Dental Anomaly Recognition on Radiographs using LLMs

Dragos A. Gavrus[1], Ioana G. Ciuciu[2]

[1]Babes-Bolyai University, Cluj-Napoca, Romania

Abstract

Manual analysis of dental radiographs is time-consuming and prone to error, requiring significant expertise. This project proposes an AI-based system that leverages deep learning and large language models (LLMs) to automate the detection and classification of dental anomalies. The system aids clinicians by identifying anatomical structures and pathological findings such as caries, implants, bone loss, apical lesions, and restorative treatments. It improves diagnostic consistency and speeds up clinical decision-making. Furthermore, a Retrieval-Augmented Generation (RAG) chatbot powered by a large language model enables personalized, accessible explanations to assist patients in understanding their diagnoses.

Keywords

Dental anomalies, radiograph analysis, retrieval-augumented generation, machine learning

1. Introduction and Motivation

Dental diagnostics rely heavily on radiographic imaging to identify anomalies such as caries, implants, bone loss, apical lesions, or root canal failures. However, manual interpretation of dental radiographs is time-consuming, error-prone, and requires substantial clinical expertise. These challenges often lead to inconsistent diagnoses and delayed treatment decisions, particularly in high-throughput or resource-limited clinical environments.

To mitigate these issues, we introduce **Orthovision**, an AI-powered web platform for automated dental anomaly recognition. The system integrates computer vision techniques, specifically convolutional neural networks (CNNs) [1], for X-ray image segmentation and object detection, with large language models (LLMs) for generating human-readable diagnostic explanations. By combining visual deep learning with retrieval-augmented natural language processing, Orthovision aims to enhance diagnostic precision, reduce clinical burden, and improve communication with patients.

While AI tools have been increasingly adopted in medical imaging, existing dental AI solutions [2, 3] often lack personalization, integration of patient history, and interactive capabilities. Most systems focus solely on image analysis, without offering interpretability or follow-up explanations tailored to the patient's context.

Orthovision is motivated by the need to move beyond static diagnostic systems [4] toward intelligent platforms that assist both practitioners and patients. By embedding a Retrieval-Augmented Generation (RAG)-based chatbot, our system enables personalized communication based on past radiographs and documented anomalies. This enhances understanding for patients and provides decision support for clinicians, particularly in practices where time or expertise may be limited.

2. Problem Statement

Although recent advances in machine learning have significantly improved medical image analysis, existing frameworks for dental radiograph interpretation still face several key limitations. Many systems are restricted to single-task models that detect only a narrow set of conditions [3], lacking the

RuleML+RR'25: Companion Proceedings of the 9th International Joint Conference on Rules and Reasoning, September 22–24, 2025, Istanbul, Turkiye

dragosgavrus17@gmail.com (D. A. Gavrus); ioana.ciuciu@ubbcluj.ro (I. G. Ciuciu)
https://cs.ubbcluj.ro/~oana (I. G. Ciuciu)
0000-0002-7126-0585 (I. G. Ciuciu)

flexibility to perform robust multi-class anomaly detection across variable image qualities. Additionally, few solutions offer precise tooth-level localization [5, 6], which is essential for generating structured diagnostic reports that link anomalies to specific anatomical regions.

Equally important, most tools lack meaningful patient-specific explanations, ignoring historical data and offering little accessible, tailored feedback. They also typically lack end-to-end deployment, focusing on isolated inference models without the infrastructure for secure user management, image processing, reporting, and patient interaction in a unified system.

This work addresses these limitations by proposing a full-stack system that combines YOLOv11-based [7] object detection for both anomaly and tooth localization with a DeepSeek-V3-powered Retrieval-Augmented Generation (RAG) [8] chatbot for patient education. The integration of deep learning and large language models enables both diagnostic precision and interpretability, offering a practical and explainable solution for modern dental care.

3. Proposed Approach

The proposed system, **Orthovision**, is a application designed to automate the detection and explanation of dental anomalies from X-ray images. It integrates computer vision models for image analysis and large language models for generating context-aware, human-readable diagnostics.

The app uses a React/TypeScript frontend and a Flask backend to support X-ray uploads, anomaly detection, chatbot interaction, and data storage with Supabase and Qdrant.

3.1. Image Analysis Pipeline

The image analysis pipeline in Orthovision is composed of two dedicated YOLOv11 models. The first model is trained to detect and localize each of the 32 human teeth within panoramic dental radiographs. This enables tooth-specific indexing and supports the generation of structured diagnostic reports. The second model focuses on detecting a wide range of dental anomalies, including caries, implants, apical lesions, bone loss, and other conditions. It operates by drawing class-specific bounding boxes around identified regions of interest.

To associate anomalies with individual teeth, bounding boxes from the anomaly and tooth detection models are aligned using the IoU metric. K-Means clustering organizes teeth into upper and lower arches for clearer visualization. Although a U-Net segmentation model was explored, the final pipeline uses only YOLOv11-based detectors for better speed and accuracy.

3.2. Report Generation

The backend synthesizes anomaly and tooth detection results into a detailed diagnostic report. Each identified anomaly is linked to the affected tooth number, enabling dentists and patients to trace conditions visually and textually. A k-means clustering algorithm is used to assign teeth to the correct dental arch. An example of a generated report is shown in Figure 1.

3.3. Chatbot Integration

A Retrieval-Augmented Generation (RAG) chatbot, powered by the DeepSeek-V3 model hosted on Hugging Face, is used to provide explanations based on prior radiographs and dental literature. The chatbot performs a k-nearest neighbors (k-NN) search in Qdrant over embedded clinical documents related to 14 dental anomaly types [9–20]. Retrieved contexts are combined with patient-specific reports to generate personalized responses to user queries.

For document embedding, the system uses the `all-MiniLM-L6-v2` sentence transformer, which provides a good trade-off between speed and semantic accuracy for dense vector retrieval in the Qdrant database.

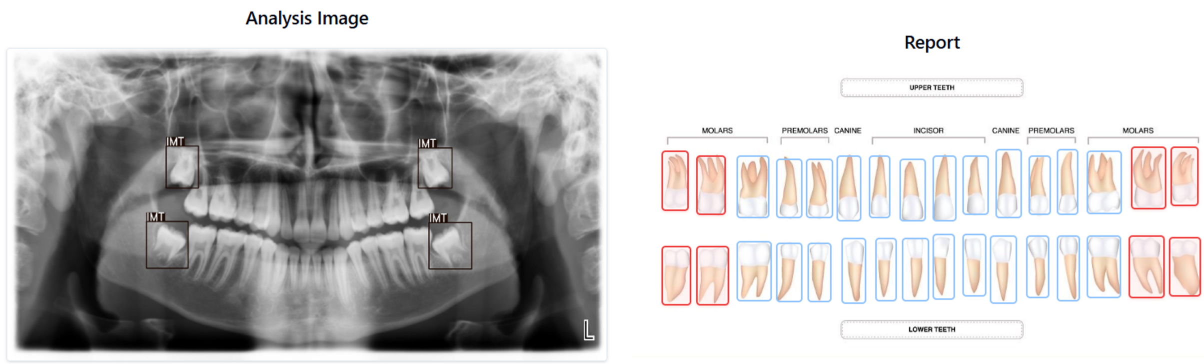

Figure 1: Generated Dental Report.

4. Results and Discussion

This section presents the evaluation of the core components deployed in the system: the YOLOv11-based models for anomaly and tooth detection, and the Retrieval-Augmented Generation (RAG) chatbot built on DeepSeek-V3. Results are discussed in terms of detection accuracy, clinical relevance, and usability in practice.

4.1. Dental Anomaly Detection with YOLOv11

The YOLOv11 anomaly detection model was trained using a proprietary dataset of 1,000 annotated panoramic dental radiographs covering 14 clinically relevant conditions. The model achieved a mean average precision (mAP) of 0.369 at an IoU threshold of 0.5, and 0.167 over the mAP@0.5:0.95 range. It performed particularly well on classes such as surgical root debridement, endodontic treatment, and dental implants, where visual characteristics were distinct and annotations consistent. These results suggest the model is robust for commonly encountered pathologies in clinical practice.

Performance was notably lower for underrepresented and hard-to-distinguish conditions like apical scars and bone loss, highlighting the need for balanced datasets and improved feature engineering or multi-modal integration. Still, the model's real-time detection of multiple anomalies with reasonable accuracy supports automating routine diagnostics, as illustrated in Figure 2.

4.2. Tooth Identification with YOLOv11

To enable anomaly-to-tooth linking, a second YOLOv11 model was trained to detect and index all 32 teeth individually. The model achieved outstanding performance, with an mAP@0.5 of 0.975 and an mAP@0.5:0.95 of 0.701. Both precision and recall exceeded 93%, and predictions generalized well across test images with varying orientations and densities.

This precise tooth-level localization allowed the system to associate each anomaly with a specific tooth and generate indexed, structured diagnostic reports. The resulting overlays and summaries (see Figure 3) provide clinicians with a clear overview of affected regions, while also making it easier to explain findings to patients. The integration of this module significantly enhances the interpretability and traceability of the system's outputs.

4.3. RAG Chatbot Performance

The chatbot is implemented using a Retrieval-Augmented Generation (RAG) architecture that combines dense semantic retrieval with natural language generation. A curated set of 12 clinical documents—covering implants [9–20], caries, bone loss, and other anomaly types—was embedded using a transformer-based sentence encoder and stored in a Qdrant vector database. When a user submits a

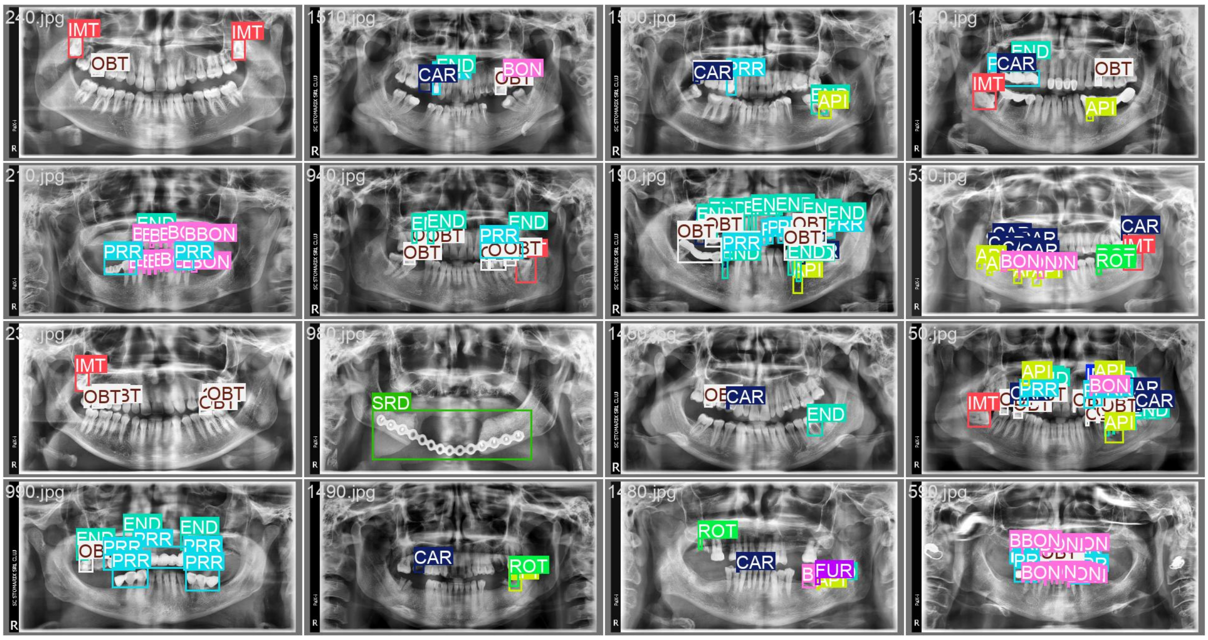

Figure 2: YOLOv11 detection visualizations on validation batch.

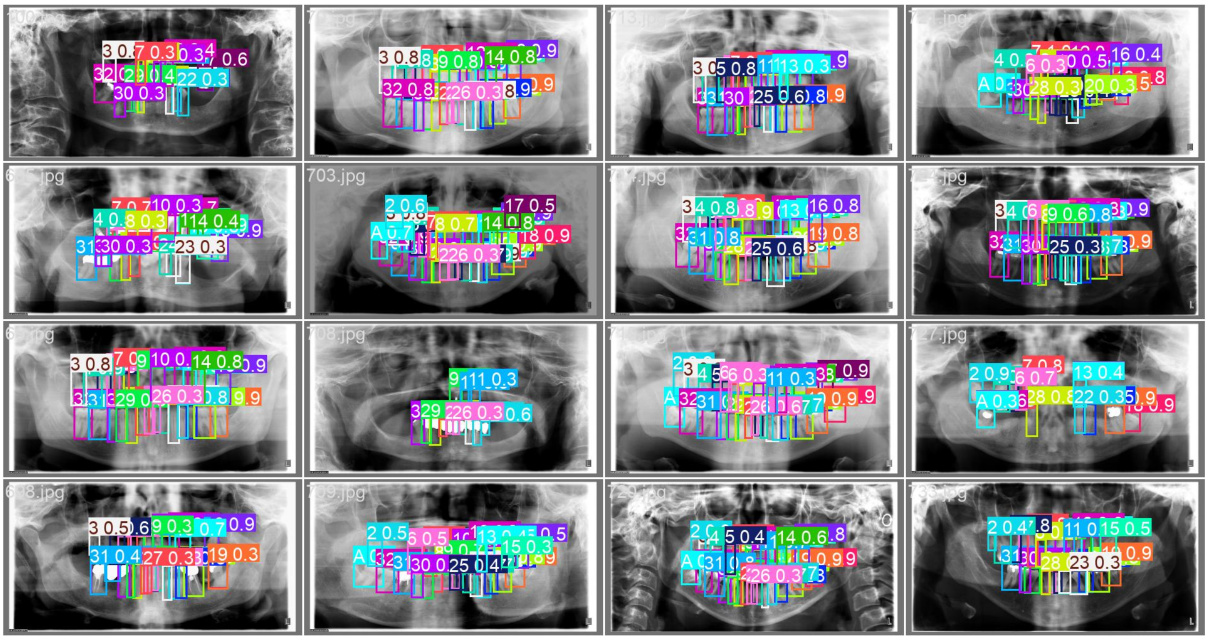

Figure 3: Predicted results showing individual tooth detections.

query, the system retrieves the top relevant document segments and integrates them into a prompt for DeepSeek-V3, a high-performance open-source LLM hosted via Hugging Face.

The chatbot was evaluated on a set of 20 queries simulating real-world patient questions about diagnoses and anomalies. The system produced context-aware, clinically relevant responses in 85% of cases. Retrieved context chunks were topically appropriate in 89% of interactions, indicating high-quality semantic matching. The LLM was especially effective in generating coherent, empathetic explanations for common anomalies such as implants, root fillings, and caries. These responses often incorporated historical context from the user's past reports, enhancing personalization and continuity of care.

Despite strong performance, some limitations were observed. The chatbot occasionally produced verbose responses or exhibited cautious overqualification in rare cases. Additionally, the system currently depends on remote inference APIs, which introduces latency and limits offline availability. Future improvements may include integrating on-device LLMs or caching frequent queries to improve responsiveness.

4.4. Discussion

Taken together, the results indicate that the system is capable of performing high-precision dental anomaly detection, reliable tooth localization, and interactive diagnostic explanation via natural language. The combination of YOLOv11 for vision tasks and DeepSeek-V3 for language generation supports both clinical decision-making and patient engagement. While rare anomalies remain challenging, and scalability could be further optimized, the current platform demonstrates strong real-world potential for augmenting radiograph-based dental workflows. Continued expansion of the training datasets, anomaly categories, and multilingual support are promising avenues for future development.

Integrating logic-based techniques would enhance the system's reliability and precision by using explicit rules and ontological reasoning to represent the structured nature of the dental domain, rather than relying on the current heuristic methods like Intersection over Union (IoU) and k-means clustering. This improvement would lead to more accurate structured reports and allow the chatbot to provide more concise, precise, and contextually relevant explanations, addressing issues of verbosity and over-qualification.

5. Conclusion

This paper presented Orthovision, a full-stack AI-driven web platform for automated dental anomaly recognition and patient-oriented explanation. By integrating YOLOv11 models for tooth and anomaly detection with a Retrieval-Augmented Generation chatbot powered by DeepSeek-V3, the system enables both accurate diagnostics and personalized natural language interactions. Experimental results confirmed the effectiveness of the object detection models across most anomaly and tooth classes, while the chatbot demonstrated strong performance in delivering clinically grounded, user-friendly responses.

The proposed approach bridges the gap between AI-assisted imaging and patient communication, offering a scalable and explainable solution for modern dental care. Future work includes expanding the training data, improving detection of rare anomalies, and enhancing chatbot responsiveness through local inference, multilingual support and ontology-based reasoning.

Acknowledgments

We gratefully acknowledge the support and resources that contributed to the development of this thesis. We thank **Prof. Mihaela Hedeșiu** for providing access to proprietary dental radiograph data through a research collaboration with **UMF University**, which played a critical role in training and evaluating the models.

Special thanks to **Lect. Dr. Laura Diosan** for her guidance and valuable input on the machine learning components of this work.

Declaration on Generative AI During the preparation of this work the authors used Chatgpt in order to improve the clarity and quality of expression. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the paper.

References

[1] M. Bhatt, et al., Review of deep learning: concepts, cnn architectures, challenges, applications, future directions, Journal of Big Data (2021).

[2] Diagnocat, Diagnocat, https://www.diagnocat.com/, 2020. Last accessed: 05.09.2025.

[3] Denti.AI, Denti.ai detect, https://www.denti.ai/, 2024. Last accessed: 05.09.2025.

[4] Overjet, Overjet, https://www.overjet.com/, 2018. Last accessed: 05.09.2025.

[5] CephX, Cephx, https://www.cephx.com/, 2024. Last accessed: 05.09.2025.

[6] CranioCatch, Craniocatch, https://www.craniocatch.com/, 2024. Last accessed: 05.09.2025.

[7] N. Rao, Yolov11 explained: Next-level object detection with enhanced speed and accuracy, https://medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71, 2024. Accessed: 2025.

[8] P. Lewis, E. Perez, A. Piktus, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, in: EMNLP, 2020.

[9] N. Veiga, F. Figueiredo, C. Pina, Dental caries: A review, Journal of Dental and Oral Health (2016). URL: https://ciencia.ucp.pt/files/37440728/dental_caries_a_review.pdf.

[10] A. Heboyan, V. L. Avetisyan, K. M. Vardanyan, et al., Tooth root resorption: A review, Science Progress 105 (2022) 1–29. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10358711/.

[11] S. W. Peeran, A. S. Alghamdi, et al., Furcation involvement in periodontal disease: A narrative review, Cureus 16 (2024) e55924. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11004587/. doi:10.7759/cureus.55924.

[12] C. Ho, C. Argáez, Endodontic Therapy Interventions for Root Canal Failure: A Review of Clinical Effectiveness and Guidelines, CADTH Rapid Response Reports, Canadian Agency for Drugs and Technologies in Health (CADTH), 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK470661/.

[13] H. Blake, F. Barros, B. Fong, E. Dhamija, Apical Periodontitis, StatPearls Publishing, 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK589656/.

[14] C.-W. Lee, et al., Clinicopathological study of periapical scars, Journal of Dental Sciences 16 (2021) 1140–1145. doi:10.1016/j.jds.2021.05.008.

[15] V. Malagnino, et al., The fate of overfilling in root canal treatments with long-term follow-up: A case series, Restorative Dentistry & Endodontics 46 (2021) e27. URL: https://pubmed.ncbi.nlm.nih.gov/33908464/. doi:10.5395/rde.2021.46.e27.

[16] P. Santosh, Impacted mandibular third molars: Review of literature and a proposed classification, Annals of Medical and Health Sciences Research 5 (2015) 229–234. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512113/.

[17] D. Gupta, et al., Dental Implants, StatPearls Publishing, 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK556031/.

[18] S.-C. Chang, et al., Recent clinical treatment and basic research on the alveolar bone, Biomedicines 11 (2023) 843. URL: https://www.mdpi.com/2227-9059/11/3/843. doi:10.3390/biomedicines11030843.

[19] K. Singh, A. Chatterjee, Oral Surgery, Extraction of Roots, StatPearls Publishing, Treasure Island (FL), 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK589696/.

[20] P. Desai, E. Solomon, Orthodontics, Malocclusion, StatPearls Publishing, Treasure Island (FL), 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK592395/.