Explainable Deep Learning-Based Multi-Scale Caries Detection

Walid Brahmi[a,d], Imen Jdey[b,d] and Fadoua Drira[c,d]

[a]National School of Electronics and Telecommunications of Sfax (ENET'Com) - University of Sfax, Tunisia
[b]Faculty of Economics and Management of Sfax - University of Sfax, Tunisia
[c]National Engineering School of Sfax (ENIS) - University of Sfax, Tunisia
[d]Research Groups in Intelligent Machines (REGIM Lab), ENIS, University of Sfax, Tunisia

ARTICLE INFO

Keywords:
Dentistry
Caries Detection
Deep Learning
YOLO
Explainable Artificial intelligence
GradCam
Lime

ABSTRACT

Early and accurate detection of dental caries from panoramic X-rays is fundamental to preventive dentistry, yet manual analysis is time-consuming and prone to subjectivity. This study presents a novel multi-level framework for automated caries detection and segmentation, designed to evaluate model robustness across varying levels of diagnostic complexity. We utilize three datasets of increasing specificity: Full Panoramic Radiographs (FPR), Cropped Panoramic Regions (CPR), and Single Tooth patches (ST). A rigorous comparative analysis is conducted between the established You Only Look Once (YOLOv8) model and the (YOLOv11) architecture, assessing their efficacy in this domain. To bridge the gap between model output and clinical trust, Explainable Artificial Intelligence (XAI) techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-agnostic Explanations (LIME) are integrated to provide transparent visual and analytical explanations. Evaluation extends beyond standard metrics (mean Average Precision (mAP), precision, recall) to encompass computational efficiency and environmental impact, quantified via CO_2 emissions. Results demonstrate that YOLOv8 consistently and significantly outperforms YOLOv11 across all dataset levels, achieving peak mean Average Precision (mAP@50) scores of 47.0% (FPR), 79.7% (CPR), and 90.0% (ST). YOLOv8 also exhibited superior training efficiency (2.06 h) and a lower environmental cost (0.041 kg Carbon Dioxide (CO_2) per 100 epochs). The integrated XAI successfully illuminated model reasoning, validating its focus on clinically critical diagnostic regions. In conclusion, YOLOv8 is the preferred architecture for accurate, efficient, and environmentally conscious caries detection. Its integration with XAI provides a transparent and reliable tool, poised for effective incorporation into clinical dental workflows.

1. Introduction

Good oral health is essential for daily activities such as eating, breathing, and speaking. It encompasses the well-being of the mouth, teeth, and orofacial structures. Beyond physical health, oral health affects mental and emotional well-being, influencing self-esteem, overall wellness, and social interactions. According to the World Health Organization (WHO), oral diseases are among the most prevalent noncommunicable diseases globally, affecting approximately 3.5 billion people. Dental caries, the most common of these diseases [1], impacts individuals of all ages, particularly children, teenagers, and the elderly. It poses a significant challenge to health, quality of life, and healthcare systems.

Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in various medical imaging applications. They excel in oncology for tumor detection [2], dermatology for skin lesion classification [3], and parasitology for diagnosing malaria [4]. CNNs have also contributed significantly to the timely detection of Alzheimer's disease through brain Magnetic Resonance Imaging (MRI) analysis [5]. In dentistry, clinicians increasingly rely on CNNs to analyze intraoral and extraoral radiographs (Figure 1). These networks can accurately classify dental structures and segment teeth in panoramic images [6]. They also help detect dental caries, reveal anatomical variations, and support precise diagnosis of oral diseases [7], ultimately improving treatment planning and enhancing patient outcomes.

*Corresponding author: Walid Brahmi
bensghaierwaleed@gmail.com (W. Brahmi); imen.jdey@fsegs.usf.tn (I. Jdey); fadoua.drira@enis.tn (F. Drira)
ORCID(s): 0009-0009-2467-322X (W. Brahmi); 0000-0001-7937-941X (I. Jdey); 0000-0001-6706-4218 (F. Drira)
https://www.linkedin.com/profile/view?id=walidbensghaier (W. Brahmi), https://www.linkedin.com/profile/view?id=imen-jdey-2886a324 (I. Jdey), https://www.linkedin.com/profile/view?id=fadoua-drira-463b7034 (F. Drira)

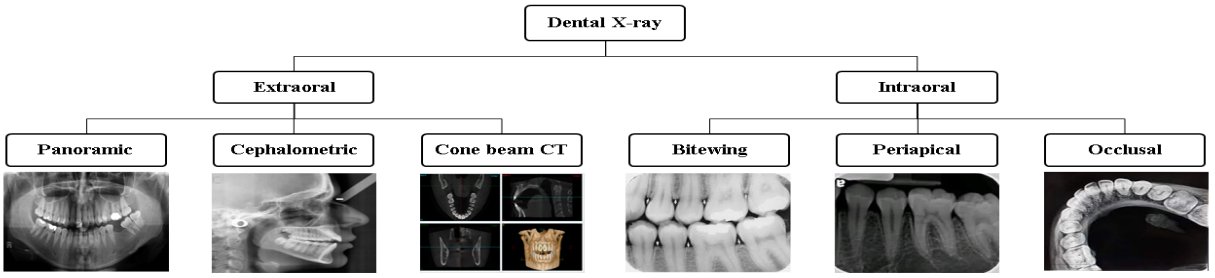


Figure 1: Taxonomy of Dental X-ray Modalities.

Explainability in healthcare Artificial Intelligence (AI) plays a crucial role in managing uncertain data and ensuring transparent collaboration with dental professionals. Explainable Artificial Intelligence (XAI) enables models to provide interpretable and trustworthy predictions, fostering clinical confidence.This emphasis on transparent AI aligns with recent advances in medical diagnostics, such as the application of SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) for explainable fatty liver disease diagnosis [8]. In this study, we compare the performance of the You Only Look Once YOLOv8 and YOLOv11 architectures for both detection and segmentation of dental caries across three different imaging modalities. To enhance model transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) and LIME are employed. These XAI techniques offer visual and analytical insights into model decision-making processes, helping to identify key diagnostic regions and potential sources of error. The proposed approach emphasizes high diagnostic accuracy while maintaining computational efficiency, making it adaptable for clinical environments with limited resources. Overall,

The remainder of this paper is structured as follows. Section 2 provides an overview of the foundational concepts, including the YOLO family of architectures and explainable AI principles. Section 3 reviews recent advancements in dental image segmentation and caries detection. Section 4 describes the dataset, experimental setup, and methodology. Section 5 presents and discusses the obtained results. Finally, Section 6 concludes the study and highlights future research directions.

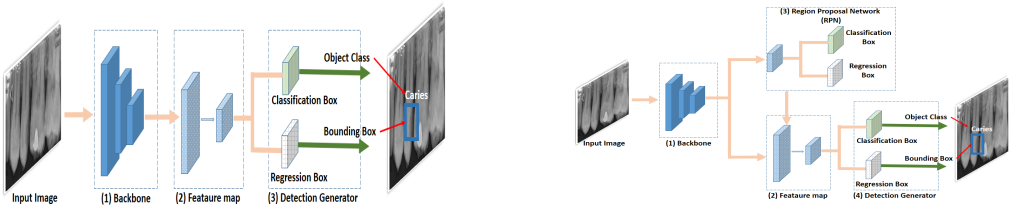
2. Fundamental Concepts and Background

To gain a comprehensive understanding of the applications and challenges of deep learning models in dentistry, it is recommended to refer to the following literature reviews: the surveys conducted by Schwendicke et al. [9], Hwang et al. [10], Kang et al. [11], and Prados-Privado et al. [12], as well as the systematic literature review conducted by Brahmi et al. [13]. Brahmi et al.'s review specifically highlights the issues related to the scarcity of public datasets and proposes the use of public databases to improve the generalization of models. These studies provide a detailed analysis of recent advancements and the challenges that need to be addressed in this rapidly evolving field.

The objective of our study is the segmentation and detection of dental caries. Therefore, it is necessary to review the currently available techniques for detection and segmentation. Object detection involves the identification and localization of specific objects in images or videos using Computer Vision (CV) and image processing techniques. It aims to determine the number of objects, track their positions, and classify them accurately. This process includes bounding-box annotation to outline the objects. The YOLO family of architectures has become a benchmark in this domain, offering real-time performance with high accuracy for diverse detection tasks, such as automatic anomaly tracking in surveillance systems [14] In medical imaging, object detection can effectively identify bone fractures, abnormal cellular activities, and other medical conditions.

There are two main types of object detection methods. Firstly, we have two-stage detectors, as illustrated in Figure 2.

- Region-based Convolutional Neural Network (R-CNN), proposed by Ross Girshick [15], is a region-based CNN for object detection. It is divided into three modules:
    1. Region proposals are generated using a selective search algorithm.
    2. Each proposal goes through five convolutional layers and two dense layers.



(a) Single Stage Object Detector Architecture

(b) Two Stage Object Detector Architecture

Figure 2: Single vs. two stage object detector architectures (adapted from [24]).

3. The third module uses linear classifiers pre-trained for each class.

- Faster R-CNN [16] and Fast R-CNN [17] fall into this category. Faster R-CNN integrates the region proposal network for quicker inference, while Fast R-CNN uses the entire image for feature extraction.

These classifiers generate class-specific scores for each object proposal. Non-max suppression is then applied to find the best fit.

The second type is one-stage detectors, as presented in Figure 2. Examples of one-stage detectors include YOLO, Single Shot Multibox Detector (SSD), EfficientDet, RetinaNet, M2Det, and RefineDet++.

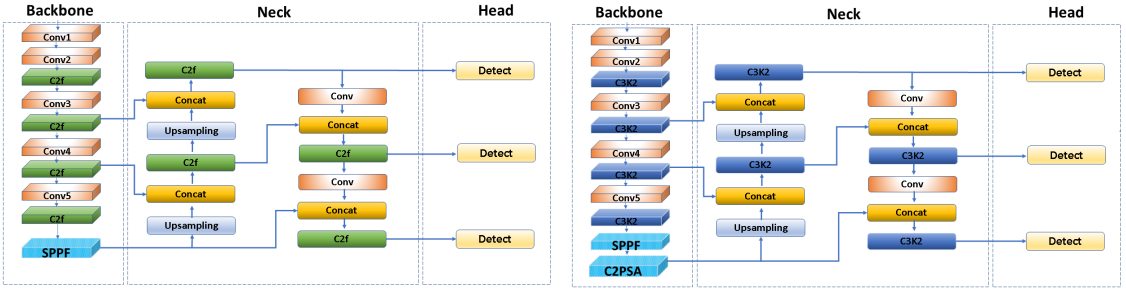
- YOLO [18] divides the image into a grid for predictions, while SSD [19] generates multiple bounding box predictions in a single pass.
- EfficientDet [20] focuses on accuracy and efficiency, while RetinaNet [21] tackles class imbalance problems. Each method brings its unique approach to object detection.
- M2Det [22] introduces an anchor-free approach with a multi-level feature pyramid network, prioritizing efficiency in object detection. On the other hand, RefineDet++ [23] enhances accuracy through refined anchor mechanisms and multi-scale feature fusion within an anchor-based framework.

2.1. Comparison Between YOLOv8 and YOLOv11 Architectures

Before exploring YOLOv8 and YOLO11, it is crucial to thoroughly review the topic. A comprehensive analysis by Terven et al. [25] offers an in-depth overview of YOLO architectures, from YOLOv1 to YOLOv8, highlighting the advancements and innovations introduced in each version. YOLO, introduced by Joseph Redmon and colleagues at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2016 [18], has had a significant impact on object detection. Its real-time, end-to-end approach eliminates the need for a Region Proposal Network (RPN) and significantly enhances speed. Unlike previous methods that required multiple network passes, YOLO can perform detection in a single pass.

Traditional algorithms like R-CNN used sliding windows, classifiers, or a two-step process involving region proposals and classification. However, the authors of YOLO have redefined object detection as a regression problem instead of a classification problem. YOLO employs a convolutional neural network to predict bounding boxes and class probabilities simultaneously in a single pass over the image, hence its name "You Only Look Once." The image is divided into $\mathbf{S}\times\mathbf{S}$ grid cells, with each cell predicting $\boldsymbol{B}$ bounding boxes, including their coordinates, the probability of an object being present, and conditional class probabilities. The bounding boxes, which can capture objects of different sizes and shapes, are referred to as anchor boxes.

Adjacent grid cells may predict overlapping bounding boxes for the same object, resulting in multiple predictions. To address this, a score threshold is applied to discard boxes with scores below a predefined value. Subsequently, the Non-Max Suppression (NMS) technique is utilized to further refine the predictions. This technique involves selecting the bounding box with the highest score and eliminating any overlapping boxes that have an Intersection over Union (IoU) greater than a predefined threshold. This process is repeated until only the most relevant bounding boxes remain, effectively reducing redundancy and improving detection accuracy.



(a) YOLOv8 architecture

(b) YOLOv11 architecture

Figure 3: Architectural schematics of the YOLOv8 (a) and YOLOv11 (b) object detection models.

Table 1
Comparative analysis of YOLOv8 and YOLOv11 architectures

Component	YOLOv8	YOLOv11
Backbone	CSPDarknet (C2f + SPPF)	Enhanced CSPDarknet (C3k2 + C2PSA)
Neck	PANet with C2f blocks	PANet with C3k2 blocks & C2PSA
Head	Standard detection head	Refined head with post-C3k2 CBS layers
Attention	None	C2PSA in backbone and neck

Table 2
YOLOv8 vs YOLO11: Layer count, parameter size, and computational complexity (Giga Floating Point Operations per Second (GFLOPs).

Characteristic	YOLOv8n	YOLO11
Layers	129	181
Parameters	3.16M	2.62M
GFLOPs	8.9	6.6

As illustrated in Figure 3, the YOLOv8 architecture consists of three primary components: the backbone, neck, and head. The **backbone** extracts hierarchical visual features using convolutional layers, C2f blocks, and a Spatial Pyramid Pooling Fast (SPPF) module that captures multi-scale contextual information. The **neck** adopts a Path Aggregation Network (PANet) structure [26] to merge features from different scales, ensuring effective information flow and better representation of small objects. The **head** performs the final prediction by generating bounding boxes, class scores, and segmentation masks, supporting both detection and instance segmentation tasks.

The YOLO11 architecture is an enhancement of the YOLOv8 architecture. As shown in Figure 3, the YOLOv11 architecture preserves the same three-part design. The **backbone** replaces the C2f blocks with lighter C3k2 blocks to improve gradient propagation and maintain efficient feature extraction, retaining the SPPF module for multi-scale representation. The **neck** combines C3k2 and a C2PSA attention module to emphasize important spatial regions and improve detection of small or partially occluded objects. The **head** integrates Convolution-BatchNorm-SiLU (CBS) layers to refine the output features and stabilize training, producing more accurate detection and segmentation results.

To highlight the improvements in backbone, neck, head, and attention mechanisms, a comparative analysis of YOLOv8 and YOLOv11 is presented in Tables 1 and 2.

2.2. Explainable artificial intelligence (XAI)

XAI, short for Explainable Artificial Intelligence, refers to AI systems that can be understood by humans in terms of their functionality, capabilities, limitations, and responses in unfamiliar situations [27, 28]. The objective is to transition from opaque "black boxes" to transparent "glass boxes" (also known as "white boxes") [29]. Unlike "black box" systems that hide their internal operations, XAI systems actively provide explanations to enhance human understanding. The term "XAI" gained significance through The Defense Advanced Research Project Agency's (DARPA) XAI program, initiated in May 2017, which emphasized the importance of explanations in human psychology.

In the context of machine learning and artificial intelligence, interpretability and explainability are related but distinct concepts. Interpretability refers to the ability to understand and predict the behavior of a system based on changes in inputs or parameters, essentially observing cause-and-effect relationships [30]. Explainability, on the other hand, involves clearly articulating the internal workings of the system in terms that are understandable to humans, providing insights into how and why decisions are made [31]. In summary, interpretability allows us to discern what is happening, while explainability provides a clear explanation of it. As shown in Figure 4, explainability techniques in AI encompass a diverse range of methods that can be classified into different categories [32–40]:

1. **Interpretation types**: When considering the type of interpretation, the categorization of XAI methods typically includes two main approaches: intrinsic, also called ante-hoc[36], and post hoc. Intrinsic interpretation involves designing models to be inherently understandable by modifying their structure or components to ensure transparency. This approach makes the model's decision-making process clear from the beginning. For example, innovative approaches like [41] use ensemble SHAP values from multiple models as prior knowledge to directly construct and stabilize Transformer models, making their attention mechanisms more transparent and predictable from the outset. Conversely, post hoc interpretation applies techniques to analyze a model after it has been trained, providing explanations about its behavior.
2. **Explanation scopes**: Another categorization in XAI relates to the scope of explanations, determining the granularity and extent of the information provided. This categorization distinguishes between techniques that offer a global understanding of the model, providing global interpretability, where the explanation is derived from the entire model, and those that focus on individual instances, known as local interpretability.
3. **Model Specificity**: According to the articles [34, 35, 37], the third category we adopt is Model Specificity, which differentiates between model-specific XAI methods and model-agnostic methods. Model-specific, as the name suggests, are tailored to work with specific types of models, utilizing their internal mechanisms to provide explanations for their behavior. By definition, intrinsic methods are model-specific.
   In contrast, according to [32], model-agnostic tools can be utilized with any machine learning model once it has been trained. These techniques generally analyze the relationships between input features and outputs without requiring access to the model's internal parameters. By definition, post hoc methods are considered model-agnostic.
4. **Explanation Forms**: The final proposed category for XAI methods is explanation forms. This category encompasses the different ways explanations can be presented and interpreted. It includes various formats such as visualizations, textual descriptions, numerical explanations, and rules-based explanations, each tailored to enhance the understanding and usability of the explanations provided by XAI methods. The article [36] provides a detailed explanation of the different formats of explanations.

To summarize this section, it is obvious that various types of XAI are critical for bridging the gap between advanced AI technologies and their practical application in healthcare, such as in dentistry for illness detection. XAI's transparency and interpretability alleviates uncertainty about AI suggestions and provides clinicians with the information they need to make informed judgments. As AI evolves, the incorporation of explainability will be critical for its general adoption and efficacy in healthcare, resulting in improved patient outcomes and a more dependable healthcare system. Finally, explainable AI increases confidence, improves decision-making, promotes continuous learning, and ensures compliance, so making AI more successful in healthcare.

3. Related Works

The timely and accurate diagnosis of dental caries from radiographic images is fundamental to contemporary preventive and restorative dentistry. Acknowledging the inherent subjectivity and substantial resource demands of manual radiographic analysis, the integration of deep-learning methodologies has triggered a paradigm shift toward automated diagnostic capabilities. This section presents a comprehensive synthesis of 12 pivotal studies, delineating key technical advances across three principal domains: object-detection architectures, segmentation methodologies, and specialized diagnostic applications. Through this systematic examination, the current state of the art is established while critical research avenues requiring further investigation are identified, thereby providing the foundational rationale for the proposed investigation.

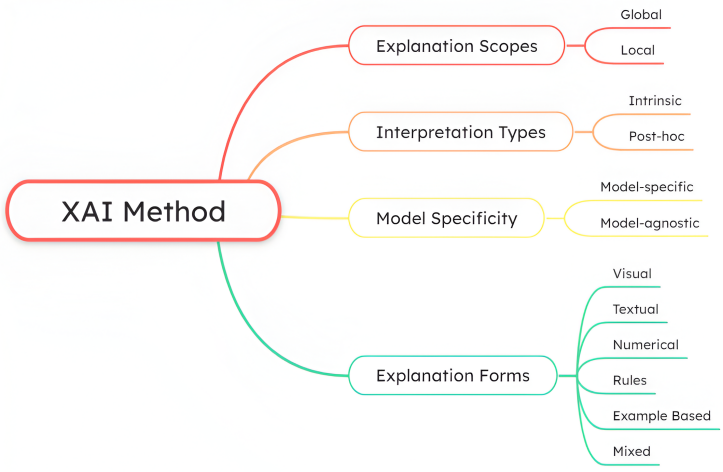


Figure 4: Categorization of eXplainable AI methods.

3.1. Advances in Object-Detection Architectures and Scale-Specific Validation

The computational efficiency and robust performance characteristics of single-stage detectors—particularly YOLO variants—have rendered them indispensable for the automated localization of carious lesions. Research efforts have progressively refined these architectures to enhance both speed and detection accuracy across diverse dental-imaging contexts.

Pérez de Frutos et al. [42] conducted a rigorous comparative evaluation of established object-detection frameworks, including YOLOv5, RetinaNet, and EfficientDet, specifically for identifying proximal caries within bitewing radiographs. Their analysis demonstrated the superior efficacy of the YOLOv5 architecture, achieving a mAP of 0.647 and outperforming human clinicians in detection accuracy. Alsolamy et al. [43] extended beyond mere lesion localization by developing a two-stage CNN framework based on EfficientDet. This comprehensive approach facilitated not only detection but also classification of proximal-caries severity, achieving a detection mAP of 94.3%. Şevik & Mutlu [44] successfully adapted the YOLOv8 framework for a distinct diagnostic application, focusing on detection of pediatric dental anomalies—such as unerupted teeth and agenesis—within the complex anatomical landscape of panoramic radiographs, reporting a mAP of 92.4%. Providing a contemporary technical benchmark, Ramírez-Pedraza et al. [45] performed a comparative analysis of recent YOLO iterations (YOLOv9, YOLOv10, and YOLOv11) for dental-plaque detection and quantification from RGB photographs. Their findings highlighted the exceptional performance of YOLOv11, which achieved a mAP of 94.62% in this specialized domain.

Thanh et al. [46] conducted one of the earliest comparative evaluations of YOLO architectures for non-radiographic dental imaging, assessing YOLOv3 alongside Faster R-CNN, RetinaNet, and SSD for caries detection with intra-oral smartphone photographs. Their analysis revealed YOLOv3 achieved the highest sensitivity (87.4%) for cavitated lesions, demonstrating the architecture's effectiveness even with unconventional imaging modalities. Similarly, Abu Tareq et al. [47] proposed a Hybrid YOLO Ensemble for caries detection from non-standardized intra-oral RGB photographs, achieving a mAP of 79.91% and illustrating YOLO's adaptability to teledentistry under variable image-quality conditions. Collectively, these studies demonstrate YOLO's versatility across imaging modalities while highlighting the need for systematic evaluation across standardized radiographic scales.

3.2. Progress in Caries-Segmentation Methodologies and Pixel-Level Precision

Segmentation models provide pixel-level delineation of carious lesions, offering superior granularity crucial for detailed treatment planning, longitudinal monitoring, and precise clinical intervention.

Dayı and al. [48] introduced the Dental Caries Detection Network (DCDNet), a specialized segmentation architecture featuring a Multi-Predicted Output structure designed to segment and classify carious lesions of the occlusal, proximal and cervical on panoramic radiographs, reporting an overall Dice score of 83.33%. Mărginean et al. [49] proposed CariSeg, an intelligent two-stage system leveraging a network ensemble comprising U-Net, Feature Pyramid Network (FPN), and Deeplabv3 architectures for sequential teeth segmentation followed by carious-lesion delineation,

Table 3
Meta-analysis of deep-learning architectures for dental diagnostics (synthesis of 12 studies).

Study	Primary Focus / Technique	Key Contribution & Best Metric	Main Limitation
Pérez de Frutos et al.	Object detection (bitewing)	YOLOv5 for proximal caries – mAP 0.647 (outperforms clinicians)	Single-scale (bitewing only)
Alsolamy et al.	Detection + severity (bitewing)	Two-stage EfficientDet – mAP 94.3%	No panoramic validation
Şevik & Mutlu	Pediatric anomalies (panoramic)	YOLOv8 – mAP 92.4%	Not caries-oriented
Ramírez-Pedraza et al.	Plaque quantification (RGB)	YOLOv11 – mAP 94.62%	Different task / modality
Thanh et al.	Smart-phone caries (RGB)	YOLOv3 – sensitivity 87.4%	Low performance on non-cavitated lesions
Abu Tareq et al.	Tele-dentistry (RGB)	Hybrid YOLO ensemble – mAP 79.91%	Variable image quality
Dayı et al.	Multi-type segmentation (pan)	DCDNet – Dice 83.33%	Single panoramic scale
Mărginean et al.	Two-stage ensemble (pan)	U-Net/FPN/Deeplabv3 – Dice 68.2%	Computationally heavy
Zhu et al. (CariesNet)	Multi-stage severity (pan)	Axial-attention U-Net – DSC 79.59%	No zoomed-in views
Adnan et al.	Semi-supervised detection	Teacher-student – Dice 0.81	No multi-scale evaluation
Oztekin et al.	Explainable classification	Grad-CAM CNN – accuracy 98.40%	Binary task only
Ramezanzade et al.	Pulp-exposure prediction	ResNet-50 multi-path – AUC 0.81	Highly specialized task

achieving a system-level accuracy of 99.42% and a mean Dice coefficient of 68.2% for lesion segmentation. Zhu et al. [50] developed CariesNet, a U-shaped network incorporating a full-scale axial-attention module specifically engineered for multi-stage caries segmentation—shallow, moderate, deep—from panoramic radiographs, attaining a Dice similarity coefficient (DSC) of 79.59%. Addressing the pervasive challenge of annotated-data scarcity, Adnan et al. [51] implemented a semi-supervised learning paradigm for caries detection and segmentation, yielding a Dice coefficient of 0.81 through a teacher–student framework trained on limited labeled data.

3.3. Specialized Diagnostic Applications and Explainable AI (XAI)

A growing body of research explores specialized clinical-prediction tasks and the integration of explainable artificial intelligence (XAI) for enhanced clinical trust and adoption, reflecting the expanding scope of computational dentistry beyond conventional detection and segmentation.

Oztekin et al. [52] promoted transparency by developing an explainable deep-learning framework for binary classification of caries presence on panoramic radiographs. Their study successfully employed Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual explanations of the model's decision-making process, achieving a prediction accuracy of 98.40%. Ramezanzade et al. [53] addressed a highly specific clinical-prediction task, utilizing a multi-path neural network based on ResNet-50 to predict the risk of pulp exposure before caries excavation, achieving an Area Under the Curve (AUC) of 0.81. These studies collectively underscore the importance of model interpretability for clinical adoption while highlighting the need for more comprehensive XAI integration within complex detection-and-segmentation workflows.

To provide a consolidated synthesis, Table 3 summarizes the core methodological contributions, performance metrics, and specific constraints identified within the current literature. The comprehensive synthesis reveals that, despite significant advances in deep learning for dental diagnostics, three critical gaps persist, preventing reliable clinical integration. First, a pronounced multi-scale-evaluation gap exists, as current models are predominantly validated under isolated scale conditions, leaving their robustness across clinically essential magnification levels—from panoramic overviews to detailed single-tooth views—largely unconfirmed. Second, an explainability-integration gap is evident, with XAI implementations primarily confined to simple classification tasks rather than complex multi-scale detection-and-segmentation workflows. Third, a notable computational-and-environmental-efficiency gap

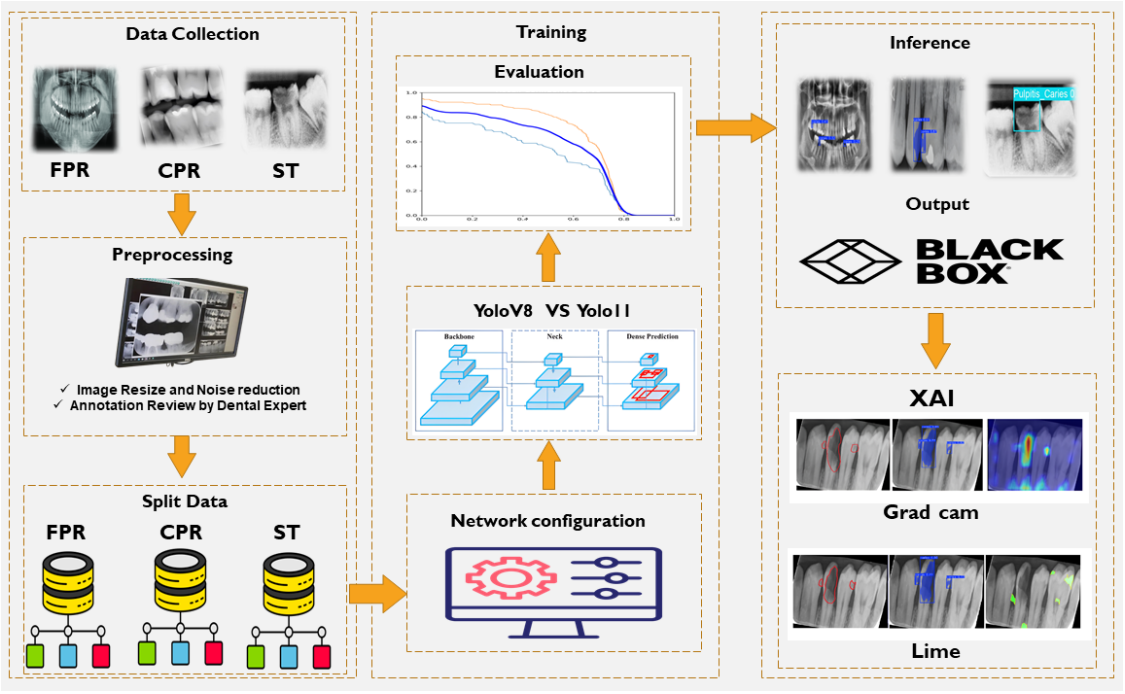


Figure 5: A block representation of the material and method used in the study.

remains unaddressed, as no prior study has incorporated systematic analysis of computational-resource consumption or associated environmental impact, an increasingly vital consideration for sustainable healthcare-AI deployment. To address these limitations, this study proposes a novel framework for explainable multi-scale caries detection that incorporates rigorous computational-efficiency and CO_2 emission analysis, thereby advancing towards clinically viable and environmentally conscious AI solutions for dental diagnostics.

4. Materials and Methods

YOLOv8 and YOLOv11 are state-of-the-art architectures supporting advanced visual tasks, including object detection, image classification, and instance segmentation. The YOLO framework has recently demonstrated superior performance across diverse real-world detection challenges, including automatic anomaly tracking in surveillance systems [14], real-time object detection for autonomous driving [54], and dental diagnostic tasks such as the automated classification of implant types [55]. Building upon this demonstrated efficacy and versatility, this study performs a direct comparative analysis of YOLOv8 and YOLOv11 for the automated detection and segmentation of dental caries in radiographs. To evaluate performance comprehensively, we designed a multi-level framework utilizing three distinct datasets of increasing specificity: FPR, CPR, and ST.approach. Figure 5 illustrates the complete approach, detailed in Algorithm 1.

4.1. Datasets

This study utilized three distinct dental imaging datasets to evaluate YOLOv8 and YOLOv11 models across different clinical contexts. Each dataset was maintained separately throughout all experiments. Figure 6 illustrates representative samples from each imaging modality.

The Full Panoramic (Pan) Dataset [56] contains 1295 panoramic radiographs with 7467 caries annotations for segmentation tasks. The Cropped Panoramic (CP) Dataset [57] includes 2595 cropped images with 4466 annotations for detection tasks. The Single Tooth (ST) X-ray Dataset [58] comprises 1295 tooth-level images with 709 annotations for detection tasks.

All images were standardized to 640×640 pixels and split 80%/10%/10% for training, validation, and testing. A certified dental radiologist verified all annotations to ensure clinical accuracy. This multi-dataset approach enables

Algorithm 1 Comparative Multi-Scale Caries Detection

Require:
$\mathbf{D_{Base}}$: Original annotated dataset
$\mathbf{D_{Sets}} = \{\mathbf{D_{FPR}}, \mathbf{D_{CPR}}, \mathbf{D_{ST}}\}$: Scale-specific datasets
$\mathbf{M_{Arch}} = \{\text{YOLOv8, YOLOv11}\}$: Model architectures
$\mathbf{H}$: Hyperparameter configuration
Ensure:
$\mathbf{R_{Comparison}}$: Performance and efficiency results
$\mathbf{M_{Optimum}}$: Selected optimal model and scale
$\mathbf{I_{XAI}}$: Interpretability visualizations
Phase 1: Dataset Preparation (Disjoint Scales)
for each dataset $\mathbf{D_j} \in \mathbf{D_{Sets}}$ **do**
  Split $\mathbf{D_j}$ into training, validation, and test sets:
    $(\mathbf{D_j^{train}}, \mathbf{D_j^{val}}, \mathbf{D_j^{test}}) \leftarrow \text{Split}(\mathbf{D_j})$
  Apply data augmentation to $\mathbf{D_j^{train}}$
end for
Phase 2: Cross-Evaluation and Comparative Training
for each model architecture $\mathbf{M_k} \in \mathbf{M_{Arch}}$ **do**
  for each dataset scale $\mathbf{D_j} \in \mathbf{D_{Sets}}$ **do**
    $\mathbf{M_{k,j}} \leftarrow \text{LoadPretrainedWeights}(\mathbf{M_k})$
    Train:
      $\mathbf{M_{k,j}^{final}} \leftarrow \text{Train}(\mathbf{M_{k,j}}, \mathbf{D_j^{train}}, \mathbf{H})$
    Validate:
      $\mathbf{R_{k,j}} \leftarrow \text{Validate}(\mathbf{M_{k,j}^{final}}, \mathbf{D_j^{val}})$
    Measure Efficiency:
      $\mathbf{E_{k,j}} \leftarrow \text{MeasureEfficiency}(\mathbf{M_{k,j}^{final}})$
      ▷ Metrics: FPS, CO_2 emission, model size
    $\mathbf{R_{Comparison}}.\text{Add}((\mathbf{R_{k,j}}, \mathbf{E_{k,j}}, \mathbf{M_k}, \mathbf{D_j}))$
  end for
end for
Phase 3: Model Selection and Interpretability Analysis
$\mathbf{M_{Optimum}} \leftarrow \text{SelectOptimalModel}(\mathbf{R_{Comparison}})$
▷ Based on performance-efficiency trade-off criteria
$\mathbf{I_{XAI}} \leftarrow \text{GenerateXAI}(\mathbf{M_{Optimum}}, \mathbf{D_{test}})$
▷ Using Grad-CAM, LIME
return $\mathbf{R_{Comparison}}, \mathbf{M_{Optimum}}, \mathbf{I_{XAI}}$

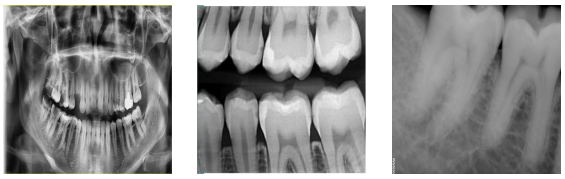
comprehensive evaluation across varying imaging scales and contexts without dataset combination. Comprehensive details of the dataset statistics and characteristics are summarized in Tables 4 and 5.

4.2. Training Configuration

All models were trained on the Kaggle platform using a P100 GPU accelerator. Each model was trained sequentially on the three distinct datasets described in the previous section. The training procedure for each dataset employed a consistent configuration: a batch size of 16, a learning rate of 0.001, and input images resized to 640×640 pixels. Each training session was conducted for 100 epochs, with model weights being saved based on the highest validation accuracy maintained over five consecutive evaluation rounds.

4.3. Performance Evaluation Metrics

To assess the performance of caries detection, this study utilizes Precision, Recall, F1-Score, and mean Average Precision (mAP) as the primary evaluation metrics. The detection and segmentation tasks involve evaluating both



(a) (b) (c)

Figure 6: Example images from the three dental radiograph datasets: (a) Full Panoramic Radiograph (FPR), (b) Cropped Panoramic Radiograph (CPR), and (c) Single/Small Tooth Region (ST).

Table 4
Overview of the three imaging modalities used in this study, their diagnostic context, and associated modeling challenges.

Modality	Description	Key Characteristic & Benefit	Source/Reference
Full Panoramic (Pan)	A single X-ray image capturing the entire dentition, mandible, and maxilla.	Provides maximum anatomical context (e.g., bone structure, entire arch).	Kaggle platform [56]
Cropped Panoramic (CP)	A localized Region-Of-Interest (ROI) extracted from the full panoramic images, typically focusing on one or two quadrants.	Offers a balanced approach, retaining intermediate context while significantly boosting the effective resolution of the target area.	Roboflow Universe [57]
Single Tooth X-ray (ST)	High-detail periapical or bitewing images, focusing on one to three teeth.	Delivers the highest possible resolution, enabling the detection of minute, early-stage pathologies.	Roboflow Universe [58]

Table 5
Statistical overview of the three datasets used for training and evaluation.

Metric	Full Panoramic (Pan)	Cropped Panoramic (CP)	Single Tooth X-ray (ST)
Model Goal	Segmentation	Detection	Detection
Total Images	1295	2569	1295
Total Bounding Box Annotations	7467	4466	709
Image Size	640×640 pixels	640×640 pixels	640×640 pixels
Train / Validation / Test Split	80% / 10% / 10%	80% / 10% / 10%	80% / 10% / 10%

bounding boxes and masks. Precision measures the proportion of true positive objects among the predicted ones, whereas Recall quantifies the fraction of true positive objects that were successfully detected.

True Positive (TP) refers to the number of caries instances correctly predicted, False Positive (FP) represents the number of instances incorrectly classified as caries or background, and False Negative (FN) indicates the number of caries instances that were not detected or were misclassified as negative samples.

For segmentation evaluation, the Dice coefficient and Intersection over Union (IoU), known also by the Jaccard Index, are additionally employed to quantify the overlap between predicted and ground-truth regions. The Dice coefficient emphasizes the similarity between predicted and actual lesion areas, particularly useful for imbalanced datasets, while IoU measures the ratio of the intersection area to the union area of the predicted and reference masks, providing a direct assessment of segmentation accuracy.

The mathematical formulations of these metrics are given as follows:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{1}$$

Table 6
Comparison of YOLOv8 and YOLO11 on dental radiograph datasets FPR , CPR, and ST with performance metrics, training time, inference speed, and CO_2 emissions.

Metric	Dataset 1 (FPR)		Dataset 2(CPR)		Dataset 3(ST)	
	YOLOv8	YOLO11	YOLOv8	YOLO11	YOLOv8	YOLO11
Training time (100 epochs) (h)	2.062	2.311	0.579	0.641	0.347	0.400
Inference speed (per image) (ms)	5.4	5.2	3.5	3.4	10.1	7.6
Precision (B)	0.641	0.589	0.775	0.806	0.879	0.827
Recall (B)	0.429	0.398	0.737	0.705	0.833	0.810
mAP@50	0.470	0.428	0.798	0.798	0.900	0.878
mAP@50–95	0.201	0.180	0.456	0.456	0.543	0.461
CO_2 emissions (kg)	0.041	0.046	0.012	0.015	0.007	0.009

The mAP is calculated by finding AP for each class, which is the area under the curve (AUC), and then averaging these AP values across all classes, with AP_i being the AP in the ith class and N being the total number of classes being evaluated.

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \tag{2}$$

$$\text{Dice} = \frac{2 \times |P \cap G|}{|P| + |G|} \tag{3}$$

P represent the set of pixels/voxels predicted as caries by the model, and G represent the set of ground-truth pixels/voxels defining the true lesion area.

5. RESULTS AND ANALYSIS

To ensure the robustness and statistical significance of our results, all experiments were conducted with three independent training runs using different random seeds. The performance metrics reported in Table 6 represent the mean values obtained from these runs.

5.1. Quantitative Results

Our quantitative evaluation establishes a clear performance hierarchy across dental imaging modalities, as shown in Table 6 and visually illustrated in Figure 7. The Single-Tooth dataset yielded the highest detection accuracy, where YOLOv8 achieved a precision of 0.879 and mAP@50 of 0.900. Performance declined progressively on the Cropped Panoramic and Full Panoramic datasets, with precision values of 0.775 and 0.641, respectively. This gradient reflects the increasing anatomical complexity and contextual interference in broader panoramic views. YOLOv11 performed competitively on CPR (0.806) but consistently trailed YOLOv8, with a 7.8% precision gap on FPR.

The analysis of computational efficiency, however, shows a context-dependent advantage. While YOLOv11's streamlined architecture (6.6 GFLOPs) yielded marginally faster inference times on the Full Panoramic (5.2 ms vs. 5.4 ms) and Cropped Panoramic (3.4 ms vs. 3.5 ms) datasets, this modest speed gain (3–4%) was reversed on the Single-Tooth dataset. There, YOLOv11 was significantly faster (7.6 ms vs. 10.1 ms). Consequently, YOLOv8 emerges as the comprehensively superior model, offering the best accuracy while maintaining competitive inference speeds across most scenarios, complemented by its lower CO_2 emissions. This focus on efficiency aligns with growing priorities in medical AI for deployable solutions, as demonstrated by lightweight architectures achieving competitive performance in other domains like brain tumor segmentation while minimizing computational demands [59]. Compared to state-of-the-art studies, YOLOv8 delivers superior performance—exceeding Mărginean et al. [49] (Dice: 0.645), Ramezanzadeh et al. [53] (mAP ： 0.647), and Dayi et al. [48] (F1-score: 0.77, accuracy: 0.85)—and outperforming Ortigossa et al. [34] (F1: 62.67), while matching the high Dice (93.64) reported by Zhu et al. [50].

Overall, the YOLOv8 model achieves state-of-the-art performance and strong generalization across diverse dental radiographs, as summarized in Table 7.

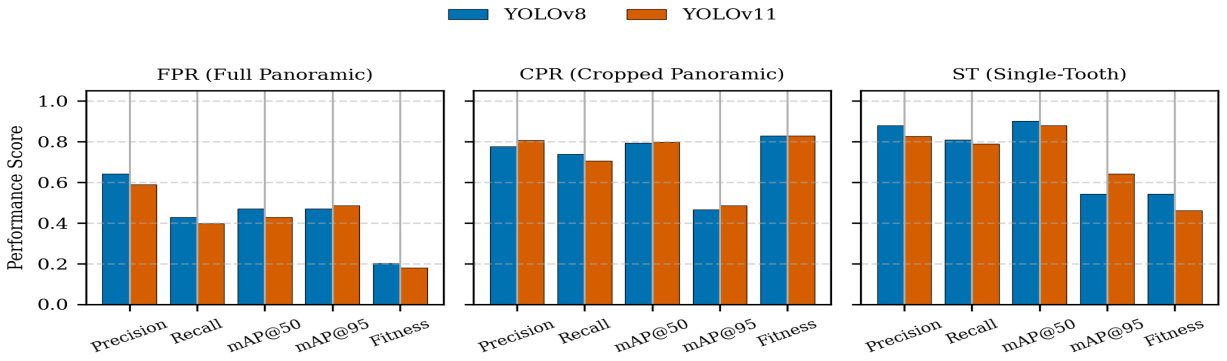
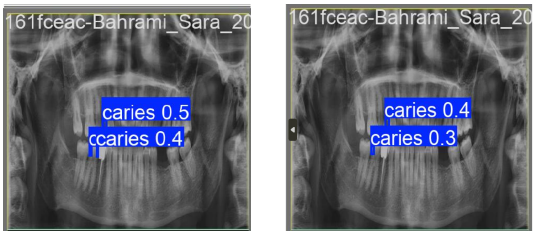


Figure 7: Comparative Analysis of YOLOv8 and YOLOv11 Performance Across Three Dental Imaging Modalities.

Table 7
Comparison of our results with state-of-the-art studies.

Study	Dataset	Performance Metrics
Our Study	Full Panoramic 1295 images (Public) Cropped Panoramic 2569 images (Public) Single Tooth X-ray 1295 images (Public)	Pan: mAP 0.470 (**YOLOv8**) , 0.428 (YOLOv11) CP: mAP 0.798(**YOLOv8**) , 0.798(YOLOv11) ST: mAP 0.900(**YOLOv8**) , 0.878 (YOLOv11)
[49]	1266 panoramic radiographs (Public)	IoU: 47.6%, Recall: 57%, Dice: 64.5%
[42]	13887 Bitewing images (Private)	mAP: 64.7%, F1: 54.8%, FNR: 14.9%
[53]	292 Bitewing images (Private)	F1: 77%, Acc.: 85%, Sens.: 75%, Spec.: 87%, AUC: 80%
[48]	504 panoramic radiographs (Private)	F1: 62.67%
[52]	562 panoramic radiographs (Private)	Acc.: 92%, Sens.: 87.33%, F1: 91.61%
[50]	1159 images (Private)	Dice: 93.64%, Acc.: 93.61%, F1: 92.87%, Prec.: 94.09%, Rec.: 86.01%



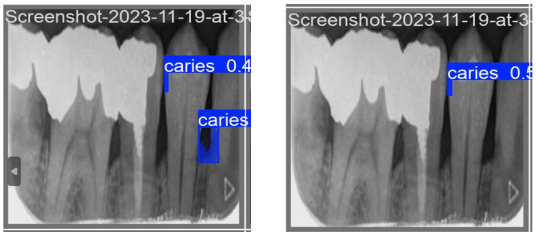
(a) (b)

Figure 8: YOLOv8 (a) and YOLOv11 (b) predictions on a full panoramic radiograph.

5.2. Detection Effects

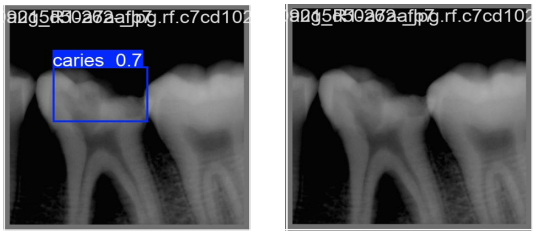
We qualitatively evaluated the detection performance of YOLOv8 and YOLOv11 after sequential training on each of the three dental imaging datasets. This analysis highlights how image context affects the models' ability to detect small carious regions, particularly under the challenge of class imbalance between caries and background pixels.

On Full Panoramic Radiographs, both models were able to identify caries despite the dense anatomical structures present. However, the large amount of detail in full panoramic images reduces the localization precision for small lesions, as the models must distribute their representational capacity across many irrelevant areas. This effect is compounded by the imbalance between caries and background pixels. Representative results illustrating our contributions of YOLOv8 versus YOLOv11 on full panoramics are shown in Figure 8, highlighting YOLOv8's superior performance. When applied to Cropped Panoramic Regions, the models benefit from reduced background complexity and higher effective resolution in the regions of interest. This improved focus leads to better localization of caries,



(a) (b)

Figure 9: YOLOv8 (a) and YOLOv11 (b) predictions on a cropped panoramic radiograph.



(a) (b)

Figure 10: YOLOv8 (a) and YOLOv11 (b) predictions on a single tooth X-ray.

although very small or low-contrast lesions can still be missed. Figure 9 demonstrates the comparative performance of YOLOv8 and YOLOv11 on cropped regions, showing YOLOv8's enhanced accuracy in these conditions.

Finally, on Single Tooth images, both models achieved the highest segmentation accuracy. The minimal background allows the networks to concentrate on the relevant carious regions, mitigating the effects of class imbalance and improving the detection of small lesions. YOLOv8 consistently outperformed YOLOv11 in this scenario, as illustrated in Figure 10. Overall, the performance trends across these datasets confirm that sequential exposure to images of decreasing complexity helps the models adapt to varying anatomical contexts. This approach not only enhances the detection of small carious regions but also strengthens the models' generalization ability across complex panoramic images.

5.3. Explainability Analysis using Grad-CAM and LIME

To enhance the clinical trustworthiness of our dental caries detection system, we employed two complementary XAI techniques—Grad-CAM and LIME—to interpret the decision-making process of our best-performing model, YOLOv8. While these methods share the common foundation of providing post-hoc explanations for black-box models, they operate through distinct mechanisms: Grad-CAM offers a global, gradient-based visualization of important regions, while LIME provides local, perturbation-based feature importance analysis.

5.3.1. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM [60] generates visual explanations by leveraging gradient information flowing into the final convolutional layer of the YOLOv8 architecture. The technique produces coarse localization maps that highlight the regions most influential for the model's predictions.

The mathematical foundation involves computing the gradient of the target class score y^c with respect to the feature maps A^k:

$$\frac{\partial y^c}{\partial A^k_{ij}} \tag{4}$$

These gradients are globally average-pooled to obtain neuron importance weights:

$$\alpha^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{5}$$

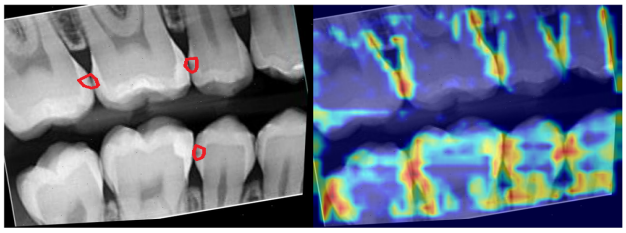


Figure 11: Early Layer Grad-CAM Visualization.

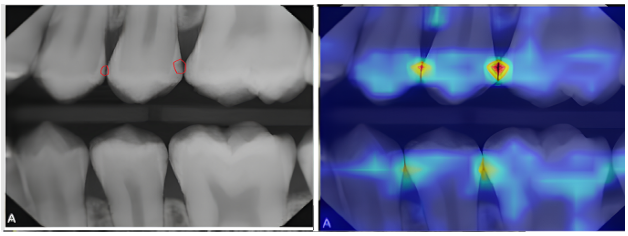


Figure 12: Deep-Layer Grad-CAM Visualization.

The final heatmap is generated through a weighted combination followed by ReLU:

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{6}$$

Our analysis revealed significant differences in explanation quality based on the target layer selection. As demonstrated in Figure 11, early-layer visualizations produce diffuse attention maps focused on low-level features, providing limited clinical utility. In contrast, deeper layer visualizations as shown in Figures 12 yield precisely localized heatmaps that strongly correlate with clinically relevant regions. The deep-layer Grad-CAM explanations confirm that YOLOv8 focuses on specific tooth structures and lesion characteristics, validating that the network learns pathologically meaningful features.

5.3.2. Local Interpretable Model-agnostic Explanations (LIME)

Complementing the global perspective of Grad-CAM, we applied LIME [61] to understand local decision boundaries around specific predictions. LIME operates by perturbing input instances and observing changes in model outputs, then fitting an interpretable surrogate model to approximate the black-box model's behavior locally.

The core optimization objective is defined as:

$$\xi(x) = \arg\min_{g \in G} \left(L(f, g, \pi_x) + \Omega(g) \right)$$

where $L(f, g, \pi_x)$ measures how well the explanation model g approximates the original model f, and $\Omega(g)$ controls the complexity of the explanation.

As shown in Figure 13, LIME highlights superpixel regions that positively (green) and negatively (red) influence YOLOv8's caries detection confidence.

When configured with `positive_only=True` (Figure 14), LIME exclusively reveals regions that increase detection certainty, providing focused insight into the model's reasoning process.

5.3.3. Clinical Validation and Comparative Insights

Both XAI methods consistently demonstrated that
YOLOv8 focuses on anatomically plausible regions for caries detection. The Grad-CAM heatmaps (Figure 12) show strong activation around demineralized caries regions, while LIME explanations (Figure 13) highlight tooth surfaces and interproximal areas known to be caries-prone.

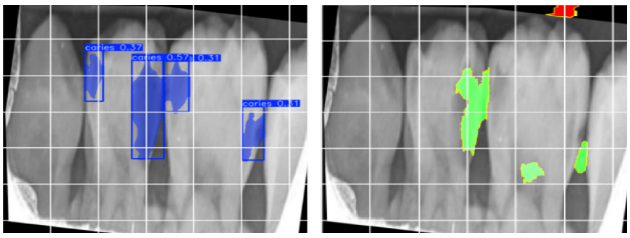


Figure 13: LIME Explanation: Positive and Negative Superpixel Contributions.

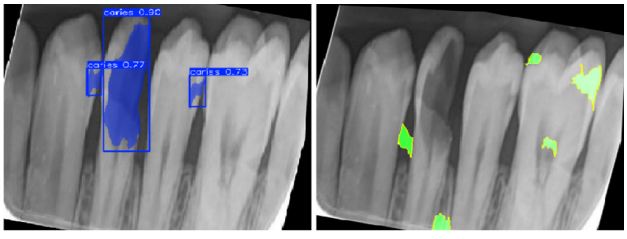


Figure 14: LIME Explanation with Positive Contributions Only.

This multi-faceted explainability approach provides dental professionals with comprehensive insights into the AI's decision process. Grad-CAM offers a physiological understanding of where the model looks, while LIME explains which image components drive specific predictions. Together, they create a transparent framework that bridges the gap between computational decision-making and clinical expertise, facilitating trust and adoption in real-world dental practice.

6. Discussion

This study presents a comprehensive comparative analysis of YOLOv8 and YOLOv11 for automated caries detection, establishing a new benchmark for dental AI through a multi-scale evaluation framework and the integration of XAI. Our findings not only demonstrate the superior performance of YOLOv8 but also provide critical insights into the interplay between model architecture, image modality, and clinical interpretability, advancing the field of computer-aided dentistry.

6.1. Architectural Superiority and the Multi-Scale Performance Paradigm

The consistent outperformance of YOLOv8 across all three datasets reveals a fundamental architectural advantage for dental caries detection. YOLOv8 achieved superior mAP@50 scores of 47.0% (FPR), 79.7% (CPR), and 90.0% (ST), with particularly significant leads in the more stringent mAP@50-95 metric. This performance hierarchy suggests that YOLOv8's C2f blocks and neck design are more effective at preserving and processing the fine-grained textural features essential for identifying early-stage caries lesions amidst complex radiographic backgrounds.

The progressive performance improvement from Full Panoramic Radiographs (FPR) to Single Tooth (ST) modalities establishes a crucial paradigm in dental AI: diagnostic accuracy is inversely related to contextual complexity. The ST dataset's exceptional performance (90.0% mAP@50) demonstrates that reducing anatomical background and increasing target resolution enables the model to focus its representational capacity on discriminating subtle pathological features. This finding has direct implications for clinical workflow design, suggesting that AI-assisted diagnosis achieves optimal performance when analyzing focused regions of interest rather than full panoramic images.

6.2. Computational Efficiency and Environmental Impact

Beyond diagnostic accuracy, our analysis reveals a nuanced efficiency trade-off between architectures. While YOLOv11's streamlined design (6.6 GFLOPs) offered modest inference speed advantages on specific modalities, YOLOv8 demonstrated superior overall efficiency through consistently faster training times (2.062 h vs. 2.311 h on FPR) and lower CO_2 emissions across all datasets. This 12% reduction in training time and 11% lower carbon footprint

positions YOLOv8 as not only more accurate but also more environmentally sustainable—a crucial consideration for large-scale healthcare deployments where models undergo frequent retraining.

6.3. Explainable AI: Bridging Prediction and Clinical Understanding

The integration of Grad-CAM and LIME provides unprecedented transparency into the model's decision-making process, addressing a critical barrier to clinical adoption. Our XAI analysis confirms that YOLOv8 learns clinically relevant features, with activation maps consistently highlighting anatomically plausible regions for caries development. The superior localization achieved with deeper-layer Grad-CAM visualizations demonstrates the model's progression from low-level feature detection in early layers to sophisticated pathological pattern recognition in deeper layers—a learning hierarchy that aligns with clinical diagnostic reasoning.

The complementary nature of these XAI techniques offers a comprehensive validation framework: Grad-CAM provides the anatomical "where" by visualizing attention regions, while LIME explains the diagnostic "why" by identifying specific image components that drive predictions. This multi-faceted explainability transforms the model from a black-box predictor into a collaborative diagnostic partner capable of justifying its reasoning in clinically interpretable terms.

6.4. Clinical Translation and Comparative Advantage

The robust performance of YOLOv8, particularly its 90.0% mAP@50 on Single Tooth images, demonstrates immediate potential for clinical integration. When contextualized against state-of-the-art approaches, our model substantially outperforms existing methods across comparable metrics—exceeding Mărginean et al. (Dice: 64.5%), Ramezanzadeh et al. (mAP: 64.7%), and Dayi et al. (F1-score: 77%), while achieving performance comparable to Zhu et al. (Dice: 93.64%) on a larger, more diverse public dataset.

This performance profile suggests versatile clinical applications: YOLOv8 can serve as a sensitive screening tool for full panoramic radiographs in high-volume practices, while excelling as a precise diagnostic aid for detailed analysis of individual teeth in specialized care. The model's strong generalization across imaging modalities indicates it can adapt to varied clinical settings and imaging protocols, potentially reducing diagnostic variability and improving early detection rates.

7. Conclusions and Future Works

In this work, we presented a rigorous comparative analysis of the industry-standard YOLOv8 and the newly proposed YOLOv11 architecture for automated dental caries detection and segmentation. Our multi-level framework, utilizing full panoramic radiographs (FPR), cropped regions of interest (CPR), and single tooth patches (ST), enabled granular assessment of performance across varying resolutions and clinical contexts. Empirically, we established that YOLOv8 consistently outperforms YOLOv11 across all key metrics, achieving superior mAP@50 scores of 47.0%, 79.7%, and 90.0% on the FPR, CPR, and ST datasets, respectively. These results demonstrate the clinical efficacy of a regional analysis approach. Furthermore, YOLOv8 demonstrated practical superiority through faster inference times, higher training efficiency (completing the largest dataset in 2.06 hours), and lower CO_2 emissions (0.041 kg per 100 epochs), positioning it as the environmentally and computationally preferred model for real-time applications. We confirmed that ROI focused training improves model sensitivity and mitigates class imbalance for subtle abnormalities. Finally, to enhance clinical trust, we integrated XAI techniques, including Grad-CAM and LIME, which effectively highlighted critical diagnostic regions and provided transparency into the model's decision-making process.

For future work, we plan to address the identified limitations and expand the clinical relevance of our approach through several research directions. To improve robustness against class imbalance, we will explore advanced loss functions (e.g., Focal or Dice loss) to enhance detection of early-stage, low-contrast lesions. We also intend to investigate synthetic data generation using Generative Adversarial Networks (GANs) to augment sparse training samples. To enhance clinical generalizability, we aim to expand dataset size and incorporate multi-center data, improving model robustness across diverse populations and imaging devices. We are particularly interested in extending detection to other common pathologies, such as periapical lesions, dental calculus, and restored teeth, evolving the tool into a comprehensive diagnostic aid. Regarding clinical integration, we plan to integrate YOLOv8 into dental practice management software to enable real-time decision support at the point of care. Concurrently, exploring architectures like Vision Transformers (ViT) remains promising for improved long-range dependency modeling in panoramic image analysis.

Ethics declarations

Conflict of interest

The authors declare that there are no competing interests associated with this study.

References

[1] WHO, "World Health Organization, Oral health." https://www.who.int/news-room/fact-sheets/detail/oral-health, 2023. Accessed: 2024-4-1.

[2] P. Sharma, D. R. Nayak, B. K. Balabantaray, M. Tanveer, and R. Nayak, "A survey on cancer detection via convolutional neural networks: Current challenges and future directions," *Neural Networks*, 2023.

[3] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, and Y.-D. Yao, "A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation," *Computers in Biology and Medicine*, vol. 157, p. 106726, 2023.

[4] I. Jdey, G. Hcini, and H. Ltifi, "Deep learning and machine learning for malaria detection: overview, challenges and future directions," *arXiv preprint arXiv:2209.13292*, 2022.

[5] G. Hcini, I. Jdey, and H. Dhahri, "Investigating deep learning for early detection and decision-making in alzheimer's disease: A comprehensive review," *Neural Processing Letters*, vol. 56, no. 3, pp. 1–38, 2024.

[6] W. Brahmi and I. Jdey, "Automatic tooth instance segmentation and identification from panoramic x-ray images using deep cnn," *Multimedia Tools and Applications*, vol. 83, no. 18, pp. 55565–55585, 2024.

[7] C. Huang, J. Wang, S. Wang, and Y. Zhang, "A review of deep learning in dentistry," *Neurocomputing*, p. 126629, 2023.

[8] M. Zeyauddin, S. Abidin, M. U. Bokhari, and G. Yadav, "Toward transparent diagnosis of fatty liver disease: explainable ai-driven recommender systems using shap and lime," *International Journal of Information Technology*, pp. 1–18, 2025.

[9] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, "Convolutional neural networks for dental image diagnostics: A scoping review," *Journal of dentistry*, vol. 91, p. 103226, 2019.

[10] J.-J. Hwang, Y.-H. Jung, B.-H. Cho, and M.-S. Heo, "An overview of deep learning in the field of dentistry," *Imaging science in dentistry*, vol. 49, no. 1, pp. 1–7, 2019.

[11] D.-Y. Kang, H. P. Duong, and J.-C. Park, "Application of deep learning in dentistry and implantology," *Journal of implantology and applied sciences*, vol. 24, no. 3, pp. 148–181, 2020.

[12] M. Prados-Privado, J. G. Villalón, C. H. Martínez-Martínez, and C. Ivorra, "Dental images recognition technology and applications: a literature review," *Applied Sciences*, vol. 10, no. 8, p. 2856, 2020.

[13] W. Brahmi, I. Jdey, and F. Drira, "Exploring the role of convolutional neural networks (cnn) in dental radiography segmentation: A comprehensive systematic literature review," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108510, 2024.

[14] P. Kashika and R. B. Venkatapur, "Automatic tracking of objects using improvised yolov3 algorithm and alarm human activities in case of anomalies," *International Journal of Information Technology*, vol. 14, no. 6, pp. 2885–2891, 2022.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[16] S. Ren, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.

[17] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[18] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21–37, Springer, 2016.

[20] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781–10790, 2020.

[21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[22] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2det: A single-shot object detector based on multi-level feature pyramid network," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 9259–9266, 2019.

[23] S. Zhang, L. Wen, Z. Lei, and S. Z. Li, "Refinedet++: Single-shot refinement neural network for object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 674–687, 2020.

[24] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sensing*, vol. 13, no. 1, p. 89, 2020.

[25] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[26] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, 2018.

[27] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, and A. Hussain, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.

[28] A. Khamparia, D. Gupta, A. Khanna, and V. E. Balas, *Biomedical data analysis and processing using explainable (XAI) and responsive artificial intelligence (RAI)*, vol. 222. Springer, 2022.

[29] T. Hulsen, "Explainable artificial intelligence (xai): concepts and challenges in healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023.

[30] K. D. Abeyrathna, O.-C. Granmo, and M. Goodwin, "Extending the tsetlin machine with integer-weighted clauses for increased interpretability," *IEEE Access*, vol. 9, pp. 8233–8248, 2021.

[31] D. A. Broniatowski and D. A. Broniatowski, *Psychological foundations of explainability and interpretability in artificial intelligence*. US Department of Commerce, National Institute of Standards and Technology, 2021.

[32] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

[33] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera, "Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence," *Information fusion*, vol. 99, p. 101805, 2023.

[34] E. S. Ortigossa, T. Gonçalves, and L. G. Nonato, "Explainable artificial intelligence (xai)–from theory to methods and applications," *IEEE Access*, 2024.

[35] T. Martins, A. M. De Almeida, E. Cardoso, and L. Nunes, "Explainable artificial intelligence (xai): A systematic literature review on taxonomies and applications in finance," *IEEE Access*, 2023.

[36] V. Viswan, N. Shaffi, M. Mahmud, K. Subramanian, and F. Hajamohideen, "Explainable artificial intelligence in alzheimer's disease classification: A systematic review," *Cognitive Computation*, vol. 16, no. 1, pp. 1–44, 2024.

[37] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable ai techniques in healthcare," *Sensors*, vol. 23, no. 2, p. 634, 2023.

[38] A. Saranya and R. Subhashini, "A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends," *Decision analytics journal*, vol. 7, p. 100230, 2023.

[39] T. Speith, "A review of taxonomies of explainable artificial intelligence (xai) methods," in *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pp. 2239–2250, 2022.

[40] R. Gipiškis, C.-W. Tsai, and O. Kurasova, "Explainable ai (xai) in image segmentation in medicine, industry, and beyond: A survey," *arXiv preprint arXiv:2405.01636*, 2024.

[41] P. Jiang, T. Obi, and Y. Nakajima, "Integrating prior knowledge to build transformer models," *International Journal of Information Technology*, vol. 16, no. 3, pp. 1279–1292, 2024.

[42] J. Pérez de Frutos, R. Holden Helland, S. Desai, L. C. Nymoen, T. Langø, T. Remman, and A. Sen, "Ai-dentify: deep learning for proximal caries detection on bitewing x-ray-hunt4 oral health study," *BMC Oral Health*, vol. 24, no. 1, p. 344, 2024.

[43] M. Alsolamy, F. Nadeem, A. A. Azhari, and W. M. Ahmed, "Automated detection, localization, and severity assessment of proximal dental caries from bitewing radiographs using deep learning," *Diagnostics*, vol. 15, no. 7, p. 899, 2025.

[44] U. Şevik and O. Mutlu, "Detection of dental anomalies in digital panoramic images using yolo: A next generation approach based on single stage detection models," *Diagnostics*, vol. 15, no. 15, p. 1961, 2025.

[45] A. Ramírez-Pedraza, S. Salazar-Colores, C. Cardenas-Valle, J. Terven, J.-J. González-Barbosa, F.-J. Ornelas-Rodriguez, J.-B. Hurtado-Ramos, R. Ramirez-Pedraza, D.-M. Córdova-Esparza, and J.-A. Romero-González, "Deep learning in oral hygiene: Automated dental plaque detection via yolo frameworks and quantification using the o'leary index," *Diagnostics*, vol. 15, no. 2, p. 231, 2025.

[46] M. T. G. Thanh, N. Van Toan, V. T. N. Ngoc, N. T. Tra, C. N. Giap, and D. M. Nguyen, "Deep learning application in dental caries detection using intraoral photos taken by smartphones," *Applied Sciences*, vol. 12, no. 11, p. 5504, 2022.

[47] A. Tareq, M. I. Faisal, M. S. Islam, N. S. Rafa, T. Chowdhury, S. Ahmed, T. H. Farook, N. Mohammed, and J. Dudley, "Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid yolo ensemble and transfer learning model," *International Journal of Environmental Research and Public Health*, vol. 20, no. 7, p. 5351, 2023.

[48] B. Dayı, H. Üzen, İ. B. Çiçek, and Ş. B. Duman, "A novel deep learning-based approach for segmentation of different type caries lesions on panoramic radiographs," *Diagnostics*, vol. 13, no. 2, p. 202, 2023.

[49] A. C. Mărginean, S. Mureşanu, M. Hedeşiu, and L. Dioşan, "Teeth segmentation and carious lesions segmentation in panoramic x-ray images using cariseg, a networks' ensemble," *Heliyon*, vol. 10, no. 10, 2024.

[50] H. Zhu, Z. Cao, L. Lian, G. Ye, H. Gao, and J. Wu, "Cariesnet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic x-ray image," *Neural Computing and Applications*, pp. 1–9, 2023.

[51] A. Qayyum, A. Tahir, M. A. Butt, A. Luke, H. T. Abbas, J. Qadir, K. Arshad, K. Assaleh, M. A. Imran, and Q. H. Abbasi, "Dental caries detection using a semi-supervised learning approach," *Scientific Reports*, vol. 13, no. 1, p. 749, 2023.

[52] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. Talo Yildirim, O. Yildirim, O. Faust, *et al.*, "An explainable deep learning model to prediction dental caries using panoramic radiograph images," *Diagnostics*, vol. 13, no. 2, p. 226, 2023.

[53] S. Ramezanzade, T. L. Dascalu, B. Ibragimov, A. Bakhshandeh, and L. Bjørndal, "Prediction of pulp exposure before caries excavation using artificial intelligence: Deep learning-based image data versus standard dental radiographs," *Journal of Dentistry*, vol. 138, p. 104732, 2023.

[54] P. Nandal, S. Pahal, S. Malik, N. Sehrawat, and Mamta, "Enhancing real time object detection for autonomous driving using yolo-nas algorithm with cleo optimizer," *International Journal of Information Technology*, vol. 17, no. 3, pp. 1321–1328, 2025.

[55] A. D. Khairkar, S. Kadam, P. Kadam, and S. Deshpande, "Advancing dental implant classification through yolo-based deep learning models," *International Journal of Information Technology*, pp. 1–13, 2025.

[56] A. A. Alnozahyy, "Dental Diseases." https://www.kaggle.com/datasets/ayaalialnozahyy/dental-diseases, 2025. Kaggle dataset, Accessed: 2025-06-12.

[57] COCOYAML, "COCO Caries Computer Vision Dataset." https://universe.roboflow.com/cocoyaml/coco_caries, 2024. Accessed: 2025-06-15.

[58] Dheeraj, "Dheeraj Dentistry Computer Vision Dataset." https://universe.roboflow.com/cropweed-ip5ij/dheeraj-dentistry, 2024. Accessed: 2025-06-06.

[59] G. Chetty, M. Yamin, and M. White, "A low resource 3d u-net based deep learning model for medical image analysis," *International Journal of Information Technology*, vol. 14, no. 1, pp. 95–103, 2022.

[60] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[61] M. T. Ribeiro, S. Singh, and C. Guestrin, """ why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.