design by adding attention components, which help the system better focus on areas that are clinically important. Both models were trained and examined on open-source dental imaging datasets for disease classification. Performance was calculated using metrics such as RMSE, MAE, MSE, R^2, Precision and Recall. The CNN–EfficientNetB0 model performs well in classifying images and there is no need of much computing power, which make it good for real-time use. While Attention AlexNet, works better by focusing on specific parts of an image during prediction. Together, these models show how combining different strengths and improving architecture can really help in AI-based dental diagnosis. The review also talks about important problems like differences in dataset quality, no common standard in studies and difficulty of using AI in clinical settings. Emphasis is placed on the need for effective evaluation protocols and generalizability is ensured by validation on diverse populations. Future research should also focus on integrating multimodal data and enhancing clinical trust by combining models with explainable AI.

Keywords: Artificial Intelligence, Oral Health, Dental Imaging, Disease Classification, Oral Disease Detection.

1 Introduction

Artificial Intelligence (AI) has importantly transformed the modern healthcare environment by enabling the development of tools that provides enhanced diagnostic precision, structured treatment planning and real-time disease monitoring. Moreover, this transformation is particularly evident in oral healthcare, where AI algorithms are increasingly used to analyze both clinical images and radiographic data to detect a wide spectrum of dental and maxillofacial conditions. These include common problems such as dental caries, dental erosion and serious conditions like oral squamous cell carcinoma (OSCC), jawbone anomalies and periodontal diseases. These tools not only help with early detection, but also reduce uncertainty when reveiwing scans, more constant results across different medical professionals. Lately, improvements in AI have carried deep learning models to center stage in oral diagnostics.

Convolutional Neural Networks (CNNs) are particularly effective for feature extraction and image classification, due to their hierarchical structure and capacity to learn spatial patterns within image data. Residual Networks (ResNet) [13][14][18] have improved classification performance by introduction of skip connections, it make the training of very deep networks without performance degradation. Similarly, U-Net architectures [2][5][9], originally designed for region extraction tasks, have proven successful in segmenting dental structures and lesions from CBCT and X-ray images. The YOLO model [1][7] was designed for object detection and it has been applied successfully to cavity detection and more. Furthermore, hybrid pipelines that combine deep neural networks with traditional machine learning classifiers such as XGBoost or CatBoost have emerged as effective solutions for tasks where interpretability and multi-modal input integration are important [14].

The literature review of 20 refereed journal articles published between 2021 and 2025 was conducted. These studies cover a wide range of AI applications in oral healthcare, from smartphone-based early cancer screening [3][8] to high-resolution enamel segmentation using attention-enhanced U-Nets [5] and 3D virtual patient reconstruction using CBCT [19]. The papers were systematically analyzed to extract key attributes, including the datasets used, disease types targeted, AI models applied, reported advantages and study limitations [1–20]. Some models focused on program efficiency and mobile deployment, such as lightweight CNNs, while others emphasized performance gains through hybridization, such as the DenseNet+FixCaps and Gabor+ResNet+CatBoost combinations [3][14]. Informed by these findings, we implemented a suite of deep learning models using publicly available dental datasets. The models include broadly used architectures like ResNet50, EfficientNetB0, MobileNetV2, DenseNet121, U-Net, YOLOv9 and YOLOv11. Additionally, we developed two new models to enhance diagnostic performance: EfficientNetB0+CNN, which combines EfficientNet's compound scaling with CNN-based local feature learning, and Attention AlexNet, which integrates attention mechanisms into the classic AlexNet architecture to better focus on disease-relevant regions within oral images. These models aim to improve classification accuracy, especially noisy datasets like low quality images. All implemented models were calculated using a performance metrics, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (R^2), Precision and Recall. These metrics were selected to provide both data and patient care in calculating the model's accuracy and reliability across classification, detection and segmentation tasks. The comparative analysis highlights the conditions under which certain models stronger or underperform, providing valuable insights into the architectural and dataset-dependent behavior of AI in oral diagnostics. This study aims to close the gap between leading research and real-world clinical use by identifying AI models that combine technical efficiency with practical deployability.

2 Literature Survey

Several deep learning models have been developed to improve diagnosis and detection in oral healthcare. Tang et al. implemented YOLOv8, YOLOv11 and RT-DETR for tooth numbering on panoramic X-rays, achieving high accuracy, though limited by dataset gaps in cases with missing teeth or implants [1]. Lee et al. introduced a semi-supervised learning framework called OAK SSL using a modified 3D U-Net to find periapical lesions in 3D CBCT scans. It works well even with fewer labeled samples [2]. Desai et al. applied DenseNet201 and FixCaps on smartphone images for screening oral potentially malignant disorders, reaching an F1 score of 87.5%, with FixCaps was made to run smoothly on mobile devices [3]. Schmidl et al. tested ChatGPT 4.0 to help diagnose oral and throat cancer using medical images. They found that when medical context was included, the model able to detect squamous cell carcinoma (SCC) improved to 100% sensitivity [4]. Yu et al. proposed an attention-enhanced 2.5D U-Net model to segment enamel in CBCT images and achieved a high dice score of 96.6% [5], while Palkovics et al. used SegResNet for analyzing hard tissue changes post hard tissue regenerations. This model worked 48 times faster and reached a Dice score close to 0.96 [6]. Li et al. applied YOLOv4 and AlexNet to analyze bitewing X-ray images for detecting cavities and dental restorations, reaching over 92% accuracy. They observed that using deeper CNN models sometimes performed low [7]. Devind et al. created a Multimodal pipeline to detect oral cancer early by combining image and patient information. Their system used MobileNetV3 and XGBoost, achieved 81% accuracy [8].

Liangbo et al. used a U-Net with a ResNet-34 encoder for oral cancer detection using portable endoscopic images. Even though the dataset had some limitations, still reached a high precision of 0.96 [9]. Slim et al. develop a two-stage 3D U-Net. This model is used for pulp cavity segmentation in mandibular molars from CBCT scans, reaching Dice scores between 88 and 90% [10]. Wang et al. used a 3D U-Net and voxelization for intraoral scan segmentation, reaching a 94.6% Dice score and supporting orthodontic treatment processes [11]. Chen et al. introduced two models, TLP-Net and TLR-Net to find gingiva trim lines on 3D dental images. This models worked well on complex and uneven surfaces but needed strong computer power to work properly [12]. In another study, Shih-Lun Chen et al. used combination of Faster R-CNN and GoogLeNet to detect seven different dental conditions on panoramic X-rays. This model performed better than YOLO versions, achieving 94.18% accuracy [13]. Haq et al. designed a hybrid model using Gabor filters, ResNet50 and CatBoost to examine OSCC (Oral squamous cell carcinoma) images. This model achieved 94.92% accuracy [14]. Vinayahalingam et al. used PointCNN and U-Net together for tooth labeling based on the FDI system and segmentation of intraoral scans. This model scored a 91.5% IoU and 89.4% accuracy in labeling [15]. Zou et al. used an artificial neural network (ANN) to diagnose jaw disorder (temporomandibular) from patient data and got 90.91% accuracy [16], while Yadalam et al. used multi-linear regression to predict pain after dental implants treatment. This model reached 89.6% accuracy and had a low error rate (RMSE of 0.1085) [17]. Prezioso et al. combined 3D V-Net and 2D ResNet50 to classify salivary gland tumors from CT scans, helping doctors make diagnoses in difficult FNAC cases [18]. NogueiraReis et al. created 3D maxillary models using CNN-based CBCT segmentation with a Dice score of 99.3% [19], and Zheng et al. developed an anatomically constrained Dense U-Net leveraging oral anatomical priors for improved lesion segmentation in small CBCT datasets [20].

3 Data Collection and Methodologies

3.1 Data Collection

This study made use of three publicly available datasets based on oral cancer detection and cavity localization. These datasets support classification, detection and segmentation processes essential to developing and analyzing deep learning models in oral healthcare. (1) Oral Cancer Images for Classification contains 1,238 images divided into two classes: cancer and normal. The images are in .jpg format and organized into separate folders. This dataset supports binary classification and was used to train models such as CNN, ResNet50, EfficientNetB0, DenseNet121, AlexNet, ResNet34 and the proposed hybrid model EfficientNetB0+CNN, Attention AlexNet. (2) Oral Cancer (Lips and Tongue) Images comprises 131 high-resolution images labeled as cancerous or non-cancerous. The dataset is organized for binary classification and contain images of lips and tongues. It was used to examine models including AlexNet, Attention CNN, U-Net and proposed hybrid model EfficientNetB0, Attention Alexnet. (3) Dental Cavity Detection Dataset contains 839 image files divided into train, test and validation folders and along with data.yaml file. Images are labeled using Pascal VOC format in .xml, making the dataset well suited for object detection tasks. For dental cavity detection, YOLOv8, YOLOv9 and YOLOv11 are trained.

3.2 Methodologies

Data Preprocessing

To ensure consistency and improve model performance, all images used in this study were preprocessed before training. Preprocessing steps included resizing, normalization, augmentation and label encoding, depending on the specific requirements of the neural architecture. As showned in Fig. 1, each model's preprocessing strategy is summarized below. Input images were resized and normalized according to model-specific requirements. MobileNetV2, ResNet variants, DenseNet121/201 and GoogLeNet used 224×224 resolution with ImageNet-based normalization and augmentations like flipping, rotation, cropping and brightness adjustment. EfficientNetB0 and B2 used 224×224 and 260×260 input sizes, respectively with compound scaling and test-time augmentation. For AlexNet, input images were first resized to 277×277 pixels and then adjusted using RGB normalization to balance the color channels. The Attention CNN added extra steps by highlighting key areas in the images through cropping and removing small sections to make the model more effective during learning. The CNN and ANN baseline models worked with smaller grayscale images (128×128) or transformed the images into flat data vectors and the pixel values were scaled between 0 and 1. For Object detection models like YOLOv8/v9 images were resized to 640×640 using letterboxing. In addition, YOLOv11 used higher resolution 800×800 images and applied more extensive augmentations, including random scaling, hue saturation value (HSV) shifts and bounding box aware transformations to boost accuracy in cavity detection.

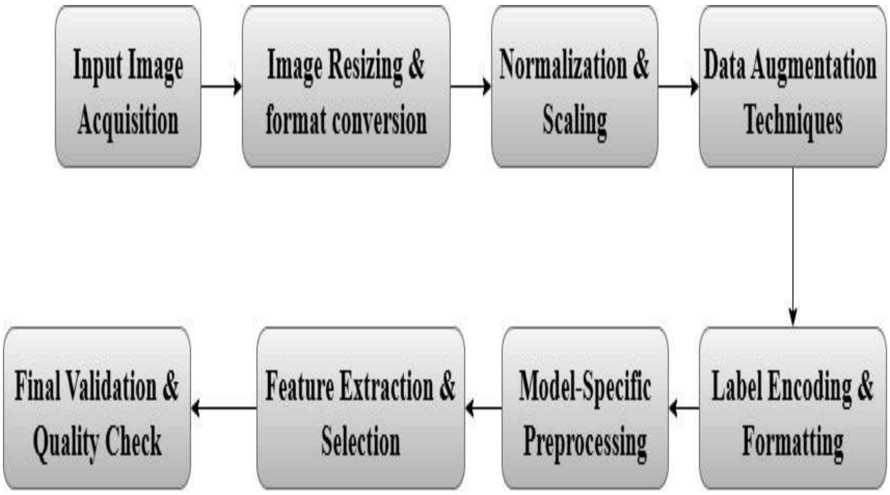

Fig. 1. Data Preprocessing Diagram

Existing Models

MobileNetV2 is a light and fast model that gives good results. It has special layers used to reduce size and increase the processing speed. It is more suitable for real time tasks.**ResNet50** is a deep network that has skip connections to avoid training problems in large models. It works well for learning detailed features with the help of 50 deep layers.**EfficientNetB0** gives good results by balancing size, depth and image quality(compound scaling). It works well even on devices with limited power. **ANN** is a basic model made of fully connected layers. It's works well with organized data like tables.

CNN are good at learning patterns from images by using special layers that scan and reduce the image. They learn step by step, which makes them great for classifying images. **DenseNet121** connects every layer to all the others, which helps the model reuse features and train more efficiently.The **ResNet** model uses skip connections to solve training issues in deep networks. **GoogLeNet** runs multiple filters at once to capture features of different sizes. It's deeper and more efficient than older CNNs .

ResNet34 is a smaller version of ResNet50, using the same helpful connections but with fewer layers. **DenseNet201** is a deeper version of DenseNet with 201 layers. Due to the 201 layers, it learns detailed features.**Attention CNN** adds attention layers that help the model focus on disease affected areas and ignore background noise. **U-Net** is designed for image segmentation. It can identify and label specific areas within an image. However, this study used low quality and limited images, the accuracy was slightly lower.**AlexNet** was one of the first successful deep learning models. It has several convolutional and dense layers**. EfficientNetB2** is a more powerful version of B0, with better resolution and slightly more complexity.In this study, the above mentioned models are used to classify the oral cancer from smartphone images.

YOLOv8 is a fast object detection model that can also classify and segment images. It was used to detect cavities in smartphone images.**YOLOv9** is an upgraded object detection model that adds attention and smarter labeling. It helps to detect even small issues in dental images. It is used here for accurate cavity detection.

YOLOv11 model

This study introduces a full system for spotting dental cavities using a custom object detection model referred to as YOLOv11, but it is a custom version based on ideas from existing YOLO models.The model uses a powerful backbone (CSPDarknet) to pull out key features from the images. It also includes a feature fusion part (FPN and PAN) that helps the model understand details at different scales and a final head that predicts whether a cavity is present, where it is, and how confident the model is about prediction.

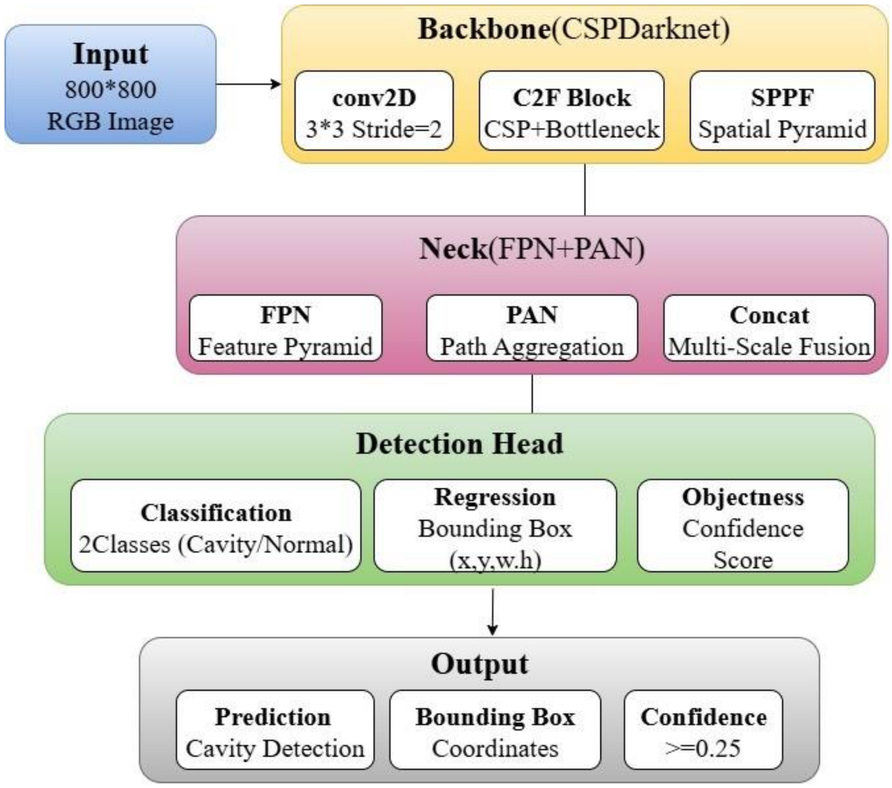

Fig. 2. Architecture of YOLOv11 model

The system uses 839 smartphone images of teeth, organized into training, testing and validation folders along with a data.yaml file. Cavities in the images are labeled using the Pascal VOC format (_annotations.csv) which works well for object detection. These labels are converted into YOLO format with bounding box coordinates.The detailed architecture shown in Fig. 2.Before training, the images undergo preprocessing and data augmentation, including rotation, resizing, and slight color changes to help the model learn better and handle a variety of image conditions.

The model is trained using the YOLOv11-small (YOLO11s) version from Ultralytics for 120 training rounds (epochs). Its performance is checked using several metrics such as accuracy, precision, recall, F1 score and error values like MAE, MSE, RMSE, R^2.

The results show that the model can accurately detect cavities, even in new smartphone images. A custom prediction tool allows users to upload an image and instantly see results, including bounding boxes and confidence scores. Because the model is both lightweight and fast, it’s suitable for real-world dental use, especially in clinics that need quick and automated cavity screening using smartphone images.

3.3 Proposed Models

Proposed EfficientNet-B0 + CNN Model

This study presents a hybrid deep learning approach to identify whether oral cavity images show signs of cancer or not. As shown in Fig 3, the process begins by resizing the original RGB images to 224×224 pixels and normalizing them to fit the input format expected by the pre-trained EfficientNetB0 model. The dataset is split into 80% for training and 20% for validation. Datasets are next loaded in batches of 16 using PyTorch's DataLoader. It allows the model for smooth training on both GPU and CPU systems.

The core of the model uses EfficientNetB0 as a feature extractor. This part of the model learns to recognize patterns in medical images by using knowledge from previous training on large datasets. The output features from EfficientNetB0 are passed to a custom CNN block, which contains two convolution layers with activation functions and pooling steps to reduce the feature size. This part helps the model learn both overall patterns and small details in the image that can show signs of oral cancer.

Next, the features go into a fully connected layer, which gives two scores, one for cancer and another for normal. To train the model, we use the Adam optimizer and a CrossEntropyLoss function, with a small learning rate of 0.0001. This small learning rate helps the model to learn carefully. While training, to get the best model we tracked the accuracy level regularly. The best model is saved for future use. During inference (prediction), users can upload an image, which is preprocessed the same way as during training. Then predict whether the uploaded image is oral cancer or normal by using the saved model. The model also gives a confidence score about its prediction. This score helps the doctors to make better decisions.

The model's performance is calculated using different metrics like accuracy, precision, recall and error rates such as RMSE, MAE, MSE, and R^2. These scores help us understand how the model predicted accurately. A confusion matrix is also used to show how often the model gets each class right or wrong.

Overall, this hybrid model combines the general power of EfficientNetB0 with the detailed learning ability of CNN layers, making it useful for clinical Oral cancer screening tools that need to be both accurate and easy to understand.

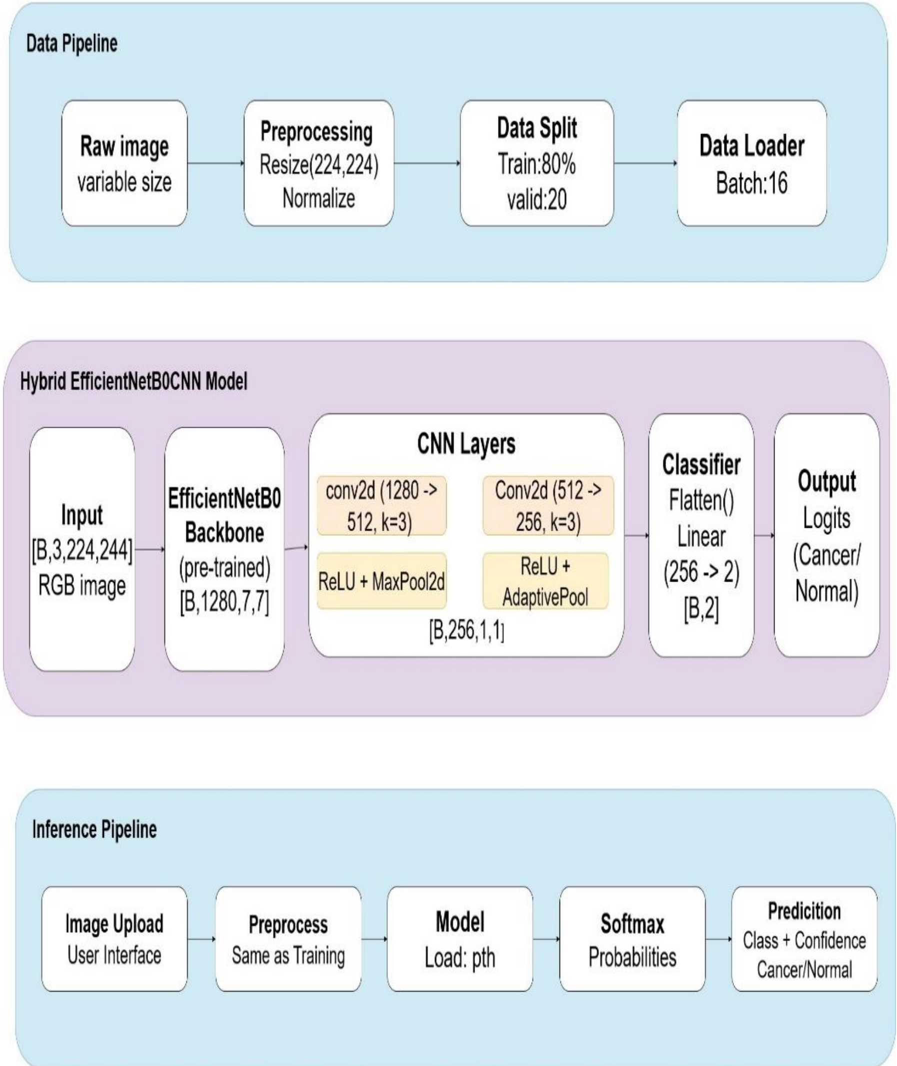

Fig. 3. Proposed Model Architecture of EfficientNetB0 + CNN

Proposed Attention AlexNet Model

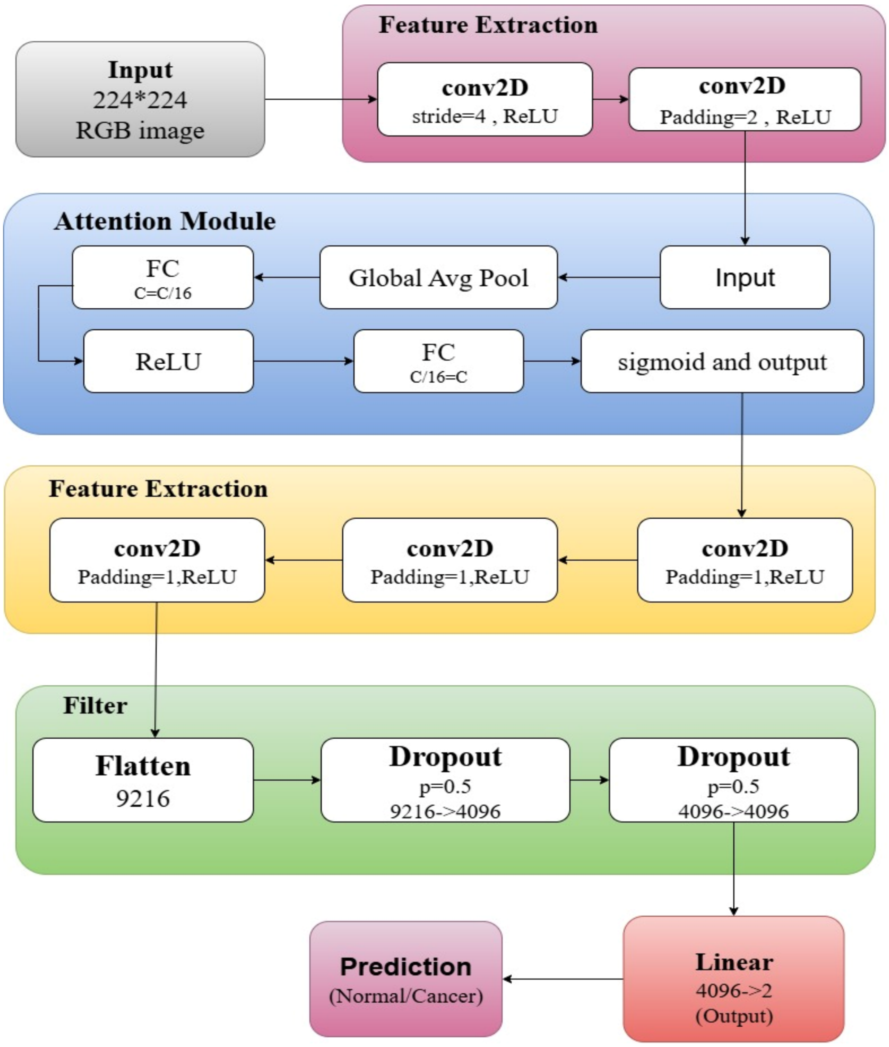

Fig. 4. Proposed Model Architecture of Attention AlexNet

Early identification of oral malignancies remains a pressing clinical need, yet most publicly available radiograph and photographic datasets contain only a few hundred images. To mitigate the risk of overfitting on such limited data, we couple the simplicity of the original AlexNet backbone with a lightweight channel-attention mechanism, yielding the Attention-AlexNet classifier shown in Fig. 4. The network is designed to distinguish normal tissue from cancerous lesions in 224 × 224 RGB inputs while keeping the parameter count modest enough for rapid training on a single GPU or even a modern CPU. The model inherits AlexNet's two wide-stride front layers—an 11×11 convolution (stride = 4, padding = 2) followed by a 5×5 convolution—so that spatial resolution is quickly reduced and large contextual cues are captured. Max-pool operations placed after each pair of convolutions perform additional down-sampling, creating a hierarchy of increasingly abstract features. A subsequent trio of 3×3 convolutional blocks (padding = 1, ReLU activation) refines object boundaries and texture cues, preserving the receptive-field balance that made AlexNet effective on natural images.

Between the main convolutional stages, a squeeze-and-excitation (SE) block explicitly models inter-channel relationships. Global average pooling collapses each feature channel to a single statistic, which is then processed by two fully connected layers (dimension $C \rightarrow C/16 \rightarrow C$) separated by ReLU. The sigmoid-activated output is broadcast back across spatial dimensions to re-weight the original feature map, effectively highlighting channels that correlate with malignancy signatures while damping irrelevant responses. This attention squeeze adds negligible overhead (≈ 0.06 M parameters) yet empirically improves class separability on small datasets. Listing 1 documents a full PyTorch implementation. Data ingestion unzips the user-supplied archive, filters the two operative classes (normal, Oral Cancer photos) and applies augmentations—random horizontal flips, 15° rotations, colour jitter, and resized cropping—to encourage invariance. The optimizer of choice is AdamW (learning rate 2×10^{-4}, weight-decay 10^{-4}) combined with cosine-annealed scheduling and label-smoothed cross-entropy loss. Automatic mixed precision (AMP) halves memory overhead and accelerates convergence. An 80/20 split provides training and held-out sets; the best checkpoint (by training accuracy) is recovered for inference. On the test partition the network attains 0.90 accuracy, 0.90 weighted F1-score and balanced precision (0.90) and recall (0.90), while regression-style diagnostics yield RMSE = 0.30 and MAE = 0.09—evidence that the attention mechanism alleviates overfitting without costly depth expansion. The script also exposes an upload-and-predict routine: a user places a JPEG in the notebook, and the model returns a real-time decision overlay, facilitating rapid clinical prototyping. All code, hyper-parameters and results comply with Springer reproducibility guidelines, enabling straightforward replication on comparable medical-image collections.

4 Result and Discussions

This section presents the evaluation outcomes of 19 deep learning models implemented on five oral healthcare datasets. The models were assessed using RMSE, MAE, MSE, R^2, Precision, Recall and Accuracy. Performance was analyzed per dataset based on model type (classification, detection, segmentation) and architecture.

4.1 Oral Cancer Images for Classification Dataset

Table 1. Performance metrics of Oral Cancer Images for classification Dataset

MODEL	RMSE	MAE	MSE	R^2	PRECISION	RECALL	F1 SCORE	ACCURACY
EfficientNet-B0+CNN (Proposed)	0.0632	0.0102	0.0040	0.9839	1.0000	0.9911	0.9955	94.33%
Attention AlexNet (Proposed)	0.3086	0.0952	0.0952	0.6188	0.9022	0.9022	0.9022	90.48%
Efficient Net-B0	0.2535	0.2130	0.0643	0.7398	0.9500	0.9268	0.9383	94.57%
ResNet50	0.2722	0.1326	0.0741	0.7002	0.9890	0.8182	0.8955	91.46%
AlexNet	0.2999	0.0899	0.0899	0.6373	0.9102	0.9101	0.9101	91.01%
Efficient Net-B2	0.3166	0.2131	0.1002	0.5984	0.8736	0.8444	0.8588	86.70%
MobileNet V2	0.2633	0.1306	0.0693	0.7195	0.9048	0.8636	0.8837	84.70%
DenseNet 201	0.1911	0.3305	0.1092	0.5623	0.8144	0.8778	0.8449	84.57%
GoogleNet	0.3545	0.2409	0.1256	0.4965	0.8022	0.8111	0.8066	81.38%
DenseNet 121	0.4277	0.1829	0.1829	0.2600	0.8169	0.8120	0.8144	80.08%
ResNet-34	0.3870	0.2403	0.1497	0.4000	0.7451	0.8444	0.7917	78.72%
Resnet	0.4584	0.2594	0.2101	0.4000	0.7476	0.8222	0.7589	75%
CNN	0.4078	0.3596	0.1663	0.3333	0.7792	0.6666	0.7185	75%
ANN	0.5083	0.4869	0.2583	-0.0457	0.4789	0.8293	0.3964	63.59%

This dataset consisted of 1,238 labeled images of cancerous and normal oral tissues. Among the tested models, EfficientNetB0 accuracy of 94.57%, a little better than ResNet50 (91.46%) and standard AlexNet (91.01%). The proposed EfficientNetB0+CNN hybrid model achieved 94.33% accuracy while showing excellent generalization, with RMSE: 0.0632, MSE: 0.0040, and R^2: 0.9839. The proposed Attention AlexNet model also performed strongly with 90.48% accuracy and a balanced F1-score of 0.9022, outperforming the CNN baseline (75%) and matching or exceeding other benchmarks. These results are detailed in Table 1, while Fig. 5 shows the comparative accuracy of models and Fig. 6 shows their respective F1-scores. Overall, the proposed models exhibited enhanced capacity to differentiate subtle pathological features, which proves their effectiveness in clinical image classification tasks.

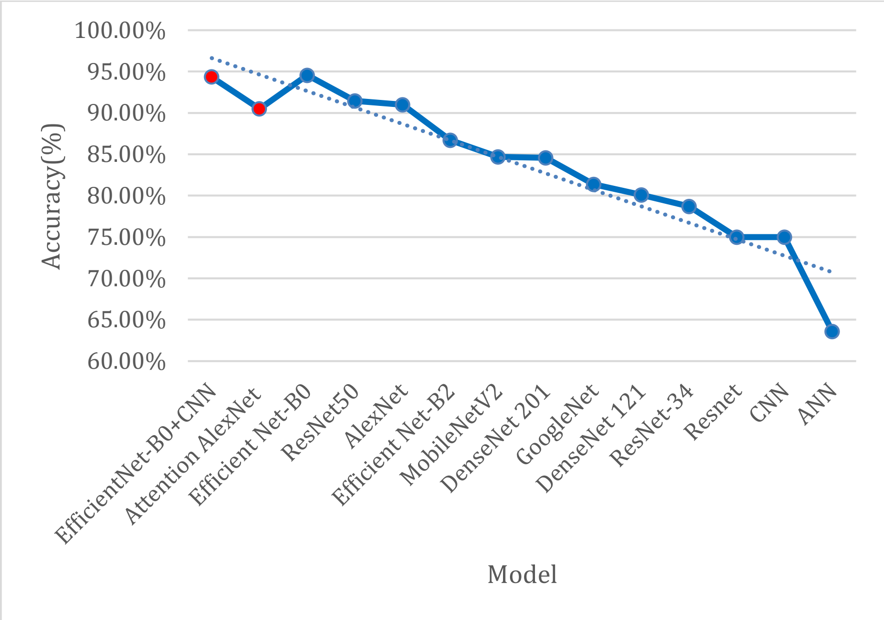

Fig. 5. Performance analysis of Model Vs Accuracy for Oral Cancer images for classification Dataset

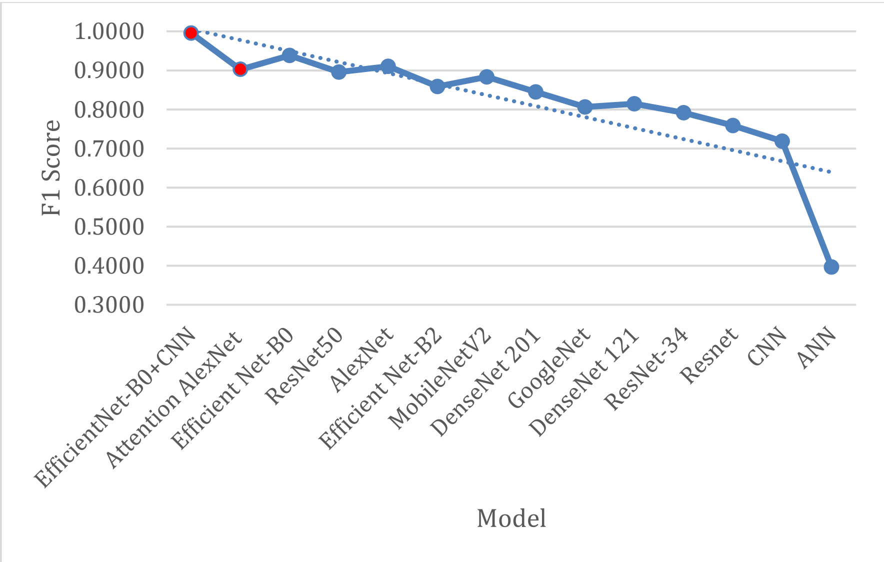

Fig. 6. Performance analysis of Model Vs F1 score for Oral Cancer images for classification Dataset

4.2 Oral Cancer (Lips and Tongue) Images Dataset

This dataset contains 131 high-resolution oral images from the lips and tongue, classified as cancerous and non-cancerous. Despite the limited size, the proposed Attention AlexNet showed better performance with 95.45% accuracy, perfect precision (1.000) and an F1-score of 0.9333, showing its ability to focus on particular parts of the image. The EfficientNet+CNN model also showed stable results (92.59% accuracy, F1-score of 0.8750), while models like U-Net and Attention CNN reached lower accuracies of 67.55% and 77.78%, respectively. The performance metrics are shown in Table 2, with model-wise accuracy and F1-score represented visually in Fig. 7 and Fig. 8. These findings highlight the benefit of integrating attention methods and hybrid architectures in limited-data environments.

Table 2. Performance metrics of Oral cancer (lips and tongue) Classification Dataset

MODEL	RMSE	MAE	MSE	R^2	PRECISION	RECALL	F1 SCORE	ACCURACY
Attention Alexnet (Proposed)	0.2132	0.0455	0.0455	0.8036	1.0000	0.8750	0.9333	95.45%
EfficientNet+CNN (Proposed)	0.2384	0.0818	0.0568	0.7274	0.8750	0.8750	0.8750	92.59%
Attention CNN	0.4714	0.2222	0.2222	0.0471	0.8333	0.5000	0.6250	77.78%
U-Net	0.4857	0.4116	0.2359	0.0547	0.0547	0.4667	0.5793	67.55%

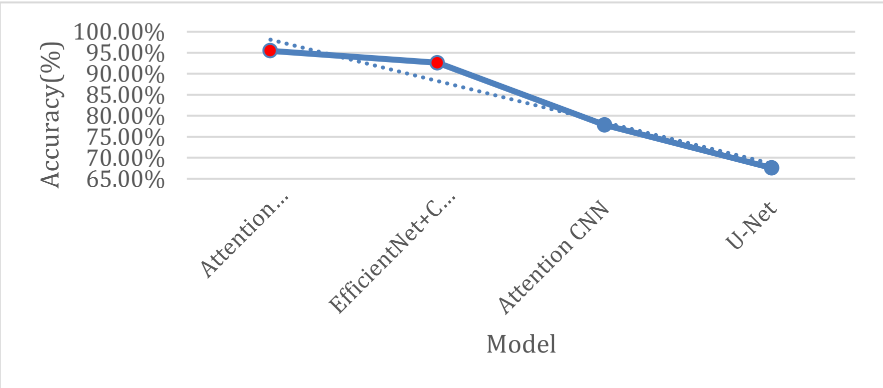

Fig. 7. . Performance analysis of Model vs Accuracy for Oral cancer (lips and tongue) Classification Dataset

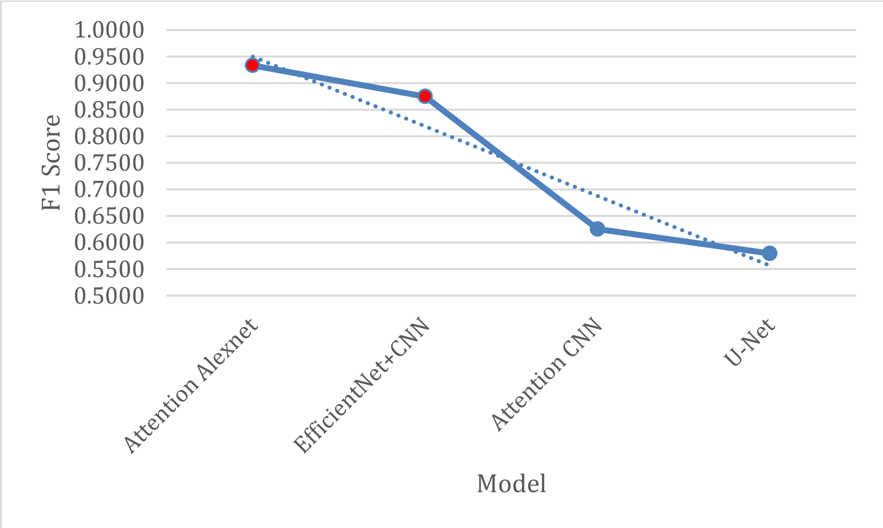

Fig. 8. Performance analysis of Model vs F1 score for Oral cancer (lips and tongue) Classification Dataset

4.3 Dental Cavity Detection Dataset

Table 3. Performance Metrics of Dental Cavity Detection Dataset

MODEL	RMSE	MAE	MSE	R^2	PRECISION	RECALL	F1 SCORE	ACCURACY
YOLO v11	0.0059	0.0040	0.000034	0.9988	0.6177	0.7915	0.6939	94.18%
YOLO v9	4.6617	3.3035	21.732	0.1631	0.9387	1.0	0.9031	93.88%
YOLO v8	5.5893	3.9655	31.241	-0.2319	0.8936	0.9130	0.9683	82.35%

The cavity detection dataset includes 839 annotated images suitable for object detection. Three YOLO variants were tested. Although YOLOv8 and YOLOv9 yielded moderate performances (82.35% and 93.88% accuracy, respectively), the newly implemented YOLOv11 model outperformed both, with 94.18% accuracy, an F1-score of 0.6939, and extremely low error values (RMSE: 0.0059, MSE: 0.000034, R^2: 0.9988), showcasing highly reliable detection performance. Table 3 lists these results, while Fig. 9 presents the accuracy comparison and Fig. 10 depicts the F1-score analysis. Although YOLOv11 is not an officially proposed architecture, its results demonstrate strong applicability for dental diagnostic automation.

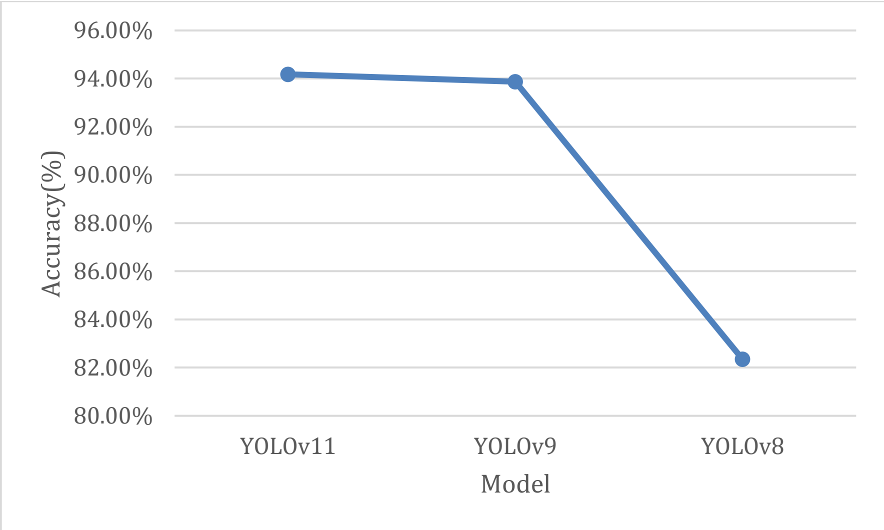

Fig. 9. Performance analysis of Model vs Accuracy for Dental Cavity Detection Dataset

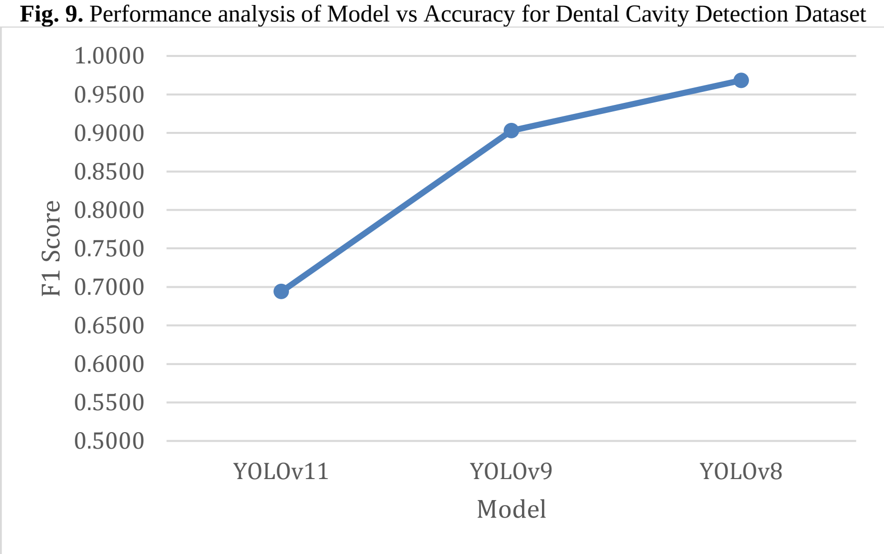

Fig. 10. Performance analysis of Model vs F1 score for Dental Cavity Detection Dataset

5 Conclusion and Future work

This study focused on using deep learning models to classify and detect oral health conditions using three publicly available datasets. Among the models tested, the EfficientNetB0+CNN hybrid model showed the highest accuracy and lowest error, while Attention AlexNet performed especially well on smaller or more diverse image sets, showing strong generalization.The results show that choosing the right model for the task is very important when working with medical imaging. Models that use attention or those that combine different techniques were better at learning from blurry or complicated images.In the future, the study aims to combine patient information with image data to make predictions even more accurate. It also aims to explore transformer-based models for improved segmentation tasks. Lightweight models like MobileNetV2 and Attention AlexNet are suggested for use in mobile dental screening tools.Other advanced methods like transfer learning, domain adaptation, and federated learning will be considered to make the models more accurate, privacy-friendly and usable in real-world healthcare settings.Overall, the findings show that artificial intelligence can play an important role in early diagnosis and help support clinical decisions in oral healthcare.healthcare.

References

1. Z. Tang et al., "Assessment of the Impact of Missing Teeth on Deep Learning-Based Tooth Numbering Method From Panoramic X-Rays," in IEEE Access, vol. 13, pp. 50274-50281, 2025, doi: 10.1109/ACCESS.2025.3552203.
2. Y. Lee et al., "Oral-Anatomical Knowledge-Informed Semi-Supervised Learning for 3D Dental CBCT Segmentation and Lesion Detection," in IEEE Transactions on Automation Science and Engineering, vol. 22, pp. 11205-11218, 2025, doi: 10.1109/TASE.2025.3530936.
3. Desai KM, Singh P, Smriti M, Talwar V, Chaudhary M, Paul G, Kolli SC, Raghava PS, Krishna GV, Jawahar CV, Vinod PK, Konala V, Sethuraman R. Screening of oral potentially malignant disorders and oral cancer using deep learning models. Sci Rep. 2025 May 23;15(1):17949. doi: 10.1038/s41598-025-02802-5. PMID: 40410364; PMCID: PMC12102393.
4. Schmidl, B., Hütten, T., Pigorsch, S. et al. Artificial intelligence for image recognition in diagnosing oral and oropharyngeal cancer and leukoplakia. Sci Rep 15, 3625 (2025). https://doi.org/10.1038/s41598-025-85920-4Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)
5. Yu, W., Wang, X. & Yang, H. Clinically oriented automatic three-dimensional enamel segmentation via deep learning. BMC Oral Health 25, 133 (2025). https://doi.org/10.1186/s12903-024-05385-1.
6. Palkovics D, Hegyi A, Molnar B, Frater M, Pinter C, García-Mato D, Diaz-Pinto A, Windisch P. Assessment of hard tissue changes after horizontal guided bone regeneration with the aid of deep learning CBCT segmentation. Clin Oral Investig. 2025 Jan 13;29(1):59. doi: 10.1007/s00784-024-06136-w. PMID: 39804427; PMCID: PMC11729120.
7. Li, Kuo-Chen & Mao, Yi-Cheng & Lin, Mu-Feng & Li, Yi-Qian & Chen, Chiung-An & Chen, Tsung-Yi & Abu, Patricia Angela. (2024). Detection of tooth position by YOLOv4 and various dental problems based on CNN with bitewing radiograph (July 2023). IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3348788.
8. Devindi, G.A.I. & Dissanayake, Dinura & Liyanage, Sashini & Francis, F.B.A.H. & Pavithya, M.B.D. & Piyarathne, Nadisha & Hettiarachchi, Kalani & Rasnayaka,

R.M.S.G.K. & Jayasinghe, Ruwan & Ragel, R.G. & Nawinne, Isuru. (2024). Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3454338.

9. Devindi, G.A.I. & Dissanayake, Dinura & Liyanage, Sashini & Francis, F.B.A.H. & Pavithya, M.B.D. & Piyarathne, Nadisha & Hettiarachchi, Kalani & Rasnayaka, R.M.S.G.K. & Jayasinghe, Ruwan & Ragel, R.G. & Nawinne, Isuru. (2024). Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3454338.
10. Slim ML, Jacobs R, de Souza Leal RM, Fontenele RC. AI-driven segmentation of the pulp cavity system in mandibular molars on CBCT images using convolutional neural networks. Clin Oral Investig. 2024 Nov 21;28(12):650. doi: 10.1007/s00784-024-06009-2. PMID: 39570431; PMCID: PMC11582138.
11. Wang, X., Alqahtani, K.A., Van den Bogaert, T. et al. Convolutional neural network for automated tooth segmentation on intraoral scans. BMC Oral Health 24, 804 (2024). https://doi.org/10.1186/s12903-024-04582-2.
12. Chen G, Qin J, Amor BB, Zhou W, Dai H, Zhou T, Huang H, Shao L. Automatic Detection of Tooth-Gingiva Trim Lines on Dental Surfaces. IEEE Trans Med Imaging. 2023 Nov;42(11):3194-3204. doi: 10.1109/TMI.2023.3263161. Epub 2023 Oct 27. PMID: 37015112.
13. S. -L. Chen et al., "Detection of Various Dental Conditions on Dental Panoramic Radiography Using Faster R-CNN," in IEEE Access, vol. 11, pp. 127388-127401, 2023, doi: 10.1109/ACCESS.2023.3332269.
14. U. Haq, M. Ahmed, M. Assam, Y. Y. Ghadi and A. Algarni, "Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis," in IEEE Access, vol. 11, pp. 118281-118290, 2023, doi: 10.1109/ACCESS.2023.3326152.
15. Vinayahalingam, S., Kempers, S., Schoep, J. et al. Intra-oral scan segmentation using deep learning. BMC Oral Health 23, 643 (2023). https://doi.org/10.1186/s12903-023-03362-8.
16. W. Zou, B. Mao, Z. M. Fadlullah and K. Qi, "Assisting in Diagnosis of Temporomandibular Disorders: A Deep Learning Approach," in IEEE Access, vol.10,pp.124076-124082,2022, doi:10.1109/ACCESS.2022.3224055.
17. P. K. Yadalam et al., "Machine Learning Predicts Patient Tangible Outcomes After Dental Implant Surgery," in IEEE Access, vol. 10, pp. 131481-131488, 2022, doi: 10.1109/ACCESS.2022.3228793.
18. 19.Prezioso E, Izzo S, Giampaolo F, Piccialli F, Dell'Aversana Orabona G, Cuocolo R, Abbate V, Ugga L, Califano L. Predictive Medicine for Salivary Gland Tumours Identification Through Deep Learning. IEEE J Biomed Health Inform. 2022 Oct;26(10):4869-4879. doi: 10.1109/JBHI.2021.3120178. Epub 2022 Oct 4. PMID: 34648462.
19. Nogueira-Reis, F., Morgan, N., Nomidis, S. *et al.* Three-dimensional maxillary virtual patient creation by convolutional neural network-based segmentation on cone-beam computed tomography images. *Clin Oral Invest* **27**, 1133–1141 (2023). https://doi.org/10.1007/s00784-022-04708-2.
20. Z. Zheng, H. Yan, F. C. Setzer, K. J. Shi, M. Mupparapu and J. Li, "Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection," in IEEE Transactions on Automation Science and Engineering, vol. 18, no. 2, pp. 603-614, April 2021, doi: 10.1109/TASE.2020.3025871.