Heliyon 10 (2024) e30836

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks' ensemble

Andra Carmen Mărginean [a,1], Sorana Mureşanu [b,*,1], Mihaela Hedeşiu [b], Laura Dioşan [a]

[a] Computer Science Department, Babes Bolyai University, Mihail Kogalniceanu 1, Cluj-Napoca, 400347, Cluj, Romania
[b] Department of Oral and Maxillofacial Surgery and Radiology, Iuliu Haţieganu University of Medicine and Pharmacy, Victor Babes, 8, Cluj-Napoca, 400012, Cluj, Romania

A R T I C L E I N F O

MSC:
0000
1111

Keywords:
Caries segmentation
Teeth segmentation
Neural network ensemble
Deep learning
U-Net
Feature pyramid network
Deeplabv3

A B S T R A C T

Background: Dental cavities are common oral diseases that can lead to pain, discomfort, and eventually, tooth loss. Early detection and treatment of cavities can prevent these negative consequences. We propose CariSeg, an intelligent system composed of four neural networks that result in the detection of cavities in dental X-rays with 99.42% accuracy.
Method: The first model of CariSeg, trained using the U-Net architecture, segments the area of interest, the teeth, and crops the radiograph around it. The next component segments the carious lesions and it is an ensemble composed of three architectures: U-Net, Feature Pyramid Network, and DeeplabV3. For tooth identification two merged datasets were used: The Tufts Dental Database consisting of 1000 panoramic radiography images and another dataset of 116 anonymized panoramic X-rays, taken at Noor Medical Imaging Center, Qom. For carious lesion segmentation, a dataset consisting of 150 panoramic X-ray images was acquired from the Department of Oral and Maxillofacial Surgery and Radiology, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca.
Results: The experiments demonstrate that our approach results in 99.42% accuracy and a mean 68.2% Dice coefficient.
Conclusions: AI helps in detecting carious lesions by analyzing dental X-rays and identifying cavities that might be missed by human observers, leading to earlier detection and treatment of cavities and resulting in better oral health outcomes.

1. Introduction

Detecting cavities in dental X-rays is a challenging task that requires a high level of expertise and experience from dental professionals. Cavities, also known as dental caries, are one of the most common dental problems affecting people of all ages. They occur when the acid produced by bacteria inside the mouth dissolves the tooth enamel, leading to a hole or cavity in the tooth. If left untreated, cavities can cause pain, infection, and even tooth loss. The process of detecting cavities in dental X-rays involves examining the images for changes in the density of the tooth structure, which appear as darker areas in the radiograph. The detection

* Corresponding author.
E-mail address: sorana.eftimie@gmail.com (S. Mureşanu).

[1] Sorana Mureşanu and Andra Carmen Mărginean contributed equally to the manuscrpt.

https://doi.org/10.1016/j.heliyon.2024.e30836
Received 11 January 2024; Received in revised form 27 March 2024; Accepted 6 May 2024

Available online 10 May 2024

of cavities is not always straightforward, as factors such as the size, shape, and location of the cavity can affect its appearance. A paper [1] presenting the results of a study that ran between 1990 and 2017 found that globally, approximately 2.3 billion individuals had dental caries in permanent teeth.

In recent years, advances in technology have led to the development of computer-aided detection (CAD) systems that use intelligent algorithms and Machine Learning techniques to assist dental professionals in detecting cavities. These systems can analyze X-ray images and provide automated diagnostic support by highlighting areas of concern or identifying potential cavities that may be missed by human examination alone. Automatic and semi-automatic segmentation techniques can be either conventional, meaning that pre-defined features are used as the inputs to a machine learning model, such as a Support Vector Machine (SVM), to classify image pixels as lesions or non-lesions, or deep learning-based, referring to the use of a deep network to segment directly the image, since they are capable of learning the features that can best distinguish lesions and non-lesions pixels, and can be trained in an end-to-end manner.

Carious lesions appear as demineralized, radiolucent (dark) areas starting at the tooth surface. The detection of cavities is not always straightforward, as factors such as the size, shape, and location of the cavity can affect its appearance. Normal tooth morphology (pits and fissures), phenomena such as the cervical radiolucency (burn-out) effect, as well as other diseases and conditions involving the dental hard tissues (for example abrasion, hypoplastic pits and pre-eruptive intracoronal resorption), may also be falsely interpreted as carious lesions [2]. Visual diagnosis also becomes more challenging with increasing levels of noise and shadows, as caries lesions can be easily mistaken for shadows and vice versa, particularly mild ones. The Mach effect, which is an optical phenomenon causing the edges of darker objects adjacent to lighter ones to appear lighter and vice versa, can also cause diagnostic difficulties. This effect can result in false shadows that may lead to misinterpretation of caries lesions that are present near dental restorations that appear whiter in dental X-rays [MRGAP11]. Furthermore, training the AI algorithms frequently requires annotated data, the so-called Ground-Truth (GT), that is difficult to build. The lack of a larger and more encompassing dataset poses a challenge to the accurate segmentation of caries and the ability of the networks to learn. The detection of cavities in dental X-rays is a complex task that requires a combination of technical skill and clinical experience. Accurate diagnosis and timely treatment of cavities are essential for maintaining good oral health and preventing more serious dental problems. Therefore, research needs to be done on the detection of cavities in teeth.

In the literature, the problem of caries identification has been investigated by different deep-learning models like U-net [3], [4] or DeepLab [5], but the performance of these simple models can be improved still. Furthermore, the previous approaches did not consider an ensemble model that can achieve higher accuracy than individual models by combining their predictions and reducing their errors. Furthermore, ensemble models can produce more reliable and confident predictions than individual models by averaging independent predictions. By enforcing model diversity, an ensemble model can reduce the over-confidence or under-confidence of individual models, which can be harmful to critical decision-making processes [6]. Ensemble models can handle complex and noisy data better than individual models by exploiting the diversity and variability of the data. Ensemble models can capture different aspects and features of the data, and cope with different challenges and limitations of individual models [7].

In this paper, a two-step CAD system, called CariSeg, is introduced for segmenting carious lesions. The first step is the segmentation of the teeth region in an X-ray image which is done using a U-shaped Neural Network. The second step involves the segmentation of carious lesions using an ensemble of three different architectures, namely the U-shaped Neural Network [8], Feature Pyramid Network [9], and DeeplabV3 [10], resulting in enhanced accuracy and precise identification. In proposing our system we have taken into account the last results related to ensemble learning, able to improve the accuracy and stability of prediction models, especially for the generalization ability on small datasets [11–13].

In this paper, we raise the following research questions:

RQ1: How does the proposed CariSeg, utilizing four neural networks, perform in terms of accuracy and efficiency for detecting dental cavities in X-ray images?

RQ2: What are the factors influencing the accuracy and reliability of the AI algorithms in detecting and segmenting dental cavities in X-ray images, and how can these factors be optimized to improve overall performance?

To answer these research questions, we have leveraged the capabilities of one deep learning model to first identify the region of interest, the teeth. We then crop the image around the teeth using the mask segmented by the model. The uniqueness of this approach lies in its pipeline-based structure. When a deep learning model is trained on a large dataset of images, it learns to identify patterns and features within the images that are relevant to the task at hand. However, if an image contains a lot of irrelevant information, such as background or surroundings, the model may struggle to identify the key features and patterns that are relevant to the specific task of cavity detection. By cropping an image around the area of interest, the model is presented with a much smaller, more focused area that is more likely to contain the relevant features and patterns necessary for accurate cavity detection.

The cropped images are used to train three different architectures, U-net, Feature Pyramid Network (FPN), and DeeplabV3. An ensemble of these models is created and is used to segment cavities by taking into account the output of each network. By fusing these models, we harness the strengths and unique characteristics of each architecture to enhance the overall performance and accuracy of the cavity segmentation task. The use of an ensemble for caries segmentation is a new approach, performs better than any of the individual models, and it results in improved generalization.

Overall, the contributions of this paper lie in the introduction of a novel pipeline-based structure, the utilization of ensemble modeling, the use of cropped images for focused analysis, and the comprehensive evaluation of the proposed approach. This results in 94.895% accuracy and a Dice score of 88.5% for teeth segmentation, as well as a 99.42% accuracy and a mean 68.2% Dice coefficient for caries segmentation. Moreover, we have done experiments using 6 different loss functions (Binary cross-entropy loss, Dice loss, Tversky loss, Focal Tversky loss, Binary cross-entropy and Dice loss, Combo loss) to identify the best one.

The paper is structured as follows. In Section 2 we present a review of related works, as well as the result of the state-of-the-art papers. Our proposed approach is introduced in Section 3, while Section 4 presents the results of our experiments. Section 5 describes the conclusions of our study.

2. Related works

2.1. Backgrounds

The use of Machine Learning in medical imaging for diagnosis of medical conditions has become increasingly popular. One such area of research is dentistry. Dental caries detection is an ongoing and important topic of research [14].

A systematic review of the use of Deep Learning for caries detection in which 48 studies were assessed was carried out by Mohammad-Rahimi et al. in 2022 [15]. The studies were of three types: classification, segmentation, and object detection. Classification studies classify an image containing one or more teeth as being healthy or diseased. Object detection studies identify the location of the carious lesions in the image and mark it with a box. Segmentation studies find the exact location and shape of the cavity. In the review, 10 studies were identified as using segmentation models for caries segmentation and only in 3 studies panoramic radiographs were used. Since then additional studies have been published, but caries segmentation remains an important topic of research.

A CAD system for cavity detection was proposed by Srivastava et al. in 2017 [16]. Using a private dataset of more than 3000 annotated bitewing radiographs for training, the fully convolutional neural network's architecture consists of over 100 layers. The result of the model is a bounding box around the cavity area. This led to an F1-score of 70%, higher than the score achieved by the dentists used for testing.

In 2020, a CNN using EfficientNet-B5 as an encoder was trained to detect cavities on 3,686 private bitewing radiographs by Cantu et al. [17], leading to an accuracy of 0.80. In the study, this was compared to the accuracy of dentists, which was lower than that of the model. Again, the output of the model is a highlighted box area where the network identifies a cavity.

Later on, in 2022, a pre-trained Cifar-10Net CNN was used by Lin et al. [18] to detect proximal cavities in periapical radiographs. The study dataset consists of 600 private images of teeth containing cavities of different progression levels ranging from 0 to 5. The dataset was preprocessed using edge extraction strategies. A bounding box is drawn around the detected diseased area and the study achieved an 85.9% accuracy.

Bayrakdar et al. [19] used U-net in combination with VGG-16 as a backbone to train the neural network on a dataset of 621 anonymized and private dental bitewing images. During the training process, 200 epochs were used and data augmentation was applied, such as horizontal and vertical flips. The study uses the U-net in order to segment cavities. It was reported an accuracy of 80% and a 73% F1-score, proving that U-net can be successfully used to segment cavities. U-net architecture was used also in [20].

In the same year, Oztekin et al. [21] introduced a model that is given a panoramic X-ray as an input produces a classification for each tooth as healthy or not. It also produces a heat map as an output, which is meant to indicate the cavities inside the image. The technique used for making the heat map is called Grad-CAM. The resulting heat map is used for segmentation purposes. They compare three pre-trained models, EfficientNet-B0, DenseNet-121, and ResNet-50. ResNet-50 is found as the better model, leading to an accuracy of 92.00%. The used dataset is private and it consists of 562 panoramic X-ray images, from which each tooth was manually cropped by a specialist.

2.2. State-of-the-art

The research paper introduced by Dayi et al. in 2023 [3] presents a new architecture, Dental Caries Detection Network (DCDNet), for detecting and segmenting carious lesions. The dataset used is private and it is composed of 504 panoramic radiographs of unknown dimensions. The ground-truth segmentation was done by a dental specialist and a radiologist. One distinctive feature of DCDNet is the Multi-Predicted Output (MPO) structure in the output layer, which splits the feature map into three separate paths for identifying occlusal, proximal, and cervical caries. The network still has an encoder-decoder architecture. The encoder uses a pre-trained backbone such as VGG16, MobileNet, and EfficientNet. The decoder consists of two modules: the Multi-level Features Concatenation (MFC) which combines the features from all the layers in the network and the Multi-Predicted Output (MPO). ResNet50 was found as the better backbone and the study obtained an average F1-Score of 62.79%.

A novel architecture proposed is CariesNet, published by Zhu et al. in 2022 [5]. CariesNet is a U-shaped encoder-decoder network, using Res2Net as a backbone feature map. The backbone's three high-level feature maps are combined with the partial decoder, which generates an initial map for dental caries, referred to as the global map. After that, both the backbone feature and the partial decoder feature are merged and fed to the attention module. The network uses a full-scale axial attention module to concatenate the features from the partial decoder and the backbone, as well as the result of the FAA from a previous level. At the end a sigmoid function is used to compute the output, a segmentation of the caries present in the oral panoramic image. The dataset used in the study is private and it has 1159 512x512 panoramic radiographs checked by three dentists and verified by a fourth. The study reports achieving a mean 93.64% Dice coefficient and 93.61% accuracy.

Another important study was published by Lian et al. in 2021 [4]. The research paper uses nnU-Net, a version of the U-net that can configure itself. It is an encoder-decoder network that uses skip connections to send information from the encoder layer to the corresponding decoder layer. Once the caries are segmented, they are classified using DenseNet into three types of lesions, according to their significance. The study reaches an accuracy of 98.6%, a Dice of 66.3%, and an IOU score of 78.5%. The dataset used in the

Table 1
Summary of advantages and disadvantages of State of the Art studies.

Study	Advantages	Disadvantages
Dayi et al. [3]	• Examined three types of caries: occlusal caries, proximal caries, cervical caries • Better results than U-Net, FPNet, Mobile-UNet, and Eff-Unet models • Good performance in segmenting occlusal caries, proximal caries • Achieved an average dice of 62.79%	• Insufficient dataset, especially for cervical caries • Panoramic radiographs used in the study could not be supported by clinical examination of the patients • Low performance in detecting cervical caries
Zhu et al. [5]	• Large and extensive dataset • Better results than U-Net, DeepLabV3+, Res-Unet, Attention-UNet, PraNet • Achieved an average dice of 93.64%	• Didn't publish the code or the dataset • The models tend to misclassify moderate caries as shallow caries or deep caries. • Lower performance on moderate caries (69.4% dice)
Lian et al. [4]	• Large and extensive dataset • The predicted caries were output as highlighted areas and presented in three different colors according to their depths • Achieved an average dice of 66.30%	• Excluded blurred panoramic films from the dataset • Labeling in the constructed reference test was not sufficiently precise, as it was not triangulated with the gold standard (histology)

study is private and it contains 1160 300x400 panoramic radiographs checked by three dentists. 40 images were used for testing while the rest were used for training.

DCDNet models consistently achieve higher success rates compared to other models. Despite the higher time consumption, DCDNet models, particularly ResNet50-DCDNet, exhibit effective detection of dental caries. The Mobilenet-DCDNet model is highlighted as the fastest among the DCDNet models, indicating that certain variations within DCDNet can offer relatively quick processing times. The proposed DCDNet models, with three outputs, require more time for processing (between 80.0 and 98.3 milliseconds) compared to single-output models like Mobile-Unet, Unet, Eff-Unet, and FPNet. Single-output models (Mobile-Unet, Unet, Eff-Unet, and FPNet) are noted as more economical in terms of time consumption. However, their success rates are comparatively lower, as highlighted in [3]. The Mobilenet-DCDNet model, while the fastest, is mentioned to excel primarily in Type 1 caries detection, potentially making it less versatile in identifying other types of caries compared to the ResNet50-DCDNet model.

CariesNet [5] achieves better results in dental caries detection. CariesNet utilizes a full-scale axial attention module, allowing it to capture wide and efficient contextual information. This suggests that the model is adept at understanding the broader context of the input data, contributing to its improved performance. While CariesNet may show limited performance on moderate caries over the backbone, it demonstrates a significant improvement on deep caries and on shallow caries. This indicates a specialized strength in detecting specific types of caries, but also that there might be challenges or room for improvement in detecting this specific type of dental caries. The limited performance on moderate caries over the backbone suggests a potential dependency on the underlying architecture or backbone used in CariesNet. This could be a limitation if the backbone is not optimized for detecting certain types of caries. Furthermore, the axial attention module, which captures wide and efficient contextual information, is susceptible to adding supplementary computational costs or complexities. The approach proposed in [4] is based on a DenseNet121, a 121-layer connected network, that is employed to identify caries stages. It addresses the vanishing gradient issue and strengthens feature propagation by connecting all preceding layers to subsequent layers, potentially improving the model's ability to capture complex relationships in data. However, the panoramic films were made on the equipment of one company, and blurred images (90 out of 1250) were excluded before model training, limiting the generalizability of the reference dataset. External verification on different equipment is deemed essential in the next steps. Furthermore, the study lacks a gold standard such as micro-CT and histology of extracted teeth, which could potentially impact the precision of the labeling and annotation process. The absence of a hard gold standard suggests a need for validation through other diagnostic approaches.

A summary of the advantages and disadvantages of each study is presented in Table 1.

3. CariSeg - the proposed CAD system

In this section, we introduce the proposed CAD system CariSeg, an ensemble approach. In our innovative approach to caries segmentation, we first segment the teeth in the radiography using a model and then segment the carious lesions using an ensemble made up of three alternative architectures. Despite having a tiny dataset, the trials show that this method produces 99.42% accuracy and a mean 68.2% Dice coefficient.

3.1. CariSeg - main idea

A two-step process is introduced to segment carious lesions in X-ray images of teeth. Initially, a model is employed to accurately identify the teeth region. The second step involves segmenting the carious lesions using an ensemble of three distinct deep-learning

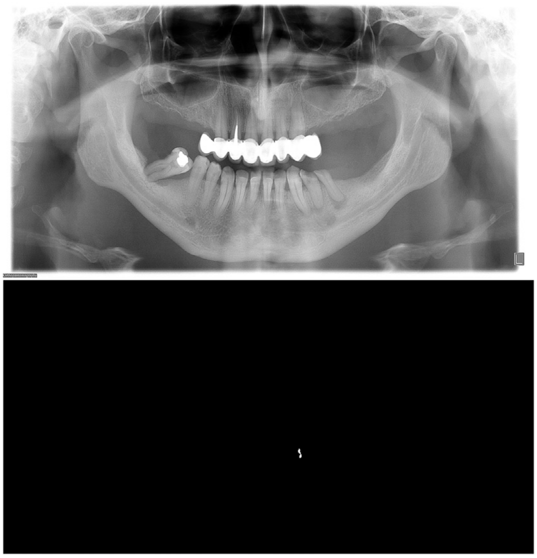

Fig. 1. A panoramic radiography and its associated mask of segmented cavities, taken from the Clujana dataset. Dental cavities are present in the left mandibular canine (tooth 3.3).

models: U-Net, Feature Pyramid Network, and DeeplabV3. This combination of models leads to improved accuracy and precise identification of the lesions.

We first identified the area of interest, the teeth, by utilizing the capabilities of one deep learning model. Using the segmented mask predicted by the model, we then crop the image around the teeth. The pipeline-based structure of this technique is what makes it special. A deep learning model learns to recognize patterns and features in the images that are pertinent to the task at hand after being trained on a large dataset of images. The model may, however, find it difficult to recognize the essential characteristics and patterns that are pertinent to the particular task of cavity identification if an image contains a lot of irrelevant information, such as background or surroundings. Cropping focuses the model's attention on a smaller area likely to contain essential information for accurate cavity detection.

Three alternative architectures — U-net, Feature Pyramid Network (FPN), and DeeplabV3 — are trained using the cropped images. By combining these models and taking into account each network's output, cavities are segmented using the ensemble of these models. We fuse these models to improve the overall performance and accuracy of the cavity segmentation task by utilizing the advantages and distinctive qualities of each architecture. An innovative method that outperforms all individual models and leads to superior generalization is the use of an ensemble-based approach for caries segmentation.

The underlying schema of the proposed ensemble is bagging with bootstrapping that considers three heterogeneous weak learners. The learners have been trained independently from each other in parallel on the same dataset, while for combining the predicted values a deterministic voting strategy is followed.

3.2. Datasets

The main dataset consists of 150 anonymous panoramic radiographs obtained from the Department of Oral and Maxillofacial Surgery and Radiology, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca. Dental experts labeled the carious lesions associated with each image. Each radiograph was annotated in collaboration with dentists. One aspect of note is that cases of severe tooth decay (destruction of the entire dental crown and parts of the root) were not annotated as carious lesions, but instead as root fragments – these were not the object of the current study and were meant to be explored in further research. Radiographs in which caries could not be detected were not included in the analysis, resulting in 108 images that could be used. Additionally, the excluded images were not used for assessing model performance. An example of a radiograph from the dataset and its associated mask is shown in Fig. 1.

For teeth segmentation two combined datasets were used. One of them is the Tufts Dental Database [22], which consists of 1000 panoramic dental radiography images. Each of the images was segmented by an expert from the Tufts University School of Dental Medicine. The radiographs were collected from January 2014 to December 2016 and were fully anonymized and de-identified.

Another dataset of 116 anonymized panoramic dental X-rays was published by Abdi et al. [23]. They were taken of patients at Noor Medical Imaging Center, Qom. The subjects in the study encompass a broad spectrum of dental conditions, including individuals with healthy teeth as well as those with partial or complete tooth loss. A segmentation of the teeth was published by Helli et al. in 2022 [24]. An example of panoramic radiography and its corresponding teeth mask resized to be of 512x512 dimensions is shown in Fig. 2.

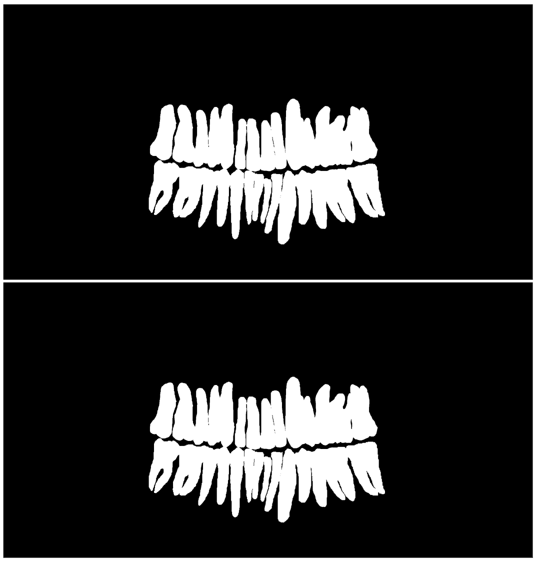

Fig. 2. A panoramic radiography and its associated mask segmenting the teeth, taken from the Tufts Dental Database [22].

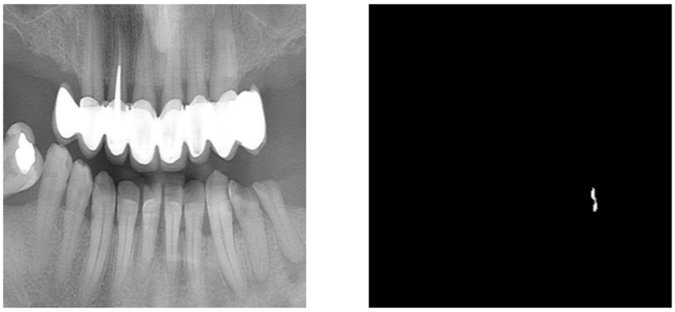

Fig. 3. A panoramic radiography and its associated mask of segmented cavities, taken from the Clujana dataset. Note the cavity present in the left mandibular canine (tooth 3.3). The radiography and its mask were cropped using the output of the teeth segmentation model.

3.3. Preprocessing

The original images in the Clujana dataset, the Tufts Dental Database dataset [22], and the Noor Medical dataset [23] have different widths and heights. They and their corresponding mask are re-dimensioned to be of the size 512x512.

The Clujana dataset cropped using the teeth segmentation model, is split into 2 subsets: a training part and a testing part. 14 out of 108 images are used for testing, while the rest 94 of them are used for training.

Gamma correction is used to adjust the brightness and contrast of the radiographs in the Clujana dataset to ensure that the image accurately represents the density and structure of the teeth. Without gamma correction, the image can appear too dark or too bright, making it challenging to identify the details needed for diagnosis.

3.4. Teeth segmentation

A model for teeth segmentation was trained using the U-net architecture and the images from Tufts Dental Database [22] and Noor Medical Imaging Center dataset [23]. The model was then used on the Clujana dataset to segment the teeth inside the radiographs and crop the images around them. Before cropping the image around the area of interest and eliminating upper and lower jaws, surrounding structures, and tissues, the cavities represented an average of 0.036% of the images. After tooth segmentation and cropping, the average surface represented by carious lesions has a 44% increase. Now cavities make up on average 1.584% of an image (Fig. 3).

3.5. Caries segmentation

For caries segmentation, the radiographs from the Clujana dataset are first pre-processed by cropping the images and applying gamma correction. On the resulting radiographs, data augmentation is applied in order to artificially increase the size of the dataset by creating new variations of existing images, making the dataset more diverse. From 94 training images, we obtained 380 images. The models that are part of the ensemble are trained using these 380 images.

The augmentation technique is particularly useful when the size of the available dataset is small and it helps with problems such as overfitting and underfitting. Overfitting occurs when a model starts to fit the training data too well, leading to poor generalization of new images. By artificially increasing the amount of training data, the model is less likely to overfit because it has more examples to learn from, allowing it to come up with more general patterns. On the other hand, underfitting occurs when a model lacks enough data to learn from. By increasing the dataset, underfitting is less likely to occur. Data augmentation can also help prevent the model from memorizing specific features of the training data and instead learn more abstract representations that apply to a wider range of examples. Data augmentation was implemented by using the Albumentations Python module. The modifications applied to the radiographs are the following: horizontal flips, brightness alteration in the range [0.85, 1.15], contrast alteration in the range [0.65, 1.35], rotation of the image of up to 15 degrees, shifting of the image of up to 10%, scaling of the image by up to 10%, sharpening of the image.

Ensemble learning involves the fusion of multiple models to make predictions or decisions, and it is a powerful technique in machine learning. Ensemble models offer several benefits, such as improved accuracy by leveraging the collective knowledge of diverse models, reduced overfitting by averaging biases and variances, increased robustness to noisy data, better generalization to unseen examples, and the ability to estimate uncertainty in predictions. Overall, ensemble models are widely used and can significantly enhance predictive performance in various Machine Learning applications.

U-Shaped Neural Networks, also called U-Nets, were first introduced in 2015 by Olaf R. et al. [8]. The overall structure of the network is composed of two paths that form a "U" shape, hence the name. It is meant to be compatible with smaller datasets and to be used in medical image segmentation tasks. In total, it is formed of 23 convolutional layers and it is formed of two paths: a contracting path and an expansive path.

Feature Pyramid Network [9] has a pyramid-like construction and it has a bottom-up pathway and a top-down pathway. In the bottom-up pathway, a pyramid of feature maps is built from the input image, with each level of the pyramid corresponding to a different scale of the image, using a scaling step of 2. Each level of the feature pyramid leads to a prediction. The FPN's top-down pathway upscales less detailed but more semantically relevant feature maps from higher pyramid levels to create higher-resolution features. These are then combined with features from the bottom-up pathway through lateral connections.

Deeplab was introduced by Chen et al. in 2016 [25]. It is a semantic architecture used for image segmentation. A modified version, Deeplabv3, was published in 2017 by Chet et al. [10]. In the original paper, the input image is passed through the network by utilizing dilated convolutions. The output generated from the network goes through bilinear interpolation and passes through a fully connected CRF (Conditional Random Field), a probabilistic graphical model, for enhancing the predictions. Deeplabv3 improves upon the Deeplab architecture and achieves better accuracy.

We have chosen to create an ensemble containing the best-trained model of each one of the three architectures: U-net, FPN, and DeeplabV3. Each of the models segments the cavities inside of the radiography with a certain probability. The probabilities are then summed up and only carious lesions that have a probability higher than 105% are shown. By setting this threshold, we aim to ensure that only carious lesions with a relatively high probability of presence are highlighted and presented to the users. Experimenting with different thresholds led us to this choice. This was chosen so that a cavity pixel is identified with a probability of over 50% by at least 2 models or over 35% over 3 models.

In Fig. 4, for a better understanding, we present the pipeline structure employed in our approach, which involves multiple steps. First, a dedicated model is utilized to segment the teeth from the radiograph, ensuring that only the relevant area is retained for further analysis. Second, the ensemble predicts the caries lesions in the radiograph, and the results from each model are combined.

4. Experiments

4.1. Metrics

The terms FP, TP, TN, and FN are used to refer to the number of samples (pixels, in our case) that were correctly or incorrectly classified. FP, TP, TN, FN mean False Positive, True Positive, True Negative, and False Negative respectively. The metrics Accuracy, Precision, Recall, F1, and IOU are used to test the performance of the presented ensemble. They are defined in Equations (1), (2), (3), (4), (5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

$$IOU = \frac{TP}{TP + FP + FN} \tag{5}$$

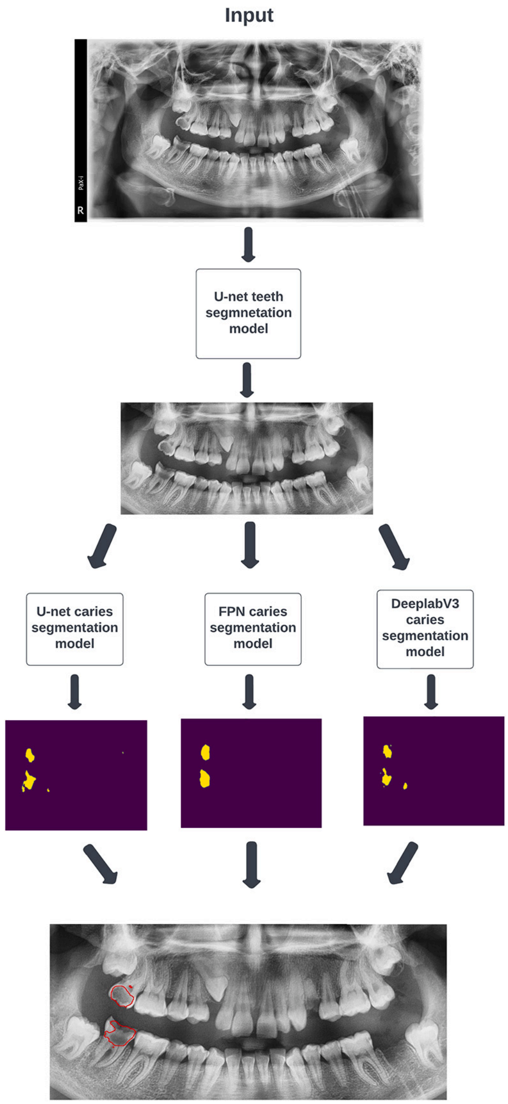

Fig. 4. Image showing the design of our two-step CariSeg system. At first, the segmentation of the teeth region in the X-ray image is done using a trained U-shaped Neural Network. The second step represents the carious lesion segmentation, which is done using an ensemble incorporating three different models, namely U-shaped Neural Network, Feature Pyramid Network, and DeeplabV3.

4.2. Experimental setup

All of our experiments were done in Kaggle and we applied the same hyperparameters to each experiment. The models were trained for 100 epochs, and we used the Adam optimizer with a learning rate of 0.0001. Adam is one of the most popular optimizers used in deep learning, it is computationally efficient and easy to implement. These hyperparameters were chosen after a parameter-tuning process and many experiments. We did experiments with learning epochs ranging from 10 to 200 and with 6 different optimizers, as well as with different learning rates. We did not include these experiments in the study to ensure the conciseness and manageability of the study.

Since we are solving a binary classification problem (cavity or non-cavity), we are using the Sigmoid activation function in the output layer of the models. Softmax is a generalization of Sigmoid for when there are more than two categories. The Sigmoid function returns the probability of its input of being in the first class. The probability value ranges between 0 and 1. Although usually the threshold for classification is 0.5, we have found that marking every pixel with a probability of being a cavity above 0.25 leads to better results. For the sake of experimentation, we tried both Softmax and Sigmoid in our experiments. The use of Softmax led to blank images.

Another important hyperparameter is the number of training epochs. A training epoch refers to a pass of the training dataset through the network. The number of training epochs refers to the number of times the model will learn from that dataset. The number of training epochs directly affects the performance of the model and how well it generalizes to unseen data. If the number is

Fig. 5. Example of teeth segmentation on the test data, showing the original mask, the predicted mask by the trained model, and the corresponding radiography. The data was taken from the Tufts Dental Database [22].

too low, it will result in underfitting, as the model won't have enough time to learn from the dataset and it will be unable to spot the patterns in the data. However, if the number is too high, it will result in overfitting. The model will learn too well from the dataset and will start memorizing the output, resulting in a model that performs well on the training data, but poorly on new data. Choosing the number of epochs is a balance between these two scenarios. In general, a high number of training epochs is more suitable for a larger dataset. On our dataset, we have found that using a higher number of training epochs (more than 100) led to fewer cavities being detected, but those were more accurate. Whereas a lower number of training epochs (less than 100) led to more cavities being detected, but among those some fake positives were present.

Another hyperparameter that can be chosen during training is the optimizer, which has the goal of minimizing the loss function by adjusting the weights of the model. Different optimizers use different strategies to update the weights, such as adjusting the learning rate, using momentum, or adapting the learning rate based on the historical gradients. Common optimizers used in deep learning include Stochastic Gradient Descent (SGD), Adagrad, Adadelta, RMSprop, and Adam [26]. We compared the performance of these 5 optimizers. Adam was the best-performing optimizer, followed closely by RMS Prop. On the other hand, Adagrad and Adadelta for our dataset led to suboptimal results. No segmentation was done on the images and the calculated dice was a very low value, 0.020, respectively 0.021. As a result, Adam is the optimizer used in the rest of the experiments.

In our experiments, we use the following loss functions: Binary Cross-entropy loss, Dice loss, Tversky loss, Focal Tversky loss, and Combo loss.

The Binary Cross-entropy (BCE) loss is commonly used for binary classification tasks [27] It measures the difference between the predicted probability distribution over the two classes and the true distribution, which is an encoded vector indicating the true class of the input.

The Dice loss function [28] is widely used in scientific research papers related to image segmentation and is effective in dealing with imbalanced datasets. It is based on the Dice coefficient, which is a measure of the similarity between two sets. The function produces a value between 0 and 1, with higher values indicating better overlap or similarity between the predicted and true segmentation masks.

An alternative to Dice loss is the Tversky loss. It is based on the Tversky index, which is a generalization of the Dice coefficient [29]. The Tversky index takes into account both false positives and false negatives in the segmentation masks. Focal Tversky loss is an extension of the Tversky loss function that introduces a focal parameter to further emphasize challenging samples during training. It helps to address the class imbalance and focus more on hard-to-segment regions.

Another alternative is Combo loss which was defined by Saeid et al. [30] in 2011. It combines Dice loss with a slightly modified cross-entropy loss and it is meant to handle both input and output class imbalance.

4.3. Results for teeth segmentation

The first part of the pipeline is the teeth segmentation model. The models were trained on 1016 images and tested on 100 images. The best model, trained with BCE loss, achieved an accuracy of 94.939% and a Dice (or F1) score of 0.889, a Jaccard Index (or IOU) of 0.798, a recall of 0.816, and a precision of 0.980. The results are comparable to the Dice overlap score of $95.4 \pm 0.3\%$ achieved by Helli et al. in 2022 [24]. A prediction example is given in Fig. 5. It is worth noting that the primary objective of our study does not revolve around optimizing the teeth segmentation model, as its primary purpose is to facilitate the cropping of images around the region of interest. Therefore, our Dice overlap score, which aligns well with our specific requirements, did not necessitate further enhancement.

The predicted results of the teeth segmentation models (for testing images from [22] and [23] are presented in Table 2. For these images, the teeth masks are available, facilitating the computation of the performance metrics. Overall, the results do not significantly differ because all the tried losses can be used for image segmentation. The Dice coefficient, also known as the F1-score, is slightly better for the binary cross-entropy function. The second-best results were obtained using the Dice loss.

4.4. Results for carious lesions segmentation

Regarding the carious segmentation, the first experiments were done for the U-net architecture, which is well known for its use in medical image segmentation. The results of the experiments are presented in Table 3. The highest Dice, 0.645, was achieved using

Table 2
Teeth segmentation results on test images.

Loss function	Dice	Accuracy	IOU	Recall	Precision
BCE	**0.889**	**94.939**	**0.798**	**0.816**	**0.980**
Dice	0.885	94.895	0.794	0.810	0.977
Tversky	0.874	94.779	0.786	0.804	0.970
Focal Tversky	0.879	94.798	0.791	0.807	0.899
BCE+Dice	0.875	94.795	0.789	0.805	0.971
Combo	0.885	94.895	0.794	0.810	0.977

Table 3
Results on the test images from the Clujana dataset using the U-net architecture.

Loss function	Dice	Accuracy	IOU	Recall	Precision
BCE	0.585	99.221	0.410	0.582	0.669
Dice	0.587	99.242	0.416	0.513	0.687
Tversky	**0.645**	**99.340**	**0.476**	**0.570**	0.743
Focal Tversky	0.554	99.291	0.383	0.419	**0.820**
BCE+Dice	0.569	99.270	0.398	0.459	0.750
Combo	0.614	99.258	0.443	0.562	0.678

Table 4
Results on the test images from the Clujana dataset using the FPN architecture.

Loss function	Dice	Accuracy	IOU	Recall	Precision
BCE	0.572	99.251	0.401	0.477	0.717
Dice	0.597	99.344	0.426	0.512	0.844
Tversky	0.629	**99.350**	0.459	0.524	0.787
Focal Tversky	**0.634**	99.281	**0.464**	**0.592**	**0.683**
BCE+Dice	0.605	99.318	0.434	0.498	0.774
Combo	0.584	99.277	0.413	0.484	0.739

Table 5
Results on the test images from the Clujana dataset using the DeeplabV3 architecture.

Loss function	Dice	Accuracy	IOU	Recall	Precision
BCE	0.558	99.233	0.387	0.460	0.710
Dice	0.569	99.193	00.397	0.506	0.650
Tversky	0.582	99.263	0.411	0.489	0.721
Focal Tversky	**0.621**	**99.310**	**0.450**	**0.538**	**0.735**
BCE+Dice	0.571	99.231	0.407	0.487	0.692
Combo	0.535	99.238	0.365	0.417	0.748

the Tversky loss function. The alpha parameter was set to 0.7 to minimize false negatives, as mentioned in the paper that introduced the Tversky loss [31].

Initially, we experimented with the U-Net architecture for our cavity segmentation task. While U-Net demonstrated promising results, we decided to explore the use of Feature Pyramid Networks (FPN) as an alternative approach. The motivation behind this decision was to leverage FPN's ability to capture multi-scale features and enhance the model's capability to detect cavities at different levels of granularity.

The experimental results are presented in Table 4. The highest dice score of 0.634 was achieved using the Focal Tversky loss, while the Tversky loss led to the highest accuracy of 99.350%.

After evaluating the performance of both FPN and U-Net, which yielded comparable results in our cavity segmentation task, we further expanded our exploration by implementing DeepLabV3. The motivation behind this decision was to leverage DeepLabv3's advanced semantic segmentation capabilities and its ability to handle fine-grained details and object boundaries more effectively. DeepLabV3 utilizes atrous convolutions and atrous spatial pyramid pooling to capture multi-scale context and improve the overall segmentation accuracy.

This time the best results were once again achieved using the Focal Tversky loss, leading to a Dice score of 0.621 and an accuracy of 99.310%. The experimental results can be viewed in Table 5.

Combining the best model from each of the three architectures, we formed an ensemble. The results achieved are better and more accurate than any of the results of the individual models. The ensemble of models yielded promising results in cavity segmentation in

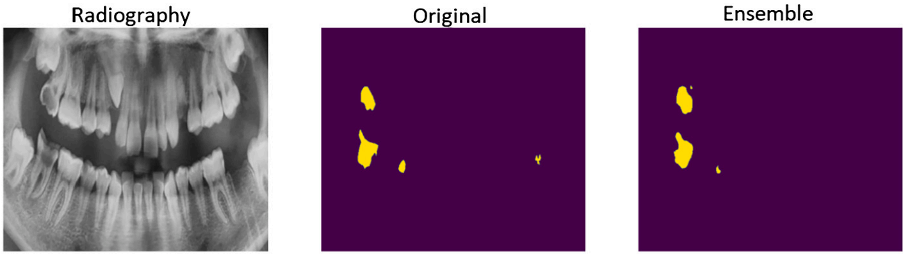

Fig. 6. Example 1 of caries segmentation on the test data, showing the radiography, the original mask, and the mask predicted by the ensemble. Caries were present in the following teeth: 1.7, 3.6, 4.4, and 4.7.

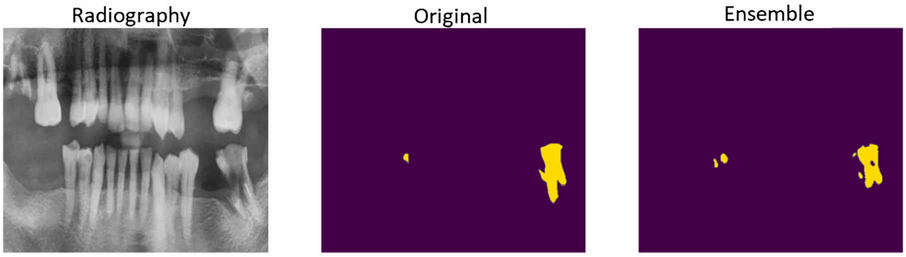

Fig. 7. Example 2 of caries segmentation on the test data, showing the radiography, the original mask, and the mask predicted by the ensemble. Caries were present in teeth 3.7 and 4.4.

radiographs. With an accuracy of 99.42%, the ensemble demonstrated highly accurate classifications. The Dice coefficient of 0.682 indicates a good overlap between predicted and ground truth segmentations. The ensemble resulted in an IOU score of 0.518, a recall of 0.591, and a precision of 0.808. Collectively, these results demonstrate the potential of ensemble approaches to enhance performance in this field.

Two of the best segmentation results on the images used for testing are presented in Fig. 6 and Fig. 7, having a Dice score of 0.832, respectively 0.803.

Because the single models have limitations in capturing the diversity and variability of the data our proposed ensemble approach combines the outputs of three different single models that are well-known and widely used for image segmentation tasks, and they have different architectures and characteristics that complement each other. The probabilistic criterion that we use to fuse the outputs of the models assigns a cavity label to a pixel if the majority of the models agree on it. By evaluating our ensemble approach and comparing it to the individual models, we showed that our ensemble approach outperforms the individual models in terms of accuracy, robustness, and consistency. We obtained promising results with our ensemble approach, which demonstrated highly accurate classifications (99.42%). The Dice coefficient (0.682) indicates a good overlap between the predicted and ground truth segmentations. The ensemble approach also achieved better performances in terms of Intersection over Union (IOU) and Area under the Curve (AUC). IOU and AUC are commonly used metrics to evaluate the quality and reliability of image segmentation methods. For example, a high IOU score means that the predicted and ground truth segmentations have a large overlap, while a high AUC value means that the method can distinguish between cavity and non-cavity pixels with high confidence. Furthermore, we obtained on the test images an average False Positive Rate of 1.34% and an average precision of 75.51%. However, for several images, the False Negative Rate was over 50% (in this case, the algorithms did not identify all the segments/caries previously manually annotated), indicating that the ensemble approach can still be improved, possible by using a cost-sensitive learning methodology that gives higher penalties to false negatives, reflecting their higher cost in the problem domain. Collectively, these results demonstrate the potential of ensemble approaches to enhance performance in cavity segmentation in radiographs. Our ensemble approach is a novel and effective method for cavity segmentation in radiographs, which leverages the strengths of different models and addresses the limitations of single models. Our approach can improve the accuracy and reliability of dental diagnosis and treatment.

4.5. Comparison with state-of-the-art results

We compared the results of our approach with three previously conducted studies. It is important to note the difference in datasets. We had a dataset of 150 images, of which only 108 contained cavities and were used for training or testing. Dayi et al. [3] had a dataset of 504 images, Lian et al. [4] of 1160, and Zhu et al. [5] of 1159. The datasets in the mentioned studies were not made public. Our results are compared to the ones achieved by state-of-the-art studies in Table 6.

It is imperative to emphasize that, due to various constraints, such as unavailability of code, private dataset usage, and inadequately described models, the task of replicating other methods with our dataset becomes a complex undertaking that is beyond the scope of this study. The achievement of superior performance of models such as that described in [5] is attributed to the utilization of an extensive training dataset. However, an impediment arises as the dataset and comprehensive implementation details are not made accessible, precluding the replication or validation of their approach on alternative datasets.

Table 6
Table comparing our experimental results to State of the Art studies.

Author	Dataset	Test set	Accuracy	Dice	IOU	Recall
Our results	108 (private)	14	99.42	68.20	51.8 0	59.10
Dayi et al. [3]	504 (private)	126	NA	62.79	NA	NA
Zhu et al. [5]	1159 (private)	124	93.61	93.64	NA	86.01
Lian et al. [4]	1160 (private)	89	98.60	66.30	78.50	NA

4.6. Discussion

In this section we aim to answer the research questions raised at the beginning of the paper.

RQ1: The proposed ensemble CariSeg leads to improved segmentation of dental cavities in X-ray radiographs. The ensemble of three neural networks leverages the strengths of each architecture, combining their outputs to achieve better segmentation performance and better generalization than any individual model.

RQ2: The accuracy and reliability of AI algorithms for dental cavity detection in X-ray images depend on factors such as the quality and quantity of training data, appropriate loss function and model architecture selection, effective pre-processing, and data augmentation, handling class imbalance, transfer learning, hyperparameter tuning, and post-processing techniques. Optimizing these factors can lead to improved overall performance, enhancing the usefulness of AI algorithms in dental diagnostics and patient care. Possible future improvements are further detailed in Section 4.7.

The strengths of our CariSeg include the utilization of ensemble learning, which capitalizes on the diversity of U-shaped Neural Networks, Feature Pyramid Networks, and DeeplabV3, harnessing their unique architectural characteristics for improved performance. This approach enhances the accuracy and robustness of our caries segmentation. Additionally, the clear separation of the caries segmentation task into two distinct stages, teeth segmentation and lesion segmentation allows for a more focused and accurate analysis of carious lesions.

Our CariSeg does come with some limitations that are important to highlight in the conclusion. Firstly, there is a limitation in terms of the patient sample used for dataset creation. The number of radiographs used for training and the number of images available for each type of lesion in our dataset was limited, which may reduce the dataset's generalizability. Secondly, our image labeling process had its limitations as well. The images were labeled by a single dentist, which might introduce subjective bias at times. Lastly, our study specifically used panoramic radiographs. Certain caries that are not visible in panoramic radiographs may be detectable with other diagnostic approaches or tools, suggesting a potential limitation when it comes to the scope of our caries detection.

4.7. Future improvements

There are several key areas for improvement in cavity segmentation in panoramic radiographs. Firstly, one possibility is to refine the network architectures by exploring variations and creating a new architecture inspired by the ones available today. This includes experimenting with advanced architectural designs and incorporating state-of-the-art techniques to further optimize the models.

Another important aspect is the exploration of diverse data enhancement techniques. By applying filters to the images, we can make caries that are originally hard to detect more visible.

Moreover, additional image augmenting techniques can be applied to the training dataset to further enhance the model's ability to generalize and handle different variations in radiographic images. One other improvement is to apply the image augmentation techniques before cropping the images to reduce the black space added by rotating/shifting the images after they were cropped.

We also recognize the need to address the class imbalance issue between the small number of cavity pixels and the larger number of non-cavity pixels. We did that by cropping the radiographs around the area of interest, the teeth, by using a teeth segmentation model. To further mitigate this, other techniques could be investigated such as using weighted loss functions to ensure a more balanced training process.

One of the most important areas of improvement is expanding the evaluation to larger and more diverse datasets that include all types of caries lesions. This is crucial for assessing the generalizability and robustness of the models. By including data from different dental clinics, populations, and imaging devices, we can validate the performance of our application across a broader range of scenarios. Neural networks heavily rely on high-quality, diverse, and well-labeled training data. Insufficient or biased datasets may result in suboptimal performance, highlighting the need for large-scale, representative datasets.

Another possible future direction is expanding into doing research for the segmentation of other teeth problems, such as apical periodontitis. By applying the same pipeline ensemble approach used here, we can see how well it generalizes and compares the results.

5. Conclusions

In conclusion, this paper presents a comprehensive investigation into the application of neural networks for cavity segmentation in radiographs. We explored different architectures, including U-net, FPN, and DeeplabV3, along with various loss functions such as Binary Cross-entropy, Dice, Tversky, and Focal Tversky. Through extensive experimentation on the Clujana dataset, we obtained promising results. Our ensemble approach, CariSeg, combining the strengths of DeeplabV3, U-net, and FPN, achieved an accuracy of 99.42% and a Dice coefficient of 68.26%. These results indicate the efficacy of our proposed methodology in accurately detecting cavities in radiographs. Future work could focus on refining the network architectures, exploring different data augmentation techniques, expanding the evaluation to larger and more diverse datasets, and finding ways to handle the class imbalance between a small number of cavity pixels and a large number of non-cavity pixels.

Even if various deep-learning models, such as U-net or DeepLab, have been used to address the issue of caries detection in the literature, these models have room for improvement in their performance. Moreover, none of the previous methods employed an ensemble model similar to our approach, which has enhanced the accuracy by combining and correcting the predictions of individual models. Our ensemble approach can generate more trustworthy and confident predictions by averaging independent predictions and can also avoid the over-confidence or under-confidence of individual models.

With continued advancements, AI cavity detection holds great promise for revolutionizing dental care, improving patient outcomes, and promoting preventive strategies.

Ethics statement

This study was reviewed and approved by the Iuliu Haţieganu University of Medicine and Pharmacy Ethics Commission, with the approval number: 34/31.01.2024. Informed consent was not required for this study due to its retrospective nature.

CRediT authorship contribution statement

Andra Carmen Mărginean: Writing – review & editing, Writing – original draft, Software, Investigation. **Sorana Mureşanu:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Mihaela Hedeşiu:** Validation, Supervision, Project administration, Data curation, Conceptualization. **Laura Dioşan:** Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

Code is available at: https://github.com/am2731/cavities-segmentation. Panoramic radiography dataset can be made available upon reasonable request to the corresponding author.

Acknowledgements

This work was supported by project PNRR-III-C9-2023-I8, “Technologically Enabled Advancements in Dental Medicine (TEAM)”, CF.80/31.07.2023, number 760235/28.12.2023.

References

[1] GBD2017-Oral-Disorders-Collaborators, E. Bernabe, W. Marcenes, C. Hernandez, J. Bailey, L. Abreu, V. Alipour, S. Amini, J. Arabloo, Z. Arefi, et al., Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the global burden of disease 2017 study, 2020.
[2] A. Wenzel, Radiographic modalities for diagnosis of caries in a historical perspective: from film to machine-intelligence supported systems, Dentomaxillofacial Radiol. 50 (5) (2021) 20210010.
[3] B. Dayı, H. Üzen, İ.B. Çiçek, Ş.B. Duman, A novel deep learning-based approach for segmentation of different type caries lesions on panoramic radiographs, Diagnostics 13 (2) (2023) 202.
[4] L. Lian, T. Zhu, F. Zhu, H. Zhu, Deep learning for caries detection and classification, Diagnostics (Basel) 11 (9) (Sep 2021).
[5] H. Zhu, Z. Cao, L. Lian, G. Ye, H. Gao, J. Wu, Cariesnet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic x-ray image, Neural Comput. Appl. (2022) 1–9.
[6] A.J. Larrazabal, C. Martínez, J. Dolz, E. Ferrante, Orthogonal ensemble networks for biomedical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24, Springer, 2021, pp. 594–603.
[7] Z. Faska, L. Khrissi, K. Haddouch, N. El Akkad, A robust and consistent stack generalized ensemble-learning framework for image segmentation, J. Eng. Appl. Sci. 70 (1) (2023) 74.
[8] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, arXiv:1505.04597, 2015.
[9] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, arXiv:1612.03144, 2017.
[10] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv:1706.05587, 2017.
[11] S. Zhang, M. Liu, J. Yan, The diversified ensemble neural network, Adv. Neural Inf. Process. Syst. 33 (2020) 16001–16011.
[12] W. Li, R.C. Paffenroth, D. Berthiaume, Neural network ensembles: theory, training, and the importance of explicit diversity, arXiv preprint, arXiv:2109.14117, 2021.
[13] S. Zhou, J. Wang, L. Wang, X. Wan, S. Hui, N. Zheng, Inverse adversarial diversity learning for network ensemble, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[14] L.G.K. Zanini, I.R.F. Rubira-Bullen, F.d.L.d.S. Nunes, A systematic review on caries detection, classification, and segmentation from x-ray images: methods, datasets, evaluation, and open opportunities, J. Imaging Inform. Med. (2024) 1–22.

[15] H. Mohammad-Rahimi, S.R. Motamedian, M.H. Rohban, J. Krois, S.E. Uribe, E. Mahmoudinia, R. Rokhshad, M. Nadimi, F. Schwendicke, Deep learning for caries detection: a systematic review, J. Dent. 122 (2022) 104115, https://doi.org/10.1016/j.jdent.2022.104115.
[16] M.M. Srivastava, P. Kumar, L. Pradhan, S. Varadarajan, Detection of tooth caries in bitewing radiographs using deep learning, arXiv:1711.07312, 2017.
[17] 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, https://doi.org/10.1016/j.jdent.2020.103425.
[18] X. Lin, D. Hong, D. Zhang, M. Huang, H. Yu, Detecting proximal caries on periapical radiographs using convolutional neural networks with different training strategies on small datasets, Diagnostics 12 (5) (2022), https://doi.org/10.3390/diagnostics12051047, https://www.mdpi.com/2075-4418/12/5/1047.
[19] I.S. Bayrakdar, K. Orhan, S. Akarsu, Ö. Çelik, S. Atasoy, A. Pekince, Y. Yasa, E. Bilgir, H. Sağlam, A.F. Aslan, A. Odabaş, Deep-learning approach for caries detection and segmentation on dental bitewing radiographs, Oral Radiol. 38 (4) (2022) 468–479, https://doi.org/10.1007/s11282-021-00577-9.
[20] S.S. Alharbi, A.A. AlRugaibah, H.F. Alhasson, R.U. Khan, Detection of cavities from dental panoramic x-ray images using nested u-net models, Appl. Sci. 13 (23) (2023) 12771.
[21] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. Talo Yildirim, O. Yildirim, O. Faust, U.R. Acharya, An explainable deep learning model to prediction dental caries using panoramic radiograph images, Diagnostics 13 (2) (2023), https://www.mdpi.com/2075-4418/13/2/226.
[22] K. Panetta, R. Rajendran, A. Ramesh, S. Rao, S. Agaian, Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems, IEEE J. Biomed. Health Inform. 26 (4) (2022 Apr) 1650–1659, https://doi.org/10.1109/JBHI.2021.3117575, epub 14 April 2022.
[23] A.H. Abdi, S. Kasaei, Panoramic dental x-rays with segmented mandibles, Mendeley Data, V2, https://doi.org/10.17632/hxt48yk462.2, online; 2 July 2020.
[24] I. Hell, A. Hamaci, Tooth instance segmentation on panoramic dental radiographs using u-nets and morphological processing, DUBİTED 10 (1) (2022) 39–50.
[25] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, arXiv:1606.00915, 2017.
[26] S. Ruder, An overview of gradient descent optimization algorithms, CoRR, arXiv:1609.04747, 2016.
[27] U. Ruby, V. Yendapalli, Binary cross entropy with deep learning technique for image classification, Int. J. Adv. Trends Comput. Sci. Eng. 9 (10 2020), https://doi.org/10.30534/ijatcse/2020/175942020.
[28] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer International Publishing, 2017, pp. 240–248.
[29] S.R. Hashemi, S.S.M. Salehi, D. Erdogmus, S.P. Prabhu, S.K. Warfield, A. Gholipour, Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: application to multiple sclerosis lesion detection, IEEE Access 7 (2019) 1721–1735, https://doi.org/10.1109/access.2018.2886371.
[30] S.A. Taghanaki, Y. Zheng, S.K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: handling input and output imbalance in multi-organ segmentation, CoRR, arXiv:1805.02798, 2018.
[31] N. Abraham, N.M. Khan, A novel focal Tversky loss function with improved attention u-net for lesion segmentation, CoRR, arXiv:1810.07842, 2018.