



ISSN ONLINE: 2447-0228



RESEARCH ARTICLE

OPEN ACCESS

AUTOMATED DIAGNOSIS OF ORAL DISEASES USING ROI DETECTION AND TRANSFER LEARNING ON PHOTOGRAPHIC DENTAL IMAGES

Ashish Gupta^{*1}, Deepak Gupta², Shweta Nishit Jain³, Bhaskar Roy⁴,
Pradeep Yadav⁵, and Prashant Sharma⁶

¹Department of Information Technology, Institute of Technology and Management, Gwalior, India.

^{2,5}Department of Computer Science and Engineering, Institute of Technology and Management, Gwalior, India.

³Department of Electronics and Telecommunication Engg., Hope Foundation's International Institute of Information Technology, Pune, India.

⁴Department of Computer Science and Engineering (AIML), Asansol Engineering College Asansol, India.

⁶Department of Management, Institute of Technology and Management, Gwalior, India.

¹<https://orcid.org/0009-0002-4460-6605>, ²<https://orcid.org/0000-0003-3929-1362>, ³<https://orcid.org/0009-0001-5883-4679>

⁴<https://orcid.org/0000-0002-6516-3682>, ⁵<https://orcid.org/0000-0003-1073-7076>, ⁶<https://orcid.org/0009-0001-5883-4679>

Email: *guptaashishnitm@gmail.com, Deepak.gupta@itmgoi.in, dodal.shweta@gmail.com, bhaskar.ec11@gmail.com,
pradeep.yadav@itmgoi.in, prashant.sharma@itmgoi.in

ARTICLE INFO

Article History

Received: January 20, 2026

Reviewed: July 23, 2026

Accepted: August 23, 2026

Published: September 30, 2026

Keywords:

Oral disease

Artificial intelligence

Machine learning

VGG19

Convolutional neural networks

ABSTRACT

The early and precise diagnosis of oral diseases is essential for effective treatment and improved patient outcomes. This research presents a novel deep learning-based framework that leverages region of interest (ROI) detection and transfer learning to automate the classification of six common dental conditions using high-resolution photographic images. A dataset of 1,800 images, annotated by certified dental practitioners, was compiled from local clinics. The proposed pipeline begins with a binary classifier to determine whether ROI localization is required. For images needing localization, a lightweight YOLOv4-tiny detector was employed to isolate the relevant regions. These cropped ROIs were then used to fine-tune multiple pre-trained convolutional neural networks (CNNs), including VGG16, VGG19, InceptionV3, DenseNet201, and ResNet50. Among these, the VGG19 model achieved the high performance, with an accuracy of 95.85%, precision of 95.47%, and recall of 96.18%. The results demonstrate that combining efficient ROI detection with transfer learning significantly enhances diagnostic accuracy while maintaining computational efficiency, making the approach suitable for integration into mobile and clinical diagnostic systems. This work highlights the potential of AI-driven dental diagnostics to support early detection, reduce consultation delays, and improve accessibility to oral healthcare.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The growing global susceptibility to diseases has placed unprecedented strain on healthcare systems worldwide. To mitigate the spread of infectious diseases, hospitals, clinics, and other medical facilities have been compelled to follow strict safety protocols, which in turn has reduced the number of patients that can be treated at any given time. In response, researchers particularly in a field of Artificial Intelligence (AI) have been creating innovative methods and technologies to enable safe diagnosis and treatment with minimal direct contact. Among the promising areas of innovation is the automated diagnosis of dental images [1]. AI methodologies have been successfully applied across diverse domains, including natural language processing [2-4] computer vision [5-7] speech recognition [8], [9], biometrics, smart home systems, medical imaging, healthcare, robotics, banking, agriculture, defense, marketing, and even oil and gas exploration. In recent years, computer vision has proven to be a powerful tool in medical applications, enabling accurate diagnostics and reducing the likelihood of human error. The integration of AI into dentistry holds considerable promise and is anticipated to become a critical component of future dental healthcare, particularly for the diagnosis of oral diseases.

Deep learning techniques have shown impressive success in interpretation and analysis of dental imagery. Convolutional Neural Networks (CNN) and other neural network architectures are ideal for utilising x-ray and photographic datasets to perform tasks such as tooth segmentation, classification, numbering, and lesion detection [10]. Given the prolonged waiting times for dental consultations and the importance of early diagnosis, we developed an AI-based dental detection and classification system designed to provide preliminary assessments from simple images captured via a smartphone. The technique system utilizes a dataset of photographic images collected from local clinics in collaboration with a team of specialized dentists. This paper's remaining sections are arranged as follows: In Section 2 reviews relevant literature; Section 3 details the dataset and proposed methodology; Section 4 presents experimental results; and Section 5 concludes the work.

II. LITERATURE SURVEY

This section reviews existing methods proposed for automating the diagnosis of dental diseases. Prior studies have primarily relied on either X-ray imaging or photographic imaging for detection and classification tasks. By [11], developed an intelligent Dental Health-IoT system that is deployed on intelligent hardware, where a mobile device served as the terminal device for image acquisition. A MASK R-CNN model was trained on 126,000 collected images to perform disease detection. The system achieved 90% accuracy in detecting and identifying seven different dental conditions. In turn [12] utilized 800 X-ray images, applying adaptive histogram equalization to separate background bone structures from foreground teeth. Hybrid graph cut segmentation was then used to isolate the oral cavity and associated tissues. The deep learning architecture trained on these images that have been segmented that achieved a prediction accuracy of 97%. According to [13], employed Orthopantomogram (OPD) images and applied preprocessing techniques prior to training a Faster R-CNN for simultaneous detection and classification. Their model achieved 90% detection accuracy and 99% classification accuracy. In turn [14], Proposed OralCam, a mobile application enabling self-examination for five types of oral diseases. A dataset of 3,182 images of oral from five participants was used to train a CNN, achieving a 78.7% detection sensitivity. Duong et al. used a smartphone to take 620 imagery of the extracted molars. Four dentists manually categorised the photos into three classes, followed by a series of preprocessing techniques for image enhancement and segmentation. Classification was performed using an SVM classifier trained on color intensity features. By [15], Collected 640 photographic images of patients' oral cavities via smartphone. Image enhancement and data augmentation increased the dataset to 3,990 images. A YOLOv3 CNN model was fine-tuned through transfer learning to detect and classify two types of dental caries. By [16], Introduced CariesNet, a U-shaped deep learning architecture designed for the segmentation of 3,127 panoramic radiographs. To improve segmentation performance, the model integrated a full-scale axial attention module, achieving a 93.61% accuracy rate in identifying the three dental cavities severity levels.

According to [17] Developed a hybrid system combining X-ray images and photographic images to localize caries regions. A MASK R-CNN deep learning model was trained for segmentation, achieving 92% accuracy in detecting cavities and other oral diseases. In turn [18], Collected 1,902 photographic images from 695 participants to identify three types of dental cavities. Four DL architectures Faster R-CNN, SSD, RetinaNet, and YOLOv3 were fine-tuned for lesion detection. By [19], conducted a retrospective study using photographic images from 65 healthy subjects and 72 cases of oral lichen planus. An InceptionV3 network trained as a binary classifier achieved 100% accuracy in distinguishing healthy tissue from lesion cases. According to [20], Compiled 5,069 images representing six types of oral mucosal lesions, which were manually labeled and cropped by specialists. Four CNN architectures—ResNet-50, VGG16, InceptionV3, as well as Xception were trained using 70% of the dataset, with the remainder used for testing. By [21], Created a labeled dataset of four hundreds seventy Panorama X-ray images. A genetic algorithm was applied for tooth segmentation, combined with image processing operations to isolate individual teeth. Features extracted from multiple deep learning networks were fed into a capsule network, yielding an accuracy of 86.05%. By [22], This research demonstrates that combining patient metadata with image-based analysis through a multimodal deep-learning pipeline significantly enhances early detection of OSCC. Using MobileNetV3-Large, the model achieved 81% accuracy, outperforming image-only approaches. These results underscore the potential of multimodal integration to improve diagnostic accuracy, support clinical decision-making, and ultimately contribute to better patient prognosis and treatment outcomes.

III. MATERIALS AND METHODS

The initial stage of the research work involved the collection of dental image data. This process was conducted in collaboration with a team of specialist dentists at local clinics, producing in a dataset of 1,800 Panorama images representing six prevalent dental diseases. The dataset included images from male and female patients aged between 8 and 68 years. Photographs were captured using smartphones, enabling easy image acquisition, including from children, whereas obtaining X-ray images for younger patients posed greater complexity and risk of error. All collected data was anonymized by removing any personally identifiable information such as names, ages, medical history, or social details. Figure 1 illustrates sample images for each disease category.



Figure 1: Photographic representations of our dataset.
Source: Authors, (2026).

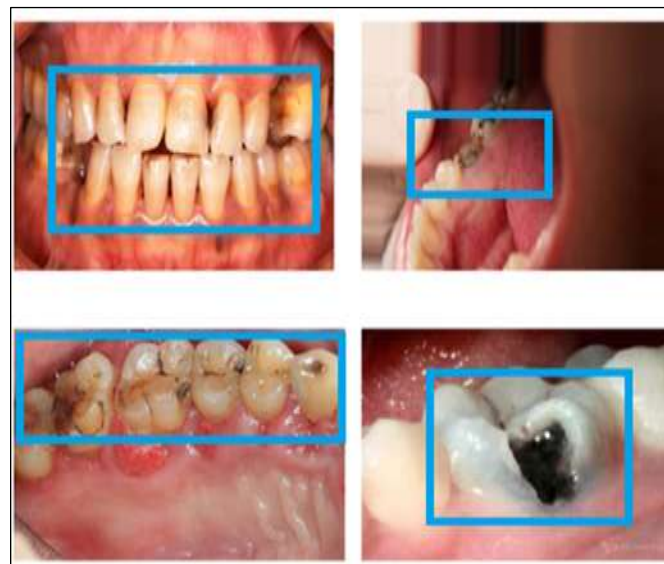


Figure 2: Data annotation for detector model training dataset.
Source: Authors, (2026).

In the subsequent stage, images were annotated by professional dentists, as the classification process required domain expertise. Three dentists manually labeled the dataset into six disease categories. The annotated dataset was split into 75% training and 25% testing subsets. The training dataset was utilised to develop predictive models for disease detection and recognition, while the testing set served for performance evaluation. Given the strong performance of Convolutional Neural Networks (CNNs) in image-based tasks, they were adopted for this project. The proposed deep learning pipeline begins with detecting and localizing the teeth region our Region of Interest (ROI) followed by cropping these regions for classification purposes. Data augmentation approach, including rotation, translation, scaling, and brightness adjustment, were used to expand the dataset, mitigate overfitting, and improve model robustness [23], [24]. The dataset contained two types of images: those with only the teeth region and those including additional facial features such as cheeks, lips, and jawlines. A binary classification step was introduced to determine whether an image required ROI detection. To train this binary classifier, 70 images per disease were selected 35 requiring ROI detection and 25 not using transfer learning (TL) with a previous trained MobileNetV2 model. Final classification layer of MobileNetV2 was replaced with a fully connected layer of sixty four nodes, and then a two-node SoftMax output layer.

During training, all original MobileNetV2 layers were frozen in order to maintain learnt weights, while only the new layers were optimized [25]. For ROI detection, a lightweight deep learning object detection model YOLOv4-tiny was employed due to its efficiency and reduced computational demands. Fifty images per disease category requiring ROI detection were labeled using the Bbox-Label-Tool-Multi-Class from the Darknet library, ensuring YOLO-compatible bounding box annotations. The pre-trained YOLOv4-tiny model was fine-tuned with these labeled images to detect and crop the ROI, which was then used for disease classification [26]. For disease classification, transfer learning was applied to the VGG16 network, modifying the architecture by replacing its final layer with a flattening operation, a fully connected layer of 512 nodes after that a dropout layer (drop rate: 0.4), another fully connected layer of 256 nodes with dropout, and a SoftMax output layer with six nodes representing the disease classes. Fine-tuning was performed to adjust network parameters for optimal feature extraction from the dental dataset. To enhance performance, the same fine-tuning approach was extended to additional pre-trained models [27], including VGG19, InceptionV3, Xception, InceptionResNetV2, GoogleNet, DenseNet201, MobileNetV2, as well as NASNetLarge. For each model, the classification head was replaced with two fully connected layers (512 and 256 nodes) separated by dropout layers, and a SoftMax layer with six outputs. The comparative results from these experiments are shown in the subsequent section.

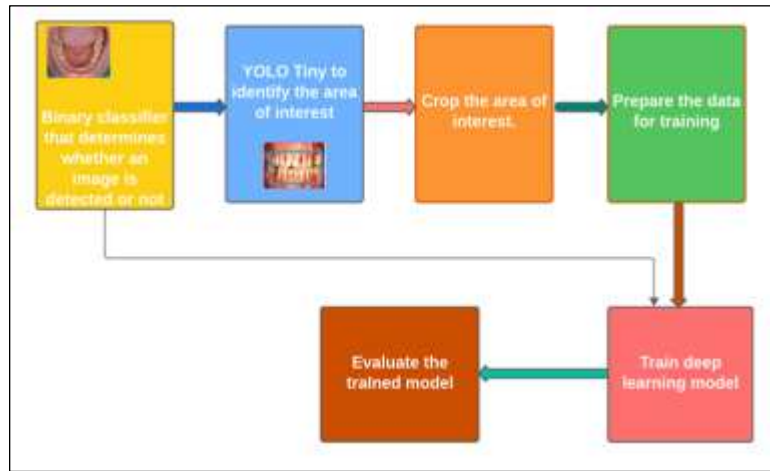


Figure 3: Workflow of proposed method.

Source: Authors, (2026).

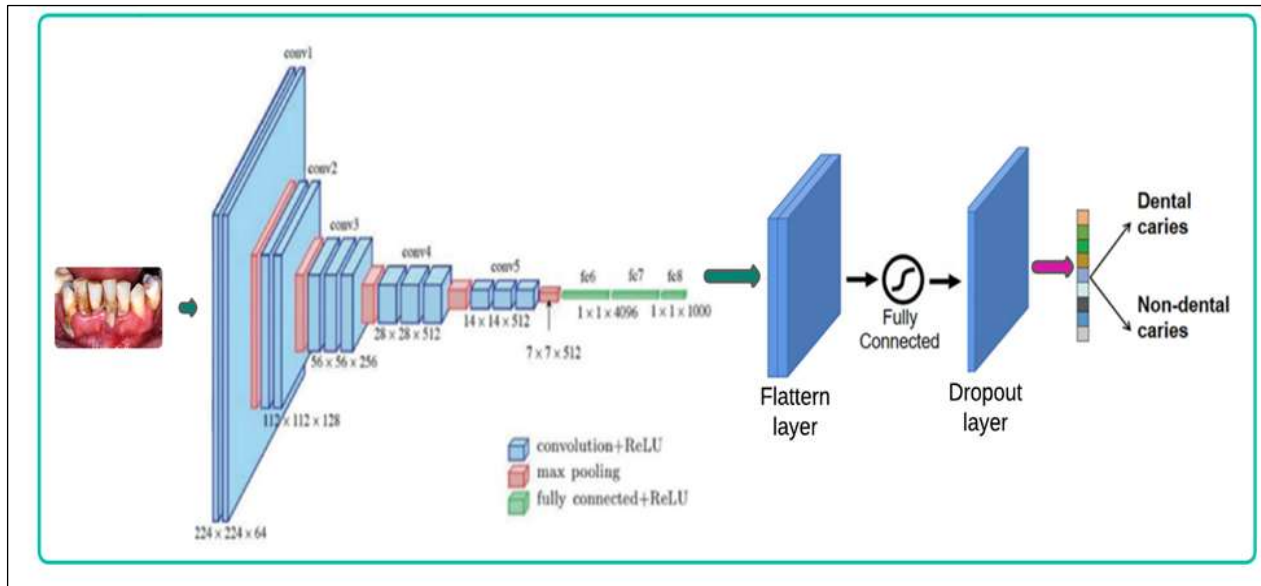


Figure 4: Shows the CNN network's architecture.

Source: Authors, (2026).

IV. RESULTS AND DISCUSSIONS

This section shows the experimental findings obtained from proposed methodology. The training subset of the dataset was utilized to extract optimal feature representations for building the dental disease classification model. Transfer learning was employed as the core recognition strategy, with the VGG16 network initially adapted through fine-tuning to reshape its weights and parameters, enabling it to capture the most relevant features of the dataset. Model performance was evaluated on the testing subset using accuracy, precision, and recall metrics, ensuring a comprehensive assessment and comparative analysis of different classifiers. Using the modified VGG16 architecture, an overall accuracy of approximately 92% was achieved, demonstrating the effectiveness of leveraging pre-trained models with large-scale prior knowledge when fine-tuned for this specific task of classifying six types of dental caries.

Table 1: Comparative results.

S.No	Architecture	Accuracy (%)	Precision (%)	Recall (%)
1	AlexNet	89.19	88.99	89.41
2	Resnet50	92.74	92.02	92.83
3	GoogleNet	91.25	91.14	90.79
4	NasNet-Mobile	85.86	86.32	85.95
5	DenseNet201	92.76	92.58	92.39
6	InceptionV3	90.41	90.97	90.39
7	MobileNetV2	87.82	88.01	88.91
8	InceptionResNetV2	90.68	90.49	90.18
9	VGG16	92.11	92.42	92.99
10	VGG19	95.85	95.47	96.18

Source: Authors, (2026).

All experiments were conducted on a Lenovo Core i7 laptop with 32 GB RAM, which emphasizes the computational efficiency of the proposed approach. The re-trained models used in this study varied in architectural design, connectivity patterns, layer depth, and parameter count, resulting in differences in performance and efficiency. As expected, models responded differently to the classification task depending on their inherent structural properties and adaptability to the unseen dataset. Comparative results in Table 1 indicate that VGG19 outperformed other models, achieving an accuracy of approximately 95.85%. The improved performance of the fine-tuned VGG19 can be assigned to its ability to learn highly discriminative feature representations for the dental dataset. VGG19 Figure 5 demonstrated a distinct trend by reaching its optimal epoch near the end of the maximum training period, at epoch 45. The model achieved an impressive accuracy of 95.85% with a notably low training loss of 0.2165. The validation loss was 0.3892, while the validation accuracy stood at 93.13%. These results indicate that VGG19 continued to learn and generalize effectively even after prolonged training, showcasing its capability to extract and leverage increasingly complex data features. The corresponding confusion matrix for the modified VGG19 is presented in Figure 6, illustrating its training and validation loss over time classification behaviour across all disease categories.



Figure 5: Demonstrated the accuracy of model performance.

Source: Authors, (2026).

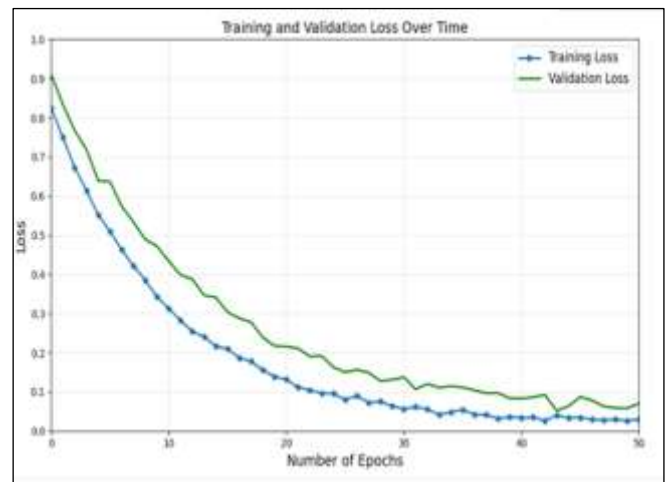


Figure 6: Demonstrated the training and validation loss over time of model.

Source: Authors, (2026).

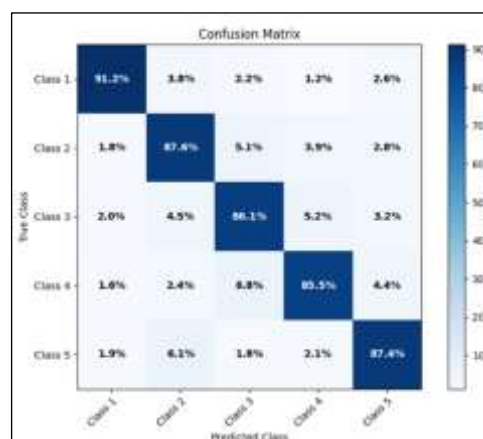


Figure 7: shows the VGG19 re-trained model's confusion matrix.

Source: Authors, (2026).

V. CONCLUSIONS

This research work introduces a high-resolution photographic images novel dataset, annotated by expert dentists, for the automated diagnosis of six distinct oral diseases. The workflow begins with the training of a binary classifier to identify the necessity of ROI detection in dental images, followed by the use of a customized YOLOv4-tiny network for precise ROI localization. Detected ROIs are cropped and utilized to construct the classification dataset. Transfer learning was then applied to multiple pre-trained CNN architectures, enabling the exploitation of prior knowledge while adapting the models to the new task. Among the tested architectures, the fine-tuned VGG19 achieved the highest performance, with an accuracy of 95.85% in classifying the six oral disease categories. These results highlight the effectiveness of integrating lightweight detection models with transfer learning-based classification networks for accurate, efficient, and scalable dental disease diagnosis.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Ashish Gupta, Deepak Gupta.

Methodology: Deepak Gupta, Shweta Nishit Jain.

Investigation: Ashish Gupta, Deepak Gupta.

Discussion of results: Pradeep Yadav.

Writing – Original Draft: Prashant Sharma.

Writing – Review and Editing: Shweta Nishit Jain, Bhaskar Roy.

Resources: Prashant Sharma.

Supervision: Ashish Gupta, Deepak Gupta.

Approval of the final text: Prashant Sharma.

VII. REFERENCES

- [1] Ashish Gupta, Deepak Gupta, Deepali Virmani, Rolly Gupta, Ashwin. M, and John Philip Bhimavarapu, "An Improved Diagnostic Classification of Breast Cancer Using Ensemble and Deep Learning Methods," KSII Transaction on Internet and Information System Vol. 19, No. 12, Dec. 2025.
- [2] Mahmoud, A., & Zrigui, M. Deep neural network models for paraphrased text classification in the Arabic language. In Natural Language Processing and Information Systems: 24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019, Salford, UK, June 26–28, 2019, Proceedings 24 (pp. 3-16). Springer International Publishing, 2019.
- [3] Merhbene, L., Zouaghi, A., & Zrigui, M., June Ambiguous Arabic words disambiguation. In 2010 11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (pp. 157-164, 2010).
- [4] W. Li, E. Guo, H. Zhao, Y. Li, L. Miao, and C. Liu, "Evaluation of transfer ensemble learning-based convolutional neural network models for the identification of chronic gingivitis from oral photographs," BMC Oral Health, vol. 24, no. 814, 2024.
- [5] Mansouri, S., Charhad, M., & Zrigui, M. Arabic Text Detection in News Video Based on Line Segment Detector. Res. Computer. Science., 132, 97-106, 2017.
- [6] Farhani, N., Terbeh, N., & Zrigui, M., Object recognition approach based on generalized hough transform and color distribution serving in generating arabic sentences. International Journal of Computer and Information Engineering, 13(6), 339-344, 2019.
- [7] Satish N. Gujar., Ashish Gupta Sanjay Kumar P. Pingat, Rashmi Pandey. "Brain tumour detection and multi classification using GNB-based machine learning architecture ", Int. J. Data Analysis Techniques and Strategies, Vol. 17, No. 1, 2025.
- [8] Bellagha, M. L., & Zrigui, M., Speaker naming in tv programs based on speaker role recognition. In 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA) pp. 1-8, . November 2020,
- [9] Slimi, A., Hamroun, M., Zrigui, M., & Nicolas, H.. "Emotion recognition from speech using spectrograms and shallow neural networks". In Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia (pp. 35-39) November 2020,
- [10] Priya S., Sudhakar, R., Patil K T., Pillai, S., Joshi V., Gupta A. and Gupta D, "Artificial Intelligent Based on Modelling for Prediction of Alzheimer's Disease for Optimal Solution ". In International International Journal of Basic and Applied Sciences 14(4), pp. 313-319, August 2025,
- [11] Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S. C., & Zheng, L.R., "A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal". IEEE journal of biomedical and health informatics, 24(3), 898-906.
- [12] Al Kheraif, A. A., Wahba, A. A., & Fouad, H. "Detection of dental diseases from radiographic 2d dental image using hybrid graph-cut technique and convolutional neural network". Measurement, 146, 333-342, 2019.
- [13] Laishram, A., & Thongam, K., "Detection and classification of dental pathologies using faster-RCNN in orthopantomogram radiography image". In 2020 7th international conference on signal processing and integrated networks (SPIN) , pp. 423-428 Feb. 2020,
- [14] Liang, Y., Fan, H. W., Fang, Z., Miao, L., Li, W., Zhang, X., ... & Chen, X. "A. OralCam: enabling self-examination and awareness of oral health using a smartphone camera. " In Proceedings of the 2020 CHI conference on human factors in computing systems (pp1-13) April. 2020,
- [15] Ding, B., Zhang, Z., Liang, Y., Wang, W., Hao, S., Meng, Z., & Lv, Y. "Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm" Annals of Translational Medicine, 9(21), (2021)..
- [16] Zhu, H., Cao, Z., Lian, L., Ye, G., Gao, H., & Wu, J., CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image. Neural Computing and Applications, 1-9, 2022.
- [17] Rashid, U., Javid, A., Khan, A. R., Liu, L., Ahmed, A., Khalid, O., & Nawaz, R., "A hybrid mask RCNN-based tool to localize dental cavities from real-time mixed photographic images". PeerJ Computer Science, 8, e888, 2022.

- [18] Thanh, M. T. G., Van Toan, N., Ngoc, V. T. N., Tra, N. T., Giap, C. N., & Nguyen, D. M., “Deep learning application in dental caries detection using intraoral photos taken by smartphones”. *Applied Sciences*, 12(11), 5504. 2022.
- [19] Keser, G., Bayrakdar, İ. Ş., Pekiner, F. N., Çelik, Ö., & Orhan, K. , “A deep learning algorithm for classification of oral lichen planus lesions from photographic images: A retrospective study”. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 124(1),101264, 2023..
- [20] Gomes, R. F. T., Schmith, J., Figueiredo, R. M. D., Freitas, S. A., Machado, G. N., Romanini, J., & Carrard, V. C., “Use of artificial intelligence in the classification of elementary oral lesions from clinical images”. *International Journal of Environmental Research and Public Health*, 20(5), 3894, 2023.
- [21] Haghanifar, A., Majdabadi, M. M., Haghanifar, S., Choi, Y., & Ko, S. B., “PaXNet: Tooth segmentation and dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier. *Multimedia Tools and Applications*, 1-21, 2023.
- [22] G.A.I Devendi, D.M.D.R. Dissanayake, S.N. Liyanage, F.B.A.H. Francis & R.G. Ragle) “Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer”, IEEE doi: 10.1109/ACCESS.2024.3454338, 2024.
- [23] Kim, H. N., Kim, K., & Lee, Y., “ Intra-oral photograph analysis for gingivitis screening in orthodontic patients. *International Journal of Environmental Research and Public Health*”, 20(4), 3705., 2023.
- [24] Laishram, A., & Thongam, K., “Detection and classification of dental pathologies using faster-RCNN in orthopantomogram radiography image”. In 2020 7th international conference on signal processing and integrated networks (SPIN) ,pp. 423-428 February 2020,
- [25] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for MobileNetV3,” *arXiv preprint arXiv:1905.02244*, 2019.,
- [26] A. Bochkovski, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” *arXiv preprint arXiv:2004.10934*, 2020..
- [27] A. Gupta, D. Gupta, M. Pathak, Sharmila K. Wagh “The advancement of ensemble deep learning architecture for the detection and classification of brain tumours with MRI images,” *Int. J. Biomedical Engineering and Technology*, Vol. 45, No. 1, 2024.