

Artificial intelligence (AI) in restorative dentistry: Performance of AI models designed for detection of interproximal carious lesions on primary and permanent dentition

DIGITAL HEALTH
Volume 9: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231216681
journals.sagepub.com/home/dhj



Amr Ahmed Azhari¹ , Narmin Helal², Leena M Sabri³
and Abeer Abduljawad³

Abstract

Objective: The objective of this study was to evaluate the effectiveness of deep learning methods in detecting dental caries from radiographic images.

Methods: A total of 771 bitewing radiographs were divided into two groups: adult (n = 554) and pediatric (n = 217). Two distinct semantic segmentation models were constructed for each group. They were manually labeled by general dentists for semantic segmentation. The inter-examiner reliability of the two examiners was also measured. Finally, the models were trained using transfer learning methodology along with computer science advanced tools, such as ensemble U-Nets with ResNet50, ResNext101, and Vgg19 as the encoders, which were all pretrained on ImageNet weights using a training dataset.

Results: Intersection over union (IoU) score was used to evaluate the outcomes of the deep learning model. For the adult dataset, the IoU averaged 98%, 23%, 19%, and 51% for zero, primary, moderate, and advanced carious lesions, respectively. For pediatric bitewings, the IoU averaged 97%, 8%, 17%, and 25% for zero, primary, moderate, and advanced caries, respectively. Advanced caries was more accurately detected than primary caries on adults and pediatric bitewings $P < 0.05$.

Conclusions: The proposed deep learning models can accurately detect advanced caries in permanent or primary bitewing radiographs. Misclassification mostly occurs between primary and moderate caries. Although the model performed well in correctly classifying the lesions, it can misclassify one as the other or does not accurately capture the depth of the lesion at this early stage.

Keywords

Deep learning, artificial intelligence, bitewings, caries, x-rays, dental caries, bioinformatics

Submission date: 20 March 2023; Acceptance date: 8 November 2023

Introduction

Dental caries is one of the most common chronic human diseases in the world¹; it is a multifactorial, infectious oral disease caused primarily by the complex interaction of cariogenic oral biofilm on the tooth surface with fermentable dietary carbohydrates over time.^{2a} The occurrence of dental caries is the ramification of the breakdown of food remains by bacteria. Acidic substances eventually degrade tooth structure, creating cavities. Two types of bacteria

¹Department of Restorative Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

²Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

³Internship Training Program, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding author:

Amr Ahmed Azhari, Department of Restorative Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia.
Email: aaaazhari@kau.edu.sa



are responsible for this phenomenon: streptococcus mutans and lactobacilli.^{2b} In severe cases, dental caries can lead to toothache, tooth structure loss, intra-oral abscess, or facial swelling.³ Although dental radiography (including panoramic, periapical, and bitewing views) and explorers (or dental probes), which are broadly utilized and respected to be exceptionally dependable indicative devices for the recognition of dental caries,⁴ they are subject to practitioner exposure, experience, and fatigue levels.

The importance of AI in caries detection lies in its potential to assist clinicians in detecting tooth caries quickly and reliably in routine clinical practice. The accuracy of early diagnosis of dental caries remains a challenge for dentists.⁵ Machine learning is a computational tool that utilizes algorithms and data inputs to self-improve and learn automatically via experience and exposure to a wide range of samples and variables. These algorithms are based on a specific computational model that directs them to retrieve results related to pre-determined tasks.^{7,8} Lee et al. developing a deep CNN model called GoogLeNet for detecting dental caries in periapical radiographs.⁹ The deep CNN algorithm demonstrated good detection and diagnostic performance, with molar models achieving diagnostic accuracies of 88.0%, 89.0%, and 82.0%, respectively. The premolar model provided the highest AUC, significantly higher than other models. Mao et al. developed a conventional network model called AlexNet for restoration and caries determinations, with accuracies of 95.56% and 90.30%, respectively.¹⁰ These studies highlight the potential of deep learning in dental caries detection and restoration.

Cantu et al. developed a CNN (U-Net) to evaluate deep learning models against individual dentists in detecting carious lesions¹¹ with mean accuracy of 0.80 compared to dentists' mean accuracy 0.71. Furthermore, Lian et al. developed CNN models, nnU-Net and DenseNet121, to detect caries lesions and classify radiographic extensions on panoramic films.¹² The results showed tantamount between expert dentists and neural networks. Moran et al. evaluated the effectiveness of deep CNN algorithms for detecting and diagnosing dental caries on periapical radiographs.¹³ Within 480 teeth images obtained, the CNN identified 18 incipient and 16 advanced lesions, with less experienced dentists reporting statistically indistinguishable results. Singh et al. developed a CNN-LSTM model using 1500 dental images as training data and 300 as testing data.¹⁴ The CNN-LSTM model demonstrated high accuracy and reliability. Likewise, Lee et al. developed a CNN model for early dental caries detection on bitewing radiographs, using 304 bitewing radiographs and 50 radiographs for training.¹⁵ The model's performance evaluation showed improved diagnostic accuracy, but more stable results are needed. Bayrakdar et al. explored the use of CNN-based AI algorithms for accurate tooth caries detection and segmentation in bitewing radiographs.¹⁷ They developed

automatic caries detection and segmentation models using VGG-16 and U-Net architecture, achieving sensitivity, precision, and F-measure rates of 0.84, 0.81, 0.84, 0.86, and 0.84, 0.84, respectively. However, the model faced limitations, such as being trained with the same parameters and having a smaller sample size. Casalegno et al.¹⁸ presented a deep-learning model based on CNN for automated tooth caries perception in NILT images, achieving an average IOU rate of 72.7%. Devito et al. evaluated the success of radiographic diagnosis of proximal caries using extracted teeth on bitewing radiographs, finding a diagnostic improvement of 39.4% using the AI model.

Because of the different anatomical morphologies of teeth and shapes of restorations, no substantial improvement can be achieved in the demonstrative strategy for distinguishing dental caries. Therefore, using deep learning in caries detection, dental professionals can potentially detect caries at an earlier stage, leading to more effective and less invasive treatment. It can also help to reduce the workload on dental professionals, allowing them to focus on other aspects of patient care.

Early stage caries detection interproximally can be missed by visual examination in both primary and permanent dentition. Consequently, it can progress into an irreversible situation that preventive measures can be difficult to demineralize it. Thus, artificial intelligence and machine learning have been used to increase the accuracy of dental practitioners' diagnosis.⁶ In this research paper, the objectives of the study were to evaluate the effectiveness of deep learning methods in the detection of dental caries from bitewing radiographic images collected from King AbdulAziz University Hospital data base in primary and permanent dentition and to determine whether the application of deep learning methods improved the dentists' accuracy in detecting proximal dental carious lesions from an intra-oral bitewings radiograph.

Materials and methods

Datasets

This research study was conducted at the Faculty of Dentistry of King Abdulaziz University Dental Hospital in Jeddah, Saudi Arabia. The bitewing radiographic image datasets were obtained from dental practitioners' historians and the Electronic Medical Record System (CS R4 Practice Management Software) of the University Hospital between August 2021 and April 2022. All images were obtained with permission from the radiology department of King Abdulaziz University Dental Hospital (proposal number.032-02-22). Labeling was performed after the dataset was collected. The dataset consisted of adult and pediatric bitewings, which included primary, moderate, and advanced dental caries. All images were selected using the following criteria.

1. The inclusion criteria were bitewing images, enamel and dentin carious lesions, and primary and permanent teeth.
2. The exclusion criteria were periapical images, overlapping images, radiographs with distortions and shadows, and images with full crowns only, as well as bridges.
3. Advanced caries: radiolucent, extending into the inner one-third of dentin D3.

Tools and environment

The obtained radiographic images were processed using Python in addition to Python-friendly environments such as Google Colab. Python is a widely used programming language. It is a complete language and platform that can be used for research and production system development. Google Colab enables programmers to write and execute arbitrary Python code through their web browsers.

It was used throughout this study because it is well-suited to machine learning and data analysis. Additionally, the Keras and Segmentation_models Library were used to develop the current model. The main features of this library can be summarized as having a high-level API, implying that complicated tasks can be performed with extremely few lines of code, which significantly reduces the time needed for code writing and enables the implementation of more experiments.

In addition, the library has four model architectures for binary and multiclass image segmentation, which have 25 available backbones for each architecture, all trained on image weights, and finally, it has a helpful set of segmentation losses (Jaccard, Dice, Focal) and metrics (IoU, F-score).

Cleaning, augmentation and labeling

Duplicated images as well as images where caries are challenging to identify because of technical errors that might interfere with caries identification were omitted from the dataset to ensure better accuracy. The process of image augmentation was applied to images with low brightness or contrast. Then, all images in the dataset were manually labeled by dentists.

Labeling is a process in which data are structured in a way that a computer can understand, and the labels correspond to the intended outcome of the machine learning model. Labeling tasks can be classified into four main types: categorization, segmentation, sequencing, and mapping.

Categorization refers to the task in which an image is assigned to a category, which could be binary or multiclass labeling; for example, classifying image data of moles into cancer or noncancer.

Segmentation is a task in which data are divided into segments. This approach can be applied to various data types. Using image data, segmentation identifies the pixels in an image that belong to a specific object or object class. For example, in a medical scan, segmentation can label different organs separately.

Criteria for excluding images from the training dataset were as the following: (1) Images with poor quality, such as those with artifacts, blurriness, or extreme overexposure, (2) Images that contain identifiable patient information to comply with privacy regulations and ethical standards, (3) Duplicate images was excluded to prevent overrepresentation and potential bias in the dataset. (4) Bitewings images with very low resolution, (5) Images with confounding factors, such as the presence of dental crowns/fixes dental prosthesis in all the teeth or orthodontic appliances that might interfere with caries detection.

The collected data were used in a manner consistent with ethical principles and legal requirements. Data were secured in a separate computer at the hospital and never accessed by any external sources. Participants were informed and consented during file opening stage to use their radiographic images for research and education purposes. The collected and labeled dataset consisted of nearly 554 adult bitewings and 217 pediatric bitewings. All images were converted from the original format to JPEG file format to unify the quality of the image processing procedure. The dataset was split into three subsets: training, validation, and test sets for adults as well as pediatrics. Using adult and pediatric bitewings to train the same model would lead to misclassification owing to differences in tooth density between adults and pediatrics. Therefore, we decided to train each individual group separately. The label distribution for each group is shown in the following figure (Figures 1–3):

After collecting the radiographic images, we measured the reliability of the two examiners, blinded from the study, to ensure that all examiners were calibrated by labeling the radiographic image and to avoid bias. Inter- and intra-rater reliability tests were performed between two examiners and graded by a restorative dentist consultant. Examiners were tested by labeling 10 radiographs per examiner. Radiographs were containing a mixture of primary, moderate and advanced caries and Kappa was over 0.8 between them. All images were manually labeled according to the International Caries Detection and Assessment System (ICIDAS) stratification according to lesion depth:

1. Primary caries: radiolucent extending into inner enamel junction E1, E2.
2. Moderate caries: radiolucent extending into the dentin-enamel junction and middle one-third of the dentin D1 and D2.

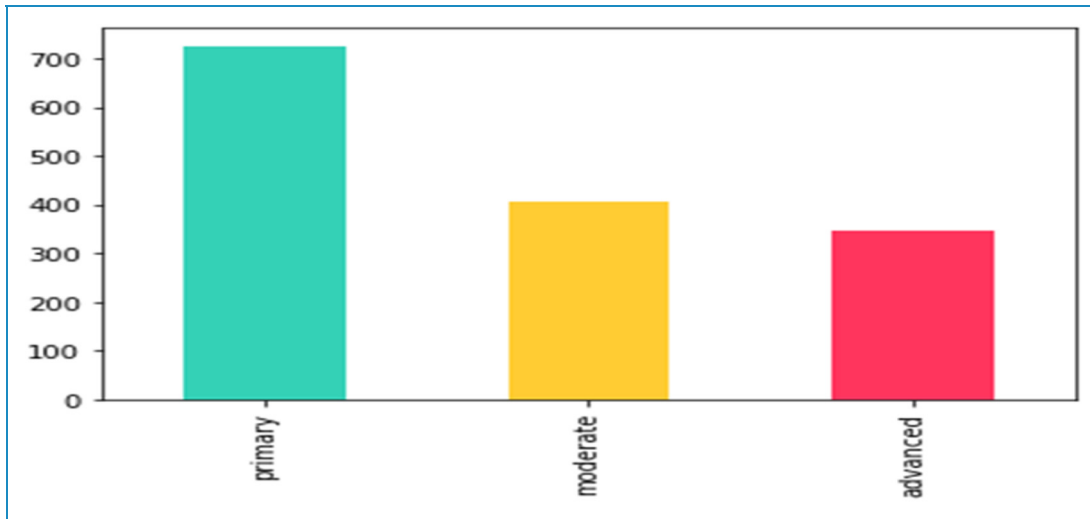


Figure 1. Adults' labels distribution: primary: 879, moderate: 559, advanced: 422. According to International Caries Detection and Assessment System (ICIDAS), primary caries is initial enamel caries, moderate caries is defined by the extension of proximal caries into the outer 2/3 of dentin and advanced caries is defined by the extension of proximal caries into the inner 1/3 of dentin.

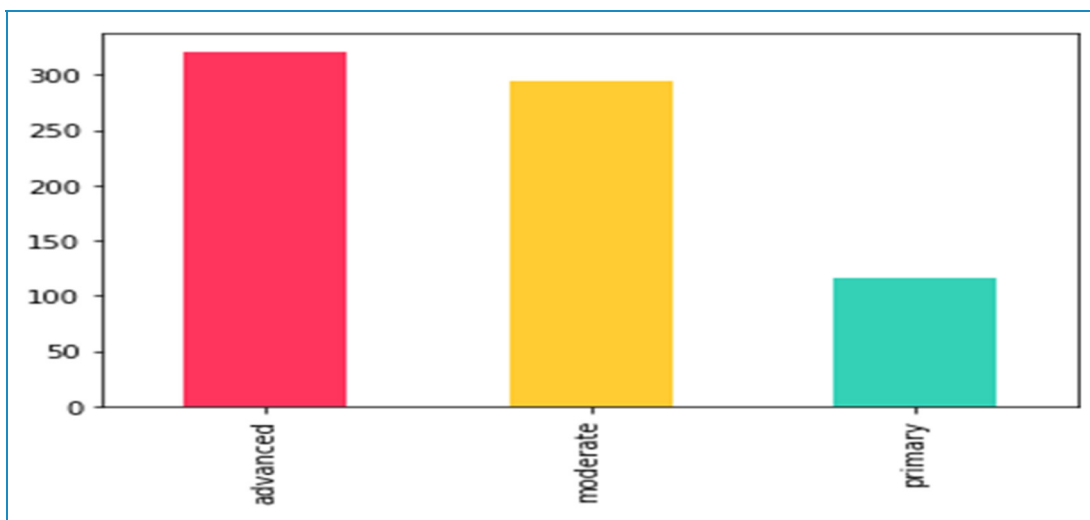


Figure 2. Pediatrics' labels distribution: primary: 116, moderate: 294, advanced: 321. According to the International Caries Detection and Assessment System (ICIDAS), primary caries is initial enamel caries, moderate caries is defined by the extension of proximal caries into the outer 2/3 of dentin and advanced caries is defined by the extension of proximal caries into the inner 1/3 of dentin.

Sequencing describes the progression of items in a series of data. This approach is particularly common when time-series modeling is used to predict future events.

Mapping operates by mapping one piece of data to another. This labeling technique is common in language-to-language translation, wherein a word in one language is mapped to a similar word in another language.

In this paper, segmentation was used to achieve the required outcome. Specifically, a type of segmentation tasks that is called "semantic segmentation" enables us to determine the location, size, and shape of an object in a given image. The goal of semantic image segmentation is to label each pixel of

an image with a corresponding label, as shown in the following labeled radiographic image (Figure 4).

Research model training

A semantic segmentation model was then developed. All training was performed within Google Colab using GPU runtime. Notably, the method of training is highly dependent on the size and quality of the data. Therefore, transfer learning, which is defined as the use of previous outcomes as a reference for future activities, was employed to obtain better accuracy with limited time and resources. To elaborate further, a

Primary Caries			Moderate Caries			Advanced Caries		
	Dentist (1)	Dentist (2)		Dentist (1)	Dentist (2)		Dentist (1)	Dentist (2)
X-ray 1	1	1	X-ray 1	3	3	X-ray 1	5	5
X-ray 2	0	0	X-ray 2	4	4	X-ray 2	4	4
X-ray 3	1	2	X-ray 3	2	2	X-ray 3	1	1
X-ray 4	1	1	X-ray 4	1	1	X-ray 4	1	1
X-ray 5	0	0	X-ray 5	1	1	X-ray 5	1	1
X-ray 6	0	0	X-ray 6	3	3	X-ray 6	1	1
X-ray 7	0	0	X-ray 7	1	1	X-ray 7	3	3
X-ray 8	2	2	X-ray 8	2	2	X-ray 8	0	0
X-ray 9	2	2	X-ray 9	1	1	X-ray 9	1	1
X-ray 10	4	4	X-ray 10	2	2	X-ray 10	0	0
Match		9	Match		10	Match		10
Total		10	Total		10	Total		10
interexaminer reliability		90%	interexaminer reliability		100%	interexaminer reliability		100%

Figure 3. Training specifications of reliability.

pretrained model, which is a saved network that was previously trained on a sufficiently large and general dataset, was utilized; this model effectively served as a generic model of the visual world. Finally, these learned feature maps can be leveraged without starting from scratch by training a large model on a large dataset.

In this study, training was performed using the U-Net. U-Net is an architecture that is used for semantic segmentation. This consists of a contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with stride 2 for downsampling. At each downsampling step, we doubled the number of feature channels. Every step in the expansive path consists of an up-sampling of the feature map followed by a 2×2 convolution (“upconvolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. Cropping is necessary because of the loss of border pixels in each convolution. In the final layer, a 1×1 convolution was used to map each 64-component feature vector to the desired number of classes. The network had 23 convolutional layers (Figure 5).

Prior to training the models, all the images were resized to 512×512 pixels. Additionally, the contrast was increased to enable primary caries to be easier to detect using the model.

The training specifications are presented in the following table (Table 1):

Hyperparameters are training variables that are set manually with a predetermined value before starting the training. Each hyperparameter is defined as follows.

1. Epochs: This is the number of times a network is to be trained during the entire training process; the training



Figure 4. Semantic image segmentation during labeling of bitewing radiographs; “blue” is early caries, “yellow” is moderate caries, and “red” is advanced caries.

should stop when the validation loss becomes significantly less than the training loss (overfitting).

2. Callbacks: Based on the Keras documentation, a callback is a set of functions to be applied at given stages of the training procedure. During training, callbacks can be used to obtain a view of the internal states and statistics of the model. These are defined to automate some of the training processes.
3. Learning Rate: The learning rate defines the rate at which a network updates its parameters. A Lower learning rate helps the model converge smoothly and may reduce overfitting, although it decelerates the training process. By contrast, a higher learning rate accelerates the training process but can miss the global minima and increase the chance of overfitting. A learning rate of 0.0001 was used for the previously stated reasons and the default value was 0.01.
4. Optimizer: Optimization is the process of finding a set of inputs for an objective function that results in a

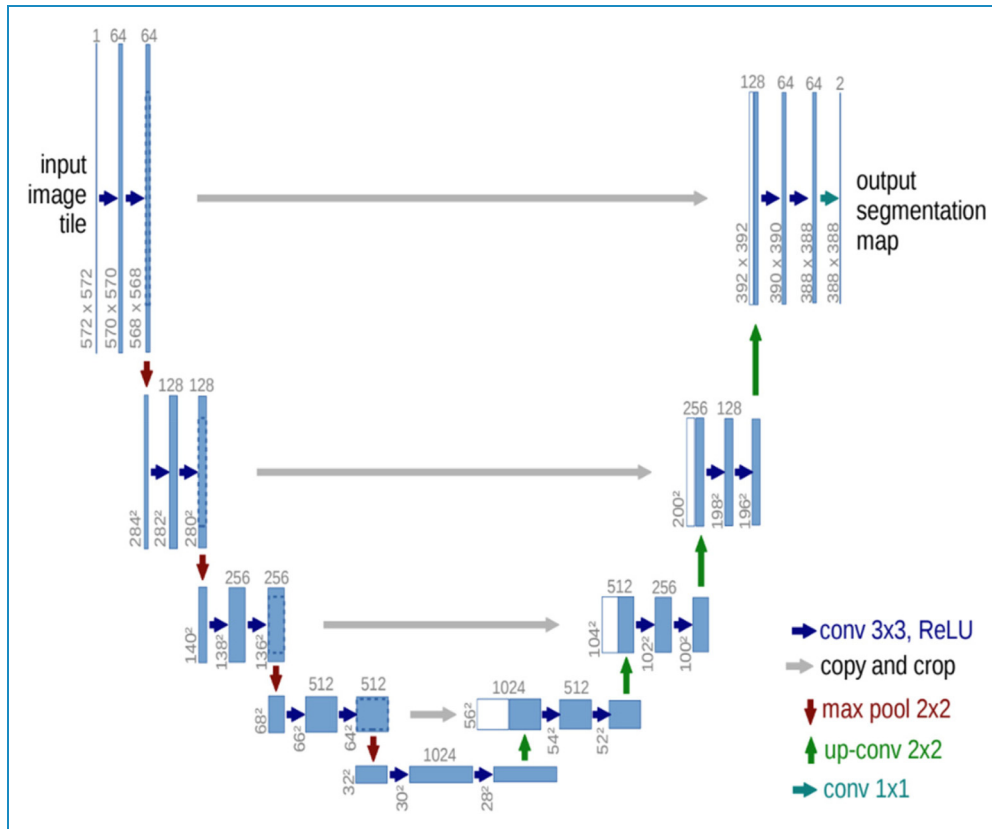


Figure 5. U-Net architecture. Three U-Net models were assembled to obtain the final results. ResNet50, ResNext101, and Vgg19. were used as encoders, which were all pretrained on the ImageNet weights. Ensemble learning is an approach that combines several weak models to produce a model with stronger predictive power.

Table 1. Hyperparameters are used in the individual and ensemble models.

Epochs	60
Callbacks	Early stopping
Learning rate	0.0001
Optimizer	Adam
Loss	Focal, and Dice loss.
Batch size	2
Encoders	ResNet50, ResNext101, Vgg19

maximum or minimum function evaluation. Many machine learning algorithms, ranging from fitting logistic regression models to training artificial neural networks, are based on this difficult problem.

5. Loss: The loss functions measure the distance between an estimated value and its true value. A loss function connects decisions to costs. Loss functions are

dynamic and alter depending on the job at hand and desired outcome.

6. Batch size: The batch size is the size of sub-samples given to the network after which parameter updates occur. Here, a batch size of two was used because of limited resources.
7. Encoders: The encoder is the first half of U-Net architecture. It is usually a pretrained classification network, such as VGG or ResNet, wherein convolution blocks are applied followed by max pool downsampling to encode the input image into feature representations at multiple different levels.

Another model configuration that was used as a callback is “early stopping”. This is a well-established regularization technique that reduces overfitting (model memorizes data instead of learning it). Another benefit of early stopping is that it provides a mechanism for preventing the waste of resources when training is not improving.

The two types of losses were combined to address class imbalance, which cannot be avoided because of the difference in lesion size across classes. For example, it is natural for advanced caries to have more pixels classified because of its depth.

First, focal loss applies a modulating term to the cross-entropy loss (the most common loss used in classification problems) to provide more weight to the difficult-to-classify examples instead of the easy-to-classify ones. This is a dynamically scaled cross-entropy loss, where the scaling factor decays to zero as the confidence in the correct class increases.

Second, Dice loss is widely used in medical image segmentation tasks to address the data imbalance problem. The issue with Dice loss is that it only handles the imbalance problem between the foreground and background, but ignores another imbalance between easy and difficult examples that also severely affects the training process of a learning model. Therefore, in this study, a combination of both was implemented to overcome for these weaknesses.

To elaborate on encoders, ResNet50 is a variant of the ResNet model that has 48 convolution layers, one max pooling layer, and one average pooling layer. It has 3.8×10^9 floating-point operations. ResNeXt repeats the building block that aggregates a set of transformations with the same topology. Compared to ResNet, it exposes a new dimension, cardinality (the size of the set of transformations) C , as an essential factor in addition to the dimensions of depth and width. The CNN Inception-ResNet-v2 was trained on over a million photos from the ImageNet collection [1]. The 164-layer network can classify photos into 1000 object categories, including keyboards, mice, pencils, and a variety of animals.

Statistical analysis

Model performance evaluation

To evaluate the model's performance, the data of the test set and statistical analysis tools were utilized, including true positive (TP), false positive (FP), false negative (FN), and IoU. To illustrate this, each individual radiograph was compared to the ground truth labeling provided previously. Notably, the testing dataset was not observed by the model during the training phase. This step is important to ensure the ability of the model to generalize to all future data. In this study, all evaluation metrics were derived from the confusion matrix, which can be represented as follows for each label (C):

1. TPs of C are all C instances classified as C.
2. True negatives of C are all non-C instances not classified as C.
3. FPs of C are all non-C instances classified as C.
4. FNs of C are all C instances not classified as C.

IoU, also known as the Jaccard index, was leveraged as the primary evaluation metric. It is a widely used metric that quantifies the similarity between the predicted area and the ground truth area in which the intersection is divided

by the union of the two areas. It can be rephrased in terms of true/false positives/negatives, as follows:

$$\text{Jaccard} = \text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

The following metrics were also calculated to evaluate model performance:

1. Recall (sensitivity) = $\text{TP} / (\text{TP} + \text{FN})$
2. Precision = $\text{TP} / (\text{TP} + \text{FP})$
3. F1-score = $(2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Notably, the following definition explains the aforementioned correlations:

1. TP: It represents true positive results.
2. FP: It represents false positive results.
3. FN: It represents false negative results.
4. Recall: This metric quantifies the number of correct positive predictions out of all positive predictions.
5. Precision: This metric quantifies the number of correct positive predictions.
6. F1-score: This metric combines recall and precision into a single score by calculating the harmonic mean of the precision and recall. It is used instead of accuracy owing to the imbalance observed in the datasets, which is natural and inevitable because primary caries are more common in adult bitewings and advanced caries are more common in pediatric bitewings, as discussed in the following sections.

Results

Two separate models were created during model creation. One model was used for adult bitewings and the other was used for pediatric bitewings. This approach is important to avoid misclassification that usually occurs because of the difference in teeth density.

The two models have the same architecture, although they are trained on different datasets, namely, the adults' and pediatrics' datasets. The results were evaluated using IoU score and F1 score.

1. For adults, the IoU averaged 98%, 23%, 19%, and 51% for no caries, primary caries, moderate caries, and advanced caries, respectively.
2. For pediatric bitewings, the IoU averaged 97%, 8%, 17%, and 25% for no caries, primary caries, moderate caries, and advanced caries, respectively. The following tables show the scores of the three models separately and when ensemble learning was implemented. All calculations were performed on the test set.

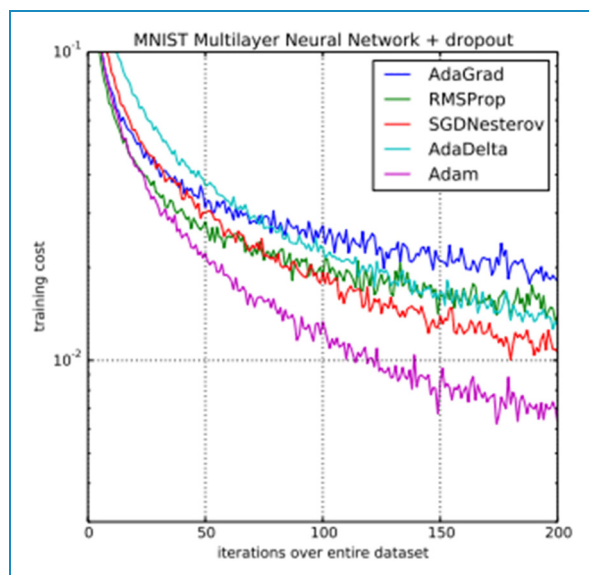


Figure 6. (Kingma, D.P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.), the difference between the performances of different optimizers is shown in comparison with the Adam optimizer, which was used in this study. As per Kingma et al., a Multilayer Neural Networks was trained on the MNIST dataset with AdaGrad, RMSProp, SGD Nesterov, AdaDelta, and Adam optimizers each time, and the values of the loss function (training cost) after each iteration were plotted. From figure 6, we can see that the Adam optimizer reaches convergence much earlier than other commonly used optimizers. This is due to its adaptive learning rate, combining momentum and RMSprop techniques, bias correction, computational efficiency, widespread adoption and support, and robustness to hyperparameter selection. Adam's adaptive learning rate adjusts parameters based on historical gradients, enabling efficient convergence. By combining momentum and RMSprop, Adam achieves superior performance. Bias correction addresses initialization biases. Adam's computational efficiency and wide support make it convenient for large-scale applications. Additionally, Adam is robust to hyperparameter selection, performing well with default settings. Thus, we chose Adam as our optimizer.

Furthermore, as shown by the scores and in screenshots shown in Figures 6–8, the model can recognize whether a shadow represents caries. In addition, it can accurately detect advanced caries in both adult and pediatric bitewings. Misclassification occurs mostly between primary and moderate caries. Although it can classify the lesions correctly many times, it can occasionally misclassify one as the other or does not capture the depth of the lesion accurately. As a standalone model, the model with ResNext101 as its encoder has the best performance on both datasets.

Discussion

In the field of dentistry, the ability to identify caries in the early stages as well as to be able to follow the progression of such caries are crucial factors to implement the most appropriate prevention and treatment methods. When the

diagnosis of dental caries is not properly performed, the potential lesions may progress to reach the pulp, causing extreme pain. It could subsequently progress to a stage requiring extensive procedures and clinical time, which could have been avoided if dental care was discovered in advance, accurately, and appropriately.

One of the most important challenges with the traditional caries detection approach is that it is performed by dentists without any technical aid from available advanced technological resources. Therefore, detection could potentially become lengthy and inconsistent in the long run. To elaborate on this, more than one criterion exist to define caries stages, with a variance in the level of experience that affects the ability of accurate detection and the effect of the quality of radiographic images, including technical errors, brightness, shadow, and contrast. Additionally, the level of fatigue that the dental practitioner might experience in one day could significantly affect caries detection during dental examinations.

One of the available solutions is to capitalize on automated assistance systems. These systems can provide the desired consistency and agility. Therefore, this potential solution was the focus of this study. In this study, bitewing radiographic images were used as an input to the model because they are primarily used to detect the presence of or monitor the progression of interproximal caries. This study has a range of strengths and limitations, which are discussed comprehensively in this section.

The model presented provided a unique contribution to the domain of caries detection by involving the integration of multiple modalities of deep learning architectures to detect the caries in both primary and permanent teeth. The combination of those two architectural frameworks is an innovative methodology that has received limited attention in prior research endeavors. The other novelty in our research is the unique dataset used that represents our patient's demographic and clinical diversity that represents our population. One of the strengths is the detailed labeling of lesion staging. The caries presented in the dataset were classified into three types: primary, moderate, and advanced. Furthermore, the model is a semantic segmentation model, which means that the model can detect the size, depth, location, and type of caries, as shown in Figures 7 and 8 in the results section. An additional strength is the comparison between detection accuracy in adult and pediatric bitewings through the utilization of statistical analysis tools, as listed in Tables 2–5.

OpenCV, Keras and segmentation_models libraries were chosen to perform our research. OpenCV enables us to open images, resize them or introduce augmentation in images very easily through library functions. Keras was chosen for the pretrained models as the segmentation_models library uses Keras/Tensorflow backbones for the U-Net architecture and any other library would have been incompatible with it. Segmentation_models library was

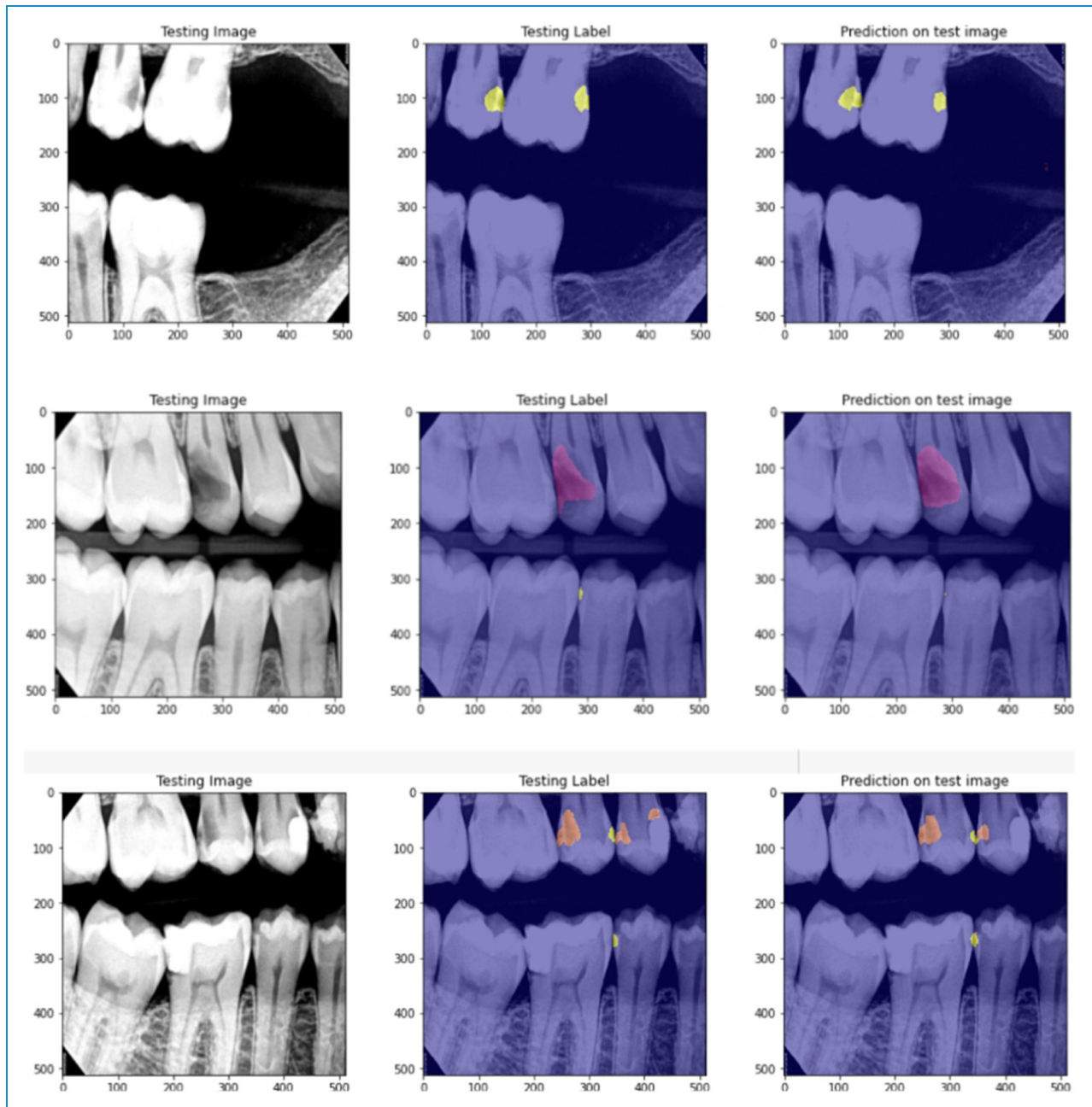


Figure 7. Advanced caries were more accurately detected than primary caries on adult bitewings $P < 0.05$.

chosen because it provides us with the U-Net architecture without having to design it from scratch. The study utilized augmentation techniques to enhance dataset diversity and performance in a model for caries detection. By rotating x-ray images, resizing images, and flipping images, the model was able to simulate variations in image resolution and magnification, addressing differences in x-ray settings. Adjustments to brightness and contrast were made to simulate exposure and lighting conditions, ensuring the model could generalize well to unseen data. Rotation reduced overfitting by exposing the model to different angles of

teeth, making it more robust in recognizing caries from various perspectives. Scaling reduced overfitting by ensuring the model was not overly sensitive to specific image sizes, encouraging the model to recognize caries regardless of image resolution. Flipping mitigated overfitting by learning features invariant to left-right or up-down reflections, enhancing the model's generalization ability. This approach reduced the risk of overfitting and allowed the model to focus on key features in radiographic images.

By contrast, one common limitation imposed upon obtaining accurate results using deep learning in the field

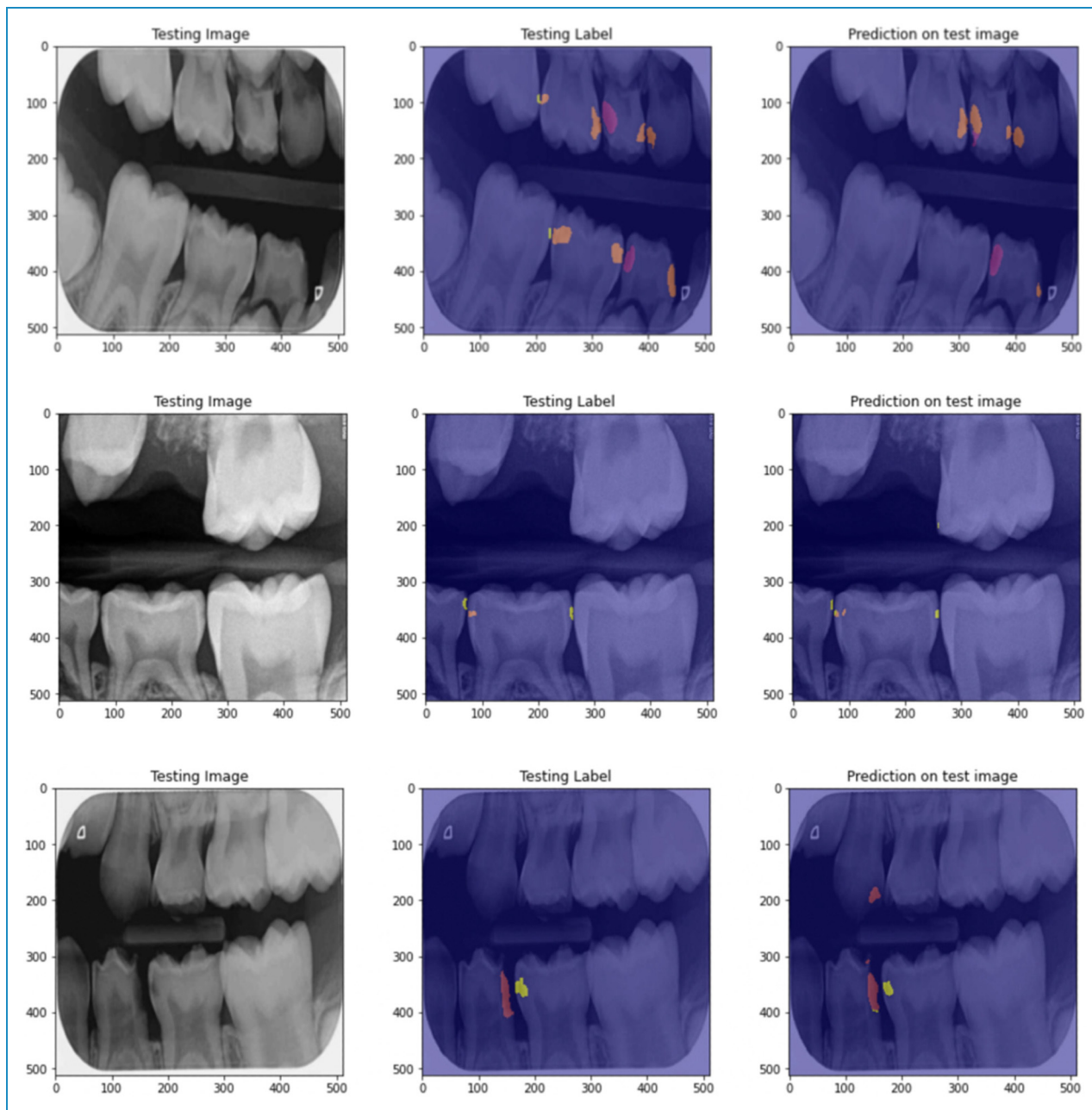


Figure 8. Advanced caries were more accurately detected than primary caries on pediatric bitewings $P < 0.05$.

Table 2. Model performance on adult bitewings.

Encoder	Mean F1	Mean IoU	Mean IoU for No Caries	Mean IoU for Advanced Caries	Mean IoU for Moderate Caries	Mean IoU for Primary Caries
Resnet50	0.517	0.4555288	0.981	0.44	0.22	0.19
ResNext101	0.552	0.4760605	0.972	0.50	0.18	0.22
Inception-ResNet-v2	0.458	0.4338291	0.98	0.34	0.22	0.19

of medicine, in general, is the small size of datasets owing to privacy concerns by the patients. However, in this study, this limitation was addressed by using transfer learning. This technique enables the use of previously learned information to retrieve insights from newly collected data. Consequently, the time and resources required to achieve the goals of the model are significantly reduced. Another limitation is the imbalance in classes. This is because moderate and primary caries are more prevalent in adult radiographs, whereas advanced caries are more prevalent in pediatric radiographs owing to negligence and delayed detection. In addition, the size of the labels is another cause of class imbalance. We observed that advanced caries would cover more pixel points than the other labels. In other words, advanced caries are considerably easier to learn and detect, and this cannot be avoided. Nevertheless, one approach that was used in this study to address this limitation was to use a combination of focal and Dice loss, which is a common technique to address such a problem in the literature. Focal loss is used for multiclass classification, wherein some classes are harder to detect than others.

Furthermore, the U-Net architecture was used to perform semantic segmentation. The U-Net architecture is a semantic segmentation architecture. It has two paths: one that contracts, and one that expands. The contracting path of the CNN follows a standard architecture. To improve the model's accuracy, ensemble learning techniques were employed, which operate on the principle that a weak learner predicts poorly when alone. However, when combined with other weak learners, they create a strong learner.

The model weakness is related to the small dataset size and the similarity between classes. In this study, three

pretrained encoders were used: ResNet50, ResNeXt101, and inceptionresnetv2. First, ResNet50 is a variant of the ResNet model, which has 48 convolution layers, along with one max pooling and one average pooling layer. It has 3.8×10^9 floating-point operations. Second, ResNeXt repeats the building block that aggregates a set of transformations with the same topology. In comparison to ResNet, it adds a new dimension, cardinality (the size of the set of transformations) C , as an essential factor in addition to the dimensions of depth and width. Finally, Inception-ResNet-v2 is a CNN that is trained on more than a million images from the ImageNet database. The 164-layer network can classify photos into 1000 object categories, including keyboards, mice, pencils, and a variety of animals.

Dental Caries is a common oral health problem that affects people of all ages. Early detection of caries is crucial for preventing the progression of the disease, which can lead to pain, infection, and tooth loss. Traditional methods of caries detection rely on visual and tactile examinations, which can be subjective and prone to errors. Recently, deep learning has emerged as a promising tool for caries detection. This approach can potentially improve the accuracy and speed of caries detection, leading to earlier interventions and better outcomes for patients. One of the clinical implications of using deep learning in caries detection is the potential to reduce the need for radiographs. Thus reduce radiation exposure and costs. Another clinical implication of using deep learning in caries detection is the potential to improve patient outcomes by allowing less invasive treatments, such as fluoride application or sealants, which can prevent the need for more extensive restorative treatments like fillings or

Table 3. Statistical results summary for model performance on adult bitewings.

Mean F1	Mean IoU	Mean IoU for No Caries	Mean IoU for Advanced Caries	Mean IoU for Moderate Caries	Mean IoU for Primary Caries
0.543	0.478	0.982	0.51	0.19	0.231

Table 5. Statistical results summary for model performance on pediatric bitewings.

Mean F1	Mean IoU	Mean IoU for No Caries	Mean IoU for Advanced Caries	Mean IoU for Moderate Caries	Mean IoU for Primary Caries
0.44	0.377	0.979	0.255	0.17	0.08

Table 4. Model performance on pediatric bitewings.

Encoder	Mean F1	Mean IoU	Mean IoU for No Caries	Mean IoU for Advanced Caries	Mean IoU for Moderate Caries	Mean IoU for Primary Caries
Resnet50	0.44	0.36	0.98	0.22	0.19	0.05
ResNext101	0.44	0.39	0.981	0.30	0.16	0.11
Inception-ResNet-v2	0.444	0.37	0.978	0.246	0.16	0.09

crowns. By detecting caries at an early stage, deep learning algorithms can help prevent the progression of the disease and improve patient outcomes.¹⁶

The number of included images was limited to 554 adult bitewings and 217 pediatric bitewings. Increasing this number will provide us with better estimation of the accuracy of DL in caries detection and will reduce the risk of biases. Moreover, separating the primary from the permanent teeth during the deep learning process can limit the applicability of this approach to pediatric patients at mix dentition stage.

After comparing the two aforementioned results, we could postulate that the deep learning model is capable of predicting the size, shape, and location of caries, although it is highly dependent on the amount of data used to train it and the consistency of labeling. For future research on the same subject, it is recommended to include balancing the number of collected labels, which could be accomplished by collecting more data. Using substantially larger amount of data would significantly increase detection accuracy, particularly for pediatric radiographs. Furthermore, with respect to this research study, another method that may be introduced for future work is a new label, root caries, to reduce the confusion of the model that may potentially occur because of unlabeled pixels that are similar to currently present labels.

As the use of AI in caries detection continues to evolve, more studies are needed to validate its efficacy and accuracy compared to traditional methods. Future studies should also explore the use of AI in different populations and settings and AI algorithms should be integrated and continued to improve with current clinical practice guidelines to ensure appropriate use and interpretation of results. This will require collaboration between dental professionals and AI experts to develop standardized protocols and guidelines. Future recommendations also should include guidelines for the responsible use of AI in dentistry, including informed consent and protection of patient data. Strategies for reducing the cost of AI technology and ensuring its availability should be tackled. By addressing these recommendations, AI can potentially revolutionize caries detection and improve oral health outcomes for patients. It would be appropriate to consider collecting the dataset of mixed dentition stage (6–12 years old) and label the primary and permanent teeth in each image un-separately and compare the accuracy of caries detection in compression to when we train the machine separately. The result will be more clinically relevant in terms of practicality.

Conclusion

When evaluated the effectiveness of deep learning methods in dental caries detection by applying deep learning technology to bitewing x-ray radiographs, the following conclusions were withdrawn:

- Model's score results show that capitalizing on machine learning for dental caries detection helps recognize

whether a shadow represents a caries in a faster manner than traditional methods.

- Using the model increased the accuracy of detecting advanced caries in both adult and pediatric bitewing radiographs.
- Misclassifications mostly occur between primary and moderate caries. Although the model showed a high ability to classify the lesions correctly, it can misclassify one as the other or does not capture the depth of the lesion accurately at this stage.

Clinical significance

Advancements in the fields of AI and machine learning are unprecedented. Therefore, applying deep learning to dental practices, such as dental caries detection could potentially increase the efficiency of the dentists while also resulting in a society with considerably better dental health. This study has proven the ability of deep learning to be used as an assistive automated diagnostic tool. If the limitations are addressed, which can be achieved if not limited by time, deep learning can accurately detect caries type, size, and location, providing consistency and reducing the workload on dentists.

Acknowledgments: The authors received no financial support for the research, authorship, or publication of this article. The authors would like to express their appreciation and gratitude to the Faculty of Dentistry, King Abdulaziz University for their continual support throughout the study. Authors would like to thank Editage for their language editing services, and Mariam Hussien for her amazing work with Deep learning. "All authors gave their final approval and agree to be accountable for all aspects of the work."

Author contributions: Conceptualization, AAA, NH; methodology, AAA, NH, LS, and AA; software, AAA and NH; validation, AAA, NH, LS, and AA; formal analysis, AAA and NH; investigation, AAA, NH, LS, and AA; resources, LS, and AA; data curation, AAA, NH, LS, and AA; writing—original draft preparation, AAA, NH, LS, and AA; writing—AAA, NH; supervision, AAA, NH; project administration, AAA; All authors have read and agreed to the published version of the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The study was conducted in accordance with the guidelines approved by the Ethics Review Committee of King Abdulaziz University, Department of Pediatric Dentistry (Proposal No. 032-02-22, 28/02/2022).

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

Guarantor: AAA

ORCID ID: Amr Ahmed Azhari  <https://orcid.org/0000-0002-8749-4714>

References

- Selwitz R, Ismail A and Pitts N. Dental caries. *Lancet* 2007; 369: 51–59.
- (a) Keyes PH. Research in dental caries. *J Am Dent Assoc* 1968; 76: 1357–1373. (b) Bacterial Diseases of the Mouth | Boundless Microbiology [Internet]. Courses.lumenlearning.com. 2021 [cited 1 November 2021]. Available from: <https://courses.lumenlearning.com/boundless-microbiology/chapter/bacterial-diseases-of-the-mouth/>.
- Selwitz R, Ismail A and Pitts N. Dental caries. *Lancet* 2007; 369: 51–59.
- Lee J, Kim D, Jeong S, et al. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018; 77: 106–111.
- Mao Y, Chen T, Chou H, et al. Caries and restoration detection using bitewing film based on transfer learning with CNNs. *Sensors* 2021; 21: 4613.
- Lee JH, Kim DH, Jeong SN, et al. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018 Oct 1; 77: 106–111. (<https://www.sciencedirect.com/science/article/pii/S0300571218302252>).
- Bengio Y, Courville A and Vincent P. Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013; 35: 1798–1828. arXiv:1206.5538. doi:10.1109/tpami.2013.50. PMID 23787338. S2CID 393948.
- Lee JH, Kim DH, Jeong SN, et al. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018 Oct 1; 77: 106–111. (<https://www.sciencedirect.com/science/article/pii/S0300571218302252>).
- Schwendicke F, Rossi J, Göstemeyer G, et al. Cost-effectiveness of artificial intelligence for proximal caries detection. *J Dent Res* 2021; 100: 369–376.
- Mao YC, Chen TY, Chou HS, et al. Caries and restoration detection using bitewing film based on transfer learning with CNNs. *Sensors* 2021; 21: 4613.
- Cantu A, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent* 2020; 100: 103425.
- Lian L, Zhu T, Zhu F, et al. Deep learning for caries detection and classification. *Diagnostics* 2021; 11: 1672.
- Moran M, Faria M, Giraldi G, et al. Classification of approximal caries in bitewing radiographs using convolutional neural networks. *Sensors* 2021; 21: 5192.
- Singh P and Sehgal P. GV Black dental caries classification and preparation technique using optimal CNN-LSTM classifier. *Multimed Tools Appl* 2021; 80: 5255–5272.
- Lee S, Oh SI, Jo J, et al. Deep learning for early dental caries detection in bitewing radiographs. *Sci Rep* 2021; 11: 16807.
- Karobari MI, et al. Evaluation of the diagnostic and prognostic accuracy of artificial intelligence in endodontic dentistry: A comprehensive review of literature. *Comput Math Methods Med* 2023; 2023: 7049360.
- Bayrakdar IS, Orhan K, Akarsu S, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. *Oral Radiol* 2022; 38: 468–479.
- Casalegno F, Newton T, Daher R, et al. Caries detection with near-infrared transillumination using deep learning. *J Dent Res* 2019; 98: 1227–1233.