DISSERTATION

Künstliche Intelligenz in der Zahnheilkunde: Scoping-Review und Schließung beobachteter Wissenslücken durch eine methodische und eine klinische Studie

Artificial intelligence in dentistry: Scoping review and bridging observed knowledge gaps via a methodological study and a clinical trial

zur Erlangung des akademischen Grades
Doctor medicinae dentariae (Dr. med. dent.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Lubaina Tayeb Arsiwala-Scheppach

Erstbetreuung: Professor Falk Schwendicke

Datum der Promotion: 30.06.2024

Table of contents

List of tables

List of figures

List of abbreviations

English language:	
A Measurement Tool to Assess Systematic Reviews	AMSTAR
Area under the receiver operating characteristic	AUROC
Artificial intelligence	AI
Cone beam computed tomography	CBCT
Consolidated Standards of Reporting Trials using Artificial Intelligence	CONSORT-AI
Convolutional neural network	CNN
False negative	FN
False positive	FP
International Telecommunication Union	ITU
Machine learning	ML
Negative predictive value	NPV
Participants Intervention Comparison Outcome and Study	PICOS
Positive predictive value	PPV
Preferred Reporting Items for Systematic Reviews and Meta-Analyses	PRISMA
Residual neural network	ResNet
Sample size	n
Standards for Reporting Diagnostic Accuracy	STARD
Support vector machine	SVM
Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis	TRIPOD
True negative	TN
True positive	TP

U neural network	U-Net
Visual geometry group	VGG
World Health Organization	WHO
Two-dimensional	2-D
Three-dimensional	3-D

Deutsche Sprache:

Maschinelles Lernen	ML

Abstract

Objectives: The aims of this dissertation were to (1) conduct a scoping review of studies on machine learning (ML) in dentistry and appraise their robustness, (2) perform a benchmarking study to systematically compare various ML algorithms for a specific dental task, and (3) evaluate the influence of a ML-based caries detection software on diagnostic accuracy and decision-making in a randomized controlled trial.

Methods: The scoping review included studies using ML in dentistry published between 1st January 2015 and 31st May 2021 on MEDLINE, IEEE Xplore, and arXiv. The risk of bias and reporting quality were assessed with the QUADAS-2 and TRIPOD checklists, respectively. In the benchmarking study, 216 ML models were built using permutations of six ML model architectures (U-Net, U-Net++, Feature Pyramid Networks, LinkNet, Pyramid Scene Parsing Network, and Mask Attention Network), 12 model backbones of varying complexities (ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, VGG13, VGG16, VGG19, DenseNet121, DenseNet161, DenseNet169, and DenseNet201), and three initialization strategies (random, ImageNet, and CheXpert weights). 1,625 dental bitewing radiographs were used for training and testing. Five-fold cross-validation was carried out and model performance assessed using F1-score. In the clinical trial, each one of 22 dentists examined 20 randomly selected bitewing images for proximal caries; 10 images were evaluated with ML and 10 images without ML. Accuracy in lesion detection and the suggested treatment were evaluated.

Results: The scoping review included 168 studies, describing different ML tasks, models, input data, methods to generate reference tests, and performance metrics, impeding comparison across studies. The studies showed considerable risk of bias and moderate adherence to reporting standards. In the benchmarking study, more complex models only minimally outperformed their simpler counterparts, if at all. Models initialized by ImageNet or CheXpert weights outperformed those using random weights ($p<0.05$). The clinical trial demonstrated that dentists using ML showed increased accuracy (area under the receiver operating characteristic [mean (95% confidence interval): 0.89 (0.87–0.90)]) compared with those not using ML [0.85 (0.83–0.86); $p<0.05$], primarily due to their higher sensitivity [0.81 (0.74–0.87) compared to 0.72 (0.64–0.79); $p<0.05$]. Notably, dentists using ML also showed a higher frequency of invasive treatment decisions than those not using it ($p<0.05$).

Conclusion: To facilitate comparisons across ML studies in dentistry, a minimum (core) set of outcomes and metrics should be developed, and researchers should strive to improve robustness and reporting quality of their studies. ML model choice should be performed on an informed basis, and simpler models may often be similarly capable as more complex ones. ML can increase dentists' diagnostic accuracy but also lead to more invasive treatment.

Zusammenfassung

Ziele: Die Ziele dieser Dissertation waren, (1) ein Scoping-Review von Studien über maschinelles Lernen (ML) in der Zahnmedizin, (2) eine Benchmarking-Studie zum systematischen Vergleich verschiedener ML-Algorithmen für eine bestimmte zahnmedizinische Aufgabe, und (3) eine randomisierte kontrollierte Studie zur Bewertung einer ML-basierten Karies-Erkennungssoftware bezüglich diagnostischer Genauigkeit und Einfluss auf den Entscheidungsprozess durchzuführen.

Methoden: Das Scoping-Review umfasste Studien über ML in der Zahnmedizin, veröffentlicht vom 1. Januar 2015 bis 31. Mai 2021 auf MEDLINE, IEEE Xplore und arXiv. Bias-Risiko und Berichtsqualität wurden mit den Checklisten QUADAS-2 beziehungsweise TRIPOD bewertet. In der Benchmarking-Studie wurden 216 ML-Modelle durch Permutationen von sechs Architekturen (U-Net, U-Net++, Feature Pyramid Networks, LinkNet, Pyramid Scene Parsing Network und Mask Attention Network), 12 Backbones (Res-Net18, ResNet34, ResNet50, ResNet101, ResNet152, VGG13, VGG16, VGG19, DenseNet121, DenseNet161, DenseNet169 und DenseNet201) und drei Initialisierungsstrategien (zufällige-, ImageNet- und CheXpert-Gewichtungen) erstellt. Zum Training und Testen wurden 1.625 Bissflügel-Röntgenaufnahmen genutzt. Es wurde eine fünffache Kreuzvalidierung durchgeführt und die Modellleistung anhand des F1-Scores bewertet.

In der klinischen Studie untersuchten 22 Zahnärzte jeweils 20 zufällig ausgewählte Bissflügelbilder auf Approximalkaries; 10 Bilder wurden mit und 10 Bilder ohne ML ausgewertet. Die Genauigkeit in der Erkennung von Läsionen sowie die abgeleitete Therapieempfehlung wurden bewertet.

Ergebnisse: Das Scoping-Review schloss 168 Studien ein, in denen verschiedene ML-Aufgaben, Modelle, Eingabedaten, Methoden zur Generierung von Referenztests und Leistungsmetriken beschrieben wurden. Die Studien zeigten ein erhebliches Bias-Risiko und eine mäßige Einhaltung der Berichtsstandards. In der Benchmarking-Studie hatten komplexere Modelle gegenüber einfachen Modellen allenfalls geringe Vorteile. Mit ImageNet- oder CheXpert-Gewichtungen initialisierte Modelle übertrafen solche mit Zufallsgewichtungen ($p<0,05$). In der klinischen Studie erreichten Zahnärzte mit ML eine höhere Genauigkeit bei der Kariesdetektion (Receiver-Operating-Charakteristik [Mittelwert (95 % Konfidenzintervall) 0,89 (0,87–0,90)]) als ohne ML [0,85 (0,83–0,86); $p<0,05$], hauptsächlich aufgrund höherer Sensitivität [0,81 (0,74–0,87) verglichen mit 0,72 (0,64–0,79); $p<0,05$]. Zahnärzte mit ML wählten auffallend häufiger invasive Behandlungen als ohne ML ($p<0,05$).

Schlussfolgerung: Zur besseren Vergleichbarkeit von ML-Studien in der Zahnmedizin, sollten Core Outcomes und Metriken definiert sowie Robustheit und Berichtsqualität verbessert werden. Die Entwicklung von ML-Modellen sollte auf informierter Basis erfolgen, bei oft ähnlicher Leistung von einfacheren und komplexeren Modellen. ML kann die diagnostische Genauigkeit erhöhen, aber auch zu mehr invasiven Behandlungen führen.

1 Introduction

1.1 Artificial intelligence – What it is and how it applies to healthcare

Artificial intelligence (AI) is the development of computer programs to be able to carry out tasks that normally require human intelligence, such as visual perception, decision making, and problem solving [1]. Machine learning (ML) is a branch of AI that involves training computer algorithms to learn patterns from data and then make predictions. In ML, it is not the human who defines the rules that a computer follows to fulfil certain tasks; instead, the computer itself learns rules from the data provided to it [2]. Examples of this in our daily lives include virtual assistants like Siri and Alexa, photograph filters on social media, algorithms that suggest online content tailored to our interests, navigation maps, autocorrect functions for messages, speech-to-text converters, language translation apps, chatbots, and self-driving cars, to name a few [3]. AI is also being used in industries such as finance, transportation, and healthcare, with an ever-increasing impact on our lives. The potential for AI to change how we work, live, and communicate is enormous, and is expected to become even more critical in the future.

In healthcare, ML is being used to improve patient outcomes and diagnosis [4, 5]. It can help physicians detect conditions such as cancer, heart disease, and neurological disorders earlier and more accurately, with medical imaging being a significant area of focus [3, 5, 6]. ML-based analysis of medical imaging has been successfully employed to help interpret medical scans with more precision, thereby reducing the chances of missed diagnoses or incorrect treatments. A range of medical fields are using ML, for example, dermatology, ophthalmology, radiology, and dentistry, where it has achieved similar or higher accuracy than experienced clinicians [5-8]. According to a recent report, the market value for AI and ML in healthcare has been projected to increase by more than sevenfold between the years 2023 and 2028 up to USD 102.7 billion [9].

In the dental clinic, many tasks can be performed by ML with greater precision and fewer errors than human counterparts; from booking and coordinating appointments to assisting with clinical diagnosis and treatment planning [10, 11]. ML algorithms can analyze large datasets of dental images (e.g., photographs, radiographs, three-dimensional [3-D] scans, and transillumination images), detect patterns, and provide insights into diagnosis, treatment, and prevention [12]. Thus, they reduce the need for manual analysis and allow dentists to focus on other areas of patient care [12]. In addition, using an ML software may aid dentists in boosting their performance, for example, by increasing the chance of detecting caries lesions early, resulting in improved patient outcomes and reduction in dental care expenditure [13].

1.2 Use of machine learning in dental research

1.2.1 Existing literature

The use of big data and improvements in computer science technology has led to an explosion of studies using ML in medicine in the last decade [11]. The number of publications alone tells a story: it doubled in the last decade, from 162,444 in 2010 to 334,497 in 2021 [14]. A similar surge in publications has been observed in dental research, reflecting the growing interest in this field [2, 15].

Since ML can be used for various dental tasks, the literature covers a wide range of applications, for example, prediction of dental complications after the extraction of a third molar, tooth classification and outlining on images such as photographs or radiographs, cephalometric landmark detection, and dental pathology detection [7]. Different clinical applications or research aims necessitate different types of ML models because they involve distinct types of data, decision-making processes, and clinical outcomes [16]. For instance, some applications may require high accuracy rates with low tolerance for false positive (FP) or false negative (FN) cases, while others may prioritize interpretability to enable clinical decision-making. Thus, an ML model can be understood to be a highly specific model that is built for a specific task, which in turn needs to be built with specific algorithms pertaining to the task. Moreover, in order to evaluate how well the ML model has accomplished the given task, a plethora of performance metrics can be used, which are again specific to the nature of the task.

As a consequence, studies in the field of dentistry using ML can differ widely [7]. There is significant heterogeneity in the study designs, clinical applications, ML models, data, and performance metrics, which hinders comparing studies and evaluating their consistency and robustness [11]. Additionally, variation has been reported in the quality of ML studies in medicine with respect to the risk of bias and reporting of the methods and results [17].

It is likely that similar variance in quality and replicability also exists for dental ML studies [18] but their comprehensive objective evaluation has not been carried out so far. Reviews are regarded as a vital component of evidence-based medicine. Conducting such a comprehensive overview of the existing literature and appraising the robustness of the studies could facilitate highlighting their strengths and weaknesses and identify future research needs.

1.2.2 Current methodological trends of machine learning in dentistry

The basic features of an ML model are:

Architecture: The arrangement pattern of the basic building blocks of a ML model defines its architecture. Examples widely used in the field are U-neural network (U-Net), U-Net++, and LinkNet.

Backbone: For certain complex tasks, specialized building blocks are added to the model structure which form its foundation and hence are called a backbone. Examples are Residual neural network (ResNet) and Visual Geometry Group (VGG).

Complexity level: ML models can have different levels of complexity or depth, which correspond to the size of the model, i.e., the number of building blocks used.

Initialization: This is akin to giving a ML model a head-start for a given task, by providing it with some basic information relevant to the task, for example, about photographs or radiographs. It can be thought of as learning to play the cello. If one already knows how to play the violin, learning the cello would be easier because some of the skills and knowledge one already has can be transferred. This technique in the context of ML is called ‘transfer learning’; Figure 1.

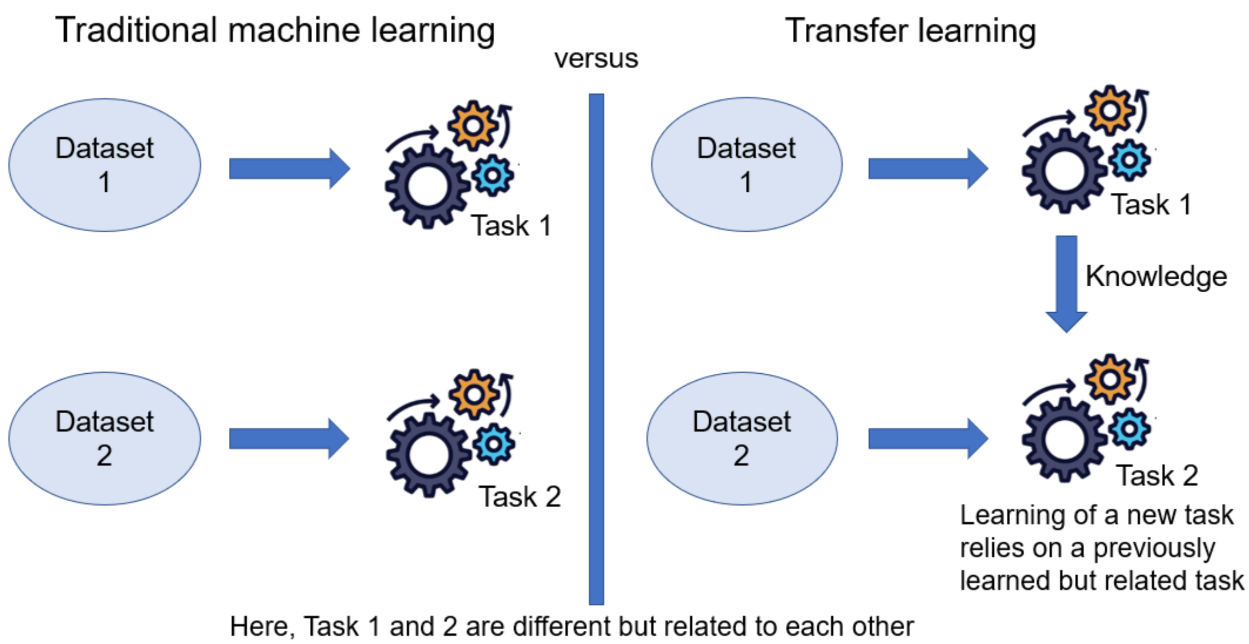

Figure 1: Representation of the concept of transfer learning in the context of machine learning and how it differs from traditional machine learning. In transfer learning, the model to be trained for a new task is supplied with information from another model which was trained for a different but related task. Source: modified from https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a Accessed on 25th January 2023.

A major weakness in ML research in dentistry is the considerable heterogeneity among the existing studies, much of which stems from the different types of ML models currently in use [2, 11, 19-22]. Building an ML model to accomplish a specific task involves choosing an architecture, backbone, complexity level or size, and initialization strategy from the numerous options available. Without a guiding framework available, researchers tend to arbitrarily select the popular options, often without empirical evidence of their suitability for the task at hand [23]. The utter number of possible configurations of ML components impedes systematic and comprehensive comparisons of the existing studies' findings and identifying the best approach for a particular task in dentistry [11, 12].

Thus, it is important to systematically compare the different model configurations on one data set. Such an evaluation is called 'benchmarking' which has a couple of advantages. First, it provides guidance for researchers in the model building process, which can improve efficiency by enabling the development of high-performing models, in shorter times, and at lower computational costs. Second, it can help establish standards for ML research in dentistry, making it easier to compare and replicate results across different studies. However, in the dental field such benchmarking initiatives are scarce [24].

1.2.3 Clinical usefulness of machine learning for dentists

Another main weakness of the dental ML field is the lack of clinical prospective comparisons [7]. The implications of this are that despite the strong advances in ML methodology over the recent years, the effectiveness and applicability of ML models in a real-world clinical setting remain unclear [2, 7]. Randomized controlled clinical trials are essential for proving generalizability and accuracy of ML systems but also for assessing their impact on diagnostic and decision-making processes as well as the resulting treatment decisions, health gains, and costs [25].

1.3 Research aims

1.3.1 Scoping review of research literature on machine learning in dentistry

As discussed, a systematic assessment of the body of evidence on ML in dentistry is required to quantify the extent of heterogeneity among the available studies as well as appraise their consistency and robustness. Such an evaluation would facilitate highlighting the current strengths and weaknesses of the existing studies and identifying future research needs. The primary aim of the scoping review was to evaluate the overall body of existing research literature on ML in dentistry with regards to the clinical and ML tasks, models, kinds of datasets, and metrics used to evaluate the performance of the models. The secondary aim was to examine the robustness of the studies, focusing on the risk of bias and reporting quality.

1.3.2 Benchmark machine learning models for a specific dental task

As described earlier, there is a plethora of options available from which one can construct an ML model for a certain task but there is a lack of a proper framework which compares the various options and guides a researcher with these decisions. The aim of this study was to systematically compare, i.e., benchmark the possible configurations of various model architectures, backbones, complexity levels, and initialization strategies on one data set. The different models were evaluated on the dental task of outlining various parts of a tooth as seen on a radiograph, such as enamel, dentin, pulp, fillings, and prosthetic crowns. This dental task was chosen because ML models have demonstrated superior performance on it. Furthermore, the tooth structures in question can be easily identified even by dentists with less experience and thus the establishment of the reference test would be considered valid.

The hypothesis was that the performances of models would improve with their complexity level and the implementation of transfer learning. The results from this study could inform dental researchers about suitable model structures for their experiments, contribute to evidence-based ML model building in the dental field, and help establish standards for research.

1.3.3 Evaluate a machine learning software in a randomized clinical trial

As discussed, there is a lack of studies that demonstrate the true usefulness of ML systems in a clinical setting. Hence it is important to generate evidence on the diagnostic accuracy and applicability of ML systems in the hands of dentists. The aim of the randomized controlled clinical trial was to quantify the differences in performance of dentists in the absence versus presence of assistance by an ML software in the task of detecting proximal caries on bitewing radiographs. For this dissertation, the analysis has been extended beyond the publication [26] to evaluate the performance of the ML software by itself for the given task. Furthermore, the influence of the ML software on the treatment decisions made by the dentists was examined.

The hypothesis was that dentists using ML would be more accurate than those not using ML. The results from this clinical trial could demonstrate the prospective usefulness and impact of ML software on dental diagnostics and treatments in a real-world setting.

2 Methods

2.1 Scoping review

2.1.1 PICOS question

The research question was framed according to the Participants Intervention Comparison Outcome and Study (PICOS) strategy and was as follows: “Which ML practices are being employed by studies in dentistry and what are their methodological quality and findings?” [18]

- Population: All kinds of population-level data with a dental or oral component [18].
- Intervention/Comparison: ML techniques applied with a dental or oral focus for the diagnosis, management, or prognosis of dental conditions or improving data quality [18].
- Outcome: Performance evaluation of the ML models in terms of certain metrics, for example, accuracy, intersection-over-union, sensitivity, precision, area under the receiver operating characteristic (AUROC), F indices, specificity, negative predictive value (NPV), rank-N recognition rate, error estimates, correlation coefficients, etc. [18]
- Study design type: All kinds of studies except reviews, editorials, and technical standards, with no language restrictions [18].

2.1.2 Search strategy

The search strategy was designed with the aim to identify all studies meeting the eligibility criteria in accordance with the objectives of the review. The varying publication norms among different academic disciplines were taken into consideration. The review [18] did not restrict the inclusion of studies with respect to the target study population, outcome of interest, or the context in which ML was used. It aimed to include all original studies related to dentistry and ML, as long as they did not contain major reporting errors, such as failing to define the type of ML used, inadequately describing the dataset employed, or omitting explicit reporting of the study findings.

Three electronic databases (MEDLINE via PubMed, Institute of Electrical and Electronics Engineers Xplore, and arXiv) were used. The archiving database arXiv was used in an effort to also include grey literature. This included studies that did not go through a formal, but a non-formal peer-review process and then were updated after peer-review [18].

The search terms used were ‘deep learning’, ‘artificial intelligence’, ‘machine learning’, ‘convolutional neural network’, ‘dental’ and ‘teeth’. The search strategies for all the three databases are defined below [18]:

- Database MEDLINE/PubMed

(“deep learning” OR “artificial intelligence” OR “machine learning” OR “convolutional neural network”) AND (“dental” OR “teeth)

- Database Institute of Electrical and Electronics Engineers Xplore

(((“Document Title”:“deep learning” OR “artificial intelligence” OR “machine learning” OR “convolutional neural network”) OR (“Keywords”:“deep learning” OR “artificial intelligence” OR “machine learning” OR “convolutional neural network”)) AND ((“Document Title”: “dental” OR “teeth) OR (“Keywords”:“dental” OR “teeth)))

- Database arXiv

“deep learning” OR “artificial intelligence” OR “machine learning” OR “convolutional neural network” AND “dental” OR “teeth

The following inclusion criteria [18] were applied:

(1) Studies with a dental/oral focus, including technical papers.

(2) Studies employing ML.

(3) Studies published between 1st January 2015 and 31st May 2021, as the aim was to gather recent studies and specifically include the most rapidly evolving ML era at present.

Owing to the lack of randomized controlled trials on ML in dentistry the scoping review was expanded to include non-randomized studies in order to gain a comprehensive overview of the field.

The inclusion of the studies was decided by two reviewers in consensus. All studies found to be potentially eligible were assessed in full text against the inclusion criteria. All the included and excluded studies were listed along with justification for exclusion for the excluded studies.

2.1.3 Data collection

Data extraction was performed by three reviewers and then finally reviewed by one of them. In case of disagreements, a consensus process was used.

The extracted data was recorded using a formerly validated Excel document. Study characteristics included country, publication year, study aim, clinical field, type of input data (covariates, photographs, or radiographs; two-dimensional [2-D] or 3-D imagery), data source and size, type of ML model, reference test (i.e., how the ground truth was defined), comparators (e.g., current standard of care, clinicians, etc.), and model performance metrics along with their values. No assumptions were made regarding missing or unclear data [18].

2.1.4 Assessment of risk of bias in the individual studies

The risk of bias was examined using the QUADAS-2 tool in four domains [27]. First, the risk of bias in data selection was examined with regard to inappropriate exclusions, case-control study design, and patient enrolment strategy. Second, the risk of bias in the index test was examined with regard to the independence of the measurement from the reference test and pre-specification of thresholds. Third, the risk of bias in the reference test was examined for its validity of assessing the ground truth and the independence of its measurement from the index test. Fourth, the risk of bias in flow and timing was examined regarding whether there was an appropriate interval between the index and reference tests, whether the reference test was used for all participants, whether the same reference test was used for all participants, and whether all participants were included in the analysis. The impact of the risk of bias in the individual studies on the overall results of the review was assessed and discussed.

Using the same tool, applicability concerns, i.e., how specific methods used by the studies influenced the generalizability of their results, were evaluated in three domains. First, applicability concerns in data selection were examined regarding a potential mismatch between the included participants and the review question. Second, applicability concerns in the index test were examined regarding a potential mismatch between the test, its conduct, or its interpretation and the review question. Third, applicability concerns in the reference test were examined regarding a potential mismatch between the target condition as defined by the reference test and the review question [18].

2.1.5 Assessment of reporting quality of the individual studies

Observance of reporting guidelines was assessed using the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) tool, a 22-item checklist that provides guidelines for reporting of prediction studies [18, 28]. TRIPOD has been used for similar assessments of studies in other medical fields [17, 29].

2.1.6 Data synthesis

Initially, a meta-analysis was planned for all studies included in the review; however only 10% of studies reported complete confusion matrices that could be used for such an analysis. Furthermore, these few studies differed from each other in terms of clinical research question/task, type of input data, model architecture, inferences from the results, etc, [18] thus making a meta-analysis not feasible. Hence, a narrative synthesis was performed instead, displaying which ML tasks were used in different clinical fields of dentistry, namely restorative dentistry and endodontics, oral medicine, oral radiology, orthodontics, oral surgery and implantology, periodontology, prosthodontics, and general dentistry.

For this dissertation, the analysis was expanded beyond the publication [18] to construct confusion matrices for studies which presented their metrics as described ahead.

When sensitivity, specificity, precision, and sample size were available:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1\text{-Sensitivity} & 0 & 0 & \text{-Sensitivity} \\ 0 & 1\text{-Specificity} & \text{-Specificity} & 0 \\ 1\text{-Precision} & 0 & \text{-Precision} & 0 \end{pmatrix} \begin{pmatrix} TP \\ TN \\ FP \\ FN \end{pmatrix} = \begin{pmatrix} n \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

When sensitivity, specificity, accuracy, and sample size were available:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1\text{-Sensitivity} & 0 & 0 & \text{-Sensitivity} \\ 0 & 1\text{-Specificity} & \text{-Specificity} & 0 \\ 1/\text{Accuracy} & 1/\text{Accuracy} & 0 & 0 \end{pmatrix} \begin{pmatrix} TP \\ TN \\ FP \\ FN \end{pmatrix} = \begin{pmatrix} n \\ 0 \\ 0 \\ n \end{pmatrix}$$

where TP = number of true positive, TN = number of true negative, FP = number of false positive, FN = number of false negative, and n = sample size.

Combining the studies who properly reported their confusion matrices along with those whose confusion matrices were reconstructed, resulted in a total of 29 studies. Among these studies, 19 studies performed classification tasks, four studies each performed object detection and semantic segmentation tasks, and two studies performed instance segmentation tasks. Since the number of studies in the latter three groups were too few for a meaningful analysis, only the classification studies were used to estimate the mean sensitivity and exact binomial 95% confidence interval for each study and displayed in a forest plot. All data management and statistical analyses were performed with R (version 4.0.3, www.r-project.org) [30].

2.1.7 Reporting protocol and ethics statement

The review methods were decided upon before the commencement of the scoping review to reduce the risk of bias. The study protocol was registered with PROSPERO (registration number CRD42021288159). Reporting of the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist. In accordance with the guidelines of the Charité Promotionsbüro, this scoping review was appraised using the checklist A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2 and it achieved a very high rating. From a total of 16 items in the checklist, 13 items were applicable to this scoping review, out of which 12 items (92%) were rated with a ‘yes’. The item rated with ‘no’ referred to the reporting of the sources of funding for the individual studies included in the review.

Ethics approval was not sought because the review was based exclusively on published literature [18].

2.2 Benchmarking study

The aim was to systematically compare the various configurations of different model architectures, backbones, complexity levels, and initialization strategies for the task of outlining tooth structures on bitewing radiographs.

2.2.1 Model components

Six architectures were selected: U-Net, U-Net++, Feature Pyramid Networks, LinkNet, Pyramid Scene Parsing Network, and Mask Attention Network. These networks were selected as they allow to employ the same backbones (i.e., ResNet, VGG, and DenseNet) with varying levels of complexity. 12 different levels of model complexity were used: ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, VGG13, VGG16, VGG19, DenseNet121, DenseNet161, DenseNet169, and DenseNet201. The numeric value at the end of the model's name indicates the complexity level. Three different initialization strategies were evaluated, i.e., random weights initialization, initialization based on ImageNet data, and initialization based on CheXpert data.

Thus, a total of 216 model configurations were evaluated. Figure 2 represents the study design. All models were trained under a five-fold cross-validation scheme, also depicted in Figure 2, which is a technique to evaluate the performance of a model on a limited dataset. The basic idea is to divide the dataset into two parts: a training set on which the model is trained and a testing set on which the model’s performance is evaluated [31]. For example, in a five-fold cross-validation, the dataset is randomly divided into five equal parts. The model is trained on four parts and tested on the remaining part. This process is repeated five times, so that each part is used as the testing set exactly once. The results from each iteration/fold are then averaged to obtain a more robust estimate of the model's performance and prevent undue data-related influence on the models.

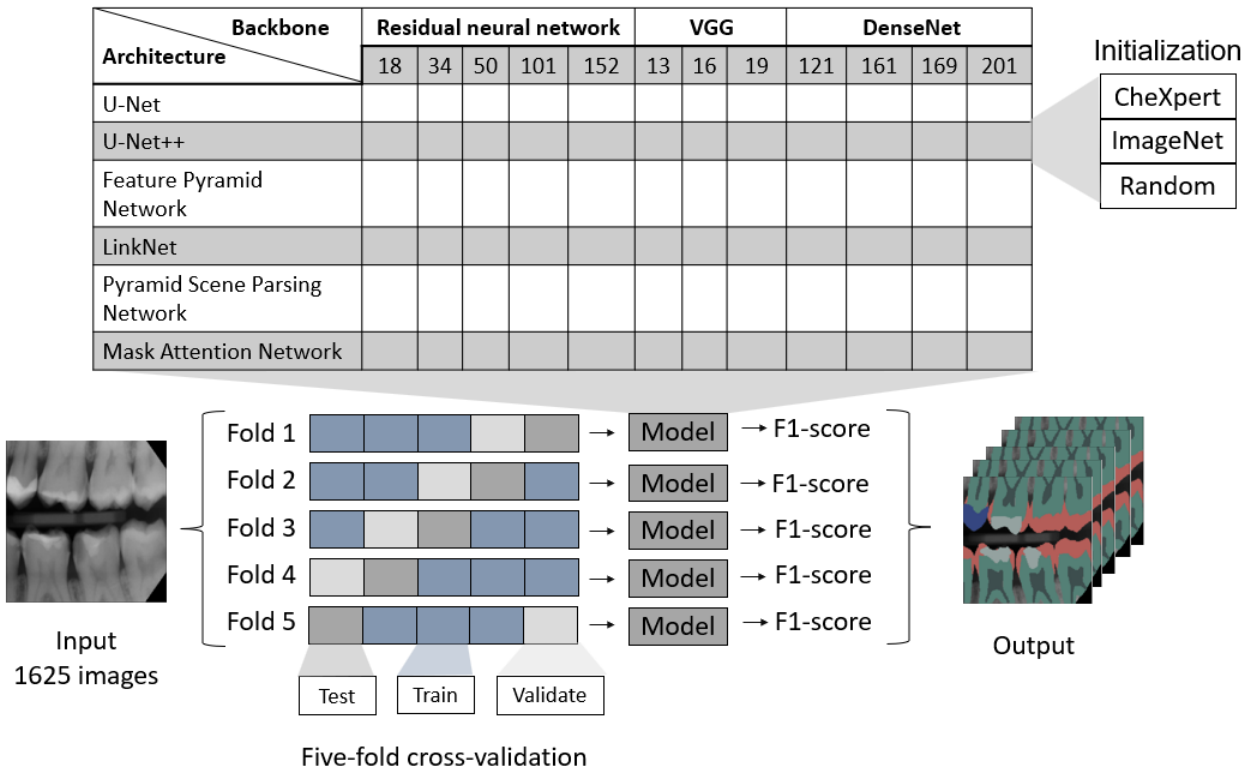

Figure 2: Illustration of the benchmarking study design. Model setups were based on different architectures, backbones, complexity levels, and initialization strategies (top) and five-fold cross-validation with varying train, test, and validation sets for each iteration/fold (bottom). Exemplary input bitewing radiograph (left) and the output image containing markings of the different tooth structures (right). The numbers below the names of the different backbone groups represent the various complexity levels. Abbreviation: VGG, visual geometry group. Source: modified from Figure 1, publication L. Schneider, L. Arsiwala-Scheppach, J. Krois, H. Meyer-Lueckel, K.K. Bressem, S.M. Niehues, F. Schwendicke, Benchmarking Deep Learning Models for Tooth Structure Segmentation, J Dent Res (101, 11) pp. 1343-1349. Copyright © 2022 International Association for Dental Research and American Association for Dental, Oral, and Craniofacial Research. doi: 10.1177/00220345221100169. Image rights for reuse in dissertation held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under Open Access category and license CC-BY NC 4.0 as per publisher policy. Additionally, kind permission for reuse was obtained from the publisher Sage Publications.

2.2.2 Data used

1,625 dental bitewing radiographs were used, each displaying up to nine teeth. One dentist annotated the parts of a tooth, such as enamel, dentin, pulp, fillings, and prosthetic crowns, on the radiographs using an in-house custom-built annotation tool. These annotations served as the reference test [32]. A second dentist reviewed the accuracy of these annotations. Both dentists were calibrated for the annotation process. Images containing implants, bridges, or root canal fillings accounted for less than one percent of the total images and hence were excluded. It should be noted that enamel, dentin, and pulp were present in all images whereas fillings and crowns were less frequent (80% and 20% images, respectively). To suit the requirements of the ML models, the images and annotations were resized to a fixed input size.

2.2.3 Statistical analysis

The performance of the models was primarily quantified by the F1-score which was calculated as described by Forman and Scholz [33]. The different model configurations regarding architectures and initialization strategies were ranked by their performance i.e., F1-score and compared using the Wilcoxon rank-sum test. Additionally, the relationship between model complexity and performance was examined by the Spearman's correlation metric. As a sensitivity analysis, model performance was evaluated on the less prevalent classes of fillings (80%) and crowns (20%). Owing to the skewed distribution of the data, non-parametric statistical tests were used. The p values were adjusted by the Benjamini–Hochberg method to account for multiple testing. The level of significance was set to $p<0.05$. Statistical analyses were performed with R (version 4.0.3, www.r-project.org) [30].

2.2.4 Reporting protocol and ethics statement

Two reporting protocols were followed for this study: the Standards for Reporting Diagnostic Accuracy (STARD) guidelines [34] and the Checklist for Artificial Intelligence in Dental Research [35]. The study was approved by the Ethics Committee of the Charité (EA4/102/14 and EA4/080/18) [36].

2.3 Clinical trial

2.3.1 Study design

A randomized controlled non-blinded clustered cross-over superiority trial was conducted with an allocation ratio of 1:1.

Randomization: Seven blocks of 20 radiographs each were randomly generated using randomize.org from a collection of 140 radiographs. Each dentist then randomly received one of these seven blocks. The sequence of radiographs in each block was also randomly determined and was identical for every dentist. Of the 20 radiographs to be viewed, half of them were randomly assigned to be viewed by the dentist with assistance from the ML software and the other half without. Owing to the nature of the intervention it was not possible to blind the dentists regarding which image belonged to which trial group. Figure 3 represents the study design of the trial.

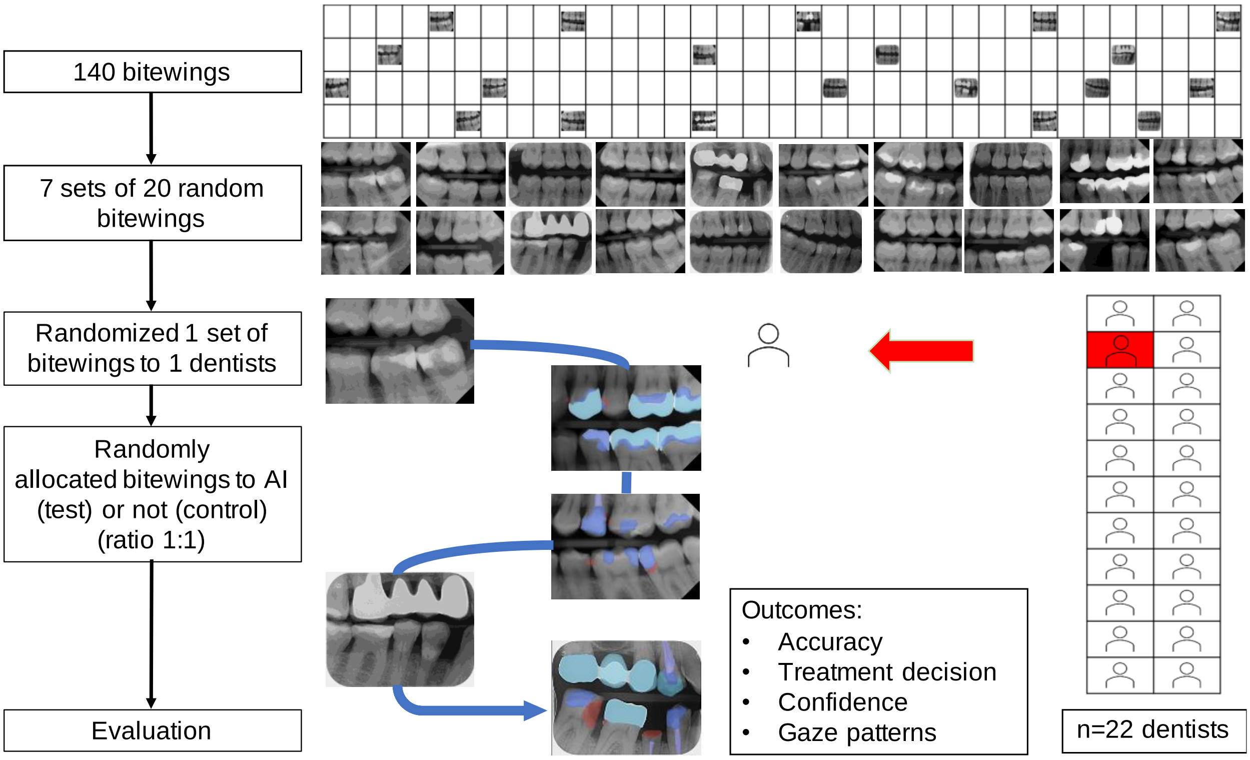

Figure 3: Flowchart of the randomized clinical trial. From 140 bitewing radiographic images, seven blocks of 20 images were randomly generated. Each of the 22 dentists randomly assessed one block, with images being randomly allocated to the intervention (with machine learning software) or control group in a 1:1 allocation ratio. Different colors on the bitewing images indicate different findings, e.g., blue indicates fillings, crowns, or root-canal fillings, while red indicates caries lesions. Abbreviation: AI, artificial intelligence. Source: Figure 1, publication S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849. Image rights for reuse in dissertation held by authors of the publication, including Lubaina T. Arsiwala, as per publisher Elsevier policy.

2.3.2 Sample size

The sample size for the trial was based on a prior study [25] which used the same ML software. The study design was a clustered trial where approximately 20 tooth surfaces were visible per radiograph and to account for this, the 'design effect' was estimated. The formula used to estimate the design effect was 1 + (cluster size - 1) * intraclass correlation coefficient, where the intraclass correlation coefficient was assumed to be 0.2, based on a prior study [37]. Thus, a cluster size of 20 surfaces resulted in a design effect of 4.8. A trial with 95% power and an 'alpha' value of 0.05 would require 1280 tooth surfaces to be included. Thus, for the present trial, the number of surfaces required was 1280 * 4.8 = 6144. Since each dentist was assigned to examine 20 radiographs (i.e., 400 surfaces), a minimum of 16 dentists had to be recruited. Note that in the protocol, recruitment of 20 dentists was planned and finally 22 were recruited. There were no predetermined stopping rules or interim analyses.

2.3.3 Data used

Study participants were recruited and the trial was conducted from October 2020 to January 2021. The participating dentists worked at Charité – Universitätsmedizin Berlin dental hospital or in private clinics in Berlin and thus the trial was conducted at these locations. Care was taken to ensure standardization of study conditions at all locations as follows: For the participants from private clinics, the study investigator brought the monitor screen used in the trial at the Charité dental hospital to their clinic and the experiment was carried out in a dimly lit room in the clinic, similarly as conducted at the Charité dental hospital. Participants were excluded if they were no longer clinically active, had less than two years of clinical experience, or had no regular experience of caries detection. Written informed consent was obtained from all participating dentists. Participants' characteristics such as age and gender were used for descriptive analyses.

The 140 bitewing radiographs of permanent teeth used in the trial were from patients treated between the years 2016 and 2018 at Charité – Universitätsmedizin Berlin dental hospital under an ethics approved protocol (EA4/080/18). Bitewing radiographs of the permanent dentition were included if, at minimum, the crowns of one dental arch were discernible. The radiographs were generated using machines produced by Dentsply Sirona or Dürr Dental companies.

The reference test was established by four dentists independently outlining proximal caries lesions on all radiographic images using an in-house customised annotation software in dimly lit rooms using diagnostic screens under standardized conditions. All annotations were reviewed and modified, if necessary, by a fifth dentist who could consult the other four dentists. The union of areas annotated by all dentists for each lesion constituted the reference standard; this is a popular method for generating a reference standard when a “hard” reference like histopathological examination is unavailable.

The caries lesions were classified into enamel lesion, early dentin lesion within the outer 1/3 of the dentin, or advanced dentin lesion expanding deeper than that, by two independent dentists in consensus.

2.3.4 Trial intervention

The intervention was an ML-based software for viewing radiographs in order to classify teeth and outline fillings and caries lesions on bitewing radiographs (dentalXrai Pro 1.0.4, dentalXrai Ltd., Berlin, Germany). The software could display the original radiograph and an augmented version with pathology detections by the ML software shown as overlays (see Figure 3 for examples of augmented radiographs). With respect to caries detection, the ML software indicated whether a caries lesion was present or absent for every surface. At least one week prior to the study, all dentists received a handbook of the ML software to be used during the trial and were advised to gain familiarity with the software by using it to analyse a minimum of four bitewing radiographs. The control group constituted the conventional radiographic detection of proximal caries without any aid from the ML software.

The intervention was applied as described: First, each dentist was assigned to a randomly chosen block of 20 bitewing radiographs, half of which were randomly assigned to be viewed along with the ML software and the other half without. In the ML group, dentists had the option to enable or disable the ML software as per their choice. Dentists verbally reported their diagnoses of proximal caries and their accompanying treatment decisions to the study assistant.

2.3.5 Outcomes

The primary outcomes were AUROC, accuracy, F1-score, sensitivity, specificity, positive predictive value (PPV), and NPV. These were calculated for both groups of the trial, i.e., dentists without ML and dentists with ML. For this dissertation, the analysis was expanded beyond the publication [26] to include the aforementioned primary outcomes for the ML software alone (i.e., without a dentist).

The secondary outcome focused on the treatment decision assigned by the dentists for each proximal surface, i.e., the number of no treatment, non-invasive (e.g., fluoride varnish), micro-invasive (e.g., caries infiltration) or invasive (e.g., filling) treatments. Since the secondary outcomes were derived from data provided by dentists, they could not be calculated for the ML software.

2.3.6 Statistical analysis

First, the design effect was estimated to account for the clustered trial design as described earlier. In addition to this, clustering by dentists was accounted for as every dentist was present in both groups of the trial. Thus, a combined design effect was estimated and applied to all analyses. The AUROC of all dentists stratified by trial group were plotted to facilitate comparison. For this dissertation, the comparison was extended beyond the publication [26] to include the performance of the ML software too. Also, a paired scatterplot was created to highlight the differences in sensitivity and specificity for each dentist across the two groups of the trial. The number of surfaces assigned to each type of treatment was calculated. Furthermore, additional analysis was incorporated into this dissertation: the inter-rater agreement between the dentists for detecting caries lesions, using the Fleiss kappa metric. The two-sided t-test, chi-squared test, and Fisher's exact test were used to compare the results between the trial groups. The level of significance was set at $p<0.05$. No deviation from the trial protocol occurred. All analyses were conducted for the total dataset (i.e., overall) and stratified by caries depth. All data management and statistical analyses were performed with R (version 4.0.3, www.r-project.org) [30].

2.3.7 Reporting protocol and ethics statement

The trial was registered at the Deutsches Register Klinischer Studien (DRKS00022357) and was ethically approved by the Charité – Universitätsmedizin Berlin (EA/144/20). Reporting of the trial followed the Consolidated Standards of Reporting Trials using Artificial Intelligence (CONSORT-AI) checklist and the Checklist for Artificial Intelligence in Dental Research [18, 35, 38].

3. Results

3.1 Scoping review

3.1.1 Individual study characteristics

183 studies were identified, out of which 168 (92%) studies were included; Figure 4 [18]. The included studies [13, 25, 31, 32, 39-202] and their characteristics are listed in Table 1 and the excluded studies along with justifications for exclusion are provided in Table 2. The included studies were published from 1st January 2015 to 31st May 2021 (median year: 2019), with their annual numbers increasing steadily as depicted in Figure 5 (for year 2021, data only until May was available) [18]. The studies were from 40 countries (Figure 6) and employed various types of input data, e.g., 2-D data (radiographs: 42% studies, photographs, or other types: 16% studies), 3-D data (radiographs: 18% studies, non-radiographs: 4% studies), survey data: 10% studies, and combinations of different kinds of data (9% studies) [18]. 97% studies used institutional data e.g., universities, hospitals, and private practices, whereas 3% studies used the National Health and Nutrition Examination Survey, M3BE database, 2013 Nationwide Readmissions Database of the USA, and the National Institute of Dental and Craniofacial Research dataset [18].

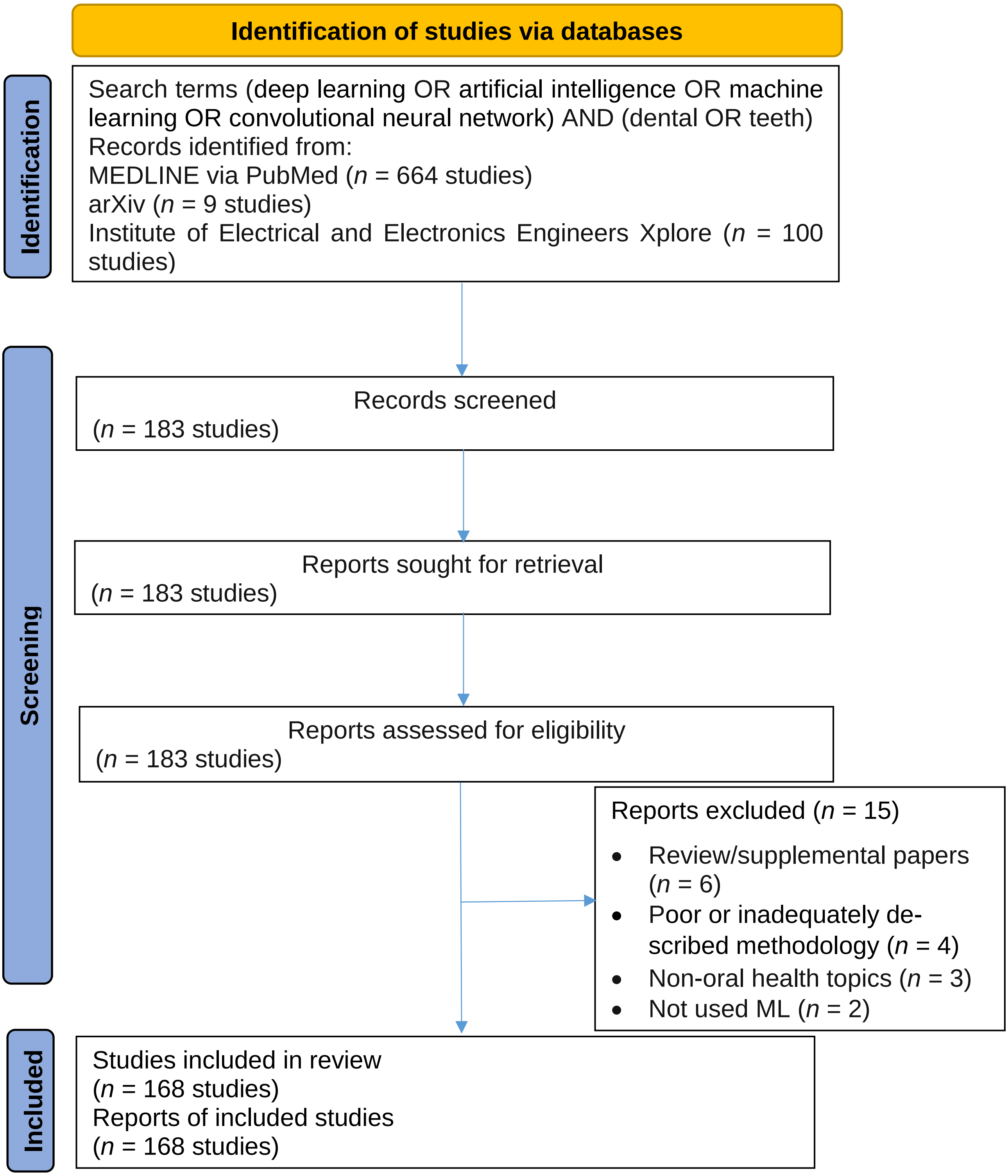

Figure 4: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study flow diagram of the scoping review on machine learning in dentistry. 168 studies were screened and included in the scoping review. Abbreviation: ML, machine learning. Source: modified from Figure 1, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937. Image rights held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under license CC BY 4.0 as per publisher MDPI policy.

Table 1: Studies included in the scoping review on machine learning in dentistry along with their characteristics (n=168).

No. [Citation]	Year	Study aim	Data type and size	ML model	Reference test	Model metrics
1 [39]	2015	Determine the most appropriate dental filling and monitor it.	4,336 charts, notes, and radiographs of fillings	Bayesian network and Multilayer Perceptrons	An expert saw the subjects annually for follow-up	Longevity error in years
2 [40]	2015	Automatic landmark detection on 3-D CBCT images	30 3-D CBCT images	Knowledge-based algorithm	Manual landmark plotting by 3 orthodontists	Overall mean error, overall landmark detection accuracy
3 [41]	2015	Evaluate accuracy of 3-D cephalometric measurements by a knowledge-based algorithm	30 3-D CBCT images transformed to DICOM format	Knowledge-based algorithm	21 cephalometric landmarks identified manually by three orthodontists	Mean error in measurements, mean error of distance ratios, inter-observer correlation
4 [42]	2015	Automatic identification of the oral transfer phase of deglutition	2,058 swallow and 3,248 non-swallow pressure measures & time periods	Deep neural network [Time-delay ANN]	A trained speech and language therapist marked the onset and offset of oral activity	Accuracy, receiver operating characteristic, mean squared error
5 [43]	2015	Differentiate osteoporotic patients from normal patients	2-D panoramic radiograph and bone mineral density of 141 females	Not deep-learning [naïve Bayes classifier, k-nearest neighbor, SVM]	The subjects were classified according to the World Health Organization	Receiver operating characteristic, sensitivity, specificity, accuracy
6 [44]	2015	Biofilm quantification independent of grader perceptual bias	2-D quantitative light-induced fluorescence images (maybe n=470)	Not deep learning [Gaussian Markov random field model]	To define clean areas, images were deemed clean by two expert graders	Confusion matrix, consistency
7 [45]	2016	Diagnosis of extractions	2-D lateral cephalograms of 156 patients	Non-deep learning neural network	Treatment plans were made by 1 orthodontic specialist	Success rate
8 [46]	2016	Diagnose females with osteoporosis	2-D dental panoramic radiographs of 141 females	Hybrid genetic swarm fuzzy classifier	Dual-energy X-ray absorptiometry and WHO guidelines on classification	Accuracy, sensitivity, specificity, PPV, NPV, likelihood ratio
9 [47]	2016	Predict if patient has teeth periapical lesion or not and its type using ML techniques	201 2-D dental x-ray images	Feed Forward Neural Networks, K-Nearest Neighbor Classifier	Not mentioned explicitly	Accuracy of classifier on two-class dataset, on four-class dataset

10 [48]	2016	Tooth segmentation and classification	3-D MicroCT-images of 3 mandibular molars (each tooth had 256 slices and 280 regions of interest)	Pulse coupled CNN	Experienced dentists manually labeled tooth structures to identify different regions in a segmented mask.	Relative error, correlation coefficient, mean absolute difference of volumes, similarity index, sensitivity, specificity
11 [49]	2017	Stage lower third molar development for age estimation	400 2-D panoramic radiographs	CNN [AlexNet]	Two observers decided about the stages. If necessary, a third observer	Rank-N recognition rate, intra-class correlation, accuracy, linearly weighted Cohen´s kappa, confusion matrix
12 [50]	2017	Osteoporosis detection using various methods on radiography	454 2-D dental panoramic radiographs	Decision tree and SVM	Radiographs classified by bone mineral density (T-score)	Accuracy, sensitivity, specificity, mean of textural features
13 [51]	2017	Classify quantitative light-induced fluorescence images to determine dental plaque level	427 quantitative light-induced fluorescence images	CNN [ResNet]	3 ground truths derived from 3 different plaque-scoring systems	F1-score
14 [52]	2017	Diagnose vertical root fractures in intact and endodontically treated teeth	240 2-D periapical dental radiographs and 3-D CBCT images	Non-deep learning neural network as a perceptron [Daubechies 3 wavelet transform, Gabor filters]	Teeth were evaluated under a microscope for presence/absence of a fracture	Accuracy, sensitivity and specificity
15 [53]	2017	Prediction of oral cancer risk in patients with oral leukoplakia	Exfoliative cytology, histopathology and clinical data from 364 patients	SVM, SVMfull, k-nearest neighbors, peaks-closed and peaks-random forest	Clinical data, exfoliative cytology, histopathology, and follow-up data were collected.	Sensitivity, specificity, area under the curve
16 [54]	2017	Investigate the application of deep CNN for classifying types of teeth on CBCTs	52 3-D CBCT images (35,259 regions were classified in 7 tooth types)	CNN [AlexNet (Caffe framework)]	The smallest bounding box for each tooth was manually placed on the CT volume	Classification accuracy, effect of augmentation on accuracy
17 [55]	2017	Teeth detection in dental panoramic radiographs with CNN	2-D dental panoramic radiographs from 100 people	CNN [modified version of AlexNet where multi-class classification is performed]	Each tooth in the images was delineated by a dentist	Accuracy for tooth class detection
18 [56]	2017	Classification of dental diseases using CNN	251 Radio Visiography x-ray images	CNN [VGG-16 (for transfer learning)]	Images were labeled by dentists and radiologists	Accuracy

19 [57]	2017	Development of an ANN to classify dental cusps with sufficient accuracy	3-D surface scans of 129 dental casts (full arches) from 69 participants	Non-deep learning neural network (Cusp Distance & Range Image Method)	Manual classification of cusps by an investigator using the modified FDI scheme	Correct classification
20 [58]	2017	Segmentation of gingival diseases from oral images	405 2-D intra-oral color-augmented fluorescent images	CNN [Auto-encoders]	A dentist drew bounding boxes around inflamed gingiva and gave a modified gingival index	AUROC, precision, recall
21 [59]	2017	Detection of tooth caries	over 3000 2-D bitewing radiographs	Fully CNN [not mentioned]	Annotations by dentists after clinical verification of caries	Recall, precision, F1 score
22 [60]	2017	22 methods were compared to analyze and improve dental age estimation in children	976 2-D panoramic X-rays	Not deep learning [22 models were used]	Teeth were divided into 14 sub-stages and assigned a numerical value	Mean absolute error, Root mean squared error
23 [61]	2017	Classify periapical cyst and keratocystic odontogenic tumor	50 3-D CBCT images from 50 patients	Deep neural network [Details not mentioned]	Experts classified and manually marked the lesions	Accuracy, F1-score, confusion matrix (not presented)
24 [62]	2018	Estimate positioning error of patient´s dental arch and correct the panoramic image	5166 pairs of 2-D dental panoramic radiographs and it's deviation value	CNN [Built on own]	Reconstruction with dental arch in predefined position	Mean absolute error, Maximum absolute error
25 [63]	2018	Mandible segmentation on a valid ground truth dataset	20 3-D CT data sets	Fully CNN [with 32s, 16s, and 8s separately]	Generated by 2 clinical experts manually	All networks: Dice coefficient; Best trained model: accuracy
26 [31]	2018	Classify normal, abscessed, and impacted teeth	60 2-D periapical dental radiographs	Not deep-learning	from dataset	Accuracy of all models in different set-ups of images
27 [64]	2018	Interactive segmentation of panoramic radiographs	2-D dental panoramic radiographs (maybe 5)	Conditional spatial fuzzy C-means clustering algorithm [Gaussian Kernel-based]	Manual generation of ground truth of 5 images by one doctor.	Misclassification error, relative foreground area
28 [65]	2018	Laser speckle image segmentation of tooth surface to detect early-stage caries	2-D laser speckle images, data size not mentioned	Not deep learning [K-means clustering algorithm]	Evaluation of samples (original treated teeth) by one trained odontologist	Accuracy in segmentation

29 [66]	2018	Find the determinant location factors of an inserted implant, which influence implant survival	Explanatory variables, survival, and complication of 53 patients (59 cases)	Not deep learning [Decision tree, SVM]	A prosthodontist evaluated the implants and categorized them according to chart records	Accuracy
30 [67]	2018	Identification of unknown people by comparing ante- and post-mortem panoramic radiographs	43467 2-D dental panoramic radiographs from 24545 persons	Speeded Up Robust Features + random sampling consensus algorithms	Given by dataset	Number of matching points, detection rate
31 [68]	2018	Classify head and neck CT for presence of dental artifacts	1417 2-D panoramic radiographs	Mask R-CNN [ResNet101 +region proposal network]	Annotation of the mouth, no additional information	Accuracy, F1-score, precision, recall, specificity
32 [69]	2018	Usage of a multi-stream deep learning framework for teeth-brushing (activity) recognition	2-D brush photos and data from smart bracelets of 74 people	CNN [VGG-19]	Data samples are manually labeled according to Bass brushing method	Accuracy of classifying 16 different movements of Bass teeth-brushing (confusion matrix)
33 [70]	2018	Predict BRONJ occurrence in at-risk patients	125 patient parameters (41 cases and 84 controls)	Logistic regression, SVM, Decision tree, ANN, Random Forest	Standard definition of BRONJ was used	AUROC, sensitivity, specificity
34 [71]	2018	Diagnose and predict periodontally compromised teeth	1740 2-D periapical radiographs	CNN [VGG-19]	3 calibrated periodontists determined the severity of periodontally compromised teeth	Diagnostic & predictive accuracy, sensitivity, specificity, PPV, NPV, AUROC, confusion matrix
35 [72]	2018	Evaluation of the efficacy of deep CNN algorithms for detection and diagnosis of dental caries on periapical radiographs	3000 2-D periapical radiographs	CNN [Inception V3, GoogLeNet]	All images were revalidated and dental caries were distinguished from non-dental caries by 4 calibrated dentists	Diagnostic accuracy, sensitivity, specificity, PPV, NPV, AUROC
36 [73]	2018	Classify incisor, canine, premolar and molar	3-D dental CT of 200 teeth (50 from each category)	Extreme learning machine [1-hidden layer network]	From dataset	Sensitivity of each class, entire accuracy
37 [74]	2018	Detect and quantify cracks using high-resolution CBCT images	42 3-D high resolution scans	Not deep learning [SVM]	Given by own dataset	Absolute maximum wavelet coefficient, AUROC, discrimin- ative sensitivity and specificity
38 [75]	2018	Screen high-risk populations for oral cancer	170 autoflourescence and white light image pairs	CNN [MatConNet]	Labeled by oral oncology specialists	Accuracy, sensitivity, specificity

39 [76]	2018	Assess the need for orthodontic treatment in patients with permanent dentition	15 variables and need of orthodontic treatment from 1000 patient datasets	Not deep learning [Bayesian network]	Stated need for orthodontic treatment mentioned in the hospital system	Accuracy, specificity, sensitivity, kappa, AUROC
40 [77]	2018	Automated clinical quality evaluation for decision making	196 2-D dental periapical radiograph pairs	CNN [modified GoogLeNet]	3 dentists classified cases based on clinical observation	Test accuracy, F1-score, recall, precision, confusion matrix
41 [78]	2018	Predict self-reported tooth mobility in urban Latinos	4623 Latinos with 9 social variables	Neural network [Multilayer Perceptron]	Self-reported by participants	Predictive accuracy, AUROC, interpretability, applicability
42 [79]	2018	A) Locate each present tooth	3-D CBCT images from 1274 studies	Combinations [Fully CNN: V-Net]	Annotated by 4 specialists and entering a tooth number	Accuracy
		B) Detect common conditions		[CNN: DenseNet]	Annotated by 5 specialists	AUROC
43 [80]	2018	Analyze the determinants that affect presence or absence of caries/restorations	105 dietary and demographic features from 9812 subjects	Deep neural network [Built on own]	Given by database	Accuracy, loss function, AUROC, processing time, PPV, NPV
44 [81]	2018	Teeth recognition using label tree and cascade network structure	1000 dental periapical radiographs	CNN [3 CNNS using VGG-16]	Annotations with bounding boxes and label the 32 teeth	Precision, recall, F-score
45 [82]	2019	Real-time recognition of dental instruments using deep learning	631 images of 11 objects	CNN [Shot Multibox Detector network, MobileNet]	Images were annotated by the researchers	Accuracy, precision, recall, true negative rate
46 [83]	2019	Age estimation using CNN on dental panoramic X-rays	2575 2-D dental panoramic radiographs	CNN [Capsule-net which is built on own]	Given by dataset	Average accuracy, recall, precision, F1-score
47 [84]	2019	Detect periodontal disease using smartphones and ML techniques	2-D gingival images from 30 subjects	SVM	Given by dataset (cases were diagnosed by dentist)	Accuracy, sensitivity, specificity
48 [85]	2019	Detect decay on dental X-ray images to predict the needed treatment	120 periapical and 116 panoramic radiographs	CNN	Manual cropping of teeth by a dental expert and classification as per the needed treatment	Accuracy
49 [86]	2019	Localize dental lesions in near-infrared transillumination images	217 near-infrared transillumination images	Fully CNN [similar to U-Net, inspired by VGG16]	Reference segmentation maps from dental experts	Overall intersection over union for 5-class, AUROC
50 [87]	2019	Detect and number teeth in dental periapical films	1250 2-D dental digital periapical films	Faster R-CNN, Deep NN [Inception Resnet version 2 (for Faster R-CNN)]	1 dentist framed each intact tooth and provided a corresponding tooth number	Mean intersection over union to obtain precisions and recalls, boxes detected

51 [88]	2019	Improve the precision of dental hard tissue segmentation	20000 2-D intraoral images from 40 videos	Sparse representation-based classifier	Manually labeling dental tissue type pixel by pixel	Precision, recall, and their harmonic average
52 [89]	2019	Label teeth and identify root canal	250 2-D dental panoramic radiographs	2 CNNs comprised the generative adversarial network	A dentist marked each tooth and the gap between them	Accuracy, structural similarity index after every iteration
53 [90]	2019	Diagnose orthognathic surgery cases	12 measurements and 6 indices from 316 patients	Not deep learning [Built on own]	1 orthodontist decided the treatment plans	Decision-making success rates
54 [91]	2019	Tooth instance segmentation and identification from CBCT images	20 3-D CBCT images	Mask R-CNN [3-D region proposal network]	Annotation with a tooth-level bounding box, mask, & label	Accuracy, Dice similarity coefficient
55 [92]	2019	Segmentation of mental foramen	1000 2-D dental panoramic radiographs	Fully CNN [U-Net]	Annotation by radiologists	Dice similarity coefficient
56 [93]	2019	Develop a complete identification system to aid dental forensics without the use of radiation	Top view images of the teeth in the upper and lower jaw from 30 persons	Recurrent Neural Network [Built on own]	Given by dataset	Percent match, reliability, confusion matrix, accuracy to compute the correct area
57 [94]	2019	Segmentation of alveolar bone	50 2-D intraoral ultrasound images of 8 mandibular incisors	Fully CNN [U-Net]	Delineation by an orthodontist, medical physicist, and biomedical engineer	Dice coefficient, sensitivity, specificity, Hausdorff distance
58 [32]	2019	Detect apical lesions on panoramic dental radiographs	2001 segments from 85 2-D panoramic radiographs	CNN [Built on own]	Majority vote of 6 dentists on manually cropped segments.	AUROC, sensitivity, specificity, PPV, NPV
59 [95]	2019	Investigate a 3-D single image super-resolution method based on tensor factorization	3-D CBCT of 13 single teeth	Tensor factorization	State-of-the-art iterative deconvolution technique with low-rank regularization	Mean of absolute difference in Feret and Area, Dice coefficient, time, peak signal-to-noise ratio
60 [96]	2019	Resolution enhancement of 2-D CBCT image slices of ex vivo teeth	5680+1824 2-D CBCT slices of 17 ex vivo teeth & in vivo microCT images	CNN [inspired by U-Net and subpixel networks]	micro-CT images were used as ground truth	Peak signal-to-noise ratio, mean squared error, structure similarity index, Dice coefficient, mean difference: Feret, area, and volume

61 [97]	2019	Propose DL metal segmentation method for metal artifact reduction in dental CT	1000 3-D CBCT images and projection images from 4 patients	Fully CNN [U-Net]	Manually segmented metal regions on the training images using Adobe Photoshop CS6	Relative error, sum of square difference, normalized absolute difference, Jaccard index, Dice similarity index
62 [98]	2019	Classification of root morphology of mandibular first molars on panoramic radiographs	3-D dental CBCT and 2-D panoramic radiographs of 760 mandibular first molar	CNN [AlexNet and GoogleNet (DIGITS library on Caffe framework)]	Observations performed by a radiologist	Diagnostic accuracy, sensitivity, specificity, PPV, NPV, area under the curve
63 [99]	2019	Address low-dose artifacts in dental CT-scanning	24024 3-D dental CT image pairs (high-dose and low-dose)	Generative adversarial network, CNN	Blind reader study with 20 groups of images	Signal-to-noise ratio, structural similarity, image quality metrics, test time
64 [100]	2019	Select the most relevant variables to classify the presence and absence of root caries	Medical, dental, and physiological measures from 5135 people	Not deep learning model	Oral examination by experts using a decayed, missing, and filled surface index	Accuracy, sensitivity, specificity, AUROC, features that are associated with root caries
65 [101]	2019	Gender estimation from panoramic dental x-ray images	4155 2-D dental panoramic radiographs	CNN [VGG16]	Given by dataset	Accuracy
66 [102]	2019	Detect atherosclerotic carotid plaques on orthopantomograms	65 2-D dental panoramic radiographs	Faster R-CNN [Resnet-101]	2 oral medicine & maxillofacial radiologists marked lesions	Accuracy, sensitivity, specificity, AUROC
67 [103]	2019	Automatic detection of atherosclerotic carotid plaques in panoramic images	65 2-D dental panoramic radiographs	Faster R-CNN [Resnet-101]	2 oral medicine & maxillofacial radiologists jointly marked lesions	Sensitivity, specificity, receiver operating characteristic
68 [104]	2019	Survival prediction of oral squamous cell carcinoma patients	255 patient medical records	Deep neural network [DeepSurv (Multi-layer feed forward network)]	Given by hospital's medical records and according to a cancer staging manual	Prediction accuracy (Harrell’s c-index)
69 [105]	2019	Detect periodontal bone loss on identified teeth in panoramic dental radiographs	12179 2-D dental panoramic radiographs	Fully CNN [U-shaped architecture]	5 dental hygienists marked lesions independently (monitored by a dentist) and numbered the tooth	F1-score, AUROC, sensitivity, specificity, PPV, NPV

70 [106]	2019	Detect Sjögren's syndrome in praotid and submandibular salivary glands	200 patients resulted in 8000 augmented 2-D ultrasonography images	CNN [VGG16]	According to the Japanese criteria and American-European Consensus Group	Accuracy, sensitivity, specificity, AUROC
71 [107]	2019	Compute and improve semantic segmentation of dental panoramic images	1500 2-D dental panoramic radiographs (10 categories)	Fully CNN [U-Net]	Annotators outlined teeth at certain anchor points and interpolated between them.	Accuracy, specificity, precision, sensitivity, Dice score
72 [108]	2019	Detection of periodontal bone loss (PBL) on panoramic dental radiographs	85 2-D panoramic images cropped into 1737 single-tooth segments	CNN [a seven-layer feed-forward CNN]	3 examiners independently determined 3 points on each tooth to estimate PBL percent	Accuracy, area under the curve, F1-score, sensitivity, specificity, PPV, NPV
73 [109]	2019	Detect osteoporosis on dental panoramic radiographs	1268 2-D dental panoramic radiographs	CNN [AlexNet]	2 oral and maxillofacial radiologists independently diagnosed osteoporosis	Confusion matrix, accuracy, precision, recall, F1 score, AUROC
74 [110]	2019	Propose a 3-stage approach to recognize tooth-marked tongue	641 2-D tongue photographs	CNN (feature extractor), multiple-instance SVM (MI-SVM; classifier) [VGG-16 (CNN)]	Positive case: marked by traditional Chinese medicine practitioners. Negative case: generated by an algorithm	Accuracy, true positive rate, true negative rate
75 [111]	2019	Explore a smart dental system for in-home dental healthcare	12600 clinical images	Mask R-CNN [ResNet-50-C4]	Training sets were calibrated by 20 dental disease experts	Diagnosis accuracy, sensitivity, specificity, mean diagnosis time
76 [112]	2019	Optimization of PointNet++ to improve classification results	3-D point cloud data of 12311 CAD models	CNN [PointNet++]	Not applicable	Accuracy of different objects
77 [113]	2019	Biological gender estimation based on deep learning	4000 2-D dental panoramic radiographs	CNN [DenseNet201, InceptionResNetV2, VGG16, VGG19, ResNet50, Xception]	Given by dataset	Mean accuracy for different 1) networks, 2) attention, 3) number of filters, and 4) number of units
78 [114]	2019	Bone segmentation in CBCT scans affected by metal artifacts	20 3-D CBCT images	Mixed-scale dense CNN	Global thresholding and post-processing by a medical engineer	Mean Dice similarity coefficients, mean absolute deviations

79 [115]	2019	Investigate the effect of different augmentation methods on a MapReduce-like model	820 dental front oral images	CNN [AlexNet, generative adversarial network for augmentation]	8 pocket depths of tooth measured by a few dentists	Accuracy, sensitivity, specificity, receiver operating characteristic
80 [116]	2019	Detect and classify occlusal caries	88 in vivo dental images	Mask R-CNN [Based on Feature Pyramid Network and ResNet101]	Superpixels comprising dental lesions were marked by the Dental Annotator version 1.5.1	Classification: micro F-measure (Accuracy); Precision / Recall / F-measure for each class
81 [117]	2019	Diagnose maxillary sinusitis on dental panoramic radiographs	920 image patches from 2-D panoramic, CT, and CBCT images	CNN [AlexNet (with DIGITS library, Caffe framework)]	Lesions were verified by their appearances on CT or CBCT	Accuracy, sensitivity, specificity, AUROC, PPV, NPV
82 [118]	2019	Describe the impact of orthognathic treatment on facial attractiveness and age look	2164 pre- and post-treatment photographs from 146 patients	CNN [VGG-16]	Age labels and attractiveness scores were derived from datasets used	Differences between real and apparent age, and real and apparent attractiveness
83 [119]	2019	Evaluate facial attractiveness of treated cleft patients & controls	60 frontal and left-profile images from 30 patients	CNN [VGG-16]		Mean difference, co-efficient of variation in rating
84 [120]	2019	Automatically detect the type of lesion in periapical x-rays	534 2-D periapical radiographs	CNN [Alexnet]	Labeled by experienced radiologist and dentist	Classification accuracy
85 [121]	2019	Facilitate diagnosis and treatment by providing easy access to the International Association for Dental Traumatology (IADT) guideline	32 forms filled according to IADT guideline	Deep neural network [Multilayer perceptron, Kstar, instance-based k-classifier, sequential minimal optimization, logistic regression]	Own dataset curated according to the rules given in IADT guideline	Accuracy of classification, kappa, root mean square error, mean absolute error
86 [122]	2019	Explore ensemble and deep learning for real-time sensors in smart toothbrush devices	1.44 million instances and 144,000 features each for 10 individuals	Details on depth of models not mentioned	The smart toothbrush saved the correct tooth and surface brushed labels in a database	Accuracy, precision, recall, F1-score, training time, prediction time, model size (bytes)
87 [123]	2019	Segmenting and classifying tooth types on 3-D dental models	600 3-D dental models	3-D CNN [O-CNN]	Given by dataset	Accuracy, specificity, recall, macro -accuracy, -specificity, and -recall

88 [124]	2019	Detect and number teeth on panoramic radiographs	1574 2-D dental panoramic radiographs	Faster R-CNN: VGG-16	5 radiologists numbered and marked all teeth (FDI system)	Sensitivity, precision, specificity
89 [125]	2019	Automated high-performance segmentation of third molars and inferior alveolar nerve canal	81 2-D digital panoramic radiographs	Fully CNN [based on U-Net]	Regions were manually segmented and labelled; another observer refined them	Dice coefficient, Jaccard index, sensitivity, specificity
90 [126]	2019	Differentiate post-cancer from healthy muscle coordination	MRI scans with time i.e., 4D data of 26 subjects	CNN	From dataset	Accuracy
91 [127]	2019	3-D dental model/ mesh segmentation	1200 3-D dental meshes	2 separate CNNs	Manually labeled dental meshes provided by company	Accuracy, mean errors
92 [128]	2019	Predict the debonding probability of CAD/CAM crowns	8640 2-D images of 3-D stereolithography die models	CNN [Built on their own]	Labels of 'trouble-free' and 'debonding' were assigned to each crown/die	Predictive accuracy probability, precision, recall, F-measure, AU-ROC, mean calculate time
93 [129]	2019	Correlation of systemic health conditions with periodontal disease	1215 2-D intraoral fluorescent images	CNN [Auto-encoders]	Physicians independently assigned localized and image-wide modified gingival indices	Area under the curve, true and false positive rates, precision, recall, mean intersection over union
94 [130]	2020	Automatic detection and classification of dental restorations	83 2-D dental panoramic radiographs (738 dental restorations)	Not deep learning [Cubic SVM with Error-Correcting Output Codes]	1 oral medicine specialist identified and labeled the existing dental restorations	Accuracy, detection rate, sensitivity, specificity, PPV, NPV
95 [131]	2020	Automatic detection of periodontal disease in orthodontic patients	134 intraoral images which were split into 804 regions	Faster R-CNN [2 models, each used ResNet-50]	Gingiva was annotated and labeled by dentists using the Löe and Silness gingival index	Detection accuracy, precision, recall, mean average precision
96 [132]	2020	Compute mandibular indices for detecting the thinning and deterioration of mandibular bone	370 2-D dental panoramic radiographs	Fuzzy K-means classification algorithm to identify artificial structures	2 dentists applied a semi-automatic process to define the required lines and points	Distances between relevant points
97 [133]	2020	Fully automated third molar development staging (localization, segmentation, and classification)	400 2-D panoramic radiographs	CNN + Fully CNN [Localization: YOLO-like CNN. Segmentation: U-Net-like	Staging by three observers. Same as in de Tobel 2017	Mean absolute error, mean Euclidean distance, precision, recall, Dice score, accuracy, linear

				CNN. Classification: 2 CNNs]		weighted Cohen´s kappa, time for analysis
98 [25]	2020	Detection of caries lesions of different radiographic extension on bitewing radiographs	3686 bitewing radiographs	Fully CNN [U-Net]	Images were annotated and labeled by 3 dentists and reviewed by a 4th dentist	Accuracy, sensitivity, specificity, F1, PPV, NPV, Matthew´s correlation
99 [134]	2020	Detect and classify/ stage periodontal bone loss of each individual tooth	340 2-D dental panoramic radiographs	Mask R-CNN [Based on a feature pyramid network and ResNet101]	Oral and maxillofacial radiologists manually delineated the relevant areas	Accuracy, Dice score, Jaccard index, Pearson correlation, mean absolute difference, intraclass correlation
100 [135]	2020	Assess maxillary variation in unilateral canine impaction	96 3-D CBCT images	Learning-based multi-source IntegratioN frame worK for Segmentation	36 CBCT images manually segmented	Average dice ratio, intraclass correlation, difference in volume
101 [136]	2020	Segment individual teeth in dental CBCT images	25 3-D CBCT images (more than 770 teeth)	Fully CNN [modified V-net architecture]	Each tooth was manually segmented and morphological operations generated the reference image	Jaccard similarity coefficient, Dice similarity coefficient, relative volume difference, average symmetric surface distance
102 [137]	2020	1) Pose-aware volume-of-interest realignment	175 3-D CBCT images	CNN: Modified VGG-16 to output a 6D tensor	Manual annotation and classification of CBCT images according to percentage of metal artifacts by clinical experts	
		2) Tooth detection		Modified Faster R-CNN (Region proposal network)		Average precision, overlapping ratio, object include ratio
		3) Individual segmentation network		CNN: Adopted the base architecture of 3-D U-Net		F1 score, aggregated Jaccard index, precision, sensitivity, Hausdorff distance, average symmetric surface distance
103 [138]	2020	Investigate how 24 oral and maxillofacial surgeons assess the presence of periapical radiolucencies	3099 2-D dental panoramic radiographs	Fully CNN [U-Net]	Pulp vitality was tested using thermal and percussion tests	Mean true positive rate (TPR), precision, F1 score, positive predictive value (PPV), area under

						PPV-TPR curve based on Riemann summation
104 [139]	2020	Automatic human identification system	15,868 2-D dental panoramic radiographs	CNN	5 bony landmarks were labeled manually	Accuracy, recall, precision, F1-score, true and false rates, AUROC, cumulative match curve
105 [140]	2020	Diagnose various periodontal diseases	300 patients and 11 variables	Not deep learning [SVM]	The professional's diagnosis	Accuracy, hypervolume under manifold value
106 [141]	2020	Predict disease-free survival in patients with oral squamous cell carcinoma	3 planes of 18F-fluorodeoxyglucose PET images from 113 patients	CNN [ResNet-101]	From patients' medical records	Accuracy, sensitivity, specificity, PPV, NPV
107 [142]	2020	Evaluate the relationship between mandibular third molar and the mandibular canal	600 2-D dental panoramic radiographs	CNN [AlexNet, GoogleNet, VGG-16]	Radiologists marked the relationship of the roots and canals in all images	Accuracy, time, storage space, sensitivity, specificity, AUROC, intra- and inter-CNN consistency
108 [143]	2020	Detect vertical root fracture on panoramic radiographs	300 2-D dental panoramic radiographs	CNN [DetectNet (with DIGITS version 5.0)]	Detection: 2 radiologists and 1 endodontist. Marked by 1 radiologist	Recall. precision, F measures
109 [144]	2020	Dental caries diagnosis using a back-propagation neural network for classification	105 intra-oral digital radiography images	Non-deep learning neural network as a perceptron [Back-propagation network]	Caries was annotated by a dentist.	False positive rate, accuracy, AUROC, precision recall curve area, learning rate, momentum, precision, recall, F measure, Matthew's correlation coefficient
110 [145]	2020	Risk prediction of unmet dental care needs in USA	33,929 participants and 237 variables	decision tree classifier [Built on own]	Given by dataset/ survey responses	Accuracy, sensitivity, specificity, precision, area under the curve
111 [146]	2020	Predict patients at risk of all-cause dental 30-day hospital readmission	Variables data on 11,341 cases	Decision Tree, SVM, k-nearest neighbor, ANN, logistic regression	Given by dataset	Area under the curve, accuracy, sensitivity, specificity, precision
112 [147]	2020	Automatic localization of the mandibular canal	637 3-D CBCT scans from 594 patients	Fully CNN [similar to U-Net]	Annotation by 2 medical professionals	Dice similarity coefficient, recall, precision, average symmetric

						surface distance, mean curve distance, Hausdorff distance
113 [148]	2020	Determine whether CNNs can judge soft tissue profiles requiring orthognathic surgery using facial photographs alone	2-D front and right facial photos + posteroanterior and lateral cephalometry from 822 patients	CNN [VGG19]	2 orthodontists, 3 maxillofacial surgeons, and 1 maxillofacial radiologist classified patients into Groups I and II	Accuracy, precision, recall, and F1 scores
114 [149]	2020	Person authentication with deep learning technique	750 2-D hand radiographs	CNN [k-nearest neighbor and SVM]	Given by dataset	Percentage of cross-validation accuracy
115 [150]	2020	Detection and segmentation of the mental foramen	112 2-D dental panoramic radiographs	Fully CNN [U-Net]	Oral medicine specialists with training in radiology	Dice similarity, recall, precision, true and false positive rates
116 [151]	2020	Investigation of automated feature detection, segmentation, and quantification of common findings in radiographs	206 2-D periapical radiographs	Computer vision, CNN + Fully CNN [U-Net, Xnet, SegNet]	Labeling by 2 oral pathologists and 1 endodontist. One expert labelled and the other 2 accepted or rejected them.	Mean intersection over union, Dice coefficient
117 [152]	2020	Develop a fully automated cephalometric analysis	2075 lateral 2-D cephalograms	CNN [stacked hourglass network]	2 orthodontists corrected and marked new landmarks	point to point error, successful detection and classification rate
118 [153]	2020	Automatically identify and classify skeletal malocclusions	218 3-D CBCT	CNN [VGG 16 and additionally Inception-V3]	Manual ground truth creation by clinical experts	Accuracy, precision, recall, F1 score
119 [154]	2020	Identify 4 different types of implant fixture systems	801 2-D periapical radiographs	CNN [SqueezeNet, GoogLeNet, ResNet-18, MobileNet-v2, ResNet-50]	From patient records	Accuracy, precision, recall, F1 score for each network
120 [155]	2020	Creation an automated cephalometric X-ray analysis	1792 cephalometric images	CNN [Built on own]	6 orthodontists and 6 in-training orthodontists marked 18 landmarks	Pearson product–moment correlation, Bland–Altman plots
121 [156]	2020	Identify how swallow sounds correspond to swallowing and how swallow times differ by viscosity	226 subjects gave rise to 1859 swallows and 2021 noise samples	CNN	Videofluoroscopic swallow images labelled by 2 medical professionals	Change in swallow duration

122 [157]	2020	Classify maxillary impacted supernumerary teeth in patients with fully erupted incisors	550 2-D dental panoramic radiographs	CNN [AlexNet, VGG-16, DetectNet]	Images reviewed by 2 radiologists	Accuracy, sensitivity, specificity, AUROC, recall, precision, F-measure
123 [158]	2020	Mandibular canal detection using a deep CNN	102 3-D CBCT images	Fully CNN [2-D SegNet, 2-D and 3-D U-Net]	2 researchers traced the canal. An oral and maxillofacial radiologist reviewed vague cases	Pixel-, global-, class-accuracy, mean intersection of union
124 [159]	2020	Reduce metal artifact for sinogram and dental CT images	3-D CT of 33 teeth phantoms with metal implants	Fully CNN	Sinograms and CT images from teeth phantoms without metal implants	Root-mean-square error, structural similarity
125 [160]	2020	Automated tooth segmentation using individual annotation	864 images from 50 2-D dental panoramic radiographs	Mask R-CNN [ResNet-101]	1 oral radiologist performed annotated teeth on 30 training panoramic radiographs	F1 score, mean intersection of union, visual analysis
126 [161]	2020	Identification and classification of dental implant systems	10,770 cropped images from 2-D panoramic and periapical radiographs	CNN [GoogLeNet Inception-v3]	Regions of interest manually cropped and labeled by 3 periodontology residents	Sensitivity, specificity, AUROC, confusion matrix
127 [162]	2020	Detection and diagnosis of odontogenic cysts	1,140 2-D panoramic radiographs and 986 3-D CBCT	CNN [GoogLeNet Inception-v3]	Histopathological examinations by an oral pathologist	AUROC, sensitivity, specificity, confusion matrix with and without normalization
128 [163]	2020	Classification of dental implant systems on panoramic and periapical radiographs	7,146 2-D dental panoramic and 4,834 periapical radiographs	CNN [Built on own]	Manual classification by 5 periodontal residents and confirmed by 3 periodontists	AUROC, standard error, Youden index, sensitivity, specificity
129 [164]	2020	Locate cephalometric landmarks with confidence regions	400 2-D lateral cephalograms	CNN [Bayesian CNN]	A junior and senior orthodontist independently annotated	Mean landmark error, successful detection rate, confusion matrix
130 [165]	2020	Classify specific osteoporosis features in dental radiographs	680 2-D dental panoramic radiographs	CNN [CNN3, VGG-16]	T-Score for osteoporosis detection	Accuracy, sensitivity, specificity, receiver operating characteristic, precision recall curve
131 [166]	2020	Tooth segmentation on CBCT images for dental implant planning	102 3-D CBCT datasets (each dataset has 264 to 727 2-D image slices)	Fully CNN [U-Net + dense block + spatial dropout]	Manually classification of images	Recall, precision, Dice score

132 [167]	2020	Identification of tongue color, fur color, crack, and tooth mark in traditional Chinese medicine	2-D tongue photos from 200 subjects	CNN [YOLO V3 optimized for this study for classification task]	2 Traditional Chinese Medicine diagnostic experts	Accuracy rate, precision rate, recall rate
133 [168]	2020	Automatic tooth root segmentation on CBCT images	1521 3-D CBCT images	CNN [A combination of Recurrent neural network + Attention U-Net]	CBCT images were classified into 3 classes	Intersection over union, average precision and recall rate, Dice similarity coefficient, average symmetrical surface distance
134 [169]	2020	Intelligent dental plaque segmentation using oral endoscope images	607 oral endoscopic images	CNN+HKS+LBP, random forest [DeepLabV3+]	Dentists cropped and marked plaque regions referring to post-stained images	Super-pixel accuracy, training time, intersection over union, out-of-bag error curves
135 [170]	2020	Automated tooth labeling on raw dental surfaces	Raw maxillary surfaces acquired by 3-D intraoral scanners of 30 subjects	CNN [MeshSegNet an extension of PointNet]	Segmentations done by a resident guided by experienced dentists	Dice similarity coefficient, sensitivity, positive predictive value
136 [171]	2020	Automated teeth recognition from panoramic images	1000 2-D dental panoramic radiographs	Faster R-CNN [ResNet-101, ResNet-50]	Each tooth with proper roots and shape was labelled	F1 score, precision, recall, mean average precision
137 [172]	2020	Predict difficulty level of endodontic cases and decide about a referral	500 filled American Association of Endodontist Endodontic Case Difficulty Assessment Forms and radiographs	2 ML algorithms were used, out of which 1 was a deep neural network	Assessment of forms by 2 endodontists, in case of conflict third endodontist´s opinion was taken	Accuracy, sensitivity, specificity, precision
138 [173]	2020	Personal identification with paired orthopantomographs obtained in relatively short period	30 pairs of orthopantomographs from 30 participants	CNN [VGG16, ResNet50, Inception-v3, Xception, InceptionResNet-v2, MobileNet-v2]	From the university hospital	Detection accuracy, precision, recall
139 [174]	2020	Automated third molar stage allocation for age estimation	400 2-D panoramic radiographs	CNN [DenseNet201]	2 observers staged FDI 38 with modified Demirjian scale. Another observer reviewed cases of disagreement.	Accuracy, mean absolute difference, linearly weighted Cohen´s kappa

140 [175]	2020	Recognize dental defect using Adaptive CNN and Bag of Visual Word	447 2-D panoramic images	Adaptive CNN [pre-trained VGG16]	Images were labeled and sorted by dentists based on 3rd molar appearance	Accuracy
141 [176]	2020	Identify CT slices for head and neck cancer with dental artifacts	1164 axial slices in pairs from 44 CT scans	Model for kernel L2-Wasserstein distance	CT slices were classified by a medical imaging expert	Prediction rate, computation time
142 [177]	2020	Detection, localization, and volume determination of periapical pathosis on CBCT	3900 3-D CBCT images	Fully CNN [U-Net-like architecture]	Annotations by maxillofacial radiologists and automatically examined to eliminate errors	Reliability of correctly detecting a periapical lesion, recall, precision, F-measure
143 [178]	2020	Automatic detection of trabecular landmarks	108 dental panoramic radiographs	CNN [statistic shape model]	8 osteoporotic regions annotated by dentists	Loss of 5-fold cross validation, mean and median loss
144 [179]	2020	Detection of caries lesions in Near-Infrared-Light Transillumination images	226 NILT images of single tooth segments	CNN [Resnet18, Resnext50]	Caries annotated by 2 dentists	Accuracy, AUROC, sensitivity, specificity, PPV, NPV
145 [180]	2020	Automated segmentation of CBCT images and detection of periapical lesions	20 3-D CBCT images	3-D CNN [multi-label U-Net]	Segmentation performed and revised by 1 maxillofacial radiologist, 1 endodontist, and 1 senior graduate in radiology honors program	Sensitivity, specificity, PPV, NPV, DICE index
146 [181]	2020	Classify and clarify the accuracy of different dental implant brands	8859 image segments from 6513 2-D dental panoramic radiographs	CNN [VGG16, VGG19]	Electronic medical records and dental implant usage ledger of the department	Accuracy, precision, recall, receiver operating characteristic, F1-score, gradient-weighted class activation maps
147 [182]	2020	Automatic and accurate segmentation and identification of individual teeth	100 3-D digital dental casts	CNN [feature steered graph CNN which used FeaStNet]	Given by dataset	Labeling accuracy, Dice similarity coefficient
148 [183]	2020	Classify partially edentulous dental arches for designing removable partial dentures	1184 oral photographs of dental arches	CNN [ResNet152 (using Tensorflow, Keras deep learning libraries)]	Arch types judged by authors	Diagnostic accuracy, precision, recall, F-measures, AUROC, percentage of correct predictions

149 [184]	2020	Tongue region and landmark detection	1838 2-D tongue photographs	CNN [combination of Image Pyramid, Coarse-Net, Fine-Net, Refine-Net]	Labeled by 2 primary physicians and a resident physician	Precision, recall, accuracy, F1-score, intersection over union, mean error rate, failure rate
150 [185]	2020	Identify periodontally compromised teeth	100 2-D digital dental panoramic radiographs	Faster R-CNN [ResNet-101]	Annotations by 3 periodontology experts	Average precision and recall rate, sensitivity, specificity, F-score
151 [186]	2020	Estimate the chronological age of a subject from panoramic image	2289 2-D dental panoramic radiographs	CNN [Built on own]	Images were labelled with the subject's date of birth and the date of image	Absolute error, coefficient of determination, accuracy, area under the interquartile coefficient of receiver operating characteristic
152 [187]	2020	Automatic segmentation of mandibular molar and predict its eruption potential	838 2-D panoramic radiographs	CNN [ResNet-101]	Human reference measurements	Accuracy, Bland Altman plot, intersection over union, recall, Hausdorff distance, analysis time, precision
153 [188]	2020	Recognition of tooth-marked tongue	1548 2-D tongue photographs each in 2 datasets	CNN [ResNet34]	Classification by 3 traditional Chinese medicine practitioners	Accuracy, sensitivity, specificity
154 [189]	2020	Predict Children's Oral Health Status Index (COHSI) score and referral for treatment needs (RFTN)	Short-form survey responses from 545 families	Extreme gradient boosting, Naïve Bayesian algorithms	A dental exam to evaluate the clinical oral health outcomes summarized as COHSI score and RFTN	Residual mean square error, correlation, sensitivity, specificity
155 [190]	2020	Classify dental artifacts status	1538 head and neck CT images	3-D CNN	Classification by an observer	Area under precision recall curve, precision, recall
156 [191]	2020	Dental artifact detection for CT	2112 head and neck CT images	CNN	A single observer	Receiver operating characteristic
157 [192]	2020	Presentation of a novel strategy to combine bounding box annotations from multiple clinicians	2155 oral cavity images	Combinations [ResNet-101 (image classification), Faster R-CNN (object detection)]	800 images annotated by 3-7 clinicians; the remaining 1355 images annotated by 1 clinician	F1-Score, precision, recall

158 [193]	2020	Evaluate diagnostic performance of CNN You Only Look Once (YOLO) v2	1603 2-D dental panoramic radiographs	CNN [YOLO v2]	Histopathologic diagnosis	Precision, recall, accuracy, F1 score, average time to evaluate test datasets, confusion matrix
159 [194]	2020	Detect plaque on primary teeth	886 tooth photos	CNN [DeepLab network, DeepLabV3+]	Plaque-disclosing agent used and identified by a researcher	Mean intersection-over-union
160 [195]	2020	Automatic detection of dental caries from oral photos	3932 oral photos	CNN [VGG-16]	Labeled by 1 of 3 dentists by clinical visual-tactile exam	AUROC, Free-Response receiver operating characteristic
161 [196]	2020	Perform segmentation and lesion detection on CBCT images	100 3-D CBCT images	CNN [Fully convolutional Densenet with some modifications (which is essentially a U-Net)]	Manual and semi-automatic segmentation; revised by 1 oral and maxillofacial radiologist, 1 endodontist, and 1 senior graduate in radiology honors program	detection accuracy, precision, recall (DICE index calculated but not presented numerically)
162 [197]	2020	Automatic tooth detection and segmentation	100 dental front photos	Mask R-CNN [deeper ResNet101+ Feature Pyramid Network]	Image labeling tool used to form multiple polygons around teeth	Pixel accuracy
163 [198]	2021	Predict global five-year survival in oral cancer and its cancer recurrence	Variables from 416 patients	Logistic regression, K-nearest neighbor, Naïve Bayes, Decision tree, Random Forest classifier	Histological diagnosis of oral squamous cell carcinoma	Receiver operating characteristic, accuracy, sensitivity, specificity, F1 score
164 [199]	2021	Tooth detection and segmentation	153 2-D panoramic radiographs	CNN [DeepLab-v3, ResNet-101]	A dentomaxillofacial radiologist labeled and segmented each tooth	Accuracy, time, sensitivity, recall, F1-score, precision, intersection over union, Hausdorff distances
165 [200]	2021	Predict genetic risk of nonsyndromic oral clefts	Nucleotide sequences from 1588 participants	Non-deep learning neural network as a perceptron	Known from dataset	Accuracy, error rate, interactions of nucleotide sequences
166 [201]	2021	Detect and classify teeth for automatic filing of dental charts	100 2-D dental panoramic radiographs	CNN [DetectNet, ResNet]	A dental radiologist localized and classified each tooth	Accuracy, detection sensitivity, number of false positives

167 [13]	2021	Compare cost-effectiveness of proximal caries detection with versus without AI	3686 bitewing radiographs	Fully CNN [U-Net]	4 dentists marked carious lesions	Accuracy, sensitivity, specificity, effectiveness, cost, incremental cost-effectiveness ratio
168 [202]	2021	Automatic detection system for numbering teeth	1125 2-D dental bitewing radiographs	Faster R-CNN [Inception v2 (COCO)]	A radiologist annotated images and with tooth numbers	Accuracy, confusion matrix, F1 score, precision, sensitivity
Abbreviations: AI, artificial intelligence; ANN, artificial neural network; AUROC, area under the receiver operating characteristic; BRONJ, bisphosphonate related osteonecrosis of the jaw; CBCT, cone beam computed tomography; CNN, convolutional neural network; FDI, Federation Dentaire Internationale; ML, machine learning; NPV, negative predictive value; PPV, positive predictive value; SVM, support vector machine; VGG, visual geometry group.						

Source: modified from Table S1, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937. Data rights held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under license CC BY 4.0 as per publisher MDPI policy.

Table 2: Studies excluded from the scoping review on machine learning in dentistry along with the reason for exclusion (n=15).

No. [Citation]	Reason for exclusion from the scoping review
1 [203]	Poor methodology/ reporting • Reference test for the training and validation datasets was generated by only one professional expert, who was not formally trained in dentistry but was a biomedical engineer. • The validation set was also utilized during training to determine when to stop the parameter update to prevent overfitting.
2 [7]	A review article
3 [204]	Not an oral health topic
4 [205]	A review article
5 [206]	A systematic review article
6 [207]	No machine learning method used
7 [208]	No machine learning method used
8 [209]	Poor methodology/ reporting • Labeled bounding boxes were generated by a software tool to serve as the reference test for the training dataset but were not checked for errors by a human expert. • Model architecture not adequately described, for example, number of convolutional layers. • Some results are shown via images which have poor resolution. • Absence of the ‘Discussion’ section of the paper. Hence placing the results in the context of the previous and current research is lacking.
9 [210]	A conceptual review article
10 [211]	Not an oral health topic
11 [212]	Poor methodology/ reporting

	• The paper does not discuss how its specific research question is tied to the larger context of oral health in USA. • The study used 6 deep neural network models for variable selection but no further details are given. • The study also used 10 data mining algorithms, whose names are listed but no further details are provided.
12 [213]	A supplement article (similar to a review article)
13 [214]	Not an oral health topic
14 [215]	Poor methodology/ reporting • The authors selected 19 feature variables or elements that characterize orthodontic problems and are assumed to be important in extraction decisions based on existing orthodontic literature. But these 19 variables or elements are not named or described further. • Performance metrics, such as accuracy and error rate, were measured and reported via bar-charts but were not specified in the text. This hampered the evaluation of the results and their interpretation.
15 [216]	A review article

Source: modified from Table S2, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937.

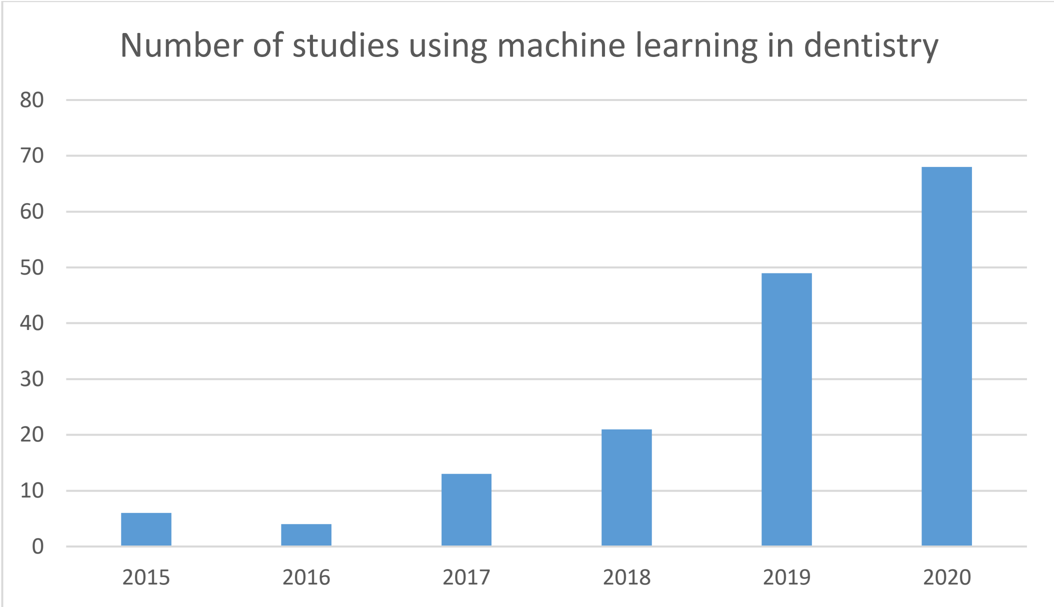

Figure 5. Temporal trend in number of publications included in the scoping review on machine learning in dentistry between 1st January 2015 and 31st May 2021. Source: own representation.

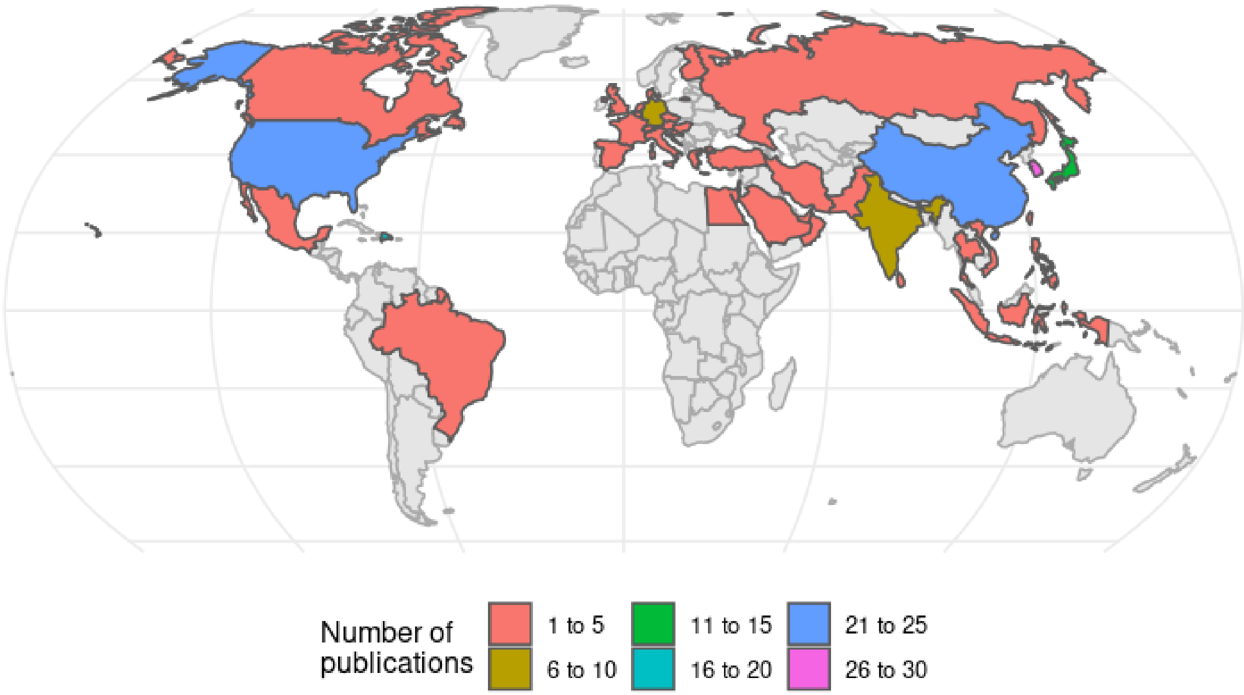

Figure 6: Geographical trend in number of publications included in the scoping review on machine learning in dentistry between 1st January 2015 and 31st May 2021. Source: modified from Figure S1, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937. Image rights held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under license CC BY 4.0 as per publisher MDPI policy.

85% studies split their datasets into training and testing subsets, while 59% studies created validation subsets too [18]. The median size of training datasets was 450 (range: 12 - 1,296,000) and of test datasets was 126 (range: 1 - 144,000) [18]. Half of the studies evaluated ML model performance on a hold-out dataset while the other half used cross-validation [18].

65% studies artificially increased their input data by using methods like image augmentation [18]. Only 20% studies externally validated their model's performance [18]. 73% studies used experts to establish the reference test (i.e., how the ground truth was defined): one expert in 18% studies, two to three experts in 11% studies each, four to five experts in 2% studies each, six to eight experts in 1% studies each, 12 and 20 experts in 0.5% studies each, and no information on number of experts in 27% studies [18]. 22% studies established the reference test from their datasets (e.g., age and diagnosis from medical records) and 1% studies used software- generated reference test [18]. The remaining 4% studies did not report on how the reference test was generated [18].

Of all studies, 70% used complex ML models, such as convolutional neural networks; further details are available in the publication [18]. Another 22% studies used simple ML models, such as random forest classifier and support vector machine [18]. In addition, 6% studies used various model combinations and 2% studies did not report information on the model structure [18]. Both, the complex and simple models were used more frequently by studies in restorative dentistry and endodontics, oral medicine, and non-specific field or general dentistry [18]; Table 3. Additionally, the simple models were often used by studies in orthodontics and periodontology [18]. Finally, 20% studies compared their model's performance to human experts [18].

Table 3: Number of studies included in the scoping review on machine learning in each field of dentistry, stratified by the types of machine learning models used (n=168).

Field of dentistry, n (%)	Models not using deep learning				Models using deep learning			
	Classifier model	Support Vector Machine	Neural networks without deep learning	Other models without deep learning	Non-convolutional neural networks	Convolutional neural networks	Combination models	Inadequate model details
n	10	4	7	16	7	111	10	3
Restorative dentistry and endodontics	2 (20%)	2 (50%)	2 (29%)	1 (6%)	2 (29%)	14 (13%)	1 (10%)	1 (33%)
Oral medicine	2 (20%)	0 (0%)	0 (0%)	5 (31%)	2 (29%)	14 (13%)	2 (20%)	0 (0%)
Oral radiology	0 (0%)	0 (0%)	0 (0%)	2 (13%)	0 (0%)	8 (7%)	1 (10%)	0 (0%)
Orthodontics	1 (10%)	0 (0%)	3 (43%)	3 (19%)	0 (0%)	10 (9%)	0 (0%)	1 (33%)
Oral surgery and implantology	1 (10%)	0 (0%)	0 (0%)	1 (6%)	1 (14%)	14 (13%)	0 (0%)	0 (0%)
Periodontology	0 (0%)	2 (50%)	1 (14%)	1 (6%)	0 (0%)	13 (12%)	1 (10%)	1 (33%)
Prosthodontics	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Others (non-specific field, general dentistry)	4 (40%)	0 (0%)	1 (14%)	3 (19%)	2 (29%)	36 (32%)	5 (50%)	0 (0%)

Source: modified from Table S3, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937. Data rights held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under license CC BY 4.0 as per publisher MDPI policy.

3.1.2 Risk of bias in the individual studies

Risk of bias was assessed in four domains, namely data selection, index test, reference standard, and flow and timing. It was found to be high for most studies with respect to data selection and reference standard [18]; Table 4. Concerns about the applicability of a study's methods and results were found to be high for most studies with respect to data selection [18].

Table 4. Evaluation of risk of bias in studies included in the scoping review (n=168) on machine learning in dentistry using the QUADAS-2 tool.

No. [Citation]	Data selection: risk of bias/ applicability concerns	Index test: risk of bias/ applicability concerns	Reference standard: risk of bias/ applicability concerns	Flow and timing: risk of bias
1 [39]	high/high	low/high	high/high	low
2 [40]	low/low	low/low	low/low	low
3 [41]	high/low	low/low	low/low	low
4 [42]	low/low	low/high	high/high	low
5 [43]	low/low	low/low	low/low	low
6 [44]	high/high	low/high	high/high	low
7 [45]	high/high	low/low	high/low	low
8 [46]	low/low	low/high	low/low	low
9 [47]	low/low	low/low	low/high	low
10 [48]	high/high	low/low	high/low	low
11 [49]	high/high	low/low	high/high	low
12 [50]	high/low	high/low	high/low	low
13 [51]	low/high	low/low	high/high	low
14 [52]	high/high	high/low	low/low	low
15 [53]	low/low	high/low	low/low	low
16 [54]	high/low	low/low	high/low	low
17 [55]	high/high	low/low	high/low	low
18 [56]	high/low	low/low	high/low	low
19 [57]	high/high	low/low	high/low	low
20 [58]	high/high	low/high	high/low	low

21 [59]	high/high	high/high	high/high	low
22 [60]	low/low	low/low	low/low	low
23 [61]	low/high	low/low	low/high	low
24 [62]	high/high	low/low	low/low	low
25 [63]	low/low	low/low	low/low	low
26 [31]	high/high	low/low	high/low	low
27 [64]	high/high	low/low	high/low	low
28 [65]	high/high	low/low	high/low	low
29 [66]	high/high	high/low	high/low	low
30 [67]	low/low	low/low	low/low	low
31 [68]	high/low	high/low	low/low	low
32 [69]	low/high	low/high	low/high	low
33 [70]	low/low	high/low	high/low	low
34 [71]	high/high	low/high	low/high	low
35 [72]	high/low	low/low	low/low	low
36 [73]	high/high	low/low	low/low	low
37 [74]	high/high	low/high	low/high	low
38 [75]	low/low	low/low	high/low	low
39 [76]	low/high	low/low	high/low	high
40 [77]	low/high	low/low	low/low	low
41 [78]	high/low	low/high	high/low	low
42 [79]	high/high	low/low	low/low	high
43 [80]	low/low	low/high	low/high	low
44 [81]	high/high	low/low	high/low	low
45 [82]	high/high	low/high	high/low	low
46 [83]	high/high	low/low	high/high	low
47 [84]	high/high	high/high	high/high	low
48 [85]	low/high	low/low	high/high	low
49 [86]	low/high	low/low	high/high	low
50 [87]	low/low	low/low	high/low	high
51 [88]	high/high	low/low	high/low	low
52 [89]	low/high	low/high	high/high	low
53 [90]	high/high	high/high	high/high	low
54 [91]	high/high	low/low	high/low	low
55 [92]	low/high	low/low	high/low	low

56 [93]	high/high	low/high	low/high	low
57 [94]	low/high	low/low	low/low	high
58 [32]	low/low	low/low	low/low	low
59 [95]	high/high	low/low	low/low	low
60 [96]	low/low	low/low	low/low	low
61 [97]	low/high	low/low	high/low	low
62 [98]	low/low	low/low	high/low	low
63 [99]	low/low	low/low	low/low	low
64 [100]	low/low	low/low	low/low	low
65 [101]	low/low	low/low	low/low	low
66 [102]	high/high	high/low	high/low	low
67 [103]	high/low	high/low	high/low	low
68 [104]	high/low	high/low	high/low	low
69 [105]	high/low	high/low	low/low	low
70 [106]	low/low	low/low	low/low	low
71 [107]	low/low	low/low	high/low	low
72 [108]	low/low	low/low	low/low	low
73 [109]	high/low	low/low	high/low	low
74 [110]	low/low	low/low	low/low	high
75 [111]	high/high	low/low	low/low	low
76 [112]	high/high	high/low	low/low	low
77 [113]	low/low	low/low	low/low	low
78 [114]	high/high	high/high	high/high	low
79 [115]	high/high	high/high	high/high	low
80 [116]	high/high	low/low	high/low	low
81 [117]	low/low	low/low	high/low	low
82 [118]	low/high	low/low	high/high	low
83 [119]	low/low	low/low	low/low	high
84 [120]	high/high	high/low	high/high	low
85 [121]	low/high	high/low	high/high	low
86 [122]	high/high	low/high	low/high	low
87 [123]	high/high	low/low	low/low	low
88 [124]	low/high	low/high	high/high	low
89 [125]	low/high	low/high	high/high	low
90 [126]	low/high	low/low	low/high	low

91 [127]	high/high	low/low	high/low	low
92 [128]	low/low	low/low	low/low	low
93 [129]	high/low	low/high	high/high	low
94 [130]	low/low	low/low	high/low	low
95 [131]	high/high	low/high	high/low	low
96 [132]	low/high	low/low	low/high	low
97 [133]	low/low	low/low	high/low	low
98 [25]	low/low	low/low	low/low	low
99 [134]	high/high	high/low	high/low	low
100 [135]	low/low	low/low	high/low	low
101 [136]	low/low	low/low	high/low	low
102 [137]	high/low	high/low	high/high	low
103 [138]	high/high	low/low	low/high	high
104 [139]	low/high	low/low	low/low	low
105 [140]	low/low	low/low	low/low	low
106 [141]	low/low	low/low	low/low	low
107 [142]	high/high	high/low	high/low	low
108 [143]	high/low	high/low	low/low	low
109 [144]	high/high	low/low	high/low	low
110 [145]	low/low	low/low	low/low	low
111 [146]	low/low	low/high	high/high	low
112 [147]	low/low	low/low	high/low	low
113 [148]	high/high	high/low	low/low	low
114 [149]	low/low	low/low	low/low	low
115 [150]	low/high	low/low	high/low	low
116 [151]	high/high	low/low	low/high	low
117 [152]	high/low	high/low	high/low	high
118 [153]	high/high	low/low	high/low	low
119 [154]	low/low	high/low	low/low	low
120 [155]	high/low	low/low	low/low	low
121 [156]	high/high	low/low	high/low	high
122 [157]	high/low	high/low	high/low	low
123 [158]	high/high	low/low	high/low	low
124 [159]	high/high	low/low	high/low	high
125 [160]	low/high	high/low	high/low	low

126 [161]	high/low	high/low	low/low	low
127 [162]	high/low	high/low	high/low	low
128 [163]	high/low	high/high	low/low	low
129 [164]	low/high	low/high	high/low	low
130 [165]	high/low	low/low	low/low	low
131 [166]	high/low	low/high	high/high	low
132 [167]	low/low	low/low	high/low	low
133 [168]	high/high	low/low	high/low	low
134 [169]	high/high	low/low	low/low	low
135 [170]	high/high	low/low	high/low	low
136 [171]	high/high	high/low	high/low	low
137 [172]	high/high	low/low	high/low	low
138 [173]	high/low	high/low	low/low	low
139 [174]	high/high	low/low	high/low	low
140 [175]	high/high	low/high	high/high	low
141 [176]	high/high	low/high	high/high	low
142 [177]	low/high	low/low	high/high	low
143 [178]	high/high	low/high	high/high	low
144 [179]	high/low	low/low	high/low	low
145 [180]	low/low	low/high	high/high	low
146 [181]	low/low	high/low	low/low	low
147 [182]	low/high	low/low	low/low	low
148 [183]	high/high	high/low	high/low	high
149 [184]	low/low	low/low	low/low	low
150 [185]	low/high	low/high	low/high	low
151 [186]	high/low	low/high	low/low	low
152 [187]	low/low	high/low	high/low	high
153 [188]	high/low	low/low	low/high	low
154 [189]	low/low	low/high	low/high	low
155 [190]	high/low	low/low	high/low	low
156 [191]	low/low	low/low	high/low	low
157 [192]	low/high	low/high	high/high	high
158 [193]	low/low	low/low	low/low	low
159 [194]	low/low	low/low	high/low	low
160 [195]	low/high	high/low	high/high	low

161 [196]	high/high	low/low	low/low	low
162 [197]	low/low	low/low	high/low	low
163 [198]	low/low	low/low	low/low	low
164 [199]	low/low	low/low	high/low	low
165 [200]	low/low	low/low	low/low	low
166 [201]	high/high	high/high	high/high	low
167 [13]	high/high	high/low	low/low	low
168 [202]	high/high	low/low	high/low	low

Source: Table 1, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937.

3.1.3 Reporting quality of the individual studies

Overall adherence to the TRIPOD checklist was 33.3%, with less than 50% studies adhering to 18 out of 22 domains [18]; Figure 7. Adherence was below 10% for sample size calculation, handling of missing data, differences between development and validation data, and details on study participants [18]. In particular, less than 20% studies adequately defined their predictors and outcomes, stratification by risk groups, presented the full prediction model and provided information on supplementary resources, such as study protocol, web calculator, or data sets [18]. Less than 40% studies adequately reported their data sources, participant eligibility, statistical methods (specifically, details on model refinement), model results, study limitations, and model performance in development data, and any other validation data [18].

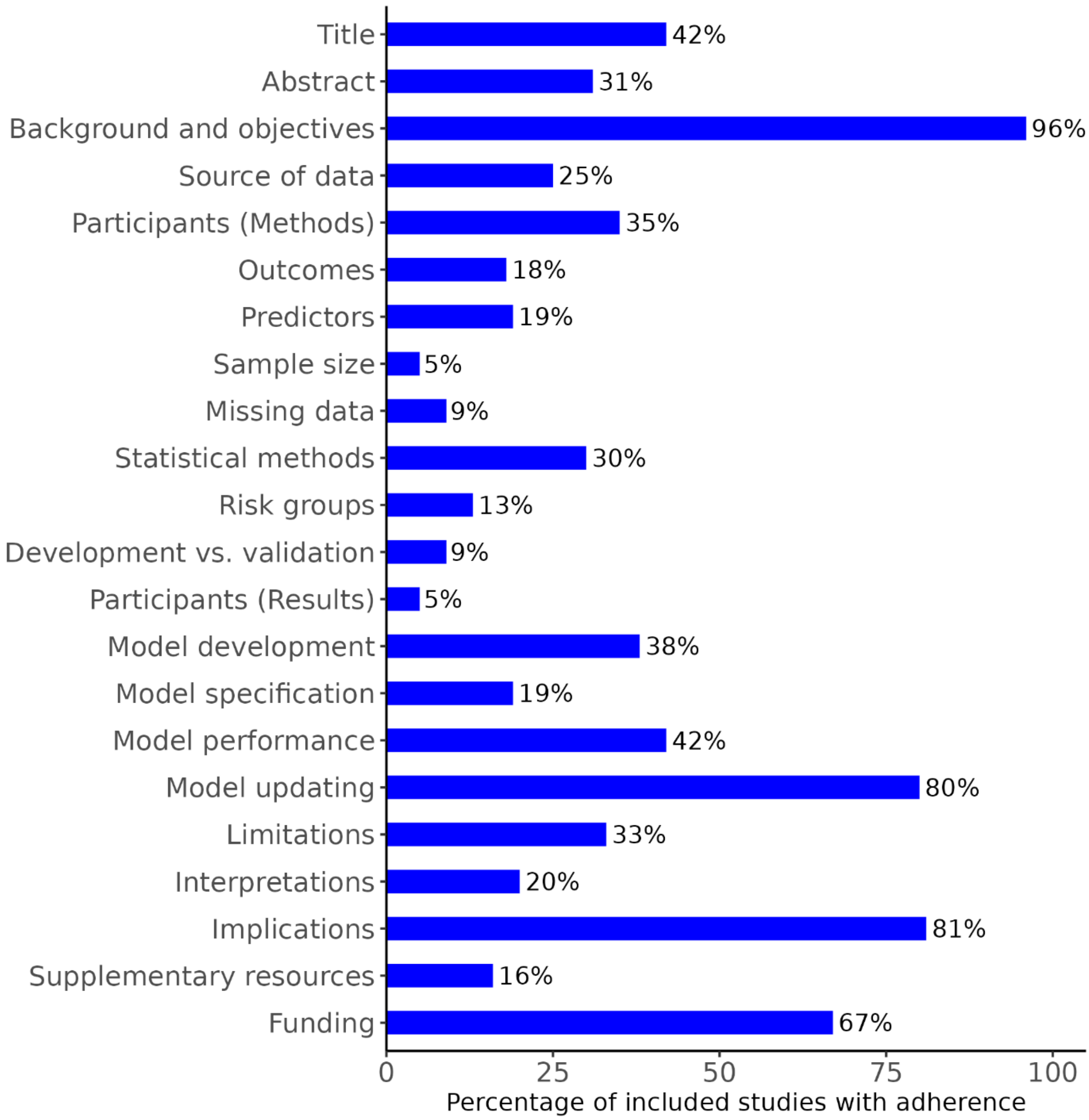

Figure 7: Percentage of studies included in the scoping review (n=168) on machine learning in dentistry that adhered to each of the 22 domains of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist. Less than half of the included studies showed high reporting quality. Source: modified from Figure 2, publication L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023), doi: 10.3390/jcm12030937. Image rights held by authors of the publication, including Lubaina T. Arsiwala-Scheppach, under license CC BY 4.0 as per publisher MDPI policy.

3.1.4 Model performance metrics

A total of 42 different metrics were used by the studies to evaluate model performance, while some of which could be grouped together, e.g., the different correlation coefficients could be combined; such grouping resulted in 26 distinct classes [18]. The most commonly reported metrics were accuracy, sensitivity, area under the receiver operating characteristic, specificity, precision, and intersection-over-union [18]. Figure 8 graphically shows the relative proportion of studies which used the different metrics, stratified by ML task. Also, the mean sensitivity values were generally high (≥0.80) in the studies performing classification tasks whose confusion matrices were either presented or reconstructed from the available data; Figure 9.

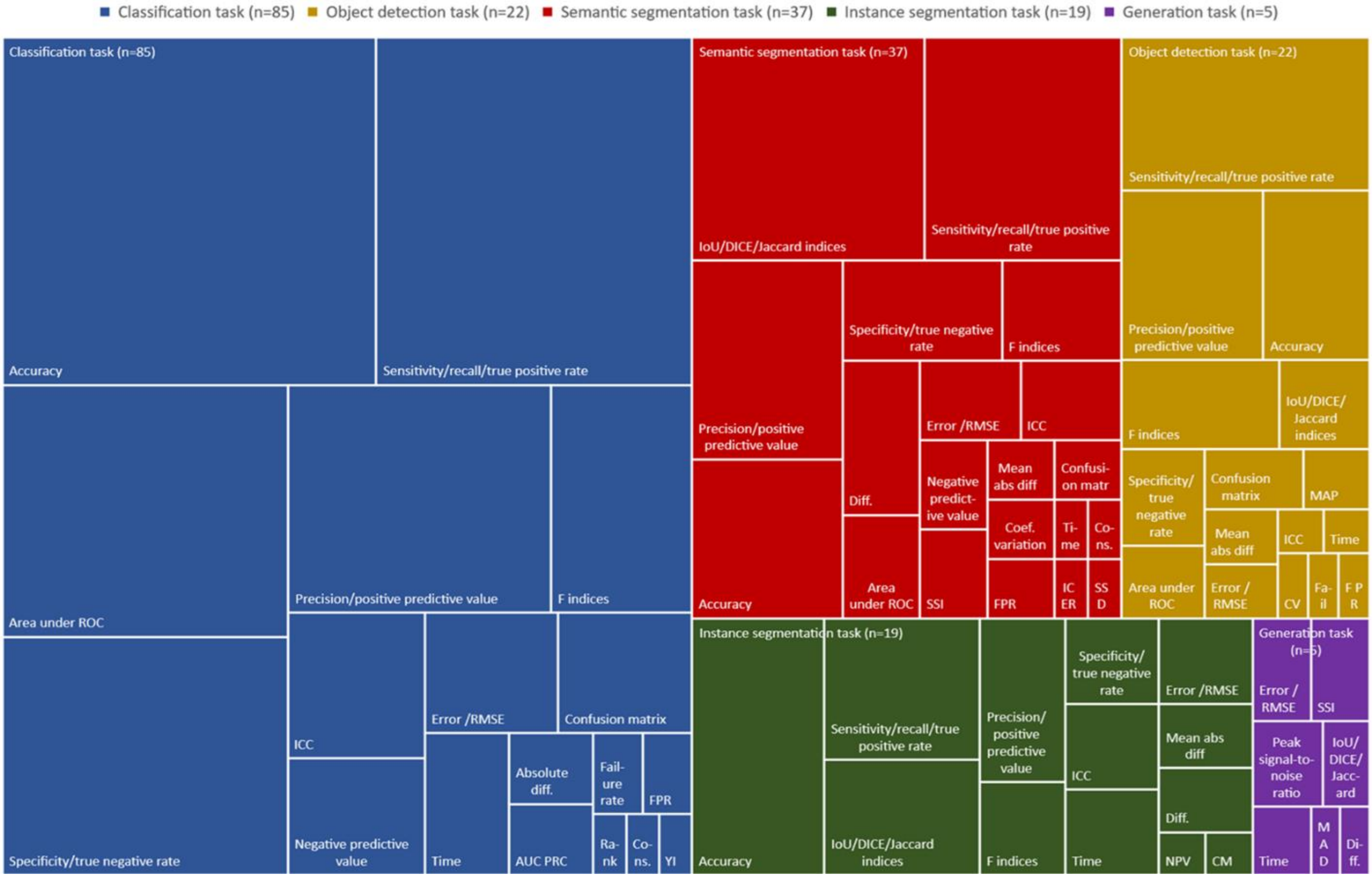

Figure 8: Treemap of performance metrics used in the studies included in the scoping review (n=168) on machine learning in dentistry stratified

by type of machine learning task. The size of each box is proportional to the number of studies included in the scoping review that used that particular metric for a machine learning task. Most studies reported multiple metrics. Abbreviations: Abs diff, mean or normalized absolute difference; AUC PRC, area under the precision-recall curve; CM, confusion matrix; Cons., model consistency (intra-CNN or inter-CNN consistency); CV, coefficient of variation; Diff., differences/distances be-tween volumes, surfaces, or points; FPR, false positive rate; ICC, correlation coefficients; ICER, incremental cost-effectiveness ratio; IoU, intersection over union; MAD, mean absolute difference; MAP, mean average precision; matr, matrix; NPV, negative predictive value; RMSE, root mean squared error; ROC, receiver operating characteristic; SSD, sum of squared difference; SSI, structural similarity index; YI, Youden's index. Key (for non-abbreviated terms): Fail, failure rate; Rank, Rank-N recognition rate; Time, time taken for analysis. Source: own representation.

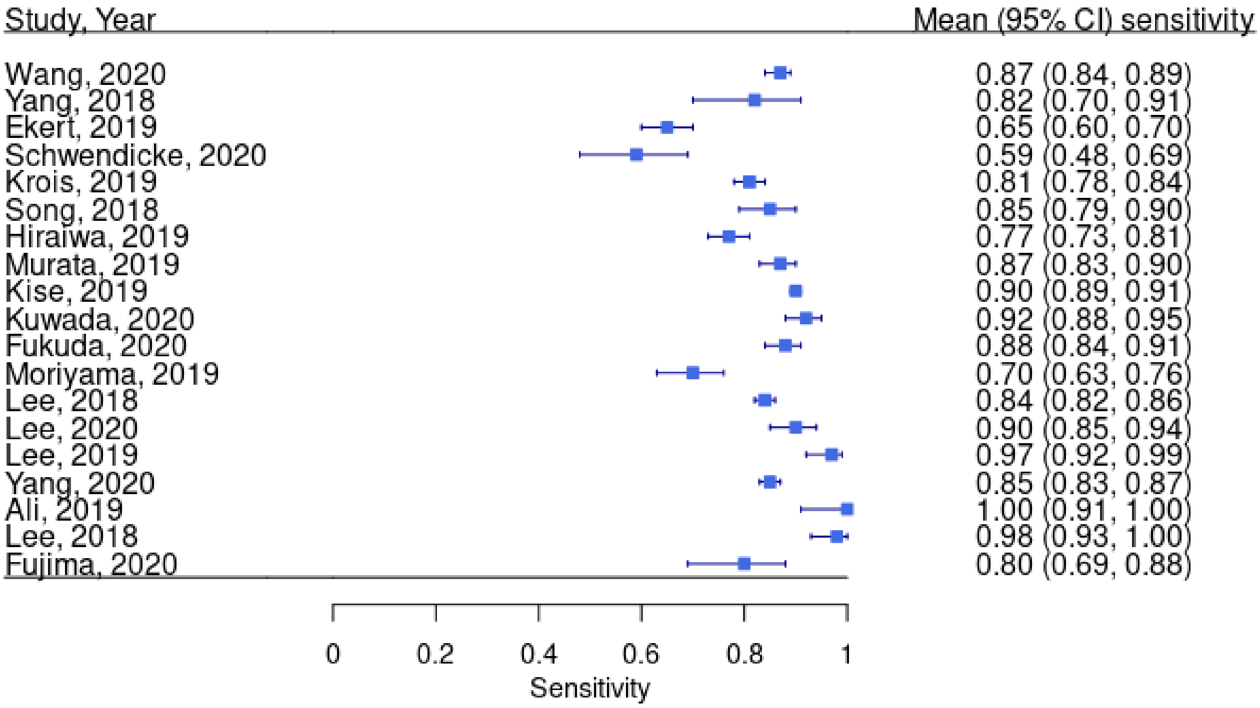

Figure 9: Forest plot displaying the mean (95% confidence interval) sensitivity of the studies (n=19) performing classification tasks whose confusion matrices were either reported or re-constructed from the available data, out of the 168 studies included in the scoping review on machine learning in dentistry. Source: own representation.

3.2 Benchmarking study

The performances of the various model configurations are depicted in Figure 10. ML architectures of U-Net++, U-Net, and LinkNet achieved a high F1-score of 0.86 (median; value has been rounded up) and outperformed their counterparts [36]. Models using the backbones of VGG group achieved a high F1-score of 0.85 (median) and outperformed the ResNet group [36]. Models initialized with ImageNet or CheXpert datasets outperformed models initialized with random weights (p< 0.001) [36]. Examination of all the 216 model combinations revealed that the highest performances was achieved by models consisting of U-Net++ or LinkNet architectures and ResNet or DenseNet backbones [36].

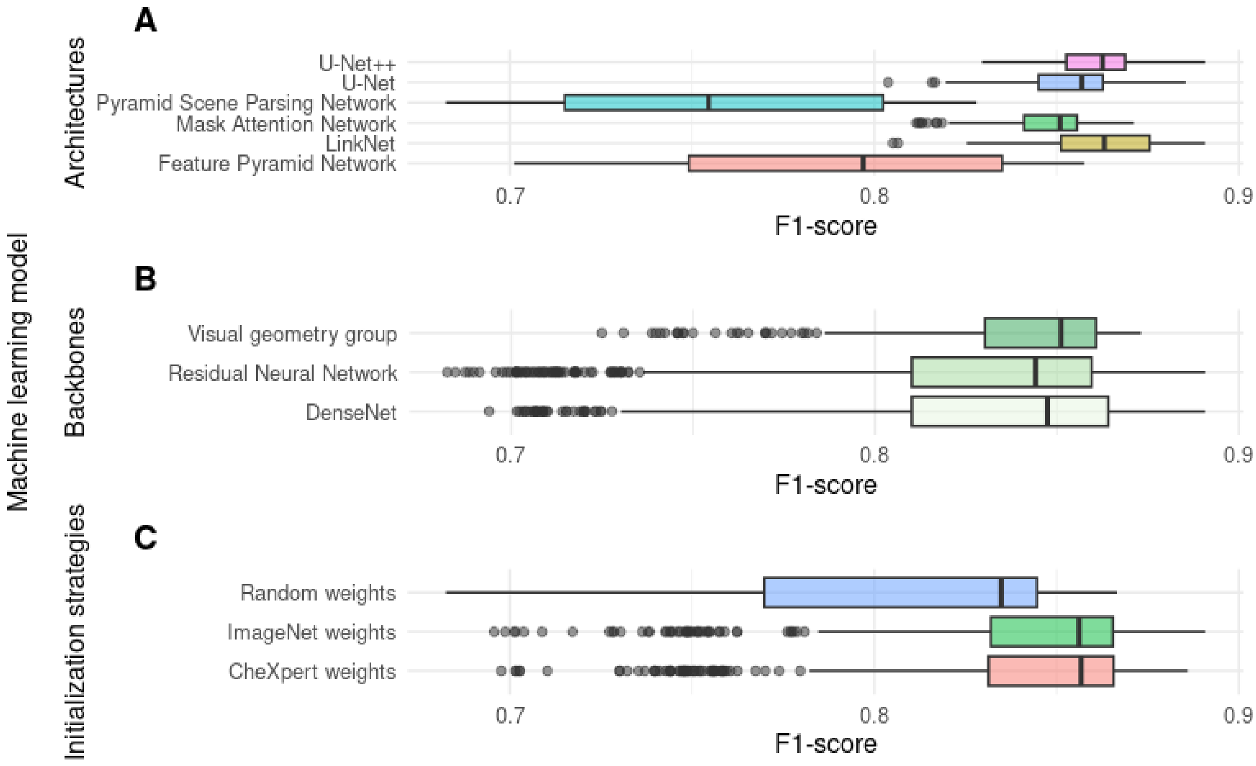

Figure 10: Boxplots showing the distributions of F1-scores of the different machine learning model configurations evaluated in the benchmarking study stratified by model architectures (A), backbone groups (B), and initialization strategies (C). The highest median F1-scores were attained by the architectures of U-Net++, U-Net, and LinkNet, backbone of visual geometry group, and models initialized with ImageNet or CheXpert datasets. Source: own representation.

A weak positive correlation between a model's complexity level and its performance was observed with r = 0.32 (p<0.001) [36].

Class imbalance: As a sensitivity analysis, model performance was evaluated on the less prevalent classes of fillings (80%) and crowns (20%). In general, the models' performance was inversely proportional to class frequencies [36].

3.3 Clinical trial

Six female and 16 male dentists participated with an average age of 38 years (range: 27 to 60 years) [26].

The performance metrics are displayed in Table 5. The overall mean (95% confidence interval) AUROC of dentists was higher in the ML group [0.89 (0.87-0.90)] than in the without ML group [0.85 (0.83-0.86)], $p<0.05$ [26]. The AUROC of all dentists stratified by trial group demonstrated that sensitivity was higher in the ML group [0.81 (0.74-0.87)] than in the without ML group [0.72 (0.64-0.79)], $p<0.05$ [26]; Figures 11 and 12. Higher values of AUROC, F1-score, and sensitivity in the ML group were observed for enamel caries but not dentin lesions [26]. For most dentists, an increase in sensitivity when using the ML software came at no or very limited decrease in specificity when compared to their own performance in the without ML group [26]; Figure 13. On comparing the dentists in either group of the trial with the ML software by itself, the latter had higher AUROC, accuracy, F1-score, specificity, and positive predictive value; Figures 11 and 12.

The inter-rater agreement between the dentists for detecting caries lesions when not using the ML software showed an expected trend. When stratified by depth of caries lesion, the inter-rater agreement was as follows; absence of caries lesion: 0.18, enamel caries: 0.03, early dentin caries: 0.14, advanced dentin caries: 0.47.

Table 5: Performance of dentists with and without support of a machine learning software and by the machine learning software by itself in the randomized clinical trial for detection of proximal caries lesions on bitewing radiographs. Mean and 95% confidence interval values shown. Comparisons between dentists with and without support of the machine learning software using the t-test where p<0.05 are indicated in bold.

Clinical trial group	Depth of caries lesion	AUROC	Accuracy	F1- score	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Dentists without machine learning	Overall	**0.85 (0.83, 0.86)**	0.93 (0.92, 0.95)	**0.76 (0.73, 0.78)**	**0.72 (0.64, 0.79)**	0.97 (0.96, 0.98)	0.80 (0.72, 0.86)	0.95 (0.94, 0.97)
	Enamel caries	**0.81 (0.78, 0.83)**	0.94 (0.92, 0.95)	**0.64 (0.60, 0.68)**	**0.64 (0.53, 0.74)**	0.97 (0.96, 0.98)	0.67 (0.56, 0.77)	0.97 (0.95, 0.98)
	Early dentin caries	0.89 (0.86, 0.91)	0.96 (0.95, 0.97)	0.65 (0.60, 0.71)	0.81 (0.66, 0.91)	0.97 (0.96, 0.98)	0.55 (0.42, 0.68)	0.99 (0.98, 1.00)
	Advanced dentin caries	0.92 (0.89, 0.96)	0.97 (0.95, 0.98)	0.58 (0.46, 0.71)	0.87 (0.66, 0.97)	0.97 (0.96, 0.98)	0.42 (0.28, 0.57)	1.00 (0.99, 1.00)
Dentists with machine learning	Overall	**0.89 (0.87, 0.90)**	0.94 (0.93, 0.96)	**0.81 (0.78, 0.84)**	**0.81 (0.74, 0.87)**	0.97 (0.95, 0.98)	0.82 (0.75, 0.88)	0.97 (0.95, 0.98)
	Enamel caries	**0.86 (0.84, 0.88)**	0.95 (0.93, 0.96)	**0.73 (0.68, 0.77)**	**0.75 (0.65, 0.83)**	0.97 (0.95, 0.98)	0.71 (0.61, 0.80)	0.97 (0.96, 0.98)
	Early dentin caries	0.92 (0.90, 0.94)	0.96 (0.95, 0.97)	0.70 (0.63, 0.77)	0.86 (0.73, 0.95)	0.97 (0.95, 0.98)	0.57 (0.44, 0.69)	0.99 (0.99, 1.00)
	Advanced dentin caries	0.95 (0.92, 0.97)	0.97 (0.95, 0.98)	0.59 (0.51, 0.67)	0.91 (0.72, 0.99)	0.97 (0.95, 0.98)	0.42 (0.28, 0.57)	1.00 (0.99, 1.00)
Artificial intelligence	Overall	0.91 (0.89, 0.93)	0.97 (0.96, 0.97)	0.88 (0.86, 0.89)	0.83 (0.79, 0.86)	0.99 (0.98, 0.99)	0.94 (0.91, 0.96)	0.97 (0.96, 0.98)

Abbreviation: AUROC, area under the receiver operating characteristic.

Source: modified from Table 1, publication S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849. Data rights for reuse in dissertation held by authors of the publication, including Lubaina T. Arsiwala, as per publisher Elsevier policy.

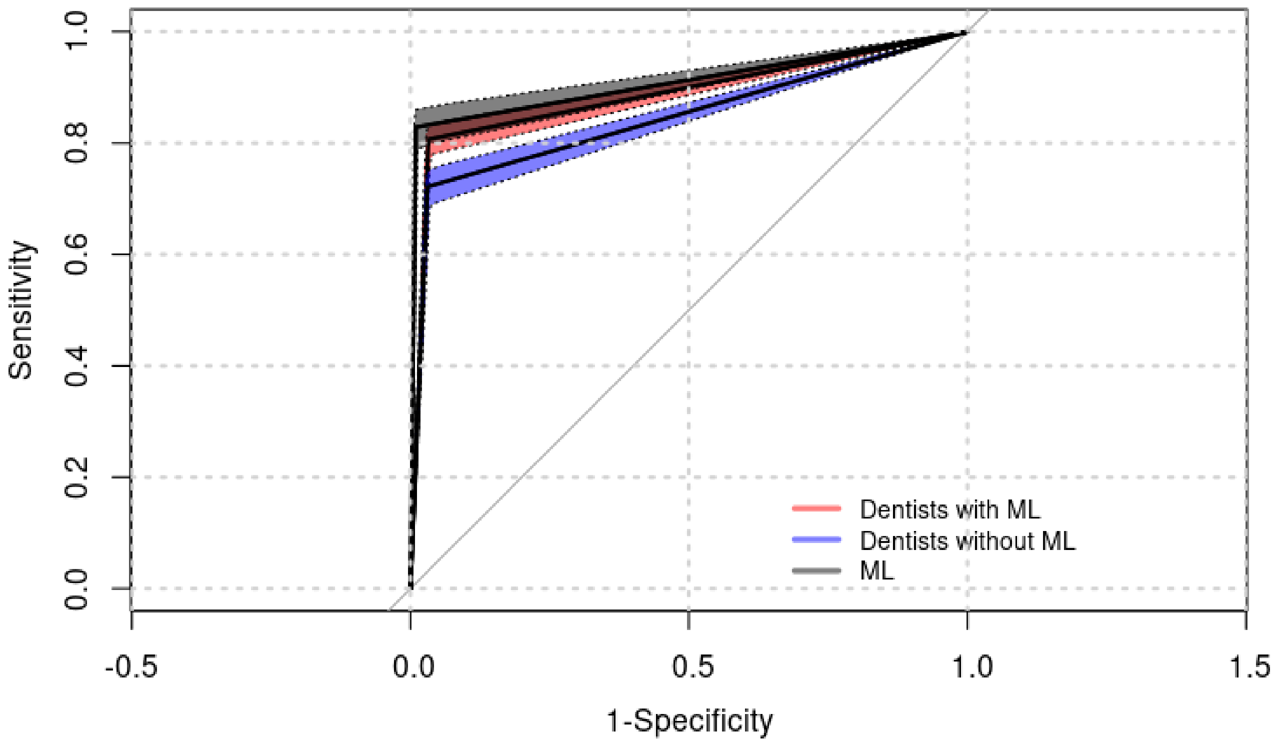

Figure 11: Receiver operating characteristic of the dentists with machine learning software (red), dentists without machine learning software (blue) and the machine learning software by itself (grey) as evaluated in the randomized clinical trial for detecting proximal caries lesions on bitewing radiographs. Mean (solid black lines) and 95% confidence intervals (coloured areas within the dotted lines) of the curves are shown. Abbreviation: ML, machine learning. Source: modified from Figure 2 (a), publication S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849. Data rights for reuse in dissertation held by the authors of publication, including Lubaina T. Arsiwala, as per publisher Elsevier policy.

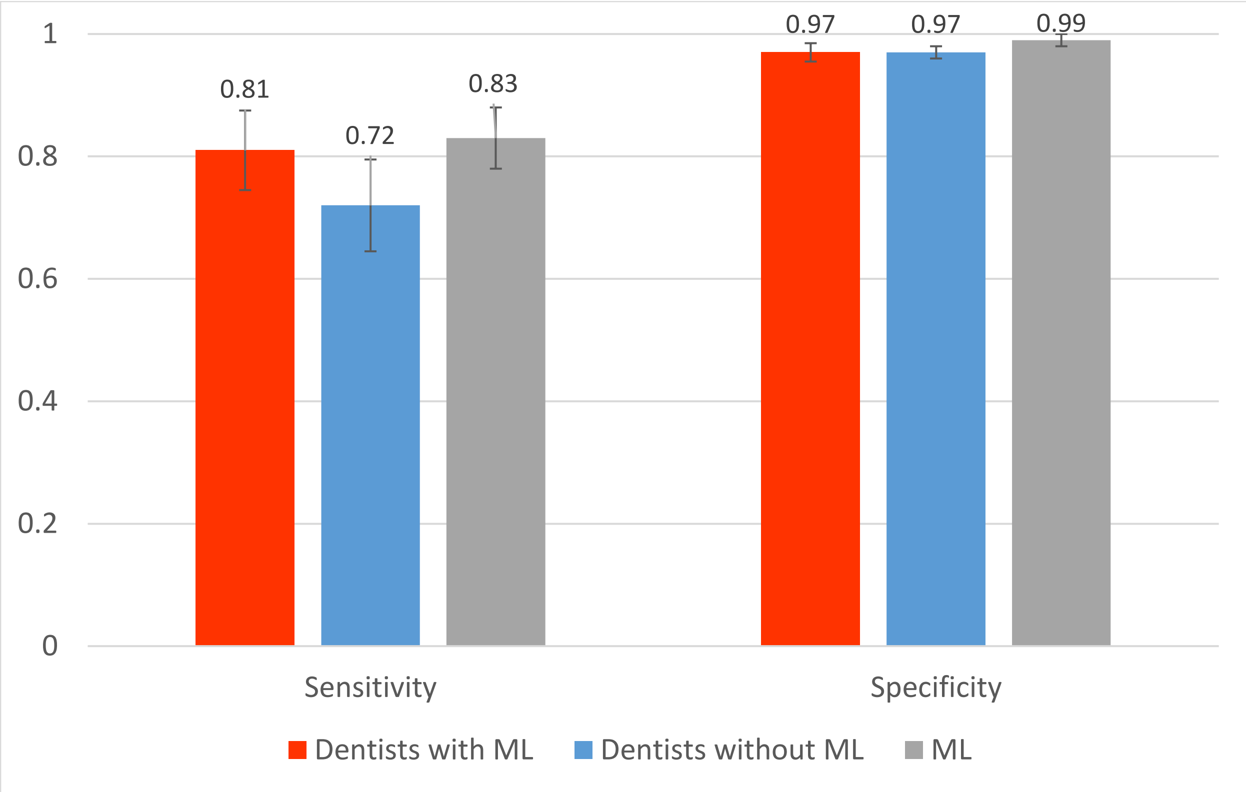

Figure 12: Comparisons of the sensitivity and specificity in detecting proximal caries lesions on bitewing radiographs in the randomized clinical trial between dentists with machine learning (red), dentists without machine learning (blue), and the machine learning software by itself (grey). Mean (numbers atop the bars) and 95% confidence intervals (black whisker lines on the bars) of the estimates are shown. Abbreviation: ML, machine learning. Source: own representation.

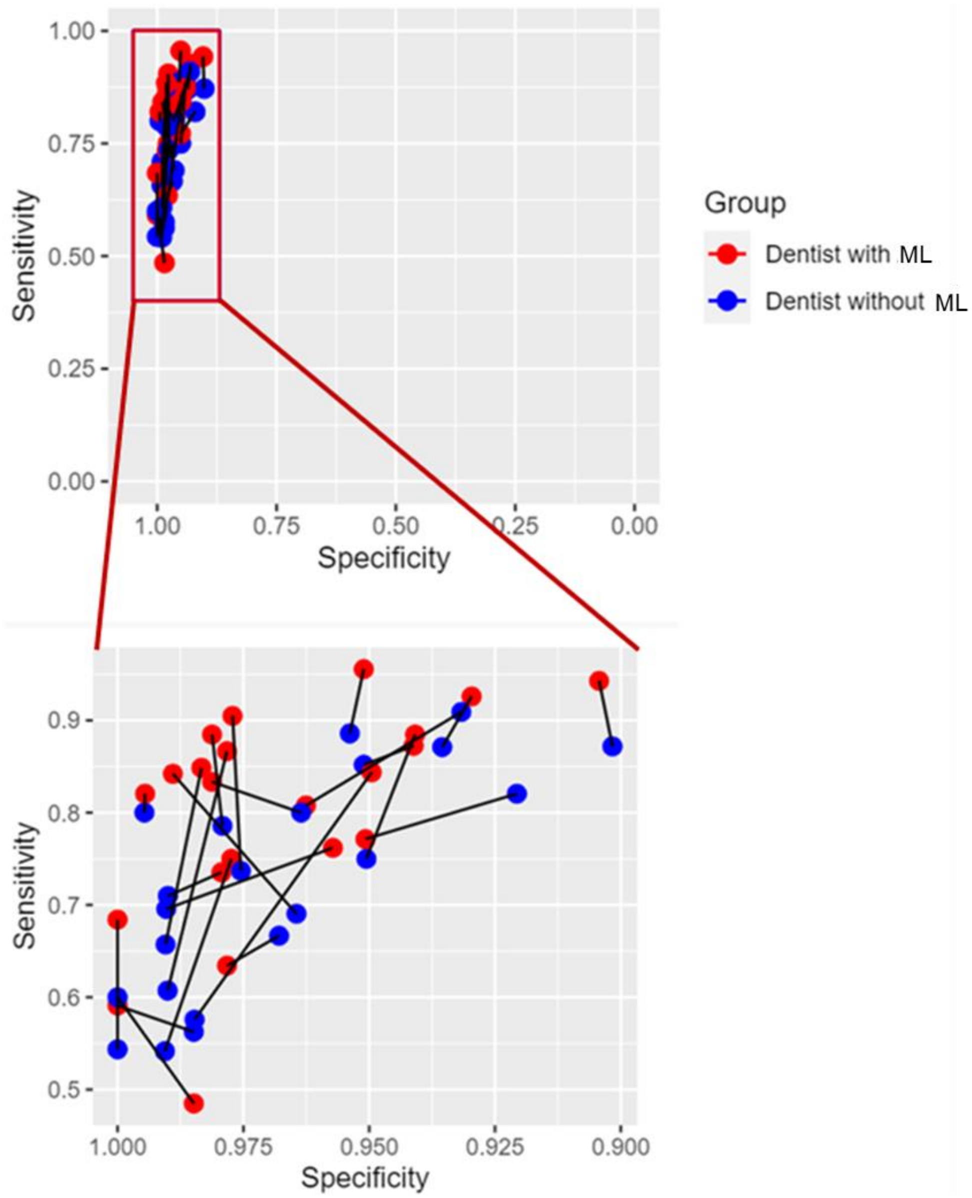

Figure 13: Differences in sensitivity and specificity of each dentist (points) across the two groups of the randomized clinical trial, i.e., with machine learning software (red) and without it (blue) for the detection of proximal caries lesions on bitewing radiographs. The top image shows the entire X- and Y-axes. The bottom image has zoomed-in where the data points are concentrated (outlined by a red box) hence note the extent of the X- and Y-axes. Each pair of data points belonging to an individual dentist is connected by a black line to highlight the dentist-wise change in sensitivity and specificity between the two groups of the clinical trial. For most dentists, an increase in sensitivity when aided by the machine learning software came at no or very limited decrease in specificity when compared to their own performance without any such support. Abbreviation: ML, machine learning. Source: modified from Figure 2 (b), publication S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849. Data rights for reuse in dissertation held by authors of the publication, including Lubaina T. Arsiwala, as per publisher Elsevier policy.

When comparing treatment decisions between the trial groups, the use of ML was found to increase the likelihood of the dentists' decision to non-invasively treat enamel caries (increase of 4%; $p<0.05$) as well as the decision to treat them invasively (increase of 7%; $p<0.05$) [26]. A similar shift was observed for early dentin caries, where the likelihood of invasive treatments increased by 11%; $p<0.05$ [26].

4. Discussion

4.1 Short summary of results

The research literature on ML in dentistry contains a large variety of clinical applications which demand a wide range of input data types, ML methodology, and performance metrics [7, 18]. The number of studies in the field is growing exponentially, however, many of them are hampered by considerable risk of bias and poor reporting quality [18]. This heterogeneity and paucity of robust evidence implies that despite an abundance of scientific evidence, we are faced with limited comparability across the studies [18].

To characterize the emerging patterns in the included studies, we first needed to examine the nature of clinical tasks which were tackled using ML. A plethora of research aims was present; from detecting artifacts in images to examining the usefulness of transfer learning, from categorizing different dental conditions to supporting decision-making and assessing cost-effectiveness of healthcare systems [18]. Classification tasks were the most common (51%) and can be used for diagnosing dental anomalies on images which is vital for early detection and successful treatment [18]. However, over the years, ML methods have improved their image classification performance at the cost of increased model complexity and opacity [217]. The inability to explain ML's methods and decisions has boosted the development of the field of explainable AI [18]; discussed in detail further ahead. Second, in the field of restorative dentistry and endodontics, the trend is starting to move away from traditional tasks, e.g., caries detection and classification of teeth in photographs or radiographs, to more complex ones. For instance, recent studies have investigated ML for diagnosing more subtle features like tooth cracks, performing image segmentation to detect early-stage caries, localizing lesions in near-infrared transillumination, characterizing root morphology, volumetric analysis, formulating treatment plans, and even assessing the cost-effectiveness of healthcare systems [8, 52, 65, 74, 85, 86, 98, 177].

An important reason behind the poor comparability across research studies is the high number of different ML model configurations in use. The abundance of model options combined with a scarcity of initiatives to benchmark them makes it challenging for researchers to select appropriate models [218]. The benchmarking study aimed to address this issue by conducting a systematic comparison of different model configurations for the specific task of outlining parts of a tooth on bitewing radiographs. ML combinations that attained the highest performance for this task consisted of U-Net++ or LinkNet architectures and ResNet or DenseNet backbones [36]. VGG backbones demonstrated consistency and stability across different model configurations [36]. Complex models performed slightly better, if at all, than simpler alternatives and were not highly efficient on imbalanced datasets [36]. The benchmarking study tested the hypothesis that model performance would be positively correlated to its complexity. While the results showed that this hypothesis was accepted, it must be highlighted that the large increases in model complexity, which came at the cost of larger computing demands, resulted in small improvements in performance [36]. It should be noted that lower computing demands allow for high resolution of input images which may be important for several dental applications [36]. Additionally, as hypothesized, the process of transfer learning improved model performance [36].

Another weakness of the existing literature on ML in dentistry is the lack of prospective clinical comparisons. The randomized clinical trial described here attempted to address this issue. It revealed that the ML software outperformed the dentists and when used by the dentists, can improve their sensitivity in identifying enamel carious lesions [26]. However, the hypothesis of the clinical trial could only be partially accepted because the ML software did not improve the specificity of the dentists or impact their diagnostic abilities for advanced lesions [26]. This could be attributed to the ML software's ability to diagnose caries by learning from multiple experts, which acted as an extra pair of eyes for the dentists and bolstered their sensitivity for incipient lesions [26]. These improvements in performances varied across the individual dentists [26]. On the other hand, the dentists did not require much assistance from the software to identify advanced lesions because they were more conspicuous on radiographs. These aspects of performance were also reflected in the inter-rater agreement between the dentists; the agreement was the lowest for enamel lesions and gradually increased with the depth of the caries lesions. Additionally, using ML increased the treatment severity for the detected lesions; significantly more enamel caries lesions were detected and then assigned non-invasive treatments or, for a notable proportion of the lesions, invasive treatments [26].

4.2 Interpretation of results

The three studies presented here are generally consistent with the findings of other studies. First, the heterogeneity in the studies included by other reviews prevented them from performing meta-analyses [11, 12, 19-22]. Second, most reviews reported that included individual studies had risk of bias and poor reporting quality [8, 11, 20, 22, 219-221]. Third, studies have noted that the superiority of ML models in one domain does not necessarily transfer to other domains [222]. Fourth, transfer learning has been shown to improve model performance [222]. Fifth, a clinical study has demonstrated that an ML software performed significantly better than dentists in detecting caries and suggested that its use may improve dentists' accuracy and sensitivity, especially for enamel caries lesions [25]. Finally, using ML increased the treatment intensity for a notable proportion of enamel lesions to invasive therapy which is in line with the fact that dentists continue to manage early lesions restoratively as demonstrated by a meta-analysis [223].

Nonetheless, some results were not consistent with other studies. First, most other reviews included far fewer studies than the scoping review described herein because they focussed on specific dental topics and thus had more restrictive inclusion criteria [8, 12, 19, 21, 22, 219-221, 224]. Second, the use of ML software in the clinical trial did not improve dentists' accuracy for advanced lesions, as suggested by a previous study [25]. The potential reason for this may be that advanced lesions show prominent radiographic changes which the dentists could identify even without ML support. Lastly, the dentists in the clinical trial exhibited higher sensitivity than that reported by a meta-analysis of over 100 studies [225]. This suggests that the dentists in the clinical trial were particularly accurate owing to possible selection bias or performance bias.

4.3 Embedding the results into the current state of research

The scoping review aimed to make ML studies in dentistry more robust and contribute to bridging the knowledge gap in the research field by identifying areas of fallacy in the current literature and suggesting methods to overcome them [18]. First, reporting of results that are generalizable is one of the cornerstones of high-quality research [2, 220, 226]. Hence, researchers should strive to generate data from multiple centres which may add diversity in terms of geographic location, racial, social, and economic status [220, 226].

Also, using a variety of types of data sources to create richer datasets could allow for cross-checking the data integrity and leveraging information from different sources [226]. Furthermore, the studies usually did not provide access to their data, except for those which used open databases, thus resulting in difficulties in replication of results [226]. Researchers are urged to follow the journals' data sharing guidelines in order to promote study replicability [18]. There may be concerns towards data sharing and privacy, for example, when anonymization of data is difficult [227]. Here, options like federated learning which eliminate the need to share data, as explained in Figure 14, should be encouraged [227]. Normally, in AI, data is collected from different local sources and sent to a central server for training an AI model. However, in federated learning, the data stays on the local devices or sources. Instead of sending the data to a central server, the AI model is sent to the local devices. Each device trains the model using its own data, and only their model is returned to the central server, which combines all the individual models into one updated model [227]. Thus, the personal data remains private and secure, which is useful for situations where privacy is important, like in healthcare or financial applications.

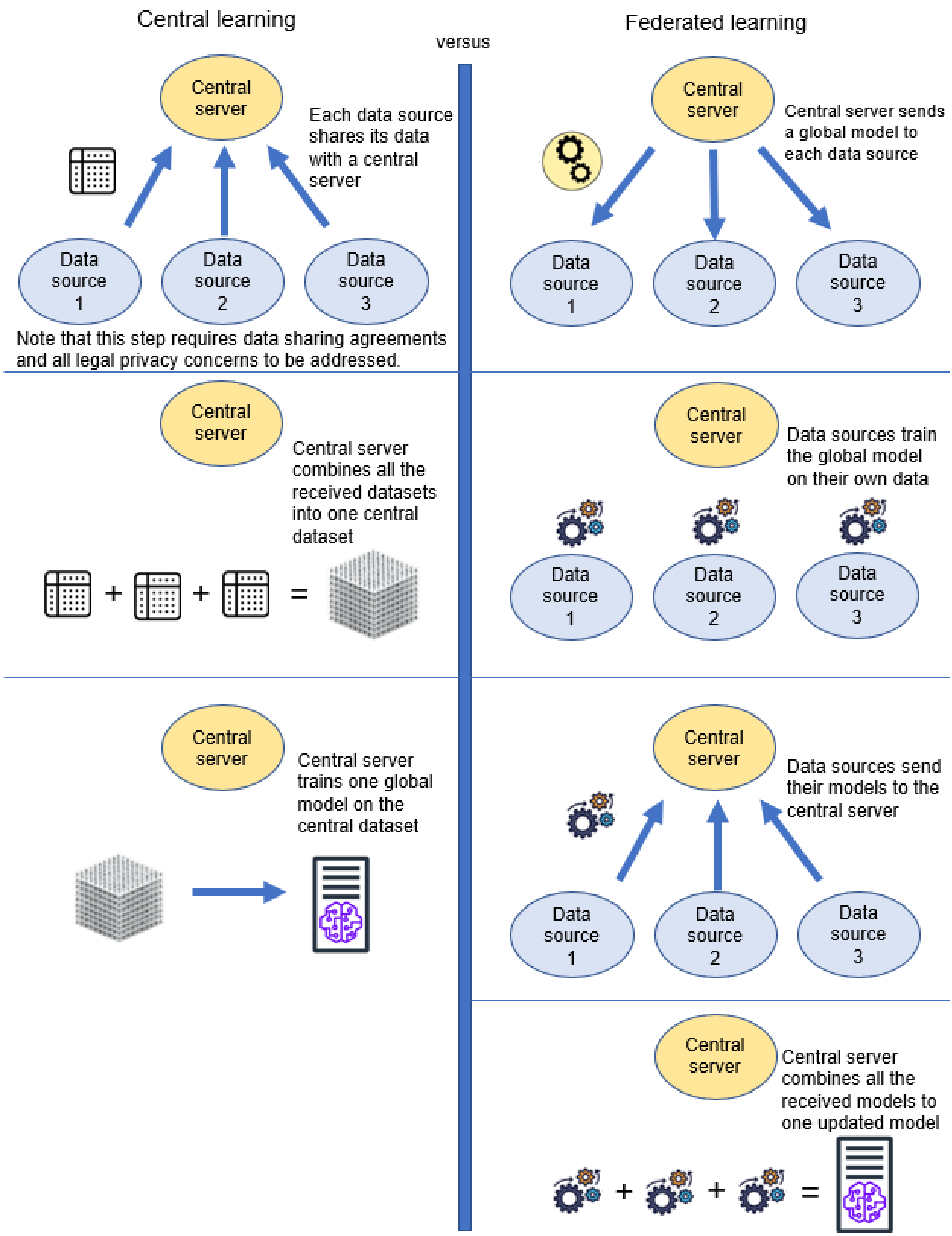

Figure 14: Representation of the concept of federated learning in the context of machine learning and how it differs from traditional central learning. In federated learning, the need to share data between institutions of universities is eliminated. Source: own representation.

Second, the high number of metrics used to measure model performance further exacerbated the limited comparability between studies [18]. It is crucial to define a standardized set of outcome metrics for specific dental subtasks in ML that encompasses diagnostic and clinical usefulness, prevalence of outcome, and various aspects of model performance. Also, studies examining the value of ML when used by dentists compared to the current standard of care are needed [2].

Third, the generation of reference tests (i.e., establishment of the ground truth) merits discussion. Overall, the studies included in the review used a variety of methods to establish reference tests but many did not provide further details [18]. It was concerning to note that a few studies had their reference test developed by only one expert, which is not ideal considering the variance in experts' annotations [8]. Additionally, datasets used to evaluate model performance should be standardized and heterogeneous to ensure balanced datasets and generalizability [18, 221, 228]. One approach is to establish benchmarking datasets that are publicly available, as attempted by the International Telecommunication Union (ITU) together with the World Health Organization (WHO) [18]. The ITU/WHO has set up a focus group to define the standards of AI applications in medicine and one of its subgroups is 'Dental Diagnostics and Digital Dentistry' [229].

Fourth, the quality of existing literature on ML in dentistry was poor to moderate [18]. The risk of bias arising from ML methods and data was insufficiently addressed, e.g., biases in data, leakage of data, or overfitting of the model. Furthermore, many studies failed to externally validate their models which is important as it speaks to the generalizability of the results [2, 221]. Generally, most studies tested applications, built models, and concluded that ML can learn and predict. However, general reporting without details hinders study replication [18]. Researchers are strongly advised to adhere to the published checklists on study conduct and reporting [35].

Current guidelines require rigorous and comprehensive planning, conducting, and reporting of ML studies in dentistry [35]. A crucial component of these guidelines is the hypothesis-driven selection of the ML model. Researchers must select a model architecture, backbone, complexity level, and initialization strategy specific for their study. However, the abundance of options for models combined with a lack of their comprehensive comparisons often result in researchers struggling to identify an ML model suitable for their specific requirements [36]. With many researchers defaulting to choosing the popularly known models, there is a lack of hypothesis-driven model selection. The benchmarking study aimed to address this knowledge gap by conducting a systematic comparison of various ML model configurations in order to provide guidance on ML design and thus contribute to evidence-based building of ML models in the field.

The clinical trial adds to the research field by providing empirical evidence of a prospective clinical comparison of an ML software. It highlighted the promising potential of combining dentists with a high-performing ML software in a real-world clinical setting to achieve diagnostic capabilities superior to the dentists alone [26]. However, the heightened sensitivity to enamel caries came with a higher proportion of them being assigned to invasive treatments [26]. These findings indicate the need to validate ML applications prospectively; ML for health should meet the criteria of evidence-based care and researchers in ML in dentistry should critically and comprehensively evaluate ML solutions [2].

4.4 Strengths and weaknesses of the studies

The scoping review presented herein has several noteworthy features. First, it is the most comprehensive overview of ML in dentistry comprising of 168 studies [18]. Second, the review is potentially generalizable to other studies as it covers the diversity in research questions, ML models, model performance metrics, and challenges related to risk of bias and poor reporting quality. Third, and as a limitation, no randomized controlled trials were included because none were available, which should be noted while interpreting the results [18]. Fourth, while the TRIPOD checklist was used to examine the reporting quality of the individual studies, it has not been specifically validated for ML applications [17]. Nevertheless, previous studies have employed it to assess ML models as it was originally designed for assessing clinical prediction tools, which are comparable to ML models [17]. Last, this review did not examine the clinical usability of the reviewed ML models as it was outside the scope of the study aim [18].

The benchmarking study has a few limitations. First, the specific task of outlining parts of a tooth on radiographs and evaluation of a certain set of ML models may restrict the generalizability of the results to similar outlining-related tasks or model structures [36]. Second, the use of data generated from different machines may have influenced the results [36]. Additionally, radiographs containing rare features, e.g., bridges, implants, and root canal fillings were excluded. However, these limitations do not undermine the study results since the primary aim was to benchmark models rather than build clinically useful or highly accurate ones [36]. Last, the study did not explore the possibility of more efficient restructuring of complex models that could reduce computing resources [36].

The clinical trial has several strengths and limitations. First, it is one of the few clinical randomized controlled trials in the field of dentistry which utilized an array of outcomes to carefully quantify the influence of ML [26]. Second, and as a limitation, the trial was not entirely conducted in a clinical setting but in a simulated clinical environment [26]. While this offered the advantage of controlling and standardizing the setup to a certain extent, it should be noted when interpreting the results. In a real clinical setup, there are other diagnostic options available and several factors affecting treatment decisions beyond the diagnoses (e.g., dentists' experience and armamentarium, patients' expectations, and costs) [26]. Third, the radiographs and participating dentists were selected from only two machines and one clinical center, respectively, and thus the results may have reduced generalizability [26]. Additionally, the participating dentists were younger than the average German dentist, primarily practicing in an urban environment, and exhibited higher accuracy than reported in other studies [26]. Fourth, the reference test was defined by human experts, a method that may have limited robustness; and additional validation using histology was not possible [26].

4.5 Implications for practice and/or future research

In the field of dentistry, ML studies should aim to reduce the risk of bias and improve adherence to reporting standards, thereby allowing for their replication and improving robustness, transparency, and generalizability of their findings [2, 7, 18, 221]. A minimum (core) set of outcomes and metrics for model performance should be established to facilitate comparisons across studies [18]. Future research should aim to showcase how ML can improve the quality and efficiency of patient care, as attempted by the clinical trial described herein [2, 7, 18]. Researchers may benefit from applying the concept of transfer learning when building ML models for dental radiographic analysis and considering less complex models as alternatives if computing resources and time needed to develop the models are constraints [36, 222, 230]. However, researchers must note that models developed on non-dental data sets may not perform similarly well on dental data sets [36, 222]. In clinical practice, ML has the potential to improve dentists' diagnostic performance, especially for detecting enamel caries lesions and should be pursued in future studies along with the implications of the nature of treatments thus assigned [26]. Furthermore, as the influence of ML may differ among dentists it warrants further investigation in order to advance towards personalized dental practice [26].

ML applications can enhance clinician-patient communication [2]. For example, ML software can generate an augmented version of the original image with the pathology being highlighted in color, as depicted in Figure 2. This can help patients to better understand their condition and thus the treatment plan [2]. However, it is essential that the ML software conveys its results in an easy-to-understand manner and not in technical or methodological jargon [2]. The most recent example of ChatGPT, an AI application that can generate meaningful language text, can be effectively used in this context [231].

Simulation is another AI application omnipresent in our everyday life. For example, autonomous driving relies heavily on simulations. As a real-world test drive on all the roads in the world, approximately 8.8 billion miles, is a significant challenge, simulation is a potential solution for this [2]. Simulation is not widely used in dentistry as of yet [2]. The dental pharmaceutical industry invests millions of dollars in the drug development process but often drug trials do not achieve the desired targets. Simulation of these processes where the experiments are run by a computer is a potential alternative to advance the drug trials [232]. The idea here is to capitalize on the ability of AI to analyze big data to identify previously unknown molecular characteristics and interactions and thus, predict the properties of the drug under trial [232].

All the implications of AI in dentistry discussed until now, namely faster, earlier, and more accurate disease diagnosis and thus less expensive treatment plans, efficient management of workflow in clinics, and better communication with patients can all come together and contribute to the bigger picture of better understanding an individual's healthcare needs [7]. This provides the basis for advancing personalized dentistry which is currently in its nascent stages [19]. Its imminent obstacle is the unavailability of data which is neither standardized nor linkable to other data sources [2]. Resolving such issues would go a long way in advancing personalized dentistry.

Furthermore, AI has been recently introduced in dental education and so far, its use has been limited to aid teaching of operative dentistry and craniofacial anatomy [233]. The integration of AI in dental education holds several implications for the field. First, AI technologies can enhance the learning experience by providing interactive and personalized educational resources [234]. Virtual simulations and augmented reality tools can allow students to practice dental procedures in a risk-free environment, improving their skills and confidence. One of my previous studies has demonstrated that augmented vision helped motivate dental students in learning to detect proximal caries lesions on bitewing radiographs [234]. Second, AI can facilitate students to access vast amounts of dental knowledge and research. Third, AI can assist in assessments and evaluations by automating tasks such as grading and feedback generation, saving time for educators and providing timely and objective assessments for all students. Overall, the adoption of AI in dental education has the potential to revolutionize teaching and learning methods, improve clinical competence, and promote continuous professional development in the dental field. Future studies should further explore how the promising potential of AI can be tapped for education.

Last, it is hard to interpret the process used by most AI systems to arrive at a final decision due to the inherently complex structure of the model. Thus, AI applications are usually regarded as ‘black boxes’ i.e., users cannot fully comprehend the criteria used by the AI to generate a certain result [7, 219]. This has boosted the field of explainable AI where attempts are made towards unravelling the underlying decision systems employed by AI models [18]. Advancements in explainable AI will certainly improve the transparency of the models and thus help clinicians to trust AI more. The clinical trial described in this dissertation also collected data on the eye movements of the dentists while they examined the bitewing radiographic images, as indicated in Figure 3. Figure 15 depicts an example of a dentist’s eye movements in this trial. Analysing this data could help to understand how the dentists extract relevant information from the images. These insights can then be transferred to AI models, enabling them to better replicate and augment human expertise. Consequently, this advancement may contribute to the development of improved AI-supported diagnostic tools and progress further towards explainable AI.

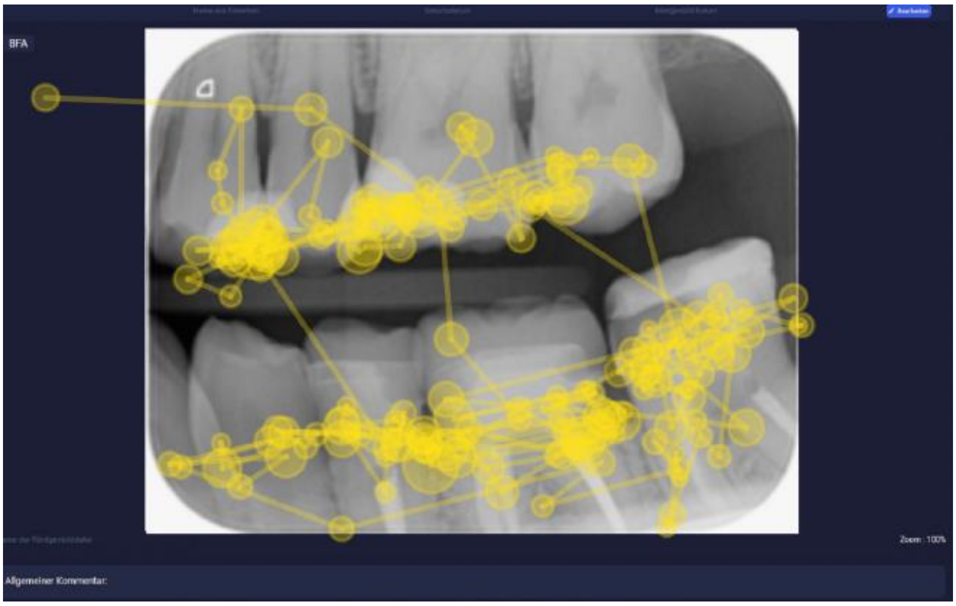

Figure 15: An exemplary gaze pattern of a participating dentist while detecting proximal caries lesions on bitewing radiographs in the randomized clinical trial. The observed gaze pattern is characteristic of the task assigned to the dentist and it shows that the dentist employed a systematic search strategy i.e., examining the proximal surfaces of the teeth in one jaw before moving on to the opposite jaw. Source: own representation.

5. Conclusions

ML techniques have been widely used in dentistry for a variety of tasks and have employed a diverse set of models and metrics to evaluate their performance [18]. The existing literature showed a considerable risk of bias as well as limited adherence to reporting guidelines [18]. While the focus of many studies was on developing ML models, their generalizability, robustness, or clinical usefulness was infrequently presented [18]. ML researchers are encouraged to adopt the practice of selecting models based on their hypothesis and optimizing their model structures with regards to transfer learning, model complexity, and computing resources [36]. Empirical evidence from a randomized controlled trial suggested that ML software can bolster dentists' performance in clinical diagnostic tasks and this advantage should be leveraged and explored further [26]. In my current research projects, I aim to expand on these results by investigating how the ML software influenced the visual search strategies used by the dentists which led to better diagnostic performance. Understanding how dentists extract information from radiographic images may serve in building improved ML-supported tools, improving transparency of the models, and thus fostering trust and acceptance of ML systems by clinicians.

Reference list

[1] O.E. Dictionary, artificial intelligence, n, Oxford English Dictionary, Oxford University Press, 2023, doi: https://doi.org/10.1093/OED/3194963277.
[2] F. Schwendicke, L.T. Arsiwala-Scheppach, J. Krois, Artificial intelligence: what it is and what it can do for dentists, Dental Update 50(4) (2023) 245-249, doi: 10.12968/denu.2023.50.4.245.
[3] X. Liu, L. Faes, A.U. Kale, S.K. Wagner, D.J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J.R. Ledsam, M.K. Schmid, K. Balaskas, E.J. Topol, L.M. Bachmann, P.A. Keane, A.K. Denniston, A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis, Lancet Digit Health 1(6) (2019) e271-e297, doi: 10.1016/S2589-7500(19)30123-2.
[4] J. Mohd, H. Abid, Ravi, S. Rajiv, R. Shanay, Significance of machine learning in healthcare: Features, pillars and applications, International Journal of Intelligent Networks 3 (2022) 58-73, doi: https://doi.org/10.1016/j.ijin.2022.05.002.
[5] O.T. Jones, R.N. Matin, M. van der Schaar, K. Prathivadi Bhayankaram, C.K.I. Ranmuthu, M.S. Islam, D. Behiyat, R. Boscott, N. Calanzani, J. Emery, H.C. Williams, F.M. Walter, Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review, Lancet Digit Health 4(6) (2022) e466-e476, doi: 10.1016/S2589-7500(22)00023-1.
[6] J. Zhang, Y. Liu, T. Mitsuhashi, T. Matsuo, Accuracy of Deep Learning Algorithms for the Diagnosis of Retinopathy of Prematurity by Fundus Images: A Systematic Review and Meta-Analysis, J Ophthalmol 2021 (2021) 8883946, doi: 10.1155/2021/8883946.
[7] M.-L. Sun, Y. Liu, G. Liu, D. Cui, A.A. Heidari, W.-Y. Jia, X. Ji, H. Chen, Y. Luo, Application of Machine Learning to Stomatology: A Comprehensive Review, IEEE Access 8 http://aliasgharheidari.com (2020) 184360-184374, doi: 10.1109/ACCESS.2020.3028600.
[8] F. Schwendicke, A. Chaurasia, L. Arsiwala, J.H. Lee, K. Elhennawy, P.G. Jost-Brinkmann, F. Demarco, J. Krois, Deep learning for cephalometric landmark detection: systematic review and meta-analysis, Clin Oral Investig 25(7) (2021) 4299-4309, doi: 10.1007/s00784-021-03990-w.
[9] ReportLinker, Artificial intelligence in healthcare market by offering, technology, end-use application, end user and geography – global forecast to 2028, ReportLinker, 2023.
[10] Y.W. Chen, K. Stanley, W. Att, Artificial intelligence in dentistry: current applications and future perspectives, Quintessence Int 51(3) (2020) 248-257, doi: 10.3290/j.qi.a43952.
[11] N. Ahmed, M.S. Abbasi, F. Zuberi, W. Qamar, M.S.B. Halim, A. Maqsood, M.K. Alam, Artificial Intelligence Techniques: Analysis, Application, and Outcome in Dentistry-A Systematic Review, Biomed Res Int 2021 (2021) 9751564, doi: 10.1155/2021/9751564.
[12] F. Schwendicke, T. Golla, M. Dreher, J. Krois, Convolutional neural networks for dental image diagnostics: A scoping review, J Dent 91 (2019) 103226, doi: 10.1016/j.jdent.2019.103226.
[13] F. Schwendicke, J.G. Rossi, G. Gostemeyer, K. Elhennawy, A.G. Cantu, R. Gaudin, A. Chaurasia, S. Gehrung, J. Krois, Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection, J Dent Res 100(4) (2021) 369-376, doi: 10.1177/0022034520972335.
[14] C. Hu, Artificial intelligence is everywhere now. This report shows how we got here., in: P. Science (Ed.) Technology, Stanford University, USA, 2022.

[15] F. Schwendicke, W. Samek, J. Krois, Artificial Intelligence in Dentistry: Chances and Challenges, J Dent Res 99(7) (2020) 769-774, doi: 10.1177/0022034520915714.
[16] J. Arbet, C. Brokamp, J. Meinzen-Derr, K.E. Trinkley, H.M. Spratt, Lessons and tips for designing a machine learning study using EHR data, J Clin Transl Sci 5(1) (2020) e21, doi: 10.1017/cts.2020.513.
[17] B. Li, T. Feridooni, C. Cuen-Ojeda, T. Kishibe, C. de Mestral, M. Mamdani, M. Al-Omran, Machine learning in vascular surgery: a systematic review and critical appraisal, NPJ Digit Med 5(1) (2022) 7, doi: 10.1038/s41746-021-00552-y.
[18] L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023) doi: 10.3390/jcm12030937.
[19] F. Carrillo-Perez, O.E. Pecho, J.C. Morales, R.D. Paravina, A. Della Bona, R. Ghinea, R. Pulgar, M.D.M. Perez, L.J. Herrera, Applications of artificial intelligence in dentistry: A comprehensive review, J Esthet Restor Dent 34(1) (2022) 259-280, doi: 10.1111/jerd.12844.
[20] A. Hasuike, T. Watanabe, S. Wakuda, K. Kogure, R. Yanagiya, K.M. Byrd, S. Sato, Machine Learning in Predicting Tooth Loss: A Systematic Review and Risk of Bias Assessment, J Pers Med 12(10) (2022) doi: 10.3390/jpm12101682.
[21] J. Scott, A.M. Biancardi, O. Jones, D. Andrew, Artificial Intelligence in Periodontology: A Scoping Review, Dent J (Basel) 11(2) (2023) doi: 10.3390/dj11020043.
[22] M. Prados-Privado, J. Garcia Villalon, C.H. Martinez-Martinez, C. Ivorra, J.C. Prados-Frutos, Dental Caries Diagnosis and Detection Using Neural Networks: A Systematic Review, J Clin Med 9(11) (2020) doi: 10.3390/jcm9113579.
[23] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, Comput Methods Programs Biomed 161 (2018) 1-13, doi: 10.1016/j.cmpb.2018.04.005.
[24] A. Ben-Hamadou, O. Smaoui, H. Chaabouni-Chouayakh, A. Rekik, S. Pujades, E. Boyer, J. Strippoli, A. Thollot, H. Setbon, C. Trosset, E. Ladroit, Teeth3DS: a benchmark for teeth segmentation and labeling from intra-oral 3D scans, 2022, p. arXiv:2210.06094, doi: 10.48550/arXiv.2210.06094.
[25] A.G. Cantu, S. Gehrung, J. Krois, A. Chaurasia, J.G. Rossi, R. Gaudin, K. Elhennawy, F. Schwendicke, Detecting caries lesions of different radiographic extension on bitewings using deep learning, J Dent 100 (2020) 103425, doi: 10.1016/j.jdent.2020.103425.
[26] S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849.
[27] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M.M. Leeflang, J.A. Sterne, P.M. Bossuyt, Q.-. Group, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann Intern Med 155(8) (2011) 529-36, doi: 10.7326/0003-4819-155-8-201110180-00009.
[28] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement, BMC Med 13 (2015) 1, doi: 10.1186/s12916-014-0241-z.
[29] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A.I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J.R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, A. Ruggiero, A. Korhonen, E. Jefferson, E. Ako, G. Langs, G. Gozaliasl, G. Yang, H. Prosch, J. Preller, J. Stanczuk, J. Tang, J. Hofmanninger, J. Babar, L.E. Sánchez, M. Thillai, P.M. Gonzalez, P. Teare, X. Zhu, M. Patel, C. Cafolla, H. Azadbakht, J. Jacob, J. Lowe, K. Zhang, K. Bradley, M. Wassin, M. Holzer, K. Ji, M.D. Ortet, T. Ai, N. Walton, P.

Lio, S. Stranks, T. Shadbahr, W. Lin, Y. Zha, Z. Niu, J.H.F. Rudd, E. Sala, C.-B. Schönlieb, C. Aix, Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nature Machine Intelligence 3(3) (2021) 199-217, doi: 10.1038/s42256-021-00307-0.

[30] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021.

[31] N.A. Fakhriy, I. Ardiyanto, H.A. Nugroho, G.N.P. Pratama, Machine Learning Algorithms for Classifying Abscessed and Impacted Tooth: Comparison Study, 2018 2nd International Conference on Biomedical Engineering (IBIOMED), 2018, pp. 88-93, doi: 10.1109/IBIOMED.2018.8534887.

[32] T. Ekert, J. Krois, L. Meinhold, K. Elhennawy, R. Emara, T. Golla, F. Schwendicke, Deep Learning for the Radiographic Detection of Apical Lesions, Journal of Endodontics 45(7) (2019) 917-922.e5, doi: 10.1016/j.joen.2019.03.016.

[33] G. Forman, M. Scholz, Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement, SIGKDD Explor. Newsl. 12(1) (2010) 49–57, doi: 10.1145/1882471.1882479.

[34] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, J.G. Lijmer, D. Moher, D. Rennie, H.C. de Vet, H.Y. Kressel, N. Rifai, R.M. Golub, D.G. Altman, L. Hooft, D.A. Korevaar, J.F. Cohen, S. Group, STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies, BMJ 351 (2015) h5527, doi: 10.1136/bmj.h5527.

[35] F. Schwendicke, T. Singh, J.H. Lee, R. Gaudin, A. Chaurasia, T. Wiegand, S. Uribe, J. Krois, I.e.-o.h. network, I.T.U.W.H.O.f.g.A.I.f.H. the, Artificial intelligence in dental research: Checklist for authors, reviewers, readers, J Dent 107 (2021) 103610, doi: 10.1016/j.jdent.2021.103610.

[36] L. Schneider, L. Arsiwala-Scheppach, J. Krois, H. Meyer-Lueckel, K.K. Bressem, S.M. Niehues, F. Schwendicke, Benchmarking Deep Learning Models for Tooth Structure Segmentation, J Dent Res 101(11) (2022) 1343-1349, doi: 10.1177/00220345221100169.

[37] L. Meinhold, J. Krois, R. Jordan, N. Nestler, F. Schwendicke, Clustering effects of oral conditions based on clinical and radiographic examinations, Clin Oral Investig 24(9) (2020) 3001-3008, doi: 10.1007/s00784-019-03164-9.

[38] X. Liu, S. Cruz Rivera, D. Moher, M.J. Calvert, A.K. Denniston, A.I. Spirit, C.-A.W. Group, Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension, Nat Med 26(9) (2020) 1364-1374, doi: 10.1038/s41591-020-1034-x.

[39] I.J. Aliaga, V. Vera, J.F. De Paz, A.E. García, M.S. Mohamad, Modelling the Longevity of Dental Restorations by means of a CBR System, BioMed Research International 2015 (2015) 540306, doi: 10.1155/2015/540306.

[40] A. Gupta, O.P. Kharbanda, V. Sardana, R. Balachandran, H.K. Sardana, A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images, Int J Comput Assist Radiol Surg 10(11) (2015) 1737-52, doi: 10.1007/s11548-015-1173-6.

[41] A. Gupta, O.P. Kharbanda, V. Sardana, R. Balachandran, H.K. Sardana, Accuracy of 3D cephalometric measurements based on an automatic knowledge-based landmark detection algorithm, Int J Comput Assist Radiol Surg 11(7) (2016) 1297-309, doi: 10.1007/s11548-015-1334-7.

[42] A.J. Hadley, K.R. Krival, A.L. Ridgel, E.C. Hahn, D.J. Tyler, Neural Network Pattern Recognition of Lingual–Palatal Pressure for Automated Detection of Swallow, Dysphagia 30(2) (2015) 176-187, doi: 10.1007/s00455-014-9593-y.

[43] M.S. Kavitha, S.Y. An, C.H. An, K.H. Huh, W.J. Yi, M.S. Heo, S.S. Lee, S.C. Choi, Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women, Oral Surg Oral Med Oral Pathol Oral Radiol 119(3) (2015) 346-56, doi: 10.1016/j.oooo.2014.11.009.
[44] A. Mansoor, V. Patsekin, D. Scherl, J.P. Robinson, B. Rajwa, A Statistical Modeling Approach to Computer-Aided Quantification of Dental Biofilm, IEEE Journal of Biomedical and Health Informatics 19(1) (2015) 358-366, doi: 10.1109/JBHI.2014.2310204.
[45] S.K. Jung, T.W. Kim, New approach for the diagnosis of extractions with neural network machine learning, Am J Orthod Dentofacial Orthop 149(1) (2016) 127-33, doi: 10.1016/j.ajodo.2015.07.030.
[46] M.S. Kavitha, P.G. Kumar, S.-Y. Park, K.-H. Huh, M.-S. Heo, T. Kurita, A. Asano, S.-Y. An, S.-I. Chien, Automatic detection of osteoporosis based on hybrid genetic swarm fuzzy classifier approaches, Dentomaxillofacial Radiology 45(7) (2016) 20160076, doi: 10.1259/dmfr.20160076.
[47] Y.E. Mahmoud, S.S. Labib, H.M.O. Mokhtar, Teeth periapical lesion prediction using machine learning techniques, 2016 SAI Computing Conference (SAI), 2016, pp. 129-134, doi: 10.1109/SAI.2016.7555972.
[48] L. Wang, S. Li, R. Chen, S.-Y. Liu, J.-C. Chen, An Automatic Segmentation and Classification Framework Based on PCNN Model for Single Tooth in MicroCT Images, PLOS ONE 11(6) (2016) e0157694, doi: 10.1371/journal.pone.0157694.
[49] J. De Tobel, P. Radesh, D. Vandermeulen, P.W. Thevissen, An automated technique to stage lower third molar development on panoramic radiographs for age estimation: a pilot study, J Forensic Odontostomatol 35(2) (2017) 42-54, doi, PMC6100230
[50] J.J. Hwang, J.H. Lee, S.S. Han, Y.H. Kim, H.G. Jeong, Y.J. Choi, W. Park, Strut analysis for osteoporosis detection model using dental panoramic radiography, Dentomaxillofac Radiol 46(7) (2017) 20170006, doi: 10.1259/dmfr.20170006.
[51] S. Imangaliyev, M.H. van der Veen, C. Volgenant, B.G. Loos, B.J. Keijser, W. Crielaard, E. Levin, Classification of quantitative light-induced fluorescence images using convolutional neural network, arXiv preprint arXiv:1705.09193 (2017) doi: 10.48550/arXiv.1705.09193.
[52] M. Johari, F. Esmaeili, A. Andalib, S. Garjani, H. Saberkari, Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: an ex vivo study, Dentomaxillofac Radiol 46(2) (2017) 20160107, doi: 10.1259/dmfr.20160107.
[53] Y. Liu, Y. Li, Y. Fu, T. Liu, X. Liu, X. Zhang, J. Fu, X. Guan, T. Chen, X. Chen, Z. Sun, Quantitative prediction of oral cancer risk in patients with oral leukoplakia, Oncotarget 8(28) (2017) doi: 10.18632/oncotarget.17550.
[54] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, H. Fujita, Classification of teeth in cone-beam CT using deep convolutional neural network, Comput Biol Med 80 (2017) 24-29, doi: 10.1016/j.compbiomed.2016.11.003.
[55] A.B. Oktay, Tooth detection with Convolutional Neural Networks, 2017 Medical Technologies National Congress (TIPTEKNO), 2017, pp. 1-4, doi: 10.1109/TIPTEKNO.2017.8238075.
[56] S.A. Prajapati, R. Nagaraj, S. Mitra, Classification of dental diseases using CNN and transfer learning, 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), 2017, pp. 70-74, doi: 10.1109/ISCBI.2017.8053547.
[57] S. Raith, E.P. Vogel, N. Anees, C. Keul, J.-F. Güth, D. Edelhoff, H. Fischer, Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data, Computers in Biology and Medicine 80 (2017) 65-76, doi: 10.1016/j.compbiomed.2016.11.013.

[58] A. Rana, G. Yauney, L.C. Wong, O. Gupta, A. Muftu, P. Shah, Automated segmentation of gingival diseases from oral images, 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), 2017, pp. 144-147, doi: 10.1109/HIC.2017.8227605.
[59] M.M. Srivastava, P. Kumar, L. Pradhan, S. Varadarajan, Detection of tooth caries in bitewing radiographs using deep learning, arXiv preprint arXiv:1711.07312 (2017) doi: 10.48550/arXiv.1711.07312.
[60] M. Štepanovský, A. Ibrová, Z. Buk, J. Velemínská, Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods, Forensic Sci Int 279 (2017) 72-82, doi: 10.1016/j.forsciint.2017.08.005.
[61] E. Yilmaz, T. Kayikcioglu, S. Kayipmaz, Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography, Comput Methods Programs Biomed 146 (2017) 91-100, doi: 10.1016/j.cmpb.2017.05.012.
[62] X. Du, Y. Chen, J. Zhao, Y. Xi, A Convolutional Neural Network Based Auto-Positioning Method For Dental Arch In Rotational Panoramic Radiography, 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 2615-2618, doi: 10.1109/EMBC.2018.8512732.
[63] J. Egger, B. Pfarrkirchner, C. Gsaxner, L. Lindner, D. Schmalstieg, J. Wallner, Fully Convolutional Mandible Segmentation on a valid Ground- Truth Dataset, 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 656-660, doi: 10.1109/EMBC.2018.8512458.
[64] A. Fariza, A.Z. Arifin, E.R. Astuti, Interactive Segmentation of Conditional Spatial FCM with Gaussian Kernel-Based for Panoramic Radiography, 2018 International Symposium on Advanced Intelligent Informatics (SAIN), 2018, pp. 157-161, doi: 10.1109/SAIN.2018.8673381.
[65] L.G. Gavinho, S.A. Araujo, S.K. Bussadori, J.V.P. Silva, A.M. Deana, Detection of white spot lesions by segmenting laser speckle images using computer vision methods, Lasers Med Sci 33(7) (2018) 1565-1571, doi: 10.1007/s10103-018-2520-y.
[66] S.-R. Ha, H.S. Park, E.-H. Kim, H.-K. Kim, J.-Y. Yang, J. Heo, I.-S.L. Yeo, A pilot study using machine learning methods about factors influencing prognosis of dental implants, J Adv Prosthodont 10(6) (2018) 395-400, doi: 10.4047/jap.2018.10.6.395.
[67] A. Heinrich, F. Güttler, S. Wendt, S. Schenkl, M. Hubig, R. Wagner, G. Mall, U. Teichgräber, Forensic Odontology: Automatic Identification of Persons Comparing Antemortem and Postmortem Panoramic Radiographs Using Computer Vision, Rofo 190(12) (2018) 1152-1158, doi: 10.1055/a-0632-4744.
[68] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira, Deep Instance Segmentation of Teeth in Panoramic X-Ray Images, 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 400-407, doi: 10.1109/SIBGRAPI.2018.00058.
[69] M.-X. Jiang, Y.-M. Chen, W.-H. Huang, P.-H. Huang, Y.-H. Tsai, H. Yu-Hsuan, C.-K. Chiang, Teeth-Brushing Recognition Based on Deep Learning, 2018, pp. 1-2, doi: 10.1109/ICCE-China.2018.8448684.
[70] D.W. Kim, H. Kim, W. Nam, H.J. Kim, I.-H. Cha, Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report, Bone 116 (2018) 207-214, doi: 10.1016/j.bone.2018.04.020.
[71] J.H. Lee, D.H. Kim, S.N. Jeong, S.H. Choi, Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm, J Periodontal Implant Sci 48(2) (2018) 114-123, doi: 10.5051/jpis.2018.48.2.114.

[72] J.H. Lee, D.H. Kim, S.N. Jeong, S.H. Choi, Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm, J Dent 77 (2018) 106-111, doi: 10.1016/j.jdent.2018.07.015.
[73] S. Lu, J. Yang, W. Wang, Z. Li, Z. Lu, Teeth Classification Based on Extreme Learning Machine, 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2018, pp. 198-202, doi: 10.1109/WorldS4.2018.8611570.
[74] H. Shah, P. Hernandez, F. Budin, D. Chittajallu, J.B. Vimort, R. Walters, A. Mol, A. Khan, B. Paniagua, Automatic quantification framework to detect cracks in teeth, Proc SPIE Int Soc Opt Eng 10578 (2018) doi: 10.1117/12.2293603.
[75] B. Song, S. Sunny, R.D. Uthoff, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, A. Anbarani, P. Wilder-Smith, M.A. Kuriakose, P. Birur, J.J. Rodriguez, R. Liang, Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning, Biomed. Opt. Express 9(11) (2018) 5318-5329, doi: 10.1364/BOE.9.005318.
[76] B. Thanathornwong, Bayesian-Based Decision Support System for Assessing the Needs for Orthodontic Treatment, Healthc Inform Res 24(1) (2018) 22-28, doi: 10.4258/hir.2018.24.1.22.
[77] J. Yang, Y. Xie, L. Liu, B. Xia, Z. Cao, C. Guo, Automated Dental Image Analysis by Deep Learning on Small Dataset, 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018, pp. 492-497, doi: 10.1109/COMPSAC.2018.00076.
[78] S. Yoon, M. Odlum, Y. Lee, T. Choi, I.M. Kronish, K.W. Davidson, J. Finkelstein, Applying Deep Learning to Understand Predictors of Tooth Mobility Among Urban Latinos, Stud Health Technol Inform 251 (2018) 241-244, doi: Not available.
[79] A. Zakirov, M. Ezhov, M. Gusarev, V. Alexandrovsky, E. Shumilov, Dental pathology detection in 3D cone-beam CT, arXiv preprint arXiv:1810.10309 (2018) doi: 10.48550/arXiv.1810.10309.
[80] L.A. Zanella-Calzada, C.E. Galván-Tejada, N.M. Chávez-Lamas, J. Rivas-Gutierrez, R. Magallanes-Quintanar, J.M. Celaya-Padilla, J.I. Galván-Tejada, H. Gamboa-Rosales, Deep Artificial Neural Networks for the Diagnostic of Caries Using Socioeconomic and Nutritional Features as Determinants: Data from NHANES 2013–2014, Bioengineering 5(2) (2018) 47, doi: 10.3390/bioengineering5020047.
[81] K. Zhang, J. Wu, H. Chen, P. Lyu, An effective teeth recognition method using label tree with cascade network structure, Computerized Medical Imaging and Graphics 68 (2018) 61-70, doi: 10.1016/j.compmedimag.2018.07.001.
[82] H. Ali, M. Khursheed, S.K. Fatima, S.M. Shuja, S. Noor, Object Recognition for Dental Instruments Using SSD-MobileNet, 2019 International Conference on Information Science and Communication Technology (ICISCT), 2019, pp. 1-6, doi: 10.1109/CISCT.2019.8777441.
[83] S. Alkaabi, S. Yussof, S. Al-Mulla, Evaluation of Convolutional Neural Network based on Dental Images for Age Estimation, 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2019, pp. 1-5, doi: 10.1109/ICECTA48151.2019.8959665.
[84] B. Askarian, F. Tabei, G.A. Tipton, J.W. Chong, Smartphone-Based Method for Detecting Periodontal Disease, 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), 2019, pp. 53-55, doi: 10.1109/HI-POCT45284.2019.8962844.
[85] M. Bouchahma, S.B. Hammouda, S. Kouki, M. Alshemaili, K. Samara, An Automatic Dental Decay Treatment Prediction using a Deep Convolutional Neural Network on X-

Ray Images, 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), 2019, pp. 1-4, doi: 10.1109/AICCSA47632.2019.9035278.
[86] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schurmann, I. Krejci, H. Markram, Caries Detection with Near-Infrared Transillumination Using Deep Learning, J Dent Res 98(11) (2019) 1227-1233, doi: 10.1177/0022034519871884.
[87] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, C.-H. Lee, A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films, Scientific Reports 9(1) (2019) 3840, doi: 10.1038/s41598-019-40414-y.
[88] B. Cheng, W. Wang, Dental hard tissue morphological segmentation with sparse representation-based classifier, Med Biol Eng Comput 57(8) (2019) 1629-1643, doi: 10.1007/s11517-019-01985-0.
[89] C.L. Chin, J.W. Lin, C.S. Wei, M.C. Hsu, Dentition Labeling And Root Canal Recognition Using Ganand Rule-Based System, 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2019, pp. 1-6, doi: 10.1109/TAAI48200.2019.8959926.
[90] H.I. Choi, S.K. Jung, S.H. Baek, W.H. Lim, S.J. Ahn, I.H. Yang, T.W. Kim, Artificial Intelligent Model With Neural Network Machine Learning for the Diagnosis of Orthognathic Surgery, J Craniofac Surg 30(7) (2019) 1986-1989, doi: 10.1097/scs.0000000000005650.
[91] Z. Cui, C. Li, W. Wang, ToothNet: Automatic Tooth Instance Segmentation and Identification From Cone Beam CT Images, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 6361-6370, doi: 10.1109/CVPR.2019.00653.
[92] C. Dasanayaka, B. Dharmasena, W.R. Bandara, M.B. Dissanayake, R. Jayasinghe, Segmentation of Mental Foramen in Dental Panoramic Tomography using Deep Learning, 2019 14th Conference on Industrial and Information Systems (ICIIS), 2019, pp. 81-84, doi: 10.1109/ICIIS47346.2019.9063312.
[93] J.C.D. Cruz, R.G. Garcia, J.C.C.V. Cueto, S.C. Pante, C.G.V. Toral, Automated Human Identification through Dental Image Enhancement and Analysis, 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2019, pp. 1-6, doi: 10.1109/HNICEM48295.2019.9072780.
[94] D.Q. Duong, K.C.T. Nguyen, N.R. Kaipatur, E.H.M. Lou, M. Noga, P.W. Major, K. Punithakumar, L.H. Le, Fully Automated Segmentation of Alveolar Bone Using Deep Convolutional Neural Networks from Intraoral Ultrasound Images, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 6632-6635, doi: 10.1109/EMBC.2019.8857060.
[95] J. Hatvani, A. Basarab, J.Y. Tourneret, M. Gyongy, D. Kouame, A Tensor Factorization Method for 3-D Super Resolution With Application to Dental CT, IEEE Trans Med Imaging 38(6) (2019) 1524-1531, doi: 10.1109/tmi.2018.2883517.
[96] J. Hatvani, A. Horváth, J. Michetti, A. Basarab, D. Kouamé, M. Gyöngy, Deep Learning-Based Super-Resolution Applied to Dental Computed Tomography, IEEE Transactions on Radiation and Plasma Medical Sciences 3(2) (2019) 120-128, doi: 10.1109/TRPMS.2018.2827239.
[97] M.A.A. Hegazy, M.H. Cho, M.H. Cho, S.Y. Lee, U-net based metal segmentation on projection domain for metal artifact reduction in dental CT, Biomedical Engineering Letters 9(3) (2019) 375-385, doi: 10.1007/s13534-019-00110-2.
[98] T. Hiraiwa, Y. Ariji, M. Fukuda, Y. Kise, K. Nakata, A. Katsumata, H. Fujita, E. Ariji, A deep-learning artificial intelligence system for assessment of root morphology of the

mandibular first molar on panoramic radiography, Dentomaxillofac Radiol 48(3) (2019) 20180218, doi: 10.1259/dmfr.20180218.
[99] Z. Hu, C. Jiang, F. Sun, Q. Zhang, Y. Ge, Y. Yang, X. Liu, H. Zheng, D. Liang, Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks, Medical Physics 46(4) (2019) 1686-1696, doi: 10.1002/mp.13415.
[100] M. Hung, M.W. Voss, M.N. Rosales, W. Li, W. Su, J. Xu, J. Bounsanga, B. Ruiz-Negrón, E. Lauren, F.W. Licari, Application of machine learning for diagnostic prediction of root caries, Gerodontology 36(4) (2019) 395-404, doi: 10.1111/ger.12432.
[101] I. Ilić, M. Vodanović, M. Subašić, Gender Estimation from Panoramic Dental X-ray Images using Deep Convolutional Networks, IEEE EUROCON 2019 -18th International Conference on Smart Technologies, 2019, pp. 1-5, doi: 10.1109/EUROCON.2019.8861726.
[102] L. Kats, M. Vered, A. Zlotogorski-Hurvitz, I. Harpaz, Atherosclerotic carotid plaque on panoramic radiographs: neural network detection, Int J Comput Dent 22(2) (2019) 163-169, doi: Not available.
[103] L. Kats, M. Vered, A. Zlotogorski-Hurvitz, I. Harpaz, Atherosclerotic carotid plaques on panoramic imaging: an automatic detection using deep learning with small dataset, arXiv preprint arXiv:1808.08093 (2018) doi: 10.48550/arXiv.1808.08093.
[104] D.W. Kim, S. Lee, S. Kwon, W. Nam, I.-H. Cha, H.J. Kim, Deep learning-based survival prediction of oral cancer patients, Scientific Reports 9(1) (2019) 6994, doi: 10.1038/s41598-019-43372-7.
[105] J. Kim, H.-S. Lee, I.-S. Song, K.-H. Jung, DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs, Scientific Reports 9(1) (2019) 17615, doi: 10.1038/s41598-019-53758-2.
[106] Y. Kise, M. Shimizu, H. Ikeda, T. Fujii, C. Kuwada, M. Nishiyama, T. Funakoshi, Y. Ariji, H. Fujita, A. Katsumata, K. Yoshiura, E. Ariji, Usefulness of a deep learning system for diagnosing Sjögren's syndrome using ultrasonography images, Dentomaxillofac Radiol 49(3) (2020) 20190348, doi: 10.1259/dmfr.20190348.
[107] T.L. Koch, M. Perslev, C. Igel, S.S. Brandt, Accurate Segmentation of Dental Panoramic Radiographs with U-NETS, 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 15-19, doi: 10.1109/ISBI.2019.8759563.
[108] J. Krois, T. Ekert, L. Meinhold, T. Golla, B. Kharbot, A. Wittemeier, C. Dörfer, F. Schwendicke, Deep Learning for the Radiographic Detection of Periodontal Bone Loss, Scientific Reports 9(1) (2019) 8495, doi: 10.1038/s41598-019-44839-3.
[109] J.S. Lee, S. Adhikari, L. Liu, H.G. Jeong, H. Kim, S.J. Yoon, Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study, Dentomaxillofac Radiol 48(1) (2019) 20170344, doi: 10.1259/dmfr.20170344.
[110] X. Li, Y. Zhang, Q. Cui, X. Yi, Y. Zhang, Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features, IEEE Trans Cybern 49(2) (2019) 380-387, doi: 10.1109/tcyb.2017.2772289.
[111] L. Liu, J. Xu, Y. Huan, Z. Zou, S.C. Yeh, L.R. Zheng, A Smart Dental Health-IoT Platform Based on Intelligent Hardware, Deep Learning, and Mobile Terminal, IEEE Journal of Biomedical and Health Informatics 24(3) (2020) 898-906, doi: 10.1109/JBHI.2019.2919916.
[112] Y. Liu, X. Shang, Z. Shen, B. Hu, Z. Wang, G. Xiong, 3D Deep Learning for 3D Printing of Tooth Model, 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 2019, pp. 274-279, doi: 10.1109/SOLI48380.2019.8955074.

[113] D. Milošević, M. Vodanović, I. Galić, M. Subašić, Estimating Biological Gender from Panoramic Dental X-Ray Images, 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 105-110, doi: 10.1109/ISPA.2019.8868804.
[114] J. Minnema, M. van Eijnatten, A.A. Hendriksen, N. Liberton, D.M. Pelt, K.J. Batenburg, T. Forouzanfar, J. Wolff, Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network, Medical Physics 46(11) (2019) 5027-5035, doi: 10.1002/mp.13793.
[115] Y. Moriyama, C. Lee, S. Date, Y. Kashiwagi, Y. Narukawa, K. Nozaki, S. Murakami, Evaluation of Dental Image Augmentation for the Severity Assessment of Periodontal Disease, 2019 International Conference on Computational Science and Computational Intelligence (CSCI), 2019, pp. 924-929, doi: 10.1109/CSCI49370.2019.00176.
[116] K. Moutselos, E. Berdouses, C. Oulis, I. Maglogiannis, Recognizing Occlusal Caries in Dental Intraoral Images Using Deep Learning, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1617-1620, doi: 10.1109/EMBC.2019.8856553.
[117] M. Murata, Y. Ariji, Y. Ohashi, T. Kawai, M. Fukuda, T. Funakoshi, Y. Kise, M. Nozawa, A. Katsumata, H. Fujita, E. Ariji, Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography, Oral Radiol 35(3) (2019) 301-307, doi: 10.1007/s11282-018-0363-7.
[118] R. Patcas, D.A.J. Bernini, A. Volokitin, E. Agustsson, R. Rothe, R. Timofte, Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age, Int J Oral Maxillofac Surg 48(1) (2019) 77-83, doi: 10.1016/j.ijom.2018.07.010.
[119] R. Patcas, R. Timofte, A. Volokitin, E. Agustsson, T. Eliades, M. Eichenberger, M.M. Bornstein, Facial attractiveness of cleft patients: a direct comparison between artificial-intelligence-based scoring and conventional rater groups, European Journal of Orthodontics 41(4) (2019) 428-433, doi: 10.1093/ejo/cjz007.
[120] M. Sajad, I. Shafi, J. Ahmad, Automatic Lesion Detection in Periapical X-rays, 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2019, pp. 1-6, doi: 10.1109/ICECCE47252.2019.8940661.
[121] G.B. Senirkentli, S. Sen, O. Farsak, E. Bostanci, A Neural Expert System Based Dental Trauma Diagnosis Application, 2019 Medical Technologies Congress (TIPTEKNO), 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895077.
[122] B. Stark, M. Samarah, Ensemble and Deep Learning for Real-time Sensors Evaluation of algorithms for real-time sensors with application for detecting brushing location, 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019, pp. 555-559, doi: 10.1109/ICCC47050.2019.9064080.
[123] S. Tian, N. Dai, B. Zhang, F. Yuan, Q. Yu, X. Cheng, Automatic Classification and Segmentation of Teeth on 3D Dental Model Using Hierarchical Deep Learning Networks, IEEE Access 7 (2019) 84817-84828, doi: 10.1109/ACCESS.2019.2924262.
[124] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S.I. Nikolenko, M.M. Sveshnikov, G.B. Bednenko, Tooth detection and numbering in panoramic radiographs using convolutional neural networks, Dentomaxillofac Radiol 48(4) (2019) 20180051, doi: 10.1259/dmfr.20180051.
[125] S. Vinayahalingam, T. Xi, S. Bergé, T. Maal, G. de Jong, Automated detection of third molars and mandibular nerve by deep learning, Scientific Reports 9(1) (2019) 9007, doi: 10.1038/s41598-019-45487-3.
[126] J. Woo, F. Xing, J.L. Prince, M. Stone, J.R. Green, T. Goldsmith, T.G. Reese, V.J. Wedeen, G.E. Fakhri, Differentiating post-cancer from healthy tongue muscle

coordination patterns during speech using deep learning, The Journal of the Acoustical Society of America 145(5) (2019) EL423-EL429, doi: 10.1121/1.5103191.
[127] X. Xu, C. Liu, Y. Zheng, 3D Tooth Segmentation and Labeling Using Deep Convolutional Neural Networks, IEEE Transactions on Visualization and Computer Graphics 25(7) (2019) 2336-2348, doi: 10.1109/TVCG.2018.2839685.
[128] S. Yamaguchi, C. Lee, O. Karaer, S. Ban, A. Mine, S. Imazato, Predicting the Debonding of CAD/CAM Composite Resin Crowns with AI, J Dent Res 98(11) (2019) 1234-1238, doi: 10.1177/0022034519867641.
[129] R. Abdalla-Aslan, T. Yeshua, D. Kabla, I. Leichter, C. Nadler, An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography, Oral Surg Oral Med Oral Pathol Oral Radiol 130(5) (2020) 593-602, doi: 10.1016/j.oooo.2020.05.012.
[130] G. Yauney, A. Rana, L.C. Wong, P. Javia, A. Muftu, P. Shah, Automated Process Incorporating Machine Learning Segmentation and Correlation of Oral Diseases with Systemic Health, Annu Int Conf IEEE Eng Med Biol Soc 2019 (2019) 3387-3393, doi: 10.1109/embc.2019.8857965.
[131] D.M. Alalharith, H.M. Alharthi, W.M. Alghamdi, Y.M. Alsenbel, N. Aslam, I.U. Khan, S.Y. Shahin, S. Dianišková, M.S. Alhareky, K.K. Barouch, A Deep Learning-Based Approach for the Detection of Early Signs of Gingivitis in Orthodontic Patients Using Faster Region-Based Convolutional Neural Networks, International Journal of Environmental Research and Public Health 17(22) (2020) 8447, doi: 10.3390/ijerph17228447.
[132] I. Aliaga, V. Vera, M. Vera, E. García, M. Pedrera, G. Pajares, Automatic computation of mandibular indices in dental panoramic radiographs for early osteoporosis detection, Artificial Intelligence in Medicine 103 (2020) 101816, doi: 10.1016/j.artmed.2020.101816.
[133] N. Banar, J. Bertels, F. Laurent, R.M. Boedi, J. De Tobel, P. Thevissen, D. Vandermeulen, Towards fully automated third molar development staging in panoramic radiographs, International Journal of Legal Medicine 134(5) (2020) 1831-1841, doi: 10.1007/s00414-020-02283-3.
[134] H.-J. Chang, S.-J. Lee, T.-H. Yong, N.-Y. Shin, B.-G. Jang, J.-E. Kim, K.-H. Huh, S.-S. Lee, M.-S. Heo, S.-C. Choi, T.-I. Kim, W.-J. Yi, Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis, Scientific Reports 10(1) (2020) 7531, doi: 10.1038/s41598-020-64509-z.
[135] S. Chen, L. Wang, G. Li, T.H. Wu, S. Diachina, B. Tejera, J.J. Kwon, F.C. Lin, Y.T. Lee, T. Xu, D. Shen, C.C. Ko, Machine learning in orthodontics: Introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients, Angle Orthod 90(1) (2020) 77-84, doi: 10.2319/012919-59.1.
[136] Y. Chen, H. Du, Z. Yun, S. Yang, Z. Dai, L. Zhong, Q. Feng, W. Yang, Automatic Segmentation of Individual Tooth in Dental CBCT Images From Tooth Surface Map by a Multi-Task FCN, IEEE Access 8 (2020) 97296-97309, doi: 10.1109/ACCESS.2020.2991799.
[137] M. Chung, M. Lee, J. Hong, S. Park, J. Lee, J. Lee, I.-H. Yang, J. Lee, Y.-G. Shin, Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation, Computers in Biology and Medicine 120 (2020) 103720, doi: 10.1016/j.compbiomed.2020.103720.
[138] M.G. Endres, F. Hillen, M. Salloumis, A.R. Sedaghat, S.M. Niehues, O. Quatela, H. Hanken, R. Smeets, B. Beck-Broichsitter, C. Rendenbach, K. Lakhani, M. Heiland, R.A.

Gaudin, Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs, Diagnostics 10(6) (2020) 430, doi: 10.3390/diagnostics10060430.
[139] M. Farhadian, P. Shokouhi, P. Torkzaban, A decision support system based on support vector machine for diagnosis of periodontal disease, BMC Res Notes 13(1) (2020) 337, doi: 10.1186/s13104-020-05180-5.
[140] F. Fan, W. Ke, W. Wu, X. Tian, T. Lyu, Y. Liu, P. Liao, X. Dai, H. Chen, Z. Deng, Automatic human identification from panoramic dental radiographs using the convolutional neural network, Forensic Science International 314 (2020) 110416, doi: 10.1016/j.forsciint.2020.110416.
[141] N. Fujima, V.C. Andreu-Arasa, S.K. Meibom, G.A. Mercier, A.R. Salama, M.T. Truong, O. Sakai, Deep learning analysis using FDG-PET to predict treatment outcome in patients with oral cavity squamous cell carcinoma, European Radiology 30(11) (2020) 6322-6330, doi: 10.1007/s00330-020-06982-8.
[142] M. Fukuda, Y. Ariji, Y. Kise, M. Nozawa, C. Kuwada, T. Funakoshi, C. Muramatsu, H. Fujita, A. Katsumata, E. Ariji, Comparison of 3 deep learning neural networks for classifying the relationship between the mandibular third molar and the mandibular canal on panoramic radiographs, Oral Surg Oral Med Oral Pathol Oral Radiol 130(3) (2020) 336-343, doi: 10.1016/j.oooo.2020.04.005.
[143] M. Fukuda, K. Inamoto, N. Shibata, Y. Ariji, Y. Yanashita, S. Kutsuna, K. Nakata, A. Katsumata, H. Fujita, E. Ariji, Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography, Oral Radiol 36(4) (2020) 337-343, doi: 10.1007/s11282-019-00409-x.
[144] V. Geetha, K.S. Aprameya, D.M. Hinduja, Dental caries diagnosis in digital radiographs using back-propagation neural network, Health Information Science and Systems 8(1) (2020) 8, doi: 10.1007/s13755-019-0096-y.
[145] M. Hung, E.S. Hon, B. Ruiz-Negron, E. Lauren, R. Moffat, W. Su, J. Xu, J. Park, D. Prince, J. Cheever, F.W. Licari, Exploring the Intersection between Social Determinants of Health and Unmet Dental Care Needs Using Deep Learning, International Journal of Environmental Research and Public Health 17(19) (2020) 7286, doi: 10.3390/ijerph17197286.
[146] M. Hung, W. Li, E.S. Hon, S. Su, W. Su, Y. He, X. Sheng, R. Holubkov, M.S. Lipsky, Prediction of 30-Day Hospital Readmissions for All-Cause Dental Conditions using Machine Learning, Risk Manag Healthc Policy 13 (2020) 2047-2056, doi: 10.2147/rmhp.S272824.
[147] J. Jaskari, J. Sahlsten, J. Järnstedt, H. Mehtonen, K. Karhu, O. Sundqvist, A. Hietanen, V. Varjonen, V. Mattila, K. Kaski, Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes, Scientific Reports 10(1) (2020) 5842, doi: 10.1038/s41598-020-62321-3.
[148] S.H. Jeong, J.P. Yun, H.-G. Yeom, H.J. Lim, J. Lee, B.C. Kim, Deep learning based discrimination of soft tissue profiles requiring orthognathic surgery by facial photographs, Scientific Reports 10(1) (2020) 16235, doi: 10.1038/s41598-020-73287-7.
[149] S.V. Joshi, R.D. Kanphade, Deep Learning Based Person Authentication Using Hand Radiographs: A Forensic Approach, IEEE Access 8 (2020) 95424-95434, doi: 10.1109/ACCESS.2020.2995788.
[150] L. Kats, M. Vered, S. Blumer, E. Kats, Neural Network Detection and Segmentation of Mental Foramen in Panoramic Imaging, J Clin Pediatr Dent 44(3) (2020) 168-173, doi: 10.17796/1053-4625-44.3.6.
[151] H.A. Khan, M.A. Haider, H.A. Ansari, H. Ishaq, A. Kiyani, K. Sohail, M. Muhammad, S.A. Khurram, Automated feature detection in dental periapical radiographs by using

deep learning, Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 131(6) (2021) 711-720, doi: 10.1016/j.oooo.2020.08.024.
[152] H. Kim, E. Shim, J. Park, Y.J. Kim, U. Lee, Y. Kim, Web-based fully automated cephalometric analysis by deep learning, Comput Methods Programs Biomed 194 (2020) 105513, doi: 10.1016/j.cmpb.2020.105513.
[153] I. Kim, D. Misra, L. Rodriguez, M. Gill, D.K. Liberton, K. Almpani, J.S. Lee, S. Antani, Malocclusion Classification on 3D Cone-Beam CT Craniofacial Images Using Multi-Channel Deep Learning Models, 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 1294-1298, doi: 10.1109/EMBC44109.2020.9176672.
[154] J.-E. Kim, N.-E. Nam, J.-S. Shim, Y.-H. Jung, B.-H. Cho, J.J. Hwang, Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs, Journal of Clinical Medicine 9(4) (2020) 1117, doi: 10.3390/jcm9041117.
[155] F. Kunz, A. Stellzig-Eisenhauer, F. Zeman, J. Boldt, Artificial intelligence in orthodontics : Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network, J Orofac Orthop 81(1) (2020) 52-68, doi: 10.1007/s00056-019-00203-8.
[156] N. Kuramoto, K. Ichimura, D. Jayatilake, T. Shimokakimoto, K. Hidaka, K. Suzuki, Deep Learning-Based Swallowing Monitor for Realtime Detection of Swallow Duration, 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 4365-4368, doi: 10.1109/EMBC44109.2020.9176721.
[157] C. Kuwada, Y. Ariji, M. Fukuda, Y. Kise, H. Fujita, A. Katsumata, E. Ariji, Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs, Oral Surg Oral Med Oral Pathol Oral Radiol 130(4) (2020) 464-469, doi: 10.1016/j.oooo.2020.04.813.
[158] G.H. Kwak, E.-J. Kwak, J.M. Song, H.R. Park, Y.-H. Jung, B.-H. Cho, P. Hui, J.J. Hwang, Automatic mandibular canal detection using a deep convolutional neural network, Scientific Reports 10(1) (2020) 5711, doi: 10.1038/s41598-020-62586-8.
[159] D. Lee, C. Park, Y. Lim, H. Cho, A Metal Artifact Reduction Method Using a Fully Convolutional Network in the Sinogram and Image Domains for Dental Computed Tomography, J Digit Imaging 33(2) (2020) 538-546, doi: 10.1007/s10278-019-00297-x.
[160] J.-H. Lee, S.-S. Han, Y.H. Kim, C. Lee, I. Kim, Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs, Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 129(6) (2020) 635-642, doi: 10.1016/j.oooo.2019.11.007.
[161] J.H. Lee, S.N. Jeong, Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: A pilot study, Medicine (Baltimore) 99(26) (2020) e20787, doi: 10.1097/md.0000000000020787.
[162] J.H. Lee, D.H. Kim, S.N. Jeong, Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network, Oral Dis 26(1) (2020) 152-158, doi: 10.1111/odi.13223.
[163] J.-H. Lee, Y.-T. Kim, J.-B. Lee, S.-N. Jeong, A Performance Comparison between Automated Deep Learning and Dental Professionals in Classification of Dental Implant Systems from Dental Imaging: A Multi-Center Study, Diagnostics 10(11) (2020) 910, doi: 10.3390/diagnostics10110910.
[164] J.-H. Lee, H.-J. Yu, M.-j. Kim, J.-W. Kim, J. Choi, Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks, BMC Oral Health 20(1) (2020) 270, doi: 10.1186/s12903-020-01256-7.

[165] K.-S. Lee, S.-K. Jung, J.-J. Ryu, S.-W. Shin, J. Choi, Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs, Journal of Clinical Medicine 9(2) (2020) 392, doi: 10.3390/jcm9020392.
[166] S. Lee, S. Woo, J. Yu, J. Seo, J. Lee, C. Lee, Automated CNN-Based Tooth Segmentation in Cone-Beam CT for Dental Implant Planning, IEEE Access 8 (2020) 50507-50518, doi: 10.1109/ACCESS.2020.2975826.
[167] C. Li, D. Zhang, S. Chen, Research about Tongue Image of Traditional Chinese Medicine(TCM) Based on Artificial Intelligence Technology, 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), 2020, pp. 633-636, doi: 10.1109/ITOEC49072.2020.9141557.
[168] Q. Li, K. Chen, L. Han, Y. Zhuang, J. Li, J. Lin, Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN, Journal of X-Ray Science and Technology 28 (2020) 905-922, doi: 10.3233/XST-200678.
[169] S. Li, Z. Pang, W. Song, Y. Guo, W. You, A. Hao, H. Qin, Low-Shot Learning of Automatic Dental Plaque Segmentation Based on Local-to-Global Feature Fusion, 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 664-668, doi: 10.1109/ISBI45749.2020.9098741.
[170] C. Lian, L. Wang, T.H. Wu, F. Wang, P.T. Yap, C.C. Ko, D. Shen, Deep Multi-Scale Mesh Feature Learning for Automated Labeling of Raw Dental Surfaces From 3D Intraoral Scanners, IEEE Transactions on Medical Imaging 39(7) (2020) 2440-2450, doi: 10.1109/TMI.2020.2971730.
[171] F.P. Mahdi, K. Motoki, S. Kobashi, Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs, Scientific Reports 10(1) (2020) 19261, doi: 10.1038/s41598-020-75887-9.
[172] S. Mallishery, P. Chhatpar, K.S. Banga, T. Shah, P. Gupta, The precision of case difficulty and referral decisions: an innovative automated approach, Clin Oral Investig 24(6) (2020) 1909-1915, doi: 10.1007/s00784-019-03050-4.
[173] S. Matsuda, T. Miyamoto, H. Yoshimura, T. Hasegawa, Personal identification with orthopantomography using simple convolutional neural networks: a preliminary study, Scientific Reports 10(1) (2020) 13559, doi: 10.1038/s41598-020-70474-4.
[174] R. Merdietio Boedi, N. Banar, J. De Tobel, J. Bertels, D. Vandermeulen, P.W. Thevissen, Effect of Lower Third Molar Segmentations on Automated Tooth Development Staging using a Convolutional Neural Network, Journal of Forensic Sciences 65(2) (2020) 481-486, doi: 10.1111/1556-4029.14182.
[175] V.T.N. Ngoc, A.C. Agwu, L.H. Son, T.M. Tuan, C. Nguyen Giap, M.T.G. Thanh, H.B. Duy, T.T. Ngan, The Combination of Adaptive Convolutional Neural Network and Bag of Visual Words in Automatic Diagnosis of Third Molar Complications on Dental X-Ray Images, Diagnostics 10(4) (2020) 209, doi: 10.3390/diagnostics10040209.
[176] J.H. Oh, M. Pouryahya, A. Iyer, A.P. Apte, J.O. Deasy, A. Tannenbaum, A novel kernel Wasserstein distance on Gaussian measures: An application of identifying dental artifacts in head and neck computed tomography, Computers in Biology and Medicine 120 (2020) 103731, doi: 10.1016/j.compbiomed.2020.103731.
[177] K. Orhan, I.S. Bayrakdar, M. Ezhov, A. Kravtsov, T. Ozyurek, Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans, Int Endod J 53(5) (2020) 680-689, doi: 10.1111/iej.13265.
[178] J. Ren, H. Fan, J. Yang, H. Ling, Detection of Trabecular Landmarks for Osteoporosis Prescreening in Dental Panoramic Radiographs, 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 2194-2197, doi: 10.1109/EMBC44109.2020.9175281.

[179] F. Schwendicke, K. Elhennawy, S. Paris, P. Friebertshäuser, J. Krois, Deep learning for caries lesion detection in near-infrared light transillumination images: A pilot study, Journal of Dentistry 92 (2020) 103260, doi: 10.1016/j.jdent.2019.103260.
[180] F.C. Setzer, K.J. Shi, Z. Zhang, H. Yan, H. Yoon, M. Mupparapu, J. Li, Artificial Intelligence for the Computer-aided Detection of Periapical Lesions in Cone-beam Computed Tomographic Images, J Endod 46(7) (2020) 987-993, doi: 10.1016/j.joen.2020.03.025.
[181] S. Sukegawa, K. Yoshii, T. Hara, K. Yamashita, K. Nakano, N. Yamamoto, H. Nagatsuka, Y. Furuki, Deep Neural Networks for Dental Implant System Classification, Biomolecules 10(7) (2020) 984, doi: 10.3390/biom10070984.
[182] D. Sun, Y. Pei, G. Song, Y. Guo, G. Ma, T. Xu, H. Zha, Tooth Segmentation and Labeling from Digital Dental Casts, 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 669-673, doi: 10.1109/ISBI45749.2020.9098397.
[183] T. Takahashi, K. Nozaki, T. Gonda, K. Ikebe, A system for designing removable partial dentures using artificial intelligence. Part 1. Classification of partially edentulous arches using a convolutional neural network, J Prosthodont Res 65(1) (2021) 115-118, doi: 10.2186/jpr.JPOR_2019_354.
[184] W. Tang, Y. Gao, L. Liu, T. Xia, L. He, S. Zhang, J. Guo, W. Li, Q. Xu, An Automatic Recognition of Tooth- Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning, IEEE Access 8 (2020) 153470-153478, doi: 10.1109/ACCESS.2020.3017725.
[185] B. Thanathornwong, S. Suebnukarn, Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks, Imaging Sci Dent 50(2) (2020) 169-174, doi: 10.5624/isd.2020.50.2.169.
[186] N. Vila-Blanco, M.J. Carreira, P. Varas-Quintana, C. Balsa-Castro, I. Tomás, Deep Neural Networks for Chronological Age Estimation From OPG Images, IEEE Transactions on Medical Imaging 39(7) (2020) 2374-2384, doi: 10.1109/TMI.2020.2968765.
[187] M. Vranckx, A. Van Gerven, H. Willems, A. Vandemeulebroucke, A. Ferreira Leite, C. Politis, R. Jacobs, Artificial Intelligence (AI)-Driven Molar Angulation Measurements to Predict Third Molar Eruption on Panoramic Radiographs, International Journal of Environmental Research and Public Health 17(10) (2020) 3716, doi: 10.3390/ijerph17103716.
[188] X. Wang, J. Liu, C. Wu, J. Liu, Q. Li, Y. Chen, X. Wang, X. Chen, X. Pang, B. Chang, J. Lin, S. Zhao, Z. Li, Q. Deng, Y. Lu, D. Zhao, J. Chen, Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark, Comput Struct Biotechnol J 18 (2020) 973-980, doi: 10.1016/j.csbj.2020.04.002.
[189] Y. Wang, R.D. Hays, M. Marcus, C.A. Maida, J. Shen, D. Xiong, I.D. Coulter, S.Y. Lee, V.W. Spolsky, J.J. Crall, H. Liu, Developing Children's Oral Health Assessment Toolkits Using Machine Learning Algorithm, JDR Clinical & Translational Research 5(3) (2020) 233-243, doi: 10.1177/2380084419885612.
[190] M.L. Welch, C. McIntosh, T.G. Purdie, L. Wee, A. Traverso, A. Dekker, B. Haibe-Kains, D.A. Jaffray, Automatic classification of dental artifact status for efficient image veracity checks: effects of image resolution and convolutional neural network depth, Physics in Medicine & Biology 65(1) (2020) 015005, doi: 10.1088/1361-6560/ab5427.
[191] M.L. Welch, C. McIntosh, A. Traverso, L. Wee, T.G. Purdie, A. Dekker, B. Haibe-Kains, D.A. Jaffray, External validation and transfer learning of convolutional neural networks for computed tomography dental artifact classification, Physics in Medicine & Biology 65(3) (2020) 035017, doi: 10.1088/1361-6560/ab63ba.

[192] R.A. Welikala, P. Remagnino, J.H. Lim, C.S. Chan, S. Rajendran, T.G. Kallarakkal, R.B. Zain, R.D. Jayasinghe, J. Rimal, A.R. Kerr, R. Amtha, K. Patil, W.M. Tilakaratne, J. Gibson, S.C. Cheong, S.A. Barman, Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer, IEEE Access 8 (2020) 132677-132693, doi: 10.1109/ACCESS.2020.3010180.
[193] H. Yang, E. Jo, H.J. Kim, I.-h. Cha, Y.-S. Jung, W. Nam, J.-Y. Kim, J.-K. Kim, Y.H. Kim, T.G. Oh, S.-S. Han, H. Kim, D. Kim, Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs, Journal of Clinical Medicine 9(6) (2020) 1839, doi: 10.3390/jcm9061839.
[194] W. You, A. Hao, S. Li, Y. Wang, B. Xia, Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments, BMC Oral Health 20(1) (2020) 141, doi: 10.1186/s12903-020-01114-6.
[195] X. Zhang, Y. Liang, W. Li, C. Liu, D. Gu, W. Sun, L. Miao, Development and evaluation of deep learning for screening dental caries from oral photographs, Oral Dis 28(1) (2022) 173-181, doi: 10.1111/odi.13735.
[196] Z. Zheng, H. Yan, F.C. Setzer, K.J. Shi, M. Mupparapu, J. Li, Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection, IEEE Transactions on Automation Science and Engineering 18(2) (2021) 603-614, doi: 10.1109/TASE.2020.3025871.
[197] G. Zhu, Z. Piao, S.C. Kim, Tooth Detection and Segmentation with Mask R-CNN, 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 070-072, doi: 10.1109/ICAIIC48513.2020.9065216.
[198] H. Alkhadar, M. Macluskey, S. White, I. Ellis, A. Gardner, Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma, Journal of Oral Pathology & Medicine 50(4) (2021) 378-384, doi: 10.1111/jop.13135.
[199] A.F. Leite, A.V. Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaêta-Araujo, M. Vranckx, R. Jacobs, Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs, Clin Oral Investig 25(4) (2021) 2257-2267, doi: 10.1007/s00784-020-03544-6.
[200] R.A. Machado, C. de Oliveira Silva, H. Martelli-Junior, L.T. das Neves, R.D. Coletta, Machine learning in prediction of genetic risk of nonsyndromic oral clefts in the Brazilian population, Clinical Oral Investigations 25(3) (2021) 1273-1280, doi: 10.1007/s00784-020-03433-y.
[201] C. Muramatsu, T. Morishita, R. Takahashi, T. Hayashi, W. Nishiyama, Y. Ariji, X. Zhou, T. Hara, A. Katsumata, E. Ariji, H. Fujita, Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data, Oral Radiol 37(1) (2021) 13-19, doi: 10.1007/s11282-019-00418-w.
[202] Y. Yasa, Ö. Çelik, I.S. Bayrakdar, A. Pekince, K. Orhan, S. Akarsu, S. Atasoy, E. Bilgir, A. Odabaş, A.F. Aslan, An artificial intelligence proposal to automatic teeth detection and numbering in dental bite-wing radiographs, Acta Odontologica Scandinavica 79(4) (2021) 275-281, doi: 10.1080/00016357.2020.1840624.
[203] K.C.T. Nguyen, D.Q. Duong, F.T. Almeida, P.W. Major, N.R. Kaipatur, T.T. Pham, E.H.M. Lou, M. Noga, K. Punithakumar, L.H. Le, Alveolar Bone Segmentation in Intraoral Ultrasonographs with Machine Learning, J Dent Res 99(9) (2020) 1054-1061, doi: 10.1177/0022034520920593.
[204] X. Min, C. Haijin, Research on Rapid Detection of Tooth Profile Parameters of the Clothing Wires Based on Image Processing, 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2020, pp. 586-590, doi: 10.1109/ICAICA50127.2020.9182581.

[205] S. Rasteau, N. Sigaux, A. Louvrier, P. Bouletreau, Three-dimensional acquisition technologies for facial soft tissues – Applications and prospects in orthognathic surgery, Journal of Stomatology, Oral and Maxillofacial Surgery 121(6) (2020) 721-728, doi: https://doi.org/10.1016/j.jormas.2020.05.013.
[206] G. Dot, F. Rafflenbeul, M. Arbotto, L. Gajny, P. Rouch, T. Schouman, Accuracy and reliability of automatic three-dimensional cephalometric landmarking, Int J Oral Maxillofac Surg 49(10) (2020) 1367-1378, doi: 10.1016/j.ijom.2020.02.015.
[207] V. Kapralos, A. Koutroulis, E. Irinakis, P. Kouros, K. Lyroudia, I. Pitas, G. Mikrogeorgis, Digital subtraction radiography in detection of vertical root fractures: accuracy evaluation for root canal filling, fracture orientation and width variables. An ex-vivo study, Clin Oral Investig 24(10) (2020) 3671-3681, doi: 10.1007/s00784-020-03245-0.
[208] R. Tanaka, T. Tanaka, A.W.K. Yeung, A. Taguchi, A. Katsumata, M.M. Bornstein, Mandibular Radiomorphometric Indices and Tooth Loss as Predictors for the Risk of Osteoporosis using Panoramic Radiographs, Oral Health Prev Dent 18(1) (2020) 773-782, doi: 10.3290/j.ohpd.a45081.
[209] A. Laishram, K. Thongam, Detection and Classification of Dental Pathologies using Faster-RCNN in Orthopantomogram Radiography Image, 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 2020, pp. 423-428, doi: 10.1109/SPIN48934.2020.9071242.
[210] G.K.L. Rao, N. Mokhtar, Y.H.P. Iskandar, A.C. Srinivasa, Learning Orthodontic Cephalometry through Augmented Reality: A Conceptual Machine Learning Validation Approach, 2018 International Conference on Electrical Engineering and Informatics (ICELTICs), 2018, pp. 133-138, doi: 10.1109/ICELTICS.2018.8548939.
[211] G. Damiani, E. Grossi, E. Berti, R.R.Z. Conic, U. Radhakrishna, A. Pacifico, N.L. Bragazzi, R. Piccinno, D. Linder, Artificial neural networks allow response prediction in squamous cell carcinoma of the scalp treated with radiotherapy, J Eur Acad Dermatol Venereol 34(6) (2020) 1369-1373, doi: 10.1111/jdv.16210.
[212] S. Yoon, T. Choi, M. Odlum, D.A. Mitchell, I.M. Kronish, K.W. Davidson, J. Finkelstein, Machine Learning to Identify Behavioral Determinants of Oral Health in Inner City Older Hispanic Adults, Stud Health Technol Inform 251 (2018) 253-256, PMC6211787
[213] M. Yatabe, J.C. Prieto, M. Styner, H. Zhu, A.C. Ruellas, B. Paniagua, F. Budin, E. Benavides, B. Shoukri, L. Michoud, N. Ribera, L. Cevidanes, 3D superimposition of craniofacial imaging-The utility of multicentre collaborations, Orthod Craniofac Res 22 Suppl 1(Suppl 1) (2019) 213-220, doi: 10.1111/ocr.12281.
[214] M. Hung, E. Lauren, E.S. Hon, W.C. Birmingham, J. Xu, S. Su, S.D. Hon, J. Park, P. Dang, M.S. Lipsky, Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence, J Med Internet Res 22(8) (2020) e22590, doi: 10.2196/22590.
[215] Y. Suhail, M. Upadhyay, A. Chhibber, Kshitiz, Machine Learning for the Diagnosis of Orthodontic Extractions: A Computational Analysis Using Ensemble Learning, Bioengineering (Basel) 7(2) (2020) doi: 10.3390/bioengineering7020055.
[216] N. Mehandru, W.L. Hicks, Jr., A.K. Singh, M.R. Markiewicz, Machine Learning for Identification of Craniomaxillofacial Radiographic Lesions, J Oral Maxillofac Surg 78(12) (2020) 2106-2107, doi: 10.1016/j.joms.2020.07.010.
[217] D.B. Tom Vermeire, Sofie Goethals, Raphael Mazzine Barbossa de Oliveira, David Martens Explainable image classification with evidence counterfactual, Pattern Anal Applic (2022) doi: 10.1007/s10044-021-01055-y.

[218] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, IEEE J Biomed Health Inform 22(5) (2018) 1589-1604, doi: 10.1109/JBHI.2017.2767063.
[219] N.K. Singh, K. Raza, Progress in deep learning-based dental and maxillofacial image analysis: A systematic review, Expert Systems with Applications 199 (2022) 116968, doi: https://doi.org/10.1016/j.eswa.2022.116968.
[220] K. Hung, A.W.K. Yeung, R. Tanaka, M.M. Bornstein, Current Applications, Opportunities, and Limitations of AI for 3D Imaging in Dental Research and Practice, Int J Environ Res Public Health 17(12) (2020) doi: 10.3390/ijerph17124424.
[221] L.T. Reyes, J.K. Knorst, F.R. Ortiz, T.M. Ardenghi, Scope and challenges of machine learning-based diagnosis and prognosis in clinical dentistry: A literature review, J Clin Transl Res 7(4) (2021) 523-539, PMC8445629
[222] A. Ke, W. Ellsworth, O. Banerjee, A.Y. Ng, P. Rajpurkar, CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-ray interpretation, Proceedings of the Conference on Health, Inference, and Learning, ACM, 2021, doi: 10.1145/3450439.3451867.
[223] N.P.T. Innes, F. Schwendicke, Restorative Thresholds for Carious Lesions: Systematic Review and Meta-analysis, J Dent Res 96(5) (2017) 501-508, doi: 10.1177/0022034517693605.
[224] S. Corbella, S. Srinivas, F. Cabitza, Applications of deep learning in dentistry, Oral Surg Oral Med Oral Pathol Oral Radiol 132(2) (2021) 225-238, doi: 10.1016/j.oooo.2020.11.003.
[225] F. Schwendicke, M. Tzschoppe, S. Paris, Radiographic caries detection: A systematic review and meta-analysis, J Dent 43(8) (2015) 924-33, doi: 10.1016/j.jdent.2015.02.009.
[226] J. Egger, C. Gsaxner, A. Pepe, K.L. Pomykala, F. Jonske, M. Kurz, J. Li, J. Kleesiek, Medical deep learning-A systematic meta-review, Comput Methods Programs Biomed 221 (2022) 106874, doi: 10.1016/j.cmpb.2022.106874.
[227] R. Rischke, L. Schneider, K. Müller, W. Samek, F. Schwendicke, J. Krois, Federated Learning in Dentistry: Chances and Challenges, Journal of Dental Research 101 (2022) 002203452211089, doi: 10.1177/00220345221108953.
[228] L.M. Stevens, B.J. Mortazavi, R.C. Deo, L. Curtis, D.P. Kao, Recommendations for Reporting Machine Learning Analyses in Clinical Research, Circ Cardiovasc Qual Outcomes 13(10) (2020) e006556, doi: 10.1161/CIRCOUTCOMES.120.006556.
[229] ITU/WHO, Focus Group on "Artificial Intelligence for Health", 2018. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx. (Accessed 31/10/2022 2022).
[230] K.K. Bressem, L.C. Adams, C. Erxleben, B. Hamm, S.M. Niehues, J.L. Vahldiek, Comparing different deep learning architectures for classification of chest radiographs, Sci Rep 10(1) (2020) 13590, doi: 10.1038/s41598-020-70479-z.
[231] Y. Wu, M. Schuster, Z. Chen, Q.V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J.R. Smith, J. Riesa, A. Rudnick, O. Vinyals, G.S. Corrado, M. Hughes, J. Dean, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, ArXiv abs/1609.08144 (2016).
[232] H. Zhu, Big Data and Artificial Intelligence Modeling for Drug Discovery, Annu Rev Pharmacol Toxicol 60 (2020) 573-589, doi: 10.1146/annurev-pharmtox-010919-023324.

[233] H. Kim-Berman, E. Karl, J. Sherbel, L. Sytek, V. Ramaswamy, Validity and User Experience in an Augmented Reality Virtual Tooth Identification Test, J Dent Educ 83(11) (2019) 1345-1352, doi: 10.21815/JDE.019.139.
[234] J.A. Rodrigues, L.T. Arsiwala-Scheppach, J. Krois, F. Schwendicke, Augmented Vision for Dental Students' Education in Detecting Proximal Carious Lesions on Bitewing Radiographs: A Randomized Controlled Trial, Caries Res 56(3) (2022) 197-205, doi: 10.1159/000525777.

Statutory Declaration

“I, Lubaina T. Arsiwala-Scheppach, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic Artificial intelligence in dentistry: Scoping review and bridging observed knowledge gaps via a methodological study and a clinical trial; Künstliche Intelligenz in der Zahnheilkunde: Scoping-Review und Schließung beobachteter Wissenslücken durch eine methodische und eine klinische Studie, independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

06.08.2023

Date Signature

Declaration of your own contribution to the publications

I, Lubaina T. Arsiwala-Scheppach, contributed the following to the below listed publications:

Publication no. 1: L.T. Arsiwala-Scheppach; A. Chaurasia; A. Muller; J. Krois; F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, Journal of Clinical Medicine, 2023

My contributions to this publication are described ahead. First, I conducted a literature survey of the existing studies to formally identify the research gaps in the field of machine in dentistry. Second, I identified the appropriate instruments and checklists to be used for this scoping review, i.e., the PRISMA checklist, QUADAS-2 tool, and TRIPOD tool. Third, the individual studies to be included in the scoping review were identified by me in accordance with the PRISMA checklist. I reviewed the 183 studies screened for inclusion, identified 15 studies to be excluded, and recorded their reasons for exclusion. Fourth, all the steps undertaken in the scoping review were documented by me to ensure validity and reproducibility of the results. Fifth, I performed the data extraction of the resulting 168 studies and then summarized them accordingly. I was in charge of interpreting and evaluating the extracted data. Sixth, the data was reviewed, validated, and curated in a consistent way by me such that it was comprehensible for all co-authors. Additionally, I contributed to the adjudication process in case of disagreements with respect to the extracted data. Seventh, I evaluated all 168 studies for risk of bias using the QUADAS-2 tool and for adherence to reporting guidelines using the TRIPOD tool. Eighth, all the formal qualitative and quantitative analyses were conducted by me. To present the results of the scoping review, I created all the tables, figures, and supplementary material for this publication, namely, Tables 1, 2, S1, S2, S3, and S4 and Figures 1, 2, and S1. Ninth, I presented this study at the 2022 Pan European Region Oral Health Research Congress held by the International Association for Dental Research and incorporated the feedback that I received into the manuscript. Tenth, the original draft of the manuscript was written independently by me. Furthermore, I revised it according to the suggestions from my co-authors. Last, I submitted the final version of the manuscript according to the journal’s requirements and, with feedback from the senior author, addressed the reviewers’ comments to their satisfaction during the peer-review process.

Publication no. 2: L. Schneider; L.T. Arsiwala-Scheppach; J. Krois; H. Meyer-Lueckel; K.K. Bressem; S.M. Niehues; F. Schwendicke, Benchmarking Deep Learning Models for Tooth Structure Segmentation, Journal of Dental Research, 2022

My contributions to this publication are listed as follows. First, I advised on and formulated the statistical analysis plan for this study, i.e., evaluate the data distributions and thus identify the appropriate statistical tests. Additionally, all the formal statistical analyses for this publication were conducted by me and I was responsible for interpreting the results. Figures 3 and 4 were created on the basis of my statistical analyses, which show the primary results of the publication, i.e., which model configurations performed better than their counterparts. Second, I supported the creation of machine learning models in collaboration with the department's data scientists by providing clinical insights for the study. Third, the section titled ‘Statistical Analysis’ in the original draft of the manuscript was written by me. Furthermore, I critically revised the entire manuscript. Last, I assisted in addressing the journal reviewers’ comments during the peer-review process.

Publication no. 3: S. Mertens; J. Krois; A.G. Cantu; L.T. Arsiwala; F. Schwendicke, Artificial intelligence for caries detection: Randomized trial; Journal of Dentistry, 2021

My contributions to this publication include, first, performing extensive data cleaning to identify and eliminate errors in the raw dataset. The cleaning of the data took place in multiple iterations. Then I proceeded to perform data management or wrangling in order to bring the dataset to an analyzable format according to the statistical software to be implemented. Second, I advised on and conducted all the formal statistical analyses for this publication. Tables 1 and 2, and Figures 2 and S1 were created on the basis of my statistical analyses, which show the primary results of the publication, i.e., how the performance and treatment decisions of the dentists differed by the use of a machine learning-based caries detection software. Third, the sections titled ‘Outcomes’, ‘Sample size’, ‘Statistical methods’ and ‘Results’ in the original draft of the manuscript were written by me. Also, I critically revised the entire manuscript. Last, I assisted in addressing the journal reviewers’ comments during the peer-review process.

Signature, date and stamp of first supervising university professor / lecturer

Signature of doctoral candidate

Printing copies of the publications

Journal of
Clinical Medicine

Review

Machine Learning in Dentistry: A Scoping Review

Lubaina T. Arsiwala-Scheppach [1,2,*], Akhilanand Chaurasia [2,3], Anne Müller [4], Joachim Krois [1,2] and Falk Schwendicke [1,2]

[1] Department of Oral Diagnostics, Digital Health and Health Services Research, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 14197 Berlin, Germany
[2] ITU/WHO Focus Group AI on Health, Topic Group Dental Diagnostics and Digital Dentistry, CH-1211 Geneva 20, Switzerland
[3] Department of Oral Medicine and Radiology, King George's Medical University, Lucknow 226003, India
[4] Pharmacovigilance Institute (Pharmakovigilanz- und Beratungszentrum, PVZ) for Embryotoxicology, Institute of Clinical Pharmacology and Toxicology, Charité—Universitätsmedizin Berlin, 13353 Berlin, Germany
* Correspondence: lubaina.arsiwala@charite.de

Abstract: Machine learning (ML) is being increasingly employed in dental research and application. We aimed to systematically compile studies using ML in dentistry and assess their methodological quality, including the risk of bias and reporting standards. We evaluated studies employing ML in dentistry published from 1 January 2015 to 31 May 2021 on MEDLINE, IEEE Xplore, and arXiv. We assessed publication trends and the distribution of ML tasks (classification, object detection, semantic segmentation, instance segmentation, and generation) in different clinical fields. We appraised the risk of bias and adherence to reporting standards, using the QUADAS-2 and TRIPOD checklists, respectively. Out of 183 identified studies, 168 were included, focusing on various ML tasks and employing a broad range of ML models, input data, data sources, strategies to generate reference tests, and performance metrics. Classification tasks were most common. Forty-two different metrics were used to evaluate model performances, with accuracy, sensitivity, precision, and intersection-over-union being the most common. We observed considerable risk of bias and moderate adherence to reporting standards which hampers replication of results. A minimum (core) set of outcome and outcome metrics is necessary to facilitate comparisons across studies.

Keywords: dental radiography; dentistry; machine learning; neural networks; scoping review

Citation: Arsiwala-Scheppach, L.T.; Chaurasia, A.; Müller, A.; Krois, J.; Schwendicke, F. Machine Learning in Dentistry: A Scoping Review. *J. Clin. Med.* **2023**, *12*, 937. https://doi.org/10.3390/jcm12030937

Academic Editor: Paul R. Cooper

Received: 10 December 2022
Revised: 6 January 2023
Accepted: 23 January 2023
Published: 25 January 2023

1. Introduction

With the advent of the big data era, machine learning (ML) methods like Support Vector Machine, Naïve Bayesian Classifier, Decision Tree, Random Forest (RF), K-Nearest Neighbor, and Deep Learning involving Convolutional Neural Network (CNN), etc., have been increasingly adopted in fields such as finance, spatial sciences, and speech recognition [1]. Additionally, in medicine and dentistry, ML has been employed for a range of applications, for example, image analysis in dermatology, ophthalmology, or radiology, with accuracy values similar or better than that of experienced clinicians [1,2].

In the field of ML, mathematical models are employed to enable computers to learn inherent structures in data and to use the learned understanding for predicting on new, unseen data [3]. For deep learning models, specifically CNNs, different types of model 'architecture' can be used. A ML workflow involves training the model, where a subset of the data is used to learn the underlying statistical patterns in the data, and testing it on a yet unseen, testing data subset. ML models tend to become more accurate, when larger training datasets are used [4]. Moreover, basic learning parameters are usually optimized on a separate data subset, referred to as validation data, a process called hyperparameter tuning. Testing the model on the test data involves a wealth of performance metrics (accuracy, sensitivity also known as recall, specificity, and F-scores, among others), while the assessment of a model's generalizability, achievable via assessing its performance on an external (independent) dataset, is not frequently performed yet.

Notably, studies in the field of dental ML can vary widely [1]. Different research questions translate into different ML tasks, which in turn necessitate different model specifications. Various input data (numerical, imagery, speech, etc.) can be employed and varied models (SVM, Extreme Learning Machine, Decision Tree, RF, K-Nearest Neighbor, Neural Network, etc.) can be used. Datasets of different sizes and partitions (training, testing, and validation sets) can be used, and a range of methods for balancing the input datasets via synthetic data generation can be conducted. Moreover, the reference test can be established either by having a "hard" ground truth (for example, for imagery, histological sectioning) or fuzzy labeling schemes (for example, multiple human annotators labeling the same image), and a variety of performance metrics can be used to evaluate the model's performance. These metrics differ with the ML task (classification or, for imagery, detection of objects, or segmentation of specific pixels in an image, or even generation of new images), and can be determined on different hierarchical levels, e.g., patient level, image level, tooth level, surface level or pixel level. Exemplary metrics are accuracy, the confusion matrix and (associated with it) sensitivity (also known as recall), specificity, positive predictive value (precision), and negative predictive value as well as the area-under-the receiver-operating-characteristics curve (c-statistic). For image segmentation tasks (where each pixel has its own classification accuracy), the intersection-over-union (IoU), i.e., the overlap between labeled and predicted pixels (DICE coefficient or Jaccard index), is often used.

As a result, there is significant heterogeneity in the data, tasks, models, and performance metrics, which makes it difficult to contrast studies and assess the robustness and consistency of the emerging body of evidence for ML in dentistry. Additionally, the quality of ML studies—both with regards to the risk of bias but also the reporting of the methods and results—has been shown to vary [5], and with a high likelihood such variance in quality and replicability is also present for dental ML studies.

We aimed to assess this quality of recent ML studies in dentistry, focusing on risk of bias and reporting quality, and to characterize the overall body of evidence with regards to the clinical and ML tasks frequently studied, the model types and underlying datasets, and the employed metrics. Having an overview about these aspects and appraising the consistency and robustness of existing ML studies in our field facilitates to highlight current strengths and weaknesses, and to identify future research needs. In comparison with recent focused reviews on certain clinical tasks (e.g., caries detection on radiographs [6], cephalometric landmark detection [2], etc.), this scoping review not only mainly targets clinical applicability and performance in a subfield of dentistry, but captures the overall picture of ML in our field with a broader focus, and thus a higher number of studies are expected to be included.

2. Materials and Methods

2.1. Search Strategy and Selection Criteria

We screened three electronic databases (MEDLINE via PubMed, Institute of Electrical and Electronics Engineers (IEEE) Xplore, and arXiv). Search terms used were 'deep learning', 'artificial intelligence', 'machine learning', 'convolutional neural network', 'dental' and 'teeth'. The search strategy for all the three databases used is specified in the Supplementary Materials. No language restrictions were applied. The search was overall designed to account for different publication cultures across disciplines. Reviews, editorials, and technical standards were excluded.

The following inclusion criteria were applied:

(1) Studies which had a dental/oral focus, including technical papers.
(2) Studies employing ML, for example, SVM, RF, Artificial Neural Network, CNN.

(3) Studies published between 1 January 2015 and 31 May 2021, as we aimed to gather recent studies and specifically include deep learning as the most rapidly evolving ML field at present.

Reporting of this scoping review followed the PRISMA checklist [7,8]. Our PICO question was as follows: Which ML practices are being employed by studies in dentistry and what are the methodological quality and findings? The question was constructed according to the Participants Intervention Comparison Outcome and Study (PICOS) strategy.

- Population: All types of data with a dental or oral component.
- Intervention/Comparison: ML techniques applied with a dental or oral focus for the diagnosis, management, prognosis of dental conditions or improving data quality. Patient-level, tooth-level, surface-level, or pixel-level.
- Outcome: Performance evaluation of the ML models in terms of metrics, for example, accuracy, IoU, sensitivity, precision, area under the receiver operating characteristic, F indices, specificity, negative predictive value, rank-N recognition rate, error estimates, correlation coefficients, etc.
- Study design type: For this review, we considered all kinds of studies except reviews, editorials, and technical standards, with no language restrictions.

Ethics approval was not sought because this study was based exclusively on published literature.

Screening of titles or abstracts was performed by one reviewer (A.C.). Inclusion or exclusion was decided by two reviewers in consensus (F.S. and A.C.). All papers which were found to be potentially eligible were assessed in full text against the inclusion criteria. We did not limit the inclusion of studies based on the target study population, outcome of interest, or the context in which ML was used. All original studies related to dentistry and ML, without gross reporting fallacies, such as failure to define the type of ML used, failure to minimally describe which dataset was employed for training and testing, and failure to report study findings, were included in this scoping review.

2.2. Data Collection, Items, and Pre-Processing

Data extraction was performed jointly by A.C., A.M., and L.T.A.-S. The extracted data was reviewed by L.T.A.-S. Adjudication in case of any disagreement was performed by discussion (L.T.A.-S. and J.K.). A pretested Excel spreadsheet was used to record the extracted data. Study characteristics included country, year of publication, aim of study and clinical field, type of input data (covariates or imagery [photographs or radiographs; 2-D or 3-D imagery]), dataset source, size and partitions (training, test, validation sets), type of model used and, for deep learning, architecture, augmentation strategies employed, reference test and its definition, comparators (if available, e.g., current standard of care, clinicians, etc.), and performance metrics and their values. In each study, all data items that were compatible with a domain of the extracted data were sought and recorded (e.g., all performance metrics, models employed). No assumptions were made regarding missing or unclear data.

2.3. Quality Assessment

The risk of bias was assessed using the QUADAS-2 tool in four domains [9]. First, risk of bias in data selection was assessed using the parameters of 'inappropriate exclusions', 'case-control design', and 'consecutive or random patient enrollment'. Second, risk of bias in the index test was assessed using the parameters of 'assessment independent of reference standard and 'pre-specification of thresholds used'. Third, risk of bias in the reference standard was assessed using the parameters of 'validity of reference standard and 'assessment independent of index test'. Fourth, risk of bias in the flow and timing was assessed using the parameters of 'appropriate interval between index test and reference standard', 'use of a reference standard for all patients', 'use of the same reference standard for all patients', and 'inclusion of all patients in the analysis'. Using the same tool, applicability concerns in three domains were also evaluated. First, applicability concerns for data selection were assessed using the parameter of 'mismatch between the included patients and the review question'. Second, applicability concerns for the index test were assessed via the parameter of 'mismatch between the test, its conduct, or its interpretation and the review question'. Last, applicability concerns for the reference standard were assessed via the parameter of 'mismatch between the target condition as defined by the reference standard and the review question'. We note that alternatively (or even complimentary), the PROBAST tool [10] could have been used for the same assessment.

Adherence to reporting standards was assessed using the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) tool, which is a 22-item checklist that provides reporting standards for prediction model studies [11]. Note that not all studies included were prediction model studies (studies varied widely in their broader approach, as discussed below), but all involved a mathematical model (ML) for a specific task, which is why we assumed that this checklist would require most studies to adhere to the large majority of domains. TRIPOD has been used for similar purposes in other domains [5]. Risk of bias and adherence to reporting standards were independently assessed by one reviewer (L.T.A.-S.).

2.4. Data Synthesis

We describe various aspects of the included studies, such as country of origin, type of input data used, source of datasets, type of ML methods used, etc. We had initially attempted to conduct a meta-analysis using the results of the confusion matrices reported by the included studies; however, out of 168 studies, only 16 (10%) studies presented their confusion matrices in a way that could be used for analysis and furthermore. These studies differed from each other in terms of their clinical research question/task, type of input data, model architecture, etc.

Instead, a narrative synthesis was performed, displaying which ML tasks (i.e., classification, object detection, semantic segmentation, instance segmentation, and generation) have been studied in different clinical fields of dentistry namely, restorative dentistry and endodontics, oral medicine, oral radiology, orthodontics, oral surgery and implantology, periodontology, prosthodontics, and others, i.e., non-specific field or general dentistry. We briefly explain the different tasks in the following section:

- In ML, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. An example is to classify a given handwritten character as one of the known characters. Algorithms popularly used for classification in the included studies were logistic regression, k-Nearest Neighbors, Decision Trees, Naïve Bayes, RF, Gradient Boosting, etc.
- In object detection tasks, one attempts to identify and locate objects within an image or video. Specifically, object detection draws bounding boxes around the detected objects, which allow to locate the said objects. Given the complexity of handling image data, deep learning based on CNNs, such as Region-based CNN, Fast Region-based CNN, You Only Look Once, Single Shot multiBox Detection, are popularly used for this task.
- In image segmentation tasks, one aims to identify the exact outline of a detected object in an image. There are two types of segmentation tasks: semantic segmentation and instance segmentation. Semantic segmentation classifies each pixel in the image into a particular class. It does not differentiate between different instances of the same object. For example, if there are two cats in an image, semantic segmentation gives the same label, for instance, 'cat', to all the pixels of both cats. Instance segmentation differs from this in the sense that it gives a unique label to every instance of a particular object in the image. Thus, in the example of an image containing two cats, each cat would receive a distinct label, for instance, 'cat1' and 'cat2'. Currently, the most

popular models for image segmentation are Fully CNNs and their variants like UNet, DeepLab, PointNet, etc.

- A fifth type of a ML task is a generation task, which is not predictive in nature. Such tasks involve the generation of new images from the input images, for example, generation of artifact-free CT images from those containing metal artifacts.

The study protocol was registered after the initial screening stage (PROSPERO registration no. CRD42021288159).

3. Results

3.1. Study Selection and Characteristics

A total of 183 studies were identified and 168 (92%) studies were included (Figure 1). The included studies [3,4,12–177] and their characteristics can be found in Table S1. The excluded studies with reasons for exclusion are listed in Table S2. The included studies were published between 1 January 2015 and 31 May 2021 (median: 2019), with the number of published studies increasing each year; 2015: six studies, 2016: four studies, 2017: 13 studies, 2018: 21 studies, 2019: 49 studies, 2020: 68 studies (for 2021, data only until May was available). The included studies stemmed from 40 countries (Figure S1) and used different kinds of input data, such as 2-D data (radiographs: 42% studies, photographs, or other kinds of images: 16% studies), 3-D data (radiographic scans: 18% studies, non-radiographic scans: 4% studies), non-image data (survey data: 10% studies, single nucleotide polymorphism sequences: 1% studies), and combinations of the aforementioned types of data (9% studies). Further, 97% studies used data from universities, hospitals, and private practices, whereas 1% studies each used data from the National Health and Nutrition Examination Survey, M3BE database, 2013 Nationwide Readmissions Database of the USA, and the National Institute of Dental and Craniofacial Research dataset.

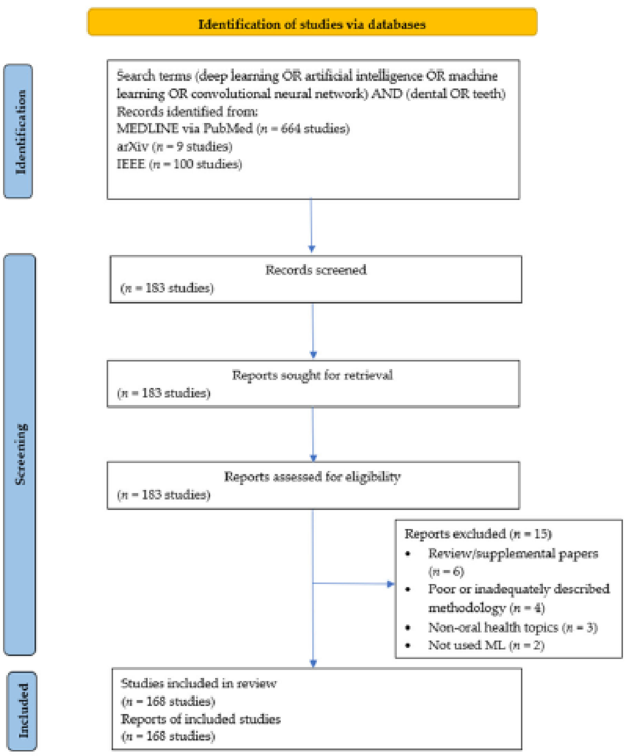

Figure 1. PRISMA study flow diagram.

Additionally, 85% studies partitioned their total dataset into training and testing data subsets, and 59% studies also created validation data subsets from the same data source. The median size of the training datasets was 450 (range: 12 to 1,296,000 data instances) and of the test datasets was 126 (range: 1 to 144,000). Nearly half of the studies tested model performance on a hold-out test dataset while the remaining used cross-validation. Cross-validation is a resampling method that uses different portions of the data to test and train a model during each iteration. For example, in a 10-fold cross-validation, the original dataset is randomly partitioned into 10 subsamples, out of which nine subsamples are used as training data and one subsample as the test data. Ten iterations of the following step are carried out; the model is trained on the nine subsamples designated as training data and tested on the one subsample of test data; but in each iteration, a different subsample is chosen to serve as the test data and thus a different combination of subsamples constitutes the training data. Eventually, the final estimation of model performance is the average of these results.

In addition, 65% studies augmented their input data, mainly the training data, but a few augmented the testing data, too. Only 20% studies used an external dataset to validate their model's performance. The reference test (i.e., how the ground truth was defined) was established by professional experts in 73% studies: one expert in 18% studies, two experts in 11% studies, three experts in 10% studies, four and five experts in 2% studies each, six experts in 1% studies, and seven, eight, 12, and 20 experts in 0.5% studies, each. Another 27% studies used experts for establishing the reference test but did not provide details on the exact numbers. Additionally, 22% studies used information from their datasets as the reference test (for example, age, diagnosis from medical records) and 1% studies used a software tool to generate the reference test. The remaining 4% studies did not provide details on how the reference test was established.

Of all studies, 70% used deep learning models; CNN as classifiers: 59 studies, CNN for other tasks: 14 studies, Faster R-CNN: seven studies, fully CNN: 19 studies, Mask R-CNN: seven studies, 3-D CNN: three studies, adaptive CNN and pulse-coupled CNN: one study each, and non-convolutional deep neural networks: seven studies (Table S1). Another 22% studies used non-deep learning models; perceptron: four studies, other neural networks: three studies, other types of models, such as, fuzzy classifier, SVM, RF, etc.: 30 studies. In addition, 6% studies used various combinations of the aforementioned models and 2% studies did not provide details of the model architecture employed. Both, models using and not using deep learning were employed in higher proportions by studies in restorative dentistry and endodontics, oral medicine, and non-specific field or general dentistry (Table S3). Additionally, models not using deep learning were frequently employed by studies in orthodontics and periodontology. Finally, 20% studies compared their model's performance with that of human comparators.

3.2. Risk of Bias and Applicability Concerns

The risk of bias was assessed in four domains, namely data selection, index test, reference standard, and flow and timing. It was found to be high for 54% of the studies regarding data selection and for 58% of the studies regarding the reference standard (Table 1). On the other hand, the risk of bias was low for the majority of studies regarding the index test (77%) and flow and timing (89%). Applicability concerns were found to be high for 53% of the studies regarding data selection but were low for most studies regarding the index test (79%) and reference standard (73%).

Table 1. Evaluation of risk of bias in studies included (n = 168) using the QUADAS-2 tool.

Sr. No. [Citation]	Data Selection: risk of Bias/Applicability Concerns	Index Test: Risk of Bias/Applicability Concerns	Reference Standard: Risk of Bias/Applicability Concerns	Flow and Timing: Risk of Bias
1. [12]	high/high	low/high	high/high	low
2. [13]	low/low	low/low	low/low	low
3. [14]	high/low	low/low	low/low	low
4. [15]	low/low	low/high	high/high	low
5. [16]	low/low	low/low	low/low	low
6. [17]	high/high	low/high	high/high	low
7. [18]	high/high	low/low	high/low	low
8. [19]	low/low	low/high	low/low	low
9. [20]	low/low	low/low	low/high	low
10. [21]	high/high	low/low	high/low	low
11. [22]	high/high	low/low	high/high	low
12. [23]	high/low	high/low	high/low	low
13. [24]	low/high	low/low	high/high	low
14. [25]	high/high	high/low	low/low	low
15. [26]	low/low	high/low	low/low	low
16. [27]	high/low	low/low	high/low	low
17. [28]	high/high	low/low	high/low	low
18. [29]	high/low	low/low	high/low	low
19. [30]	high/high	low/low	high/low	low
20. [31]	high/high	low/high	high/low	low
21. [32]	high/high	high/high	high/high	low
22. [33]	low/low	low/low	low/low	low
23. [34]	low/high	low/low	low/high	low
24. [35]	high/high	low/low	low/low	low
25. [36]	low/low	low/low	low/low	low
26. [37]	high/high	low/low	high/low	low
27. [38]	high/high	low/low	high/low	low
28. [39]	high/high	low/low	high/low	low
29. [40]	high/high	high/low	high/low	low
30. [41]	low/low	low/low	low/low	low
31. [42]	high/low	high/low	low/low	low
32. [43]	low/high	low/high	low/high	low
33. [44]	low/low	high/low	high/low	low
34. [45]	high/high	low/high	low/high	low
35. [46]	high/low	low/low	low/low	low
36. [47]	high/high	low/low	low/low	low
37. [48]	high/high	low/high	low/high	low
38. [49]	low/low	low/low	high/low	low
39. [50]	low/high	low/low	high/low	high
40. [51]	low/high	low/low	low/low	low
41. [52]	high/low	low/high	high/low	low
42. [53]	high/high	low/low	low/low	high
43. [54]	low/low	low/high	low/high	low
44. [55]	high/high	low/low	high/low	low

45. [56]	high/high	low/high	high/low	low
46. [57]	high/high	low/low	high/high	low
47. [58]	high/high	high/high	high/high	low
48. [59]	low/high	low/low	high/high	low
49. [60]	low/high	low/low	high/high	low
50. [61]	low/low	low/low	high/low	high
51. [62]	high/high	low/low	high/low	low
52. [63]	low/high	low/high	high/high	low
53. [64]	high/high	high/high	high/high	low
54. [65]	high/high	low/low	high/low	low
55. [66]	low/high	low/low	high/low	low
56. [67]	high/high	low/high	low/high	low
57. [68]	low/high	low/low	low/low	high
58. [69]	low/low	low/low	low/low	low
59. [70]	high/high	low/low	low/low	low
60. [71]	low/low	low/low	low/low	low
61. [72]	low/high	low/low	high/low	low
62. [73]	low/low	low/low	high/low	low
63. [74]	low/low	low/low	low/low	low
64. [75]	low/low	low/low	low/low	low
65. [76]	low/low	low/low	low/low	low
66. [77]	high/high	high/low	high/low	low
67. [78]	high/low	high/low	high/low	low
68. [79]	high/low	high/low	high/low	low
69. [80]	high/low	high/low	low/low	low
70. [81]	low/low	low/low	low/low	low
71. [82]	low/low	low/low	high/low	low
72. [83]	low/low	low/low	low/low	low
73. [84]	high/low	low/low	high/low	low
74. [85]	low/low	low/low	low/low	high
75. [86]	high/high	low/low	low/low	low
76. [87]	high/high	high/low	low/low	low
77. [88]	low/low	low/low	low/low	low
78. [89]	high/high	high/high	high/high	low
79. [90]	high/high	high/high	high/high	low
80. [91]	high/high	low/low	high/low	low
81. [92]	low/low	low/low	high/low	low
82. [93]	low/high	low/low	high/high	low
83. [94]	low/low	low/low	low/low	high
84. [95]	high/high	high/low	high/high	low
85. [96]	low/high	high/low	high/high	low
86. [97]	high/high	low/high	low/high	low
87. [98]	high/high	low/low	low/low	low
88. [99]	low/high	low/high	high/high	low
89. [100]	low/high	low/high	high/high	low
90. [101]	low/high	low/low	low/high	low
91. [102]	high/high	low/low	high/low	low
92. [103]	low/low	low/low	low/low	low
93. [4]	high/low	low/high	high/high	low

94. [104]	low/low	low/low	high/low	low
95. [105]	high/high	low/high	high/low	low
96. [106]	low/high	low/low	low/high	low
97. [107]	low/low	low/low	high/low	low
98. [108]	low/low	low/low	low/low	low
99. [109]	high/high	high/low	high/low	low
100. [110]	low/low	low/low	high/low	low
101. [111]	low/low	low/low	high/low	low
102. [112]	high/low	high/low	high/high	low
103. [113]	high/high	low/low	low/high	high
104. [3]	low/high	low/low	low/low	low
105. [114]	low/low	low/low	low/low	low
106. [115]	low/low	low/low	low/low	low
107. [116]	high/high	high/low	high/low	low
108. [117]	high/low	high/low	low/low	low
109. [118]	high/high	low/low	high/low	low
110. [119]	low/low	low/low	low/low	low
111. [120]	low/low	low/high	high/high	low
112. [121]	low/low	low/low	high/low	low
113. [122]	high/high	high/low	low/low	low
114. [123]	low/low	low/low	low/low	low
115. [124]	low/high	low/low	high/low	low
116. [125]	high/high	low/low	low/high	low
117. [126]	high/low	high/low	high/low	high
118. [127]	high/high	low/low	high/low	low
119. [128]	low/low	high/low	low/low	low
120. [129]	high/low	low/low	low/low	low
121. [130]	high/high	low/low	high/low	high
122. [131]	high/low	high/low	high/low	low
123. [132]	high/high	low/low	high/low	low
124. [133]	high/high	low/low	high/low	high
125. [134]	low/high	high/low	high/low	low
126. [135]	high/low	high/low	low/low	low
127. [136]	high/low	high/low	high/low	low
128. [137]	high/low	high/high	low/low	low
129. [138]	low/high	low/high	high/low	low
130. [139]	high/low	low/low	low/low	low
131. [140]	high/low	low/high	high/high	low
132. [141]	low/low	low/low	high/low	low
133. [142]	high/high	low/low	high/low	low
134. [143]	high/high	low/low	low/low	low
135. [144]	high/high	low/low	high/low	low
136. [145]	high/high	high/low	high/low	low
137. [146]	high/high	low/low	high/low	low
138. [147]	high/low	high/low	low/low	low
139. [148]	high/high	low/low	high/low	low
140. [149]	high/high	low/high	high/high	low
141. [150]	high/high	low/high	high/high	low
142. [151]	low/high	low/low	high/high	low

143. [152]	high/high	low/high	high/high	low
144. [153]	high/low	low/low	high/low	low
145. [154]	low/low	low/high	high/high	low
146. [155]	low/low	high/low	low/low	low
147. [156]	low/high	low/low	low/low	low
148. [157]	high/high	high/low	high/low	high
149. [158]	low/low	low/low	low/low	low
150. [159]	low/high	low/high	low/high	low
151. [160]	high/low	low/high	low/low	low
152. [161]	low/low	high/low	high/low	high
153. [162]	high/low	low/low	low/high	low
154. [163]	low/low	low/high	low/high	low
155. [164]	high/low	low/low	high/low	low
156. [165]	low/low	low/low	high/low	low
157. [166]	low/high	low/high	high/high	high
158. [167]	low/low	low/low	low/low	low
159. [168]	low/low	low/low	high/low	low
160. [169]	low/high	high/low	high/high	low
161. [170]	high/high	low/low	low/low	low
162. [171]	low/low	low/low	high/low	low
163. [172]	low/low	low/low	low/low	low
164. [173]	low/low	low/low	high/low	low
165. [174]	low/low	low/low	low/low	low
166. [175]	high/high	high/high	high/high	low
167. [176]	high/high	high/low	low/low	low
168. [177]	high/high	low/low	high/low	low

3.3. Adherence to Reporting Standards

Overall adherence to the TRIPOD reporting checklist was 33.3%, with 18/22 domains having an adherence rate less than 50% (Figure 2). Reporting adherence was at or above 80% for background and objectives, and potential clinical use of the model and implications for future research, but below 10% for sample size calculation, handling of missing data, differences between development and validation data, and details on participants. In particular, less than 20% of studies adequately defined their predictors and outcomes (in terms of their blinded assessments), stratification into risk groups, presented the full prediction model and provided information on supplementary resources, such as study protocol, web calculator, or data sets. Less than 40% of the studies adequately reported about their data sources (i.e., study dates), participant eligibility, statistical methods (specifically, details on model refinement), model results (in terms of results from crude models), study limitations, and results with reference to performance in the development data, and any other validation data.

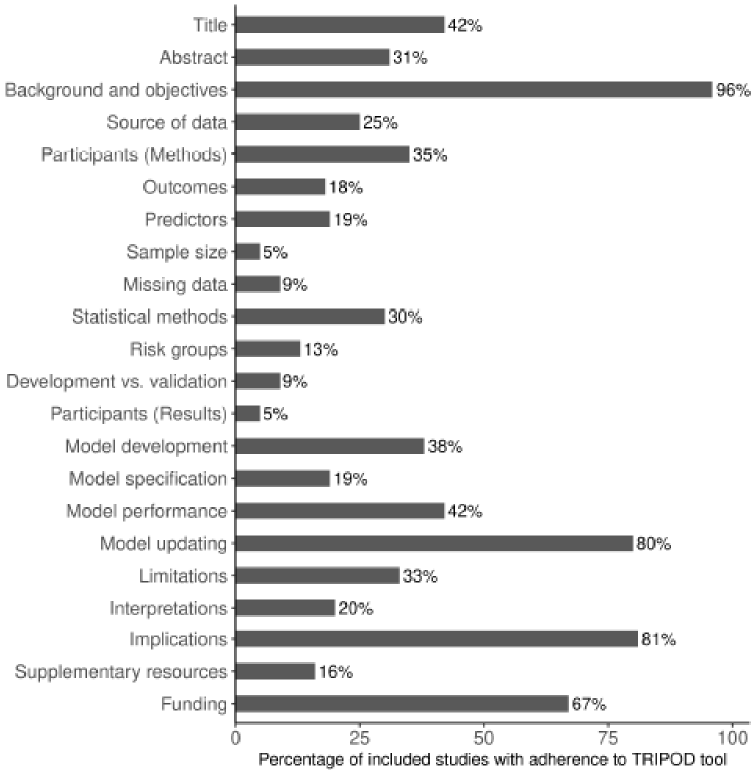

Figure 2. Reporting adherence of studies (n = 168) to Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) tool.

3.4. Tasks, Metrics, and Findings of the Studies

Based on the nature of the ML task formulated, the 168 included studies could be classified into five major categories of ML tasks; classification task, n = 85; object detection task, n = 22; semantic segmentation task, n = 37; instance segmentation task, n = 19; and generation task, n = 5. Classification tasks were most commonly used in oral medicine studies (22%), whereas object detection, semantic segmentation, and instance segmentation tasks, each were most commonly used in non-specific field or general dentistry studies (36%, 38%, and 58%, respectively), Table 2. Generation tasks, though small in number, were most commonly used in oral radiology studies (80%).

Table 2. Number of studies in each field of dentistry, stratified by type of machine learning task (n = 168).

	Classification Task	Object Detection Task	Semantic Segmentation Task	Instance Segmentation Task	Generation Task
n	85	22	37	19	5
Field of dentistry, n (%)					
Restorative dentistry and endodontics	13 (15%)	1 (4%)	9 (24%)	2 (11%)	0 (0%)
Oral medicine	19 (22%)	5 (23%)	1 (3%)	0 (0%)	0 (0%)
Oral radiology	3 (4%)	0 (0%)	2 (5%)	2 (11%)	4 (80%)
Orthodontics	10 (12%)	3 (14%)	1 (3%)	3 (15%)	1 (20%)
Oral surgery and implantology	11 (13%)	3 (14%)	3 (8%)	0 (0%)	0 (0%)

Periodontology	9 (11%)	2 (9%)	7 (19%)	1 (5%)	0 (0%)
Prosthodontics	2 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Others (non-specific field, general dentistry)	18 (21%)	8 (36%)	14 (38%)	11 (58%)	0 (0%)

A total of 42 different metrics were used by the studies to evaluate model performance and some of these could be grouped into one class, for example, the various correlation coefficients could be combined. Such grouping (or consolidation) resulted in 26 distinct classes of metrics. Note that most studies reported multiple metrics. Studies on classification tasks commonly reported accuracy, sensitivity, area under the receiver-operating characteristic, specificity, and precision, and those on object detection reported on sensitivity, precision, and accuracy. Studies on semantic segmentation reported on IoU and sensitivity, and those on instance segmentation reported on accuracy, sensitivity, and IoU. Lastly, studies using generation tasks commonly reported on peak signal-to-noise ratio, structural similarity index, and relative error. Table S4 shows the number of studies which used the different metrics, stratified by ML task.

After stratifying the studies by ML task and clinical field of dentistry, we attempted to evaluate studies that reported on accuracy, or mean average precision, or IoU. A formal comparison was inhibited by the large variability at the level of clinical or diagnostic tasks amongst the studies.

4. Discussion

ML in dentistry is characterized by the availability of a plethora of clinical tasks which necessitate the use of a wide range of input data types, ML models, performance metrics, etc. This has given rise to a large body of evidence with limited comparability. The present scoping review synthesized this evidence and allowed to comprehensively assess this body. We will begin by discussing our findings in detail.

First, the included studies aimed for different ML tasks on a wide variety of data. These data then differed once more within specific subtypes (e.g., imagery, with radiographs, scans, photographs, each of them being sub-classified again, and differing in resolution, contrast, etc.). Moreover, data usually stemmed from single centers, representing only a limited population (and diversity in terms of data generation strategy or technique), all of which likely adversely impacts generalizability of results. The data used were nearly never available, except for the few studies employing data from open databases, leading to difficulties in replication of results. Researchers are urged to comply with journals' data sharing policies and make their data available upon reasonable request. We acknowledge that there may be data sharing and privacy concerns across institutions and countries. Alternatives to centralized learning of ML models, like federated learning, which do not require data sharing may be of relevance especially for data which are hard to de-identify [178]. Practices of data linkage and triangulation, i.e., using a variety of data sources to create a richer dataset, were almost non-existent. Thus, limiting options for verification of data integrity and increasing the learning output of a ML model by leveraging information from multiple data sources on hierarchical structures and correlations.

Second, a wide range of outcome measures was used by the included studies. These can be measured on different levels, such as patient-level, tooth-level, and surface-level, and while this is relevant for any comparison or synthesis across studies, it was not always reported on what level the outcomes were assessed. Another issue was the high number of performance metrics in use, as evident from our results, leading to only a few studies being comparable to each other. Defining an agreed-upon set of outcome metrics for specific subtasks in ML in dentistry (e.g., classification, detection, segmentation on images) along with standards towards the level of outcome assessment seems warranted. This outcome set should reflect various aspects of performance (e.g., under- and over-detection), consider the impact of prevalence (e.g., predictive values), and attempt to transport not only diagnostic value, but also clinical usefulness. For the latter, studies attempting to assess the value of ML in the hands of clinicians against the current standard of care are needed.

Third, the use of reference tests (i.e., how the ground truth was established) warrants discussion. A wide range of strategies to establish reference tests were employed. In many studies, no details towards the definition of the reference tests were provided. A few studies using image data used only one human annotator as the reference test, a decision which may be criticized given the known wide variability in experts' annotations [2]. Alternative concepts of applying the reference test to training datasets should be employed and compared to gauge the impact of different approaches and validate the one eventually selected. Additionally, testing datasets should be standardized and heterogeneous to ensure class balance and generalizability. One approach is to establish open benchmarking datasets, as attempted by the ITU/WHO Focus Group on Artificial Intelligence for Health [179].

Fourth, the quality of conducting and reporting ML studies in dentistry remains problematic. Notably, the specific risks emanating from ML and the underlying data are insufficiently addressed, e.g., biases, data leakage, or overfitting of the model. Furthermore, many studies suffered from unclear or a lack of validation of their results on external datasets. The evaluation of a model's performance on unseen data is a crucial aspect as it relates to the generalizability of ML models regarding performance on data from other sources. Exploration of why some models were not generalizable was even less common, thus preventing identification of steps required to better the models. Generally, the majority of studies performed application testing, developed models, and showed that ML can learn and, in many studies, predict. Understanding why this is, how it could be improved, what the clinical domain needs, or which safeguards for ML in dentistry are required, was seldom an issue. General reporting did not allow full replication, as many details were not presented, and additionally, the display of the model performance remained, as discussed, insufficient. Researchers need to adhere to the published guidelines on study conduct and reporting [180–182].

In an effort to characterize the emerging pattern in the included studies, first, we would like to elaborate on the nature of clinical tasks employed by the studies. A wide array of research questions were present; from detecting dental artifacts in images to investigating the benefits of transfer-learning, from classifying different dental conditions to aiding in decision-making and assessing cost-effectiveness. Thus, there is evidence of broadening of avenues where ML could be exploited. As stated earlier, classification tasks were the most common and this may be because diagnosing dental structures or anomalies on images is a vital step towards successful treatment outcomes and prognosis. However, over the years, ML methods have improved their classification performance on images at the cost of increased model complexity and opacity [183]. The inability to explain ML's methods and decisions is one of the contributing factors towards development of explainable AI, i.e., a set of processes that allows human users to comprehend and trust the results created by ML algorithms. Second, more recent studies tended to employ image segmentation models [2,25,39,48,59,60,73,151].

The presented scoping review has a few salient features. First, it is the most comprehensive overview on ML in dentistry with 168 studies being included. Second, and as a limitation, we could not include randomized controlled trials because none were available and found the included studies to have a considerable risk of bias, both of which should be considered when interpreting our results. Third, to our knowledge this study is the first to employ TRIPOD for gauging the reporting quality of studies using ML in dentistry. TRIPOD is a checklist designed to assess prediction models which has not been validated specifically for ML applications [5]. However, previous studies have used it to evaluate ML models since the quality assessment criteria for clinical prediction tools and ML models are similar [5]. At present, a TRIPOD-ML tool is under-construction [5]. Fourth, we included studies until May 2021 only, as the systematic critique of the 168 studies required considerable time and effort since then. We acknowledge that inclusion of recently published studies may have strengthened our review. Furthermore, we acknowledge that arXiv, an archiving database, may include studies which did not undergo a formal peer-review process and this may be a limitation for our study. However, studies on arXiv are reviewed by peers in a non-formal process and updated after peer-review. Last, any clinical usability cannot be inferred from this study because it was not the focus of this comprehensive review.

5. Conclusions

In conclusion, we demonstrated that ML has been employed for a large number of tasks in dentistry, building on a wide range of methods and employing highly heterogeneous reporting metrics. As a result, comparisons across studies or benchmarking of the developed ML models are only possible to a limited extent. A minimum (core) set of defined outcomes and outcome metrics would help to overcome this and facilitate comparisons, whenever appropriate. The overall body of evidence showed considerable risk of bias as well as moderate adherence to reporting standards. Researchers are urged to adhere more closely to reporting standards and plan their studies with even greater scientific rigor to reduce any risk of bias. Last, the included studies mainly focused on developing ML models, while presenting their generalizability, robustness, or clinical usefulness was uncommon. Future studies should aim to demonstrate that ML positively impacts the quality and efficiency of healthcare.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcm12030937/s1, Search strategy, Figure S1: Geographical trends in number of publications of machine learning methods in dentistry between 1 January 2015 and 31 May 2021; Table S1: Studies included in the scoping review along with their characteristics (n = 168); Table S2: Studies excluded from the scoping review along with the reason for exclusion (n = 15); Table S3: Number of studies in each field of dentistry, stratified by the machine learning model used (n = 168); Table S4: Number of studies using the various performance metrics stratified by type of machine learning task. References [1,184–197] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, F.S., A.C. and J.K.; methodology, L.T.A.-S., A.C. and A.M.; software, A.C. and L.T.A.-S.; validation, L.T.A.-S. and J.K.; formal analysis, L.T.A.-S.; investigation, A.C.; resources, F.S. and A.C.; data curation, L.T.A.-S. and A.M.; writing—original draft preparation, L.T.A.-S.; writing—review and editing, F.S.; visualization, L.T.A.-S.; supervision, F.S.; project administration, J.K.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge financial support from the Open Access Publication Fund of Charité—Universitätsmedizin Berlin and the German Research Foundation (DFG).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are available through the paper and supplementary material. Additional information is available from the authors upon reasonable request.

Conflicts of Interest: F.S. and J.K. are co-founders of the startup dentalXrai GmbH. dentalXrai GmbH had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Sun, M.-L.; Liu, Y.; Liu, G.-M.; Cui, D.; Heidari, A.A.; Jia, W.-Y.; Ji, X.; Chen, H.-L.; Luo, Y.-G. Application of Machine Learning to Stomatology: A Comprehensive Review. *IEEE Access* **2020**, *8*, 184360–184374. https://doi.org/10.1109/access.2020.3028600.
2. Schwendicke, F.; Chaurasia, A.; Arsiwala, L.; Lee, J.-H.; Elhennawy, K.; Jost-Brinkmann, P.-G.; Demarco, F.; Krois, J. Deep learning for cephalometric landmark detection: Systematic review and meta-analysis. *Clin. Oral Investig.* **2021**, *25*, 4299–4309. https://doi.org/10.1007/s00784-021-03990-w.
3. Farhadian, M.; Shokouhi, P.; Torkzaban, P. A decision support system based on support vector machine for diagnosis of periodontal disease. *BMC Res. Notes* **2020**, *13*, 337. https://doi.org/10.1186/s13104-020-05180-5.

4. Abdalla-Aslan, R.; Yeshua, T.; Kabla, D.; Leichter, I.; Nadler, C. An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2020**, *130*, 593–602. https://doi.org/10.1016/j.oooo.2020.05.012.
5. Ben Li, B.; Feridooni, T.; Cuen-Ojeda, C.; Kishibe, T.; de Mestral, C.; Mamdani, M.; Al-Omran, M. Machine learning in vascular surgery: A systematic review and critical appraisal. *NPJ Digit. Med.* **2022**, *5*, 7. https://doi.org/10.1038/s41746-021-00552-y.
6. Schwendicke, F.; Tzschoppe, M.; Paris, S. Radiographic caries detection: A systematic review and meta-analysis. *J. Dent.* **2015**, *43*, 924–933. https://doi.org/10.1016/j.jdent.2015.02.009.
7. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. https://doi.org/10.1371/journal.pmed.1000097.
8. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. https://doi.org/10.1136/bmj.n71.
9. Whiting, P.F.; Rutjes, A.W.S.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.G.; Sterne, J.A.C.; Bossuyt, P.M.M.; QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
10. Wolff, R.F.; Moons, K.G.; Riley, R.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; for the PROBAST Group. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. https://doi.org/10.7326/m18-1376.
11. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med.* **2015**, *13*, 214. https://doi.org/10.1186/s12916-014-0241-z.
12. Aliaga, I.J.; Vera, V.; De Paz, J.F.; García, A.E.; Mohamad, M.S. Modelling the Longevity of Dental Restorations by means of a CBR System. *BioMed Res. Int.* **2015**, *2015*, 540306. https://doi.org/10.1155/2015/540306.
13. Gupta, A.; Kharbanda, O.P.; Sardana, V.; Balachandran, R.; Sardana, H.K. A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images. *Int. J. Comput. Assist. Radiol. Surg.* **2015**, *10*, 1737–1752. https://doi.org/10.1007/s11548-015-1173-6.
14. Gupta, A.; Kharbanda, O.P.; Sardana, V.; Balachandran, R.; Sardana, H.K. Accuracy of 3D cephalometric measurements based on an automatic knowledge-based landmark detection algorithm. *Int. J. Comput. Assist. Radiol. Surg.* **2016**, *11*, 1297–1309. https://doi.org/10.1007/s11548-015-1334-7.
15. Hadley, A.J.; Krival, K.R.; Ridgel, A.L.; Hahn, E.C.; Tyler, D.J. Neural Network Pattern Recognition of Lingual–Palatal Pressure for Automated Detection of Swallow. *Dysphagia* **2015**, *30*, 176–187. https://doi.org/10.1007/s00455-014-9593-y.
16. Kavitha, M.S.; An, S.-Y.; An, C.-H.; Huh, K.-H.; Yi, W.-J.; Heo, M.-S.; Lee, S.-S.; Choi, S.-C. Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2015**, *119*, 346–356. https://doi.org/10.1016/j.oooo.2014.11.009.
17. Mansoor, A.; Patsekin, V.; Scherl, D.; Robinson, J.P.; Rajwa, B. A Statistical Modeling Approach to Computer-Aided Quantification of Dental Biofilm. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 358–366. https://doi.org/10.1109/jbhi.2014.2310204.
18. Jung, S.-K.; Kim, T.-W. New approach for the diagnosis of extractions with neural network machine learning. *Am. J. Orthod. Dentofac. Orthop.* **2016**, *149*, 127–133. https://doi.org/10.1016/j.ajodo.2015.07.030.
19. Kavitha, M.S.; Kumar, P.G.; Park, S.-Y.; Huh, K.-H.; Heo, M.-S.; Kurita, T.; Asano, A.; An, S.-Y.; Chien, S.-I. Automatic detection of osteoporosis based on hybrid genetic swarm fuzzy classifier approaches. *Dentomaxillofac. Radiol.* **2016**, *45*, 20160076. https://doi.org/10.1259/dmfr.20160076.
20. Mahmoud, Y.E.; Labib, S.S.; Mokhtar, H.M.O. Teeth periapical lesion prediction using machine learning techniques. In Proceedings of the 2016 SAI Computing Conference (SAI), London, UK, 13–15 July 2016; pp. 129–134. https://doi.org/10.1109/sai.2016.7555972.
21. Wang, L.; Li, S.; Chen, R.; Liu, S.-Y.; Chen, J.-C. An Automatic Segmentation and Classification Framework Based on PCNN Model for Single Tooth in MicroCT Images. *PLoS ONE* **2016**, *11*, e0157694. https://doi.org/10.1371/journal.pone.0157694.
22. De Tobel, J.; Radesh, P.; Vandermeulen, D.; Thevissen, P.W. An automated technique to stage lower third molar development on panoramic radiographs for age estimation: A pilot study. *J. Forensic Odonto-Stomatol.* **2017**, *35*, 42–54.
23. Hwang, J.J.; Lee, J.-H.; Han, S.-S.; Kim, Y.H.; Jeong, H.-G.; Choi, Y.J.; Park, W. Strut analysis for osteoporosis detection model using dental panoramic radiography. *Dentomaxillofac. Radiol.* **2017**, *46*, 20170006. https://doi.org/10.1259/dmfr.20170006.
24. Imangaliyev, S.; van der Veen, M.H.; Volgenant, C.; Loos, B.G.; Keijser, B.J.; Crielaard, W.; Levin, E. Classification of quantitative light-induced fluorescence images using convolutional neural network. *arXiv* **2017**, arXiv:1705.09193.
25. Johari, M.; Esmaeili, F.; Andalib, A.; Garjani, S.; Saberkari, H. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: An ex vivo study. *Dentomaxillofac. Radiol.* **2017**, *46*, 20160107–20160107. https://doi.org/10.1259/dmfr.20160107.
26. Liu, Y.; Li, Y.; Fu, Y.; Liu, T.; Liu, X.; Zhang, X.; Fu, J.; Guan, X.; Chen, T.; Chen, X.; et al. Quantitative prediction of oral cancer risk in patients with oral leukoplakia. *Oncotarget* **2017**, *8*, 46057–46064. https://doi.org/10.18632/oncotarget.17550.
27. Miki, Y.; Muramatsu, C.; Hayashi, T.; Zhou, X.; Hara, T.; Katsumata, A.; Fujita, H. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput. Biol. Med.* **2017**, *80*, 24–29. https://doi.org/10.1016/j.compbiomed.2016.11.003.

28. Oktay, A.B. Tooth detection with Convolutional Neural Networks. In Proceedings of the 2017 Medical Technologies National Congress (TIPTEKNO), Trabzon, Turkey, 12–14 October 2017; pp. 1–4. https://doi.org/10.1109/tiptekno.2017.8238075.
29. Prajapati, A.S.; Nagaraj, R.; Mitra, S. Classification of dental diseases using CNN and transfer learning. In Proceedings of the 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, United Arab Emirates, 11–14 August 2017; pp. 70–74.
30. Raith, S.; Vogel, E.P.; Anees, N.; Keul, C.; Güth, J.-F.; Edelhoff, D.; Fischer, H. Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data. *Comput. Biol. Med.* **2017**, *80*, 65–76. https://doi.org/10.1016/j.compbiomed.2016.11.013.
31. Rana, A.; Yauney, G.; Wong, L.C.; Gupta, O.; Muftu, A.; Shah, P. Automated segmentation of gingival diseases from oral images. In Proceedings of the 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), Bethesda, MD, USA, 6–8 November 2017; pp. 144–147. https://doi.org/10.1109/hic.2017.8227605.
32. Srivastava, M.M.; Kumar, P.; Pradhan, L.; Varadarajan, S. Detection of tooth caries in bitewing radiographs using deep learning. *arXiv* **2017**, arXiv:1711.07312.
33. Štepanovský, M.; Ibrová, A.; Buk, Z.; Velemínská, J. Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic Sci. Int.* **2017**, *279*, 72–82. https://doi.org/10.1016/j.forsciint.2017.08.005.
34. Yilmaz, E.; Kayikcioglu, T.; Kayipmaz, S. Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography. *Comput. Methods Programs Biomed.* **2017**, *146*, 91–100. https://doi.org/10.1016/j.cmpb.2017.05.012.
35. Du, X.; Chen, Y.; Zhao, J.; Xi, Y. A Convolutional Neural Network Based Auto-Positioning Method for Dental Arch In Rotational Panoramic Radiography. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 2615–2618. https://doi.org/10.1109/embc.2018.8512732.
36. Egger, J.; Pfarrkirchner, B.; Gsaxner, C.; Lindner, L.; Schmalstieg, D.; Wallner, J. Fully Convolutional Mandible Segmentation on a valid Ground- Truth Dataset. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 656–660. https://doi.org/10.1109/embc.2018.8512458.
37. Fakhriy, N.A.A.; Ardiyanto, I.; Nugroho, H.A.; Pratama, G.N.P. Machine Learning Algorithms for Classifying Abscessed and Impacted Tooth: Comparison Study. In Proceedings of the 2018 2nd International Conference on Biomedical Engineering (IBIOMED), Bali, Indonesia, 24–26 July 2018; pp. 88–93. https://doi.org/10.1109/ibiomed.2018.8534887.
38. Fariza, A.; Arifin, A.Z.; Astuti, E.R. Interactive Segmentation of Conditional Spatial FCM with Gaussian Kernel-Based for Panoramic Radiography. In Proceedings of the 2018 International Symposium on Advanced Intelligent Informatics (SAIN), Yogyakarta, Indonesia, 29–30 August 2018; pp. 157–161. https://doi.org/10.1109/sain.2018.8673381.
39. Gavinho, L.G.; Araujo, S.A.; Bussadori, S.K.; Silva, J.V.P.; Deana, A.M. Detection of white spot lesions by segmenting laser speckle images using computer vision methods. *Lasers Med. Sci.* **2018**, *33*, 1565–1571. https://doi.org/10.1007/s10103-018-2520-y.
40. Ha, S.-R.; Park, H.S.; Kim, E.-H.; Kim, H.-K.; Yang, J.-Y.; Heo, J.; Yeo, I.-S.L. A pilot study using machine learning methods about factors influencing prognosis of dental implants. *J. Adv. Prosthodont.* **2018**, *10*, 395–400. https://doi.org/10.4047/jap.2018.10.6.395.
41. Heinrich, A.; Güttler, F.; Wendt, S.; Schenkl, S.; Hubig, M.; Wagner, R.; Mall, G.; Teichgräber, U. Forensic Odontology: Automatic Identification of Persons Comparing Antemortem and Postmortem Panoramic Radiographs Using Computer Vision. *Rofo* **2018**, *190*, 1152–1158. https://doi.org/10.1055/a-0632-4744.
42. Jader, G.; Fontineli, J.; Ruiz, M.; Abdalla, K.; Pithon, M.; Oliveira, L. Deep Instance Segmentation of Teeth in Panoramic X-ray Images. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, 29 October–1 November 2018; pp. 400–407. https://doi.org/10.1109/sibgrapi.2018.00058.
43. Jiang, M.-X.; Chen, Y.-M.; Huang, W.-H.; Huang, P.-H.; Tsai, Y.-H.; Huang, Y.-H.; Chiang, C.-K. Teeth-Brushing Recognition Based on Deep Learning. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taichung, Taiwan, 19–21 May 2018; pp. 1–2. https://doi.org/10.1109/icce-china.2018.8448684.
44. Kim, D.W.; Kim, H.; Nam, W.; Kim, H.J.; Cha, I.-H. Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report. *Bone* **2018**, *116*, 207–214. https://doi.org/10.1016/j.bone.2018.04.020.
45. Lee, J.-H.; Kim, D.-H.; Jeong, S.-N.; Choi, S.-H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J. Periodontal Implant. Sci.* **2018**, *48*, 114–123. https://doi.org/10.5051/jpis.2018.48.2.114.
46. 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. *J. Dent.* **2018**, *77*, 106–111. https://doi.org/10.1016/j.jdent.2018.07.015.
47. Lu, S.; Yang, J.; Wang, W.; Li, Z.; Lu, Z. Teeth Classification Based on Extreme Learning Machine. In Proceedings of the 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 30–31 October 2018; pp. 198–202. https://doi.org/10.1109/worlds4.2018.8611570.
48. Shah, H.; Hernandez, P.; Budin, F.; Chittajallu, D.; Vimort, J.B.; Walters, R.; Mol, A.; Khan, A.; Paniagua, B. Automatic quantification framework to detect cracks in teeth. *Proc. SPIE Int. Soc. Opt. Eng.* **2018**, *10578*, 105781K. https://doi.org/10.1117/12.2293603.
49. Song, B.; Sunny, S.; Uthoff, R.; Patrick, S.; Suresh, A.; Kolur, T.; Keerthi, G.; Anbarani, A.; Wilder-Smith, P.; Kuriakose, M.A.; et al. Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning. *Biomed. Opt. Express* **2018**, *9*, 5318–5329. https://doi.org/10.1364/boe.9.005318.
50. Thanathornwong, B. Bayesian-Based Decision Support System for Assessing the Needs for Orthodontic Treatment. *Health Inform. Res.* **2018**, *24*, 22–28. https://doi.org/10.4258/hir.2018.24.1.22.

51. Yang, J.; Xie, Y.; Liu, L.; Xia, B.; Cao, Z.; Guo, C. Automated Dental Image Analysis by Deep Learning on Small Dataset. In Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 July 2018; pp. 492–497. https://doi.org/10.1109/compsac.2018.00076.
52. Yoon, S.; Odlum, M.; Lee, Y.; Choi, T.; Kronish, I.M.; Davidson, K.W.; Finkelstein, J. Applying Deep Learning to Understand Predictors of Tooth Mobility Among Urban Latinos. *Stud. Health Technol. Inform.* **2018**, *251*, 241–244.
53. Zakirov, A.; Ezhov, M.; Gusarev, M.; Alexandrovsky, V.; Shumilov, E. Dental pathology detection in 3D cone-beam CT. *arXiv* **2018**, arXiv:1810.10309.
54. Zanella-Calzada, L.A.; Galván-Tejada, C.E.; Chávez-Lamas, N.M.; Rivas-Gutierrez, J.; Magallanes-Quintanar, R.; Celaya-Padilla, J.M.; Galván-Tejada, J.I.; Gamboa-Rosales, H. Deep Artificial Neural Networks for the Diagnostic of Caries Using Socioeconomic and Nutritional Features as Determinants: Data from NHANES 2013–2014. *Bioengineering* **2018**, *5*, 47. https://doi.org/10.3390/bioengineering5020047.
55. Zhang, K.; Wu, J.; Chen, H.; Lyu, P. An effective teeth recognition method using label tree with cascade network structure. *Comput. Med. Imaging Graph.* **2018**, *68*, 61–70. https://doi.org/10.1016/j.compmedimag.2018.07.001.
56. Ali, H.; Khursheed, M.; Fatima, S.K.; Shuja, S.M.; Noor, S. Object Recognition for Dental Instruments Using SSD-MobileNet. In Proceedings of the 2019 International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, 9–10 March 2019; pp. 1–6. https://doi.org/10.1109/cisct.2019.8777441.
57. Alkaabi, S.; Yussof, S.; Al-Mulla, S. Evaluation of Convolutional Neural Network based on Dental Images for Age Estimation. In Proceedings of the 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates, 19–21 November 2019; pp. 1–5. https://doi.org/10.1109/icecta48151.2019.8959665.
58. Askarian, B.; Tabei, F.; Tipton, G.A.; Chong, J.W. Smartphone-Based Method for Detecting Periodontal Disease. In Proceedings of the 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), Bethesda, MD, USA, 20–22 November 2019; pp. 53–55. https://doi.org/10.1109/hi-poct45284.2019.8962844.
59. Bouchahma, M.; Ben Hammouda, S.; Kouki, S.; Alshemaili, M.; Samara, K. An Automatic Dental Decay Treatment Prediction using a Deep Convolutional Neural Network on X-ray Images. In Proceedings of the 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 3–7 November 2019; pp. 1–4. https://doi.org/10.1109/aiccsa47632.2019.9035278.
60. Casalegno, F.; Newton, T.; Daher, R.; Abdelaziz, M.; Lodi-Rizzini, A.; Schürmann, F.; Krejci, I.; Markram, H. Caries Detection with Near-Infrared Transillumination Using Deep Learning. *J. Dent. Res.* **2019**, *98*, 1227–1233. https://doi.org/10.1177/0022034519871884.
61. 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. *Sci. Rep.* **2019**, *9*, 3840. https://doi.org/10.1038/s41598-019-40414-y.
62. Cheng, B.; Wang, W. Dental hard tissue morphological segmentation with sparse representation-based classifier. *Med. Biol. Eng. Comput.* **2019**, *57*, 1629–1643. https://doi.org/10.1007/s11517-019-01985-0.
63. Chin, C.-L.; Lin, J.-W.; Wei, C.-S.; Hsu, M.-C. Dentition Labeling and Root Canal Recognition Using Ganand Rule-Based System. In Proceedings of the 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Kaohsiung, Taiwan, 21–23 November 2019; pp. 1–6. https://doi.org/10.1109/taai48200.2019.8959926.
64. Choi, H.-I.; Jung, S.-K.; Baek, S.-H.; Lim, W.H.; Ahn, S.-J.; Yang, I.-H.; Kim, T.-W. Artificial Intelligent Model with Neural Network Machine Learning for the Diagnosis of Orthognathic Surgery. *J. Craniofac. Surg.* **2019**, *30*, 1986–1989. https://doi.org/10.1097/scs.0000000000005650.
65. Cui, Z.; Li, C.; Wang, W. ToothNet: Automatic Tooth Instance Segmentation and Identification from Cone Beam CT Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6361–6370. https://doi.org/10.1109/cvpr.2019.00653.
66. Dasanayaka, C.; Dharmasena, B.; Bandara, W.R.; Dissanayake, M.B.; Jayasinghe, R. Segmentation of Mental Foramen in Dental Panoramic Tomography using Deep Learning. In Proceedings of the 2019 14th Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 18–20 December 2019; pp. 81–84. https://doi.org/10.1109/iciis47346.2019.9063312.
67. Cruz, J.C.D.; Garcia, R.G.; Cueto, J.C.C.V.; Pante, S.C.; Toral, C.G.V. Automated Human Identification through Dental Image Enhancement and Analysis. In Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, 29 November–1 December 2019; pp. 1–6. https://doi.org/10.1109/hnicem48295.2019.9072780.
68. Duong, D.Q.; Nguyen, K.-C.T.; Kaipatur, N.R.; Lou, E.H.M.; Noga, M.; Major, P.W.; Punithakumar, K.; Le, L.H. Fully Automated Segmentation of Alveolar Bone Using Deep Convolutional Neural Networks from Intraoral Ultrasound Images. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6632–6635. https://doi.org/10.1109/embc.2019.8857060.
69. Ekert, T.; Krois, J.; Meinhold, L.; Elhennawy, K.; Emara, R.; Golla, T.; Schwendicke, F. Deep Learning for the Radiographic Detection of Apical Lesions. *J. Endod.* **2019**, *45*, 917–922.e5. https://doi.org/10.1016/j.joen.2019.03.016.
70. Hatvani, J.; Basarab, A.; Tourneret, J.-Y.; Gyongy, M.; Kouame, D. A Tensor Factorization Method for 3-D Super Resolution with Application to Dental CT. *IEEE Trans. Med. Imaging* **2019**, *38*, 1524–1531. https://doi.org/10.1109/tmi.2018.2883517.
71. Hatvani, J.; Horvath, A.; Michetti, J.; Basarab, A.; Kouame, D.; Gyongy, M. Deep Learning-Based Super-Resolution Applied to Dental Computed Tomography. *IEEE Trans. Radiat. Plasma Med. Sci.* **2019**, *3*, 120–128. https://doi.org/10.1109/trpms.2018.2827239.

72. Hegazy, M.A.A.; Cho, M.H.; Cho, M.H.; Lee, S.Y. U-net based metal segmentation on projection domain for metal artifact reduction in dental CT. *Biomed. Eng. Lett.* **2019**, *9*, 375–385. https://doi.org/10.1007/s13534-019-00110-2.
73. Hiraiwa, T.; Ariji, Y.; Fukuda, M.; Kise, Y.; Nakata, K.; Katsumata, A.; Fujita, H.; Ariji, E. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac. Radiol.* **2019**, *48*, 20180218. https://doi.org/10.1259/dmfr.20180218.
74. Hu, Z.; Jiang, C.; Sun, F.; Zhang, Q.; Ge, Y.; Yang, Y.; Liu, X.; Zheng, H.; Liang, D. Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks. *Med. Phys.* **2019**, *46*, 1686–1696. https://doi.org/10.1002/mp.13415.
75. Hung, M.; Voss, M.W.; Rosales, M.N.; Li, W.; Su, W.; Xu, J.; Bounsanga, J.; Ruiz-Negrón, B.; Lauren, E.; Licari, F.W. Application of machine learning for diagnostic prediction of root caries. *Gerodontology* **2019**, *36*, 395–404. https://doi.org/10.1111/ger.12432.
76. Ilic, I.; Vodanovic, M.; Subasic, M. Gender Estimation from Panoramic Dental X-ray Images using Deep Convolutional Networks. In Proceedings of the IEEE EUROCON 2019 -18th International Conference on Smart Technologies, Novi Sad, Serbia, 1–4 July 2019; pp. 1–5. https://doi.org/10.1109/eurocon.2019.8861726.
77. Kats, L.; Vered, M.; Zlotogorski-Hurvitz, A.; Harpaz, I. Atherosclerotic carotid plaque on panoramic radiographs: Neural network detection. *Int. J. Comput. Dent.* **2019**, *22*, 163–169.
78. Kats, L.; Vered, M.; Zlotogorski-Hurvitz, A.; Harpaz, I. Atherosclerotic carotid plaques on panoramic imaging: An automatic detection using deep learning with small dataset. *arXiv* **2018**, arXiv:1808.08093.
79. Kim, D.W.; Lee, S.; Kwon, S.; Nam, W.; Cha, I.-H.; Kim, H.J. Deep learning-based survival prediction of oral cancer patients. *Sci. Rep.* **2019**, *9*, 6994. https://doi.org/10.1038/s41598-019-43372-7.
80. Kim, J.; Lee, H.-S.; Song, I.-S.; Jung, K.-H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci. Rep.* **2019**, *9*, 17615. https://doi.org/10.1038/s41598-019-53758-2.
81. Kise, Y.; Shimizu, M.; Ikeda, H.; Fujii, T.; Kuwada, C.; Nishiyama, M.; Funakoshi, T.; Ariji, Y.; Fujita, H.; Katsumata, A.; et al. Usefulness of a deep learning system for diagnosing Sjögren's syndrome using ultrasonography images. *Dentomaxillofac. Radiol.* **2020**, *49*, 20190348. https://doi.org/10.1259/dmfr.20190348.
82. Koch, T.L.; Perslev, M.; Igel, C.; Brandt, S.S. Accurate segmentation of dental panoramic radiographs with U-NETS. In Proceedings of 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 15–19.
83. Krois, J.; Ekert, T.; Meinhold, L.; Golla, T.; Kharbot, B.; Wittemeier, A.; Dörfer, C.; Schwendicke, F. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci. Rep.* **2019**, *9*, 8495. https://doi.org/10.1038/s41598-019-44839-3.
84. Lee, J.-S.; Adhikari, S.; Liu, L.; Jeong, H.-G.; Kim, H.; Yoon, S.-J. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: A preliminary study. *Dentomaxillofac. Radiol.* **2019**, *48*, 20170344. https://doi.org/10.1259/dmfr.20170344.
85. Li, X.; Zhang, Y.; Cui, Q.; Yi, X.; Zhang, Y. Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features. *IEEE Trans. Cybern.* **2019**, *49*, 380–387. https://doi.org/10.1109/tcyb.2017.2772289.
86. Liu, L.; Xu, J.; Huan, Y.; Zou, Z.; Yeh, S.-C.; Zheng, L.-R. A Smart Dental Health-IoT Platform Based on Intelligent Hardware, Deep Learning, and Mobile Terminal. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 898–906. https://doi.org/10.1109/jbhi.2019.2919916.
87. Liu, Y.; Shang, X.; Shen, Z.; Hu, B.; Wang, Z.; Xiong, G. 3D Deep Learning for 3D Printing of Tooth Model. In Proceedings of the 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Zhengzhou, China, 6–8 November 2019; pp. 274–279. https://doi.org/10.1109/soli48380.2019.8955074.
88. Milosevic, D.; Vodanovic, M.; Galic, I.; Subasic, M. Estimating Biological Gender from Panoramic Dental X-ray Images. In Proceedings of the 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 23–25 September 2019; pp. 105–110. https://doi.org/10.1109/ispa.2019.8868804.
89. Minnema, J.; van Eijnatten, M.; Hendriksen, A.A.; Liberton, N.; Pelt, D.M.; Batenburg, K.J.; Forouzanfar, T.; Wolff, J. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med. Phys.* **2019**, *46*, 5027–5035. https://doi.org/10.1002/mp.13793.
90. Moriyama, Y.; Lee, C.; Date, S.; Kashiwagi, Y.; Narukawa, Y.; Nozaki, K.; Murakami, S. Evaluation of Dental Image Augmentation for the Severity Assessment of Periodontal Disease. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; pp. 924–929. https://doi.org/10.1109/csci49370.2019.00176.
91. Moutselos, K.; Berdouses, E.; Oulis, C.; Maglogiannis, I. Recognizing Occlusal Caries in Dental Intraoral Images Using Deep Learning. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 1617–1620. https://doi.org/10.1109/embc.2019.8856553.
92. Murata, M.; Ariji, Y.; Ohashi, Y.; Kawai, T.; Fukuda, M.; Funakoshi, T.; Kise, Y.; Nozawa, M.; Katsumata, A.; Fujita, H.; et al. Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiol.* **2019**, *35*, 301–307. https://doi.org/10.1007/s11282-018-0363-7.
93. Patcas, R.; Bernini, D.; Volokitin, A.; Agustsson, E.; Rothe, R.; Timofte, R. Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. *Int. J. Oral Maxillofac. Surg.* **2019**, *48*, 77–83. https://doi.org/10.1016/j.ijom.2018.07.010.
94. Patcas, R.; Timofte, R.; Volokitin, A.; Agustsson, E.; Eliades, T.; Eichenberger, M.; Bornstein, M.M. Facial attractiveness of cleft patients: A direct comparison between artificial-intelligence-based scoring and conventional rater groups. *Eur. J. Orthod.* **2019**, *41*, 428–433. https://doi.org/10.1093/ejo/cjz007.

95. Sajad, M.; Shafi, I.; Ahmad, J. Automatic Lesion Detection in Periapical X-rays. In Proceedings of the 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Swat, Pakistan, 24–25 July 2019; pp. 1–6. https://doi.org/10.1109/icecce47252.2019.8940661.
96. Senirkentli, G.B.; Sen, S.; Farsak, O.; Bostanci, E. A Neural Expert System Based Dental Trauma Diagnosis Application. In Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 3–5 October 2019; pp. 1–4. https://doi.org/10.1109/tiptekno.2019.8895077.
97. Stark, B.; Samarah, M. Ensemble and Deep Learning for Real-time Sensors Evaluation of algorithms for real-time sensors with application for detecting brushing location. In Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 5–9 December 2019; pp. 555–559. https://doi.org/10.1109/iccc47050.2019.9064080.
98. 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* **2019**, *7*, 84817–84828. https://doi.org/10.1109/access.2019.2924262.
99. Tuzoff, D.V.; Tuzova, L.N.; Bornstein, M.M.; Krasnov, A.S.; Kharchenko, M.A.; Nikolenko, S.I.; Sveshnikov, M.M.; Bednenko, G.B. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac. Radiol.* **2019**, *48*, 20180051. https://doi.org/10.1259/dmfr.20180051.
100. Vinayahalingam, S.; Xi, T.; Bergé, S.; Maal, T.; de Jong, G. Automated detection of third molars and mandibular nerve by deep learning. *Sci. Rep.* **2019**, *9*, 9007. https://doi.org/10.1038/s41598-019-45487-3.
101. Woo, J.; Xing, F.; Prince, J.L.; Stone, M.; Green, J.R.; Goldsmith, T.; Reese, T.G.; Wedeen, V.J.; El Fakhri, G. Differentiating post-cancer from healthy tongue muscle coordination patterns during speech using deep learning. *J. Acoust. Soc. Am.* **2019**, *145*, EL423–EL429. https://doi.org/10.1121/1.5103191.
102. Xu, X.; Liu, C.; Zheng, Y. 3D Tooth Segmentation and Labeling Using Deep Convolutional Neural Networks. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 2336–2348. https://doi.org/10.1109/tvcg.2018.2839685.
103. Yamaguchi, S.; Lee, C.; Karaer, O.; Ban, S.; Mine, A.; Imazato, S. Predicting the Debonding of CAD/CAM Composite Resin Crowns with AI. *J. Dent. Res.* **2019**, *98*, 1234–1238. https://doi.org/10.1177/0022034519867641.
104. Yauney, G.; Rana, A.; Wong, L.C.; Javia, P.; Muftu, A.; Shah, P. Automated Process Incorporating Machine Learning Segmentation and Correlation of Oral Diseases with Systemic Health. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 3387–3393. https://doi.org/10.1109/embc.2019.8857965.
105. Alalharith, D.M.; Alharthi, H.M.; Alghamdi, W.M.; Alsenbel, Y.M.; Aslam, N.; Khan, I.U.; Shahin, S.Y.; Dianišková, S.; Alhareky, M.S.; Barouch, K.K. A Deep Learning-Based Approach for the Detection of Early Signs of Gingivitis in Orthodontic Patients Using Faster Region-Based Convolutional Neural Networks. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8447. https://doi.org/10.3390/ijerph17228447.
106. Aliaga, I.; Vera, V.; Vera, M.; García, E.; Pedrera, M.; Pajares, G. Automatic computation of mandibular indices in dental panoramic radiographs for early osteoporosis detection. *Artif. Intell. Med.* **2020**, *103*, 101816. https://doi.org/10.1016/j.artmed.2020.101816.
107. Banar, N.; Bertels, J.; Laurent, F.; Boedi, R.M.; De Tobel, J.; Thevissen, P.; Vandermeulen, D. Towards fully automated third molar development staging in panoramic radiographs. *Int. J. Leg. Med.* **2020**, *134*, 1831–1841. https://doi.org/10.1007/s00414-020-02283-3.
108. 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. *J. Dent.* **2020**, *100*, 103425. https://doi.org/10.1016/j.jdent.2020.103425.
109. Chang, H.-J.; Lee, S.-J.; Yong, T.-H.; Shin, N.-Y.; Jang, B.-G.; Kim, J.-E.; Huh, K.-H.; Lee, S.-S.; Heo, M.-S.; Choi, S.-C.; et al. Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. *Sci. Rep.* **2020**, *10*, 7531. https://doi.org/10.1038/s41598-020-64509-z.
110. Chen, S.; Wang, L.; Li, G.; Wu, T.-H.; Diachina, S.; Tejera, B.; Kwon, J.J.; Lin, F.-C.; Lee, Y.-T.; Xu, T.; et al. Machine Learning in Orthodontics: Introducing a 3D Auto-segmentation and Auto-landmark Finder of Cbct Images To Assess Maxillary Constriction in Unilateral Impacted Canine patients. *Angle Orthod.* **2020**, *90*, 77–84. https://doi.org/10.2319/012919-59.1.
111. Chen, Y.; Du, H.; Yun, Z.; Yang, S.; Dai, Z.; Zhong, L.; Feng, Q.; Yang, W. Automatic Segmentation of Individual Tooth in Dental CBCT Images from Tooth Surface Map by a Multi-Task FCN. *IEEE Access* **2020**, *8*, 97296–97309. https://doi.org/10.1109/access.2020.2991799.
112. Chung, M.; Lee, M.; Hong, J.; Park, S.; Lee, J.; Lee, J.; Yang, I.-H.; Lee, J.; Shin, Y.-G. Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. *Comput. Biol. Med.* **2020**, *120*, 103720. https://doi.org/10.1016/j.compbiomed.2020.103720.
113. Endres, M.; Hillen, F.; Salloumis, M.; Sedaghat, A.; Niehues, S.; Quatela, O.; Hanken, H.; Smeets, R.; Beck-Broichsitter, B.; Rendenbach, C.; et al. Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs. *Diagnostics* **2020**, *10*, 430. https://doi.org/10.3390/diagnostics10060430.
114. Fan, F.; Ke, W.; Wu, W.; Tian, X.; Lyu, T.; Liu, Y.; Liao, P.; Dai, X.; Chen, H.; Deng, Z. Automatic human identification from panoramic dental radiographs using the convolutional neural network. *Forensic Sci. Int.* **2020**, *314*, 110416. https://doi.org/10.1016/j.forsciint.2020.110416.
115. Fujima, N.; Andreu-Arasa, V.C.; Meibom, S.K.; Mercier, G.A.; Salama, A.R.; Truong, M.T.; Sakai, O. Deep learning analysis using FDG-PET to predict treatment outcome in patients with oral cavity squamous cell carcinoma. *Eur. Radiol.* **2020**, *30*, 6322–6330. https://doi.org/10.1007/s00330-020-06982-8.
116. Fukuda, M.; Ariji, Y.; Kise, Y.; Nozawa, M.; Kuwada, C.; Funakoshi, T.; Muramatsu, C.; Fujita, H.; Katsumata, A.; Ariji, E. Comparison of 3 deep learning neural networks for classifying the relationship between the mandibular third molar and the

mandibular canal on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2020**, *130*, 336–343. https://doi.org/10.1016/j.oooo.2020.04.005.

117. Fukuda, M.; Inamoto, K.; Shibata, N.; Ariji, Y.; Yanashita, Y.; Kutsuna, S.; Nakata, K.; Katsumata, A.; Fujita, H.; Ariji, E. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol.* **2020**, *36*, 337–343. https://doi.org/10.1007/s11282-019-00409-x.
118. Geetha, V.; Aprameya, K.S.; Hinduja, D.M. Dental caries diagnosis in digital radiographs using back-propagation neural network. *Health Inf. Sci. Syst.* **2020**, *8*, 8. https://doi.org/10.1007/s13755-019-0096-y.
119. Hung, M.; Hon, E.S.; Ruiz-Negron, B.; Lauren, E.; Moffat, R.; Su, W.; Xu, J.; Park, J.; Prince, D.; Cheever, J.; et al. Exploring the Intersection between Social Determinants of Health and Unmet Dental Care Needs Using Deep Learning. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7286. https://doi.org/10.3390/ijerph17197286.
120. Hung, M.; Li, W.; Hon, E.S.; Su, S.; Su, W.; He, Y.; Sheng, X.; Holubkov, R.; Lipsky, M.S. Prediction of 30-Day Hospital Readmissions for All-Cause Dental Conditions using Machine Learning. *Risk Manag. Health Policy* **2020**, *13*, 2047–2056. https://doi.org/10.2147/rmhp.s272824.
121. Jaskari, J.; Sahlsten, J.; Järnstedt, J.; Mehtonen, H.; Karhu, K.; Sundqvist, O.; Hietanen, A.; Varjonen, V.; Mattila, V.; Kaski, K. Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes. *Sci. Rep.* **2020**, *10*, 5842. https://doi.org/10.1038/s41598-020-62321-3.
122. Jeong, S.H.; Yun, J.P.; Yeom, H.-G.; Lim, H.J.; Lee, J.; Kim, B.C. Deep learning based discrimination of soft tissue profiles requiring orthognathic surgery by facial photographs. *Sci. Rep.* **2020**, *10*, 16235. https://doi.org/10.1038/s41598-020-73287-7.
123. Joshi, S.V.; Kanphade, R.D. Deep Learning Based Person Authentication Using Hand Radiographs: A Forensic Approach. *IEEE Access* **2020**, *8*, 95424–95434. https://doi.org/10.1109/access.2020.2995788.
124. Kats, L.; Vered, M.; Blumer, S.; Kats, E. Neural Network Detection and Segmentation of Mental Foramen in Panoramic Imaging. *J. Clin. Pediatr. Dent.* **2020**, *44*, 168–173. https://doi.org/10.17796/1053-4625-44.3.6.
125. Khan, H.A.; Haider, M.A.; Ansari, H.A.; Ishaq, H.; Kiyani, A.; Sohail, K.; Muhammad, M.; Khurram, S.A. Automated feature detection in dental periapical radiographs by using deep learning. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2021**, *131*, 711–720. https://doi.org/10.1016/j.oooo.2020.08.024.
126. Kim, H.; Shim, E.; Park, J.; Kim, Y.-J.; Lee, U.; Kim, Y. Web-based fully automated cephalometric analysis by deep learning. *Comput. Methods Programs Biomed.* **2020**, *194*, 105513. https://doi.org/10.1016/j.cmpb.2020.105513.
127. Kim, I.; Misra, D.; Rodriguez, L.; Gill, M.; Liberton, D.K.; Almpani, K.; Lee, J.S.; Antani, S. Malocclusion Classification on 3D Cone-Beam CT Craniofacial Images Using Multi-Channel Deep Learning Models. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 1294–1298. https://doi.org/10.1109/embc44109.2020.9176672.
128. Kim, J.-E.; Nam, N.-E.; Shim, J.-S.; Jung, Y.-H.; Cho, B.-H.; Hwang, J.J. Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs. *J. Clin. Med.* **2020**, *9*, 1117. https://doi.org/10.3390/jcm9041117.
129. Kunz, F.; Stellzig-Eisenhauer, A.; Zeman, F.; Boldt, J. Artificial intelligence in orthodontics: Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *J. Orofac. Orthop.* **2020**, *81*, 52–68. https://doi.org/10.1007/s00056-019-00203-8.
130. Kuramoto, N.; Ichimura, K.; Jayatilake, D.; Shimokakimoto, T.; Hidaka, K.; Suzuki, K. Deep Learning-Based Swallowing Monitor for Realtime Detection of Swallow Duration. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4365–4368. https://doi.org/10.1109/embc44109.2020.9176721.
131. Kuwada, C.; Ariji, Y.; Fukuda, M.; Kise, Y.; Fujita, H.; Katsumata, A.; Ariji, E. Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2020**, *130*, 464–469. https://doi.org/10.1016/j.oooo.2020.04.813.
132. 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. *Sci. Rep.* **2020**, *10*, 5711. https://doi.org/10.1038/s41598-020-62586-8.
133. Lee, D.; Park, C.; Lim, Y.; Cho, H. A Metal Artifact Reduction Method Using a Fully Convolutional Network in the Sinogram and Image Domains for Dental Computed Tomography. *J. Digit. Imaging* **2020**, *33*, 538–546. https://doi.org/10.1007/s10278-019-00297-x.
134. Lee, J.-H.; Han, S.-S.; Kim, Y.H.; Lee, C.; Kim, I. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2020**, *129*, 635–642. https://doi.org/10.1016/j.oooo.2019.11.007.
135. Lee, J.-H.D.; Jeong, S.-N. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: A pilot study. *Medicine* **2020**, *99*, e20787. https://doi.org/10.1097/md.0000000000020787.
136. Lee, J.-H.; Kim, D.-H.; Jeong, S.-N. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis.* **2020**, *26*, 152–158. https://doi.org/10.1111/odi.13223.
137. Lee, J.-H.; Kim, Y.-T.; Lee, J.-B.; Jeong, S.-N. A Performance Comparison between Automated Deep Learning and Dental Professionals in Classification of Dental Implant Systems from Dental Imaging: A Multi-Center Study. *Diagnostics* **2020**, *10*, 910. https://doi.org/10.3390/diagnostics10110910.
138. Lee, J.-H.; Yu, H.-J.; Kim, M.-J.; Kim, J.-W.; Choi, J. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. *BMC Oral Health* **2020**, *20*, 270. https://doi.org/10.1186/s12903-020-01256-7.

139. Lee, K.-S.; Jung, S.-K.; Ryu, J.-J.; Shin, S.-W.; Choi, J. Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs. *J. Clin. Med.* **2020**, *9*, 392. https://doi.org/10.3390/jcm9020392.
140. Lee, S.; Woo, S.; Yu, J.; Seo, J.; Lee, J.; Lee, C. Automated CNN-Based Tooth Segmentation in Cone-Beam CT for Dental Implant Planning. *IEEE Access* **2020**, *8*, 50507–50518. https://doi.org/10.1109/access.2020.2975826.
141. Li, C.; Zhang, D.; Chen, S. Research about Tongue Image of Traditional Chinese Medicine(TCM) Based on Artificial Intelligence Technology. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 633–636. https://doi.org/10.1109/itoec49072.2020.9141557.
142. Li, Q.; Chen, K.; Han, L.; Zhuang, Y.; Li, J.; Lin, J. Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN. *J. X-ray Sci. Technol.* **2020**, *28*, 905–922. https://doi.org/10.3233/XST-200678.
143. Li, S.; Pang, Z.; Song, W.; Guo, Y.; You, W.; Hao, A.; Qin, H. Low-Shot Learning of Automatic Dental Plaque Segmentation Based on Local-to-Global Feature Fusion. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 664–668. https://doi.org/10.1109/isbi45749.2020.9098741.
144. Lian, C.; Wang, L.; Wu, T.-H.; Wang, F.; Yap, P.-T.; Ko, C.-C.; Shen, D. Deep Multi-Scale Mesh Feature Learning for Automated Labeling of Raw Dental Surfaces From 3D Intraoral Scanners. *IEEE Trans. Med. Imaging* **2020**, *39*, 2440–2450. https://doi.org/10.1109/tmi.2020.2971730.
145. Mahdi, F.P.; Motoki, K.; Kobashi, S. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. *Sci. Rep.* **2020**, *10*, 19261. https://doi.org/10.1038/s41598-020-75887-9.
146. Mallishery, S.; Chhatpar, P.; Banga, K.S.; Shah, T.; Gupta, P. The precision of case difficulty and referral decisions: An innovative automated approach. *Clin. Oral Investig.* **2020**, *24*, 1909–1915. https://doi.org/10.1007/s00784-019-03050-4.
147. Matsuda, S.; Miyamoto, T.; Yoshimura, H.; Hasegawa, T. Personal identification with orthopantomography using simple convolutional neural networks: A preliminary study. *Sci. Rep.* **2020**, *10*, 13559. https://doi.org/10.1038/s41598-020-70474-4.
148. Boedi, R.M.; Banar, N.; De Tobel, J.; Bertels, J.; Vandermeulen, D.; Thevissen, P.W. Effect of Lower Third Molar Segmentations on Automated Tooth Development Staging using a Convolutional Neural Network. *J. Forensic Sci.* **2020**, *65*, 481–486. https://doi.org/10.1111/1556-4029.14182.
149. Ngoc, V.T.N.; Agwu, A.C.; Son, L.H.; Tuan, T.M.; Giap, C.N.; Thanh, M.T.G.; Duy, H.B.; Ngan, T.T. The Combination of Adaptive Convolutional Neural Network and Bag of Visual Words in Automatic Diagnosis of Third Molar Complications on Dental X-ray Images. *Diagnostics* **2020**, *10*, 209. https://doi.org/10.3390/diagnostics10040209.
150. Oh, J.H.; Pouryahya, M.; Iyer, A.; Apte, A.P.; Deasy, J.O.; Tannenbaum, A. A novel kernel Wasserstein distance on Gaussian measures: An application of identifying dental artifacts in head and neck computed tomography. *Comput. Biol. Med.* **2020**, *120*, 103731. https://doi.org/10.1016/j.compbiomed.2020.103731.
151. Orhan, K.; Bayrakdar, I.S.; Ezhov, M.; Kravtsov, A.; Özyürek, T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int. Endod. J.* **2020**, *53*, 680–689. https://doi.org/10.1111/iej.13265.
152. Ren, J.; Fan, H.; Yang, J.; Ling, H. Detection of Trabecular Landmarks for Osteoporosis Prescreening in Dental Panoramic Radiographs. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 2194–2197. https://doi.org/10.1109/embc44109.2020.9175281.
153. Schwendicke, F.; Elhennawy, K.; Paris, S.; Friebertshäuser, P.; Krois, J. Deep learning for caries lesion detection in near-infrared light transillumination images: A pilot study. *J. Dent.* **2020**, *92*, 103260. https://doi.org/10.1016/j.jdent.2019.103260.
154. 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. *J. Endod.* **2020**, *46*, 987–993. https://doi.org/10.1016/j.joen.2020.03.025.
155. Sukegawa, S.; Yoshii, K.; Hara, T.; Yamashita, K.; Nakano, K.; Yamamoto, N.; Nagatsuka, H.; Furuki, Y. Deep Neural Networks for Dental Implant System Classification. *Biomolecules* **2020**, *10*, 984. https://doi.org/10.3390/biom10070984.
156. Sun, D.; Pei, Y.; Song, G.; Guo, Y.; Ma, G.; Xu, T.; Zha, H. Tooth Segmentation and Labeling from Digital Dental Casts. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 669–673. https://doi.org/10.1109/isbi45749.2020.9098397.
157. Takahashi, T.; Nozaki, K.; Gonda, T.; Ikebe, K. A system for designing removable partial dentures using artificial intelligence. Part 1. Classification of partially edentulous arches using a convolutional neural network. *J. Prosthodont. Res.* **2021**, *65*, 115–118. https://doi.org/10.2186/jpr.jpor_2019_354.
158. Tang, W.; Gao, Y.; Liu, L.; Xia, T.; He, L.; Zhang, S.; Guo, J.; Li, W.; Xu, Q. An Automatic Recognition of Tooth- Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning. *IEEE Access* **2020**, *8*, 153470–153478. https://doi.org/10.1109/access.2020.3017725.
159. Thanathornwong, B.; Suebnukarn, S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Sci. Dent.* **2020**, *50*, 169–174. https://doi.org/10.5624/isd.2020.50.2.169.
160. Vila-Blanco, N.; Carreira, M.J.; Varas-Quintana, P.; Balsa-Castro, C.; Tomas, I. Deep Neural Networks for Chronological Age Estimation From OPG Images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2374–2384. https://doi.org/10.1109/tmi.2020.2968765.
161. Vranckx, M.; Van Gerven, A.; Willems, H.; Vandemeulebroucke, A.; Leite, A.F.; Politis, C.; Jacobs, R. Artificial Intelligence (AI)-Driven Molar Angulation Measurements to Predict Third Molar Eruption on Panoramic Radiographs. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3716. https://doi.org/10.3390/ijerph17103716.

162. Wang, X.; Liu, J.; Wu, C.; Liu, J.; Li, Q.; Chen, Y.; Wang, X.; Chen, X.; Pang, X.; Chang, B.; et al. Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 973–980. https://doi.org/10.1016/j.csbj.2020.04.002.
163. Wang, Y.; Hays, R.D.; Marcus, M.; Maida, C.A.; Shen, J.; Xiong, D.; Coulter, I.D.; Lee, S.Y.; Spolsky, V.W.; Crall, J.J.; et al. Developing Children's Oral Health Assessment Toolkits Using Machine Learning Algorithm. *JDR Clin. Transl. Res.* **2020**, *5*, 233–243. https://doi.org/10.1177/2380084419885612.
164. Welch, M.L.; McIntosh, C.; Purdie, T.G.; Wee, L.; Traverso, A.; Dekker, A.; Haibe-Kains, B.; Jaffray, D.A. Automatic classification of dental artifact status for efficient image veracity checks: Effects of image resolution and convolutional neural network depth. *Phys. Med. Biol.* **2020**, *65*, 015005. https://doi.org/10.1088/1361-6560/ab5427.
165. Welch, M.L.; McIntosh, C.; Traverso, A.; Wee, L.; Purdie, T.G.; Dekker, A.; Haibe-Kains, B.; Jaffray, D.A. External validation and transfer learning of convolutional neural networks for computed tomography dental artifact classification. *Phys. Med. Biol.* **2020**, *65*, 035017. https://doi.org/10.1088/1361-6560/ab63ba.
166. Welikala, R.A.; Remagnino, P.; Lim, J.H.; Chan, C.S.; Rajendran, S.; Kallarakkal, T.G.; Zain, R.B.; Jayasinghe, R.D.; Rimal, J.; Kerr, A.R.; et al. Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer. *IEEE Access* **2020**, *8*, 132677–132693. https://doi.org/10.1109/access.2020.3010180.
167. Yang, H.; Jo, E.; Kim, H.J.; Cha, I.-H.; Jung, Y.-S.; Nam, W.; Kim, J.-Y.; Kim, J.-K.; Kim, Y.H.; Oh, T.G.; et al. Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs. *J. Clin. Med.* **2020**, *9*, 1839. https://doi.org/10.3390/jcm9061839.
168. 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* **2020**, *20*, 141. https://doi.org/10.1186/s12903-020-01114-6.
169. 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 Dis.* **2022**, *28*, 173–181. https://doi.org/10.1111/odi.13735.
170. Zheng, Z.; Yan, H.; Setzer, F.C.; Shi, K.J.; Mupparapu, M.; Li, J. Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 603–614. https://doi.org/10.1109/tase.2020.3025871.
171. Zhu, G.; Piao, Z.; Kim, S.C. Tooth Detection and Segmentation with Mask R-CNN. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 19–21 February 2020; pp. 070–072. https://doi.org/10.1109/icaiic48513.2020.9065216.
172. Alkhadar, H.; Macluskey, M.; White, S.; Ellis, I.; Gardner, A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J. Oral Pathol. Med.* **2021**, *50*, 378–384. https://doi.org/10.1111/jop.13135.
173. Leite, A.F.; Van Gerven, A.; Willems, H.; Beznik, T.; Lahoud, P.; Gaêta-Araujo, H.; Vranckx, M.; Jacobs, R. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clin. Oral Investig.* **2021**, *25*, 2257–2267. https://doi.org/10.1007/s00784-020-03544-6.
174. Machado, R.A.; De Oliveira Silva, C.; Martelli-Junior, H.; das Neves, L.T.; Coletta, R.D. Machine learning in prediction of genetic risk of nonsyndromic oral clefts in the Brazilian population. *Clin. Oral Investig.* **2021**, *25*, 1273–1280. https://doi.org/10.1007/s00784-020-03433-y.
175. Muramatsu, C.; Morishita, T.; Takahashi, R.; Hayashi, T.; Nishiyama, W.; Ariji, Y.; Zhou, X.; Hara, T.; Katsumata, A.; Ariji, E.; et al. Tooth detection and classification on panoramic radiographs for automatic dental chart filing: Improved classification by multi-sized input data. *Oral Radiol.* **2021**, *37*, 13–19. https://doi.org/10.1007/s11282-019-00418-w.
176. Schwendicke, F.; Rossi, J.; Göstemeyer, G.; Elhennawy, K.; Cantu, A.; Gaudin, R.; Chaurasia, A.; Gehrung, S.; Krois, J. Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection. *J. Dent. Res.* **2021**, *100*, 369–376. https://doi.org/10.1177/0022034520972335.
177. Yasa, Y.; Çelik, Ö.; Bayrakdar, I.S.; Pekince, A.; Orhan, K.; Akarsu, S.; Atasoy, S.; Bilgir, E.; Odabaş, A.; Aslan, A.F. An artificial intelligence proposal to automatic teeth detection and numbering in dental bite-wing radiographs. *Acta Odontol. Scand.* **2021**, *79*, 275–281. https://doi.org/10.1080/00016357.2020.1840624.
178. Rischke, R.; Schneider, L.; Müller, K.; Samek, W.; Schwendicke, F.; Krois, J. Federated Learning in Dentistry: Chances and Challenges. *J. Dent. Res.* **2022**, *101*, 1269–1273. https://doi.org/10.1177/00220345221108953.
179. ITU/WHO. Focus Group on "Artificial Intelligence for Health". 2018. Available online: https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx (accessed on 31 October 2022).
180. Schwendicke, F.; Singh, T.; Lee, J.-H.; Gaudin, R.; Chaurasia, A.; Wiegand, T.; Uribe, S.; Krois, J.; on behalf of the IADR e-Oral Health Network; the ITU WHO Focus Group AI for Health. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J. Dent.* **2021**, *107*, 103610. https://doi.org/10.1016/j.jdent.2021.103610.
181. Lones, M. How to avoid machine learning pitfalls: A guide for academic researchers. *arXiv* **2021**, arXiv:2108.02497.
182. Stevens, L.M.; Mortazavi, B.J.; Deo, R.C.; Curtis, L.; Kao, D.P. Recommendations for Reporting Machine Learning Analyses in Clinical Research. *Circ. Cardiovasc. Qual. Outcomes* **2020**, *13*, e006556. https://doi.org/10.1161/circoutcomes.120.006556.
183. Vermeire, T.; Brughmans, D.; Goethals, S.; de Oliveira, R.M.B.; Martens, D. Explainable image classification with evidence counterfactual. *Pattern Anal. Appl.* **2022**, *25*, 315–335. https://doi.org/10.1007/s10044-021-01055-y.
184. Nguyen, K.; Duong, D.; Almeida, F.; Major, P.; Kaipatur, N.; Pham, T.; Lou, E.; Noga, M.; Punithakumar, K.; Le, L. Alveolar Bone Segmentation in Intraoral Ultrasonographs with Machine Learning. *J. Dent. Res.* **2020**, *99*, 1054–1061. https://doi.org/10.1177/0022034520920593.

185. Min, X.; Haijin, C. Research on Rapid Detection of Tooth Profile Parameters of the Clothing Wires Based on Image Processing. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; pp. 586–590. https://doi.org/10.1109/icaica50127.2020.9182581.
186. Rasteau, S.; Sigaux, N.; Louvrier, A.; Bouletreau, P. Three-dimensional acquisition technologies for facial soft tissues—Applications and prospects in orthognathic surgery. *J. Stomatol. Oral Maxillofac. Surg.* **2020**, *121*, 721–728. https://doi.org/10.1016/j.jormas.2020.05.013.
187. Dot, G.; Rafflenbeul, F.; Arbotto, M.; Gajny, L.; Rouch, P.; Schouman, T. Accuracy and reliability of automatic three-dimensional cephalometric landmarking. *Int. J. Oral Maxillofac. Surg.* **2020**, *49*, 1367–1378. https://doi.org/10.1016/j.ijom.2020.02.015.
188. Kapralos, V.; Koutroulis, A.; Irinakis, E.; Kouros, P.; Lyroudia, K.; Pitas, I.; Mikrogeorgis, G. Digital subtraction radiography in detection of vertical root fractures: Accuracy evaluation for root canal filling, fracture orientation and width variables. An ex-vivo study. *Clin. Oral Investig.* **2020**, *24*, 3671–3681. https://doi.org/10.1007/s00784-020-03245-0.
189. Tanaka, R.; Tanaka, T.; Yeung, A.W.K.; Taguchi, A.; Katsumata, A.; Bornstein, M.M. Mandibular Radiomorphometric Indices and Tooth Loss as Predictors for the Risk of Osteoporosis using Panoramic Radiographs. *Oral Health Prev. Dent.* **2020**, *18*, 773–782. https://doi.org/10.3290/J.OHPD.A45081.
190. Laishram, A.; Thongam, K. Detection and Classification of Dental Pathologies using Faster-RCNN in Orthopantomogram Radiography Image. In Proceedings of the 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 27–28 February 2020; pp. 423–428. https://doi.org/10.1109/spin48934.2020.9071242.
191. Rao, G.K.L.; Mokhtar, N.; Iskandar, Y.H.P.; Srinivasa, A.C. Learning Orthodontic Cephalometry through Augmented Reality: A Conceptual Machine Learning Validation Approach. In Proceedings of the 2018 International Conference on Electrical Engineering and Informatics (ICELTICs), Banda Aceh, Indonesia, 19–20 September 2018; pp. 133–138. https://doi.org/10.1109/iceltics.2018.8548939.
192. Damiani, G.; Grossi, E.; Berti, E.; Conic, R.; Radhakrishna, U.; Pacifico, A.; Bragazzi, N.; Piccinno, R.; Linder, D. Artificial neural networks allow response prediction in squamous cell carcinoma of the scalp treated with radiotherapy. *J. Eur. Acad. Dermatol. Venereol.* **2020**, *34*, 1369–1373. https://doi.org/10.1111/jdv.16210.
193. Yoon, S.; Choi, T.; Odlum, M.; Mitchell, D.A.; Kronish, I.M.; Davidson, K.W.; Finkelstein, J. Machine Learning to Identify Behavioral Determinants of Oral Health in Inner City Older Hispanic Adults. *Stud. Health Technol. Inform.* **2018**, *251*, 253–256. https://doi.org/10.3233/978-1-61499-880-8-253.
194. Yatabe, M.; Prieto, J.C.; Styner, M.; Zhu, H.; Ruellas, A.C.; Paniagua, B.; Budin, F.; Benavides, E.; Shoukri, B.; Michoud, L.; et al. 3D superimposition of craniofacial imaging—The utility of multicentre collaborations. *Orthod. Craniofacial Res.* **2019**, *22* (Suppl. S1), 213–220. https://doi.org/10.1111/ocr.12281.
195. Hung, M.; Lauren, E.; Hon, E.S.; Birmingham, W.C.; Xu, J.; Su, S.; Hon, S.D.; Park, J.; Dang, P.; Lipsky, M.S. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. *J. Med. Internet Res.* **2020**, *22*, e22590. https://doi.org/10.2196/22590.
196. Suhail, Y.; Upadhyay, M.; Chhibber, A.; Kshitiz Machine Learning for the Diagnosis of Orthodontic Extractions: A Computational Analysis Using Ensemble Learning. *Bioengineering* **2020**, *7*, 55. https://doi.org/10.3390/bioengineering7020055.
197. Mehandru, N.; Hicks, W.L.; Singh, A.K.; Markiewicz, M.R. Machine Learning for Identification of Craniomaxillofacial Radiographic Lesions. *J. Oral Maxillofac. Surg.* **2020**, *78*, 2106–2107. https://doi.org/10.1016/j.joms.2020.07.010.

Research Reports: Clinical

Benchmarking Deep Learning Models for Tooth Structure Segmentation

Journal of Dental Research
2022, Vol. 101(11) 1343–1349

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00220345221100169
journals.sagepub.com/home/jdr

L. Schneider[1,2], L. Arsiwala-Scheppach[1,2], J. Krois[1,2], H. Meyer-Lueckel[3], K.K. Bressem[4,5], S.M. Niehues[4], and F. Schwendicke[1,2]

Abstract

A wide range of deep learning (DL) architectures with varying depths are available, with developers usually choosing one or a few of them for their specific task in a nonsystematic way. Benchmarking (i.e., the systematic comparison of state-of-the art architectures on a specific task) may provide guidance in the model development process and may allow developers to make better decisions. However, comprehensive benchmarking has not been performed in dentistry yet. We aimed to benchmark a range of architecture designs for 1 specific, exemplary case: tooth structure segmentation on dental bitewing radiographs. We built 72 models for tooth structure (enamel, dentin, pulp, fillings, crowns) segmentation by combining 6 different DL network architectures (U-Net, U-Net++, Feature Pyramid Networks, LinkNet, Pyramid Scene Parsing Network, Mask Attention Network) with 12 encoders from 3 different encoder families (ResNet, VGG, DenseNet) of varying depth (e.g., VGG13, VGG16, VGG19). On each model design, 3 initialization strategies (ImageNet, CheXpert, random initialization) were applied, resulting overall into 216 trained models, which were trained up to 200 epochs with the Adam optimizer (learning rate = 0.0001) and a batch size of 32. Our data set consisted of 1,625 human-annotated dental bitewing radiographs. We used a 5-fold cross-validation scheme and quantified model performances primarily by the F1-score. Initialization with ImageNet or CheXpert weights significantly outperformed random initialization ($P < 0.05$). Deeper and more complex models did not necessarily perform better than less complex alternatives. VGG-based models were more robust across model configurations, while more complex models (e.g., from the ResNet family) achieved peak performances. In conclusion, initializing models with pretrained weights may be recommended when training models for dental radiographic analysis. Less complex model architectures may be competitive alternatives if computational resources and training time are restricting factors. Models developed and found superior on nondental data sets may not show this behavior for dental domain-specific tasks.

Keywords: computer vision, artificial intelligence, segmentation, tooth structures, transfer learning, neural networks

Introduction

Deep learning (DL) has been widely employed for image analytics in dermatology (skin photographs) (Jafari et al. 2016), ophthalmology (retina imagery) (Son et al. 2020), or pathology (histological specimens) (Kather et al. 2019). Also in dentistry, DL classification models have been employed to predict the modality of radiographs (Cejudo et al. 2021), the presence of caries lesions (Lee et al. 2018), periodontal bone loss (Krois et al. 2019), and apical lesions (Ekert et al. 2019) on dental radiographs. DL segmentation models, which perform a classification task at the pixel level, were used for the segmentation of anatomical structures in panoramic images (Cha et al. 2021), apical lesions on cone beam computed tomography scans (Orhan et al. 2020), periodontal bone loss on panoramic radiographs (Kim et al. 2019), and caries lesions on bitewings (Cantu et al. 2020).

Recent guidelines in the field call for rigorous and comprehensive planning, conducting, and reporting of DL studies in dentistry (Schwendicke et al. 2021). One key element in those guidelines is a hypothesis-driven selection of the DL model configuration, which includes, among others, its architecture, its complexity, and the initialization strategy for the model weights (e.g., via transfer learning). (1) **Architecture:** The basic unit of an artificial neural network is a neuron, which is a nonlinear mathematical model inspired by the biological neuron (McCulloch and Pitts 1943). These units are stacked to build layers that are connected via mathematical operations

[1]Department of Oral Diagnostics, Digital Health and Health Services Research, Charité–Universitätsmedizin, Berlin, Germany
[2]ITU/WHO Focus Group on AI for Health, Topic Group Dental Diagnostics and Digital Dentistry, Geneva, Switzerland
[3]Department of Restorative, Preventive and Pediatric Dentistry, Zahnmedizinische Kliniken der Universität Bern, University of Bern, Bern, Switzerland
[4]Charité–Universitätsmedizin Berlin, Klinik für Radiologie, Berlin, Germany
[5]Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Berlin, Germany

A supplemental appendix to this article is available online.

Corresponding Author:
F. Schwendicke, Department of Oral Diagnostics, Digital Health and Health Services Research, Charité–Universitätsmedizin Berlin, Aßmannshauser Str. 4-6, Berlin, 14197, Germany.
Email: falk.schwendicke@charite.de

with other layers of neurons. The arrangement of these layers and operations defines the model architecture. Model architectures such as ResNet (He et al. 2016) or VGG (Simonyan and Zisserman 2015) are widely used in the field of machine learning. For image segmentation, specialized layers extend the basic model architectures, which in such a setting are referred to as backbone. This allows one to plug in different backbones and benchmark them for image segmentation tasks. (2) **Complexity:** Most model architectures are available in different degrees of complexities, which reflects the depth of the neural network (i.e., the number of layers included and the number of neurons and connections between them). Deeper models are more complex as they consist of more parameters (i.e., connections between neurons). (3) **Initialization:** The connections between neurons and layers of neurons, which are also referred to as model weights, are basically digits that correspond to the strength of the connection. During model training, these weights are adjusted to find a set of values that are most suitable to solve the underlying task. Starting with a predefined setting of these weights enhances the efficiency of the training process and improves model convergence. Using a predefined setting of weights that stem from a previously trained neural network provides a meaningful starting point for the training process. This technique is referred to as transfer learning (Tan et al. 2018).

The sheer number of possible configurations of model architecture, including backbones, complexity, and initialization strategies, impedes systematic and comprehensive comparisons of existing study findings (Schwendicke et al. 2019). One strategy to overcome this issue is to perform benchmarking, which involves the systematic comparison of different model architectures and model configurations on an identical data set. Such benchmarking studies provide guidance for researchers in the model design process, which improves research efficiency by enabling the development of high-performing models in a shorter time at lower development costs. However, in the medical domain and, more so, dentistry, benchmarking initiatives are scarce, owing to limited data availability and high costs for establishing solid and accepted ground truth labels and annotations. To cope with these difficulties, the ITU/WHO Focus Group Artificial Intelligence for Health (FG-AI4H) is developing a standard evaluation process and benchmarking framework for artificial intelligence (AI) models in health. The present study will inform this initiative.

In a recent benchmarking study, Bressem et al. (2020) benchmarked 16 different model architectures for classification tasks on 2 openly available chest radiograph data sets: CheXpert (Irvin et al. 2019) and the COVID-19 Image Data Collection. They showed that complex and deep models do not necessary outperform simpler architectures. Similarly, Ke et al. (2021) addressed the assumption that model architectures that perform better on the ImageNet data set (Deng et al. 2009), a popular open-source benchmark data set containing millions of labeled images, also generally perform better on CheXpert. This assumption was not found to be valid based on the comparison of 16 convolutional architectures on 5 classification tasks.

In the present study, we aim to expand the studies of Bressem et al. (2020) and Ke et al. (2021) to a dental segmentation task. We benchmarked 216 DL models defined by their architecture, complexity, and initialization strategy. We evaluated these model configurations for a specific dental task: tooth structure (enamel, dentin, pulpal cavity, fillings, and crowns) segmentation on dental bitewing radiographs. We deliberately decided to use this application since first, there is evidence that segmentation models perform well on this task (Ronneberger et al. 2015a) and, second, there is less ambiguity about the establishment of the ground truth for this task, with tooth structures being easily discriminated even by nonsenior clinicians. We expect our results to inform dental researchers about suitable model configurations for their experiments and aim to contribute to evidence-guided DL model selection in dental research.

Materials and Methods

Benchmarking Tasks

This analysis is based on a segmentation task for tooth structures on dental bitewing radiographs. Several model development aspects were benchmarked. (1) **Architecture:** First, we assessed different DL model architectures, since to date, most neural networks have mainly been benchmarked on openly available data sets such as ImageNet. However, it is not yet determined whether the best-performing networks on ImageNet will also perform best for dental radiographic images. Hence, we benchmarked architectures such as U-Net (Ronneberger et al. 2015b), U-Net++ (Zhou et al. 2018), Feature Pyramid Networks (FPN) (Kirillov et al. 2019), LinkNet (Chaurasia and Culurciello 2017), Pyramid Scene Parsing Network (PSPNet) (Zhao et al. 2017), and Mask Attention Network (MAnet) (Fan et al. 2020), among others. These networks were selected, as they all allow to employ the same established backbones of varying depths of model layers (ResNet50 [He et al. 2016], VGG13 [Simonyan and Zisserman 2015], DenseNet121 [Huang et al. 2017]). The depth of the encoder is conventionally represented by the digits behind the name of the architecture (e.g., ResNet18, ResNet34). All model implementations were taken from the same software package (Yakubovskiy 2020). (2) **Complexity:** Second, we investigated the model performances emanating from model complexity. Supposedly, deeper DL models, which have more trainable parameters, outperform shallower alternatives if enough data and computational resources are available. However, deeper models are more likely to overfit training data, and model convergence may not be reached. Furthermore, limited computational resources imply restrictions regarding image resolution or batch size; both may negatively affect the model performance. (3) **Initialization:** Third, we analyzed different initialization strategies, such as random weights initialization or initialization based on pretrained weights from the ImageNet as well as the CheXpert data set. The latter strategies are referred to as transfer learning. Thereby, features learned on large, open data sets are directly transferred to a new task and hence do not have to be learned from scratch. This technique speeds up model convergence and improves model performance. Initialization with ImageNet is one of the most popular transfer learning strategies. Even for tasks on medical radiographs, transferring knowledge from models trained on ImageNet yields a boost in performance (Ke et al. 2021). However, the feature space learned on ImageNet differs fundamentally from medical features of radiographs. ImageNet consists of natural RGB color images that are classified into more than 20,000 classes, while radiographic images contain grayscale images and are usually classified in only a few categories. Hence, an initialization with pretrained models on radiographic images such as the CheXpert data set (Irvin et al. 2019) may potentially be more suitable for medical segmentation tasks of, for instance, dental radiographs.

Ethics Statement

This study was ethically approved by the ethics committee of the Charité (EA4/102/14 and EA4/080/18).

Study Design

In the present study, 72 models were built from a combination of varying architectures and encoder backbones and were each trained with 3 different initialization strategies on a tooth structure segmentation task. Each model was trained with 5-fold cross-validation with varying train, validation, and test sets for each fold. Hence, for each model run, the data were randomly split into training, validation, and test data with proportions of 60% (3 folds), 20% (1 fold), and 20% (1 fold), respectively. We additionally applied a sensitivity analysis and assessed model performances on underrepresented classes (in our case, fillings and crowns), as in real life, medical data set class imbalance is likely the rule and not the exception. Reporting of this study follows the Standards for Reporting Diagnostic Accuracy guideline (STARD) (Bossuyt et al. 2015) and the Checklist for Artificial Intelligence in Dental Research (Schwendicke et al. 2021).

Performance Metrics

Model performances were primarily quantified by the F1-score, which captures the harmonic mean of recall (specificity) and precision (positive predictive value [PPV]). F1-scores are computed from the sum of true positives, false positives, and false negatives over all channels of segmentation masks and cross-validation folds. This method was described by Forman and Scholz (2010) and results in unbiased F-scores in cross-validation schemes. Secondary metrics were accuracy, sensitivity, precision, and intersection of union (IoU). Based on the distribution of the results, the median was chosen as a descriptive statistic.

Data Set, Sample Size, and Reference Test

The available data set consisted of 1,625 dental bitewing radiographs with a maximum of 8 to 9 teeth per image and is described in detail in the Appendix. Tooth structures visible on bitewing radiographs (namely, enamel, dentin, the pulp cavity, and nonnatural "structures" like fillings and crowns) were annotated in a pixel-wise fashion (as masks) by 1 dental expert. These masks represent the ground truth for each data sample. In a second iteration, those annotations were reviewed by another dental expert for validity and correctness. Each annotator independently assessed each image using an in-house custom-built annotation tool described in Ekert et al. (2019). All examiners were calibrated and advised on how to perform the segmentation. Images with implants, bridges, or root canal fillings were very rare (<1%) and therefore excluded.

Notably, enamel, dentin, and pulpal areas were present in every radiograph, while fillings and crowns were only available in 80% and 20% of images, respectively. Images and segmentation masks were resized to a resolution of 224 × 224 to provide a fixed input size of the images as required by the model architectures.

Models and Training

As represented in Figure 1, models were built by combining different model architectures (U-Net, U-Net++, FPN, LinkNet, PSPNet, MAnet) with backbones from 3 different families (ResNet, VGG, DenseNet) of different depths (ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, VGG13, VGG16, VGG19, DenseNet121, DenseNet161, DenseNet169, DenseNet201). This led to a total of 72 model designs, which were each initialized with 3 different strategies (random, ImageNet, CheXpert), resulting into 216 trained models in total. All models were trained under a 5-fold cross-validation scheme, where the combination of samples in training, validation, and test set was varied for each fold to achieve a reasonable estimate of the model performance independent from the data split. Details on training are described in the Appendix.

Statistical Analysis

Model configurations with respect to initialization strategies and architectures were ranked according to their median F1-score and formally tested for differences between configurations with the nonparametric Wilcoxon rank-sum test. The nonparametric Spearman's rank-order correlation was estimated to determine the relationship between complexity and model performance (F1-score). To account for multiple comparisons, we adjusted the P values using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). P values below 0.05 were considered statistically significant. The number of pairwise comparisons C of conditions k was computed via equation (1).

$$C = \frac{k(k-1)}{2} \qquad (1)$$

Results

Figure 2 presents an overview of segmentation outputs generated by different model architectures in comparison to the ground truth. Figure 3 shows the F1-scores of different model

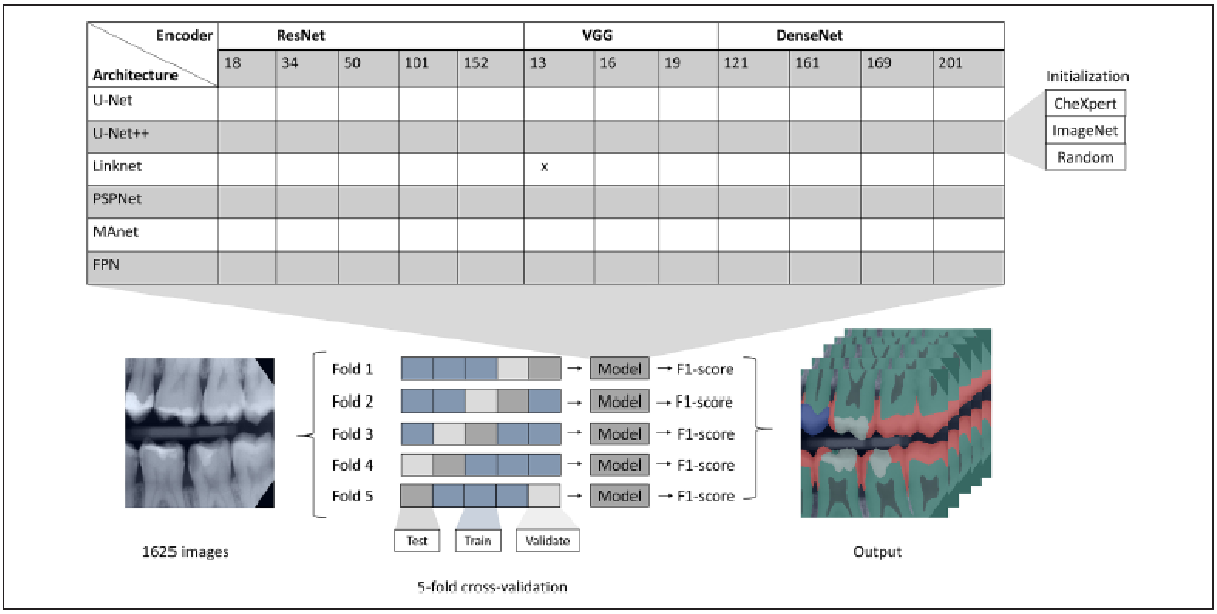

Figure 1. Illustration of the study design. Model setups were based on different architectures, encoder backbones, and initialization strategies (top) and 5-fold cross-validation with varying train, validation, and test sets for each fold (bottom). Exemplary bitewing radiograph (left) and tooth structure components overlaid on an input image (right).

configurations grouped by architecture, backbone family, and initialization strategy.

(1) **Architecture:** Out of 15 pairwise comparisons of model architectures, 14 turned out to be statistically significantly different. U-Net++, U-Net, and LinkNet achieved a median (interquartile range [IQR]) F1-score of 0.86 (0.85, 0.87), (0.84, 0.86), and (0.85, 0.88), respectively, and outperformed MAnet, PSPNet, and FPN with statistical significance. Backbones from the VGG and DenseNet group reached a median (IQR) of 0.85 (0.83, 0.86) and (0.81, 0.86), respectively, while the ResNet group reached a median (IQR) F1-score of 0.84 (0.81, 0.86). Models with backbones from the VGG group outperformed models with backbones of the ResNet group with statistical significance.

(2) **Complexity:** We found a statistically significant weak positive monotonic relationship between the network size and its performance with $r = 0.32$ ($P < 0.001$).

(3) **Initialization:** Different initialization strategies computed over all architectures and backbones achieved F1-scores of 0.86 (0.83, 0.87) (ImageNet), 0.86 (0.83, 0.87) (CheXpert), and 0.83 (0.77, 0.84) (random initialization). Models initialized with ImageNet or CheXpert outperformed models initialized with random weights ($P_{\text{ImageNet}} < 0.001$, $P_{\text{CheXpert}} < 0.001$). No significant difference was observed between ImageNet and CheXpert ($P = 0.85$).

(4) **Class imbalances:** In a sensitivity analysis, the model performance was evaluated on the minority classes of filling (80%) and crown (20%). In general, models' performance was inversely related to class frequencies (Fig. 4).

(4.1) **Architecture:** Models based on a VGG backbone outperformed models with a ResNet backbone on the minority classes of filling ($P = 0.009$) and crown ($P = 0.013$). Notably, there was no statistical difference between the 3 backbones on the majority classes of pulpal cavity and dentin.

(4.2) **Complexity:** We found a statistically significant weak positive monotonic relationship between the network size and its performance for class dentin ($r = 0.245$, $P < 0.001$), enamel ($r = 0.239$, $P < 0.001$), filling ($r = 0.195$, $P = 0.004$), pulpa ($r = 0.218$, $P < 0.001$), and class crown ($r = 0.154$, $P < 0.023$).

(4.3) **Initialization:** Models with ImageNet and CheXpert initialization consistently outperformed models with random initialization. There was no statistically significant difference between ImageNet and CheXpert initializations.

Discussion

We benchmarked 216 models defined by their architecture, complexity, and initialization strategy on a tooth structure segmentation task of dental bitewing radiographs. Several findings require a more detailed discussion.

First, we aimed to evaluate whether there are superior model architectures for the tooth segmentation task at hand. We discovered a performance advantage of models with backbones from the VGG family over models with backbones from the ResNet family. Our findings are consistent with those from Ke et al. (2021), who reported that architecture improvements reported on ImageNet may not always be translated to performances on medical imaging tasks. New model architectures and model improvements seem to be prone to overfitting on ImageNet data sets. Hence, transferability of newest AI research results into other domains, here the dental domain, may not be guaranteed.

The statistically significant performance advantage of models with VGG encoder backbones plead for the usage of VGG encoders, when solid baseline models are required, which perform reasonably well across different model configurations and settings. This may be relevant for the implementation of proof of concepts, for example. The top 10 performing models on the tooth structure segmentation task were built with backbones from the ResNet and DenseNet family. Consequently, if the focus is on model performance, it seems warranted to invest time to find an optimal model configuration based on more complex models (e.g., from the ResNet family). If, however, the validation of general concepts or benchmarking is the focus of the study, VGG-based models seem a reasonable choice as they are more robust across model configurations.

Second, one of our objectives evolved around the effect of the model complexity on the model performance. One of the key findings was a weak positive relationship between model depth and model performance. Therefore, we accept our hypothesis. Notably, however, the number of parameters increased in large steps, with only incremental improvements of model performance. Hence, the performance improvement was oftentimes disproportionate to the increasing demands for computational resources, training time, or the need to reduce image resolutions. The largest network in the present study was MAnet combined with a ResNet152 backbone, which reached an F1-score of 0.85 (0.85, 0.85) over all folds (ImageNet initialization). LinkNet in combination with a ResNet50 backbone was 5 times smaller but reached an F-score of 0.88 (0.88, 0.88) in comparison. It should be highlighted that lower computational costs allow for input imagery of higher resolution, which may be relevant for many dental applications.

Our third objective, aimed to give insights whether initializing with ImageNet or CheXpert, is consistently superior even when there is a difference in performance between both initialization strategies. We found statistically significant performance boosts for models initialized with ImageNet or CheXpert weights in comparison to a random initialization. These findings are consistent with those from Ke et al. (2021), who reported that 12 of 16 architectures benefited from an initialization with ImageNet weights for a classification task of chest radiographs. The comparison of ImageNet and CheXpert initialization showed no significant differences.

Fourth, we additionally found predictions on the minority class of filling (80%) to be generally more stable over different model configurations than predictions on class crowns (20%). Our results showed that there are superior architectures for segmenting minority classes (e.g., U-Net, U-Net++, LinkNet), but choosing a reasonable architecture may not be sufficient to

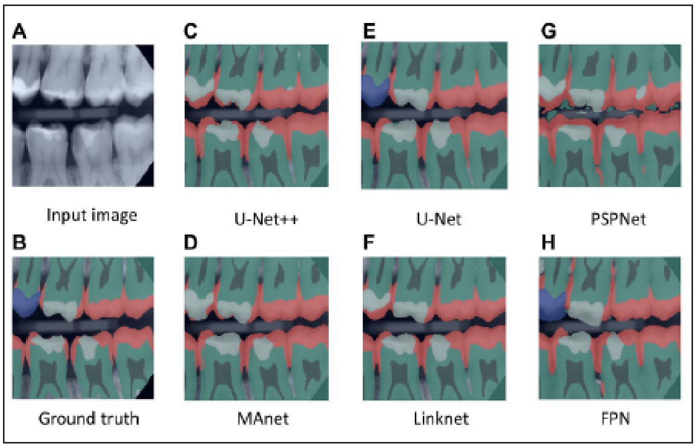

Figure 2. Examples of segmented bitewing radiographs. (**A**) Naive input image. (**B**) Ground truth and (**C–H**) output of tooth structure segmentation by different model architectures. The red, dark green, light green, gray, and blue colors indicate enamel, pulp cavity and root canals, dentin, filling, and crown classes, respectively. All models in this example were built with a ResNet50 backbone and initialized with pretrained CheXpert weights. This figure is available in color online.

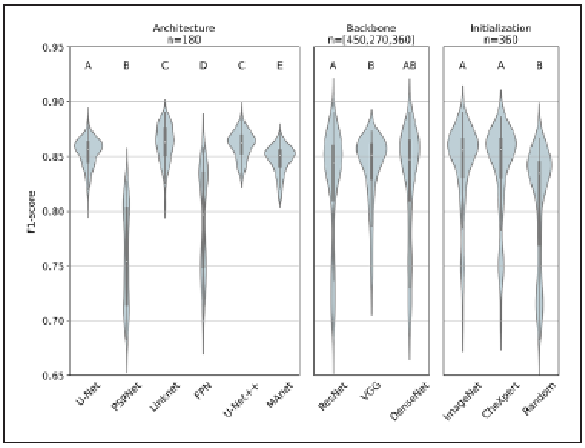

Figure 3. F1-scores stratified by initialization strategy, architecture, and backbone family based on sample sizes *n*. Median, interquartile range, and 95% confidence interval are represented by the white dot, the black box, and the black line, respectively. Different superscript letters indicate statistically significant difference (e.g., between U-Net and LinkNet), while the same superscript letters represent no significant difference (e.g., between LinkNet and U-Net++) (see Appendix for more details).

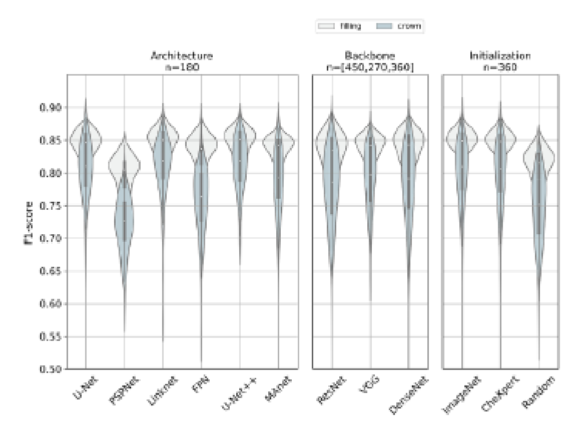

Figure 4. F1-scores of different models in the minority classes, filling (white) and crown (steel blue), respectively. We stratified the analyses by initialization strategy, architecture, and backbone family. Median, interquartile range, and 95% confidence interval are represented by the white dot, the black box, and the black line, respectively. Results are based on a sample size *n*. This figure is available in color online.

overcome class imbalance. Hence, it could be recommended to address this problem with weighted loss functions (Guerrero-Penã et al. 2018) or oversampling (Buda et al. 2018).

This study comes with several limitations. First, our results were based on 1 specific DL task, a tooth structure segmentation on bitewing radiographs, and are limited to the examined model architectures. Hence, we do not claim generalizability of our findings across other segmentation tasks or over all existing model architectures. Second, images of our data set originate from varying machines, which may lead to different behavior of the models. Furthermore, radiographs with bridges, implants, and root canal fillings were not considered in the present study as they were very rare. We accept this as our aim was to benchmark models and not to build clinically useful ones in this study. In line with this, we were only aiming at a model comparison instead of proposing a high-precision model. Hence, we did not take any actions against the existing class imbalance and did not perform an extensive hyperparameter search. Finally, we based our analysis of the relationship between model performances and model complexity exclusively on the number of model parameters. It may be the case that model architectures with more parameters require less computational power through more efficient structures of layers. Furthermore, we did not evaluate the effect of minor differences in performance within the dental environment or how computational resources are affected by differences in the number of parameters of the models.

Conclusion

We benchmarked different configurations of DL models based on their architecture, backbone, and initialization strategy regarding their performance on a tooth structure segmentation task of dental bitewing radiographs to provide guidance for researchers in their DL model selection process. Regarding the superiority of certain model architectures, we found that VGG backbones provided solid baseline models across different model configurations, while peak performances were reached through combinations of U-Net++, LinkNet, and ResNet or DenseNet encoders. Superior architectures did not overcome class imbalance. Models known to perform better than others on a nondental data set like ImageNet did not demonstrate such superiority on our dental imaging task. The analysis of the relationship between model complexity and performance showed that deeper models did not necessarily perform better than shallow alternatives with lower demands in computational resources. Finally, we found that transfer learning boosts model performance, independent of the origin of transferred knowledge.

Author Contributions

L. Schneider, contributed to conception, design, data analysis, and interpretation, drafted and critically revised the manuscript; L. Arsiwala-Scheppach, contributed to analysis, critically revised the manuscript; J. Krois, contributed to conception, design, and data analysis, drafted and critically revised the manuscript; H. Meyer-Lueckel, contributed to interpretation, critically revised the manuscript; K.K. Bressem, contributed to acquisition and interpretation, critically revised the manuscript; S.M. Niehues, contributed to acquisition, critically revised the manuscript; F. Schwendicke, contributed to conception, design, data acquisition, and interpretation, drafted and critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: F. Schwendicke and J. Krois are cofounders of the dentalXrai Ltd., a startup. dentalXrai Ltd. did not have any role in conceiving, conducting, or reporting this study.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

L. Schneider https://orcid.org/0000-0002-4431-2669
L. T. Arsiwala https://orcid.org/0000-0002-1428-6543
J. Krois https://orcid.org/0000-0002-6010-8940
F. Schwendicke https://orcid.org/0000-0003-1223-1669

References

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Method. 57(1):289–300.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, De Vet HC, et al. 2015. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem. 61(12):1446–1452.

Bressem KK, Adams LC, Erxleben C, Hamm B, Niehues SM, Vahldiek JL. 2020. Comparing different deep learning architectures for classification of chest radiographs. Sci Rep. 10(1):13590.

Buda M, Maki A, Mazurowski MA. 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 106:249–259.

Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, Elhennawy K, Schwendicke F. 2020. Detecting caries lesions of different radiographic extension on bitewings using deep learning. J Dent. 100:103425.

Cejudo JE, Chaurasia A, Feldberg B, Krois J, Schwendicke F. 2021. Classification of dental radiographs using deep learning. J Clin Med. 10(7):1496.

Cha JY, Yoon HI, Yeo IS, Huh KH, Han JS. 2021. Panoptic segmentation on panoramic radiographs: deep learning–based segmentation of various structures including maxillary sinus and mandibular canal. J Clin Med. 10(12):2577.

Chaurasia A, Culurciello E. 2017. Linknet: exploiting encoder representations for efficient semantic segmentation. In: IEEE Visual Communications and Image Processing (VCIP); December 10–13, 2017; St. Petersburg, FL. doi:10.1109/vcip.2017.8305148

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. 2009. Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition; June 20–25, 2009; Miami, FL. p. 248–255. doi:10.1109/CVPR.2009.5206848

Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F. 2019. Deep learning for the radiographic detection of apical lesions. J Endod. 45(7):917–922.

Fan T, Wang G, Li Y, Wang H. 2020. Ma-net: a multi-scale attention network for liver and tumor segmentation. IEEE Access. 8:179656–179665.

Forman G, Scholz M. 2010. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. ACM SIGKDD Explor Newsl. 12(1):49–57.

Guerrero-Penã FA, Marrero Fernandez PD, Ing Ren T, Yui M, Rothenberg E, Cunha A. 2018. Multiclass weighted loss for instance segmentation of cluttered cells. In: 25th IEEE International Conference on Image Processing (ICIP); October 7–10, 2018; Athens, Greece. p. 2451–2455. doi:10.1109/ICIP.2018.8451187

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27–30, 2016; Las Vegas, NV. p. 770–778. doi:10.1109/CVPR.2016.90

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 2017. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21–26, 2017; Honolulu, HI. p. 2261–2269. doi:10.1109/CVPR.2017.243

Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K, et al. 2019. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. Proc Conf AAAI Artif Intell. 33(1):590–597.

Jafari MH, Karimi N, Nasr-Esfahani E, Samavi S, Soroushmehr SMR, Ward K, Najarian K. 2016. Skin lesion segmentation in clinical images using deep learning. Presented at: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE; Cancún, Mexico. p. 337–342. doi: 10.1109/ICPR.2016.7899656.

Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, Marx A, Boor P, Tacke F, Neumann UP, Grabsch HI, Yoshikawa T, Brenner H, Chang-Claude J, Hoffmeister M, Trautwein C, Luedde T. 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 25(7):1054–1056.

Ke A, Ellsworth W, Banerjee O, Ng AY, Rajpurkar P. 2021. CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-ray interpretation. In: Proceedings of the Conference on Health, Inference, and Learning; April 8, 2021. New York (NY): Association for Computing Machinery. p. 116–124. doi:10.1145/3450439.3451867

Kim J, Lee HS, Song IS, Jung KH. 2019. DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. Sci Rep. 9(1):1–9.

Kirillov A, Girshick R, He K, Dollár P. 2019. Panoptic feature pyramid networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 6392–6401. doi:10.1109/CVPR. 2019.00656. https://arxiv.org/abs/1901.02446.

Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, Dörfer C, Schwendicke F. 2019. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 9(1):1–6.

Lee JH, Kim DH, Jeong SN, Choi SH. 2018. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 77:106–111.

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 5(4):115–133.

Orhan K, Bayrakdar I, Ezhov M, Kravtsov A, Ozyürek T. 2020. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J. 53(5):680–689.

Ronneberger O, Fischer P, Brox T. 2015a. Dental X-ray image segmentation using a U-shaped deep convolutional network. In: International Symposium on Biomedical Imaging; Brooklyn, NY. Vol. 1. p. 3. http://www-o.ntust.edu.tw/~cweiwang/ISBI2015/challenge2/isbi2015_Ronneberger.pdf

Ronneberger O, Fischer P, Brox T. 2015b. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention (MICCAI); Lecture Notes in Computer Science. Vol. 9351. Cham (Switzerland): Springer. p. 234–241.

Schwendicke F, Golla T, Dreher M, Krois J. 2019. Convolutional neural networks for dental image diagnostics: a scoping review. J Dent. 91:103226.

Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J; IADR e-Oral Health Network and the ITU WHO Focus Group AI for Health. 2021. Artificial intelligence in dental research: checklist for authors, reviewers, readers. J Dent. 107:103610.

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. arXiv preprint. https://arxiv.org/abs/1409.1556.

Son J, Shin JY, Kim HD, Jung KH, Park KH, Park SJ. 2020. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. Ophthalmology. 127(1):85–94.

Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. 2018. A survey on deep transfer learning. In: Kurková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, editors. Artificial Neural Networks and Machine Learning–ICANN 2018. Cham (Switzerland): Springer International Publishing. p. 270–279.

Yakubovskiy P. 2020. Segmentation models pytorch. pytorch [accessed 2021 April 22]. https://segmentation-modelspytorch.readthedocs.io/en/latest/.

Zhao H, Shi J, Qi X, Wang X, Jia J. 2017. Pyramid scene parsing network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21–26, 2017; Honolulu, HI. p. 6230–6239. doi:10.1109/CVPR.2017.660

Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. 2018. Unet++: a nested U-net architecture for medical image segmentation. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, Tavares JMR, Bradley A, Papa JP, Belagiannis V, et al., editors. Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham (Switzerland): Springer International. p. 3–11. doi:10.1007/978-3-030-00889-51.

Article titled “Artificial intelligence for caries detection: Randomized trial”

Source reference: S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849.

DOI link: https://doi.org/10.1016/j.jdent.2021.103849

Article titled “Artificial intelligence for caries detection: Randomized trial” contd.

Article titled “Artificial intelligence for caries detection: Randomized trial” contd.

Article titled “Artificial intelligence for caries detection: Randomized trial” contd.

Article titled “Artificial intelligence for caries detection: Randomized trial” contd.

Article titled “Artificial intelligence for caries detection: Randomized trial” contd.

Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my dissertation for reasons of data protection.

Curriculum Vitae contd.

Abstract of Master thesis

Periodontal disease measures and risk of incident peripheral artery disease: The Atherosclerosis Risk in Communities (ARIC) Study

Background: The association of periodontal disease with atherosclerotic cardiovascular diseases is well known, but not specifically with incident peripheral artery disease (PAD). Therefore, we studied the associations of periodontal disease with incident PAD in a population-based setting.

Methods: Among 9,793 participants (aged 53-75 years) without prevalent PAD, self-reported history of periodontal disease was ascertained. Of these, 5,872 participants underwent full-mouth examinations from which periodontal status was defined using the US Centers for Disease Control and Prevention-American Academy of Periodontology (CDC-AAP) definition. We quantified the association of periodontal disease with incident PAD (defined by hospital admission diagnosis or procedures) using multivariable Cox regression models.

Results: During a median follow-up of 20.1 years, 360 participants (3.6%) developed PAD. In models accounting for potential confounders including diabetes and smoking pack-years, there was higher hazard of PAD in participants with self-reported tooth loss because of periodontal disease (hazard ratio:1.54 [95% CI:1.20-1.98]), history of periodontal disease treatment (1.37 [1.05-1.80]), and periodontal disease diagnosis (1.38 [1.09-1.74]), compared to their respective counterparts. The clinical measure of periodontal disease (n = 5,872) was not significantly associated with incident PAD in the fully adjusted model (e.g., 1.53 [0.94-2.50] in CDC-AAP-defined severe periodontal disease versus no disease).

Conclusion: We observed a modest association of self-reported periodontal disease, especially when resulting in tooth loss, with incident PAD in the general population. Nonetheless, a larger study with the clinical measure of periodontal disease is warranted.

Citation: Arsiwala LT, Mok Y, Yang C, Ishigami J, Selvin E, Beck JD, Allison MA, Heiss G, Demmer RT, Matsushita K. Periodontal disease measures and risk of incident peripheral artery disease: The Atherosclerosis Risk in Communities (ARIC) Study. J Periodontol. 2022 Jul;93(7):943-953. doi: 10.1002/JPER.21-0342.

Publication list

Note. The applicant had a change of family name in the year 2022 from ‘Arsiwala’ to ‘Arsiwala-Scheppach’.

First-author publications:

1. **L.T. Arsiwala-Scheppach**, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine Learning in Dentistry: A Scoping Review, J Clin Med 12(3) (2023) doi: 10.3390/jcm12030937.
   Impact factor: 4.9
   Selected as the Editor's Choice Article in Q2 2023

2. **L.T. Arsiwala-Scheppach**, N. Castner, C. Rohrer, S. Mertens, E. Kasneci, J. de Oro, J. Krois, F. Schwendicke, Gaze Patterns of Dentists while Evaluating Bitewing Radiographs, J Dent (2023) 104585, doi: 10.1016/j.jdent.2023.104585.
   Impact factor: 4.9

3. **L.T. Arsiwala**, Y. Mok, C. Yang, J. Ishigami, E. Selvin, J.D. Beck, M.A. Allison, G. Heiss, R.T. Demmer, K. Matsushita, Periodontal disease measures and risk of incident peripheral artery disease: The Atherosclerosis Risk in Communities (ARIC) Study, J Periodontol 93(7) (2022) 943-953, doi: 10.1002/JPER.21-0342.
   Impact factor: 4.5

4. **L.T. Arsiwala-Scheppach**, P.Y. Ramulu, A.R. Sharrett, V. Kamath, J.A. Deal, X. Guo, S. Du, E.E. Garcia Morales, A. Mihailovic, H. Chen, A.G. Abraham, Associations among Visual, Auditory, and Olfactory Functions in Community-Based Older Adults: The Atherosclerosis Risk in Communities (ARIC) Study, Transl Vis Sci Technol 11(11) (2022) 2, doi: 10.1167/tvst.11.11.2.
   Impact factor: 3.0

5. **L.T. Arsiwala-Scheppach**, P.Y. Ramulu, A.G. Abraham, An Eye on the Brain: Insights from the Eye Determinants of Cognition Study, J Gerontol Geriatr Med 8: 132 (2022), doi: 10.24966/GGM-8662/100132.
   Impact factor: 1.5

6. **L.T. Arsiwala**, X. Guo, P.Y. Ramulu, A.R. Sharrett, A. Mihailovic, B.K. Swenor, T. Mosley, Y. Dong, A.G. Abraham, Associations of Visual Function with Cognitive Performance in Community-Based Older Adults: The Eye Determinants of Cognition (EyeDOC) Study, The Journals of Gerontology: Series A (2021) doi: 10.1093/gerona/glab349.
   Impact factor: 5.1

Co-author publications:

7. F. Schwendicke, **L.T. Arsiwala-Scheppach**, J. Krois, Artificial intelligence: What it is and what it can do for dentists, Dental Update 50(4) (2023) 245-249, doi: 10.12968/denu.2023.50.4.245.
   Impact factor: 0.2

8. N. Pan-Doh, X. Guo, **L.T. Arsiwala-Scheppach**, K.A. Walker, A.R. Sharrett, A.G. Abraham, P.Y. Ramulu, Associations of Midlife and Late-Life Blood Pressure Status With Late-Life Retinal OCT Measures, Transl Vis Sci Technol 12(2) (2023) 3, doi: 10.1167/tvst.12.2.3.
   Impact factor: 3.0

9. E. Cha, **L.T. Arsiwala-Scheppach**, D. Srikumaran, C.R. Prescott, Patient Utilization of Premium Intraocular Lenses Before and During the COVID-19 Pandemic, Eye Contact Lens 49(7) (2023) 292-295, doi: 10.1097/ICL.0000000000001000.
   Impact factor: 2.3

10. J.B. Scheppach, A. Wu, R.F. Gottesman, T.H. Mosley, **L.T. Arsiwala-Scheppach**, D.S. Knopman, M.E. Grams, A.R. Sharrett, J. Coresh, S. Koton, Association of Kidney Function Measures With Signs of Neurodegeneration and Small Vessel Disease on Brain Magnetic Resonance Imaging: The Atherosclerosis Risk in Communities (ARIC) Study, Am J Kidney Dis 81(3) (2023) 261-269 e1, doi: 10.1053/j.ajkd.2022.07.013.
    Impact factor: 13.2

11. S.S. Ong, I. Ahmed, A. Gonzales, U.T. Aguwa, B. Beatson, X. Dai, A.T. Pham, Y.S. Shah, A. Zhou, **L.T. Arsiwala**, J. Wang, J.T. Handa, Management of uncomplicated rhegmatogenous retinal detachments: A comparison of practice patterns and clinical

outcomes in a real-world setting, Eye (Lond) 37(4) (2023) 684-691, doi: 10.1038/s41433-022-02028-z.
Impact factor: 4.0

12. L. Schneider, **L. Arsiwala-Scheppach**, J. Krois, H. Meyer-Lueckel, K.K. Bressem, S.M. Niehues, F. Schwendicke, Benchmarking Deep Learning Models for Tooth Structure Segmentation, J Dent Res 101(11) (2022) 1343-1349, doi: 10.1177/00220345221100169.
Impact factor: 8.9

13. J.A. Rodrigues, **L.T. Arsiwala-Scheppach**, J. Krois, F. Schwendicke, Augmented Vision for Dental Students' Education in Detecting Proximal Carious Lesions on Bitewing Radiographs: A Randomized Controlled Trial, Caries Res 56(3) (2022) 197-205, doi: 10.1159/000525777.
Impact factor: 3.9

14. E. Burton, **L.T. Arsiwala**, T.V. Johnson, D. Srikumaran, S. Zafar, F.A. Woreta, Applicant Characteristics Associated with Glaucoma Fellowship Match from 2010 to 2017, Ophthalmol Glaucoma 5(2) (2022) 233-240, doi: 10.1016/j.ogla.2021.08.004.
Impact factor: 2.9

15. B.C. Tsou, U.T. Aguwa, **L.T. Arsiwala**, E. Burton, K. Mishra, S. Zafar, F. Woreta, Trends in Cornea Fellowship Applications and Applicant Characteristics: A San Francisco Match Analysis, Journal of Academic Ophthalmology 14(02) (2022) e216-e223, doi: 10.1055/s-0042-1756199.
Impact factor: 0.7

16. E. Burton, K. Mishra, **L.T. Arsiwala**, S. Zafar, G.A. Justin, P. Mruthyunjaya, F.A. Woreta, Applicant Characteristics Associated with Retina Fellowship Match from 2010-2019, Ophthalmol Retina 6(12) (2022) 1253-1259, doi: 10.1016/j.oret.2022.06.008.
Impact factor: 4.5

17. Y. Dong, X. Guo, **L.T. Arsiwala-Scheppach**, A.R. Sharrett, P.Y. Ramulu, A. Mihailovic, N. Pan-Doh, T. Mosley, J. Coresh, A.G. Abraham, Association of Optical Coherence Tomography and Optical Coherence Tomography Angiography Retinal

Features With Visual Function in Older Adults, JAMA Ophthalmol 140(8) (2022) 809-817, doi: 10.1001/jamaophthalmol.2022.2099.
Impact factor: 8.1

18. S.S. Ong, I. Ahmed, A. Gonzales, A.S. Al-Fakhri, H.F. Al-Subaie, F.S. Al-Qhatani, S.M. Alsulaiman, M. Mura, M. Maia, D.A. Kondo Kuroiwa, N.T. Maia, M.H. Berrocal, L. Wu, M. Zas, J.P. Francos, J.M. Cubero-Parra, **L.T. Arsiwala**, J.T. Handa, J.F. Arevalo, Vitrectomy versus Vitrectomy with Scleral Buckling in the Treatment of Giant Retinal Tear Related Retinal Detachments: An International Multicenter Study, Ophthalmol Retina 6(7) (2022) 595-606, doi: 10.1016/j.oret.2022.03.004.
Impact factor: 4.5

19. F. Schwendicke, **L.T. Arsiwala**, J. Krois, A. Baumer, B. Pretzl, P. Eickholz, H. Petsos, T. Kocher, B. Holtfreter, C. Graetz, Association, prediction, generalizability: Cross-center validity of predicting tooth loss in periodontitis patients, J Dent 109 (2021) 103662, doi: 10.1016/j.jdent.2021.103662.
Impact factor: 4.9

20. X. Guo, **L.T. Arsiwala**, Y. Dong, A. Mihailovic, P.Y. Ramulu, A.R. Sharrett, T. Mosley, A.G. Abraham, Visual Function, Physical Function, and Activities of Daily Living in Two Aging Communities, Transl Vis Sci Technol 10(14) (2021) 15, doi: 10.1167/tvst.10.14.15.
Impact factor: 3.0

21. S. Mertens, J. Krois, A.G. Cantu, **L.T. Arsiwala**, F. Schwendicke, Artificial intelligence for caries detection: Randomized trial, J Dent 115 (2021) 103849, doi: 10.1016/j.jdent.2021.103849.
Impact factor: 4.9

22. F. Schwendicke, A. Chaurasia, **L. Arsiwala**, J.H. Lee, K. Elhennawy, P.G. Jost-Brinkmann, F. Demarco, J. Krois, Deep learning for cephalometric landmark detection: Systematic review and meta-analysis, Clin Oral Investig 25(7) (2021) 4299-4309, doi: 10.1007/s00784-021-03990-w.
Impact factor: 3.6

23. L. Schneider, P. Dave, **L. Arsiwala-Scheppach**, F. Schwendicke and Joachim Krois. 2021. Exploring bias in F-score computation methods of multi-class segmentation models. In 2021 The 5th International Conference on Video and Image Processing (ICVIP 2021), December 22-25, 2021, Guangzhou, China. ACM, New York, NY, USA, 14 Pages. https://doi.org/10.1145/3511176.3511189.
Impact factor: Not applicable

24. A.G. Abraham, X. Guo, **L.T. Arsiwala**, Y. Dong, A.R. Sharrett, D. Huang, Q. You, L. Liu, B.J. Lujan, A. Tomlinson, T. Mosley, J. Coresh, Y. Jia, A. Mihailovic, P.Y. Ramulu, Cognitive decline in older adults: What can we learn from optical coherence tomography (OCT)-based retinal vascular imaging? J Am Geriatr Soc 69(9) (2021) 2524-2535, doi: 10.1111/jgs.17272.
Impact factor: 6.3

Conference abstracts

1. **L.T. Arsiwala-Scheppach**; Gaze pattern analysis and Artificial Intelligence in dentistry. ITU/WHO Focus Group on Artificial Intelligence for Health 2023, Meeting R: FGAI4H-R-040-A04.
https://www.itu.int/en/ITUT/focusgroups/ai4h/Documents/all/20230321-Meeting_R.htm

2. **L.T. Arsiwala-Scheppach**, A. Chaurasia, A. Mueller, J. Krois, F. Schwendicke; Machine Learning in Dentistry: Systematic Review. J Dent Res Vol 101 (Spec Iss C): O176 (2022 Pan European Region Oral Health Congress), https://iadr.abstractarchives.com/home.

3. **L.T. Arsiwala-Scheppach**; Impact of AI on gaze patterns of dentists. ITU/WHO Focus Group on Artificial Intelligence for Health 2022, Meeting P: FGAI4H-P-046-A07. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/all/20220919-Meeting_P.htm.

4. **L.T. Arsiwala**, C. Graetz, F. Schwendicke, J. Krois, A. Bäumer-König, B. Pretzl, P. Eickholz, H. Petsos, T. Kocher, B. Holtfreter; Cross-Center Validity and

Generalizability in Predicting Tooth Loss in Periodontitis. J Dent Res Vol 100 (Spec Iss B): 0043 (2021 Continental European and Scandinavian Divisions), https://iadr.abstractarchives.com/home.

5. **L.T. Arsiwala**, X. Guo, A.R. Sharrett, P.Y. Ramulu, A. Mihailovic, B.K. Swenor, Y. Dong, A. Abraham; Associations of visual function with cognitive measures in an older adult community-based cohort: The Eye Determinants of Cognition (EyeDOC) Study. *Invest. Ophthalmol. Vis. Sci.* 2021;62(8):3521.

6. **L.T. Arsiwala**, X. Guo, A.R. Sharrett, Y. Dong, E.E. Garcia, P.Y. Ramulu, J.A Deal, A. Abraham; Associations between visual and hearing function in an older adult population: The Eye Determinants of Cognition (EyeDOC) Study. *Invest. Ophthalmol. Vis. Sci.* 2020;61(7):2661.

7. **L.T. Arsiwala**, Y. Mok, C. Yang, J. Ishigami, K. Matsushita; Poster titled 'Periodontal disease measures and risk of peripheral artery disease: The Atherosclerosis Risk in Communities (ARIC) Study'. National Institute of Aging research showcase.

8. **L.T. Arsiwala**, V. Mave, M. Robinson; Poster titled 'Antimicrobial use and diagnosis among hospitalized febrile patients in India'. Global Health Day research showcase.

Acknowledgments

First and foremost, I express my heartfelt gratitude to my first supervisor, Professor Falk Schwendicke, for his continuous encouragement and unwavering support throughout the development of my research projects and the shaping of my dissertation. His availability to provide prompt feedback on my written drafts has been invaluable, and his commitment to perfection and patience made him the perfect supervisor for me.

I would also like to extend my gratitude to my second supervisor, Dr. Joachim Krois, for his support in technical areas, trust in my abilities, and enthusiastic explanations of unfamiliar concepts. Furthermore, I extend my sincere thanks to all the co-authors of the three publications included in this dissertation for their support and co-operation in completing the respective projects.

To my family and friends, I am deeply grateful for their patience and understanding throughout my doctoral journey and for their steadfast belief in me. A special mention goes to my husband who has consistently supported my professional aspirations and been my source of strength.