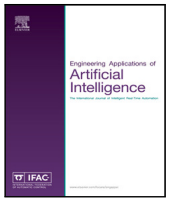
Engineering Applications of Artificial Intelligence 133 (2024) 108510

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Survey paper

Exploring the role of Convolutional Neural Networks (CNN) in dental radiography segmentation: A comprehensive Systematic Literature Review

Walid Brahmi [a,b,*], Imen Jdey [a], Fadoua Drira [a]

[a] Research Groups in Intelligent Machines (REGIM Lab), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, 3038 Sfax, Tunisia
[b] National School of Electronics and Telecommunications of Sfax (ENET'Com), University of Sfax, Tunisia

ARTICLE INFO

Keywords:
Deep learning
Convolutional Neural Network
Dental imaging
Segmentation
Evaluation metrics

ABSTRACT

In dentistry, there is a growing need for accurate diagnostic tools, particularly advanced imaging techniques such as Computed Tomography (CT), Cone Beam Computed Tomography (CBCT), Magnetic Resonance Imaging (MRI), ultrasound, and traditional intraoral periapical X-rays. Deep Learning (DL) has emerged as a pivotal tool, facilitating automated segmentation that is crucial for extracting essential diagnostic data.

This integration of cutting-edge technology addresses the urgent need for effective management of dental conditions. If left undetected, these conditions can significantly affect human health. Deep Learning (DL) has an impressive track record in various domains, including dentistry, underscoring its potential to revolutionize early detection and treatment of oral health issues. Convolutional Neural Networks (CNNs) have demonstrated significant results in diagnosis and prediction, representing an emerging field of multidisciplinary research. The goals of this study were to provide a concise overview of the state of the art, standardize ongoing debates, and establish baselines for future research.

The methodology employed in this study involved a Systematic Literature Review (SLR) to identify and select relevant studies that specifically investigated deep learning techniques for dental imaging analysis. This study elucidates a methodological approach that includes systematic data collection, statistical analysis, and outcome dissemination.

By incorporating 45 studies, we identified the selection criteria and research objectives, addressing significant gaps in the existing literature. These studies will assist clinicians in examining dental conditions and classifying dental structures, such as detecting cavities and identifying different types of teeth. The model performance was evaluated by addressing the identified gaps using a variety of metrics that were outlined and explained.

This study demonstrated the effectiveness of using CNNs to analyze images and serves as an effective tool for detecting dental pathologies. Despite acknowledging some limitations, CNNs used for segmenting and categorizing teeth demonstrated their highest level of performance overall.

1. Introduction

Medical information plays a vital role in healthcare by assisting healthcare professionals in making accurate diagnoses, providing effective treatments, and making informed decisions about patient care (Mogli, 2009; Pickering et al., 2013). In the realm of medical information, medical imaging is an invaluable resource (Jdey et al., 2023).

In contemporary healthcare, medical imaging is indispensable, providing crucial visual insights into the internal structures and conditions of the human body (Abedalla et al., 2021). In dentistry, dental imaging refers to the use of various imaging techniques to capture detailed images of oral and maxillofacial structures. These images are invaluable for diagnosing dental conditions, planning treatment, and monitoring oral health. Dental imaging plays a vital role in enabling dental professionals to visualize and evaluate the teeth, jaws, and supporting structures. It encompasses various modalities such as X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) (Hcini et al., 2024), and intraoral cameras to capture detailed images of the teeth and surrounding structures. Each imaging modality has distinct strengths and is selected based on specific diagnostic requirements and the clinical context.

Deep learning (DL), particularly Convolutional Neural Networks (CNNs) and Computer Vision (CV), has revolutionized dentistry. Cutt-

* Corresponding author at: National School of Electronics and Telecommunications of Sfax (ENET'Com), University of Sfax, Tunisia.
E-mail addresses: bensghaierwaleed@gmail.com (W. Brahmi), imen.jdey@fstsbz.u-kairouan.tn (I. Jdey), fadoua.drira@enis.tn (F. Drira).

https://doi.org/10.1016/j.engappai.2024.108510
Received 9 January 2024; Received in revised form 27 March 2024; Accepted 23 April 2024
Available online 11 May 2024

ing-edge technologies are transforming the dental industry, where dental image analysis and AI integration reshape various aspects of dental care (Khanna and Dhaimade, 2017). Harnessing DL algorithms and CV techniques, dentistry can significantly improve diagnostic accuracy, optimize treatment planning, and ultimately enhance patient outcomes.

The application of CNNs in dentistry involves detecting structures, such as teeth and bones, as well as identifying pathologies, such as caries and apical lesions. Additionally, CNNs are used to segment images by isolating the areas of interest and classifying them based on specific features, such as enamel caries lesions or cysts. However, it is important to acknowledge the limitations of this study. A significant weakness is the relatively small and private nature of the available datasets, which restricts the diversity and size of the training data. Moreover, some of the Artificial Intelligence (AI) solutions developed in this domain may lack robustness and stability, raising concerns regarding their reliability and performance.

Tooth segmentation plays a vital role in analyzing dental images for tasks textcolorbluesuch aslesion detection, age and sex determination, and human identification. In oral medicine, the automatic segmentation of teeth in panoramic radiographs is a significant focus of image analysis research. Segmenting teeth in panoramic radiographs presents challenges due to the presence of other anatomical structures such as the chin, spine, and jaws.

In the existing literature, numerous studies have applied CNNs to tooth segmentation and identification in dental images. These advanced DL techniques have exhibited promising outcomes by leveraging the capabilities of CNNs to extract meaningful features and accurately segment teeth in dental images.

This paper is organized as follows: Section 2 provides a summary of the current state-of-the-art in automatic tooth segmentation and references relevant surveys in the field. In Section 3, we delve into the principles of DL, exploring its various categories, common applications, and, specifically, focusing in the architecture of CNNs within the context of medical image analysis in CV. The subsequent Sections 4, 5, and 6 present the study methodologies, research topics, and synthesis findings, offering detailed insights into the research. Finally, Sections 7 and 8 discuss the limitations of the study and provide conclusions.

2. Related work

In recent years, several surveys on deep learning-based dental image segmentation have been conducted. This Section provides a summary of related studies and highlights the key distinctions between this study and existing surveys.

Schwendicke et al. (2019) conducted a survey on the application of CNNs to dental and oral medicine imagery. Their review included 36 relevant articles and conference proceedings published between 2015 and 2019. It addressed clinical challenges in various areas, such as general odontology. Cariology, endodontics, periodontics, orthodontics, dental radiology, and general medicine.

Hwang et al. (2019) published a survey on deep learning in oral and maxillofacial radiology, identifying 25 relevant papers up to December 2018 using PubMed, Scopus, and IEEE Explore databases. The study collected data on the deep learning architecture, training dataset size, evaluation results, advantages and disadvantages, study objectives, and imaging modalities.

Kang et al. (2020) reviewed a previous study that explored the use of deep-learning algorithms in dentistry and implantology. This study analyzed 62 articles from MEDLINE and IEEEXplore, categorizing them into tooth detection, numbering, segmentation, and bone segmentation. Articles published before October 24, 2019, were included. The study provided information on author, year of publication, architecture, input, output, and performance metrics

Prados-Privado et al. (2020) systematically investigated the current state of AI in dental applications, covering the detection of diverse dental conditions, such as teeth, caries, filled teeth, crowns, prostheses, dental implants, and endodontic treatments. The researchers used three digital databases (PubMed, IEEE Xplore, and arXiv.org) to identify 18 relevant papers. It is noteworthy that the study deliberately excluded any segmentation methods based on deep learning.

Table 1 summarizes the central distinguishing factors between this review and existing related work. Our contributions are as follows:

- Embarking on a comprehensive exploration into the realm of Convolutional Neural Networks (CNNs), we delve into the intricate tapestry of concepts, theoretical foundations, and architectures.
- Our study thoroughly examines crucial challenges in Deep Learning, including the shortage of training data, data imbalance, and sample data quality. We also delved into the proposed solutions to address these challenges.
- We are trying to categorize a comprehensive list of medical imaging applications using deep learning based on specific tasks.
- By exploring well-established platforms such as Kaggle and GitHub and specialized resources like Roboflow, researchers can efficiently access a variety of public datasets mentioned in our paper. This streamlined approach promotes collaborative research by providing a diverse range of resources to support and enhance their work.

3. Background

3.1. Convolutional Neural Networks (CNNs)

A Convolutional Neural Networks (CNNs) are deep learning models tailored for analyzing grid-structured data, such as images. Composed of multiple layers with distinct functions, the CNN architecture includes convolutional layers, ReLU layers, pooling layers, and a fully connected layer (Fig. 1) (Jlassi et al., 2021). CNNs automate feature extraction from images, thereby eliminating the need for manual feature engineering. The CNN architecture is composed of two primary Sections: convolutional layers and densely connected layers.

In the convolutional layers, tensors, known as feature maps, are used. A color image, denoted as a 3-dimensional tensor with Red, Green, and Blue (RGB) channels (height, width, and channels), is represented. For illustration, consider Fig. 1 using a grayscale image of width × height pixels (Hcini et al., 2023). This corresponds to a shape tensor (width, height, 1), indicating the number of neurons. With a width and height of 28, the tensor has 784 neurons. For a color image, the tensor shape is (28, 28, 3), totaling 2352 neurons. Crucially, the image, which is a matrix with pixel colors ranging from 0 to 255, undergoes normalization to a color range between 0 and 1 before entering the network.

A typical CNN architecture involves multiple repetitions of convolutional layers and pooling layers, followed by one or more fully connected layers (Alzubaidi et al., 2021). The process of transforming the input data into output data through these layers is known as forward propagation. The first two layers, convolution and pooling, play a crucial role in feature extraction. The convolution layer applies filters to the input data to capture the distinct features or patterns. Subsequently, the pooling layer reduces the dimensionality of the feature maps and preserves the essential information while discarding some spatial details.

The final layer, that is, the fully connected layer, is responsible for mapping the extracted features to the ultimate output, such as the classification. It establishes connections between every neuron in the previous layer and the neurons in the current layer, enabling the network to learn intricate relationships and make predictions based on extracted features.

Table 1
Comparative Analysis of Our SLR and Existing Literature Reviews on Dental Image Segmentation..

Study	Year	Total number of papers	Limitations
Schwendicke et al. (2019)	2019	36	Studies not employing deep learning for processing dental images have been included in the research. Furthermore, the study lacks a comprehensive addressing of the elements influencing the performance of the model.
Hwang et al. (2019)	2019	25	The factors influencing the model's performance were not addressed.
Kang et al. (2020)	2020	62	The research incorporates studies unrelated to dental images. Additionally, it fails to address the factors affecting the model's performance.
Prados-Privado et al. (2020)	2020	18	The paper does not deal with the segmentation task. The elements impacting the model's performance were not addressed.

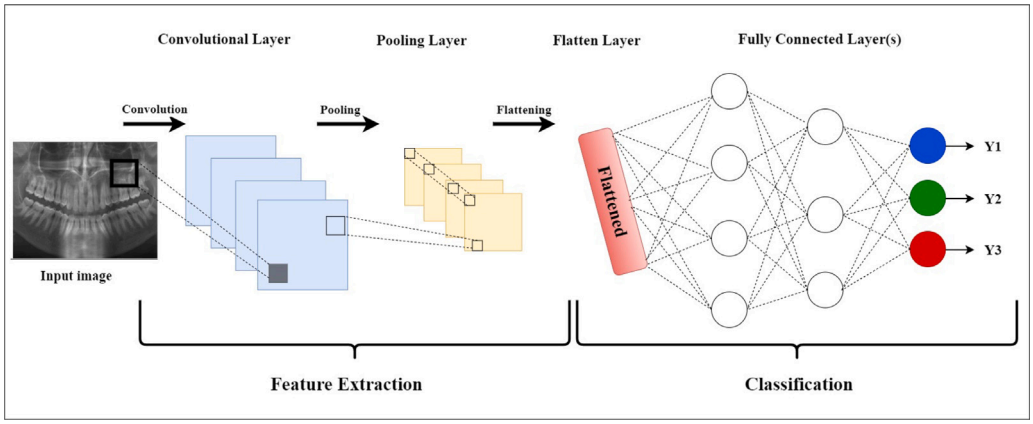


Fig. 1. Outline of CNN.

3.2. Medical images analysis

In medical image analysis, tasks involving CV, such as image classification, medical image segmentation, and object detection and recognition, are widely prominent. DL approaches, especially those relying on CNNs, have demonstrated remarkable performance in addressing these complex challenges across various medical applications. In this Section, we present a thorough and technical assessment of current research that exploits the latest advancements in DL and CNN-based algorithms to propel the field of medical image analysis and enhance our understanding.

3.2.1. Object detection

Object detection has garnered significant interest from researchers over the past few decades. With remarkable progress in DL techniques, the integration of AI in healthcare, where object detection plays a crucial role, has become more prominent.

Object detection refers to the identification and localization of instances of specific objects in an image. This technique uses CV and image processing to recognize and locate objects in still frames and videos. Object detection systems accurately determine the number of objects present in a given area, track their positions, and classify them accordingly.

The object detection process involves bounding-box annotation, in which objects are outlined with bounding boxes. The task involves recognizing objects captured in an image, with each object category possessing distinct characteristics for accurate categorization. Automated medical imaging can effectively identify bone fractures, abnormal cellular activities, and other medical conditions.

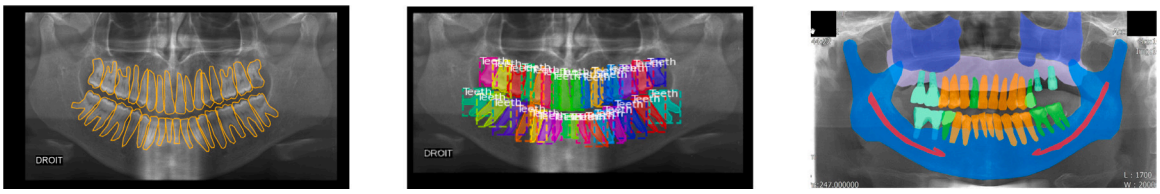
Several models have been developed in dentistry to identify various dental features, such as tooth decay detection (Bayraktar and Ayan, 2022; Lee et al., 2021; Cantu et al., 2020; Lee et al., 2018), periapical lesion detection (Setzer et al., 2020), and dental plaque detection (You et al., 2020). Common architectures utilized by researchers in the reviewed papers include Faster R-CNN (Jang et al., 2022; Chen et al., 2019; Bilgir et al., 2021), Single Shot MultiBox Detector (SSD) (Zhang et al., 2022), and You Only Look Once (YOLO) (Bayraktar and Ayan, 2022; Moidu et al., 2022; Jiang et al., 2022). Object detection involves two main stages: feature extraction from the target, and subsequent object classification and localization.

3.2.2. Classification

Image classification is commonly employed to label an image or a series of images as having specific diseases or not (Hcini et al., 2022). Traditionally, image classification involves extracting low or mid-level features to represent the image, followed by employing a trainable classifier to determine the correct label. In recent years, deep CNNs have demonstrated their superiority over manually designed low-level and mid-level features in terms of high-level feature representation. By combining feature extraction and classification networks, deep CNNs provide a unified approach that allows simultaneous training. For detailed information on DL-based medical image classification methods in clinical applications, two excellent reviews by Ker et al. (2017), Litjens et al. (2017) are recommended. CNNs have also found applications in dentistry for classifying medical images (Alotaibi et al., 2022; Aljabri et al., 2022; Hcini et al., 2021; Chen et al., 2021b).

3.2.3. Image segmentation

Image segmentation is the process of dividing an image into meaningful subregions or segments to identify the regions of interest (Heni et al., 2023). This simplifies the analysis and understanding of the entire scene in further image-processing stages. Segmentation involves



(a) Semantic segmentation (b) Instance segmentation (c) Panoptic segmentation

Fig. 2. The three common type of image segmentation.

partitioning a digital image into connected pixels or regions that share common visual characteristics such as intensity, color, texture, histogram, or features (Zhu et al., 2021). As indicated in Fig. 2, three primary segmentation strategies are commonly used in building image segmentation models: semantic segmentation 2(a), instance segmentation 2(b), and panoptic segmentation 2(c). Semantic segmentation involves labeling each pixel with a class label, indicating the category to which it belongs. Instance segmentation goes a step further by not only assigning class labels to pixels, but also distinguishing different instances of the same class. Panoptic segmentation aims to combine the advantages of semantic and instance segmentation to provide a comprehensive understanding of an image scene.

These segmentation strategies play crucial roles in various medical image analysis tasks, enabling the precise identification and delineation of regions or objects of interest, which can greatly aid in diagnosis, treatment planning, and research in dentistry and other medical fields.

3.2.4. Others

In the realm of medical imaging, various tasks harness DL techniques for enhanced outcomes. Image Reconstruction (Pauwels, 2021) employs various techniques to enhance the quality of medical images, including improving resolution, reducing noise, and optimizing image quality. In parallel, Image Registration (Kaji and Kida, 2019) aligns multiple medical images from different modalities or time points, facilitating comparison for treatment planning and disease monitoring. Image denoising (Kaji and Kida, 2019) leverages DL to minimize noise and artifacts in medical images, resulting in clearer and more accurate representations. Collectively, these applications showcase the versatility of DL in advancing various aspects of medical image analysis and interpretation.

4. Materials and methods

4.1. Method of study

In this study, we used the Systematic Literature Review (SLR) method—a rigorous and structured approach involving the assessment, interpretation, and identification of all existing research findings to address specific research questions (Keele, 2007). This method adheres to a systematic process and protocols to minimize bias and ensure an objective understanding (Wahono, 2020). The objective of our SLR was to identify and analyze research trends, methodologies, datasets, and frameworks pertaining to deep CNN approaches for dental segmentation.

4.2. Search and selection process

The literature review process involved several steps, as depicted in Fig. 3, outlining the study selection procedure for SLR. Initially, all articles relevant to deep CNN-based approaches for dental segmentation were identified through preliminary screening, using specific search strings. Subsequently, the articles were subjected to inclusion and exclusion criteria. Data extraction was conducted based on predefined criteria and pertinent information was extracted from the selected articles. Finally, the collected data were analyzed.

Table 2
Total number of papers found in a database.

	Database name	Total number of papers found (N)
1	ACM Digital Library	81
2	IEEE Xplore Digital Library	33
3	Sciences Direct Elsevier	269
4	Springer link	178
5	Willey Online Library	81
Total	642	

Of the 642 articles initially identified, 45 were selected for further study based on their relevance to the research domain. The inclusion and exclusion criteria along with the defined research protocol were applied consistently throughout the selection process to ensure the quality and relevance of the selected articles.

By employing the SLR method, we aimed to provide a comprehensive and objective analysis of existing research on deep CNN-based approaches for dental segmentation.

4.2.1. Initial search

Before commencing the search, we conducted an initial exploration to ensure the presence of a sufficient number of articles in the targeted field. In this stage, we identified 45 articles discussing the enhancement of dental imaging segmentation using deep CNN methods. This confirmation validated the suitability of the topic for conducting a Systematic Literature Review (SLR).

4.2.2. Manual search

For the manual search, we selected the appropriate search strings to capture a broad spectrum of related articles. We utilized five online databases: the IEEE Xplore Digital Library (IEEE, 2023), ACM Digital Library (ACM, 2023), Wiley Online Library (Wiley Online Library, 2023), ScienceDirect (ScienceDirect, 2023), and SpringerLink (SpringerLink, 2023).

To mitigate false-positive results, our search was confined to the titles, abstracts, and keywords of the articles. When necessary, we adjusted the search query to align it with the requirements of each search engine. The search command employed was: **((“dental radiography” OR “dentistry”) AND (“Deep CNN based”) AND (“segmentation”))**. A manual search was conducted separately in each database, and the results were amalgamated into a comprehensive spreadsheet (see Table 2).

Using the outlined search strategy, we identified 642 relevant papers. The distribution across the databases is as follows: 81 from ACM Digital Library, 33 from IEEE Xplore Digital Library, 269 from ScienceDirect Elsevier, 178 from SpringerLink, and 81 from Wiley Online Library. After eliminating duplicate articles, 45 papers were included and excluded (as detailed in Section 4.2).

4.2.3. Selection of studies

The study selection for our SLR involved the application of predefined inclusion and exclusion criteria to the collected data. The inclusion criteria encompassed papers published within a 10-year timeframe (2014–2023), written in English, and specifically addressing

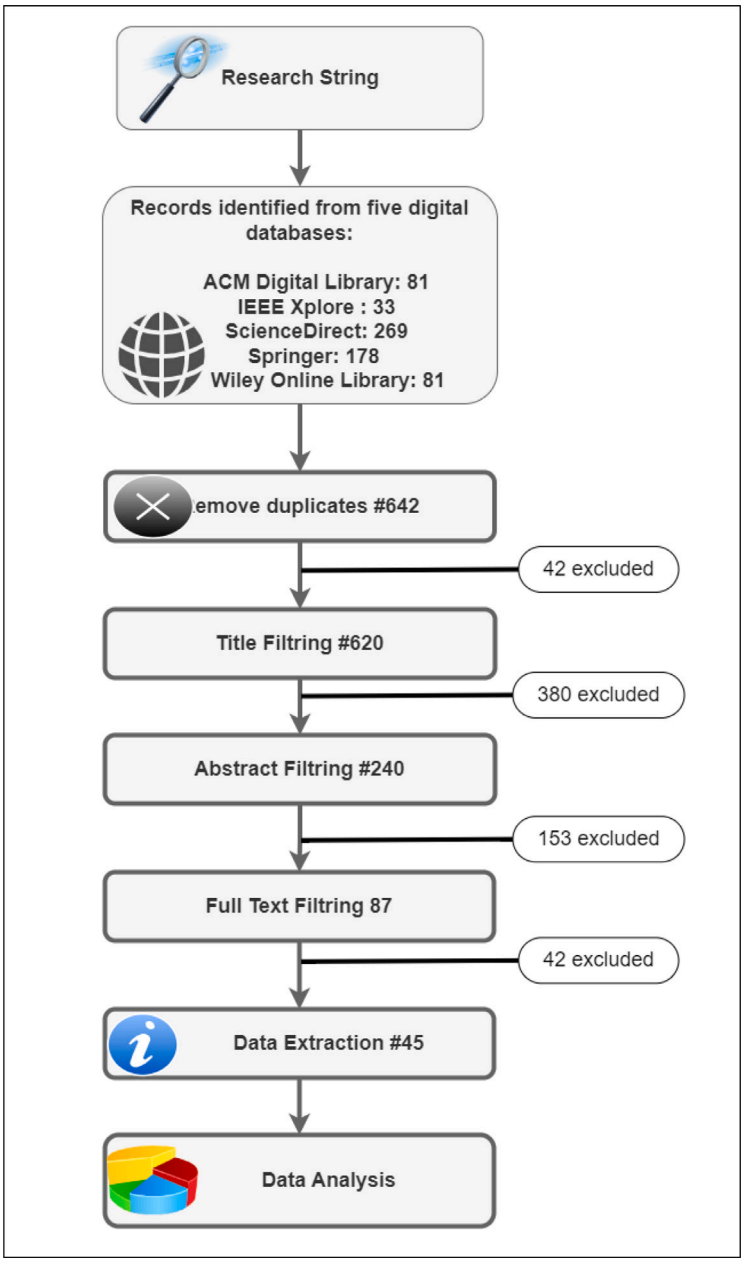


Fig. 3. The systematic search and selection process.

Table 3
Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Ten years (2014–2023)	More than ten years (2013 and less)
English language	Other languages
Must include dental segmentation	Not including dental segmentation
Must include Deep CNN	Not including Deep CNN
Review articles	Other types of scholarly publication

plant disease detection using deep CNNs. Papers failing to meet these criteria, such as those published more than ten years ago, written in languages other than English, or not primarily focused on plant pathology detection with deep CNN, were excluded. The application of these criteria facilitated the identification of relevant studies for subsequent analysis and synthesis (see Table 3).

The selection process for relevant studies was unfolded in three stages. Initially, screening was conducted based on the titles of the papers, with each title assessed for relevance to the study according to the predefined inclusion/exclusion criteria. Subsequently, screening was extended to the abstracts of the papers, where those meeting the inclusion criteria were included, and those not meeting the criteria were excluded. Finally, the same process was repeated using the full text of the papers. In instances where full texts were not accessible through online databases, efforts were made to contact the authors or locate texts from alternative sources. If full text could not be obtained, it was excluded from the study.

5. Aims and research questions

The formulation of research questions can be organized based on four elements: Population, Intervention, Comparison, Outcomes, and Context (PICOC). Population refers to the targeted research group. Intervention is a detailed aspect of research or issues that attracts researchers. A comparison is defined as the aspect of research in which the intervention will be compared. Outcomes are the results of the

Table 4
Research Questions.

R_Q Id	Research question	Motivations
RQ1	What is the status of this field of study?	Presents a chronological summary of the studies selected for this SLR and shows the trends of studies in this area in recent years.
RQ2	What was the key motivation for applying DL to dental image analysis?	Analyze different problems addressed by DL algorithms in the field of dental imaging.
RQ3	What DL algorithms have been applied?	To identify widely utilized DL models in dentistry, examining their applications, strengths, and potential contributions to diagnostics, treatment planning, and overall patient care.
RQ4	Which evaluation metrics are most frequently used to assess deep Convolutional Neural Network segmentation models?	Display the key ML or DL performance measures to select the most appropriate ones for the use case.
RQ5	How can future studies build on the findings of the reviewed papers to contribute novel insights or address existing limitations in the field?	Assess the present state of the literature and recognize any deficiencies or constraints in the current study. Acknowledging these gaps will create a platform for future studies to address these challenges.
RQ6	Which data sources were used?	To train ML or DL models and to offer a baseline for evaluating the efficacy of the suggested architectures, numerous public datasets are presented.

intervention, and the context is the environment of the research (Booth et al., 2021). Before starting the search, we determined the research questions and created a search string. According to Table 4, six research questions are proposed for this study.

6. Results

In this Section, we expound upon the overall statistical findings derived from the encompassed primary investigations and subsequently elucidate the outcomes pertinent to the research inquiries within the subsequent Sections. The 45 primary studies incorporated into this Systematic Literature Review are delineated in Table 11.

6.1. The current landscape of dental radiography segmentation

The 45 included studies were in the form of journal articles. We analyzed the affiliations of the authors of the included articles to associate them with their country of origin. The ranking highlights the prominence of Asian countries, with China leading 10 articles, followed by Korea with eight articles, and India with four articles. The following are Turkey with five articles and the USA with five and three articles, respectively. The list also features European countries, including Germany, with three articles, and Spain with two articles. Additionally, other representative countries include Saudi Arabia with three articles, and from the Asian continent, Malaysia, South Korea, and Taiwan, with one article each. South Africa and Tunisia are represented by one article each (see Fig. 4).

Fig. 5 provides an overview of the annual growth rate of publications in this field of study. An increase in the number of publications can indicate a growing interest in the field, whereas a decrease could suggest that the field is a closed research area. Notably, there is a limited representation of papers published in 2023, possibly because of restricted access to the most recent publications at the time of the study, indicating that the actual number of publications for that year may be higher. We observed an overall upward trend in publications, implying that, although the field was a relatively unpopular research topic with no publications between 2014 and 2017, it has recently attracted considerable attention from researchers.

6.2. Purposes of deep learning in dental image segmentation

CNNs have emerged as powerful tools for image recognition, particularly in the specialized domain of dental image segmentation. Leveraging CNN-based approaches enables rapid and accurate recognition and classification of various diagnostic processes (Jdey et al., 2012), thereby advancing the field of dental image processing.2D and 3D radiographic dental images play a pivotal role in addressing a wide spectrum of dental issues. Table 5 outlines the objectives and diseases addressed by the 45 articles in our survey.

The tooth numbering process assigns a unique number to each tooth by using a universal tooth numbering system. In a dental panoramic image encompassing molars, premolars, canines, and incisors in both the maxilla and mandible, the maximum tooth count reaches 16 (Fig. 2(a)). The distribution included six molars, four premolars, four canines, and eight incisors in each jaw. Researchers have successfully applied these techniques to tasks such as tooth detection and numbering, periapical lesion detection, age and sex determination, caries detection and segmentation, identification of dental structures, and the diagnosis of various dental conditions. These studies highlighted the effectiveness of CNN-based models in achieving high precision and recall rates.

Notably, these applications extend to forensic medicine, aiding age estimation and gender identification based on dental characteristics. Growing interest in this field is evident from the increasing number of publications over the years.

Overall, the integration of DL approaches, as shown in the surveyed studies, holds promise for enhancing the accuracy and efficiency of dental image analysis, thereby contributing to improved diagnostic processes and patient care.

6.3. Applied deep learning algorithms

In the background Section, we have described CNNs, which are highly relevant for DL . Below, we will discuss the most commonly used types of DL networks that have been employed in the reviewed studies, based on the publications we have selected (see Table 6).

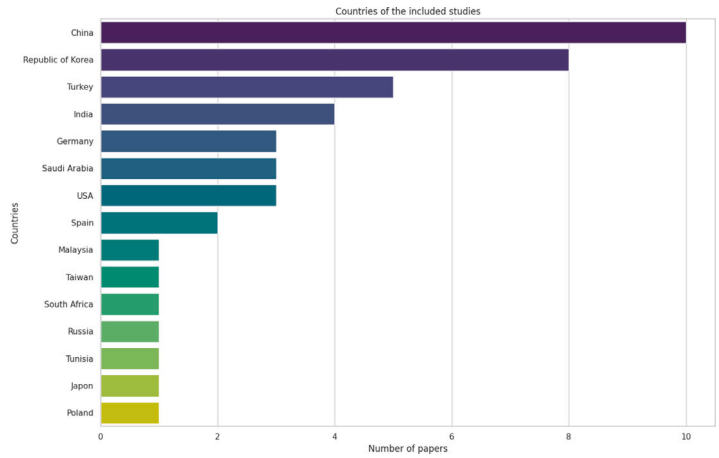


Fig. 4. Countries of the included studies.

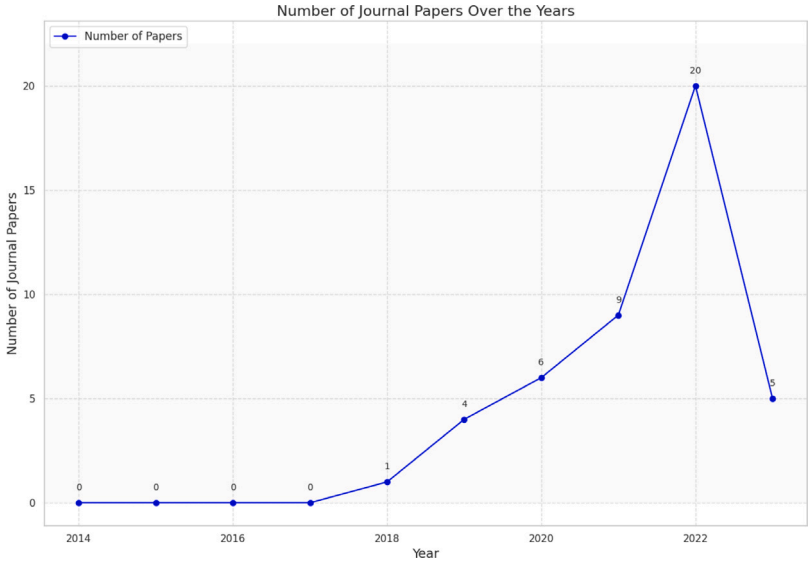


Fig. 5. Yearly papers on CNNs for dental X-ray segmentation.

Table 5
Automatic dental imaging segmentation purpose.

Basic detection and identification	
Tooth detection and numbering	Chen et al. (2019), Bilgir et al. (2021), Singh and Sehgal (2020), Kabir et al. (2022), Brahmi and Jdey (2023)
Detection of Periapical Lesions and segmentation of cysts	Moidu et al. (2022), Sivasundaram and Pandian (2021), Krois et al. (2021), Setzer et al. (2020), Ekert et al. (2019)
Structural Analysis and Classification	
Age and Sex determination	Franco et al. (2022), Mohammad et al. (2022), Vila-Blanco et al. (2022), Rajee and Mythili (2021)
Caries detection and segmentation	Bayraktar and Ayan (2022), Lee et al. (2021), Cantu et al. (2020), Lee et al. (2018), Zhang et al. (2022), Zhu et al. (2022), Park et al. (2022), Imak et al. (2022)
Dental calculus identification	You et al. (2020), Hsiao et al. (2021)
Detection and classification of dental structures	Jang et al. (2022), Aljabri et al. (2022), Chen et al. (2021b), Liu et al. (2022), Zhao et al. (2020), Kwak et al. (2020), Tian et al. (2019), Bonfanti-Gris et al. (2022), Oztekin et al. (2022), Kohlakala et al. (2022)
Advanced Condition Detection	
Detecting dental conditions	Jiang et al. (2022), Alotaibi et al. (2022), Başaran et al. (2022), Lee et al. (2022), Chen et al. (2021a), Al Kheraif et al. (2019), Park and Lee (2023)

Table 6
Summary of Pre-Trained Models.

Model	Year	Architecture	Main contribution	Limitations
LeNet (LeCun et al., 1998)	1998	Two convolutional (conv) layers for feature extraction. Two subsampling layers were used for the spatial dimension reduction. Two fully connected layers for global understanding of input. An output layer featuring a Gaussian connection suggests the use of either a Gaussian activation function or Gaussian-weighted connections.	Recognition of handwritten digits.	• Inadequate adaptation to a variety of image classes. • Utilization of oversized filters • Limited capability in low-level feature extraction
Visual geometry group (VGG) (Simonyan and Zisserman, 2014)	2014	Multiple layers of small 3 × 3 convolutional filters followed by max Pool_ Layers. The fully connected layers used ReLU activation and dropout regularization to prevent overfitting.	An increase in the depth of a network (number of layers) is a critical factor in improving performance.	• Incorporation of computationally demanding fully connected layers.
GoogLeNet (Inception V1) (Szegedy et al., 2015)	2015	Introduced the inception block in CNN. This block uses filters of different sizes to capture spatial information at various scales, aiming for high accuracy with reduced computational cost. It replaces the traditional convolutional layers with these blocks by incorporating a 1 × 1 convolutional filter bottleneck for computation regulation. Sparse connections and global average pooling reduce the redundancy and cutting parameters from 138 million to 4 million.	Development of an inception module that significantly reduces the network's parameter count, leading to improved computational efficiency.	• Cumbersome parameter customization owing to a heterogeneous topology • Potential loss of valuable information due to a representational bottleneck
Inception V3 (Szegedy et al., 2016)	2015	This architecture incorporates factorized 7 × 7 convolutions, spatial factorization using asymmetric convolutions, and an auxiliary classifier for label information propagation. 0In addition, batch normalization and ReLU activation functions were applied. With around 23 million parameters	Addresses the representational bottleneck by replacing large filters with smaller ones.	• Elaborate architecture design • Absence of homogeneity
Faster R-CNN (Ren et al., 2015)	2015	A single-stage model with end-to-end training, utilizing a novel Region Proposal Network (RPN) for efficient region proposal generation. It features key elements, such as the fully convolutional RPN, which generates proposals at a low computational cost, and the sharing of convolutional features between the RPN and the detection network. Training involves stochastic gradient descent (SGD) optimization for convolution layers, RPN weights, and the last fully connected layer weights.	Region proposal network (RPN)	• Complexity • Training Time • Detection Time • Resource Intensiveness • Model Complexity
U-Net (Ronneberger et al., 2015)	2015	A U-shaped encoder–decoder network architecture, consisting of four encoder blocks and four decoder blocks connected via a bridge. Each encoder block involved two 3 × 3 convolutions, followed by batch normalization and a ReLU activation function. The decoder network employs transposed convolutional layers for upsampling in order to restore the original image size. The inclusion of skip connections between encoder and decoder blocks facilitates the recovery of spatial information lost during downsampling.	Skip connections.	•Sensitivity to Input Size • Hardware Memory Requirements
ResNet (Residual Network) (He et al., 2016)	2015	ResNet uses residual blocks that are stacked together to form an overall network. The original ResNet had 34 layers with 2-layer blocks, whereas advanced variants, such as ResNet50, used 3-layer bottleneck blocks to improve accuracy and reduce training time. Skip connections in ResNet add outputs from previous layers to the stacked layers, enabling the training of much deeper networks than previously possible.	Residual layer.	• Complexity in architecture • Degradation of feature-map information during feedforward • Over-adaptation of hyperparameters for specific tasks due to the repeated stacking of identical modules • Numerous layers may contribute minimal or no information. • Relearning of redundant feature maps may occur.

(*continued on next page*)

Table 6 (*continued*).

MobileNet (Howard et al., 2017)	2017	The architecture comprises convolution layers, depthwise convolution layers with BN and ReLU, pointwise convolution layers with BN and ReLU, global average pooling, reshaping, dropout, additional convolutional layers, softmax, and reshaping. With approximately four million parameters, MobileNet was designed to be lightweight. It features two basic units: 3 × 3 convolution and 3 × 3 depth-wise convolution followed by 1 × 1 convolution, using separable filters to reduce the number of input channels.	MobileNet's architecture is optimized for mobile and embedded devices by utilizing depth-wise separable convolutions. The sequential layers include depthwise and pointwise convolutions with Batch Normalization and ReLU.	• Challenges arising from resource-constrained microcontrollers. • Sacrifices in accuracy compared to larger architectures, especially in detailed tasks. • Inherent tradeoff in accuracy due to its small size, though often minimal.
DenseNet (Huang et al., 2017)	2017	DenseNet's architecture utilizes Dense Blocks with densely connected layers to maximize the feature reuse. Each block includes an input, output, and growth rate (k) that regulate the information flow to the next layer. Transition Layers connect dense blocks using 1 × 1 convolutions, 1 × 1 pooling, and 3 × 3 convolutions, striking a balance between model complexity and information flow. The Growth Rate (k) governs the number of additional features in each dense block, impacting the information flow and network generalization. Finally, the Output Layer is typically a fully connected layer 000that maps000 the input features to the desired class labels.	Densely connected architecture	• Memory Usage
You Only Look Once (YOLO) (Redmon et al., 2016)	2016	YOLO uses a fully convolutional neural network that passes an image through a series of convolutional layers to extract features. It predicts object boundaries and class labels simultaneously, making it a powerful and efficient DL model for object-detection tasks.	Real-time object detection	• Detecting small or complex-shaped objects and accurately handling very large objects
Mask R-CNN (He et al., 2017)	2017	Incorporates a backbone network for feature extraction, a Region Proposal Network (RPN) for candidate region proposals, and a parallel branch for predicting segmentation masks alongside bounding box coordinates and class probabilities. This comprehensive framework excels in instance segmentation tasks and provides pixel-level masks for accurate object delineation.	RoIAlign layer.	• False alerts • Missing labels

Table 7
Confusion Matrix.

	Predicted class 0	Predicted class 1
Actual Class 0	True Negative (TN)	False Positive (FP)
Actual Class 1	False Negative (FN)	True Positive (TP)

6.4. Evaluation metrics for deep CNN segmentation in dental imaging

Tools or mechanisms for measuring model quality are known as evaluation metrics and are crucial components for developing effective ML or DL models. These metrics vary based on the tasks, applications, and models (Awassa et al., 2022; Egger et al., 2022). In the context of dental image segmentation, we outlined the evaluation metrics employed in the selected studies.

Table 8 presents the various performance measurement formulas. For a comprehensive overview of the commonly used evaluation measures along with their interpretation and implementation, refer to Table 7. These metrics are derived from the attributes in the Confusion Matrix (Table 8), which is a two-dimensional matrix that provides information about Actual and Predicted classes. Key attributes include:

- **TP** True-Positive is the number of Positive cases classified correctly.
- **TN** True-Negative is the number of Negative cases classified correctly.
- **FP** False-Positive is the number of cases where the model predicted Positive class, but the actual class was Negative. Also known as Type 1 Error.
- **FN** False-Negative is the number of cases where the model predicted Negative class, but the actual class was Positive. Also known as Type 2 Error.

To evaluate dental image segmentation models, it is important to note that in addition to the metrics discussed in this Section, some studies have utilized specific measures such as Volumetric Overlap Error (VOE) and Relative Volume Difference (RVD) (Chun et al., 2023). These measures provide a thorough assessment of model performance, especially when applied to volumetric tomographic images, such as Cone Beam Computed Tomography (CBCT). Volumetric Overlap Error (VOE) measures the degree of mismatch between segmented and actual regions, offering insights into the accuracy of segmentation in terms of spatial localization. Relative Volume Difference (RVD) quantifies volume differences between segmented regions and true anatomical volumes, providing an assessment of accuracy in terms of quantity. Thus, by considering these additional metrics, researchers can obtain a more comprehensive evaluation of the performance of dental image segmentation models, particularly when applied to volumetric tomographic images, such as CBCT scans Fig. 6).

Specificity and Sensitivity

Specificity and sensitivity are standardized and recognized measures for evaluating the performance of medical-image segmentation. Sensitivity, also known as recall or true-positive rate, focuses on the capability of detecting true positives in pixel classification, whereas

Table 8
Performance Evaluation Metrics.

Metric	Formula	Description	Study
Accuracy (Pixel Accuracy (PA))	$\frac{Number of Correct Predictions(TP+TN)}{Total Number of Predictions(TP+TN+FP+FN)} = \frac{(Number of correctly classified pixels)}{(Total number of pixels)} * 100\%$	Determines how frequently an algorithm correctly classifies a data point. This is the number of items correctly identified as true positives or true negatives of the total number of items.	Bayraktar and Ayan (2022), Cantu et al. (2020), Lee et al. (2018), Moidu et al. (2022), Jiang et al. (2022), Alotaibi et al. (2022), Aljabri et al. (2022), Chen et al. (2021b), Singh and Sehgal (2020), Kabir et al. (2022), Sivasundaram and Pandian (2021), Franco et al. (2022), Mohammad et al. (2022), Vila-Blanco et al. (2022), Rajee and Mythili (2021), Zhu et al. (2022), Park et al. (2022), Imak et al. (2022), Hsiao et al. (2021), Liu et al. (2022), Zhao et al. (2020), Kwak et al. (2020), Tian et al. (2019), Oztekin et al. (2022), Kohlakala et al. (2022), Lee et al. (2022), Al Kheraif et al. (2019), Choi et al. (2023), Chun et al. (2023), Amasya et al. (2024)
Precision	$\frac{TP}{TP+FP}$	Agreement between true class labels and machine predictions. It is calculated by adding all true positives and false positives in the system across all classes.	Lee et al. (2021), Jang et al. (2022), Chen et al. (2019), Bilgir et al. (2021), Jiang et al. (2022), Alotaibi et al. (2022), Aljabri et al. (2022), Chen et al. (2021b), Franco et al. (2022), Rajee and Mythili (2021), Zhu et al. (2022), Park et al. (2022), Imak et al. (2022), Zhao et al. (2020), Kohlakala et al. (2022), Başaran et al. (2022), Al Kheraif et al. (2019), Brahmi and Jdey (2023), Choi et al. (2023), Slimani et al. (2023), Chun et al. (2023), Ali et al. (2023)
Recall [(Sensitivity) (True Positive Rate TPR)]	$\frac{TP}{TP+FN}$	Ability of a classifier to identify class labels. It is calculated by adding all true positives and false negatives in the system across all classes.	Bayraktar and Ayan (2022), Lee et al. (2021), Cantu et al. (2020), Lee et al. (2018), Chen et al. (2019), Bilgir et al. (2021), Moidu et al. (2022), Jiang et al. (2022), Alotaibi et al. (2022), Aljabri et al. (2022), Chen et al. (2021b), Singh and Sehgal (2020), Kabir et al. (2022), Sivasundaram and Pandian (2021), Krois et al. (2021), Ekert et al. (2019), Franco et al. (2022), Rajee and Mythili (2021), Zhu et al. (2022), Park et al. (2022), Imak et al. (2022), Liu et al. (2022), Zhao et al. (2020), Bonfanti-Gris et al. (2022), Kohlakala et al. (2022), Slimani. et al. (2024), Al Kheraif et al. (2019), Choi et al. (2023), Heni et al. (2022), Chun et al. (2023), Bouzidi et al.
Specifcity	$\frac{TN}{FP+TN}$	is referred to as the true-negative rate. This function computes the proportion of actual negative cases predicted as negative by "the" model.	Bayraktar and Ayan (2022), Cantu et al. (2020), Lee et al. (2018), Jang et al. (2022), Moidu et al. (2022), Alotaibi et al. (2022), Aljabri et al. (2022), Singh and Sehgal (2020), Kabir et al. (2022), Sivasundaram and Pandian (2021), Krois et al. (2021), Ekert et al. (2019), Franco et al. (2022), Rajee and Mythili (2021), Imak et al. (2022), Liu et al. (2022), Zhao et al. (2020), Bonfanti-Gris et al. (2022), Al Kheraif et al. (2019), Chun et al. (2023)
F1-score	$2 * \frac{Precision * Recall}{Precision + Recall}$	It measures the effectiveness of identification when recall and precision are equally important.	Lee et al. (2021), Cantu et al. (2020), Jang et al. (2022), Bilgir et al. (2021), Moidu et al. (2022), Jiang et al. (2022), Alotaibi et al. (2022), Aljabri et al. (2022), Chen et al. (2021b), Sivasundaram and Pandian (2021), Krois et al. (2021), Zhu et al. (2022), Kohlakala et al. (2022), Başaran et al. (2022), Al Kheraif et al. (2019), Brahmi and Jdey (2023), Choi et al. (2023), Amasya et al. (2024)
Cohen's Kappa	$\frac{P_0 - P_e}{1 - P_e}$ where P_0 is the observed relative agreement between raters and P_e is the hypothetical probability of chance agreement.	Represents the level of precision and reliability in a statistical classification and assesses the degree of agreement between two raters (judges) who each assign items to mutually exclusive categories.	Alotaibi et al. (2022), Mohammad et al. (2022), Choi et al. (2023), Amasya et al. (2024)
Dice similarity coefficient (DSC)	$\frac{(2 * area of overlap)}{(2 * area of overlap)(Total number of pixels in both images),}$	It is used to assess how well a predicted segmentation matches the corresponding ground truth in terms of the pixel-level agreement..	Setzer et al. (2020), Moidu et al. (2022), Kabir et al. (2022), Sivasundaram and Pandian (2021), Rajee and Mythili (2021), Zhu et al. (2022), Liu et al. (2022), Zhao et al. (2020), Lee et al. (2022)
Receiver operating characteristic (ROC)	Does not have a specific formula	The ROC curve is a graphical display of sensitivity (TPR) on the y-axis and (1 – specificity) (FPR) on the x-axis for varying cut-off points of the test values.	Lee et al. (2018), Zhang et al. (2022), Bonfanti-Gris et al. (2022), Chun et al. (2023)

(*continued on next page*)

Table 8 (*continued*).

Metric	Formula	Description	Study
Area under the curve (AUC) Average precision (AP))	$AUC = 1 - \frac{1}{2}(\frac{FP}{FP+TN} + \frac{FN}{FN+TP})$ AP $= \int_{r=0}^{1} p(r)\,dr$ mean Average precision (mAP) $mAP = \frac{1}{N}\sum_{k=1}^{K=n} AP_k$	As a single scalar value, the AUC measures the performance of a binary classifier. It lies between [0.5–1.0], where the minimum value represents random classification, and the maximum value represents perfect classification.	Bayraktar and Ayan (2022), Lee et al. (2018), Chen et al. (2019, 2021b), Ekert et al. (2019), Vila-Blanco et al. (2022), Park et al. (2022), Hsiao et al. (2021), Bonfanti-Gris et al. (2022), Chen et al. (2021a), Brahmi and Jdey (2023), Chun et al. (2023), Ali et al. (2023)
Intersection over Union (IoU) [Jaccard index (JI)]	$\frac{Area_{pred} \bigcap Area_{gt}}{Area_{pred} \bigcup Area_{gt}} = \frac{TP}{TP+FP+FN}$	It is used to evaluate the performance of object detection by comparing the ground truth bounding box to the predicted bounding box.	You et al. (2020), Jang et al. (2022), Chen et al. (2019), Moidu et al. (2022), Kabir et al. (2022), Sivasundaram and Pandian (2021), Kwak et al. (2020), Oztekin et al. (2022), Lee et al. (2022), Chen et al. (2021a), Park and Lee (2023)
Matthews correlation coefficient (MCC)	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}}$	The MCC is used for binary classification and is considered particularly useful in unbalanced class situations. takes values between 1 (perfect correlation between the ground truth and predicted outcome) and −1 (inverse or negative correlation); a value of 0 denotes a random prediction.	Cantu et al. (2020), Moidu et al. (2022), Alotaibi et al. (2022)

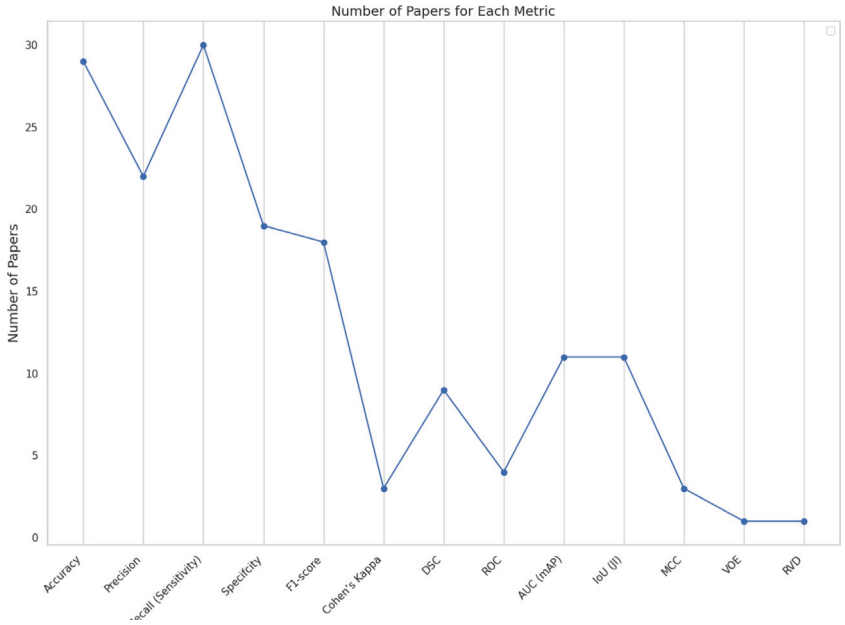


Fig. 6. Summary of performance evaluation measures reviewed in SLR.

specificity, also known as the true-negative rate, evaluates the capabilities of correctly identifying true-negative classes (such as the background class). The specificity in image segmentation tasks demonstrates the ability of the model to recognize the background class in an image. Specificity ratios near 1 are expected because of the large percentage of pixels labeled as background compared with the region of interest (ROI). Therefore, specificity is a good statistic for checking the model's operation, but not for its performance.

Accuracy/Rand Index (Müller et al., 2022)

Accuracy, also known as the Rand Index or Pixel Accuracy, is defined as the number of correct predictions including positive and negative predictions divided by the total number of predictions. However, in the case of unbalanced data, the accuracy may not be the best measure of performance (Alotaibi et al., 2022; Ekert et al., 2019). Measuring accuracy will always produce an unrealistic result owing to the inclusion of true-negatives. Even when predicting the segmentation of an entire image as a background class, the accuracy rates are often greater than or close to 100

F-measure, IoU, and DSC

The F-measure, often known as the F-score, is one of the most extensively used metrics for quantifying the performance in computer vision. It is derived from the sensitivity and precision of a prediction, which analyzes the overlap between the expected segmentation and ground truth. However, by considering accuracy, it also penalizes false positives, which are prevalent in datasets with a significant class imbalance. There are two prominent metrics based on the F-measure: intersection over union (IoU), often known as the Jaccard index or Jaccard similarity coefficient, and the Dice similarity coefficient (DSC). DSC is referred to as the harmonic mean of the sensitivity and precision. The IoU penalizes under- and over-segmentation more than the DSC. Although these scores are suitable metrics and the most common metric in validating medical image segmentation (Liu et al., 2022), they are not employed for evaluating the performance of the suggested models in the majority of the articles analyzed in our SLR.

ROC and AUC

The ROC curve, short for Receiver Operating Characteristic, is a line plot that visualizes a classifier's performance with different discriminating thresholds. Performance was measured by comparing the true

positive rate (TPR) to the false positive rate (FPR). In the medical field, ROC curves are widely accepted as standard metrics for comparing several classifiers when assessing diagnostic tests. Hanley and McNeil (1982) proposed the area under the ROC curve (AUC) as a single-value performance assessment tool for diagnostic radiology. The AUC measure has become a popular tool for validating ML classifiers ranges from 0 to 1. A model with 100% incorrect predictions had an AUC of 0.0, whereas that with 100% correct predictions had an AUC of 1.0.

Other metrics

For a segmentation system to be helpful and make a significant contribution to the field, its performance must be systematically tested. Furthermore, the evaluation must be performed using conventional and well-known measures that allow for fair comparison (Jdey, 2022). Other metrics exist and can be used based on the research topic and interpretive emphasis of the study. We refer the reader to the excellent studies by Taha and Hanbury (2015). Additionally, Wang et al. (2020) described the application of metrics in the evaluation of natural, medical, and remote sensing images, and presented supervised and unsupervised evaluation methods in the application.

6.5. Public datasets for dental radiography analysis

Dental research heavily relies on diverse datasets to achieve optimal performance. In addressing the last Research Question (RQ6), our examination, summarized in Table 11, highlights that most studies predominantly used private datasets, with only three studies opting for public datasets. Panoramic radiography is the most commonly used type of dental imaging, providing a comprehensive view of both the upper and lower jaws on a single film. This makes them essential for orthodontic treatment planning and monitoring tooth development. Furthermore, several studies have utilized CBCT to provide detailed 3D visualization. Meanwhile, classic modalities, such as periapical and bitewing radiographs, continue to offer critical diagnostic information about numerous dental diseases. Our systematic literature review underscores a significant dearth of publicly available datasets suitable for extracting relevant features related to oral medical conditions. Three datasets have been identified in the public domain (Silva et al., 2018; Abdi and Kasaei, 2020a; Walid and Imen, 2023). Recognizing the crucial impact of data on shaping the efficacy of ML and DL models, delving into publicly accessible datasets presents a significant potential to bolster the resilience and applicability of DL models in dentistry. Given this shortcoming, we aimed to provide comprehensive insights into specific publicly available reference datasets. Table 9 delineates the characteristics of public datasets for diagnosing dental problems.

1. **MLUA dataset** (Wang et al., 2023)**:** made up of 100 panoramic radiographs, each measuring 2943 × 1435 pixels, and accompanied by PNG masks. This dataset, which is available on GitHub (Wang et al., 2024), includes full-size X-ray images along with segmentation caries masks specifically designed for the segmentation of dental caries in oral panoramic radiographs. As a valuable source for research on accurate caries identification, the repository provides test samples and implementable codes.
2. **The Tufts Multimodal Panoramic X-ray Dataset** (Panetta et al., 2021b)**:** originating from Tufts University and available on Kaggle (Panetta et al., 2024), comprises 1000 labeled panoramic dental radiography images. These images, saved in TIFF/JPEG format with dimensions of 840 × 1615 pixels, were categorized into periapical, odontogenic, pericoronal, inter-radicular, and normal cases. The dataset included radiographs, labeling masks, gray and quantized maps generated with eye-tracking software, tooth masks with labels, and a region-of-interest mask outlining the maxillomandibular region. This dataset is valuable for refining tooth segmentation algorithms and incorporating radiologist expertise into precise diagnosis systems.
3. **The Dental Radiography Analysis and Diagnosis Dataset:** accessible on Kaggle (Momeni, 2024), comprises 1269 labeled panoramic radiographs, with each image measuring 512 *times* 256 pixels. This dataset plays a crucial role in the classification of dental conditions, encompassing categories such as cavities, fillings, impacted teeth, and implants. Its significance is underscored by its substantial contribution to the advancement and refinement of tooth-classification algorithms.
4. **The Teeth Segmentation on Dental X-ray Images** (Humans in the Loop, 2024a)**:** It is an annotation conducted by a team of 12 trainees in the Democratic Republic of Congo. This annotation process focused on the "Panoramic Radiography Database", which encompasses 598 images. Originally published by Román et al. (2021), this database is specifically tailored for dental segmentation within radiological scans. The images are in JPEG format, with dimensions of 2041 x 1024 pixels, and were captured using the Owandy I-max Touch panoramic radiography equipment at the Departamento de Radiología de la Facultad de Odontología, Universidad Nacional de Asunción, in Asuncion, Paraguay. This valuable resource plays a significant role in advancing dental segmentation techniques, and serves as a beneficial tool for research in dental radiology.
5. **The Panoramic Dental X-ray With Segmented Mandibles Dataset** (Abdi and Kasaei, 2020b)**:** comprises 2000 BMP format images, each with dimensions of 2900 × 1250 pixels. This dataset features anonymized and de-identified panoramic dental X-rays obtained from 116 patients at the Noor Medical Imaging Center in Qom, Iran, using the Soredex Cranex D digital X-ray unit. The subjects in the dataset encompassed a diverse range of dental conditions including healthy, partial, and completely edentulous cases. To enhance versatility, derived subsets were created, and images were categorized based on features such as the vertical distance between the alveolar process, ramus width, gonial angle acuteness, and inferior mandible characteristics.
6. **The DENTEX dataset** (Hamamci et al., 2023)**:** is a meticulously curated compilation of panoramic dental X-rays sourced from diverse clinical settings, ensuring authenticity in image quality. Prioritizing patient confidentiality, the dataset is accessible on the Zenodo open-access archival deposition platform at Hamamci (2024) and is hierarchically organized into three main categories: is hierarchically organized into three main categories:

   - Quadrant Detection (693 X-rays).
   - Tooth Detection and Enumeration (634 X-rays).
   - Fully Annotated Abnormal Tooth Detection (1005 X-rays), featuring diagnosis classes such as caries, deep caries, periapical lesions, and impacted teeth.

   Additionally, an extra set of 1571 unlabeled X-rays was provided for pretraining purposes. This dataset serves as a valuable resource for advancing research on dental image analysis and the application of artificial intelligence in dentistry.

Following an exploration of the most common public datasets, it is imperative to introduce another noteworthy public dataset source. The Roboflow Universe (Humans in the Loop, 2024b) is a pivotal resource that consolidates a compilation of open-source datasets and Application Programming Interfaces (APIs) meticulously crafted for applications in computer vision. This contribution enriches academic discourse by providing researchers with a diverse and accessible set of resources tailored to advance the field of computer vision.

6.6. How can future studies build upon the findings of the reviewed papers to contribute novel insights or address existing limitations in the field?

This Section explores the gaps revealed through the analysis of fundamental studies in our systematic literature review, specifically

Table 9
Summary of Characteristics of Public Datasets.

Dataset (Ref)	Size, modality and format	Disease category	Dataset challenges
MLUA dataset (Wang et al., 2024)	100 panoramic radiographs with their labeling masks in the PNG format, each measuring 2943 × 1435 pixels.	Dental caries.	Accurate segmentation of dental caries in oral panoramic medical images.
Tufts Multimodal Panoramic X-ray Datase (Panetta et al., 2024)	1000 panoramic radiographs with their labeling masks in TIFF/JPEG format, each measuring 840 × 1615 pixels.	Tooth abnormalities.	Image enhancement, tooth segmentation, and abnormality detection.
Dental Radiography Analysis and Diagnosis Dataset (Momeni, 2024)	1269 labeled panoramic radiograph in JPG format s, each image measuring 512 × 256 pixels	Tooth classification	Dental classification algorithms
Teeth Segmentation on Dental X-ray Images (Humans in the Loop, 2024a)	598 panoramic radiographs are provided in JPG format, along with their corresponding labeling masks in PNG format, each image measuring 2041 × 1024 pixels.	No disease categories were identified.	Tooth numbering.
Panoramic Dental X-ray With Segmented Mandibles Dataset (Abdi and Kasaei, 2020b)	2000 Panoramic BMP format images, each with dimensions of 2900 × 1250 pixels.	Mandible lesions	Mandible Segmentation for the Identification of Anatomical Structures.
The DENTEX dataset (Hamamci, 2024)	Panoramic radiographs are available in the PNG format, with each image measuring 2872 × 1504 pixels.	caries, deep caries, periapical lesions, impacted teeth	quadrant detection, tooth detection with quadrant and tooth enumeration classes, and abnormal tooth detection with quadrant, tooth enumeration, and diagnosis classes.

focusing on the segmentation of dental panoramic images. By highlighting the current deficits, our objective was to establish a clear framework for future research in this specific domain. This critical synthesis of gaps will serve as a robust foundation for discussing potential directions in future research, paving the way for innovative contributions to enhance and enrich dental panoramic-image segmentation. Following this, we briefly present the main identified gaps and envisioned avenues for future research. The prevalent gaps identified in the literature review entailed challenges linked to diverse facets. These challenges are described in Table 10.

As shown in Fig. 7,in assessing the landscape of dental image analysis, our investigation revealed critical considerations in the development and deployment of deep learning (DL) models. Notably, the dataset size and source emerged as pivotal factors influencing model development, with seven studies highlighting the challenges related to small dataset sizes. Recommendations include the implementation of data augmentation techniques, transfer learning, and the enrichment of datasets with diverse images from multiple sources.

Image quality is another substantial challenge, as acknowledged in two studies. These studies advocate preprocessing methods to address poor-quality images and improve model robustness to effectively handle variations in image quality. Additionally, three studies addressed the complexities associated with data processing and labeling, emphasizing challenges in manual mask creation and proposing solutions such as automated mask generation and unsupervised learning for region identification.

Recognizing that DL models are not a replacement for clinical data, four studies underscored the importance of integrating model predictions seamlessly with periodontal charting and other clinical data. The specific limitations and factors impacting the development, deployment, and effectiveness of the DL models were thoroughly examined. Challenges include difficulties in accurately identifying the number of teeth, complexities in localizing dental issues, and increased complexity and training time.

This comprehensive evaluation also sheds light on platform-related issues such as image dimensions and memory constraints, with two studies addressing these concerns. Furthermore, the impact of lower radiation doses on AI performance in dental imaging was explored in one study, whereas another study investigated the potential of automated image analysis based on 3D imaging for lesion detection. Collectively, these findings contribute to a nuanced understanding of the challenges and opportunities in the application of DL models in dental image analysis.

7. Limitations of the study

Numerous techniques based on DL have been identified for dental image segmentation. However, it is essential to acknowledge the limitations of the SLR. These limitations are described below.

1. Only journal articles discussing dental image segmentation were included in this SLR. Our search technique led to the discovery and subsequent exclusion of several unrelated publications during the preliminary stages of the study. This stringent approach ensured that the selected research papers aligned with the requirements of the investigation. However, the inclusion of additional sources, such as books or conference papers, could further enrich the review.
2. This review is confined to papers available in the English language.
3. There is a possibility that other digital libraries containing relevant studies may have been overlooked, even though digital databases were thoroughly considered during the review of the study articles.

Table 10
Challenges, Gaps and Future Work.

Aspects	Limitations and studies	Improving solutions and charting future paths
Dataset Size and source	The dataset size is pivotal in DL model development because it directly impacts the model's capacity to learn and generalize. A larger dataset provides more diverse examples, enabling the model to capture a broader range of patterns and variations in real-world scenarios. – The dataset used was relatively small (Bayraktar and Ayan, 2022; Moidu et al., 2022; Aljabri et al., 2022; Lee et al., 2021; Bilgir et al., 2021; You et al., 2020; Tian et al., 2019). – To assess tooth-level or case-level Residual Bone Loss (RBL) for a large cohort, the model requires more images for further optimization. Lee et al. (2022) – The dataset is collected from a single organization or single device, limiting its diversity (Zhang et al., 2022; Oztekin et al., 2022; Liu et al., 2022; Moidu et al., 2022; Jiang et al., 2022; Bonfanti-Gris et al., 2022; Krois et al., 2021; Başaran et al., 2022; Lee et al., 2021). – Absence of external test group (Başaran et al., 2022; Cantu et al., 2020). – Heterogeneity in Image Sources (Amasya et al., 2024) or Distribution of Heterogeneous Images (Alotaibi et al., 2022). – The data collected for the study were unbalanced due to high labeling costs. Park and Lee (2023)	–Employing data-augmentation methods (Bayraktar and Ayan, 2022; Mohammad et al., 2022). – Exploring Fine-tuning with transfer learning (Jang et al., 2022; Setzer et al., 2020). – Enriching the dataset with diverse data from multiple global sites and from various devices (Zhang et al., 2022; Oztekin et al., 2022). – Exploring the integration of non-imagery data, such as patient history, clinical signs, and symptoms, with imagery data within a DL system (Moidu et al., 2022; Alotaibi et al., 2022; Jiang et al., 2022; Moulahi et al., 2023). – Seek external validation of the neural network on diverse test sets to assess generalizability. (Amasya et al., 2024) showcased DiagnoCat AI's adaptability across diverse image sources and configurations, independent of specific devices or imaging parameters https://diagnocat.com.
Image Quality	The defined by attributes such as clarity, sharpness, and minimal noise. – Poor-quality images, such as those with overlapping teeth or distorted tooth length, can mislead the diagnosis, and the model does not automatically exclude them (Lee et al., 2022; Aljabri et al., 2022). – Capturing various aspects of a tooth and the difficulties associated with internal tooth structure and the interproximal tooth surface in intraoral photography (Park et al., 2022; Kabir et al., 2022; Chen et al., 2021b; Amasya et al., 2024). – Handling Incorrectly Oriented Radiographs (Kabir et al., 2022). – CBCT provides high resolution, but its accuracy in density measurement is inferior to traditional CT (Kwak et al., 2020)	– Explore methods for preprocessing or filtering poor-quality images prior to analysis. – Improve the model's robustness to handle variations in image quality. – Conduct an investigation into a more realistic simulated X-ray acquisition process (Kohlakala et al., 2022).
Data Processing and Labeling	Refines raw data through operations such as resizing and normalization, and optimizing it for effective model training. Simultaneously, data labeling assigns categorical or numerical tags to each data point, providing a crucial ground truth for supervised learning and enhancing the model's ability to generalize and make accurate predictions. – Manual Mask Creation and Pre-processing poses a challenge due to its time-consuming nature and the need for expertise, consequently restricting scalability and potentially introducing errors (Oztekin et al., 2022; Kwak et al., 2020). – In the benchmark dataset, certain categories are labeled more coarsely than others (Zhao et al., 2020). –Complexities in Labeling: especially when premolar teeth are missing without clear spacing. Ali et al. (2023)	– Investigate automated and semi-automated methods for mask generation. Explore unsupervised learning techniques for region identification (Oztekin et al., 2022). – The ambiguity in bounding box boundaries persists in object detection (Jang et al., 2022).
Not a Replacement for Clinical Data	While advantageous, DL models are not designed to replace the vital clinical data and expertise provided by healthcare professionals. – The resulting model has not been used in a clinical setting, and it is not designed to replace periodontal charting and other clinical data. Lee et al. (2022), Jiang et al. (2022), Vila-Blanco et al. (2022), Cantu et al. (2020). – The study acknowledges a low success rate in identifying dental caries and calculus, making the AI models impractical for clinical use (Başaran et al., 2022).	– Explore seamless integration of the model's predictions with periodontal charting and data.

(*continued on next page*)

Table 10 (*continued*).

Specific limitations	Additional factors that impact the development, deployment, and effectiveness of DL models. – The model faces challenges in accurately identifying tooth numbers or positions, particularly when multiple teeth are missing (Lee et al., 2022). – The model faces difficulties in precisely localizing specific dental issues. Zhang et al. (2022), Oztekin et al. (2022) – Increased Complexity and Training Time (Park and Lee, 2023). – The study did not compare the diagnostic performance of CNNs with that of dentists (Bayraktar and Ayan, 2022). – Image Dimensions and platform Memory Constraints (Google Colab) (Alotaibi et al., 2022; Brahmi and Jdey, 2023). – Overfitting (Mohammad et al., 2022). – While capable of estimating age and sex with a single tooth, the models' performance is anticipated to enhance with multiple teeth, potentially limiting applicability in scenarios with few visible or accessible teeth (Vila-Blanco et al., 2022). – Exemplary Use-Case Focus (Krois et al., 2021). – Determining the estimation success of each type of tooth (incisors, canines, premolars, molars) (Bilgir et al., 2021). – The Faster R-CNN network, while generally successful, encountered issues in recognizing two 'half teeth' as a single intact tooth due to the inherent limitations of CNNs not considering spatial relationships between image features (Chen et al., 2019).	Distinguishing implants with different types of internal connections (Kohlakala et al., 2022). – Understanding the trade-offs between image resolution and memory constraints (Alotaibi et al., 2022). – Dropout (Mohammad et al., 2022). – Investigating the impact of lower radiation doses on AI performance in dental imaging (Bilgir et al., 2021). – Automated image analysis based on 3D imaging to assist clinicians in lesion detection and contribute to the differential diagnosis of periapical pathology and nonodontogenic lesions (Setzer et al., 2020).

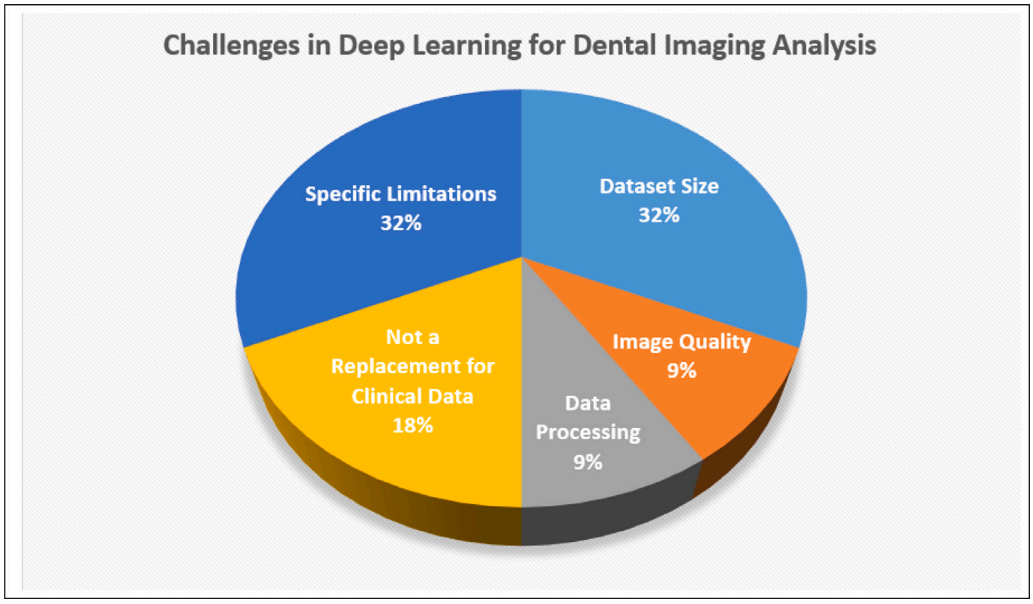


Fig. 7. Challenges in deep learning for dental imaging analysis.

4. The impact of hyperparameters, such as the number of epochs, batch size, and the number of layers, on model performance has not been addressed in this study.

8. Conclusion and future work

In conclusion, the amalgamation of Deep Learning (DL) and oral healthcare analysis has witnessed a substantial surge in research interest over the past decades, culminating in a continuous evolution of innovative models and methodologies. These advancements, characterized by the potential of DL algorithms, empower dentists in various dimensions of clinical practice, from precise tooth identification to the categorization of dental pathologies, assessment of prosthetic structures, and reduction of errors. The promising prospects of DL algorithms extend further to the optimization of clinical workflows, ultimately contributing to enhanced treatment outcomes.

Through a meticulous review of seminal articles focusing on DL methodologies in dental image segmentation, recent publications have been systematically categorized based on specific objectives. An in-depth exploration of the predominant evaluation metrics was conducted, accompanied by the identification and analysis of significant challenges across facets, such as dataset size, image quality, data processing, and labeling. Each challenge was scrutinized along with its limitations, pertinent studies, and proposed solutions, collectively shaping a comprehensive understanding of the field. Anticipating that this study provides readers with nuanced insights, we underscore potential avenues for future research and highlight the paramount importance of the quality and scale of databases in enhancing the efficacy of machine learning models in dentistry. Although three public databases were identified in the reviewed studies, we emphasize the need to expand our resources. A proactive exploration of emerging platforms such as Kaggle, Roboflow, and GitHub is advocated to discover and incorporate new public databases. This strategic approach seeks to enrich the diversity of available data, foster substantial improvements in models, and yield more representative examples of the challenges encountered in real clinical scenarios.

Our upcoming research projects are inspired by the insights gained from this comprehensive Systematic Literature Review (SLR). Our primary objective was to develop an innovative algorithm for dental pathology segmentation that harnesses the power of DL. A particular

Table 11
List of reviewed studies.

Authors	Dataset characteristics	Data Augmentation?	Main objective& Architecture	Task	Performance
Brahmi and Jdey (2023)	**A public dataset** consisting of 107 panoramic x-ray images (Walid and Imen, 2023).	No	The Mask-RCNN network was utilized to perform pixel-level instance segmentation on a dataset1 containing 107 panoramic X-ray images of both young and adult patients. Subsequently, the Mask-RCNN network underwent training on dataset2, which comprises 73 panoramic X-ray images excluding those of young patients.	Instance segmentation and Detection	mAP = 90%, F1-sore = 63% and precision = 96%
Choi et al. (2023)	Private dataset consisting of 402 periapical radiography (PA), 138 Dens evaginatus and 264 normal cases. Divided into training set (N = 282, 70%), validation set (N = 70, 18%) and test set (N = 50, 12%).	Yes	Five DL models in image classification, including a simple CNN model, VGG, DenseNet, ResNet, abd InceptionResNetV2, were selected for the diagnosis of Dens evaginatus (DE).	Classification	F Model means (Full Image) and C Model mans (Cropped Image) AUC: F Model = 0.895, C Model = 0.901 Accuracy: F Model = 0.828, C Model = 0.832 Precision: F Model = 0.869, C Model = 0.856 Recall: F Model = 0.871, C Model = 0.898 F1 Score: F Model = 0.869, C Model = 0.876 Cohen's Kappa: F Model = 0.774, C Model = 0.684
Chun et al. (2023)	A Private dataset included CBCT images from 50 patients, which were divided into a training set (64 volumes, 3546 images) and a test set (36 volumes, 1804 images). Each axial image, cropped to 512 ∗ 512 pixels, served as input for the segmentation network.	No	Development and implementation of a distance-aware network for the automated classification of the three-dimensional (3D) positional relationship between an impacted mandibular third molar (M3) and the inferior alveolar canal (MC) in cone-beam CT (CBCT) images, incorporating the accurate segmentation capabilities provided by the Dense121 U-Net.	Segmentation Classification	Segmentation performance: IoU = 0.872, DSC = 0.920, Precision = 0.946, Recall = 0.918 Classification performance: Accuracy = 1.00, Sensitivity = 1.00, Specificity = 1.00, AUC = 1.00
Amasya et al. (2024)	Private dataset consisting of 6000 panoramic radiographs with 45,161 annotated instances.	No	Two separate models are trained for different tasks. The first model uses Mask R-CNN with a pretrained ResNet-101 backbone to detect teeth, segment their masks, and define their numbering. The second model, based on the Cascade R-CNN architecture, is used for predicting periodontal bone loss.	Segmentation Classification	For tooth conditions: Overall F-score, accuracy, and Cohen's kappa coefficients were found to be 0.948, 0.977, and 0.933 for the binary results. For multiclass results, the corresponding values were 0.992, 0.988, and 0.961. For bone loss detection: Overall F-score, accuracy, and Cohen's kappa coefficients were found to be 0.985, 0.980, and 0.956 for the binary results. For multiclass results, the corresponding values were 0.996, 0.993, and 0.974.%
Ali et al. (2023)	Private dataset, initially comprising 3818 panoramic radiographs, was refined to 3138 images, with Experiment 1 utilizing 2615 for training and 523 for testing, while Experiment 2 employed a six-fold cross-validation on the entire set, essential for evaluating tooth and prosthesis detection models.	Yes	A novel method using two YOLOv7-based object detectors to enhance teeth detection in dental panoramic X-rays, addressing challenges posed by prosthetic treatments. The approach involves detecting tooth and prosthesis candidates, assigning approximate tooth numbers to prosthetic candidates, and optimizing the results, significantly improving detection efficiency.	Detection Classification	Mean Average Precision (mAP) for tooth detection = 0.982. Mean Average Precision (mAP) for prosthesis detection = 0.983.

(*continued on next page*)

Table 11 (*continued*).

Lee et al. (2022)	Private dataset. 693 periapical radiographic images, divided to 70%, 10%, and 20% for training, validation, and testing.	No	The task was to develop and evaluate different variations of the U-Net architecture for the segmentation of three different areas: the bone area, tooth, and cemento-enamel junction (CEJ) line.	Segmentation	bone area DSC = 0.96, JI = 0.93 and PA = 0.96 tooth shape DSC = 0.95, JI = 0.91 and PA = 0.89 CEJ line segmentations DSC = 0.91, JI = 0.88 and PA = 0.99
Imak et al. (2022)	Private dataset. 380 periapical images.	No	A score-based multi-input CNN ensemble (MI-DCNNE) for the automatic diagnosis of dental caries	classification	accuracy = 99.13, sensitivity = 98%, specificity = 100%, precision = 100% and F1-score = 98.99%
Zhang et al. (2022)	Private dataset. 3932 oral photographs. Divided into training set (N = 2507, 63.76%), validation set (N = 300, 7.63%) and test set (N = 1125, 28,6%).	Yes	Dental caries DL model adapted from Single Shot MultiBox Detector (SSD)	Segmentation Classification	ROC = 85.65%.
Oztekin et al. (2022)	Private dataset. 250 panoramic images of 2048 × 1024px. Randomly divided into 70% for training, 20% for validation, and 10% for testing	No	A U-Net-based deep learning model for the automatic detection and classification of amalgam and composite fillings in panoramic images.	Segmentation	mean IoU = 0.767 Pixel Accuracy = 99.81%
Park and Lee (2023)	Private dataset.1414 panoramic X-ray	No	a five-axis-based tooth recognition model using Faster R-CNN	Segmentation	IoU = 72%
Franco et al. (2022)	4003 panoramic radiographs. divided into the age groups “under 15 years“ (n = 2,254) and “equal or older 15 years” (n = 1,749).	Yes	Eight CNN architectures was used InceptionV3, Xception, Inception ResNetV2, ResNet50, ResNet101, MobileNetV2, VGG16 and DenseNet121 to distinguish females and males using dentomaxillofacial features from a radiographic image	Segmentation Classification	DenseNet121- from scratch Accuracy = 71.64%, F1-score = 70.64%, Precision = 71.59%, Recall = 69.8%, Specificity = 87.47% DenseNet121-transfer learning. Accuracy = 82.18%, F1-score = 8.064%, Precision = 0.8072%, Recall = 0.8056%, Specificity = 0.9220%
Bayraktar and Ayan (2022)	Private dataset. 1000 bitewing radiographic divided into training (80%) and testing-validation (20%) groups	Yes	Modified (YOLO) model to detect caries lesions	Segmentation Classification	accuracy = 94.59%, sensitivity = 72.26%, specificity = 98.19%and AUC = 87.19
Liu et al. (2022)	Private dataset. 254 CBCT. Divided into three subsets: the training set (154, 60.6%), the validation set (30, 11.8%), and the test set (70, 27.6%)	Yes	U-Nets used to mandibular third molar and mandibular canal segmentation ResNet-34 used tomandibular third molar and mandibular canal relation classification: 3 classes	Segmentation Classification	segmentation of mandibular third molar: mean Dice similarity coefficient (mDSC) = 97% and a mean intersection over union (mIoU) = 96.06%; segmentation of mandibular canal: mDSC = 92% and a mIoU of 90%. classification: sensitivity = 90.2%, specificity = 95.0%, accuracy = 93.3%
Kohlakala et al. (2022)	Private dataset. 483 simulated X-ray images, which contain implants inserted into either human or pig jaws of size 512 × 512 and saved in JPEG format.	Yes	Two fully convolutional networks model FCN1 and FCN2 used to dental implant detection and recognition	Segmentation Classification	segmentation accuracy = 94.0% classification accuracy = 71.7%
Moidu et al. (2022)	Private dataset. 3540 periapical root areas (PRA). Divided into two subsets: the training and validation set (3000 PRA) and the test set (540 PRA)	Yes	A CNN model based on YOLO.v3 architecture to score the periapical lesion of mandibular teeth	Segmentation Classification	Segmentation: Dice coefficient = 89% mean intersection over union (Jaccard index) = 90%, recall = 89% precision = 90% Classification: sensitivity = 92.1% , specificity = 76%, accuracy = 86.3%, F1 score = 89% Matthews correlation coefficient (MCC) = 71%.

(*continued on next page*)

Table 11 (*continued*).

Zhu et al. (2022)	Private dataset. 1159 panoramic images.Divided into three subsets: the training set (900), the validation set (135) and the test set (124).	No	CariesNet: an automated system for caries diagnosis that employs a U-shape encoder–decoder framework, a reverse attention mechanism, and a Res2Net backbone to achieve precise segmentation.	Segmentation	Dice coefficient = 93.64%,accuracy = 93.61%, precision = 94.09% and recall = 86.01%,
Park et al. (2022)	Private dataset. 2348 intraoral photographic randomly assigned to training (1638), validation (410), and test (300) datasets	No	Deep learning algorithm was utilized to perform tooth surface segmentation using U-Net, caries classification using ResNet-18, and lesion localization using Faster R-CNN. These techniques are applied in the context of caries detection..	segmentation Classification	segmentation accuracy = 81.39%, AUC = 83.7%, classification accuracy = 9 2.5%, sensitivity = 89.0% and precision = 87.4%
Alotaibi et al. (2022)	Private dataset. 1724 periapical images. Divided in a training dataset (n = 1206; 70%), a validation dataset (n = 345; 20%), and a test dataset (n = 173; 10%).	No	A CNN-based model VGG-16 to detect periodontal bone loss and classify the alveolar bone levels in teeth affected by periodontal disease.	classification	binary classification accuracy = 73.04%, precision = 70%, recall = 70% and Matthews correlation coefcient (MCC) = 51% multi-classification: accuracy = 59.42%, precision = 83%, recall = 70% and Matthews correlation coefcient (MCC) = 0.65
Jang et al. (2022)	Data available on request from the authors. 300 periapical radiographic image. split into 80:20 ratio (train/test).	No	Evaluate the diagnostic performance of the Faster R-CNN model with ResNet 101 backbone for the tasks of object detection, classification, and localization.	Detection Classification Localization	classification performance: precision = 97.7%, recall = 99.2% and F1 score = 98.4% Segmentation performance IoU=90.7%.
Mohammad et al. (2022)	Private dataset. 240 panoramic images.	Yes	Dynamic programming of active contour (DP-AC) and convolutions neural network model to segment the maturity development of the mandibular premolars.	segmentation Classification	Classification : Cohen's Kappa = 58%, accuracy = 77%
Jiang et al. (2022)	Data available on request from the authors. 640 panoramic images. separated into a training set (80%) and a test set (20%)	Yes	UNet and YOLO-v4 were used to train a deep learning model for comprehensively diagnosing and staging periodontal alveolar bone loss.	Segmentation Classification	accuracy = 77%, Precision = 77%, Sensitivity = 77%, Specificity = 88% and F1-score = 77%.
Kabir et al. (2022)	A public repository contained 116 panoramic images (Abdi and Kasaei, 2020a). 1240 intraoral radiographs obtained from a private database.	No	U-Net segmentation model with ResNet-34 as the backbone was utilized to accurately segment different structures and regions of interest in dental images, such as teeth, CEJ lines (Cementoenamel Junction), and the bone area.	segmentation Numbering	Detection Precision = 0.99 Detection Recall = 99% Numbering Precision = 96% Numbering Recall = 96%
Vila-Blanco et al. (2022)	Private dataset. 1746 dental panoramic radiographs (orthopantomograms or OPGs) was split into training, validation, and test subsets,which contained 60%, 20%, and 20% of the cases, respectively.	Yes	XAS: Automatic yet eXplainable Age and Sex determination.	segmentation Classification	Tooth segmentation mAP@0.5 = 96.4%, mAP@0.75 =94.4% Sex classification Accuracy = 91.8%
Bonfanti-Gris et al. (2022)	Private dataset. 300 panoramic radiographs exported in jpeg file	No	A pre-trained CNN called Denti.Ai to perform the tasks of detecting and classifying dental structures in panoramic radiographs. .	segmentation Classification	Sensitivity = 56.92% Specificity = 92.20%
Aljabri et al. (2022)	Private dataset. 416 panoramic radiographs	Yes	4 deep learning models were developed to classify the type of canine impaction (Type I or Type II) from panoramic dental radiographic images: DenseNet-121, VGG-16, Inception V3, and ResNet-50.	Classification	Experiment 1 Accuracy (unbalanced data : 282 samples represent Type I and 134 Type II) Inception V3 0.8095 ResNet-50 0.7619 DenseNet-121 0.7976 VGG-16 0.6548 Experiment 2 Accuracy (balanced data : 134 samples represent Type I and 134 Type II) Inception V3 0.9259 ResNet-50 87.04% DenseNet-121 68.52% VGG-16 57.41%.

(*continued on next page*)

Table 11 (*continued*).

Sivasundaram and Pandian (2021)	Data available on request from the authors. 1171 panoramic dental images, augmented into 3513.	Yes	Modified LeNet architecture for classifying the oral cyst images and a morphology-based segmentation method for segmenting the cyst regions in the classified cyst images	segmentation Classification	sensitivity = 98.3%, specificity = 98.8%, an accuracy = 98.5%, precision = 97.53%, F1 score = 98.06%, average DSC = 98.04 and IoU = 97.6.
Hsiao et al. (2021)	Private dataset. Optical coherence tomography images. 32000 B-scans for training, 8000 B-scans for validation data, and 8000 B-scans for testing	No	A VGG16-based automated model is employed to detect the presence of subgingival calculus and accurately identify the location of the lesion in optical coherence tomography (OCT) images.	Detection Localization	accuracy = 95.06%. AUC = 0.973.
Krois et al. (2021)	Two datasets each have 650 images available if needed within data protection regulation boundaries.	Yes	U-Net was used to evaluate the generalizability of deep learning models for the detection of apical lesions on panoramic radiographs and to identify methods for improving their performance.	segmentation	Model trained only on First dataset image : F1-score = 54.1% ‘ Model trained only on second dataset image : F1-score = 32.7% Cross-Dataset training F1-score (First dataset) = 50.9% F1-score (Second dataset) = 46.1%.
Başaran et al. (2022)	Private dataset. optical coherence tomography images. 1084 dental panoramic radiographs.	No	AI model based on Faster R-CNN and Google Net Inception v2 to detect dental conditions.	segmentation	best sensitivity = 96.74%, best precision = 92.59% and the F1-score = 94.33%.
Chen et al. (2021a)	Private dataset. optical coherence tomography images. 2900 digital dental periapical radiographs	No	Faster R-CNNs to detect diseases including decay, periapical periodontitis, and periodontitis in dental periapical radiographs.	Detection localization	decay IoU = 0.71, Precision = 0.61, Recall = 0.54 and AP = 0.45 periapi IoU = 0.69, Precision = 0.51, Recall = 0.51 and AP = 0.36 periodo IoU = 0.68, Precision = 0.56, Recall = 0.61 and AP = 0.43.
Lee et al. (2021)	Private dataset. 304 bitewing radiographs. Divided into 149 radiographs for training on both structure and caries segmentation, 105 radiographs for training on caries segmentation, and 50 radiographs with no dental caries.	Yes	The U-Net architecture is used for the early detection of initial dental caries (tooth decay).	Detection	precision = 63.29%; recall = 65.02%; F1-score = 64.14%.
Bilgir et al. (2021)	Data available on request from the authors. 2482 panoramic radiographs. Divided into training (80%), validation (10%), and test (10%) groups.	No	Faster R-CNN Inception v2 model, was used to automatically detect and number teeth on the panoramic radiographs.	segmentation	Sensitivity (Recall) = 0.9559 Precision = 0.9652 F1Score = 0.9606.
Chen et al. (2021b)	Private dataset. 175 full-jaw 3D tooth models	Yes	A basic CNN model for 8-class tooth type classification (first premolar, second premolar, first molar, second molar in maxilla and mandible respectively)	classification	accuracy = 91.35%, precision = 91.49%, recall = 91.29%, F1-score = 0.91
Rajee and Mythili (2021)	Private dataset. 1000 panoramic dental images divided into 600 training samples and 400 testing samples	No	Gender classification on digital dental x-ray images using deep convolutional neural network model based on Resnet50 architecture.	Segmentation Classification	dice = 0.94 accuracy = 98.27 Sensitivity = 98.04 Specificity = 98.51 Precision = 98.52
Singh and Sehgal (2020)	Private dataset. 400 panoramic dental images, divided into 240 training samples and 160 testing samples	Yes	Numbering and Classification of Panoramic Dental Images Using 6-Layer Convolutional Neural Network (DCNN)	Pre-processing Segmentation Numbering classification (4 classes molar, premolar, canine and incisor).	Augmented database accuracy = 95% Sensitivity = 98% Specificity = 92% Original dataset accuracy = 92% Sensitivity = 90% Specificity = 85%
You et al. (2020)	Data available on request from the authors. 984 intraoral divided into 886 training samples and 98 testing samples.	No	Deep learning-based dental plaque detection on primary teeth based on DeepLab and DeepLabV3 architectures .	segmentation	mean intersection-over-union (MIoU) = 0.726

(*continued on next page*)

Table 11 (*continued*).

Kwak et al. (2020)	Data available on request from the authors. 49094 CBCT images, divided into train:valid:test sets with the ratio of 6:2:2.	No	An Automatic mandibular canal detection based on 2D SegNet, 2D and 3D U-Nets architectures.	Segmentation	2D U-Net accuracy = 82% 2D SegNet accuracy = 96% 3D U-Net accuracy = 99%.
Zhao et al. (2020)	A benchmark public dataset. 1500 panoramic X-ray images divided into 1200 training samples, 150 validation samples and 150 testing samples.	No	A Two-Stage Attention Segmentation Network (TSASNet) on dental panoramic X-ray images	segmentation	Accuracy = 96.94 Specificity = 97.81 Precision = 94.97 Recall = 93.77 Dice = 92.72
Cantu et al. (2020)	Private dataset. 3686 bitewing radiographs. The data was divided into a training (3293 images), validation (252 images) and test dataset (141 images)	No	A deep learning model based on U-net to detect caries lesions on bitewing radiographs	Segmentation	Accuracy 0.80 Sensitivity 0.75 Specificity 0.83 F1score 0.73 MCC 0.57
Setzer et al. (2020)	Private dataset. 20 CBCT images were divided into 16 CBCT images) were used for training, and 4 CBCT images was used for validation	Yes	A Deep Learning model for the automated segmentation of cone-beam computed tomographic (CBCT) images and the detection of periapical lesions. Based on a U-Net architecture.	Segmentation	DICE index Lesion 0.52 Tooth structure 0.74 Bone 0.78 Restorative materials 0.58 Background 0.95
Chen et al. (2019)	Data available on request from the authors. 1250 periapical film divided into training (800), validation (20.), and test (250) groups	No	A deep learning approach for automatic teeth detection and numbering based on faster R-CNN (A total of 32 teeth classes were required to be recognized in the X-ray images)	segmentation	Numbering precision = 0.98 recall = 0.98 Mean IOU = 0.91 Numbering precision = 0.91 recall = 0.91
Al Kheraif et al. (2019)	Private dataset. 1500 panoramic images divided into training of 800 with testing of 7000	No	Histogram enhancement with a CNN to detect dental diseases	segmentation classification	Segmentation : Accuracy = 0.912; Specificity = 0.967; Precision = 0.89; Recall = 0.92 and F1 score = 0.94 classification Accuracy = 97.07%
Ekert et al. (2019)	Private dataset. A synthesized data set of 2001 tooth segments from panoramic radiographs	Yes	A deep CNNs to detect apical lesions (ALs).	segmentation	AUC = 0.85, sensitivity = 0.65 and specificity = 0.87
Tian et al. (2019)	Private dataset. 600 dental models	No	An automatic segmentation and classification method for 3D dental model via 3D CNN.	Segmentation classification : seven tooth types which are the central incisors, lateral incisors, canines, first premolars, second premolars, first molars and second molars.	classification accuracy in Level1 network is 95.96%, the average classification accuracy in Level2 network is 88.06%, and the accuracy of tooth segmentation is 89.81%
Lee et al. (2018)	Private dataset. 3000 periapical radiographic, divided into a training and validation dataset (n = 2400 [80%]) and a test dataset (n = 600 [20%])	Yes	GoogLeNet Inception v3 for detection and diagnosis of dental caries on periapical radiographs.	Segmentation classification	premolar accuracy = 89.0%, molar accuracy = 88.0%, both premolar and molar accuracy = 82.0% AUC = 0.917 on premolar, AUC = 0.890 on molar, and AUC = 0.845 on both premolar and molar models.

focus in our DL methodologies is interpretability, which enhances clarity in our model's decision-making process.

Anticipating the dynamic nature of this field, we foresee that this article is evolving into an ongoing review that is regularly updated to capture emerging publications. This dedicated commitment actively contributes to the continuous advancement of research on dental panoramic-image segmentation.

Simultaneously, our forthcoming initiatives will extend beyond refining model performance to achieve robust generalization across diverse dental centers and automating medical reports, pushing the boundaries of dental image analysis. We are committed to exploring Explainable Artificial Intelligence (XAI) to provide transparency and clarity to our models. Furthermore, we see potential in extending the impact of our model to dental planning and providing valuable support to dental students in their studies.

CRediT authorship contribution statement

Walid Brahmi: Resources, Writing – original draft. **Imen Jdey:** Writing – review & editing, Conceptualization, Supervision. **Fadoua Drira:** Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relation- ships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the REGIM-Lab (Research Laboratory in Intelligent Machines, code LR11ES48) at the National Engineering School of Sfax (ENIS), University of Sfax, and the National School of Electronics and Telecommunications of Sfax (ENET'Com), University of Sfax, Tunisia.

References

Abdi, Amir, Kasaei, Shohreh, 2020a. Panoramic dental x-rays with segmented mandibles. Mendeley Data 2, https://data.mendeley.com/datasets/hxt48yk462/1.

Abdi, Amir, Kasaei, Shohreh, 2020b. Panoramic dental x-rays with segmented mandibles. Mendeley Data 2, https://www.kaggle.com/datasets/daverattan/dental-xrary-tfrecords/.

Abedalla, Ayat, et al., 2021. Chest X-ray pneumothorax segmentation using U-net with EfficientNet and ResNet architectures. PeerJ Comput. Sci. 7, e607. http://dx.doi.org/10.7717/peerj-cs.607.

ACM, 2023. A research, discovery and networking platform. URL https://dl.acm.org/.

Al Kheraif, Abdulaziz A., Wahba, Ashraf A., Fouad, H., 2019. Detection of dental diseases from radiographic 2D dental image using hybrid graph-cut technique and convolutional neural network. Measurement 146, 333–342.

Ali, Md Anas, Fujita, Daisuke, Kobashi, Syoji, 2023. Teeth and prostheses detection in dental panoramic X-rays using CNN-based object detector and a priori knowledge-based algorithm. Sci. Rep. 13 (1), 16542.

Aljabri, Malak, et al., 2022. Canine impaction classification from panoramic dental radiographic images using deep learning models. Inform. Med. Unlocked 30, 100918. http://dx.doi.org/10.1016/j.imu.2022.100918.

Alotaibi, Ghala, et al., 2022. Artificial intelligence (AI) diagnostic tools: utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically–A retrospective study. BMC Oral Health 22 (1), 399. http://dx.doi.org/10.1186/s12903-022-02436-3.

Alzubaidi, Laith, et al., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 1–74.

Amasya, Hakan, et al., 2024. Development and validation of an artificial intelligence software for periodontal bone loss in panoramic imaging. Int. J. Imaging Syst. Technol. 34 (1), e22973.

Awassa, Lamia, et al., 2022. Study of different deep learning methods for coronavirus (COVID-19) pandemic: taxonomy, survey and insights. Sensors 22 (5), 1890. http://dx.doi.org/10.3390/s22051890.

Başaran, Melike, et al., 2022. Diagnostic charting of panoramic radiography using deep-learning artificial intelligence system. Oral Radiol. 1–7.

Bayraktar, Yusuf, Ayan, Enes, 2022. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. Clin. Oral Investig. 26 (1), 623–632. http://dx.doi.org/10.1007/s00784-021-04040-1.

Bilgir, Elif, et al., 2021. An artificial ıntelligence approach to automatic tooth detection and numbering in panoramic radiographs. BMC Med. Imaging 21, 1–9. http://dx.doi.org/10.1186/s12880-021-00656-7.

Bonfanti-Gris, Monica, et al., 2022. Evaluation of an artificial intelligence web-based software to detect and classify dental structures and treatments in panoramic radiographs. J. Dent. 126, 104301. http://dx.doi.org/10.1016/j.jdent.2022.104301.

Booth, Andrew, et al., 2021. Systematic approaches to a successful literature review. pp. 1–100.

Bouzidi, Sonia, Jdey, Imen, Alimi, Adel, A Vision Transformer Approach with L2 Regularization for Sustainable Fashion Classification. Available at SSRN 4686032.

Brahmi, Walid, Jdey, Imen, 2023. Automatic tooth instance segmentation and identification from panoramic X-Ray images using deep CNN. Multimedia Tools Appl. 1–21. http://dx.doi.org/10.1007/s11042-023-17568-z.

Cantu, Anselmo Garcia, et al., 2020. Detecting caries lesions of different radiographic extension on bitewings using deep learning. J. Dent. 100, 103425.

Chen, Hu, et al., 2019. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci. Rep. 9 (1), 3840.

Chen, Hu, et al., 2021a. Dental disease detection on periapical radiographs based on deep convolutional neural networks. Int. J. Comput. Assist. Radiol. Surg. 16, 649–661.

Chen, Qingguang, et al., 2021b. Hierarchical CNN-based occlusal surface morphology analysis for classifying posterior tooth type using augmented images from 3D dental surface models. Comput. Methods Programs Biomed. 208, 106295.

Choi, Eunhye, et al., 2023. Artificial intelligence in diagnosing dens evaginatus on periapical radiography with limited data availability. Sci. Rep. 13 (1), 13232.

Chun, So-Young, et al., 2023. Automatic classification of 3D positional relationship between mandibular third molar and inferior alveolar canal using a distance-aware network. BMC Oral Health 23 (1), 794.

Egger, Jan, et al., 2022. Medical deep learning– A systematic meta-review. Comput. Methods Programs Biomed. 221, 106874. http://dx.doi.org/10.1016/j.cmpb.2022.106874.

Ekert, Thomas, et al., 2019. Deep learning for the radiographic detection of apical lesions. J. Endod. 45 (7), 917–922.

Franco, Ademir, et al., 2022. Diagnostic performance of convolutional neural networks for dental sexual dimorphism. Sci. Rep. 12 (1), 17279.

Hamamci, Ibrahim Ethem, 2024. Dentex challenge. https://zenodo.org/records/7812323#.ZDQE1uxBwUG.

Hamamci, Ibrahim Ethem, et al., 2023. DENTEX: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic X-rays. arXiv preprint arXiv:2305.19112.

Hanley, James A., McNeil, Barbara J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1), 29–36.

Hcini, Ghazala, Jdey, Imen, Dhahri, Habib, 2024. Investigating deep learning for early detection and decision-making in Alzheimer's disease: A comprehensive review. Neural Process. Lett. 56 (3), 1–38.

Hcini, Ghazala, Jdey, Imen, Ltifi, Hela, 2022. Improving malaria detection using L1 regularization neural network. JUCS: J. Univers. Comput. Sci. 28 (10), 1087–1107. http://dx.doi.org/10.3897/jucs.81681.

Hcini, Ghazala, Jdey, Imen, Ltifi, Hela, 2023. HSV-net: A custom CNN for malaria detection with enhanced color representation. In: 2023 International Conference on Cyberworlds. (CW), IEEE, Sousse, Tunisia.

Hcini, Ghazala, et al., 2021. Hyperparameter optimization in customized convolutional neural network for blood cells classification. J. Theor. Appl. Inf. Technol. 99, 5425–5435.

He, Kaiming, et al., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada.

He, Kaiming, et al., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy.

Heni, Ashraf, Jdey, Imen, Ltifi, Hela, 2022. K-means and fuzzy c-means fusion for object clustering. In: 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT). 1, IEEE, pp. 177–182.

Heni, Ashraf, Jdey, Imen, Ltifi, Hela, 2023. Blood cells classification using deep learning with customized data augmentation and EK-means segmentation. J. Theor. Appl. Inf. Technol. 101 (3).

Howard, Andrew G., et al., 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hsiao, Tien-Yu, et al., 2021. Disease activation maps for subgingival dental calculus identification based on intelligent dental optical coherence tomography. Transl. Biophotonics 3 (3), e202100001.

Huang, Gao, et al., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA.

Humans in the Loop, 2024a. Dental radiography. https://www.kaggle.com/datasets/humansintheloop/teeth-segmentation-on-dental-x-ray-images.

Humans in the Loop, 2024b. Roboflow an end-to-end computer vision platform. https://universe.roboflow.com/.

Hwang, Jae-Joon, et al., 2019. An overview of deep learning in the field of dentistry. Imaging Sci. Dentistry 49 (1), 1.

IEEE, 2023. Institute of electrical and electronics engineers. URL https://ieeexplore.ieee.org/Xplore/home.jsp.

Imak, Andac, et al., 2022. Dental caries detection using score-based multi-input deep convolutional neural network. IEEE Access 10, 18320–18329. http://dx.doi.org/10.1109/ACCESS.2022.3150358.

Jang, Woo Sung, et al., 2022. Accurate detection for dental implant and peri-implant tissue by transfer learning of faster R-CNN: A diagnostic accuracy study. BMC Oral Health 22 (1), 591. http://dx.doi.org/10.1186/s12903-022-02539-x.

Jdey, Imen, 2022. Trusted smart irrigation system based on fuzzy iot and blockchain. In: International Conference on Service-Oriented Computing. Springer, pp. 154–165.

Jdey, Imen, Hcini, Ghazala, Ltifi, Hela, 2023. Deep learning and machine learning for malaria detection: Overview, challenges and future directions. Int. J. Inf. Technol. Decis. Mak. 1–32. http://dx.doi.org/10.1142/S0219622023300045.

Jdey, Imen, Toumi, Abdelmalek, Dhibi, Mounir, Khenchaf, Ali, 2012. The contributionof fusion techniques in the recognition systems of radar targets. IET International Conference on Radar Systems (Radar 2012), Glasgow, UK. IET, pp. 1–5. http://dx.doi.org/10.1049/cp.2012.1663.

Jiang, Linhong, et al., 2022. A two-stage deep learning architecture for radiographic staging of periodontal bone loss. BMC Oral Health 22 (1), 106. http://dx.doi.org/10.1186/s12903-022-02119-z.

Jlassi, Sinda, Jdey, Imen, Ltifi, Hela, 2021. Bayesian hyperparameter optimization of deep neural network algorithms based on ant colony optimization. In: Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September (2021) 5–10, Proceedings, Part III 16. Springer International Publishing.

Kabir, Tanjida, et al., 2022. A comprehensive artificial intelligence framework for dental diagnosis and charting. BMC Oral Health 22 (1), 480. http://dx.doi.org/10.1186/s12903-022-02514-6.

Kaji, Shizuo, Kida, Satoshi, 2019. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiol. Phys. Technol. 12 (3), 235–248.

Kang, Dae-Young, Duong, Hieu Pham, Park, Jung-Chul, 2020. Application of deep learning in dentistry and implantology. J. Implantol. Appl. Sci. 24 (3), 148–181.

Keele, Staffs, 2007. Guidelines for performing systematic literature reviews in software engineering. pp. 1–57.

Ker, Justin, et al., 2017. Deep learning applications in medical image analysis. IEEE Access 6, 9375–9389. http://dx.doi.org/10.1109/ACCESS.2017.2788044.

Khanna, Sunali S., Dhaimade, Prita A., 2017. Artificial intelligence: transforming dentistry today. Indian J. Basic Appl. Med. Res. 6 (3), 161–167.

Kohlakala, Aviwe, et al., 2022. Deep learning-based dental implant recognition using synthetic X-ray images. Med. Biol. Eng. Comput. 60 (10), 2951–2968.

Krois, Joachim, et al., 2021. Generalizability of deep learning models for dental image analysis. Sci. Rep. 11 (1), 6102.

Kwak, Gloria Hyunjung, et al., 2020. Automatic mandibular canal detection using a deep convolutional neural network. Sci. Rep. 10 (1), 5711.

LeCun, Yann, et al., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Lee, Jae-Hong, et al., 2018. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J. Dent. 77, 106–111.

Lee, Shinae, et al., 2021. Deep learning for early dental caries detection in bitewing radiographs. Sci. Rep. 11 (1), 16807.

Lee, Chun-Teh, et al., 2022. Use of the deep learning approach to measure alveolar bone level. J. Clin. Periodontol. 49 (3), 260–269. http://dx.doi.org/10.1111/jcpe.13574.

Litjens, Geert, et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, Mu-Qing, et al., 2022. Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on CBCT. Clin. Oral Investig. 1–11.

Mogli, G.D., 2009. Medical records role in healthcare delivery in 21st century. Acta Inform. Med. 17 (4), 209–212.

Mohammad, Norhasmira, et al., 2022. Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging. BMC Med. Imaging 22 (1), 66. http://dx.doi.org/10.1186/s12880-022-00794-6.

Moidu, Navas P., et al., 2022. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. Clin. Oral Investig. 26 (1), 651–658.

Momeni, Mohamdreza, 2024. Dental radiography. https://www.kaggle.com/datasets/imtkaggleteam/dental-radiography.

Moulahi, Wided, Jdey, Imen, Moulahi, Tarek, Alawida, Moatsum, Alabdulatif, Abdulatif, 2023. A blockchain-based federated learning mechanism for privacy preservation of healthcare IoT data. Comput. Biol. Med. 167, 107630.

Müller, Dominik, Soto-Rey, Iñaki, Kramer, Frank, 2022. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res. Notes 15 (1), 210. http://dx.doi.org/10.1186/s13104-022-06096-y.

Oztekin, Faruk, et al., 2022. Automatic semantic segmentation for dental restorations in panoramic radiography images using U-net model. Int. J. Imaging Syst. Technol. 32 (6), 1990–2001.

Panetta, Karen, Rajendran, Rahul, Ramesh, Aruna, Rao, Shishir Paramathma, Agaian, Sos, 2024. Tufts dental database. https://www.kaggle.com/datasets/deepologylab/tufts-dental-database.

Panetta, Karen, et al., 2021b. Tufts dental database: A multimodal panoramic x-ray dataset for benchmarking diagnostic systems. IEEE J. Biomed. Health Inform. 26 (4), 1650–1659.

Park, Jonghwan, Lee, Younghoon, 2023. Oriented-tooth recognition using a five-axis object-detection approach. Appl. Intell. 53 (9), 9846–9857.

Park, Eun Young, et al., 2022. Caries detection with tooth surface segmentation on intraoral photographic images using deep learning. BMC Oral Health 22 (1), 573. http://dx.doi.org/10.1186/s12903-022-02589-1.

Pauwels, Ruben, 2021. A brief introduction to concepts and applications of artificial intelligence in dental imaging. Oral Radiol. 37 (1), 153–160.

Pickering, Brian W., et al., 2013. Data utilization for medical decision making at the time of patient admission to ICU. Crit. Care Med. 41 (6), 1502–1510.

Prados-Privado, María, et al., 2020. Dental images recognition technology and applications: A literature review. Appl. Sci. 10 (8), 2856. http://dx.doi.org/10.3390/app10082856.

Rajee, M.V., Mythili, C., 2021. Gender classification on digital dental x-ray images using deep convolutional neural network. Biomed. Signal Process. Control 69, 102939.

Redmon, Joseph, et al., 2016. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada.

Ren, Shaoqing, et al., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28.

Román, Julio César Mello, et al., 2021. Panoramic dental radiography image enhancement using multiscale mathematical morphology. Sensors 21 (9), 3110.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing.

Schwendicke, Falk, et al., 2019. Convolutional neural networks for dental image diagnostics: A scoping review. J. Dent. 91, 103226.

ScienceDirect, 2023. Elsevier's premier platform of peer-reviewed scholarly literature. URL https://www.sciencedirect.com/.

Setzer, Frank C., et al., 2020. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. J. Endod. 46 (7), 987–993.

Silva, Gil, Oliveira, Luciano, Pithon, Matheus, 2018. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. Expert Syst. Appl. 107, 15–31.

Simonyan, Karen, Zisserman, Andrew, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, Prerna, Sehgal, Priti, 2020. Numbering and classification of panoramic dental images using 6-layer convolutional neural network. Pattern Recognit. Image Anal. 30, 125–133.

Sivasundaram, Sivagami, Pandian, Chitra, 2021. Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture. Int. J. Imaging Syst. Technol. 31 (4), 2214–2225.

Slimani, Nawel, Jdey, Imen, Kherallah, Monji, 2023. Performance comparison of machine learning methods based on cnn for satellite imagery classification. In: 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, pp. 185–189.

Slimani., Nawel, Jdey., Imen, Kherallah., Monji, 2024. Improvement of satellite image classification using attention-based vision transformer. In: Proceedings of the 16th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART. SciTePress, INSTICC, (ISSN: 2184-433X) ISBN: 978-989-758-680-4, pp. 80–87. http://dx.doi.org/10.5220/0012298400003636.

SpringerLink, 2023. A reading platform of choice for hundreds of thousands of researchers worldwide. URL https://link.springer.com/.

Szegedy, Christian, et al., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts.

Szegedy, Christian, et al., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada.

Taha, Abdel Aziz, Hanbury, Allan, 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging 15, 1–28. http://dx.doi.org/10.1186/s12880-015-0068-x.

Tian, Sukun, et al., 2019. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access 7, 84817–84828. http://dx.doi.org/10.1109/ACCESS.2019.2924262.

Vila-Blanco, Nicolás, et al., 2022. XAS: Automatic yet explainable age and sex determination by combining imprecise per-tooth predictions. Comput. Biol. Med. 149, 106072. http://dx.doi.org/10.1016/j.compbiomed.2022.106072.

Wahono, Romi Satria, 2020. Systematic literature review: pengantar, tahapan dan studi kasus. pp. 1–126, Romisatriawahono. Net.

Walid, Brahmi, Imen, Jdey, 2023. Panoramic dental xray dataset. Mendeley Data, V2, https://data.mendeley.com/datasets/73n3kz2k4k/2.

Wang, Xianyun, Gao, Sizhe, Jiang, Kaisheng, Zhang, Huicong, Wang, Linhong, Chen, Feng, Yu, Jun, Yang, Fan, 2024. Panoramic-caries-segmentation. https://github.com/Zzz512/MLUA.

Wang, Zhaobin, Wang, E., Zhu, Ying, 2020. Image segmentation evaluation: A survey of methods. Artif. Intell. Rev. 53 (8), 5637–5674.

Wang, Xianyun, et al., 2023. Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation. Neurocomputing 540, 126208.

Wiley Online Library, 2023. A digital library of scholarly resources, journals, books, and references. URL https://onlinelibrary.wiley.com/.

You, Wenzhe, et al., 2020. Deep learning-based dental plaque detection on primary teeth: A comparison with clinical assessments. BMC Oral Health 20, 1–7. http://dx.doi.org/10.1186/s12903-020-01114-6.

Zhang, Xuan, et al., 2022. Development and evaluation of deep learning for screening dental caries from oral photographs. Oral Dis. 28 (1), 173–181.

Zhao, Yue, et al., 2020. TSASNet: Tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network. Knowl.-Based Syst. 206, 106338.

Zhu, Zijiang, et al., 2021. Indoor scene segmentation algorithm based on full convolutional neural network. Neural Comput. Appl. 33, 8261–8273.

Zhu, Haihua, et al., 2022. CariesNet: A deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image. Neural Comput. Appl. 1–9.