PaXNet: Dental Caries Detection in Panoramic X-ray using Ensemble Transfer Learning and Capsule Classifier

Arman Haghanifar*‡, Mahdiyar Molahasani Majdabadi†‡, Seok-Bum Ko*†
* Div. of Biomedical Engineering, University of Saskatchewan
† Dept. of Electrical and Computer Engineering, University of Saskatchewan
‡ Both authors contributed equally to this manuscript

Abstract

Dental caries is one of the most chronic diseases involving the majority of the population during their lifetime. Caries lesions are typically diagnosed by radiologists relying only on their visual inspection to detect via dental x-rays. In many cases, dental caries is hard to identify using x-rays and can be misinterpreted as shadows due to different reasons such as low image quality. Hence, developing a decision support system for caries detection has been a topic of interest in recent years. Here, we propose an automatic diagnosis system to detect dental caries in Panoramic images for the first time, to the best of authors' knowledge. The proposed model benefits from various pretrained deep learning models through transfer learning to extract relevant features from x-rays and uses a capsule network to draw prediction results. On a dataset of 470 Panoramic images used for features extraction, including 240 labeled images for classification, our model achieved an accuracy score of 86.05% on the test set. The obtained score demonstrates acceptable detection performance and an increase in caries detection speed, as long as the challenges of using Panoramic x-rays of real patients are taken into account. Among images with caries lesions in the test set, our model acquired recall scores of 69.44% and 90.52% for mild and severe ones, confirming the fact that severe caries spots are more straightforward to detect and efficient mild caries detection needs a more robust and larger dataset. Considering the novelty of current research study as using Panoramic images, this work is a step towards developing a fully automated efficient decision support system to assist domain experts.

Index Terms

dental caries, panoramic radiography, deep learning, convolutional neural networks

I. Introduction

Dental caries, also known as tooth decay, is one of the most prevalent infectious chronic dental diseases in humans, affecting individuals throughout their lifetime [1]. According to the National Health and Nutrition Examination Survey, dental caries involves approximately 90% of adults in the United States [2], [3]. Dental caries is a dynamic disease procedure resulting from dental biofilm's metabolic activity, which gradually demineralizes enamel and dentine [4]. Tooth decay is a preventable disease, and if detected, can be stopped and potentially reversed in its early stages [5]. A standard tool for radiologists to distinguish dental diseases, such as caries, is x-ray radiography. X-rays are an essential complementary diagnosis solution to help identifying dental problems that are hard to detect via visual inspection only [6].

Radiography is one of the methods usually performed to help radiologists with the oral health assessment and diagnosis of dental diseases, such as proximal dental caries [7]. Since proximal tooth surfaces are hard to be approached or visualized directly, caries lesions in these surfaces are diagnosed with the aid of radiographs [8]. There are several dental x-ray types, each of which records a different view of dental anatomies, such as Bitewing, Panoramic, and Periapical. Basically, Bitewing radiography is the most widely used method for caries detection and has the highest diagnostic accuracy, while Panoramic has the lowest [9]. However, Bitewing has several disadvantages, like patient discomfort. Due to operators' insufficient expertise, Bitewing imaging usually results in increased patient radiation dose because of the need for image retakes [10]. On the other hand, Panoramic x-ray images, also known as Orthopantomogram (OPG), are widely used to capture the entire mouth using a very small dose of radiation [11]. OPG has a low radiation dose, simplicity of application, less time requirement, and also great patient comfort. Thus, pediatric, handicapped, and senior patients would benefit greatly from a Panoramic imaging system compared to intraoral systems [12].

Since a Panoramic image covers the entire patient dentition along with surrounding bones and jaw structure, it can not give a detailed view of each tooth. Hence, structures in Panoramic images lack specific boundaries, and visual quality is extremely low in comparison with other types of dental radiographs [13]. Besides, Panoramic images also include other parts of the mouth, such as jawbones, which make the image analysis difficult [14]. Hence, image preprocessing steps and teeth extraction are needed to facilitate visual interpretations and enhance model performance on dental disease detection [13]. A limited number of research studies have been conducted on extracting single tooth from Panoramic images using various methods, such as deep learning [15] or evolutionary algorithms [16]. The latter mentioned research is introduced as the first step toward creating a fully automated decision support system. In this work, the teeth extraction model is expanded to an end-to-end caries diagnosis

This work is the extended version of "Automated Teeth Extraction from Dental Panoramic X-Ray Images using Genetic Algorithm," [16] in 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, Oct. 2020.

model. This system has been utilized in order to provide the required data for training the proposed classifier in this paper. Bringing all together, Panoramic imaging is an affordable method accessible to most patients, covering a large maxillofacial segment including all the teeth. However, these images are noisy, low-resolution, and also need further preprocessing steps.

Recently, a number of research studies have proposed deep learning-based Computer-Aided Diagnosis (CAD) systems to detect dental caries based on various types of data, including clinical assessments [17], infrared light [18], or near-infrared transillumination imaging [19]. Since x-ray radiography is the most common imaging modality in dental clinical practice, the majority of studies have utilized x-rays to develop decision support systems for tooth decay diagnosis. Srivastava *et al.* developed a deep fully Convolutional Neural Network (CNN)-based CAD system and applied their model on a large dataset of 3000 Bitewing x-rays to detect dental caries, which outperformed certified dentists in terms of overall f1-score [20]. Regarding Periapical x-rays, there have been some works in recent years. In 2016, Choi *et al.* trained a model based on simple CNN architecture along with crown extraction algorithm using 475 Periapical images to boost the detection rate of proximal dental caries [21]. Later in 2018, Lee *et al.* used a pretrained Inception V3 for transfer learning on a set of 3000 Periapical images to diagnose dental caries [22]. In the most recent research study, Khan *et al.* benefited from a specialist-labeled dataset of 206 Periapical radiographs and trained a U-Net to segment three different dental abnormalities, such as caries [23]. While there have been some works utilizing OPG images, such as classification of tooth types with a Faster-Regional Convolutional Neural Network (Faster-RCNN) [24], to the best of authors' knowledge, there are no studies applying deep neural networks on Panoramic images to detect dental caries.

Teeth extraction is an essential part needed for developing an automatic dental caries detection system that helps increasing model accuracy by preparing extracted tooth images as model input. Teeth extraction or isolation is the process of extracting image parts, each containing one tooth's boundaries, from a dental x-ray image that also contains other unwanted parts of the mouth, like gingivae or jawbones. Automatic teeth extraction module eliminates the need for manual annotation of teeth in Panoramic images needed for both developing dental disease detection systems or training deep learning-based teeth segmentation models. In [16], we have proposed a novel genetic-based approach for teeth extraction in Panoramic dental images. Based on a dataset of Panoramic x-rays, our teeth extractor could isolate teeth with accuracy scores in line with previous works utilizing either Bitewing or Periapical x-rays. Various methods were introduced for jaw separation, teeth isolation, and accuracy improvement of the system. Considering the aforementioned teeth extraction system as the preliminary step to prepare single tooth images for the disease classification model, current study aims to extend the previous system to build an end-to-end caries detection system where a digital OPG image is the input and teeth suspicious of having caries are the final output.

Since there have been limited attempts to develop automatic deep learning-based dental disease diagnosis systems, there is a need to perform further research in this area. The objective of this study is to develop a specialized model architecture based on pretrained models and the capsule network to detect tooth decay on Panoramic x-rays efficiently. To the best of authors' knowledge,

- This research is the first to perform teeth extraction as well as dental caries detection on Panoramic images, using a relatively large dataset. Most previous works addressed other types of dental x-rays with higher quality in terms of noise level and resolution.
- Genetic algorithm is applied for the first time to isolate teeth in Panoramic images. Previous studies rely mainly on manually defined methods. This evolutionary algorithm demonstrates robust performance even on challenging jaws with several missing teeth.
- Capsule network is used for the first time as the classifier for dental caries diagnosis. Experimental results demonstrate its superiority over CNNs because of the fact that the Capsule network is capable of learning the geometrical relationships between features.
- Feature extraction module is constructed by a voting system from different pretrained architectures. CheXNet [25] is applied for the first time in dental disease detection tasks using x-rays.

The paper is structured as follows. After the introduction, section II is about the materials used for this study and the labeling process required. Section III explains the model architecture both for feature extraction and classification parts. Results are presented and thoroughly discussed in section IV. Finally, section V is the conclusion.

II. Materials and Dataset Preparation

Most related studies in the field of dental problem detection using x-rays lack a sufficient number of images in their datasets. Large datasets let the models have more sophisticated architectures, including more parameters. Hence, developed models can handle more complicated features and detect subtle abnormalities that appeared in the tooth texture, such as dental caries in the early stages. Annotation is an essential and time-consuming part that needs to be performed by the field specialists, e.g. dentists or radiologists.

A. Dataset Collection

Our dataset of 470 Panoramic x-rays is collected from two main sources along with a few number of images from publicly available resources. 280 images are obtained from the Diagnostic Imaging Center of the Southwest State University of Bahia

(UESB) [26]. Images are acquired from the x-ray camera model ORTHOPHOS XG 5 DS/Ceph from Sirona Dental Systems GmbH. Images are randomly selected from different categories with an initial size of 1991×1127. All images are in "JPG" format. Annotation masks related to the images are available in the UESB dataset. An example OPG image from the dataset is shown in Fig. 1.

(a) (b)

Fig. 1: (a) An example image from the UESB dataset with (b) the related tooth annotation mask [26]

Besides, we also collected 120 images from a local dentistry clinic taken with x-ray camera model Cranex®3D from Soredex. Images are anonymized and obtained with a contrast pre-enhancement applied by the radiologist. All images are in "BMP" format with an initial size of 3292×1536 with the bit-depth of 8. 42 images from our dataset are randomly selected to validate the performance of the teeth extraction system. 70 images are downloaded from publicly available medical image sources, such as Radiopaedia [1]. Public images are taken with unknown camera models and are available in different sizes and formats. Some of the collected images are associated with a radiologist report indicating locations of tooth decays. The dataset includes 11769 images from single tooth, which is used for training the encoder model. Among the dataset, 240 images have labeled teeth, including 742 carious and 5206 healthy.

B. Labeling Process

UESB images have tooth masks manually prepared along with the dataset, which is used to extract each tooth from Panoramic images. For our collected dataset, a genetic-based method is applied to isolate the teeth by finding the optimum lines which fall inside gaps between teeth in both maxillary, and mandibular jaws [16]. To perform labeling, a radiologist commented on Panoramic images one by one. Extracted teeth are categorized into two groups; healthy and carious. Carious teeth are also classified into mild and severe caries. Mild ones are caries lesions in their early stages, mostly located in enamel or Dentine-Enamel Junction (DEJ). In contrast, severe decays are developed dental plaques that have been spread to the internal dentine or have involved the pulpitis. Pulpitis caries lesions result in a collapsed tooth.

The labeling process is a time-consuming and challenging task that requires huge amount of time. Since Panoramic images include all teeth in one image, it helps radiologists meticulously detect caries by not only inspecting opaque areas on the tooth but also considering the type of the tooth and its location in the jaw. On the other hand, higher levels of noise and shadows make the visual diagnosis more challenging. Caries, especially mild ones, can easily be misinterpreted as shadows and vice versa. Another problem is the Mach effect. Mach effect or Mach bands is an optical phenomenon that makes the edges of darker objects next to lighter ones appear lighter and vice versa. Mach effect results in a false shadow that may bring diagnostic misinterpretation with dental caries present very close to dental restoration regions that are appeared to be whiter in dental x-rays [27].

III. Model Architecture

In this section, firstly Genetic Algorithm (GA) is introduced, and its usage in the field of image processing is briefly discussed. A teeth extraction system is then presented, and details of different modules are investigated one by one. Afterwards, the architecture and advantages of using capsule network are reviewed. Finally, the feature extraction unit's architecture and the classifier are explained, followed by a detailed illustration of the proposed PaXNet.

A. Genetic Algorithm

A genetic algorithm is considered a blind optimization technique that mimics natural evolution's learning process using selection, crossover, and mutation. These three procedures are transformed into numerical functions to help solve an optimization problem without calculating derivatives. Using a random exploration of the search space, GA is more robust in terms of being

[1] https://radiopaedia.org

stuck in the local extrema [28]. In image processing, GA is proven as a powerful search method that converts an image segmentation problem into an optimization problem [29].

In this research study, teeth are isolated using separator lines that pass the gap between two teeth next to each other. To perform the task, GA is proposed to explore the solution space, in the sense that separator lines fit into the paths with the lowest integral intensity. These paths are considered as including gaps between proximal teeth surfaces. The GA can dynamically change the image segmentation task controlling parameters to reach the best performance. The GA-based teeth extraction process is summarized as follows:

- Initial population: A number of random lines with a limited degree of freedom are considered with a certain distance from each other
- Cost function: The integral intensity projection of all the pixels in each line is considered as the cost function which needs to be minimized; meaning that the lines pass through the darkest available path in the area of proximal surfaces of two neighbor teeth. Cost function is formulated as follows:

$$C(x) = \sum_{i=1}^{n} I(x_i) \tag{1}$$

  where x is the position of each line, n is the number of lines, and $I(x_i)$ is the average of the intensity of the pixels on the line x_i.
- Genetic cycle: Produced lines are changed during iterations and are ranked based on the above-mentioned cost function. Crossover and mutation functions are specified as Scattered and Gaussian, respectively. This cycle is performed iteratively until the maximum fitness or one of the termination criteria is reached. The genetic cycle is shown in Fig. 2.

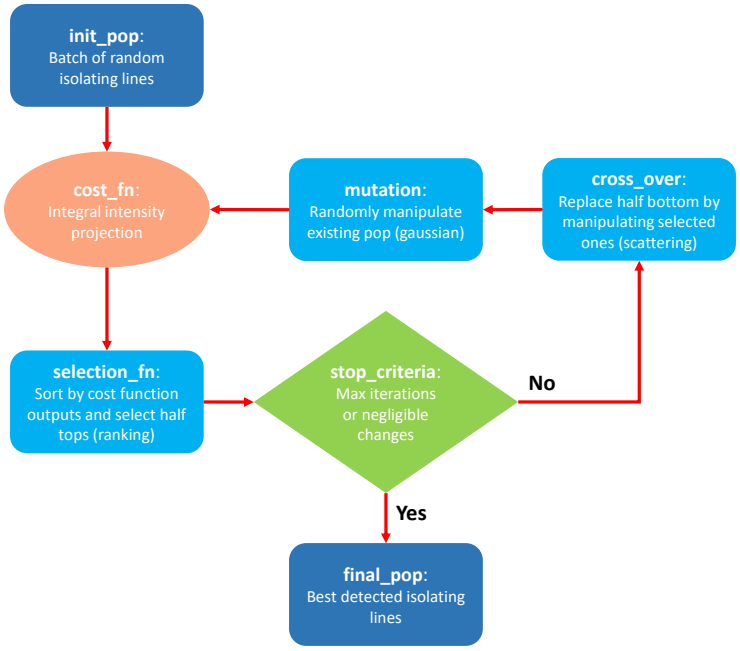


Fig. 2: Flow chart of the proposed GA-based teeth isolation method

B. Teeth Extraction

Since Panoramic images contain unwanted regions around the jaw, some segmentation algorithms should be applied to the image before performing jaw separation. Unlike Periapical images, in Panoramic images, upper jaws (maxilla) and lower jaws (mandible) need to be separated apart before tooth isolation. Each step requires its own preprocessing method to help increase performance efficiency. A detailed high-level illustration of the proposed teeth extraction pipeline is depicted in Fig. 3.

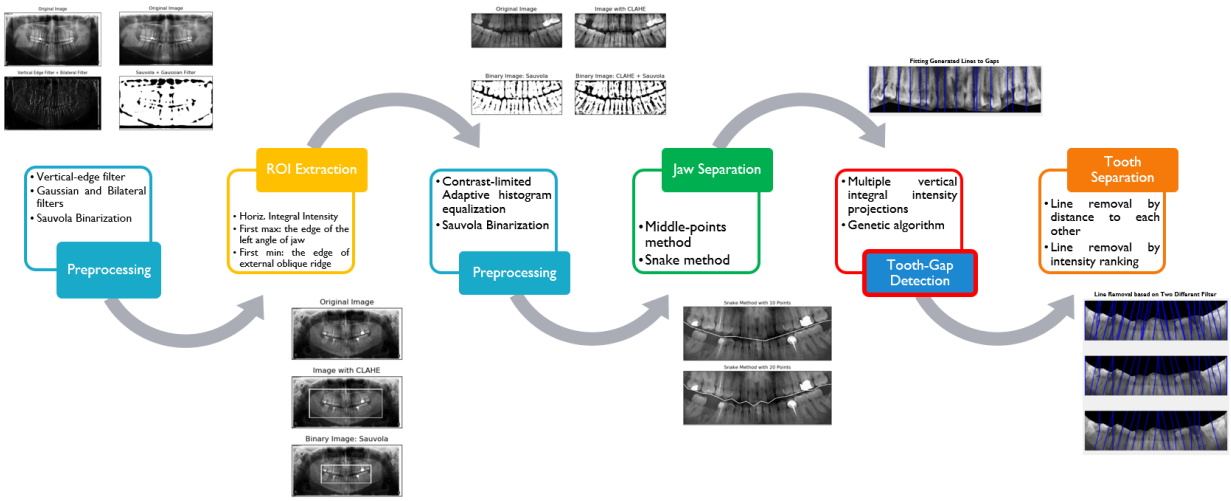


Fig. 3: Detailed diagram of the proposed teeth extraction system

As shown in Fig. 3, Panoramic raw images are the input of the system. Preprocessing, ROI extraction, and jaw separation are done, one after another, to make the image ready for the genetic algorithm to generate vertical lines for teeth separation. Line removal is then applied to output images, each including one tooth. These generated images are then fed to the PaXNet to detect dental caries and their severity. A brief description of teeth extraction methodology is described in the following subsections.

1) Preprocessing: To enhance the image quality and highlight useful details, preprocessing steps have been taken. For extracting the jaw as the Region of Interest (ROI) within the main image, two steps are considered. To extract initial ROI, a vertical-edge-detecting filter is applied to highlight the vertical edges. Then, a bilateral filter is used for sharpening the edges. After cropping the initial ROI, the Sauvola algorithm is applied to binarize the image in order to blacken the two ends of the gap between maxilla and mandible. Finally, a Gaussian filter is used to reduce the noise.

2) ROI Extraction: To reach the jaw from the main image, unwanted surroundings need to be discarded. ROI extraction is performed in two steps. After the preprocessing for the first step described above, horizontal integral intensity projection is applied to the image. The first significant positive slope represents the edge of the left angle of the jaw. On the next step, horizontal integral intensity projection is performed on the cropped image to distinguish the external oblique ridge, the starting point for the gap between maxilla and mandible, which appears to be the first significant negative slope in the intensity graph.

3) Jaw Separation: After extracting the jaw from the main image, maxilla and mandible are expected to become separate using a line that passes through the gap while maintaining the biggest possible distance from both jaws. Two procedures are employed to create the separating line: middle points and snake. In the middle points method, the image is divided into several parts. Then, vertical integral intensity is computed, and the minimum value is supposed to be the point representing the gap. The final line is eventually formed by connecting these points. Another method is the snake algorithm. First, a starting point is determined. Afterwards, it crawls through both left and right directions, looking for paths with minimum integral intensity. Each path-step length is a controlling parameter to restrict the snake from getting trapped in teeth-gap valleys.

4) Tooth Isolation: The last step is separating maxilla and mandible into a batch of isolated teeth. The process is similar to the jaw separation, whereas here, we have multiple lines. To find all the lines simultaneously and without any predefined parameters, a GA-based method is employed. Tooth morphology varies among the dentition, and the genetic algorithm can detect the best fitting lines due to its randomness. At first, 30 vertical lines are randomly generated over each jaw's image as the chromosomes of the initial population. The genetic cycle discussed earlier in subsection III-A is then performed to find the best population indicating lines fitting inside gaps between teeth. Since generated lines are more than available gaps, various line removal methods are implemented afterwards to reach the number of lines precisely equal to the number of gaps in each jaw.

C. Capsule Network

Since the introduction of capsule network [30], many studies have benefited from its advantages in various applied deep learning tasks [31]–[34]. Unlike convolutional networks, capsule layers represent each feature using a vector in the way that the length of the vector corresponds to the probability of the presence of a certain feature or class. A weight matrix is multiplied to each vector to predict the probability and the corresponding pose of the next level feature in the form of a multidimensional vector. Then, an algorithm called dynamic routing is applied to all the predictions of one class to determine the coherency of the predictions. This algorithm calculates a weighted average of predictions and reduces the impact of those vectors incoherent with others, iteratively. Since the average vector's length represents the probability of the class, it should be ranged between

0 and 1. In order to make sure of that, the Squash function is applied to the prediction vector after each iteration of dynamic routing, as follows:

$$V_j = \frac{\|S_j\|^2}{1+\|S_j\|^2}\frac{S_j}{\|S_j\|} \tag{2}$$

where, S_j is the vector after dynamic routing. Fig 4 indicates the structure of the tow-layer Capsule network utilized for caries detection.

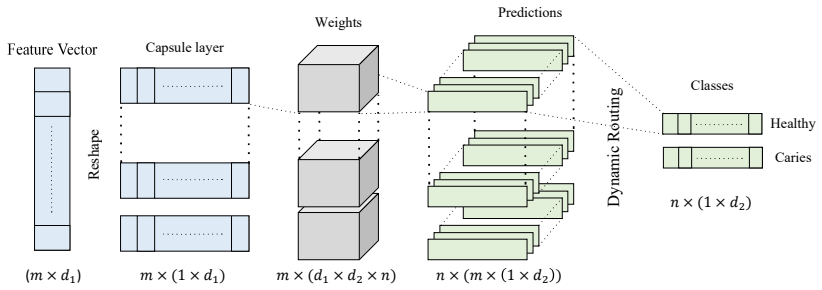


Fig. 4: The architecture of the CapsNet for binary caries classification from extracted features

Where m is the number of capsules in the first layer, d_1 is the size of each capsule in the first layer, n is the number of capsules in the number of classes (2 in this case), and d_2 is the size of each capsule in the second layer. The feature vector is the input of the CapsNet. This feature vector is extracted from feature extraction unit, which will be reviewed in detail in III-D1. The output of the CapsNet is two vectors representing two classes, healthy and caries. The length of these two vectors corresponds to the probability of each class.

Moreover, this network is capable of learning the geometrical relationship between features. Hence, it can offer unprecedented robustness in a variety of tasks [35]–[37]. The main challenge in dental caries detection in Panoramic images is distinguishing between dark parts caused by shadows and caries accurately. Geometrical features such as edges, textures and localization of these darker regions are essential for correct classification. This is why we believe the capsule network is the perfect choice for this task.

D. Proposed Panoramic Dental X-ray Network: PaXNet

The proposed network for caries detection in Panoramic dental x-rays (PaXNet) consists of two main modules; feature extraction and classifier. The aforementioned modules are explained in detail in the following subsections.

1) Feature Extraction: As far as feature extraction is concerned, to overcome the problem of the small number of samples in the dataset, the transfer learning paradigm is used in the proposed architecture. Three pretrained models are utilized in the feature extraction block: Encoder, CheXNet, and InceptionNet. All of these three models are non-trainable, and their top layers are excluded. InceptionNet is a powerful model trained on the ImageNet dataset [38]. Since caries appears in different sizes, this model's ability to learn features with multiple sizes is beneficial for caries detection. However, the model is trained on RGB images that share no similarities with Panoramic dental x-rays. This is why CheXNet is included in feature extraction block as well. CheXNet is a robust model for lung disease detection based on chest x-ray images [25]. This model is trained on CXR-14, the largest publicly available dataset of chest x-rays from adult cases with 14 different diseases. Although Panoramic dental x-rays are different from chest x-rays, since they are both x-rays, they share many similar features. There might be some features specific to a tooth that is not covered by the two models mentioned above. An encoder is used to benefit from these types of features.

The number of labeled tooth images is limited and expanding the dataset requires time and effort of specialists. In contrast, there is a considerable number of unlabeled images available. To benefit from this vast dataset, a new approach is proposed. An unsupervised side-task is designed, and a model is trained using an unlabeled dataset. Then, trained weights are used in the main model for the caries detection task. Through this approach, transfer learning enables us to take advantage of the unlabeled data as well. An auto-encoder is developed and trained to encode the image and reconstruct it from the coded version in this work.

The auto-encoder consists of two networks, Encoder and Decoder. Both of these models are used in PaXNet through transfer learning. Since all the image information should be preserved through the encoding process, an encoder can learn the most informative tooth images' features. Later, this model is used as a pretrained network for feature extraction in PaXNet. The model is benefiting from the decoder as well as the encoder, as it is explained in the next subsection.

Finally, since all these three models are non-trainable, a trainable convolutional feature extractor is embedded in PaXNet so that the model could learn task-related specific features as well. Table I presents a comparison between these four feature extraction units.

TABLE I: Comparison between feature extractor models

Model	Number of Layers	Number of Parameters	Trainable	Dataset	Dataset size
InceptionNet	42	451,160	non-trainable	ImageNet	1,281,167
CheXNet	140	1,444,928	non-trainable	CXR-14	112,000
Encoder	4	14,085	non-trainable	Unlabeled extracted teeth	11,769
CNN	8	35,808	trainable	-	-

As the similarity of the training dataset to the cries section dataset increases, the number of available samples is receded. However, more informative features can be obtained from these networks trained with more similar samples.

As far as the activation function is concerned, all convolutional layers benefit from the Swish activation function. This function is a continuous replacement of leaky-ReLU, which improves the performance of the network [39]. The swish activation function is formulated as follows:

$$f(x) = \frac{x}{1 + e^{-x}} \tag{3}$$

This activation function is basically the multiplication of the Sigmoid function with the input value. The behaviour of this activation function is similar to the ReLU and leaky-ReLU in positive values. However, in large negative values, its output is converging to zero, unlike Leaky-ReLU.

2) Classifier: In PaXNet, all extracted features are concatenated, and higher-lever features are created based on them using a CNN. Then, the last convolutions layer is flattened, followed by a fully-connected layer with 180 neurons. These 180 neurons are reshaped to 10 capsules with eight dimensions called the primary capsule layer. There are two 32 dimensional capsules in the second capsule layer representing two classes, caries and healthy. Each capsule in the primary capsule layer makes a prediction for each capsule in the second layer. Routing by agreement is performed on these predictions for three iterations. Each vector's length is then computed, and a softmax function is applied to these two values. The output of the softmax layer is the probability of each class.

Furthermore, the capsule corresponds to the class with a higher probability is extracted using a mask. This 32D value is passed to a CNN followed by the decoder. The decoder is extracted from the auto-encoder explained in subsection III-D1. It is suggested by [30] that image reconstruction can improve the capsule network's performance. Since the dataset is relatively small, the network is not able to learn the proper image reconstruction. So the decoder is utilized, and CNN is responsible for mapping the latent space of the capsule network to the decoder's latent space.

The aforementioned feature extractor and classifier modules are assembled together to form PaXNet. The high-level architecture of the proposed model is illustrated in Fig. 5.

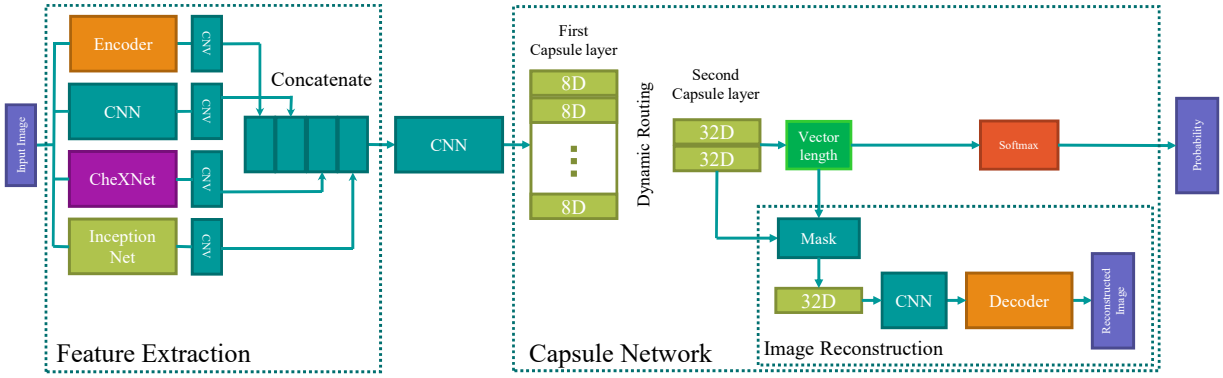


Fig. 5: Architecture of the proposed PaXNet, consisting of four networks in feature extraction module and a 2-layered capsule network to generate probability, connected with a CNN.

First, informative and valuable features of the input image are located using the feature extraction unit. Then, all these features are combined to form higher level and more complex features by a CNN. Finally, CapsNet classifies the input into two healthy or caries classes based on the extracted features. The Image reconstruction module is implemented for the training purpose. The difference between the reconstructed image and the input is used as a loss function in order to force the model to learn better features [30].

IV. Results and Discussions

In this section, the performance of each module in the proposed teeth extraction approach is briefly addressed. Also, the proposed teeth extraction system is compared with other studies using different types of dental x-rays. Worth mentioning that as far as the accuracy is concerned, the ratio of correctly separated segments to the total number of parts is considered as

the accuracy reported for teeth separation. Next, the results of caries detection with PaXNet is presented. Then, performance of the model on detecting caries in different stages is investigated. The contribution of each feature extractor network is also addressed by visual illustration of the proposed model's robustness.

A. Experimental results

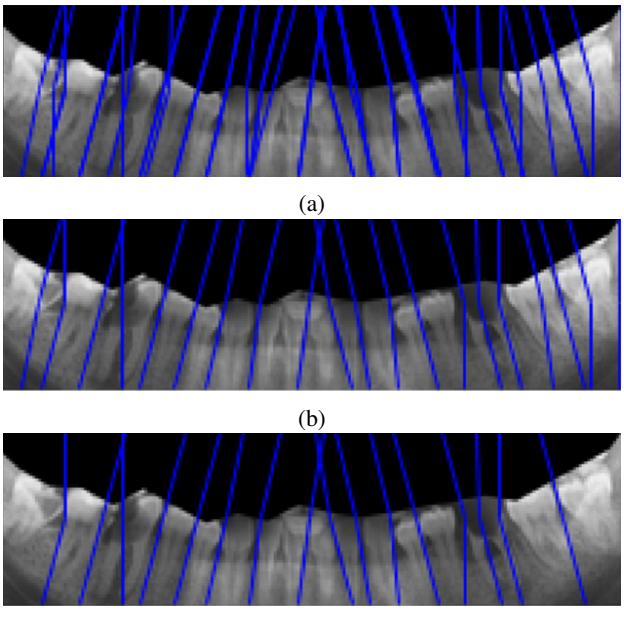
For the teeth extraction task, the algorithm is applied on 42 Panoramic images, where the total number of teeth is 1229; 616 maxillary and 613 mandibular. Jaw region separation from the surrounding unwanted area is performed on the dataset, and the success rate of 40 out of 42 images is achieved. Hence, jaw extraction accuracy score is 95.23%. It also failed to identify some of the wisdom teeth correctly. Thus, 4 maxillary and 2 mandibular wisdom teeth were also missed. The final number of remaining teeth is 582 in maxilla and 581 in mandible jaws.

Next, jaw separation is applied to 40 images comparing middle points and snake methods explained in the previous section. While middle points approach fails to separate the jaws correctly, and the crowns of one or more teeth are misclassified in many samples, snake approach demonstrates better performance and consistency. It also proves to work well even on closely-stacked-together jaws.

After jaw separation, genetic algorithm is applied on extracted jaws, followed by line removal techniques to reduce the number of wrong lines, mostly passing through the teeth instead of teeth-gap valleys. Final results on maxillary and mandibular teeth, after applying line removal techniques, are presented in Table II. Sample tooth isolation result after initial and final line removal steps is shown in Fig. 6.

TABLE II: Accuracy of tooth isolation in maxilla and mandible

Jaw Type	Total Teeth	Isolated Teeth	Accuracy
Maxilla	582	474	81.44%
Mandible	581	428	73.67%



(a) (b) (c)

Fig. 6: Tooth extraction procedure for a sample jaw: (a) preliminary genetic algorithm output (b) initial line removal output (c) final line removal output

In the maxilla, the proposed algorithm works great isolating incisor teeth; however, molar and premolar teeth performance is mediocre. Most of the time, two central incisors are missed due to the lack of clarity in boundaries. This lack of transparency, like dark shadows, is typical for central incisors in Panoramic images, which is caused by the imaging method deficiency. To improve accuracy in mandibular teeth, model capability is increased by changing the cost function and line removal technique when dealing with mandibular images. Table III briefly explains a comparison between our proposed teeth extraction method and other relevant research studies.

TABLE III: Comparison between proposed method and other teeth extraction research works

Tooth Extraction System	Image Type	Algorithm	Correct Upper	Correct Lower
Abdel-Mottaleb et al. [40]	Bitewing	Integral Intensity Method	169/195 - 85%	149/181 - 81%
Olberg and Goodwin [41]		Path-based Method	300/336 - 89.3%	270/306 - 88.2%
Nomir et al. [42]		Integral Intensity Method	329/391 - 84%	293/361 - 81%
Al-Sherif [43]		Energy-based Method	1604/1833 - 87.5%	1422/1692 - 84%
Ehsani Rad et al. [44]	Periapical	Integral Intensity Method	Overall: 90.83%	
Proposed Study	Panoramic	Genetic-based Method	474/582 - 81.44%	428/581 - 73.67%

Although we are applying teeth extraction on Panoramic images, which are noisy, include unwanted parts, and the structure boundaries of segments are ambiguous, acquired accuracy is in line with previous studies. Extracted teeth along with tooth images from the UESB dataset are labeled and used as the input to the proposed caries detection network.

PaXNet is trained using 319 samples with caries and 1519 healthy samples. In order to deal with class imbalance, the smaller class is re-sampled. Then, 80% of data is used for training and 20% is excluded as test set. Moreover, data augmentation is applied to the samples in the training process. Each sample in the dataset is rotated randomly and a random zoom and shift are applied as well, as listed in Table IV.

TABLE IV: Image augmentation functions

Attribute	Parameter	Value
Rotation	Angle	0° to 90°
Flip	Axis	Vertical and Horizontal
Brightness	Scale	70% to 130%
Zoom	Scale	90% to 150%
Width Shift	Scale	-20% to 20%
Height shift	Scale	-20% to 20%

A rotation range of 90° covers the total possible rotation range with the help of horizontal and vertical flip. Darkening or brightening an image with a large scale will result in information loss, hence we adjust the brightness to mostly 30% higher or lower than the raw input image. Zooming out of the image will help the model to see caries lesions with different scales, while zooming in can result in missing the caries of the image. Thus, we selected a zoom-in range of 10% and a zoom-out range of 50%. The same rule of preventing the caries miss for images with positive label applies to the width/shift range. Since caries mostly happen in the edges of a tooth, shifting must be set to a small value, which is set to 20% in this case. Worth mentioning that applying other augmentation methods, such as adding noise or shearing the image, whether resulted in worsening the accuracy or a negligible change in the model performance. Hence, these methods are excluded from the augmentation procedure.

Since CapsNet is very sensitive to learning rate, the optimal learning rate is calculated using the approach introduced in [45]. Fig. 7 illustrates the loss versus learning rate during 10 epochs of training while the learning is changing exponentially.

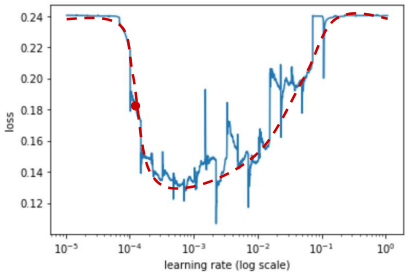


Fig. 7: loss vs learning rate and the optimal loss

According to Fig. 7 the learning rate is adjusted to 10^{-4}. The model is trained for 850 epochs with a batch size of 32. After the training process, performance of the network is evaluated. Table V presents the results of the PaXNet over a dataset of 5948 extracted tooth images.

TABLE V: Statistical performance of PaXNet

Dataset	Accuracy	Loss	Precision	Recall	F0.5-score
Training	91.23%	0.13	-	-	-
Test	86.05%	0.15	89.41%	50.67%	0.78

While the accuracy score is high, considering the relatively smaller number of positive samples, the effect of outnumbering negative images must be decreased. Hence, f0.5-score is reported as the model result. Since carious teeth misclassified as

healthy are more important than the false-positive cases, we should put more attention on minimizing false-positive ones. Thus, to increase the weight on precision and decrease the importance of recall, we selected f0.5-score as the best metric to measure the model performance.

To have a further look at how PaXNet is diagnosing caries based on the teeth location inside the jaw, a location-based accuracy map is drawn according to the accuracy of the model in the correct classification of each tooth category. To define a criterion, teeth are categorized into two classes that appear both in mandible and maxilla: molars-premolars and canines-incisors. As such, jaw is divided into 6 regions. Fig. 8 shows the above-mentioned map.

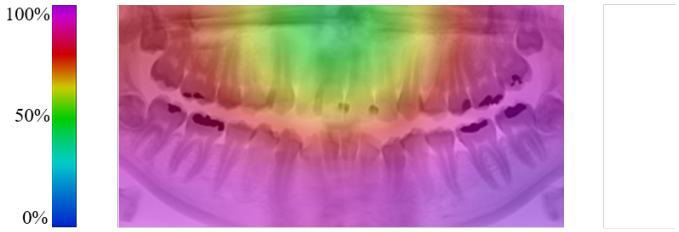


Fig. 8: Location-based accuracy map of PaXNet. The average accuracy of the model in different parts of the jaw is plotted. Purple is interpreted as having the highest score, while blue is considered the lowest

Maxillary molars-premolars achieve the highest detection rate, while mandibular molars-premolars have a lower rate. Canines-incisors in both jaws have lower accuracy scores, resulting from a lack of sufficient carious teeth in the dataset. Worth mentioning that caries occurs in molars-premolars more than canines-incisors because of certain factors such as the salivary flow.

B. Discussions

Different stages for caries in the infected tooth can be considered. At first, the infected part is small. Then, the caries lesion grows, and a larger area is affected. The detection of severe cases has higher priority since they more likely need immediate treatment. To evaluate PaXNet performance in different tooth infection levels, samples with caries in the test set are divided into two categories, mild and severe. Then, the accuracy is computed for each group, as shown in Table VI.

TABLE VI: The accuracy of PaXNet for different infection stages

Category	**Recall**
Mild	69.44%
Severe	90.52%
Total	86.05%

Severe decays usually appear as a larger demineralized area in tooth penetrating through enamel and dentine. In severe cases, it results in a total tooth collapse by destroying the pulpitis. Hence, As expected, the accuracy of the proposed model is notably higher in severe cases.

The *"right decision with wrong reason"* phenomenon can make the accuracy metrics distracting, especially when the dataset is relatively small. The computed accuracy can be a result of overfitting on this small dataset. Hence, the model might not perform this good on other samples. To make sure that the evaluated accuracy is a good reflection of the model's performance in facing new samples, the features contributing to the network's decision should be investigated. One of the most popular and effective approaches for feature visualization in CNN is Gradient-weighted Class Activation Mapping (Grad-CAM) [46]. This method computed the heatmap regarding the location of the features most contributing to the final output using the gradient. The Grad-CAM of the last convolutional layer before the capsule network is plotted. Moreover, since these high-level features are the combination of the extracted features from four different models, the Grad-CAM of the convolutional layer before the concatenation is visualized as well. Fig. 9 exhibits Grad-CAMs of five samples with caries.

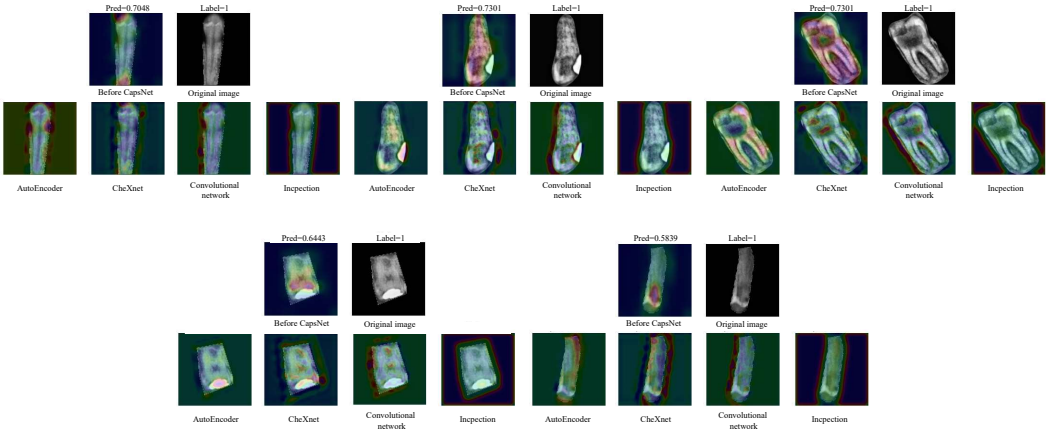


Fig. 9: Five examples of Grad-CAM after each feature extraction network and before the capsule classifier

The network is classifying these samples based on the true infected area in the tooth. Moreover, the larger the infected area is, the more confident the network becomes in caries detection. As far as feature extractors are concerned, each network is sensitive to a different type of feature. Table VII presents the test accuracy of PaXNet with different feature extraction units.

TABLE VII: The accuracy of PaXNet with different feature extractors

Feature Extractor	**Test Accuracy**
CNN	80.13%
CNN-InceptionNet	82.32%
CNN-InceptionNet-CheXNet	84.67%
CNN-InceptionNet-CheXNet-Autoencoder	86.05%

By combining these features, PaXNet detects caries based on the true infected area in the tooth. As a result, this model is capable of detecting caries correctly, and the reported accuracy is not the result of overfitting. Most importantly, in the face of new samples, a similar robust behaviour from the network is expected.

The pose of a single tooth in x-ray images is typically vertical. However, there are some unusually posed teeth in some jaws. These problematic teeth are at higher risk of infection despite the smaller number of them in the dataset. In order to address this issue, data augmentation is performed in the training process. Fig. 10 depicts the Grad-CAM of a sample with caries in various transformation.

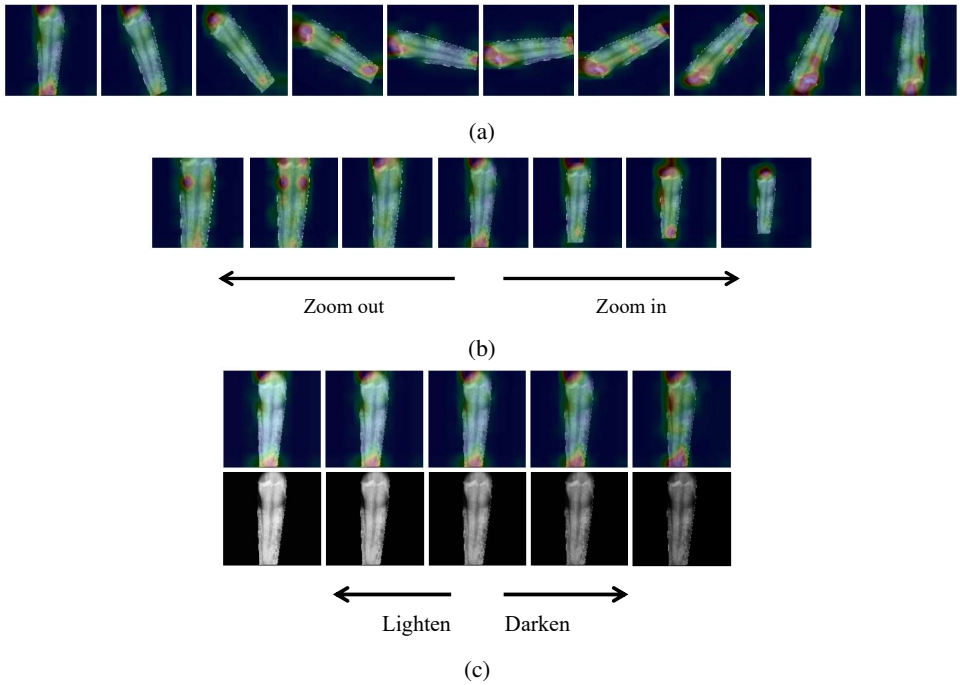


Fig. 10: The Grad-CAM visualization of an infected sample after transformation with (a) rotation, (b) zoom, and (c) brightness

The image is rotated from 0 to 180 degrees in 9 steps. Then, the image is scaled from 40% to 160%. Finally, the mid-level brightness is altered $\pm 30\%$ using a quadratic function. Fig. 10 illustrates that by applying various transformations, the proposed network is still classifying the sample accurately using correctly distinguished image features.

V. Conclusion

Dental decision support systems can help dentists by providing high-throughput diagnostic assistance to better detect dental diseases, such as caries, using radiographs. While capturing Panoramic dental radiography could be very helpful to see full patient dentition in a single image, detection of dental caries using dental Panoramic images is a very challenging task due to the low quality of the image and the ambiguity of decay regions. To help radiologists with dental caries diagnosis, we propose an automatic caries detection model using several feature extraction architectures and capsule network to predict labels from the extracted features. Grad-CAM visualization is used to validate our extracted features in terms of correct localization. Developed model could successfully hit an accuracy score of 86.05% and an f0.5-score of 0.78 on the test set. The achieved scores demonstrate model efficiency in concentrating on not missing carious teeth. Experimental results illustrate that the proposed combination of feature extraction modules achieves higher performance compared to a naïve CNN approach. Besides, PaXNet is also capable of categorizing caries lesions into two groups of mild ones with a recall of 69.44% and severe ones with a recall score equal to 90.52%. Considering PaXNet as the first deep learning-based model to detect dental caries using Panoramic x-rays, achieved results are promising and acceptable, although not outstanding in comparison with previous works on image types with higher quality. To improve the results, two attributes are to be addressed: (1) increasing the number of Panoramic images, and specifically, expanding the number of carious teeth in the dataset. And (2) benefiting from a more sophisticated neural network, such as EfficientNet-based architectures, to boost the model's capability. Radiologist-provided precise annotations for caries regions will lead to accurate segmentation of caries lesions using U-Net-based segmentation models.

REFERENCES

[1] R. H. Selwitz, A. I. Ismail, and N. B. Pitts, "Dental caries," *The Lancet*, vol. 369, no. 9555, pp. 51–59, 2007.

[2] P. Amrollahi, B. Shah, A. Seifi, and L. Tayebi, "Recent advancements in regenerative dentistry: A review," *Materials Science and Engineering: C*, vol. 69, pp. 1383–1390, 2016.

[3] E. D. Beltrán-Aguilar, L. K. Barker, M. T. Canto, B. A. Dye, B. F. Gooch, S. O. Griffin, J. Hyman, F. Jaramillo, A. Kingman, R. Nowjack-Raymer *et al.*, "Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis; united states, 1988-1994 and 1999-2002," 2005.

[4] N. Pitts and D. Zero, "White paper on dental caries prevention and management," *FDI World Dental Federation*, 2016.

[5] O. Fejerskov and E. Kidd, *Dental caries: the disease and its clinical management*. John Wiley & Sons, 2009.

[6] P. H. Lira, G. A. Giraldi, and L. A. Neves, "Panoramic dental x-ray image segmentation and feature extraction," in *Proceedings of V workshop of computing vision, Sao Paulo, Brazil*, 2009.

[7] E. Tagliaferro, A. V. Junior, F. L. Rosell, S. Silva, J. L. Riley, G. H. Gilbert, and V. V. Gordan, "Caries diagnosis in dental practices: Results from dentists in a brazilian community," *Operative dentistry*, vol. 44, no. 1, pp. E23–E31, 2019.

[8] X. Qu, G. Li, Z. Zhang, and X. Ma, "Detection accuracy of in vitro approximal caries by cone beam computed tomography images," *European journal of radiology*, vol. 79, no. 2, pp. e24–e27, 2011.

[9] Z. Akarslan, M. Akdevelioglu, K. Gungor, and H. Erten, "A comparison of the diagnostic accuracy of bitewing, periapical, unfiltered and filtered digital panoramic images for approximal caries detection in posterior teeth," *Dentomaxillofacial Radiology*, vol. 37, no. 8, pp. 458–463, 2008.

[10] P. S. Casamassimo, "Radiographic considerations for special patients-modifications, adjuncts, and alternatives," *Ped Dent*, vol. 3, no. 2, pp. 448–54, 1981.

[11] T. E. Underhill, I. Chilvarquer, K. Kimura, R. P. Langlais, W. D. McDavid, J. W. Preece, and G. Barnwell, "Radiobiologic risk estimation from dental radiology: Part i. absorbed doses to critical organs," *Oral Surgery, Oral Medicine, Oral Pathology*, vol. 66, no. 1, pp. 111–120, 1988.

[12] N. Akkaya, O. Kansu, H. Kansu, L. Cagirankaya, and U. Arslan, "Comparing the accuracy of panoramic and intraoral radiography in the diagnosis of proximal caries," *Dentomaxillofacial Radiology*, vol. 35, no. 3, pp. 170–174, 2006.

[13] J. Naam, J. Harlan, S. Madenda, and E. P. Wibowo, "The algorithm of image edge detection on panoramic dental x-ray using multiple morphological gradient (mmg) method," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 1012–1018, 2016.

[14] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018, pp. 400–407.

[15] J.-H. Lee, S.-S. Han, Y. H. Kim, C. Lee, and I. Kim, "Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 129, no. 6, pp. 635–642, 2020.

[16] A. Haghanifar, M. M. Majdabadi, and S.-B. Ko, "Automated teeth extraction from dental panoramic x-ray images using genetic algorithm," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2020, pp. 1–5.

[17] A. Haghanifar, A. Amirkhani, and M. R. Mosavi, "Dental caries degree detection based on fuzzy cognitive maps and genetic algorithm," in *Electrical Engineering (ICEE), Iranian Conference on*. IEEE, 2018, pp. 976–981.

[18] D. Fried, "Detecting dental decay with infrared light," *Optics and Photonics News*, vol. 31, no. 5, pp. 48–53, 2020.

[19] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci, and H. Markram, "Caries detection with near-infrared transillumination using deep learning," *Journal of dental research*, vol. 98, no. 11, pp. 1227–1233, 2019.

[20] M. M. Srivastava, P. Kumar, L. Pradhan, and S. Varadarajan, "Detection of tooth caries in bitewing radiographs using deep learning," *arXiv preprint arXiv:1711.07312*, 2017.

[21] J. Choi, H. Eun, and C. Kim, "Boosting proximal dental caries detection via combination of variational methods and convolutional neural network," *Journal of Signal Processing Systems*, vol. 90, no. 1, pp. 87–97, 2016.

[22] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of dentistry*, vol. 77, pp. 106–111, 2018.

[23] H. A. Khan, M. A. Haider, H. A. Ansari, H. Ishaq, A. Kiyani, K. Sohail, M. Muhammad, and S. A. Khurram, "Automated feature detection in dental periapical radiographs by using deep learning," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2020.

[24] A. Laishram and K. Thongam, "Detection and classification of dental pathologies using faster-rcnn in orthopantomogram radiography image," in *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2020, pp. 423–428.

[25] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[26] G. Silva, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives," *Expert Systems with Applications*, vol. 107, pp. 15–31, 2018.

[27] F. Martínez-Rus, A. M. García, A. H. de Aza, and G. Pradíes, "Radiopacity of zirconia-based all-ceramic crown systems." *International Journal of Prosthodontics*, vol. 24, no. 2, 2011.

[28] D. E. Goldberg, *Genetic algorithms*. Pearson Education India, 2006.

[29] A. Sheta, M. S. Braik, and S. Aljahdali, "Genetic algorithms: a tool for image segmentation," in *2012 international conference on multimedia computing and systems*. IEEE, 2012, pp. 84–90.

[30] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3856–3866. [Online]. Available: http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules.pdf

[31] M. M. Majdabadi and S.-B. Ko, "Msg-capsgan: Multi-scale gradient capsule gan for face super resolution," in *2020 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 2020, pp. 1–3.

[32] A. Pal, A. Chaturvedi, U. Garain, A. Chandra, R. Chatterjee, and S. Senapati, "Capsdemm: capsule network for detection of munro's microabscess in skin biopsy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 389–397.

[33] B. Tang, A. Li, B. Li, and M. Wang, "Capsurv: capsule network for survival analysis with whole slide pathological images," *IEEE Access*, vol. 7, pp. 26 022–26 030, 2019.

[34] T. Iesmantas and R. Alzbutas, "Convolutional capsule network for classification of breast cancer histology images," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 853–860.

[35] M. M. Majdabadi and S.-B. Ko, "Capsule gan for robust face super resolution," *Multimedia Tools and Applications*, pp. 1–14, 2020.

[36] T. Zhao, Y. Liu, G. Huo, and X. Zhu, "A deep learning iris recognition method based on capsule network architecture," *IEEE Access*, vol. 7, pp. 49 691–49 701, 2019.

[37] N. Frosst, S. Sabour, and G. Hinton, "Darccc: Detecting adversaries by reconstruction from class conditional capsules," *arXiv preprint arXiv:1811.06969*, 2018.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[39] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[40] M. Abdel-Mottaleb, O. Nomir, D. E. Nassar, G. Fahmy, and H. H. Ammar, "Challenges of developing an automated dental identification system," in *2003 46th Midwest Symposium on Circuits and Systems*, vol. 1. IEEE, 2003, pp. 411–414.

[41] J.-V. Ølberg and M. Goodwin, "Automated dental identification with lowest cost path-based teeth and jaw separation," *Scandinavian Journal of Forensic Science*, vol. 22, no. 2, pp. 44–56, 2016.
[42] O. Nomir and M. Abdel-Mottaleb, "A system for human identification from x-ray dental radiographs," *Pattern Recognition*, vol. 38, no. 8, pp. 1295–1305, 2005.
[43] N. Al-Sherif, G. Guo, and H. H. Ammar, "A new approach to teeth segmentation," in *2012 IEEE International Symposium on Multimedia*. IEEE, 2012, pp. 145–148.
[44] A. E. Rad, M. S. M. Rahim, H. Kolivand, and A. Norouzi, "Automatic computer-aided caries detection from dental x-ray images using intelligent level set," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28 843–28 862, 2018.
[45] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 464–472.
[46] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.