

Exploring the Role of Convolutional Neural Networks (CNN) in Dental Radiography Segmentation: A Comprehensive Systematic Literature Review

Walid Brahmi¹, Imen Jdey¹, Adel Alimi¹

^{1*}Research Groups in Intelligent Machines (REGIM Lab), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, 3038 Sfax,, Tunisia.

Contributing authors: bensghaierwaleed@gmail.com;
imen.jdey@fstsbz.u-kairouan.tn; adel.alimi@enis.tn;

Abstract

In the field of dentistry, there is a growing demand for increased precision in diagnostic tools, with a specific focus on advanced imaging techniques such as computed tomography, cone beam computed tomography, magnetic resonance imaging, ultrasound, and traditional intra-oral periapical X-rays. Deep learning has emerged as a pivotal tool in this context, enabling the implementation of automated segmentation techniques crucial for extracting essential diagnostic data. This integration of cutting-edge technology addresses the urgent need for effective management of dental conditions, which, if left undetected, can have a significant impact on human health. The impressive track record of deep learning across various domains, including dentistry, underscores its potential to revolutionize early detection and treatment of oral health issues.

Objective: Having demonstrated significant results in diagnosis and prediction, deep convolutional neural networks (CNNs) represent an emerging field of multidisciplinary research. The goals of this study were to provide a concise overview of the state of the art, standardize the current debate, and establish baselines for future research.

Method: In this study, a systematic literature review is employed as a methodology to identify and select relevant studies that specifically investigate the deep learning technique for dental imaging analysis. This study elucidates the methodological approach, including the systematic collection of data, statistical analysis, and subsequent dissemination of outcomes.

Results: In incorporating 45 studies, we identified selection criteria and research objectives, addressing significant gaps in the existing literature. These studies assist clinicians in examining dental conditions and classifying dental structures, including caries detection and the identification of various tooth types. We evaluated model performance, addressing the identified gaps, using diverse metrics that we strive to list and explain.

Conclusion: This work demonstrates how Convolutional Neural Networks (CNNs) can be employed to analyze images, serving as effective tools for detecting dental pathologies. Although this research acknowledged some limitations, CNNs utilized for segmenting and categorizing teeth exhibited their highest level of performance overall.

Keywords: Deep learning, Convolutional neural network, Dental imaging, Segmentation, Evaluation Metrics

1 Introduction

Medical information plays a vital role in healthcare by aiding healthcare professionals in making precise diagnoses, delivering effective treatments, and informed decisions about patient care [1, 2]. Within the realm of medical information, medical imaging stands out as an invaluable source.

In modern healthcare, medical imaging has become an essential tool, offering crucial visual insights into the internal structures and conditions of the human body [3]. Within dentistry, dental imaging refers to the use of various imaging techniques to capture detailed images of the oral and maxillofacial structures. These images are invaluable for diagnosing dental conditions, planning treatments, and monitoring oral health. Dental imaging plays a vital role in enabling dental professionals to visualize and evaluate the teeth, jaws, and supporting structures.

Dental images can be acquired through different modalities, including X-rays, computed tomography (CT), magnetic resonance imaging (MRI), and intraoral cameras. Each modality offers specific advantages and is selected based on the specific diagnostic requirements and clinical scenario.

Deep learning, in particular convolutional neural networks (CNNs) and computer vision, has become a revolutionary area of dentistry. Thanks to the use of these cutting-edge technologies, the dental industry is undergoing a period of transformation in which the analysis of dental images and the integration of artificial intelligence are revolutionizing many aspects of dental care [4]. By harnessing the capabilities of deep learning algorithms and computer vision techniques, dentistry has the potential to dramatically improve diagnostic performance, streamline treatment planning processes and, ultimately, improve patient outcomes.

The utilization of CNNs in dentistry encompasses the detection of structures like teeth and bone, as well as the identification of pathologies such as caries and apical lesions. Additionally, CNNs are employed to segment images by isolating the areas of interest and classify them based on specific features like enamel caries lesions or cysts. However, it is important to acknowledge the limitations within this field. One of the significant weaknesses lies in the relatively small and private nature of the available datasets, which restricts the diversity and size of the training data. Moreover, some of the AI solutions developed in this domain may lack robustness and stability, raising concerns about their reliability and performance.

To analyze dental images for tasks like lesion detection, age or sex determination, and human identification, tooth segmentation plays a vital role. In the field of oral medicine, the automatic segmentation of teeth in panoramic radiographs is a significant focus of research in image analysis. However, segmenting teeth in panoramic radiographs presents challenges due to the presence of other anatomical structures such as the chin, spine, and jaws.

Within the existing literature, a multitude of studies have employed convolutional neural networks (CNNs) for tooth segmentation and identification in dental images. These advanced deep learning 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 deep learning, exploring its various categories, common applications, and specifically zooming in on the architecture of Convolutional Neural Networks (CNNs) within the context of medical image analysis in computer vision. The subsequent sections 4, 5, and 6 present the study methodologies, research topics, and synthesis findings, offering detailed insights into the conducted research. Finally, Sections 7 and 8 discuss the study’s limitations and provide a conclusion.

2 Related work

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

Schwendicke et al. (2019) [5] conducted a survey on the application of Convolutional Neural Networks (CNNs) in dental and oral medicine imagery. Their review encompassed 36 relevant articles and conference proceedings spanning from 2015 to 2019, addressing clinical challenges in areas such as general odontology, cariology, endodontics, periodontics, orthodontics, dental radiology, forensic odontology, and general medicine.

Hwang et al. (2019) [6] published a survey on deep learning in oral and maxillofacial radiology, identifying 25 pertinent papers up to December 2018 using PubMed, Scopus, and IEEE Explore databases. The study collected data on deep learning architecture, training dataset size, evaluation results, advantages and disadvantages, study objectives, and imaging modalities.

Kang et al. (2020) [7] reported on a prior study focusing on the application of deep learning algorithms in dentistry and implantology. They analyzed 62 articles from MEDLINE and IEEE Xplore, categorizing them into tooth detection, numbering, segmentation, and bone segmentation. The study included articles in all languages and published before October 24, 2019, providing details on Author Year, Architecture, Input, Output, and Performance metrics.

Prados-Privado et al. (2020) [8] conducted a systematic study visualizing the state of artificial intelligence in dental applications, covering the detection of teeth, caries, filled teeth, crowns, prostheses, dental implants, and endodontic treatments. Utilizing three digital databases (PubMed, IEEE Xplore, and arXiv.org), they identified 18 relevant papers. Notably, the study did not include any segmentation methods based on deep learning.

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

- We delve into a comprehensive exploration of Convolutional Neural Networks (CNN), the pre-eminent deep learning algorithm. This involves an in-depth explanation of concepts, theory, and contemporary architectures.
- Our study thoroughly examines crucial challenges in Deep Learning, including the shortage of training data, data imbalance, and sample data quality. We also delve into 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.

Table 1: Comparison of this study with existing literature reviews on deep learning for dental image segmentation.

Study	Release year	Total number of papers	Limitations
[5]	2019	36	Study includes studies that do not process dental images with deep learning The elements impacting the model's performance were not addressed
[6]	2019	25	The elements impacting the model's performance were not addressed
[7]	2020	62	The study includes studies that do not deal with dental images. The elements impacting the model's performance were not addressed
[8]	2020	18	The article does not deal with the segmentation task. The elements impacting the model's performance were not addressed

3 Background

3.1 Deep Learning

Deep learning (DL) holds significant potential to advance applications with a real impact on the field of dentistry. Positioned within Machine Learning (ML) as illustrated in Figure 1, DL

empowers machines to emulate human intelligence in increasingly sophisticated ways [9, 10]. To assess and distill knowledge from extensive datasets, DL employs multiple layers of non-linear units. In a basic scenario, there are two sets of neurons: one set receives input signals, and the other set sends output signals. Upon receiving an input, the input layer modifies it before passing it on to the next layer. While the layers in a deep network may not be neurons, conceptualizing them as such proves beneficial. This approach enables the algorithm to leverage multiple processing layers comprising diverse linear and non-linear transformations [11]. Deep learning is part of the

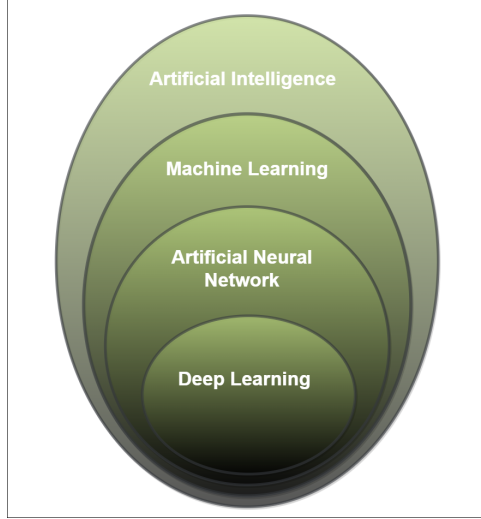


Fig. 1: The Subsets of Artificial Intelligence.

broader family of machine learning techniques centered around learning data representations. An observation, such as an image, can take various forms, including a set of edges, specific shape areas, or, more abstractly, a vector of intensity values per pixel.

Several deep learning architectures, such as deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), and others, have found application in a wide array of fields. As depicted in Figure 2, deep learning is frequently employed by partnerships to address diverse and challenging issues falling into several categories, including:

- **Computer Vision:** Deep learning is widely used in computer vision tasks such as image classification, object detection, image segmentation, and facial recognition [12].
- **Natural Language Processing (NLP):** Deep learning models are used to process and understand human language, including tasks such as language translation, sentiment analysis, text generation, and question answering [13].
- **Speech Recognition:** Deep learning techniques are employed in speech recognition systems to convert spoken language into written text. This technology is used in virtual assistants, transcription services, and voice-controlled devices [14].
- **Healthcare:** Deep learning is used in medical imaging analysis, disease diagnosis, drug discovery, and personalized medicine. It helps in improving accuracy, efficiency, and speed in various healthcare applications [15, 16].
- **Agriculture:** Finding problematic environmental conditions [17].
- **Financial Services:** Deep learning is employed in fraud detection, risk assessment, algorithmic trading, credit scoring, and customer behavior analysis in the financial industry [18].
- **Economics:** Deep learning is employed in the field of economics to analyze and comprehend intricate economic phenomena, predict economic behavior, analyze economic data, and evaluate the effectiveness of policy interventions [19].

- **Robotics:** Deep learning techniques are utilized in robotics for tasks such as object recognition, grasping and manipulation, navigation, and human-robot interaction [20].
- **Autonomous Driving :**Significant improvements in the field of autonomous driving have been made possible by the development of sensing, perception, signal processing technologies and deep learning approaches. This has reduced the effort required of human drivers and enhanced the safety of autonomous driving [21].
- **Additional uses:** These days, deep learning is applied in nearly every industry. There are other further deep learning uses, including automated text creation [22], game play [23], and picture captioning [24].

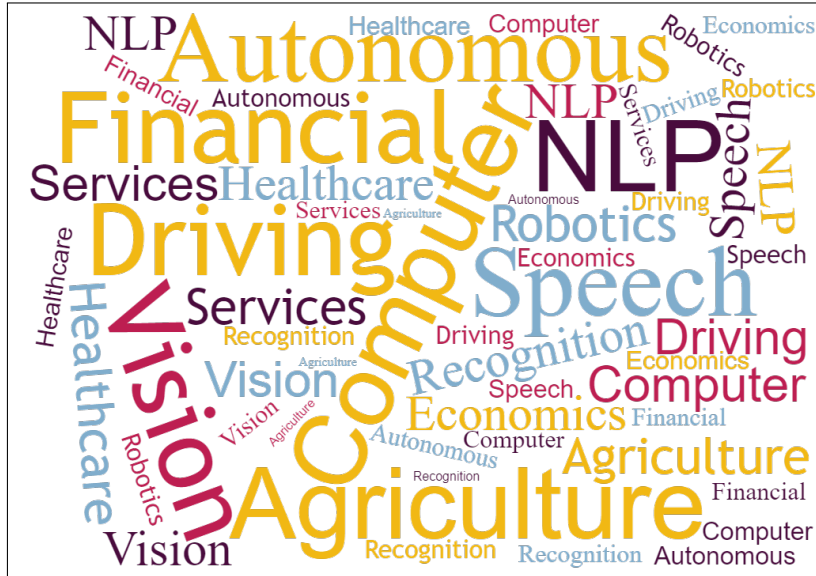


Fig. 2: Common Deep Learning Applications.

3.1.1 Convolutional Neural Networks (CNNs)

A Convolutional Neural Network (CNN) is a specialized deep learning model designed for processing data with a grid pattern, particularly images. Comprising multiple layers, each with a distinct purpose, the CNN architecture includes Convolutional layers, Rectified Linear Unit (ReLU) layers, pooling layers, and a fully connected layer, as depicted in Figure 3. CNNs automate feature extraction from images, eliminating the need for manual intervention. The CNN architecture is bifurcated into two segments: convolution and densely connected.

In convolutional layers, tensors known as feature maps are employed. A color image, represented as a 3-dimensional tensor with three channels (red, green, and blue - RGB), is denoted as (height, width, channels). To illustrate the theoretical aspects of this network type, consider the example presented in Figure 3, using a grayscale image of Width*Height pixels as input data. This image corresponds to a shape tensor (Width, Height, 1), indicating the number of neurons.

If Width=28 and Height=28, the tensor would consist of 784 neurons. For a color image, the tensor shape would be (28, 28, 3), resulting in 2352 neurons. It is important to note that the image, essentially a matrix where pixel colors range from 0 to 255, is normalized to a color range between 0 and 1 before entering the network.

In a typical CNN architecture [25], multiple convolution layers and a pooling layer are repeated several times, followed by one or more fully connected layers. The process of transforming input data into output data through these layers is known as forward propagation. The initial two layers, convolution, and pooling are responsible for feature extraction. The convolution layer applies filters

to the input data, capturing distinct features or patterns. The pooling layer then reduces the dimensionality of the feature maps, retaining essential information while discarding some spatial details.

The final layer, the fully connected layer, maps the extracted features to the ultimate output, such as classification. It connects every neuron in the previous layer to the neurons in the current layer, allowing the network to learn intricate relationships and make predictions based on the extracted features.

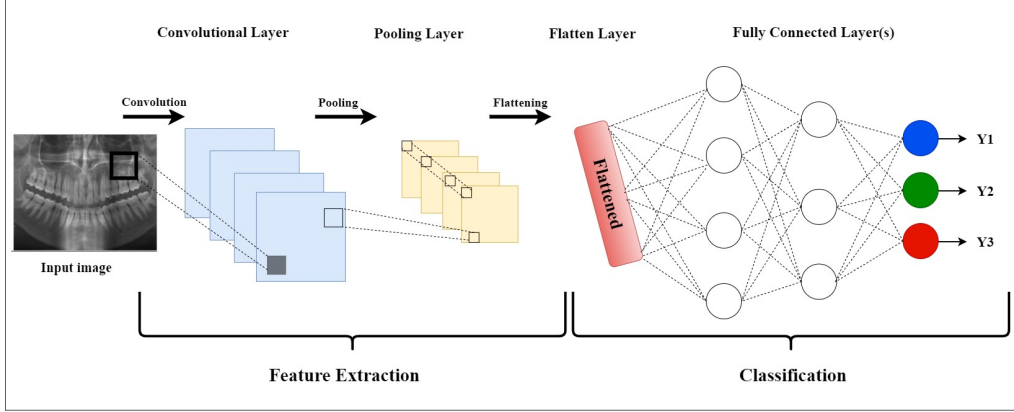


Fig. 3: Outline of CNN.

3.1.2 Categorization of Deep Learning methods

Similar to the field of machine learning, deep learning encompasses a diverse range of methodologies aimed at deriving meaningful insights from data. As figure 4 illustrates, deep learning techniques fall into four main categories [25, 26]: supervised learning, unsupervised learning, semi-supervised learning (also known as hybrid learning), and reinforcement learning.

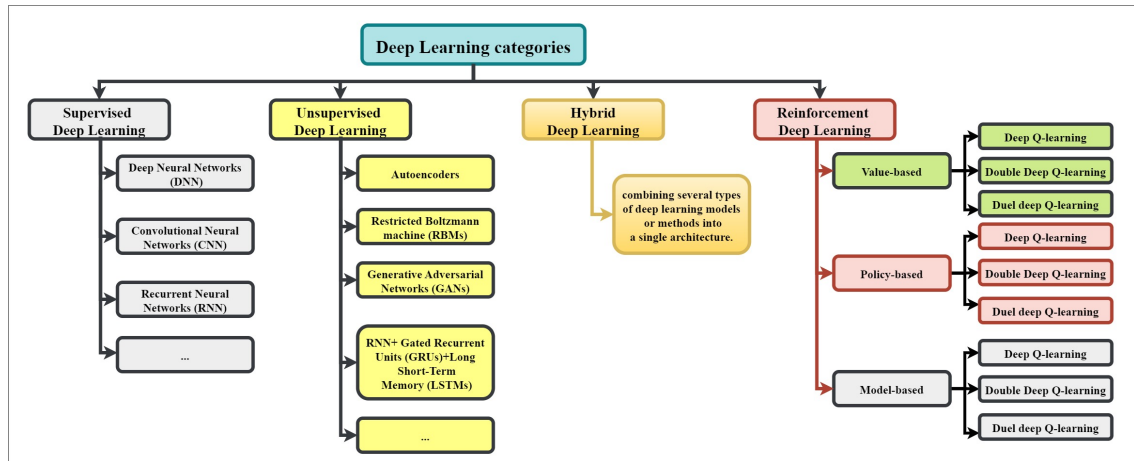


Fig. 4: Taxonomy of Deep Learning models.

Supervised Deep Learning

Supervised deep learning involves labeled data being used to train a model, with each data point assigned a label or target value. The model learns from labeled data by adjusting internal parameters to reduce the difference between expected and actual labels. The ultimate aim of supervised learning is for the model to generalize and predict the labels of previously unknown data points. A second test data set is used to assess the performance of the trained model. This test data set contains information that was not used during training but has known true labels. Several supervised learning approaches are available for DL, including recurrent neural networks (RNNs), convolutional neural networks (CNNs), and deep neural networks (DNNs). In addition, the RNN category covers techniques such as gated recurrent units (GRUs) and long short-term memory (LSTM).

Unsupervised Deep Learning

Unsupervised learning emerges as a valuable technique when labeled data is scarce or unavailable. In such scenarios, algorithms aim to identify inherent patterns or relationships within the data without relying on predefined labels. The primary focus is on acquiring essential features or representations in the absence of explicit labels, contributing to the exploration of underlying structures or relationships within the input data. A variety of techniques are utilized in unsupervised learning, encompassing generative networks like Generative Adversarial Networks (GANs), dimensionality reduction methods such as auto-encoders (AE) and restricted Boltzmann machines (RBM) known as Gibbs distribution, as well as clustering algorithms [26]. These approaches play a crucial role in diverse tasks like data clustering, dimensionality reduction, and anomaly detection. Despite the advantages associated with unsupervised learning, particularly concerning clustering – a method grouping similar data points based on specific criteria – it has its limitations. These include the incapacity to provide precise information about data categorization and the inherent computational complexity.

Hybrid Deep Learning

In this method, the learning process relies on datasets that are only partially labeled. Occasionally, both generative adversarial networks (GANs) and DRL are applied similarly to this approach. One of the benefits of this approach is its capacity to reduce the necessity for a large amount of labeled data. On the flip side, a drawback of this technique is that irrelevant input features in the training data may lead to inaccurate decisions.

Reinforcement Deep Learning

Reinforcement learning (RL) is the area focused on decision-making through the acquisition of effective behavior in an environment with the goal of maximizing rewards. This learned optimal behavior is a result of interactions with the environment. Within RL, an agent possesses the ability to make decisions, observe their consequences, and adjust strategies to develop an optimal policy. In the initial stages, the agent observes the current state, takes actions, and is rewarded along with the updated state. In this process figure 5a, the immediate reward and the new state hold the potential to influence adjustments to the agent’s policy. This iterative sequence continues until the agent’s policy gradually approaches proximity to the optimal policy. DRL (Deep Reinforcement Learning) addresses the primary limitations of RL, such as extended processing time required to achieve an optimal policy, thereby introducing new possibilities within the DRL framework. Broadly depicted in figure 5b, DRL leverages the characteristics of deep neural networks to enhance the learning process, resulting in improved speed and algorithm performance. In DRL, deep neural networks maintain the internal policy of the agent during interactions with the environment, determining the next action based on the current state. DRL encompasses three main methods: value-based, where the agent learns state or state-action values and acts based on the best action in the given state; policy-based, which aims to discover an optimal policy, whether stochastic or deterministic, for better convergence in high-dimensional or continuous action spaces; and model-based, which focuses on learning the functionality and dynamics of the environment from previous observations, often using a specific model. In value-based and policy-based methods, exploration of

the environment is a crucial step. Model-based DRL involves updating the model and replanning the process. Instances of model-based DRL include imagination-augmented agents, model-based priors for model-free approaches, and model-based value expansion. While these methods are efficient when a model is available, challenges may arise with large state spaces. In such cases, the model in model-based DRL is frequently updated, and the process is replanned to achieve optimal outcomes [26].

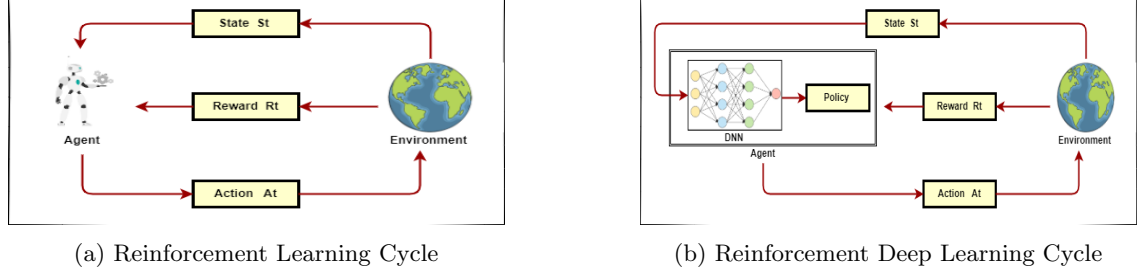


Fig. 5: Comparison of Reinforcement Deep Learning Cycle and Reinforcement Learning Cycle

3.2 Medical Images Analysis

In the field of medical image analysis, computer vision tasks such as image classification, medical image segmentation, and object detection and recognition are prevalent. Deep learning (DL) approaches, particularly those based on convolutional neural networks (CNN), have shown superior performance in handling these challenging problems across various medical applications. In this section, we provide a comprehensive and technical evaluation of current research that leverages the latest advancements in DL and CNN-based algorithms to advance medical image analysis and understanding.

3.2.1 Object detection

Object detection has garnered significant interest among researchers in the past few decades. With the remarkable progress in deep learning techniques, the integration of artificial intelligence 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 within an image. This technique utilizes computer vision 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.

The process of object detection involves bounding box annotation, where objects are outlined with bounding boxes. The task revolves around 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 activity, and other medical conditions.

In the field of dentistry, several models have been developed to identify various dental features, such as tooth decay detection [27–30], periapical lesion detection [31], and dental plaque detection [32]. Common architectures utilized by researchers in the reviewed papers include Faster R-CNN [33–35], SSD [36], and YOLO [27, 37, 38]. 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. 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.

to minimize bias and ensure an objective understanding [48]. The objective of this SLR was to identify and analyze research trends, methodologies, datasets, and frameworks pertaining to deep convolutional neural network (CNN) approaches for dental segmentation.

4.2 Search and Selection Process

The literature review process involved several steps, as depicted in Figure 7, outlining the study selection procedure for the SLR. Initially, all articles relevant to deep CNN-based approaches for dental segmentation were identified through a preliminary screening using specific search strings. Subsequently, these articles were subjected to inclusion and exclusion criteria. Following that, data extraction was conducted based on predefined criteria, and pertinent information was extracted from the selected articles. Finally, the collected data underwent analysis.

Out of a total of 642 articles initially identified, 45 articles 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 the existing research on deep CNN-based approaches for dental segmentation.

4.2.1 Initial Search

Before commencing the search process, we conducted an initial exploration to ensure the presence of a sufficient number of articles in the targeted field. At this stage, we identified 45 articles discussing the enhancement of dental imaging segmentation using deep Convolutional Neural Network (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 appropriate search strings to capture a broad spectrum of related articles. We utilized five online databases: IEEE Xplore Digital Library [49], ACM Digital Library [50], Wiley online library [51], ScienceDirect [52], and SpringerLink [53].

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 with the requirements of each search engine. The search command employed was: **(("dental radiography" OR "dentistry") AND ("Deep CNN based") AND ("segmentation"))**. The manual search was conducted separately in each database, and the results were amalgamated into a comprehensive spreadsheet.

With the outlined search strategies, we identified a total of 642 relevant papers. The distribution across 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 proceeded to the inclusion and exclusion process (as detailed in Section 3.2).

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	Wiley Online Library	81
Total		642

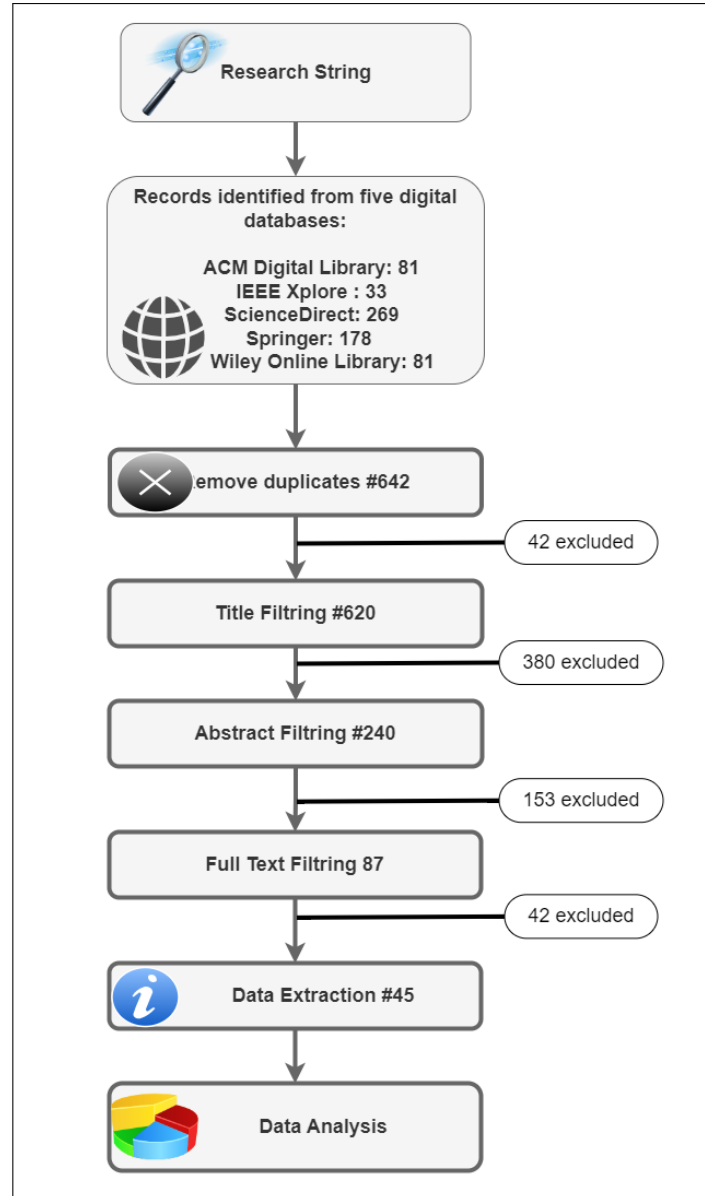


Fig. 7: The systematic search and selection process.

4.2.3 Selection of studies

The study selection for our systematic literature review involved the application of predefined inclusion and exclusion criteria to the gathered data. Inclusion criteria encompassed papers published within a 10-year timeframe (2014-2023), written in English, and specifically addressing plant disease detection using deep Convolutional Neural Networks (CNN). 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 a set of relevant studies for subsequent analysis and synthesis.

The selection process for relevant studies 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 predefined inclusion/exclusion criteria. Subsequently, screening extended to the abstracts of the

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

papers, where those meeting the inclusion criteria were included, and those not meeting the criteria were excluded. Finally, the same process was reiterated, this time utilizing 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 the texts from alternative sources. If the full text could not be obtained, the paper was excluded from the study.

5 Aims and Research Questions

The formulation of research questions can be organized based on four elements, known as PICOC (Population, Intervention, Comparison, Outcomes, and Context). Population refers to the targeted group of the research. Intervention is a detailed aspect of the research or issues that attract researchers. Comparison is defined as the aspect of the research where the intervention will be compared. Outcomes are the results of the intervention, and Context is the environment of the research [54]. Before starting the search, we determined the research questions and formed the search string. According to Table 4, seven research questions are proposed for this study.

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 that were selected for this SLR and shows the trend of studies in this area in recent years.
RQ2	What was the key motivation for applying deep learning for dental image analysis?	Analyze the different problems addressed by deep learning algorithms in the field of dental imaging.
RQ3	What deep learning implementation frameworks were used?	Understanding the technical tools and frameworks utilized in a study or project is crucial, as it provides a comprehensive insight into the realm of deep learning.
RQ4	What deep learning algorithms were applied?	To identify widely utilized deep learning models in dentistry, examining their applications, strengths, and potential contributions to diagnostics, treatment planning, and overall patient care. .
RQ5	What evaluation metrics are most frequently used to assess a deep Convolutional Neural Network segmentation model?	Display the key ML or DL performance measures so that you may select the most appropriate ones for your use case.
RQ6	How can future studies build upon 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 research. Acknowledging these gaps creates a platform for upcoming studies to confront and surmount these challenges.
RQ7	What 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.

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 roster of the 45 primary studies incorporated into this Systematic Literature Review is delineated in table 10.

6.1 The Current Landscape of Dental Radiography Segmentation

The 45 included studies are only in the form of journal articles. To associate the articles with their country of origin, we analyzed the affiliations of the authors of the included articles, ranking

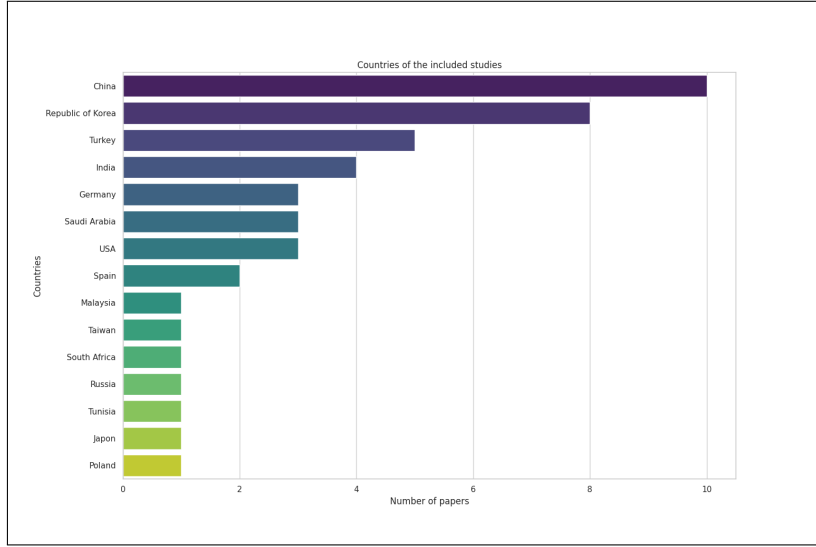


Fig. 8: Countries of the included studies.

highlights the prominence of Asian countries, with China leading with 10 articles, followed by Korea with 8 articles, and India with 4 articles. Following closely are Turkey with 5 articles and the USA with 3 articles. The list also features European countries, including Germany with 3 articles and Spain with 2 articles. Additionally, other represented countries include Saudi Arabia with 3 articles, and from the Asian continent, Malaysia, South Korea, and Taiwan with 1 article each. From the African continent, South Africa and Tunisia are represented with 1 article each.

Figure 9 provides an overview of the yearly growth rate of publications in this field of study. An increase in the number of publications can indicate a growing interest in the field, while a decrease could suggest that the field is a closed research area. Notably, there is a limited representation of papers from 2023, possibly due to 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 observe 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 in 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, thereby advancing the field of dental image processing. Radiographic 2D and 3D dental images play a pivotal role in addressing a wide spectrum of dental issues.

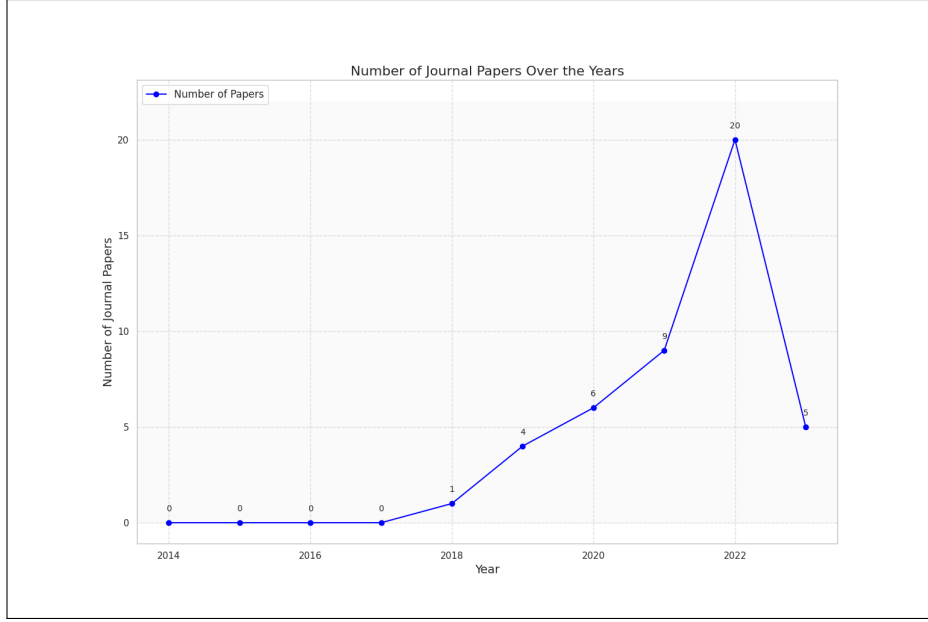


Fig. 9: Yearly papers on CNNs for dental X-ray segmentation.

In Table 5, we endeavor to outline the objectives and diseases addressed by the 45 articles in our survey. The tooth numbering process assigns a unique number to each tooth using the universal

Basic Detection and Identification	
Tooth detection and numbering	[34, 35, 55–57]
Detection of Periapical Lesions and segmentation of cysts	[31, 37, 58–60]
Structural Analysis and Classification	
Age and Sex determination	[61–64]
Caries detection and segmentation	[27–30, 36, 65–67]
Dental calculus identification	[32, 68]
Detection and classification of dental structures	[33, 42, 43, 69–75]
Advanced Condition Detection	
Detecting dental conditions	[38, 41, 76–80]

Table 5: Automatic dental imaging segmentation purpose.

tooth numbering system. In a dental panoramic image, encompassing molars, premolars, canines, and incisors in both maxilla and mandible, the maximum tooth count reaches 16 (Figure 6a). The distribution includes 6 molars, 4 premolars, 4 canines, and 8 incisors in each jaw.

In summary, recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly impacted dental image processing. 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 diagnosis of various dental conditions. The studies highlighted demonstrate the effectiveness of CNN-based models, achieving high precision and recall rates.

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

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

6.3 Popular Deep Learning Frameworks

Continuous efforts from both the industry and the academic community have led to the development of several widely adopted deep learning frameworks. Deep learning frameworks serve as software libraries or platforms designed to streamline the development, training, and deployment of deep neural networks. A plethora of software frameworks are at our disposal for implementing Deep Learning Models (DLM), and they undergo constant updates as fresh methodologies and concepts surface. There exists a multitude of deep learning frameworks, encompassing Theano, TensorFlow, Caffe, PyTorch, MXNet, MATLAB and many others [81].

Indeed, it is crucial to emphasize that there is no universal solution that can address every problem within the realm of machine learning. Quite often, achieving success necessitates the utilization of a combination of diverse tools and frameworks. Drawing from the results of our comprehensive literature review, illustrated in Figure 10, it becomes evident that Keras, PyTorch, and TensorFlow have surfaced as the dominant libraries in the realm of research. This prominence can be attributed to the fact that both Keras and TensorFlow platforms are open-source and readily accessible tools. Researchers tend to favor these alternatives due to their lack of subscription fees, a significant departure from MATLAB, where users often encounter expenses for access. In

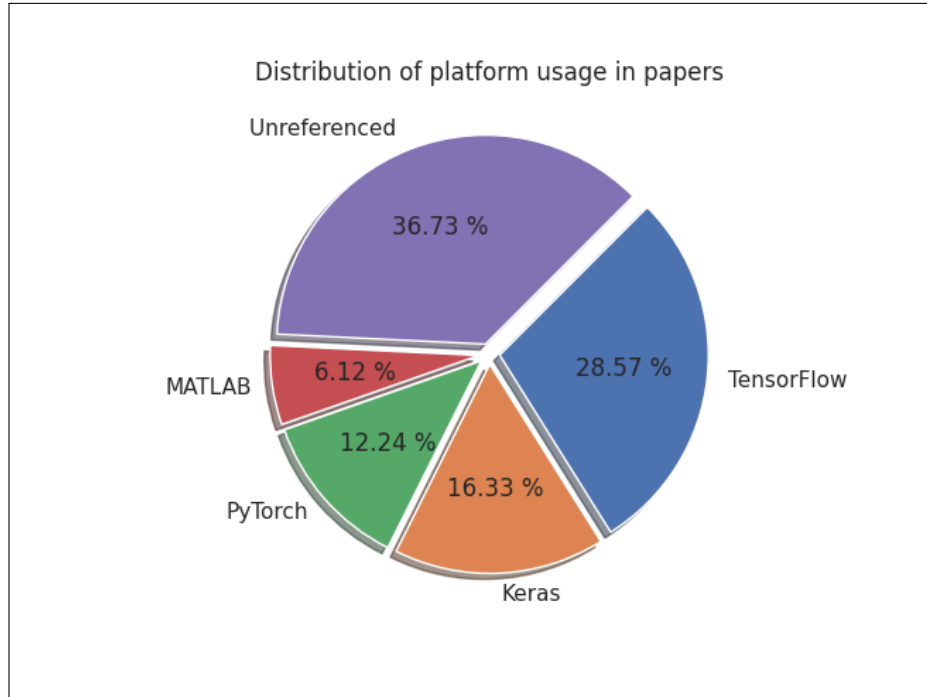


Fig. 10: The distribution of papers per implementation framework.

this section, we will examine various popular machine learning and deep learning frameworks and libraries.

6.3.1 TensorFlow

TensorFlow is an open-source software library designed for numerical computations using data flow graphs. It was created and is maintained by Google's Brain team in their Machine Intelligence research organization for machine learning and deep learning applications. This library is tailored for large-scale distributed training and inference tasks. In TensorFlow, nodes in the computational graph represent mathematical operations, while the edges represent the multidimensional

data arrays known as tensors that flow between these operations. The architecture of distributed TensorFlow includes master and worker services with kernel implementations.

TensorFlow, as the most widely adopted deep learning framework [82], boasts numerous unparalleled features. However, its development remains ongoing, addressing various shortcomings. Here are some of TensorFlow’s notable advantages:

- **Multi-language Support:** TensorFlow offers support for multiple programming languages, with the potential for future expansion to include even more languages.
- **High Performance:** TensorFlow delivers excellent performance and offers robust support for various hardware configurations, including multi-CPU, GPU, or hybrid platforms.
- **Real Portability:** TensorFlow exhibits a high level of portability, surpassing many other frameworks. Its close integration of research and product development, coupled with tools like TensorFlow Serving, enables seamless translation of innovative ideas into practical products.

6.3.2 Keras

Keras [83] is an open-source neural network library renowned for its simplicity and modularity. It is predominantly implemented in pure Python, making it accessible to a wide range of users. Keras was specifically designed to facilitate the process of experimenting with deep neural networks, and it offers a comprehensive toolkit for building and training these networks. Some key aspects of Keras include [84]:

- **Minimalist Framework:** Keras is celebrated for its minimalist framework, which reduces complexity and enables users to focus on the core aspects of deep learning model development.
- **Modularity:** Its modular design allows users to build neural networks by assembling standalone, fully-configurable modules. This modularity extends to layers, loss functions, activation functions, and more.
- **Ease of Experimentation:** Keras simplifies the experimentation process, making it easier to explore various deep learning architectures and ideas.
- **Compatibility:** Keras exhibits compatibility with multiple backends or deep learning frameworks, providing users with the flexibility to select the backend that aligns best with their specific requirements and preferences. This adaptability allows users to seamlessly integrate Keras into their existing infrastructure and leverage the strengths of different frameworks for their deep learning projects.
- **Versatility:** Keras supports Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and other deep learning models. It accommodates multiple input and output structures, making it versatile for various tasks.

6.3.3 PyTorch

Facebook, Twitter, and Yandex have employed the Torch framework, available in two iterations: the traditional version based on the LuaJIT language, and the Python-based variant known as PyTorch, launched by Facebook in January 2017. PyTorch [85], functioning as a Python package, offers two prominent features: GPU-accelerated tensor calculations and the ability to construct dynamic neural networks. These features distinguish PyTorch as a versatile and powerful tool in the realm of deep learning.

In contrast to many other frameworks such as TensorFlow, which exhibit a primarily static nature necessitating users to build and reuse fixed network structures [82], PyTorch stands out with its dynamic approach. It leverages Reverse-mode auto-differentiation, allowing users to modify network performance with minimal delay or computational overhead. While this technology is not exclusive to PyTorch, it boasts one of the most rapid and efficient implementations, establishing PyTorch as a prominent player among deep learning frameworks

6.3.4 Caffe

Caffe [86], developed by the University of California, Berkeley Vision and Learning Center (BVLC), is a convolutional neural network framework designed around a C++/CUDA/Python architecture. Jia Yangqing is credited as its creator [87]. Built upon feed-forward convolutional neural network architectures, such as CNNs, Caffe offers multiple interfaces for command line, MATLAB, and Python. Notably, it excels in feature extraction and seamlessly switches between CPU and GPU, harnessing GPU acceleration for accelerated CNN learning. On April 18, 2017, Facebook introduced an enhanced iteration of Caffe known as Caffe2, operating under the BSD license agreement. With its expanded feature set, Caffe2 has the potential to surpass Caffe in popularity, emerging as a more prominent deep learning framework in the future.

6.3.5 MXNet

In November 2016, Amazon introduced MXNet [88], an open source, lightweight, portable, and flexible deep learning library developed by DMLC (Distributed Machine Learning Community), which allows users to mix symbolic programming mode and command programming mode to maximize efficiency and flexibility. It quickly gained prominence as one of the most performant frameworks. Following its endorsement, Amazon expanded its services to include text-to-speech recognition and an image analysis service called Recognition. MXNet is officially recommended by Amazon Web Services (AWS) and is notably the first framework to offer support for multi-GPU and distributed training, distinguishing it from its peers. What sets MXNet apart is its wide range of language bindings, making it accessible to developers using C++, Python, R, Julia, MATLAB, and JavaScript. MXNet represents a powerful tool in the realm of deep learning, offering efficiency, flexibility, and versatility to its users.

It is also noteworthy that many research articles did not mention the specific implementation platform used. It is essential to provide comprehensive details regarding the implementation or to include a link to the code. Unfortunately, some articles overlooked this aspect. We recommend that researchers not only specify the deep learning implementation platform but also delve into other technical aspects employed in their model development when writing their papers for academic journals.

6.4 Applied Deep Learning Algorithms

In the background section, we have described Convolutional Neural Networks (CNNs), which are highly relevant for deep learning. Below, we will discuss the most commonly used types of deep learning networks that have been employed in the reviewed studies, based on the publications we have selected.

Table 6: Summary of Pre-Trained Models

Model	Year	Architecture	Main contribution	Limitations
LeNet [89]	1998	Two convolutional (conv) layers for feature extraction. Two sub-sampling layers for spatial dimension reduction. Two fully connected layers for a global understanding of the input. An output layer featuring a Gaussian connection, suggesting the use of either a Gaussian activation function or Gaussian-weighted connections.	The recognition of handwritten digits.	<ul style="list-style-type: none">• Inadequate adaptation to a variety of image classes.• Utilization of over-sized filters• Limited capability in low-level feature extraction

VGG (Visual geometry group) [90]	2014	Multiple layers of small 3x3 convolutional filters, followed by max Pool Layers. The fully connected layers use ReLU activation and dropout regularization to prevent overfitting.	The increase in the depth of a network (number of layers) is a critical factor in improving performance.	<ul style="list-style-type: none"> • Incorporation of computationally demanding fully connected layers.
GoogLeNet (Inception-V1) [91]	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 traditional convolutional layers with these blocks, incorporating a 1x1 convolutional filter bottleneck for computation regulation. Sparse connections and global average pooling reduce redundancy, cutting parameters from 138 million to 4 million.	The development of an inception module, which significantly reduces the network's parameter count, leading to improved computational efficiency.	<ul style="list-style-type: none"> • Cumbersome parameter customization owing to a heterogeneous topology • Potential loss of valuable information due to a representational bottleneck
Inception V3 [92]	2015	The architecture incorporates factorized 7x7 convolutions, spatial factorization using asymmetric convolutions, and an auxiliary classifier for label information propagation. Additionally, batch normalization and ReLU activation functions are applied. With around 23 million parameters	Addresses the representational bottleneck by replacing large size filters with smaller ones.	<ul style="list-style-type: none"> • Elaborate architecture design • Absence of homogeneity
Faster R-CNN [93]	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 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)	<ul style="list-style-type: none"> • Complexity • Training Time • Detection Time • Resource Intensive-ness • Model Complexity

U-Net [94]	2015	A U-shaped encoder-decoder network architecture, consisting of four encoder blocks and four decoder blocks connected via a bridge. Each encoder block involves two 3x3 convolutions, followed by batch normalization and a ReLU activation function. The decoder network employs transposed convolutional layers for upsampling, aiming 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.	<ul style="list-style-type: none"> • Sensitivity to Input Size • Hardware Memory Requirements
ResNet (Residual Network) [95]	2015	ResNet utilizes residual blocks, which are stacked together to form the overall network. The original ResNet had 34 layers with 2-layer blocks, while advanced variants, like ResNet50, used 3-layer bottleneck blocks for improved accuracy and reduced training time. Skip connections in ResNet add outputs from previous layers to stacked layers, enabling the training of much deeper networks than was previously possible.	Residual layer.	<ul style="list-style-type: none"> • 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.

MobileNet [96]	2017	The architecture comprises convolution layers, depthwise convolution layers with BN and ReLU, pointwise convolution layers with BN and ReLU, Global Average Pooling, Reshape, Dropout, additional Convolutional layers, Softmax, and Reshape. With approximately four million parameters, MobileNet is designed to be lightweight. It features two basic units: 3x3 Convolution and 3x3 Depthwise Convolution followed by 1x1 Convolution, using separable filters to reduce input channel numbers.	MobileNet's architecture is optimized for mobile and embedded devices, utilizing depthwise separable convolutions. Sequential layers include depthwise and pointwise convolutions with Batch Normalization and ReLU.	<ul style="list-style-type: none"> • Challenges arise on resource-constrained micro-controllers. • Sacrifices in accuracy compared to larger architectures, especially in detailed tasks. • Inherent tradeoff in accuracy due to its small size, though often minimal.
DenseNet [97]	2017	DenseNet's architecture utilizes Dense Blocks with densely connected layers to maximize feature reuse. Each block includes an input, output, and a growth rate (k) that regulates information flow to the next layer. Transition Layers connect dense blocks using 1x1 convolutions, 1x1 pooling, and 3x3 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 information flow and network generalization. Finally, the Output Layer is typically a fully connected layer mapping input features to the desired class labels.	Densely connected architecture	<ul style="list-style-type: none"> • Memory Usage
YOLO (You Only Look Once) [98]	2016	YOLO uses a fully convolutional neural network that passes the image through a series of convolutional layers to extract features. It predicts object boundaries and class labels simultaneously, making it a powerful and efficient deep learning model for object detection tasks.	Real-time object detection	<ul style="list-style-type: none"> • Detecting small or complex-shaped objects and accurately handling very large objects

Mask RCNN [99]	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, providing pixel-level masks for accurate object delineation.	RoIAlign layer.	<ul style="list-style-type: none"> • False alerts • Missing labels
----------------	------	--	-----------------	--

6.5 Evaluation Metrics for Deep CNN Segmentation in Dental Imaging

The tools or mechanisms for measuring model quality are known as evaluation metrics, a crucial component in developing effective ML or DL models. These metrics vary based on tasks, applications, and models. In the context of dental image segmentation, we outline the evaluation metrics employed in the selected papers.

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

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

Table 7: Confusion Matrix

- **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.

Table 8: Performance Evaluation Metrics.

Metric	Formula	Description	Study
Accuracy (Pixel Accuracy (PA))	$\frac{\text{Number of Correct Predictions}(TP + TN)}{\text{Total Number of Predictions}(TP + TN + FP + FN)}$ $= \frac{\text{Number of correctly classified pixels}}{\text{Total number of pixels}} * 100\%$	Is one way to determine how frequently the algorithm correctly classifies a data point. This is the number of items correctly identified as true positives or true negatives out of the total number of items.	[27, 29, 30, 37, 38, 41-43, 55, 56, 58, 61-72, 74, 75, 77, 79, 100-102]
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.	[28, 33-35, 38, 41-43, 57, 61, 64-67, 70, 75, 76, 78, 79, 100, 101, 103]
Recall [(Sensitivity) (True Positive Rate TPR)]	$\frac{TP}{TP + FN}$	The 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.	[27-30, 34, 35, 37, 38, 41-43, 55, 56, 58-61, 64-67, 69, 70, 73, 75, 76, 78, 79, 100, 101]
Specificity	$\frac{TN}{FP + TN}$	This is referred to as the true negative rate. This function computes the proportion of actual negative cases predicted as negative by our model.	[27, 29, 30, 33, 37, 41, 42, 55, 56, 58-61, 64, 67, 69, 70, 73, 79, 101]
F1-score	$2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$	Measures the effectiveness of identification when recall and precision are of equal importance.	[28, 29, 33, 35, 37, 38, 41-43, 57-59, 65, 75, 76, 79, 100, 102]
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.	[41, 62, 100, 102]
Dice similarity coefficient (DSC)	$\frac{(2 * \text{area of overlap})}{(Total number of pixels in both images)}$	Used to assess how well a predicted segmentation matches the corresponding ground truth in terms of pixel-level agreement.	[31, 37, 56, 58, 64, 65, 69, 70, 77]

Receiver operating characteristic (ROC)	Doesn't have a specific formula	ROC curve is graphical display of sensitivity (TPR) on y-axis and (1 - specificity) (FPR) on x-axis for varying cut-off points of test values.	[30, 36, 73, 101]
Area under the curve (AUC)	$AUC = 1 - \frac{1}{2} \left(\frac{FP}{FP+TN} + \frac{FN}{FN+TP} \right)$	As a single scalar value, 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.	[27, 30, 34, 43, 57, 60, 63, 66, 68, 73, 78, 101, 103]
Average precision (AP)	$AP = \int_{r=0}^1 p(r) dr$ mean Average precision (mAP) $mAP = \frac{1}{N} \sum_{k=1}^K AP_k$		
Intersection over Union (IoU)	$\frac{Area_{pred} \cap Area_{gt}}{Area_{pred} \cup Area_{gt}}$	Used to evaluate the performance of object detection by comparing the ground truth bounding box to the predicted bounding box.	[32-34, 37, 56, 58, 71, 74, 77, 78, 80]
Jaccard index (JI)	$= \frac{TP}{TP+FP+FN}$	The MCC is used for binary classification and is considered particularly useful in unbalanced class situations. takes values between 1 (perfect correlation between ground truth and predicted outcome) and -1 (inverse or negative correlation) - a value of 0 denotes a random prediction.	[29] [37] [41]
Matthews correlation coefficient (MCC)	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}}$		

In the field of evaluating dental image segmentation models, it is important to note that, in addition to the mentioned metrics, some papers also utilize specific measures such as Volumetric Overlap Error (VOE) and Relative Volume Difference (RVD) [101]. These measures provide a thorough assessment of model performance, especially when applied to volumetric tomographic images like Cone Beam Computed Tomography (CBCT). Volumetric Overlap Error (VOE) measures the degree of mismatch between segmented regions 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.

As shown in Figure 11, and according to the papers studied we found that the 4 evaluation metrics used to assess the proposed models are Recall, Accuracy, Precision and Specificity.

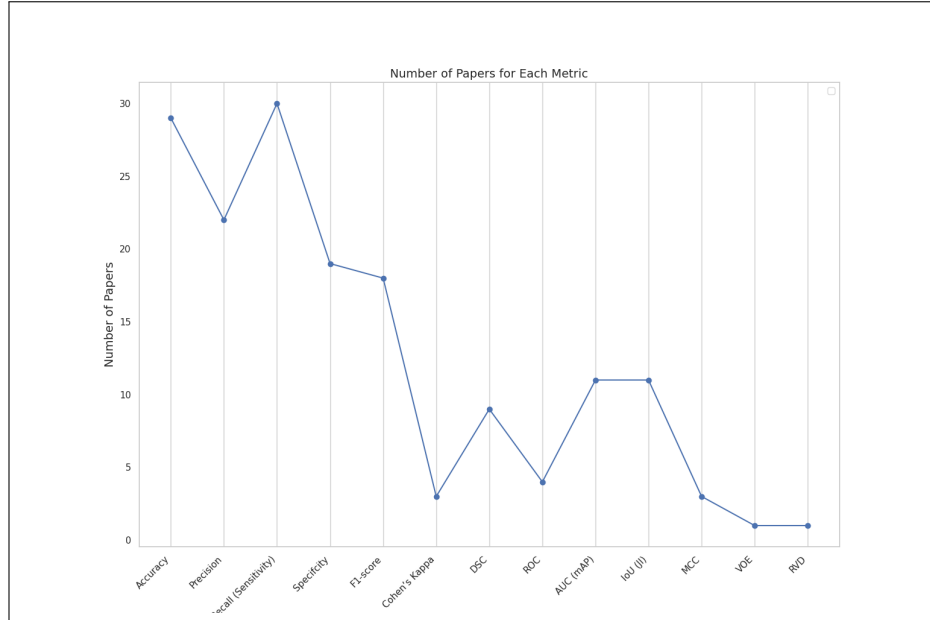


Fig. 11: The summary of Performance evaluation measures reviewed in this SLR.

Specificity and Sensitivity

Specificity and sensitivity are standardized and recognized measures for evaluating performance in medical image segmentation. Sensitivity, also known as recall or true-positive rate, focuses on the capabilities of detecting true positives in pixel classification, while specificity, also known as the true-negative rate, evaluates the capabilities of correctly identifying true-negative classes (such as the background class). 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 due to the large percentage of pixels labeled as background compared to the region of interest (ROI). Therefore, specificity is a good statistic for checking the model's operation, but not so much for its performance.

Accuracy / Rand Index [104]

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, accuracy may not be the best measure of performance

[41, 60]. Measuring accuracy will always produce an unrealistic result due to the inclusion of true negatives. Even when predicting the segmentation of an entire image as a background class, accuracy rates are often greater than or close to 100%. Therefore, the misleading accuracy measure is not suitable for evaluating medical models based on the deep learning approach.

F-measure, IoU, and DSC

The F-measure, often known as the F-score, is one of the most extensively used metrics for quantifying performance in computer vision. It is derived from the sensitivity as well as the precision of a prediction, which analyzes the overlap between the expected segmentation and the ground truth. However, by considering accuracy, it also penalizes false positives, which are prevalent in datasets with 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 sensitivity and precision. The IoU penalizes under- and over-segmentation more than the DSC. Even though these scores are suitable metrics and the most common metric in validating medical image segmentation [69], they are not employed for evaluating the performance of the suggested models in a 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. The performance is measured by comparing the true positive rate (TPR) to the false positive rate (FPR). In the medical field, ROC curves are widely accepted as a standard metric for comparing several classifiers when assessing diagnostic tests. Hanley and McNeil [105] proposed the area under the ROC curve (AUC) as a single-value performance assessment for diagnostic radiology. The AUC measure is becoming a popular tool for validating machine learning classifiers. It should be noted that an AUC has a value ranging from 0 to 1. A model with 100% incorrect predictions has an AUC of 0.0, while one with 100% correct predictions has an AUC of 1.0.

Other metrics

For a segmentation system to be helpful and make a significant contribution to the area, its performance must be systematically tested. Furthermore, the evaluation must be carried out using conventional and well-known measures that allow for fair comparison. Other metrics exist and can be used based on the research topic and interpretive emphasis of the study. We refer to the excellent studies of Taha et al. [106]. Additionally, Wang et al. [107] describe the application of metrics in the evaluation of natural, medical, and remote sensing images and present the supervised and unsupervised evaluation methods in the application.

6.6 Public Datasets for Dental Radiography Analysis

Dental studies heavily rely on diverse datasets for optimal performance. Addressing the last research question (RQ7), our investigation, as summarized in Table 10, underscores that the majority of studies predominantly utilized private datasets. Remarkably, only three studies opted for public datasets. Among the dataset types, panoramic radiographs emerged as the most frequently employed, providing a comprehensive single-film view of the maxilla and mandible, crucial for analyses such as orthodontic treatment and tooth growth. Additionally, a subset of studies utilized Cone-beam computed tomography (CBCT) for detailed three-dimensional (3D) imaging, while other types, including periapical radiographs and bitewing radiographs, offered vital diagnostic information for various dental conditions.

Our systematic literature review underscored a significant dearth of publicly available reference datasets suitable for extracting pertinent features related to oral medical conditions. Only three datasets were identified in the public domain [108], [109], and [110]. Recognizing the pivotal role of data in influencing the performance of machine learning (ML) and deep learning (DL) models,

the exploration of publicly available datasets, even in limited quantity, offers immense potential to enhance the robustness and generalizability of deep learning models in dentistry.

In the context of this scarcity, we would like to provide more detailed insights into specific publicly available reference datasets, each contributing uniquely to dental research:

- The MLUA dataset, designed for dental caries segmentation in oral panoramic images, provides full-size X-ray images, corresponding segmentation masks, cropped images at tooth borders, and bounding box annotations. Accessible at <https://github.com/Zzz512/MLUA>, this resource supports research on accurate lesion detection, and its code and test samples are available for implementation.
- Panetta et al [111], introduces The Tufts Dental Database comprises 1000 panoramic dental radiography images with expert annotations, facilitating classification based on various criteria. Notably, it includes the radiologist’s expertise through eye-tracking and think-aloud protocol recordings. This dataset serves as a benchmark for state-of-the-art systems, aiming to propel advancements in AI-driven abnormality detection, tooth segmentation algorithms, and the integration of radiologist expertise into AI applications.
- Dental Radiography Dataset, packed in a zip file, includes 1269 .jpg images along with three accompanying CSV files. Access the dataset at <https://www.kaggle.com/datasets/intkaggleteam/dental-radiography/data>.
- the authors [112] collected dental panoramic radiographs from 106 pediatric patients (2 to 13 years old). Using EISeg and LabelMe, they created the world’s first dataset for caries segmentation and dental disease detection. Additionally, 93 pediatric radiographs were combined with three adult datasets (2,692 images) for a comprehensive segmentation dataset suitable for deep learning
- Lopez et al [113] introduces a database comprises 598 panoramic radiographs with dimensions of 2041 x 1024, available in JPEG format. The annotations for these panoramic radiographs are accessible at <https://www.kaggle.com/datasets/humansintheloop/teeth-segmentation-on-dental-x-ray-images>

By highlighting these specific datasets, despite their limited number, we emphasize their critical importance in developing more reliable and applicable models. Access to these diverse datasets from different centers continues to play an essential role in cross-center validation, ensuring the models’ performance in various clinical contexts. This approach, encouraging cross-center collaboration and promoting generalization, remains essential to advance the field of deep learning in dentistry.

6.7 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 focusing on the segmentation of dental panoramic images. By highlighting the current deficits, our objective is 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 will briefly present the main identified gaps and envisioned avenues for future research. The prevalent gaps identified in the literature review entail challenges linked to diverse facets. These challenges are delineated in the subsequent table 9.

Table 9: Challenges, Gaps and Future Work

Aspects	Limitations and studies	Improving Solutions and Charting Future Paths
Dataset Size and source	<p>Dataset size is pivotal in deep learning model development as 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 present in real-world scenarios.</p> <ul style="list-style-type: none"> – The dataset used was relatively small [27, 28, 32, 35, 37, 42, 72]. – To assess tooth-level or case-level Residual Bone Loss (RBL) for a large cohort, the model requires more images for further optimization.[77] – The dataset is collected from a single organization or single device , limiting its diversity [28, 36–38, 59, 69, 73, 74, 76]. – Absence of external test group [29, 76]. – Heterogeneity in Image Sources [102] or Distribution of Heterogeneous Images [41]. – The data collected for the study were unbalanced due to high labeling costs. [80] 	<ul style="list-style-type: none"> –Employing data augmentation methods [27, 62]. – Exploring Fine-tuning with transfer learning [31, 33]. – Enriching the dataset with diverse data from multiple global sites and from various devices [36, 74]. – Exploring the integration of non-imagery data, such as patient history, clinical signs, and symptoms, with imagery data within a deep learning system [37, 38, 41]. – Seek external validation of the neural network on diverse test sets to assess generalizability. <p>Amasya et al. (2023)[102] showcased DiagnoCat AI's adaptability across diverse image sources and configurations, independent of specific devices or imaging parameters https://diagnocat.com.</p>
Image Quality	<p>The image quality defined by attributes such as clarity, sharpness, and minimal noise.</p> <ul style="list-style-type: none"> – 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 [42, 77]. – Capturing various aspects of a tooth and the difficulties associated with internal tooth structure and the interproximal tooth surface in intraoral photography [43, 56, 66, 102]. – Handling Incorrectly Oriented Radiographs [56]. – CBCT provides high resolution, but its accuracy in density measurement is inferior to traditional CT [71] 	<ul style="list-style-type: none"> – Explore methods for preprocessing or filtering poor-quality images before analysis. – Improve the model's robustness to handle variations in image quality. – Conduct an investigation into a more realistic simulated X-ray acquisition process [75].

Data Processing and Labeling	<p>Refines raw data through operations like resizing and normalization, optimizing it for effective model training. Simultaneously, data labeling assigns categorical or numerical tags to each data point, providing crucial ground truth for supervised learning and enhancing the model's ability to generalize and make accurate predictions.</p> <ul style="list-style-type: none"> – 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 [71, 74]. – In the benchmark dataset, certain categories are labeled more coarsely than others [70]. – Complexities in Labeling: especially when premolar teeth are missing without clear spacing. [103] 	<ul style="list-style-type: none"> – Investigate automated or semi-automated methods for mask generation. Explore unsupervised learning techniques for region identification [74]. – The ambiguity in bounding box boundaries persists in object detection [33].
Not a Replacement for Clinical Data	<p>While advantageous, deep learning models are not designed to replace the vital clinical data and expertise provided by healthcare professionals.</p> <ul style="list-style-type: none"> – The resulting model has not been used in a clinical setting, and it is not designed to replace periodontal charting and other clinical data.[29, 38, 63, 77]. – The study acknowledges a low success rate in identifying dental caries and calculus, making the AI models impractical for clinical use [76]. – – 	<ul style="list-style-type: none"> – Explore seamless integration of the model's predictions with periodontal charting and data.

Specific limitation	<p>Additional factors that impact the development, deployment, and effectiveness of deep learning models.</p> <ul style="list-style-type: none"> – The model faces challenges in accurately identifying tooth numbers or positions, particularly when multiple teeth are missing [77]. – The model faces difficulties in precisely localizing specific dental issues.[36, 74] – Increased Complexity and Training Time [80]. – The study did not compare the diagnostic performance of CNNs with that of dentists [27]. – Image Dimensions and platform Memory Constraints (Google Colab) [41, 57]. – Overfitting [62]. – 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 [63]. – Exemplary Use-Case Focus [59]. – Determining the estimation success of each type of tooth (incisors, canines, premolars, molars) [35]. – 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 Convolutional Neural Networks (CNNs) not considering spatial relationships between image features [34]. 	<p>Distinguishing Implants with Different Internal Connection Types [75] .</p> <ul style="list-style-type: none"> – Understanding the trade-offs between image resolution and memory constraints [41]. – Dropout [62]. – Investigating the impact of lower radiation doses on AI performance in dental imaging [35]. – 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 [31].
---------------------	--	--

7 Limitations of the Study

In this systematic literature review (SLR), various deep learning-based techniques for dental image segmentation were identified. However, there are a few limitations of our SLR that should be taken into consideration. These limitations are listed as follows:

- Only journal articles discussing dental image segmentation were included in this SLR. Through our search technique, several unrelated research publications were discovered and excluded from this review during the preliminary stages of the study. This ensured that the selected research papers aligned with the investigation’s requirements. However, it is believed that including additional sources such as books or conference papers could have enhanced the review.
- This review is limited to papers available in the English language.
- It is possible that other digital libraries with relevant studies were overlooked, although digital databases were considered when reviewing the study articles.
- The effect 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

Over the past decades, there has been a notable increase in research interest at the intersection of deep learning and oral healthcare analysis, leading to continuous advancement in innovative models and methodologies year after year. These algorithms have the capability to assist dentists in various aspects of their clinical practice, including precise tooth identification, detection and categorization of dental pathologies, assessment of prosthetic structures, and error reduction. Deep

learning algorithms hold considerable potential for streamlining the clinical workflows of dentists, thereby resulting in improved treatment outcomes.

The present study conducted a meticulous review of noteworthy articles focusing on deep learning methodologies applied to dental image segmentation. We systematically categorized these recent publications based on their specific objectives and conducted an in-depth exploration of the predominant evaluation metrics commonly used. Additionally, we addressed some of the significant challenges that influence the performance of these models. We anticipate that this study will provide readers with a comprehensive understanding of the essential dimensions of this field, highlight the most significant advancements, and shed light on potential avenues for future research.

In our forthcoming research endeavors, we aspire to capitalize on the insights gained from this rigorous systematic literature review (SLR). Our primary goal is to design a novel algorithm for dental pathology segmentation by harnessing the capabilities of deep learning. Importantly, we will also prioritize the interpretability of deep learning methodologies to mitigate their inherent opacity. Given the dynamic nature of this field, we envisage that this article will take the form of an ongoing review, with regular updates to reflect the emergence of new publications in this domain.

Declarations

- **Conflicts of Interest**

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

- **Ethical Approval**

Not applicable

- **Availability of supporting data**

Data will be made available on request.

- **Competing interests**

There are no conflicts of interest to report

- **Funding**

Not applicable

- **Authors' contributions**

- All authors conceptualized and wrote the main manuscript text.
- Imen jdey conceptualized all tables and figures.
- Walid Brahmi prepared all figures and tables .
- Adel Alimi reviewed the final version of the manuscript.
- All authors approved the final version of the manuscript.

- **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 Faculty of Science and Techniques of Sidi Bouzid, University of Kairouan, Tunisia.”

Table 10: list of reviewed studies.

Authors	Dataset characteristics	Data Augmentation?	Main objective& Architecture	Task	Performance
Brahmi et al[57], 2023	A public dataset consisting of 107 panoramic x-ray images[110].	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-score = 63% and precision = 96%
Choi et al[100], 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 convolutional neural network (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 means (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[101], 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 [102], 2023	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 [103], 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.
Lee, C. et al [77], 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, A et al [67], 2022	Private dataset. 380 periapical images.	No	A score-based multi-input CNN ensemble (ML-DCNNE) for the automatic diagnosis of dental caries	classification	accuracy = 99.13, sensitivity = 98%, specificity = 100%, precision = 100% and F1-score = 98.99%
Zhang, X. et al [36], 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%.
Oztek, F. et al [74], 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, J. et al [80], 2022	Private dataset. 1414 panoramic X-ray	No	a five-axis-based tooth recognition model using Faster R-CNN	Segmentation	IoU = 72%

Franco, A. et al [61], 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 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, Y. et al [27], 2022	Private dataset. 1000 bitewing radiographic divided into training (80%) groups 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, M. et al [69], 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, A. et al [75], 2022	Private dataset. 483 simulated X-ray images, which contain implants inserted into either human or pig jaws of size 512x512 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, N. et al [37], 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%.
Zhu, H. et al [65], 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, E. et al [66], 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, G. et al [41], 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 coefficient (MCC) = 51% multi-classification: accuracy = 59.42%, precision = 83%, recall = 70% and Matthews correlation coefficient (MCC) = 0.65
Jang, W. et al [33], 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, N. et al [62], 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, L. et al [38], 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, T. et al [56], 2022	A public repository contained 116 panoramic images [109]. 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, N. et al [63], 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, M. et al [73], 2022	Private dataset. 300 panoramic radiographs exported in jpeg file	No	A pre-trained Convolutional Neural Network (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, M. et al [42], 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%.
Sivasundaram, S. et al [58], 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, T. et al [68], 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, J. et al [59], 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%.
Basaran, M. et al [76], 2021	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, H. et al [78], 2021	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, S. et al [28], 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%.
Blgair, E. et al [35], 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, Q. et al [43], 2021	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, M. et al [64], 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, P. et al [55], 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, W. et al [32], 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
Kwak, G. et al [71], 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, Y. et al [70], 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, A. et al [29], 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, F. et al [31], 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, H. et al [34], 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, A. et al [79], 2019	Private dataset. 1500 panoramic images divided into training of 800 with testing of 7000	No	Histogram enhancement with a convolutional neural network (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, T. et al [60], 2019	Private dataset. A synthesized data set of 2001 tooth segments from panoramic radiographs	Yes	A deep convolutional neural networks (CNNs) to detect apical lesions (ALs).	segmentation	AUC = 0.85, sensitivity = 0.65 and specificity = 0.87
Tian, S. et al [72], 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, and 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, J. et al [30], 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	pre-molar accuracy = 89.0% , molar accuracy = 88.0% , both pre-molar and molar accuracy = 82.0% AUC = 0.917 on pre-molar, AUC = 0.890 on molar, and AUC = 0.845 on both pre-molar and molar models.

References

- [1] Mogli, G.: Medical records role in healthcare delivery in 21st century. *Acta Informatica Medica* **17**(4), 209–212 (2009)
- [2] Pickering, B.W., Gajic, O., Ahmed, A., Herasevich, V., Keegan, M.T.: Data utilization for medical decision making at the time of patient admission to icu. *Critical care medicine* **41**(6), 1502–1510 (2013)
- [3] Abedalla, A., Abdullah, M., Al-Ayyoub, M., Benkhelifa, E.: Chest x-ray pneumothorax segmentation using u-net with efficientnet and resnet architectures. *PeerJ Computer Science* **7**, 607 (2021)
- [4] Khanna, S.S., Dhaimade, P.A.: Artificial intelligence: transforming dentistry today. *Indian J Basic Appl Med Res* **6**(3), 161–167 (2017)
- [5] Schwendicke, F., Golla, T., Dreher, M., Krois, J.: Convolutional neural networks for dental image diagnostics: A scoping review. *Journal of dentistry* **91**, 103226 (2019)
- [6] Hwang, J.-J., Jung, Y.-H., Cho, B.-H., Heo, M.-S.: An overview of deep learning in the field of dentistry. *Imaging science in dentistry* **49**(1), 1–7 (2019)
- [7] Kang, D.-Y., Duong, H.P., Park, J.-C.: Application of deep learning in dentistry and implantology. *Journal of implantology and applied sciences* **24**(3), 148–181 (2020)
- [8] Prados-Privado, M., Villalón, J.G., Martínez-Martínez, C.H., Ivorra, C.: Dental images recognition technology and applications: a literature review. *Applied Sciences* **10**(8), 2856 (2020)
- [9] Heini, G., Jdey, I., Heni, A., Ltifi, H.: Hyperparameter optimization in customized convolutional neural network for blood cells classification. *J. Theor. Appl. Inf. Technol* **99**, 5425–5435 (2021)
- [10] Hinton, G.: Deep learning—a technology with the potential to transform health care. *Jama* **320**(11), 1101–1102 (2018)
- [11] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *nature* **521**(7553), 436–444 (2015)
- [12] Szeliski, R.: *Computer Vision: Algorithms and Applications*. Springer, ??? (2022)
- [13] Bharadiya, J.: A comprehensive survey of deep learning techniques natural language processing. *European Journal of Technology* **7**(1), 58–66 (2023)
- [14] Roger, V., Farinas, J., Pinquier, J.: Deep neural networks for automatic speech processing: a survey from large corpora to limited data. *EURASIP Journal on Audio, Speech, and Music Processing* **2022**(1), 19 (2022)
- [15] Yu, Z., Wang, K., Wan, Z., Xie, S., Lv, Z.: Popular deep learning algorithms for disease prediction: a review. *Cluster Computing* **26**(2), 1231–1251 (2023)
- [16] Egger, J., Gsaxner, C., Pepe, A., Pomykala, K.L., Jonske, F., Kurz, M., Li, J., Kleesiek, J.: Medical deep learning—a systematic meta-review. *Computer methods and programs in biomedicine* **221**, 106874 (2022)

- [17] Ahmad, A., Saraswat, D., El Gamal, A.: A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology* **3**, 100083 (2023)
- [18] Huang, J., Chai, J., Cho, S.: Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China* **14**(1), 1–24 (2020)
- [19] Zheng, Y., Xu, Z., Xiao, A.: Deep learning in economics: a systematic and critical review. *Artificial Intelligence Review*, 1–43 (2023)
- [20] Soori, M., Arezoo, B., Dastres, R.: Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics* (2023)
- [21] Muhammad, K., Ullah, A., Lloret, J., Del Ser, J., Albuquerque, V.H.C.: Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Transactions on Intelligent Transportation Systems* **22**(7), 4316–4336 (2020)
- [22] Narayan, S., Gardent, C.: *Deep Learning Approaches to Text Production*. Springer, ??? (2022)
- [23] Hazra, T., Anjaria, K.: Applications of game theory in deep learning: a survey. *Multimedia Tools and Applications* **81**(6), 8963–8994 (2022)
- [24] Xiao, F., Xue, W., Shen, Y., Gao, X.: A new attention-based lstm for image captioning. *Neural Processing Letters* **54**(4), 3157–3171 (2022)
- [25] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L.: Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data* **8**, 1–74 (2021)
- [26] Talaei Khoei, T., Ould Slimane, H., Kaabouch, N.: Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, 1–22 (2023)
- [27] Bayraktar, Y., Ayan, E.: Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clinical oral investigations* **26**(1), 623–632 (2022)
- [28] Lee, S., Oh, S.-i., Jo, J., Kang, S., Shin, Y., Park, J.-w.: Deep learning for early dental caries detection in bitewing radiographs. *Scientific reports* **11**(1), 16807 (2021)
- [29] Cantu, A.G., Gehrung, S., Krois, J., Chaurasia, A., Rossi, J.G., Gaudin, R., Elhennawy, K., Schwendicke, F.: Detecting caries lesions of different radiographic extension on bitewings using deep learning. *Journal of dentistry* **100**, 103425 (2020)
- [30] Lee, J.-H., Kim, D.-H., Jeong, S.-N., Choi, S.-H.: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry* **77**, 106–111 (2018)
- [31] Setzer, F.C., Shi, K.J., Zhang, Z., Yan, H., Yoon, H., Mupparapu, M., Li, J.: Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *Journal of endodontics* **46**(7), 987–993 (2020)
- [32] You, W., Hao, A., Li, S., Wang, Y., Xia, B.: Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health* **20**, 1–7 (2020)
- [33] Jang, W.S., Kim, S., Yun, P.S., Jang, H.S., Seong, Y.W., Yang, H.S., Chang, J.-S.: Accurate

- p detection for dental implant and peri-implant tissue by transfer learning of faster r-cnn: a diagnostic accuracy study.
- BMC Oral Health*
- 22**
- (1), 1–7 (2022)
- [34] Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., Lee, C.-H.: A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Scientific reports* **9**(1), 1–11 (2019)
 - [35] Bilgir, E., Bayrakdar, İ.Ş., Çelik, Ö., Orhan, K., Akkoca, F., Sağlam, H., Odabaş, A., Aslan, A.F., Ozcetin, C., Killi, M., *et al.*: An artificial intelligence approach to automatic tooth detection and numbering in panoramic radiographs. *BMC medical imaging* **21**, 1–9 (2021)
 - [36] Zhang, X., Liang, Y., Li, W., Liu, C., Gu, D., Sun, W., Miao, L.: Development and evaluation of deep learning for screening dental caries from oral photographs. *Oral diseases* **28**(1), 173–181 (2022)
 - [37] Moidu, N.P., Sharma, S., Chawla, A., Kumar, V., Logani, A.: Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clinical Oral Investigations* **26**(1), 651–658 (2022)
 - [38] Jiang, L., Chen, D., Cao, Z., Wu, F., Zhu, H., Zhu, F.: A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC Oral Health* **22**(1), 106 (2022)
 - [39] Ker, J., Wang, L., Rao, J., Lim, c.: Deep learning applications in medical image analysis. *Ieee Access* **6**, 9375–9389 (2017)
 - [40] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60–88 (2017)
 - [41] Alotaibi, G., Awawdeh, M., Farook, F.F., Aljohani, M., Aldhafiri, R.M., Aldhoayan, M.: 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 (2022)
 - [42] Aljabri, M., Aljameel, S.S., Min-Allah, N., Alhuthayfi, J., Alghamdi, L., Alduhailan, N., Alfahaid, R., Alqarawi, R., Alhareky, M., Shahin, S.Y., *et al.*: Canine impaction classification from panoramic dental radiographic images using deep learning models. *Informatics in Medicine Unlocked* **30**, 100918 (2022)
 - [43] Chen, Q., Huang, J., Salehi, H.S., Zhu, H., Lian, L., Lai, X., Wei, K.: Hierarchical cnn-based occlusal surface morphology analysis for classifying posterior tooth type using augmented images from 3d dental surface models. *Computer Methods and Programs in Biomedicine* **208**, 106295 (2021)
 - [44] Zhu, Z., Li, D., Hu, Y., Li, J., Liu, D., Li, J.: Indoor scene segmentation algorithm based on full convolutional neural network. *Neural Computing and Applications* **33**, 8261–8273 (2021)
 - [45] Pauwels, R.: A brief introduction to concepts and applications of artificial intelligence in dental imaging. *Oral radiology* **37**(1), 153–160 (2021)
 - [46] Kaji, S., Kida, S.: Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology* **12**(3), 235–248 (2019)
 - [47] Keele, S., *et al.*: Guidelines for performing systematic literature reviews in software

engineering. Technical report, ver. 2.3 ebse technical report. ebse (2007)

- [48] Wahono, R.S.: Systematic literature review: Pengantar, tahapan dan studi kasus. Romisatriawahono. Net (2020)
- [49] IEEE: Institute of Electrical and Electronics Engineers. [Online; accessed 13-January-2023]. <https://ieeexplore.ieee.org/Xplore/home.jsp>
- [50] ACM: a research, discovery and networking platform. [Online; accessed 13-January-2023]. <https://dl.acm.org/>
- [51] Library, W.O.: a digital library of scholarly resources, journals, books, and references. [Online; accessed 13-January-2023]. <https://onlinelibrary.wiley.com/>
- [52] ScienceDirect: Elsevier's premier platform of peer-reviewed scholarly literature. [Online; accessed 13-January-2023]. <https://www.sciencedirect.com/>
- [53] SpringerLink: a reading platform of choice for hundreds of thousands of researchers worldwide. [Online; accessed 13-January-2023]. <https://link.springer.com/>
- [54] James, M.-S., Marrissa, C., Mark, S., Anthea, B., et al.: Systematic approaches to a successful literature review. *Systematic Approaches to a Successful Literature Review*, 1–100 (2021)
- [55] Singh, P., Sehgal, P.: Numbering and classification of panoramic dental images using 6-layer convolutional neural network. *Pattern Recognition and Image Analysis* **30**, 125–133 (2020)
- [56] Kabir, T., Lee, C.-T., Chen, L., Jiang, X., Shams, S.: A comprehensive artificial intelligence framework for dental diagnosis and charting. *BMC Oral Health* **22**(1), 1–13 (2022)
- [57] Brahmi, W., Jdey, I.: Automatic tooth instance segmentation and identification from panoramic x-ray images using deep cnn. *Multimedia Tools and Applications*, 1–21 (2023)
- [58] Sivasundaram, S., Pandian, C.: Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture. *International Journal of Imaging Systems and Technology* **31**(4), 2214–2225 (2021)
- [59] Krois, J., Garcia Cantu, A., Chaurasia, A., Patil, R., Chaudhari, P.K., Gaudin, R., Gehrung, S., Schwendicke, F.: Generalizability of deep learning models for dental image analysis. *Scientific reports* **11**(1), 1–7 (2021)
- [60] Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., Schwendicke, F.: Deep learning for the radiographic detection of apical lesions. *Journal of endodontics* **45**(7), 917–922 (2019)
- [61] Franco, A., Porto, L., Heng, D., Murray, J., Lygate, A., Franco, R., Bueno, J., Sobania, M., Costa, M.M., Paranhos, L.R., et al.: Diagnostic performance of convolutional neural networks for dental sexual dimorphism. *Scientific Reports* **12**(1), 17279 (2022)
- [62] Mohammad, N., Muad, A.M., Ahmad, R., Yusof, M.Y.P.M.: Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging. *BMC Medical Imaging* **22**(1), 66 (2022)
- [63] Vila-Blanco, N., Varas-Quintana, P., Aneiros-Ardao, Á., Tomás, I., Carreira, M.J.: Xas: Automatic yet explainable age and sex determination by combining imprecise per-tooth predictions. *Computers in Biology and Medicine* **149**, 106072 (2022)

- [64] Rajee, M., Mythili, C.: Gender classification on digital dental x-ray images using deep convolutional neural network. *Biomedical Signal Processing and Control* **69**, 102939 (2021)
- [65] Zhu, H., Cao, Z., Lian, L., Ye, G., Gao, H., Wu, J.: Cariesnet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic x-ray image. *Neural Computing and Applications*, 1–9 (2022)
- [66] Park, E.Y., Cho, H., Kang, S., Jeong, S., Kim, E.-K.: Caries detection with tooth surface segmentation on intraoral photographic images using deep learning. *BMC Oral Health* **22**(1), 1–9 (2022)
- [67] Imak, A., Celebi, A., Siddique, K., Turkoglu, M., Sengur, A., Salam, I.: Dental caries detection using score-based multi-input deep convolutional neural network. *IEEE Access* **10**, 18320–18329 (2022)
- [68] Hsiao, T.-Y., Ho, Y.-C., Chen, M.-R., Lee, S.-Y., Sun, C.-W.: Disease activation maps for subgingival dental calculus identification based on intelligent dental optical coherence tomography. *Translational Biophotonics* **3**(3), 202100001 (2021)
- [69] Liu, M.-Q., Xu, Z.-N., Mao, W.-Y., Li, Y., Zhang, X.-H., Bai, H.-L., Ding, P., Fu, K.-Y.: Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on cbct. *Clinical Oral Investigations*, 1–11 (2022)
- [70] Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., Meng, D.: Tsasnet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network. *Knowledge-Based Systems* **206**, 106338 (2020)
- [71] Kwak, G.H., Kwak, E.-J., Song, J.M., Park, H.R., Jung, Y.-H., Cho, B.-H., Hui, P., Hwang, J.J.: Automatic mandibular canal detection using a deep convolutional neural network. *Scientific Reports* **10**(1), 5711 (2020)
- [72] Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., Cheng, X.: Automatic classification and segmentation of teeth on 3d dental model using hierarchical deep learning networks. *IEEE Access* **7**, 84817–84828 (2019)
- [73] Bonfanti-Gris, M., Garcia-Canas, A., Alonso-Calvo, R., Rodriguez-Manzaneque, M.P.S., Ramiro, G.P.: Evaluation of an artificial intelligence web-based software to detect and classify dental structures and treatments in panoramic radiographs. *Journal of Dentistry* **126**, 104301 (2022)
- [74] Oztekin, F., Katar, O., Sadak, F., Aydogan, M., Yildirim, T.T., Plawiak, P., Yildirim, O., Talo, M., Karabatak, M.: Automatic semantic segmentation for dental restorations in panoramic radiography images using u-net model. *International Journal of Imaging Systems and Technology* **32**(6), 1990–2001 (2022)
- [75] Kohlakala, A., Coetzer, J., Bertels, J., Vandermeulen, D.: Deep learning-based dental implant recognition using synthetic x-ray images. *Medical & Biological Engineering & Computing* **60**(10), 2951–2968 (2022)
- [76] Başaran, M., Çelik, Ö., Bayrakdar, I.S., Bilgir, E., Orhan, K., Odabaş, A., Aslan, A.F., Jagtap, R.: Diagnostic charting of panoramic radiography using deep-learning artificial intelligence system. *Oral radiology*, 1–7 (2021)
- [77] Lee, C.-T., Kabir, T., Nelson, J., Sheng, S., Meng, H.-W., Van Dyke, T.E., Walji, M.F., Jiang, X., Shams, S.: Use of the deep learning approach to measure alveolar bone level.

- Journal of clinical periodontology **49**(3), 260–269 (2022)
- [78] Chen, H., Li, H., Zhao, Y., Zhao, J., Wang, Y.: Dental disease detection on periapical radiographs based on deep convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery* **16**, 649–661 (2021)
 - [79] Al Kheraif, A.A., Wahba, A.A., Fouad, H.: Detection of dental diseases from radiographic 2d dental image using hybrid graph-cut technique and convolutional neural network. *Measurement* **146**, 333–342 (2019)
 - [80] Park, J., Lee, Y.: Oriented-tooth recognition using a five-axis object-detection approach. *Applied Intelligence*, 1–12 (2022)
 - [81] Gheisari, M., Ebrahimzadeh, F., Rahimi, M., Moazzamigodarzi, M., Liu, Y., Dutta Pramanik, P.K., Heravi, M.A., Mehbodniya, A., Ghaderzadeh, M., Feylizadeh, M.R., et al.: Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. *CAAI Transactions on Intelligence Technology* (2023)
 - [82] Wang, Z., Liu, K., Li, J., Zhu, Y., Zhang, Y.: Various frameworks and libraries of machine learning and deep learning: a survey. *Archives of computational methods in engineering*, 1–24 (2019)
 - [83] Tensorflow: an end-to-end machine learning platform. [Online; accessed 21-July-2023]. <https://www.tensorflow.org>
 - [84] Murad, M.A.A., Khalid, F., Manshor, N., Al-Jumaily, A., et al.: Medical image analysis of lung tumor diagnosis based on generalized fuzzy c-means clustering algorithm. In: *Journal of Physics: Conference Series*, vol. 2467, p. 012004 (2023). IOP Publishing
 - [85] PyTorch: Tensors and Dynamic neural networks in Python with strong GPU acceleration. [Online; accessed 19-July-2023]. <https://pytorch.org/>
 - [86] Caffe: a fast open framework for deep learning. [Online; accessed 22-July-2023]. <https://caffe.berkeleyvision.org>
 - [87] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 675–678 (2014)
 - [88] MXNET: a flexible and efficient library for deep learning. [Online; accessed 19-July-2023]. <https://mxnet.apache.org/versions/1.9.1/>
 - [89] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
 - [90] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
 - [91] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
 - [92] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826 (2016)

- [93] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28** (2015)
- [94] Ronneberger, O., Fischer, P., Brox, T.: 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, pp. 234–241 (2015). Springer
- [95] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
- [96] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017)
- [97] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)
- [98] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
- [99] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969 (2017)
- [100] Choi, E., Pang, K., Jeong, E., Lee, S., Son, Y., Seo, M.-S.: Artificial intelligence in diagnosing dens evaginatus on periapical radiography with limited data availability. *Scientific Reports* **13**(1), 13232 (2023)
- [101] Chun, S.-Y., Kang, Y.-H., Yang, S., Kang, S.-R., Lee, S.-J., Kim, J.-M., Kim, J.-E., Huh, K.-H., Lee, S.-S., Heo, M.-S., Yi, W.-J.: 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 (2023)
- [102] Amasya, H., Jaju, P.P., Ezhov, M., Gusarev, M., Atakan, C., Sanders, A., Manulius, D., Golitskya, M., Shrivastava, K., Singh, A., et al.: Development and validation of an artificial intelligence software for periodontal bone loss in panoramic imaging. *International Journal of Imaging Systems and Technology* (2023)
- [103] Ali, M.A., Fujita, D., Kobashi, S.: Teeth and prostheses detection in dental panoramic x-rays using cnn-based object detector and a priori knowledge-based algorithm. *Scientific Reports* **13**(1), 16542 (2023)
- [104] Müller, D., Soto-Rey, I., Kramer, F.: Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes* **15**(1), 1–8 (2022)
- [105] Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* **143**(1), 29–36 (1982)
- [106] Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging* **15**(1), 1–28 (2015)
- [107] Wang, Z., Wang, E., Zhu, Y.: Image segmentation evaluation: a survey of methods. *Artificial Intelligence Review* **53**, 5637–5674 (2020)

- [108] Silva, G., Oliveira, L., Pithon, M.: Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications* **107**, 15–31 (2018)
- [109] Abdi, S. Amir; Kasaei: “Panoramic Dental X-rays With Segmented Mandibles”, Mendeley Data, V1, doi: 10.17632/hxt48yk462.1. ”<https://data.mendeley.com/datasets/hxt48yk462/1>” (2017)
- [110] Walid, B., Imen, J.: Panoramic Dental Xray Dataset. Mendeley (2023)
- [111] Panetta, K., Rajendran, R., Ramesh, A., Rao, S.P., Agaian, S.: Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. *IEEE journal of biomedical and health informatics* **26**(4), 1650–1659 (2021)
- [112] Zhang, Y., Ye, F., Chen, L., Xu, F., Chen, X., Wu, H., Cao, M., Li, Y., Wang, Y., Huang, X.: Children’s dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data* **10**(1), 380 (2023)
- [113] López, V.R.F., Adorno, C.G., Román, J.C.M., Noguera, J.L.V., Silva, R.G., Legal-Ayala, H., Mello-Román, J.D., Torres, R.D.E., Facon, J.: Panoramic radiography database. Zenodo (2021). <https://doi.org/10.5281/zenodo.4457648> . <https://doi.org/10.5281/zenodo.4457648>