Automated Teeth Extraction and Dental Caries Detection in Panoramic X-ray

A thesis submitted to the
College of Graduate and Postdoctoral Studies (CGPS)
in partial fulfillment of the requirements for the degree of
Master of Science in the Division of Biomedical Engineering
University of Saskatchewan
Saskatoon, Canada

By
Arman Haghanifar

Permission to Use

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Post-graduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Disclaimer

Reference in this thesis to any specific commercial products,process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes

Request for permission to copy or to make any other use of material in this thesis in whole or in part should be addressed to:

Head of the Division of Biomedical Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A9

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Dental caries is one of the most chronic diseases that involves the majority of people at least once during their lifetime. This expensive disease accounts for $5 - 10\%$ of the health-care budget in developing countries. Caries lesions appear as the result of dental biofilm metabolic activity, caused by bacteria (most prominently Streptococcus mutans) feeding on uncleaned sugars and starches in oral cavity. Also known as tooth decay, they are primarily diagnosed by general dentists solely based on clinical assessments. Since in many cases dental problems cannot be detected with simple observations, dental x-ray imaging is introduced as a standard tool for domain experts, i.e. dentists and radiologists, to distinguish dental diseases, such as proximal caries. Among different dental radiography methods, Panoramic or Orthopantomogram (OPG) images are commonly performed as the initial step toward assessment. OPG images are captured with a small dose of radiation and can depict the entire patient dentition in a single image. Dental caries can sometimes be hard to identify by general dentists relying only on their visual inspection using dental radiography. Tooth decays can easily be misinterpreted as shadows due to various reasons, such as low image quality. Besides, OPG images have poor quality and structures are not presented with strong edges due to low contrast, uneven exposure, etc. Thus, disease detection is a very challenging task using Panoramic radiography. With the recent development of Artificial Intelligence (AI) in dentistry, and with the introduction of Convolutional Neural Network (CNN) for image classification, developing medical decision support systems is becoming a topic of interest in both academia and industry. Providing more accurate decision support systems using CNNs to assist dentists can enhance their diagnosis performance, resulting in providing improved dental care assistance for patients.

In the following thesis, the first automated teeth extraction system for Panoramic images, using evolutionary algorithms, is proposed. In contrast to other intraoral radiography methods, Panoramic is captured with x-ray film outside the patient mouth. Therefore, Panoramic x-rays contain regions outside of the jaw, which make teeth segmentation extremely difficult. Considering that we solely need an image of each tooth separately to build a caries detection model, segmentation of teeth from the OPG image is essential. Due to the ab-

sence of significant pixel intensity difference between different regions in OPG radiography, teeth segmentation becomes very hard to implement. Consequently, an automated system is introduced to get an OPG as input and gives images of single teeth as the output. Since only a few research studies are utilizing similar task for Panoramic radiography, there is room for improvement. A genetic algorithm is applied along with different image processing methods to perform teeth extraction by jaw extraction, jaw separation, and teeth-gap valley detection, respectively. The proposed system is compared to the state-of-the-art in teeth extraction on other image types.

After teeth are segmented from each image, a model based on various untrained and pretrained CNN-based architectures is proposed to detect dental caries for each tooth. Autoencoder-based model along with famous CNN architectures are used for feature extraction, followed by capsule networks to perform classification. The dataset of Panoramic x-rays is prepared by the authors, with help from an expert radiologist to provide labels. The proposed model has demonstrated an acceptable detection rate of 86.05%, and an increase in caries detection speed. Considering the challenges of performing such task on low quality OPG images, this work is a step towards developing a fully automated efficient caries detection model to assist domain experts.

Acknowledgments

I would like to express my gratitude to my supervisor, professor Seok-Bum Ko, for his support and guidance throughout my M.Sc. program. It would have been impossible to write this dissertation without his encouragement. I also would like to thank my lab-mates, specifically Mahdiyar Molahasani Majdabadi, who helped during all stages, from ideation and brainstorming to publishing papers. Moreover, I would like to thank my B.Sc. supervisor Dr. Abdollah Amirkhani for his support, my family for paving the road for me to continue my studies, and my friends for their encouragement. My sincere gratitude to the College of Engineering, Div. of Biomedical, at the University of Saskatchewan.

Table of Contents

List of Abbreviations

AI	Artificial Intelligence
CAD	Computer-Aided Diagnosis
CapsNet	Capsule Network
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DEJ	Dentine-Enamel Junction
GA	Genetic Algorithm
Grad-CAM	Gradient-weighted Class Activation Map
LOF	List of Figures
LOT	List of Tables
MLP	Multi-Layer Perceptron
OPG	Orthopantomogram
ReLU	Rectified Linear Unit
ReLU	Rectified linear unit
RNN	Recurrent Neural Networks
ROI	Region of Interest
VAE	Variational Auto-Encoder

List of Tables

List of Figures

1. Introduction

Dental Caries is a chronic dental disease that can be efficiently diagnosed using dental radiography. Artificial Intelligence (AI)-based Computer-Aided Diagnosis (CAD) systems have recently shown to be helpful in the detection of dental diseases. In the following thesis, a novel approach is proposed to detect caries from dental Panoramic radiography. Research problem, including definition of dental caries, the importance of detecting it and the challenges of detection using radiography, is presented in this chapter. After discussing the problem, motivations for conducting the following research work are mentioned, followed by a summary of contributions of this thesis and the publications and submissions during M.Sc. studies.

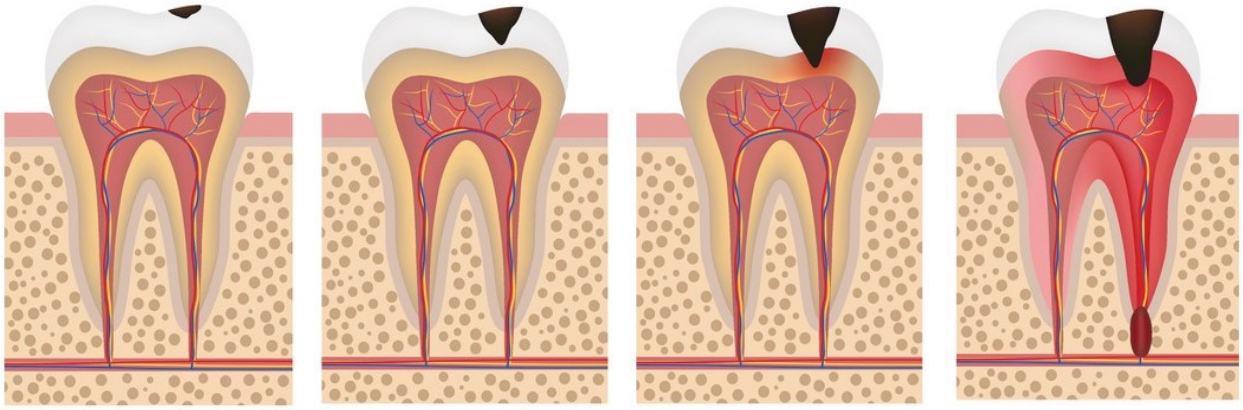
1.1 Research problem and objectives

Dental caries is the most epidemic non-contagious disease worldwide. It affects all age groups and segments of the population [5]. Dental caries, which often leads to tooth decay, and causes the tooth to collapse, is also considered as an infectious disease that is quite common in a way that each person will suffer from it at least once [6]. While caries has become more prevalent among human being since the increase of fermentable sugars in daily diet in recent decades, it has a long story. Early hominids had suffered from cavities over a million years ago. Skulls dating back from the Paleolithic to Neolithic eras indicate signs of tooth decay and alveolar bone loss present in the jaws [7]. The increase of cavitation during the Neolithic era might be attributed to the increase of plant foods in human dietary, containing carbohydrates [8].

Dental caries results from interaction between three main factors, including tooth, mi-

crobes on its surface, and the diet. Bacteria gradually accumulate in a specific spot on the teeth to form a bacterial plaque. Fermentable carbohydrates, obtained through a diet pattern including sugars, help the bacterial plaque with creating lactic acid from fermentation. The produced acid results in dissolution of dental hard tissues. Dissolution gradually becomes intense forming dental cavity, where bacteria penetrate into the hard tissues to result in creating a lesion of caries. In fact, caries represents a previous or ongoing microbial activity in the biofilm [9].

A dental biofilm first appears on the enamel of a tooth. It slowly penetrates the tissue with the presence of fermentable carbohydrates. Through continuous dissolution, the lesion reaches the inner parts of a tooth, including dentin (yellowish tissue around the roots and pulp) and pulp (central part of the tooth). With caries penetrating to the internal parts, severe pain will be suffered by the patient. When caries involves dental pulp, the tooth is collapsed and needs to be extracted. Dental caries stages are depicted by Fig. 1.1.



(a) (b) (c) (d)

Figure 1.1: Different stages of dental caries. Decay grows stage by stage from left to right: a) decay in enamel, b) advanced enamel caries (known as DEJ decay), c) decay in dentin, and d) decay in dental pulp. Third and last stages are painful, requiring extensive treatment plans, such as tooth extraction.

As seen above, dental caries in the first and second stages is not associated with extreme pain, which is called a caries in mild stages. If caries progresses to the severe stages, it is likely that the tooth must be removed from the mouth.

Dental radiography is a standard tool used in dentistry to diagnose several dental problems, such as caries, infections and bone abnormalities. It can be used to detect dental caries in initial, moderate, and extensive stages as shown in Fig. 1.1 [10]. Considered as an important supplementary diagnosis solution, x-ray imaging helps dentists to identify problems that are hard or impossible to detect using visual inspection during clinical assessments [11]. Among dental x-ray image types, Panoramic or Orthopantomogram (OPG) images are widely utilized. Panoramic images are taken using a very small dose of ionizing radiation to capture the entire mouth in one image [12]. They have simplicity of application, less time requirement, and also great patient comfort. Thus, pediatric, people with disabilities, and senior patients would benefit greatly from a Panoramic imaging system compared to intraoral systems [13]. This imaging method is mainly used for automating the diagnosis of dental diseases, such as dental caries diagnosis [14] [15], human identification [16] [17], and lesion detection [18].

However, there are several challenges in using this method for automated disease detection. One important obstacle is the need to separate each tooth, which is called teeth extraction, before applying any decision-making algorithm to the data. While separation and segmentation are often both the most difficult and the most important tasks of a machine vision system [19], implementing them on panoramic images is significantly harder than other types of x-ray images, like Bitewing or Periapical [20]. The reason behind this problem is that Panoramic is an extra-oral imaging method, where both the image detector and the x-ray machine are placed outside the patient's mouth [21]. Thus, the captured image contains other information, like jaw-skull joints, jawbones, etc., which make it difficult to be analyzed [22].

Another challenge is the lack of rich information in Panoramic radiography. Structures in OPG images do not have strong boundaries by which they can be segmented [23]. Since the quality of the information in Panoramic images is extremely poor, image preprocessing steps are needed to elevate visual interpretations [20]. Consequently, image analysis is significantly harder on Panoramic images in comparison with Bitewing or Periapical, which have drawn more research attention up until now.

So far, a limited number of research studies have been conducted to develop a Deep Learning (DL)-based decision support system for dental diseases. The main reason, which is also a major challenge in using Artificial Intelligence (AI) for medical applications, is the lack of open datasets with sufficient and confidently-labeled images. There are currently a few public datasets with images of dental radiography, i.e. A dataset of 120 Periapical dental x-rays with ground truth provided by Rad *et al.*[1] for caries screening, and a dataset of 120 Bitewing dental x-rays provided in the ISBI 2015 Challenge[2]. To date, there are currently no public datasets available for Panoramic dental radiography or other types of radiography. To alleviate the shortage of labeled data, more datasets are needed to be publicly available, specifically from unavailable types of x-ray.

1.2 Motivations

As mentioned previously, dental caries is the most common non-communicable disease in the world. It is also considered to be an expensive disease to treat due to the fact that it is responsible for $5 - 10\%$ of healthcare budget in developing countries [24]. On the other hand, efficient and early detection of dental caries is known to be the key to having an effective therapeutic method as well as a key metric of preventive approach in dentistry [25]. Therefore, increasing the accuracy of caries detection, and specifically when caries is in early stages, lead to providing better treatment decisions that can help the patient both in terms of healthiness and financially.

There have been some research studies to utilize computational models to provide automated diagnosis and prognosis tools for dental diseases. However, the low quality of information in dental radiography has made it a very challenging task to extract patterns with traditional techniques. With the advancement of DL and evolutionary algorithms and their usability in biomedical domain, they can be used to efficiently process dental radiography and provide diagnosis results for supporting, and in the future, substituting the domain experts. Promising results of recent DL architectures and the possibility of transfer learning

[1] `http://dx.doi.org/10.6070/H47H1GJ4`

[2] `http://www-o.ntust.edu.tw/~cweiwang/ISBI2015/challenge2/`

of pretrained DL models motivated us to employ these techniques alongside image processing methods and evolutionary algorithms to develop an end-to-end detection model to propose a tooth segmentation and dental caries detection system for the challenging task of caries detection in Panoramic radiography.

In this thesis, we have proposed a novel genetic-based approach for tooth segmentation in Panoramic dental images. Based on a dataset of Panoramic x-rays, our tooth segmenter could isolate teeth with accuracy scores in line with previous academic works utilizing either Bitewing or Periapical radiography. Various methods were introduced for jaw separation, tooth isolation, and accuracy improvement of the system. Considering the aforementioned tooth segmentation system as the preliminary step to prepare single tooth images for the disease classification model, next chapters of the thesis aim 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.

1.3 Contributions of the Thesis

Since there have been limited attempts to develop automatic DL-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. Contributions of this thesis are summarized as follows:

- This research is the first to perform tooth segmentation 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.
- To compensate the problem of data shortage in this field, we have introduced our dataset of Panoramic x-rays, consisting of 470 images. Out of the dataset, 120 x-rays are collected from a local dentistry clinic; anonymized and preprocessed under the supervision of an expert radiologist.
- 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 Convolutional Neural Networks (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 [26] is applied for the first time in dental disease detection.

1.4 Publications and Submissions During M.Sc. Study

1.4.1 Published Conference

1. Molahasani Majdabadi, Mahdiyar, Haghanifar, Arman, and Seok-Bum Ko. "Msg-capsgan: Multi-scale Gradient Capsule GAN for Face Super Resolution." In 2020 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1-3. IEEE, 2020. DOI: 10.1109/ICEIC49074.2020.9051244

   A minor portion of this paper is included in Chapter 4

2. Haghanifar, Arman, Mahdiyar Molahasani Majdabadi, and Seok-Bum Ko. "Automated Teeth Extraction from Dental Panoramic X-Ray Images using Genetic Algorithm." In 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1-5. IEEE, 2020. DOI: 10.1109/ISCAS45731.2020.9180937

   A major portion of this paper is included in chapters 3

1.4.2 Preprints

1. Haghanifar, Arman, Mahdiyar Molahasani Majdabadi, and Seokbum Ko. "Covid-cxnet: Detecting Covid-19 in Frontal Chest X-ray Images using Deep Learning." arXiv preprint arXiv:2006.13807 (2020).

2. Haghanifar, Arman, Mahdiyar Molahasani Majdabadi, and Seok-Bum Ko. "PaXNet: Dental Caries Detection in Panoramic X-ray using Ensemble Transfer Learning and Capsule Classifier." arXiv preprint arXiv:2012.13666 (2020).

   A major portion of this paper is included in Chapter 4

1.5 Organization of the Thesis

The thesis is organized as follows:

- **Chapter 1: Introduction** gave a clear view of the problem and encountered challenges. In this chapter, we also explained the motivations behind this research, followed by a description of teeth extraction and dental caries with deep learning. Then, the contributions of this thesis are presented, and finally, publications and submissions during the M.Sc. program are listed.

- **Chapter 2: Literature Review** provides a review of previous works regarding teeth extraction/segmentation and dental caries classification using image processing techniques as well as deep learning. The latest advancements in the application of deep learning and transfer learning in dentistry are also addressed.

- **Chapter 3: Teeth Extraction using Genetic Algorithm and Image Processing** proposes a framework based on different methods, including a genetic algorithm to convert an OPG radiography into images, each including a single tooth. Then, different approaches are employed to enhance model performance on mandibular teeth. The performance of the final model is evaluated and compared with the state-of-the-art works.

- **Chapter 4: Ensemble Transfer Learning for Dental Caries Detection** explains the steps for expanding the model proposed in Chapter 3 to include dental caries detection. The architecture of the proposed model, named PaXNet, is reviewed. Both this and previous chapters also come as sections dedicated to explaining the dataset and the preprocessing pipeline, followed by the experimental results and comparison with related works.

- **Chapter 5: Conclusions and Future work** summarizes this thesis, remaining challenges, and discusses potential future works.

2. Literature Review

In this chapter, a review of prominent machine learning-based and image processing-based research works in dentistry is presented at first. Then, previous studies in both teeth extraction and dental caries classification are discussed and performance metrics of the most recent proposed models are compared. Next, we probe into the newest works in the same field utilizing deep learning and transfer learning as their primary tool.

2.1 Artificial Intelligence in Dentistry

For decades, computational models have been used in the field of medical and dental image processing for producing highly accurate results. The primary motivation of developing these models is to develop decision support systems to help domain experts produce more accurate diagnosis decisions. Since the introduction of neural networks and traditional machine learning algorithms, artificial intelligence has played a significant role in publishing high-quality academic works involving computational models in medical and dental fields. The growing role of computational models, and artificial intelligence at the heart of it, which has happened since the 1980s, stems from a couple of reasons.

First, diagnostic information in medicine is mainly derived from medical imaging from different sources, such as computed tomography or radiography. With the introduction of CNNs for image processing, artificial intelligence models could also be developed to handle the images. Since differential diagnosis of medical images could even be impossible for human specialists, artificial intelligence-based image processing systems have proven to perform on par or even superior to the human-level performance benchmark.

Secondly, after the advancements in computer hardware design and with the introduction

of more efficient computers equipped with faster processing power, deeper neural networks could be run on usual systems, and much larger datasets could also be handled. As mentioned previously, most medicine-related tasks do require images, and medical images are relatively larger than typical structured data. Thus, the ability to handle large imaging sets together with extracting more complicated features out of them (which directly correlates with more layers in the feature extraction part of a neural network) participated in a breakthrough in how computational models are used in medical image processing, as well as generally every domain. To be more specific, traditional rule-based expert systems, which were based on a rigid framework of rules and had hardly achieved optimal discriminative performance, were outdated to bring data-driven models into practice. Data-driven models, which emphasize on data-centric model design in comparison with the rule-centric design of expert systems, are capable of achieving superhuman performance and are generally known as machine learning [1].

Machine learning is a term used to represent a set of more recent statistical techniques and computational methods which output a specific model based on their input, which is supposed to be a set of data records with labels. The constructed model usually is trained on a large dataset and optimized through feedback to classify input data into certain categories (classification) or predict continuous values (regression). The term machine learning is used for a set of numerous methods, among which the most popular one is known to be neural networks. Neural network, also known as artificial neural network, is a network of neurons shaped as layers placed one after another. Similar to other machine learning algorithms, neural net has connected layers, with weights and biases as parameters, which are tuned by incremental trial and error to explore the parameter space to minimize a predefined loss function. Loss functions are usually derived from a formula containing the difference between predicted values and actual values as the main term. To have a more clarified view, neural networks often have activation functions and operate similar to regressions, where intercept is referred as bias, and coefficient is named as weights. Each layer has its own set of weights and biases. After passing through all the layers, the final predicted value is compared to the expected value, and the difference is used to update the parameters by back-propagation.

This procedure happens iteratively for many epochs so that finally, the model converges to a steady-state. An example of a neural network is shown in Fig. 2.1.

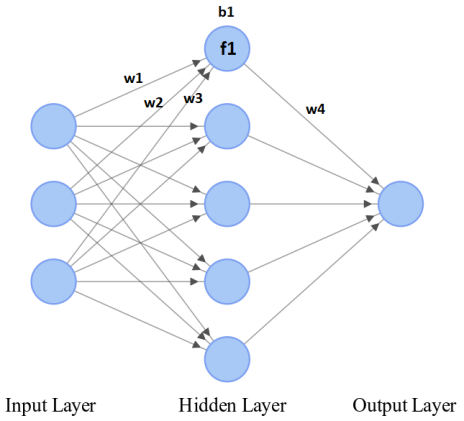


Figure 2.1: A typical three-layered neural network.

As seen above, input features are fed through the input layer, and the output layer generates a prediction value. Hidden layers increase complexity through which a model can handle more complicated patterns. $w1$, $w2$, and $w3$ are weights which are combined with the bias $b1$ and then an activation function $f1$ is applied on them to generate the weight $w4$ for the next layer.

As problems become more complicated and datasets are becoming larger, to handle the complexity, we need to add more hidden layers to a neural network. A multi-layer neural network is often called a deep neural network, which is the main technique used in deep learning. Deep learning is considered as an evolution to traditional machine learning that aims to automate data preprocessing and create an end-to-end learning model. Deep neural networks partially resemble the structure of neurons in the human brain. Deep learning became widely utilized with the introduction of convolutional layers as feature extractors. Convolutional layers, which together with neural networks form CNNs, are designed to sweep over images and extract the information layer by layer. Convolutional blocks in CNNs mainly consist of convolutional layers, pooling layers, and normalization layers. Convolutional blocks turn the input image into a rich set of numerical values representing important features in them. Assembling more convolutional blocks as a hierarchical structure will lead to extraction of more complex patterns, with the size of the dataset as the limitation for overfitting phenomenon.

An example of a CNN is depicted in Fig. 2.2.

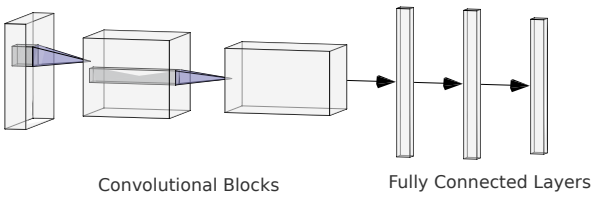


Figure 2.2: A CNN with three convolutional blocks.

Fig. 2.2 illustrates how kernels sweep over inputs to extract higher-level features and to finally pass the most high-level elements to classification block, consisting of a multilayer perceptron neural network. Final nodes of the classification block are decision-makers, each responsible for a class. Classes usually stand for generalized concepts, such as common objects, or, in case of dental imaging, a radiograph that positively shows a certain pathological condition.

In deep learning, there are many task-oriented CNN architectures developed with higher efficiency on specific tasks. Main applications of CNN are object detection and segmentation. Image segmentation is a further variation of object detection that uses a similar approach to extract the target object from its background. Image segmentation includes semantic and instance segmentation. In semantic segmentation, multiple objects belonging to the same class are considered as one target, whereas in instance segmentation, each object is assigned with a specific label. As an example, in dental x-rays, semantic segmentation is to segment the teeth from the background, while instance segmentation is to segment each tooth with a different label. The difference is more intuitively represented in Fig. 2.3.

In semantic segmentation, all extracted tooth regions are labeled the same, while in instance segmentation, each tooth is separated. Popular image segmentation CNN-based architectures include fully convolutional networks (FCNs) [27] and U-Net [28]. These networks usually consist of a down-sampling block, another name for a typical set of convolutional blocks, and an up-sampling block to reconstruct the image from the feature vector.

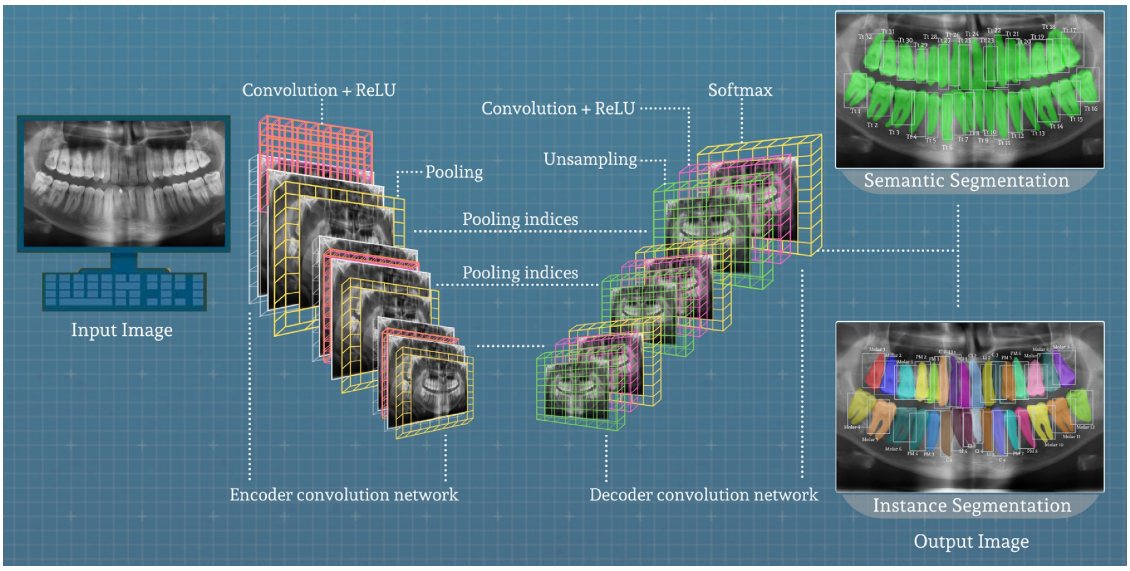


Figure 2.3: Image semantic and instance segmentation on dental radiography using fully convolutional neural network architecture [1].

Most DL-based studies in the field of dentistry have applied CNN-based architectures to analyze radiographic images, from x-rays to Computer Tomography (CT) images. The main goal of developing such models to build a decision support system that helps the dentist or radiologist detect abnormalities, rather than substituting them with the artificial model. Thus, computational models are mostly constructed to operate as a diagnostic aiding tool. In the field of dentistry, photographic and radiographic imaging is widely used as the first step in clinical assessment and diagnosis process. Hence, research studies so far have tried to develop CNNs to recognize, classify, and segment certain anatomic structures or pathological issues from dental images.

In general, DL-based models have been used for many applications in dentistry. CNNs have proven to be able to segment and classify teeth in both 2D radiographs, such as Panoramic [29] or Periapical [30], and 3D images, such as CBCT [31]. As an instance of a specific application, CNN has been used to detect the anatomy of first molar roots and also investigate it in terms of having any abnormalities [32], which is in the field of endodontics. Another example includes developing a CNN to identify the course of the inferior alveolar nerve and initiate its spatial relationship with the roots of third molars [33]. Understanding such relationship is essential in order to preplan the process of surgical extraction of

third molars. This is an application of CNN in dentistry improving surgical preplanning procedures.

An important application of AI in dentistry is to detect dental caries, which is a prevalent chronic dental disease in humans, affecting most individuals at least once during their lifetime [34]. In clinical practice, dental caries is diagnosed by a dentist or radiologist using visual observation and also capturing dental radiographs as a complementary diagnosis material. When the loss of enamel, dentin, and/or pulpitis occurs, damaged tooth could be indicated by analyzing alterations in its appearance in the radiography. Several studies employing CNN-based models have been developed to help clinicians detect caries from different dental images, from near-infrared transillumination [35] to Periapical radiography [36]. Research has identified Periapical x-rays as the most effective method for caries detection in experimental settings. However, Panoramic radiography is widely used as the first step for monitoring dental caries. To the best of our knowledge, there are very few research studies conducted to detect caries in Panoramic images, and none of them have utilized CNNs for such a purpose.

Further applications of CNN in dental image processing include, but are not limited to, detecting apical lesions in Panoramic images [37], detecting periodontal bone loss in Panoramic x-rays [38], detection of maxillary sinusitis in Panoramic radiography [39], and survival prediction of patients involved in oral cancer by analyzing Periapical radiographs [40]. In conclusion, while existing research studies are sparse and there are several issues hindering the wide usage of DL and CNNs as the decision support system for dental disease diagnosis, results are encouraging and academic interest is growing year by year.

2.2 Teeth Extraction/Segmentation Literature

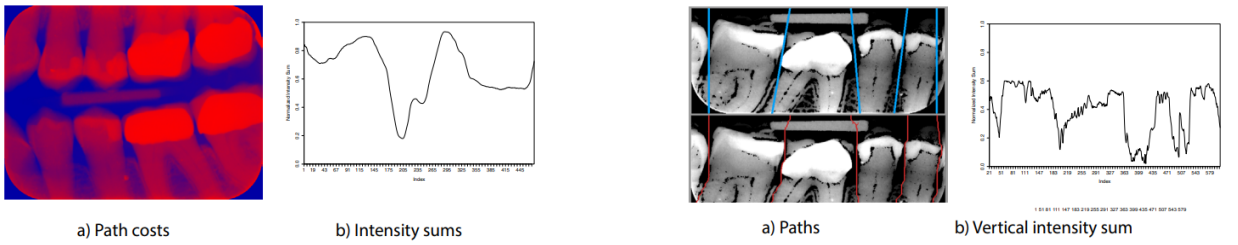
While OPG radiography has certain limitations in image quality and lower diagnostic accuracy, it has been widely used in common practice to diagnose different diseases, such as dental caries and periodontal diseases [41]. Several research studies have utilized OPG for automating different procedures involving dental radiography, e.g. dental caries diagnosis [14], human identification [42], lesion detection [18], etc. Teeth extraction is an essential part for most of the mentioned tasks, which is either mandatory to implement or is beneficial to

implement in terms of result improvement [43].

While separation and segmentation are often both the most difficult and the most important tasks of a machine vision system [19], implementing them on Panoramic dental radiography is significantly more challenging than other types of dental x-rays, like Bitewing or Periapical [20]. Image texture in OPG radiography tends not to have strong edges by which the structures can be segmented [23]. Moreover, Panoramic images contain other information, like jaw-skull joints, bones, etc., which make the image difficult to be analyzed [22]. Besides, the quality of the information in Panoramic images is extremely low, therefore, image preprocessing steps are needed to elevate visual interpretations [20]. Consequently, image analysis is significantly harder on Panoramic images in comparison with Bitewing or Periapical images, which have drawn more research attentions so far.

Teeth separation or isolation is the process of extracting images, each containing the boundaries of one tooth, from a dental x-ray image, which also contains some other unwanted parts of the mouth, like gingivae or jawbones [44]. While there are recent research studies conducted on teeth segmentation [22], [45], [46], there are few works done on teeth separation.

Olberg and Goodwin [2] introduced a lowest-cost path-based method for jaw and teeth separation as part of their system for automated human identification from Bitewing dental x-rays. Ehsani Rad et al. [15] used the integral intensity value method for tooth isolation, which was initially suggested by Jain and Chen [17]. Many other research studies have also applied the same method for tooth extraction; Abdel-Mottaleb and his colleagues [47] have applied horizontal and vertical integral projection for separating jaws and teeth, respectively. Vertical integral projection followed by spline method was utilized by [48] for teeth separation. Morphological and image cropping operations were also utilized for the same operation [49]. The idea of finding the lowest-cost path for separating teeth based on their brighter appearance in comparison with the darker gap valleys is quite the same as the integral intensity projection method and has been used as the baseline for almost every research study proposing teeth extraction methods. A simple illustration of this method is depicted in Fig. 2.4 originally by the authors of [2].



(a) (b)

Figure 2.4: Examples of path costs used to separate teeth using minima of the intensity sums graph, proposed by [2]. a) horizontal integral intensity to separate upper and lower jaws, and b) vertical integral intensity to separate teeth.

Although there have been a handful of studies performing teeth extraction, most of them have applied their algorithms on either Bitewing or Periapical images. Few studies have applied methods on Panoramic dental x-rays. Authors of [18] did the teeth isolation process on a very small dataset of Panoramic images using discrete wavelet transformation along with polynomial regression for jaw separation and angular radial scanning method for tooth segmentation, but no precise results were reported for their teeth isolation method. Furthermore, their method is not generalized, which means that parameter values may differ for each input image. Their method is relatively dependent on the number of sample points, which leads to slow software runtime.

The aforementioned reasons are also applicable to [50], where the authors have proposed an automatic segmentation method that jaws are separated with a shape-free model and teeth are segmented with projecting lines. Issues with the mentioned model include jaw segmentation using a 9-degree polynomial curve that will not work with closely-stacked-together jaws, no provided results in terms of accuracy or other evaluation metrics, etc.

2.3 Dental Caries Classification Literature

Dental caries, also known as tooth decay, is considered as one of the most prevalent infectious chronic dental diseases in human beings, which affects almost every individual during their lifetime for at least once [34]. A preliminary step to detect dental caries is the visual inspection of a domain expert, i.e., radiologist or dentist. Since a number of critical dental abnormalities, proximal dental caries as mentioned above, are hard to detect

with visual inspection only, dental radiography has been utilized in most cases [51]. Dental radiography is a standard tool to distinguish dental issues, which helps the experts with better oral health assessment and decreases the clinical observation error to a certain extent [52].

With the introduction of deep learning and its applications in medicine and dentistry, it quickly became a topic of interest for researchers to build disease detection models using medical images as input. Recently, several research studies have proposed deep learning-based Computer-Aided Diagnosis (CAD) systems to detect and screen dental caries based on various types of data, including clinical assessments [53], in-vivo dental images [54], or near-infrared transillumination imaging [35]. Since 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.* [55] developed a deep fully Convolutional Neural Network (FCNN)-based CAD system and applied their model on a large dataset of 3000 Bitewing x-rays to mark dental caries. Their proposed model outperformed the average performance of three certified dentists in terms of overall f1-score on a test set of 500 images. The aforementioned research study is one of the few works in this field to perform segmentation and validate the results using segmentation metrics, i.e., intersection over union (IoU). The proposed FCNN model consisting of more than 100 layers achieved an f1-score of 0.7, while the average score of dentists was 0.53.

Regarding Periapical x-rays, there have been some research works in recent years. In 2016, Choi *et al.* [3] trained a model based on simple CNN architecture along with a crown extraction algorithm using 475 Periapical images to boost the detection rate of proximal dental caries. As authors have claimed, their work was the first research study to propose an automatic detection system for dental Periapical images. Another worth-mentioning point is that the above-mentioned study is one of the few studies in this field to demonstrate the results as probability maps. Examples of their maps are depicted in Fig. 2.5.

Later in 2018, Lee *et al.* [36] used a pretrained Inception V3 for transfer learning on a set

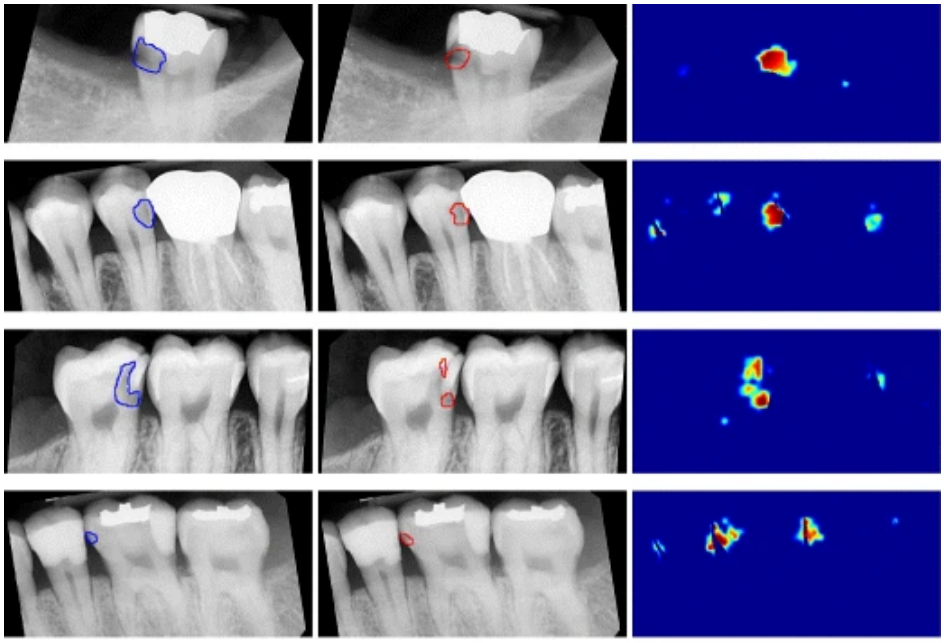


Figure 2.5: Proximal dental caries detection results. a) Ground truth regions (in blue), b) Detected proximal dental caries regions (in red), and c) Caries probability maps [3]

of 3000 Periapical images to diagnose dental caries. This study is marked as one of the first studies that have employed pretrained models to overcome the classification of dental caries using digital radiographs, which was able to achieve promising results. The final accuracy scores were 89.0% and 88.0% for premolar and molar teeth, respectively.

In the most recent research study, Khan *et al.* [56] benefited from a specialist-labeled dataset of 206 Periapical radiographs and trained a U-Net to segment three different dental abnormalities, including caries. They have compared the results of their U-Net with those of SegNet and Xnet (well-known neural network architectures used for segmentation), and have demonstrated its supremacy. While there were no false caries predictions, the reported dice coefficient on the test set was 0.239, which confirms the rigorousness of caries segmentation even on Periapical dental radiographs. The authors have indicated that caries segmentation is a challenging problem with significant room for improvement.

While there have been some works utilizing OPG images, such as classification of tooth types with a Faster-Regional Convolutional Neural Network (Faster-RCNN) [57], there are no studies applying deep neural networks on Panoramic images to detect dental caries. To

be more specific, only one study has aimed to perform tooth segmentation and dental caries detection on Panoramic radiography so far [58]. The authors of the mentioned research work have used simple image processing techniques for tooth segmentation and applied traditional machine learning algorithms for caries detection. They have introduced a dataset of 1392 dental Panoramic x-rays, however, they have used a subset of 700 images only for tooth segmentation which led to producing 1098 cropped images of single tooth for caries classification. While they have reported tooth segmentation accuracy of 71.91% and caries detection accuracy of 98.70%, they indicate their work as a case study concentrating on introducing a new dataset. The mentioned research lacks sufficient explanation of results and no prediction examples of dental x-rays are shown in the discussion. Therefore, it is hard to measure their result transparency and model's generalization, specifically when achieving high accuracy in a very challenging task using simple algorithms.

3. Automated Teeth Extraction from Dental Panoramic X-Ray Images using Genetic Algorithm & Image Processing Techniques

Dental x-ray imaging helps dentists and radiologists to diagnose dental diseases and to provide patients with appropriate treatment planning schedules. In many cases, dental diseases are hard or even impossible to be detected by relying only on visual inspection in clinical assessments. Therefore, automating the diagnosis process has been a topic of interest for dental problems. Teeth extraction is the basic task needed for almost all dentistry decision support systems relying on radiographic images as the inputs. The most challenging type of image to perform extraction on is the Panoramic x-ray, since it includes other parts of the patient's mouth and its structures lack explicit boundaries. The proposed method is the first automated teeth extraction system based on dental Panoramic images using evolutionary algorithms. First, the jaw is extracted from the main image. Then, upper and lower jaws are separated, followed by a genetic algorithm to detect teeth gap valleys. The method is assessed applying to 42 images, where the results are comparable with previous methods developed based on more straightforward dental x-ray types.

Most of the context in this chapter have been published in Haghanifar A, Majdabadi MM, and Ko SB. "Automated teeth extraction from dental panoramic x-ray images using genetic algorithm." 2020 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2020. and Haghanifar A, Majdabadi MM, and Ko SB. "Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier." arXiv preprint arXiv:2012.13666 (2020).

3.1 Background

In this section, a brief review of genetic algorithm (GA) is presented. Afterwards, the Sauvola image enhancement algorithm and its applications are briefly discussed.

3.1.1 Genetic Algorithm

Genetic algorithms are adaptive search methods which were firstly introduced by Holland [59] and then extensively studied by others to solve a wide variety of difficult numerical optimization problems. A genetic algorithm is considered a blind optimization technique that mimics natural evolution's learning procedure using selection, crossover, and mutation. These three procedures are transformed into numerical functions to help solve complex optimization problems without calculating derivatives. Using a random exploration of the search space, GA is more robust in terms of getting trapped in the local extrema [60] as well as being insensitive to the noise due to its random initialization approach [61]. In image processing, GA is proven as a powerful search method that converts an image segmentation problem into an optimization problem [62]. While there have been different genetic algorithms being proposed in research studies, the main phases are considered as follows:

- 1. Initial population: In GA, each possible solution to the predefined problem is coded into a chromosome. A set of chromosomes are called a population, which are actually a group of potential solutions. Each chromosome is represented as a string of characters which can be binary digits or real numbers. At first, the initial population is generally produced with a random function.

- 2. Fitness function: Fitness or cost function is a mathematical representation of the goal problem that the optimization algorithm is trying to find its best possible solution (finding the global exterma on multimodal surface). Every chromosome is assigned a fitness score based on how it fits as the best solution to the problem.

- 3. Selection: The best solutions are selected based on their fitness score. Practically, ranking method is used to rank the solutions and certain number of them are selected to be included in the next iteration.

- 4. Crossover: At each generation, each individual is evaluated and recombined with others to produce new chromosomes, only if its fitness score is higher than the selection threshold. Crossover is performed by randomly selecting a location in the genetic string of the parent chromosomes (current population) and concatenating the selected segment with the final segment of another parent, and the created string is called as a child (next population). The remaining segments of the parents are also merged to create the second child.

- 5. Mutation: Based on the mutation in biological evolution, in rare situations, a small segment of any of the individuals can be randomly changed to a new value, regardless of its helpfulness toward finding the best population. It provides occasional change after crossover operation by altering one or more genes in each child chromosome.

GA is an iterative procedure that tries to find the optimal solution to a complex problem by randomly accessing the multimodal surface. An abstract-level demonstration of the process is shown in the pseudo-code shown in Algorithm 1.

Algorithm 1 Genetic algorithm procedure

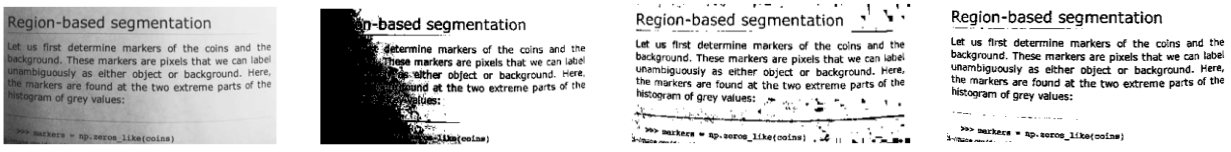
```
i ← 0 ;                                  // assign init value to the number of iterations
fitness_func() ;                                  // determine the fitness function
P(i) ;                                      // create initial population randomly
criterion ;                                // determine the criterion for termination
fitness_func() ← P(i) ;          // evaluate init population using fitness function
while !(criterion) do
    i ← i + 1 ;                                                // increment the counter
    select individuals of P(i) from P(i-1) ;                                // selection
    change individuals of P(i) to generate new children ;                   // crossover
    randomly change some of the digits of P(i) ;                             // mutation
    fitness_func() ← P(i) ;                          // ranking current population
return min(P(i)) ;                  // return best individual from final population
```

The iterative process of population generation and ranking based on the fitness function continues until the chromosomes converge to the minima of the solution space, which is considered to be the global exterma. Another point to mention here is the termination criteria. While the process is expected to converge, in some situations the multimodal space

has multiple exterma or it simply takes too long to be found by random search. Termination criteria must be assigned in these cases in order to restrict the procedure from exhaustion.

3.1.2 Sauvola

Sauvola binarization algorithm is one of the thresholding methods used for image segmentation, originally proposed by Sauvola and Pietikäinen [63]. The original Sauvola method has been presented as a document image binarization algorithm, which divides the page into several sub-components like text, background and picture. The main advantage of the aforementioned algorithm is for text extraction from degraded documents, such as scanned documents. Most conventional binarization algorithms use thresholding only, either global or local, to segment the image into certain clusters. By contrast, Sauvola method firstly performs rapid classification on the image content to categorize into different classes, such as background and text in documents. Then, two different techniques are used to define a threshold for each pixel. First thresholding technique is Soft Decision Method (SDM) which has noise filtering and signal tracking capabilities and is used for background and pictures. Second technique is a specialized Text Binarization Method (TBM) that is utilized for textual and line-drawing areas. TBM is capable of handling text component separation from background in case of the presence of noise or uneven lumination. The results of these two techniques are then combined to produce the algorithm's output image [63]. An example for document binarization is shown in Fig. 3.1, illustrating the superiority of Sauvola algorithm over other well-known methods[1].



(a) (b) (c) (d)

Figure 3.1: Document thresholding example, demonstrating better performance of the Sauvola method. a) original image, b) global thresholding, c) Niblack thresholding, and d) Sauvola thresholding.

[1] https://scikit-image.org/docs/dev/auto_examples/segmentation/plot_niblack_sauvola.html

As seen above, Sauvola algorithm results in a cleaner image than the output of the Niblack method. Sauvola is actually inherited from Niblack method, which could also successfully resolve the black noise issue. Sauvola thresholding value for each window is calculated as shown in the following:

$$T = m \times (1 - k \times (1 - \frac{\delta}{R})) \tag{3.1}$$

where m and δ are mean and standard deviation of the pixel values in each specific window, respectively. k is a control factor restricted to be in the range of $[0.2, 0.5]$, and R is a predetermined image grey-level value. The result T is the threshold value to be applied on each specific window [64].

Despite being initially proposed for document thresholding, Sauvola thresholding has been employed for a variety of segmentation tasks, including biomedical applications. For instance, Mustafa *et al.* [65] have proposed a new segmentation method for automated malaria detection inspired by Sauvola's primary segmentation method. Moreover, Sauvola algorithm has also been mentioned in literature as a benchmark for comparing custom segmentation algorithms, together with other common algorithms like Niblack, Nick, etc.

3.2 Material

The dataset for teeth extraction consists of 42 dental Panoramic x-rays randomly collected from a radiology clinic. With respect to the privacy of the patients, the name and the age of each person are anonymized. Each radiograph is a grey-scale image formatted as Bitmap (BMP) file. Initial resolution for each image is 3292×1536 pixels with a bit-depth of 8. All images are taken by SOREDEX's CRANEX ®3D system. A sample image is shown in Fig. 3.2.

The employed data is a subset of the main dataset proposed for dental caries detection model, and a more detailed investigation though the dataset features is provided in the next section.

Figure 3.2: A raw Panoramic x-ray from the proposed dataset

3.3 Method

Panoramic images contain unwanted regions around the jaw, which should be eliminated before jaw separation. Unlike Periapical images, in Panoramic images, upper jaws (maxilla) and lower jaws (mandible) need to be separated before tooth isolation. Each step also requires its own preprocessing method to increase the performance efficiency. A high-level illustration of the proposed pipeline is depicted in Fig. 3.3.

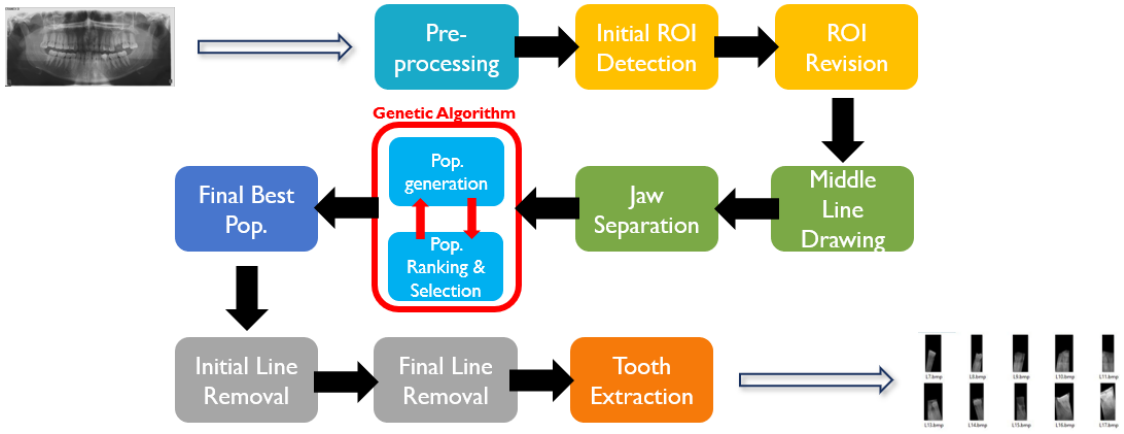


Figure 3.3: General diagram of the teeth extraction model.

As shown in Fig. 3.3, raw Panoramic images are the input of the system. Preprocessing, Region of Interest (ROI) extraction, and jaw separation are done, one after another, to make the image ready for the genetic algorithm to generate vertical lines for tooth separation. Line removal is then applied to output images, each including one tooth. The generated images

are then fed to the caries detection model to detect dental caries and their severity, which will be discussed in the next section.

In this section, we investigate the procedure of teeth extraction step-by-step. Firstly, raw images are enhanced using various preprocessing steps. Then, extraction of the ROI is discussed. On the next step, jaw separation using the proposed snake-inspired image crawler is introduced and the details are mentioned. Finally, a custom redefined GA method is discussed and the process of teeth extraction is investigated.

3.3.1 Preprocessing

In order to enhance the image and highlight the useful details, few preprocessing steps have been taken. For extracting the jaw as the region of interest (ROI) within the main image, two steps are considered. Fig. 3.4 illustrates the preprocessings performed for ROI extraction.

1) Preprocessing for Initial ROI Detection: First, a vertical-edge detection filter is applied to highlight the vertical edges. Then, a bilateral filter is used for sharpening the edges.

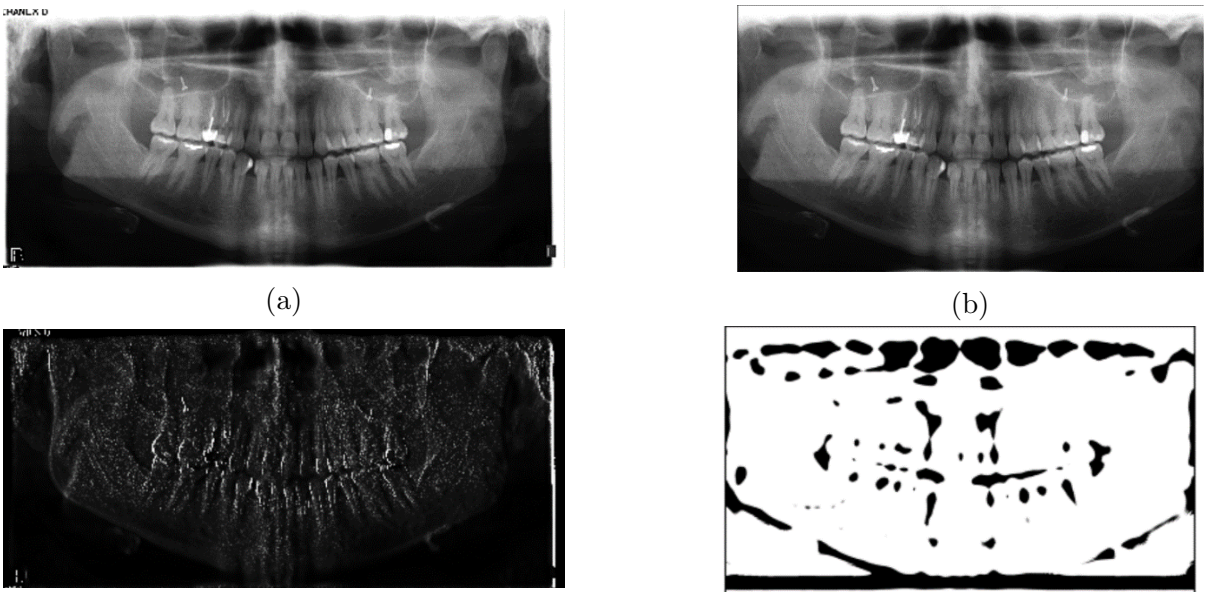
2) Preprocessing for Final ROI Extraction: After cropping the initial ROI through the main image, Sauvola segmentation algorithm is applied to binarize the image in order to blacken the two ends of the gap between maxilla and mandible. On next, a Gaussian filter is used to reduce the noise.

Fig. 3.4 demonstrates how vertical edges are whitened in the main image to find both angles of jaw. In the last image, external oblique ridges are recognizable as bigger black spots on both sides of the jaw.

3.3.2 ROI Extraction

To reach the jaw from the main image, unwanted surroundings need to be discarded. ROI extraction is performed in two steps, as illustrated in Fig. 3.5.

1) Initial ROI Detection: After the preprocessing, horizontal integral intensity projection is applied on the image, where the first significant positive slope represents the edge of the



(a) (b) (c) (d)

Figure 3.4: Preprocessing for ROI extraction. a) main image, b) initial ROI detected, c) vertical edges extracted, and c) Sauvola binarization applied on the cropped image.

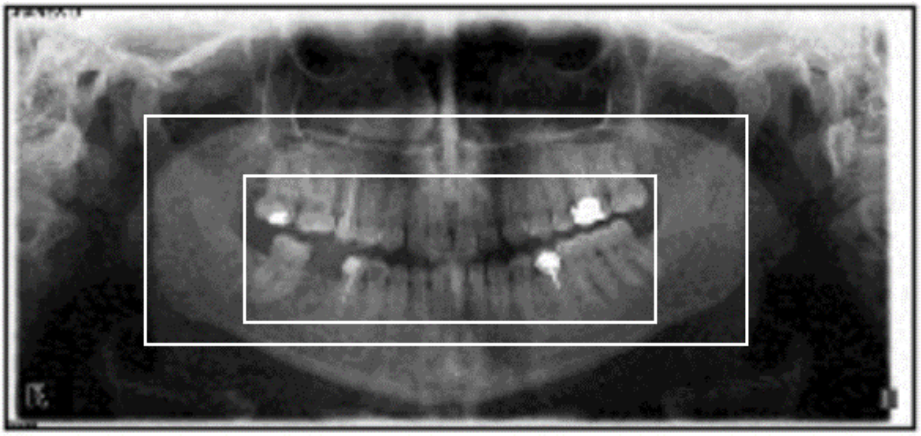


Figure 3.5: Initial ROI detection and final ROI extraction displayed as white rectangular boxes on the main image.

left angle of jaw.

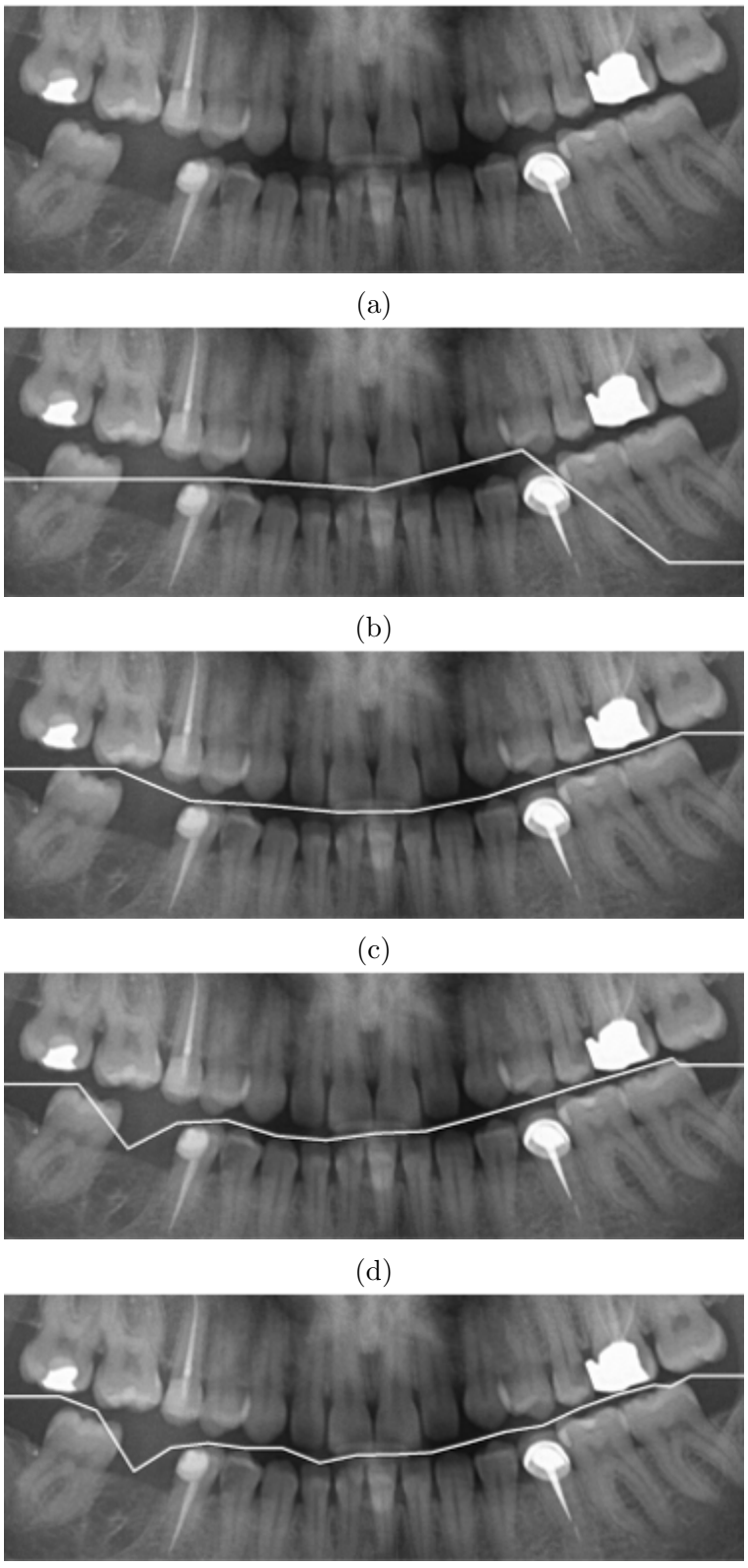
2) Final ROI Extraction: Horizontal integral intensity projection is performed on the cropped image to distinguish the external oblique ridge, starting point for the gap between maxilla and mandible, which appears to be the first significant negative slope in the intensity graph.

3.3.3 Jaw Separation

As the jaw is extracted out of the main image, maxilla and mandible are expected to become separate using a line which passes through the gap while maintaining largest possible distance with tooth crowns from both jaws. In order to create the separating line, two procedures are proposed to be evaluated: middle points and snake methods. Middle points method divides the image into several parts. Then, vertical integral intensity is computed and the minimum value is considered to be the point representing the gap. Final line is formed by connecting these points to each other. Middle points solution is a relatively simple approach and at the same time, is computationally fast. On the other hand, it is prone to the darker background around the tooth roots. Another problem with the middle points method is its inflexibility with missing teeth. An intuitive example with different values for the hyperparameter is demonstrated in Fig. 3.6.

As seen in Fig. 3.6, small values for the number of points hyperparameter may result in missing multiple tooth crowns. With increasing the number of points, separating line get better form, while the model needs more computations. Even with high number of points, like 20, missing teeth lead to the separating line cutting part of the neighboring teeth due to the straightness of the connecting lines.

Another method is the snake algorithm. First, a starting point is determined, due to the fact that snakes are sensitive to local exterma and get stuck often [23]. Afterwards, it crawls through both left and right directions; looking for paths with minimum integral intensity. The length of each path-step is a controlling parameter to restrict the crawling snake from getting trapped in teeth-gap valleys. A similar concept of snake method has been utilized in [16]. Fig. 3.7 depicts the concept of snake method for jaw separation.



(a)

(b)

(c)

(d)

(e)

Figure 3.6: Illustration of middle points method implementation results on a sample preprocessed image. a) original image, b) middle points method with number of points set to 5, c) number of points set to 10, d) number of points set to 15, and e) number of points set to 20.

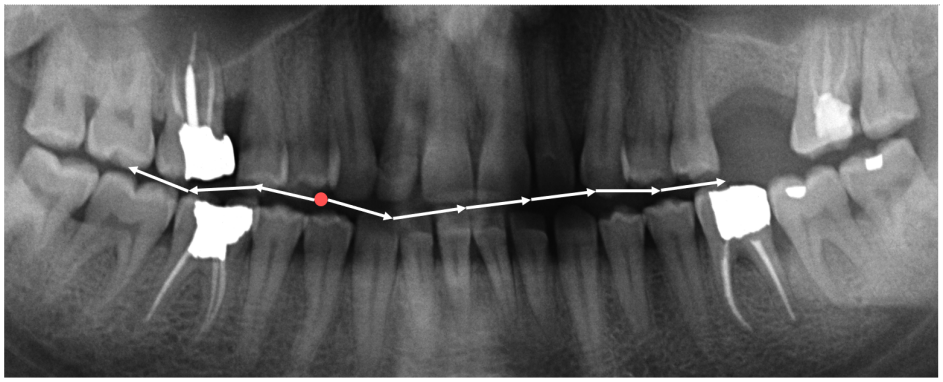


Figure 3.7: Concept of the snake method on jaw separation in a sample preprocessed OPG image. Starting point is shown in red, while path-steps are in white.

While less computationally efficient than the middle points method, it has drawn a better separation line between the jaws. To have a better comparison with the former method, snake method with different hyperparameter values are implemented on the same image for the sake of comparability. Fig. 3.8 shows the results.

As seen in Fig. 3.8, unlike the middle points method, snake provides acceptable results even with small step length values. Besides, the missing teeth problem is also resolved, since snake crawls through the gap between the jaws, and never falls inside a teeth-gap valley.

Another advantage of the snake method is that it also works flawlessly when applied on closely-stacked-together mandible and maxilla. In these images, the patient has firmly bitten the bite block of the extra-oral imaging unit, which is a common phenomenon. Examples of snake method on these images are shown by the Fig. 3.9.

3.3.4 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 in 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. On one hand, the number of chromosomes here is more than the average number of teeth in each jaw, to

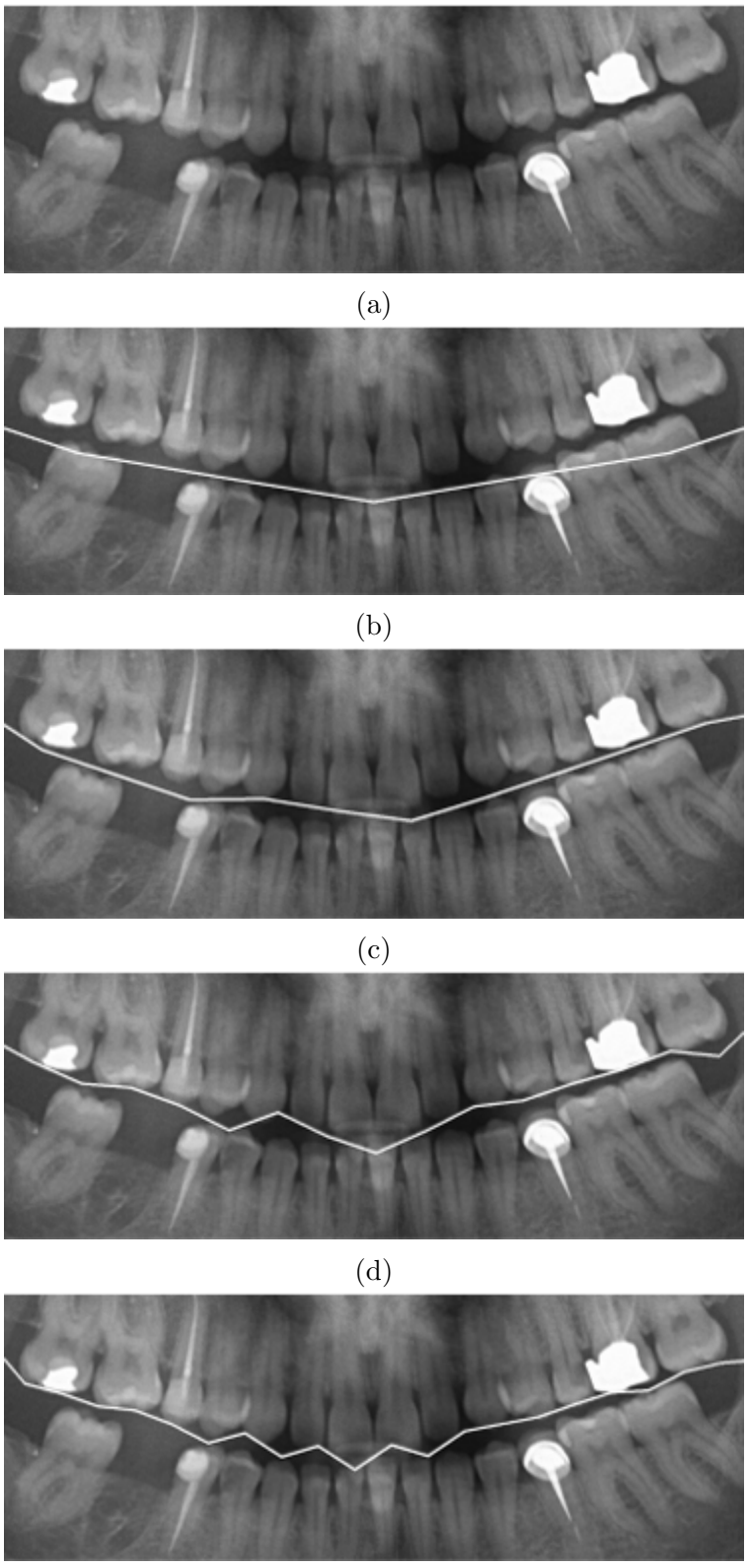
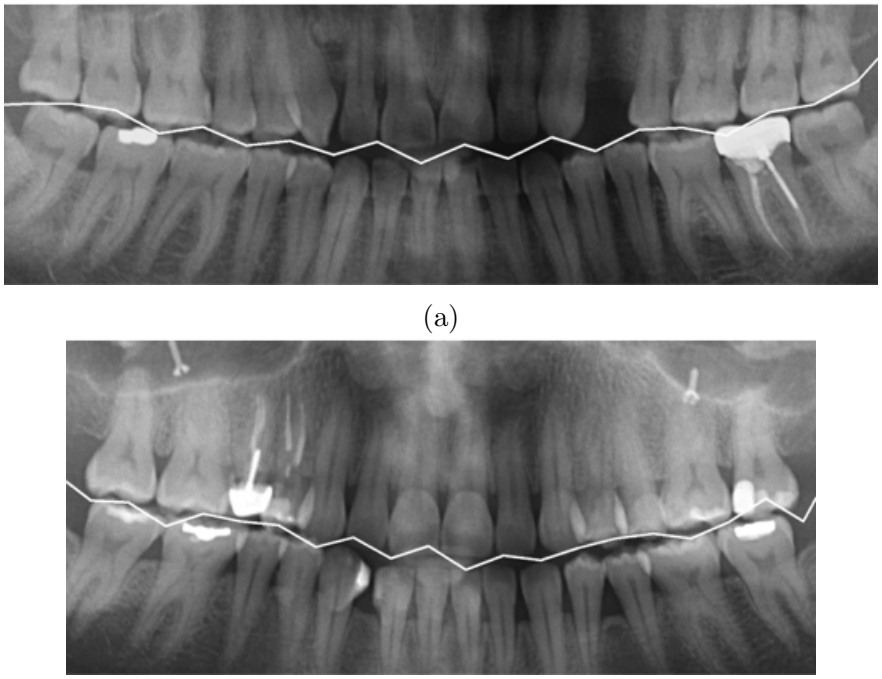


Figure 3.8: Illustration of snake method implementation results on the sample image. a) original image, b) snake method with step length set to 5, c) step length set to 10, d) step length set to 15, and e) step length set to 20.



(a)

(b)

Figure 3.9: Examples of snake method applied on closely-stacked-together jaws, which shows its good performance.

make sure that most of the gaps between teeth are included. On the other hand, taking into consideration the width of a typical tooth body and the line removal techniques being discussed further, assigning any number higher than 30 will lead to the same result. These paths are considered as including gaps between proximal teeth. The GA can dynamically change the image segmentation task controlling parameters to reach the best performance. The GA-based tooth segmentation 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:

$$C(x) = \sum_{i=1}^{n} I(x_i) \tag{3.2}$$

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. It is worth mentioning that the padded pixels after the cropping process are not considered in this function.

- 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. 3.10.

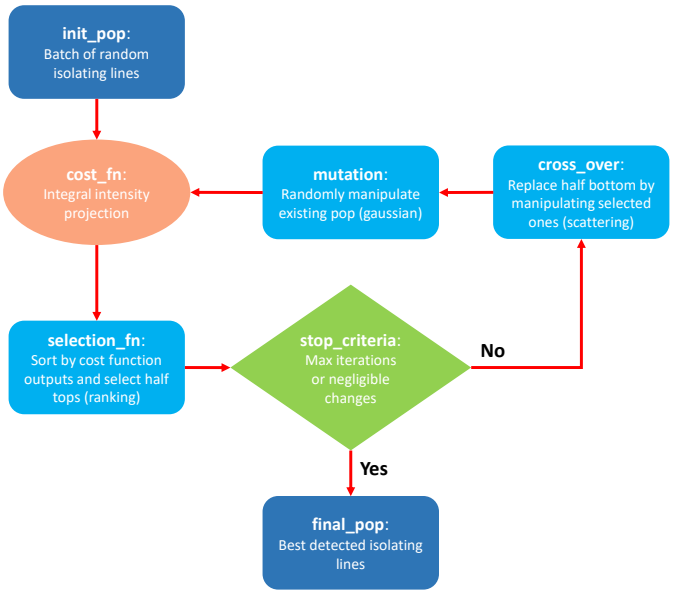


Figure 3.10: Flow chart of the proposed GA-based tooth isolation method

The genetic cycle mentioned above is finally 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. Initial line removal technique is to look

forward lines which are closer than a certain distance; keeping the best line and eliminating others. Supplementary line removal method is to select a certain number of remaining lines, which have the lowest cost function values.

3.4 Results and Discussion

In this section, the results of the proposed approach and intermediate outputs are presented followed by discussions. The algorithm is applied on 42 images, where the total number of teeth is 1229; 616 maxillary and 613 mandibular. At first, the jaw extraction and separation are tested. Subsequently, the efficiency of teeth extraction is investigated. As far as the accuracy is concerned, the ratio of correctly separated segments to total number of parts is considered as the accuracy reported for separation. Note that this is a common accuracy definition in separation approaches, while in segmentation approaches the accuracy is calculated pixel-wise [22].

3.4.1 Jaw Separation

The process of separating jaw region from the surrounding unwanted area is performed on the dataset, and the success rate of 40 out of 42 images is acquired. Thus, the accuracy of jaw extraction is calculated as 95.23%. It also failed to correctly identify some of the wisdom teeth. As the result, 4 maxillary and 2 mandibular wisdom teeth were also missed. Final number of remaining teeth is 582 in maxilla and 581 in mandible.

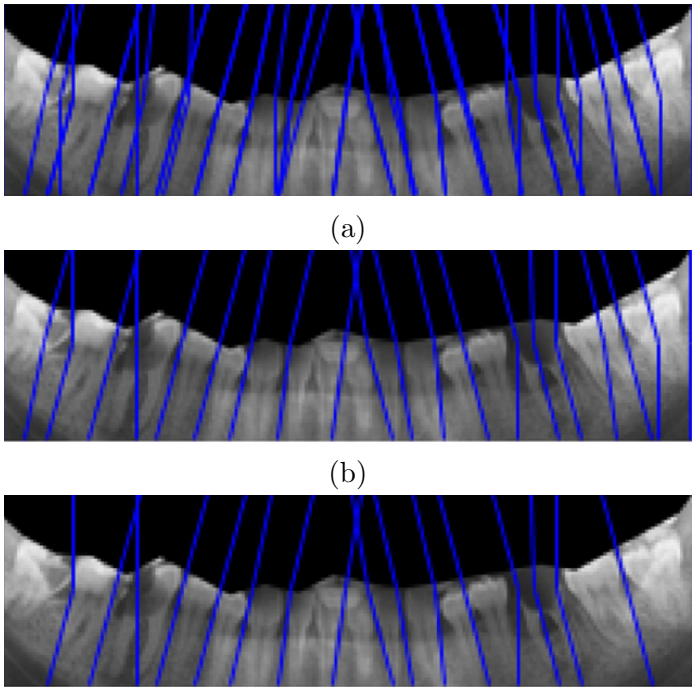
Next, jaw separation is applied on 40 images comparing middle points and snake method explained in previous section. While middle points approach failed to correctly separate the jaws, and the crowns of one or more teeth are misclassified many times, snake approach demonstrates better performance and consistency. It also proves to work well even on closely-stacked-together jaws, as mentioned earlier in the previous section. Sample results of middle points and snake method are illustrated in Fig. 3.6 and Fig. 3.8 in section 3.3.3, respectively.

3.4.2 Tooth Isolation

Genetic algorithm is applied on extracted jaws, followed by line removal techniques to reduce the number of wrong lines, which are mostly the ones passing through the teeth instead of teeth-gap valleys. Results on maxillary and mandibular teeth are presented in Table 3.1. Concept of the procedure and the output of tooth isolation for sample jaws are depicted in Fig. 3.11. and Fig. 3.12, respectively.

Table 3.1: Accuracy of the proposed tooth extraction method in Maxilla and Mandible

Jaw Type	**Total Teeth**	**Separated Teeth**	**Accuracy**
Maxilla	582	474	81.44%
Mandible	581	332	57.14%



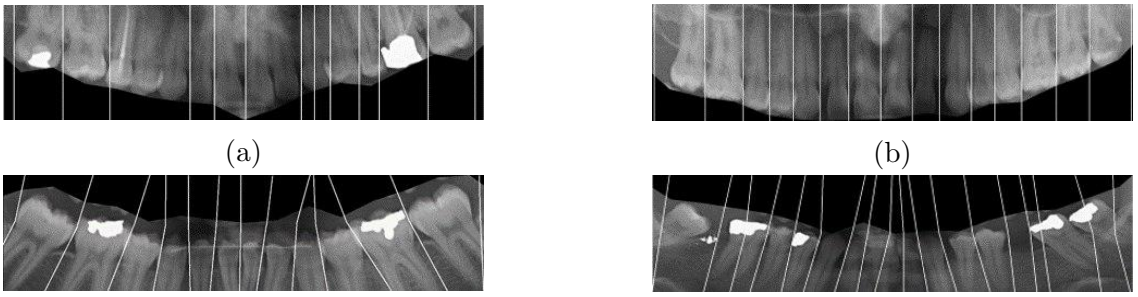
(a)

(b)

(c)

Figure 3.11: Results of the proposed GA-based line fitting method for a random mandible image. a) raw output of the genetic algorithm, b) initial line removal applied, and c) final line removal applied.

In the maxilla, the proposed algorithm performs well for isolating incisor teeth. However, the performance on molar and premolar teeth is mediocre. Most of the times two central incisors are missed, due to the lack of clarity in boundaries. This lack of clarity, like dark



(a) (b) (c) (d)

Figure 3.12: Tooth separation sample results on a sample OPG image. a) wrong maxillary separation, b) correct maxillary separation, c) correct mandibular separation, and d) wrong mandibular separation.

shadows, is common for central incisors in Panoramic images, which is caused by the imaging method. Although the applied approach has shown good capability in terms of isolating incisor teeth, the accuracy of molar and premolar separation is quite low. Mandibular Wisdom teeth are also not correctly isolated. Since molar and premolar teeth are more skewed, generated lines need to have more freedom during population generation. Enhancement strategies for mandibular teeth isolation are investigated next.

To improve the accuracy, the cost function is intentionally changed so that it tries to highlight the intensity of lines where the generated teeth separation lines are close. In order to implement this idea, a compression coefficient is multiplied to each average intensity. This coefficient is ranged between 1.0 and 2.0. New cost function is formulated in the equation 3.3.

$$C(x) = \sum_{n}^{i=1} I(x_i) * (0.5 \times tanh(3.5 - D) + 1.5) \tag{3.3}$$

where D is the average distance of the line x_i with its neighbors. Other definitions are the same as those in the equation 3.2 in the previous section.

Moreover, line removal is also changed. Previously, teeth separation lines which were closer than a minimum distance were being grouped together, and the best ranked line remained and others were eliminated. New line removal method features a position-variant

minimum distance definition. As the molar and premolar teeth are larger than the middle teeth, minimum distance gradually increases for corner lines. Comparison of accuracy values for three above-mentioned teeth separation methods is illustrated in Table 3.2.

Table 3.2: Comparison between different combinations of cost function and line removal methods

Cost Function	**Line Removal Technique**	**Total Accuracy**
Mean Intensity	Minimum Distance	57.14%
Mean Intensity × Compression Coefficient	Minimum Distance	63.16%
Mean Intensity × Compression Coefficient	Position-Variant Minimum Distance	73.67%

Table 3.2 demonstrates how improving the cost function followed by line removal function has led to accuracy enhancement of approximately 16.5%.

To resolve any ambiguity related to the metric for tooth isolation evaluation, the definition of the confusion matrix elements are elaborated as well. True Positive (TP) happens when the extracted image is actually a tooth along with its crown and roots. A false positive (FP) output is when the extracted image shows a tooth, but also contains a portion of the neighboring tooth. A false negative (FN) occurs when the extracted image does not even cover a complete tooth area. An intuitive example is depicted in Fig. 3.13.

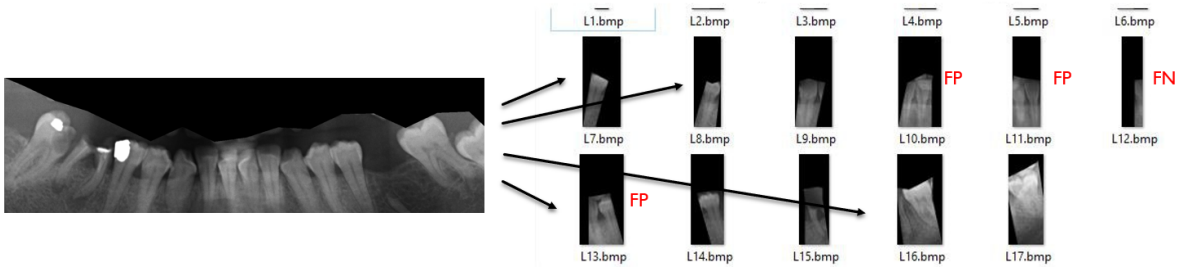


Figure 3.13: Example of the extracted images from an OPG radiography with the wrong classifications labeled in red.

To compare our proposed model with the literature mentioned in section 2.2, accuracy for both upper jaw and lower jaw are investigated. Table 3.3 briefly explains the comparison between our proposed method and other research studies.

Table 3.3: Comparison between proposed method and other teeth extraction research works.

	Image Type	**Algorithm**	**Correct Upper**	**Correct Lower**
Abdel-Mottaleb et al. [47]	Bitewing	Integral Intensity Method	169/195 – 85%	149/181 – 81%
Olberg and Goodwin [2]	Bitewing	Path-based Method	300/336 – 89.3%	270/306 – 88.2%
Nomir et al. [66]	Bitewing	Integral Intensity Method	329/391 – 84%	293/361 – 81%
Al-Sherif et al. [43]	Bitewing	Energy-based Method	1604/1833 – 87.5%	1422/1692 – 84%
Ehsani Rad et al. [15]	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, the resulting accuracy is in the line of previous works. The segmented teeth from this model will be used as the input for the dental caries detection model, which will be explained in the next section.

3.5 Summary

Dental x-ray imaging helps dentists and radiologists to diagnose dental diseases and to provide patients with treatment plannings. In many cases, dental diseases are hard to detect by relying only on visual inspection in clinical assessments. Therefore, automating the diagnosis process has been a topic of interest for dental problems. These automated models are considered as dentistry decision support systems or dentistry computer-aided diagnosis models, as a subcategory of medical systems in general. While there have been some research studies and even some commercialized systems operating in some healthcare institutions, developing a robust dentistry CAD system is a long way.

Teeth extraction is the basic task needed for nearly all dentistry decision support systems relying on radiographic images as the inputs. The most challenging type of image to perform extraction on is the Panoramic image since it includes other parts of the patient's mouth, and structures lack explicit boundaries. Unlike intra-oral imaging techniques, i.e., Bitewing and Periapical, in extra-oral imaging the raw image needs further preprocessing and ROI detection steps. In Panoramic x-ray, a considerable part of the image is filled by background and jaw-bones, which could be helpful in detecting bone-related problems, but not needed in our case. To provide a dental caries detection system, the only essential part of the OPG

image is the tooth crown and body, and more preferably, the roots with complete boundaries.

The proposed method in this section is the first automated teeth extraction system from dental panoramic images using evolutionary algorithms. First, the jaw is extracted from the main image. Then, upper and lower jaws are separated, followed by a genetic algorithm to detect teeth gap valleys. A more detailed diagram is presented by Fig. 3.14.

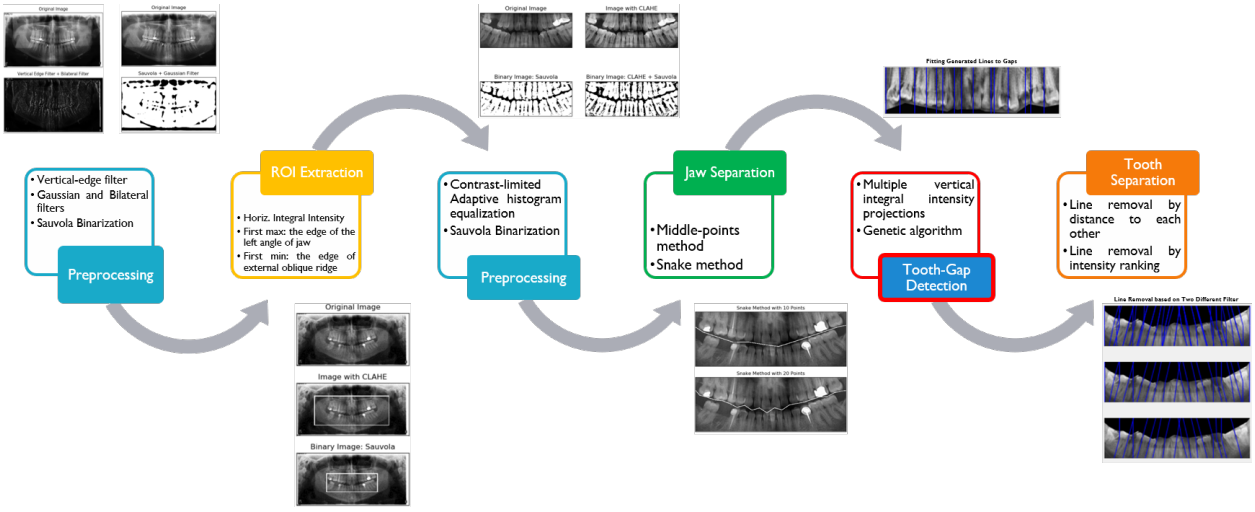


Figure 3.14: Detailed diagram of the proposed teeth extraction system

As shown in Fig. 3.14, unpreprocessed Panoramic 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 tooth separation. Line removal is then applied to output images, each including one tooth. These generated images are then fed to the dental caries detection model to detect dental caries and their severity, which will be addressed in the upcoming section.

Aforementioned earlier in this section, teeth isolation is a mandatory preliminary part needed for developing an automatic dental decision support system. Most previous research works utilized their teeth extraction algorithms on Bitewing or Periapical images. In this study, we presented a genetic-based approach for teeth extraction in Panoramic images. By using compression coefficient along with average intensity as the improved cost function and position variant minimum-distance approach for enhanced line removal, the accuracy is notably increased. On a dataset of 42 images, it has achieved the overall accuracy of

77.56%; where the accuracy values are 81.44% and 73.67% for maxillary and mandibular teeth, respectively. The acquired scores are comparable with other studies as compared in Table 3.3. Although the accuracy score could be improved with more complex models and a larger dataset, this system is a step towards a fully automated dental disease diagnosis and treatment suggestion system. The results demonstrate the efficiency of our method on a small portion of input data.

4. Dental Caries Detection in Panoramic X-ray using Ensemble Transfer Learning & Capsule Classifier

Dental caries is one of the most chronic diseases involving the majority of the population during their lifetime. Caries lesions are typically diagnosed by general dentists relying only on their visual inspection using dental x-rays. In many cases, dental caries is hard to identify in 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, which benefits from various deep pretrained models through transfer learning to extract relevant features 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 lesions are more straightforward to detect and efficient mild caries detection needs a more robust and larger dataset. Considering the novelty of current study as using Panoramic images, this work is a step towards developing a fully automated efficient decision support

Most of the context in this chapter have been published in Haghanifar A, Majdabadi MM, and Ko SB. "Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier." arXiv preprint arXiv:2012.13666 (2020).

system to assist domain experts with clinical observations.

4.1 Background

In this section, capsule network and its general architecture are reviewed. Then, pre-trained models used in this study are investigated. Firstly, InceptionNet and the ImageNet dataset are explained. Secondly, the CheXNet model and its performance on other types of x-ray images are reviewed.

4.1.1 Capsule Network

Since the introduction of capsule network [67], many studies have benefited from its advantages in various applied deep learning tasks [68–71]. Capsule network consists of computational units called capsules. Each capsule is a group of neurons, nested together to represent the substantiation parameters of a particular feature by using a vector. 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 by 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. Squash function is one of the relatively recent activation functions being introduced and widely used in the literature, and is formulated as follows:

$$V_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \frac{S_j}{\|S_j\|} \tag{4.1}$$

where, S_j is the vector after applying dynamic routing.

As mentioned earlier, the length of the vector corresponds to the probability of the

presence of that particular feature. Each vector is multiplied by a weight matrix to predict vectors that each corresponds to the higher-level feature. Then, dynamic routing examines the agreement between predictions, and outputs the final vector for each capsule in the second layer. Fig. 4.1 indicates the structure of a typical two-layer Capsule network utilized for caries detection.

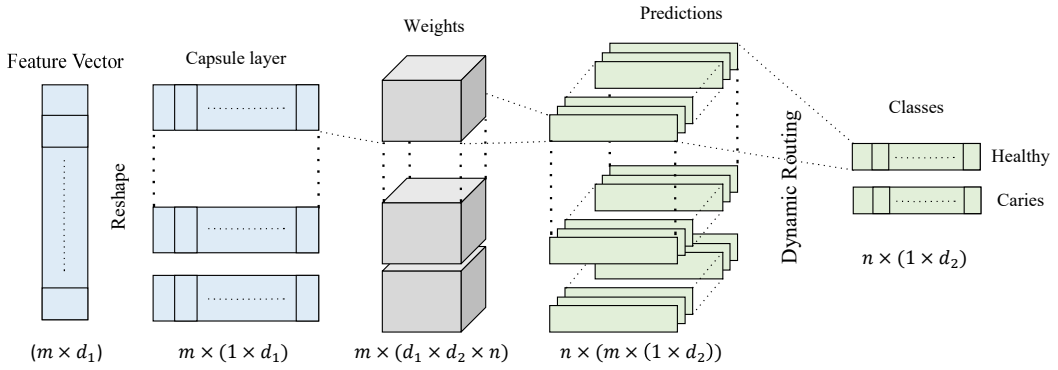


Figure 4.1: The architecture of the CapsNet for binary classification of dental caries from the extracted features.

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 further in this section. The outputs of the CapsNet are two vectors representing two classes, healthy and caries. The length of these two vectors corresponds to the probability of each class.

Besides, this network is capable of learning geometrical relationship between features as well. Hence, it can offer unprecedented robustness in a variety of tasks [72–74]. The main challenge in dental caries detection in Panoramic images is accurately distinguishing between dark regions caused by shadows and the actual caries spots. Geometrical features such as edges, textures and location of these darker regions are essential for correct classification. This is the reason for why it is believed that capsule network is the perfect choice for this task.

4.1.2 InceptionNet

Inception networks are a group of neural network architectures consisting of a similar Inception block. Originally known as GoogLeNet, Inception has been able to set a new state-of-the-art (SoTA) top-5 and top-1 accuracy score on the well-known ImageNet dataset. InceptionNet has not only outperformed the previous best architecture's top-5 accuracy (VGG with 7.32% while GoogLeNet reached 6.67%), it is also more computationally efficient. The basic idea of the original authors was to introduce wider networks instead of going deeper and deeper by concatenating more convolutional layers [75].

The motivation behind approaching wider networks was the large size variations between salient parts in a sample image. While the final goal is to detect particular objects in images, the objects might appear with different sizes in each image. Thus, choosing the right kernel size for convolution operation becomes troublesome. Larger kernels are preferred for globally-distributed information, and smaller kernels are more favorable for information that is distributed more locally. Introducing a very deep network requires large data and is highly prone to overfitting, as well as being computationally expensive. The authors of GoogLeNet came with the idea of choosing filters with multiple sizes in each level to overcome the aforementioned problems with deep networks. The initially proposed Inception block, depicted in Fig. 4.2, utilized kernels with three different sizes on the same level.

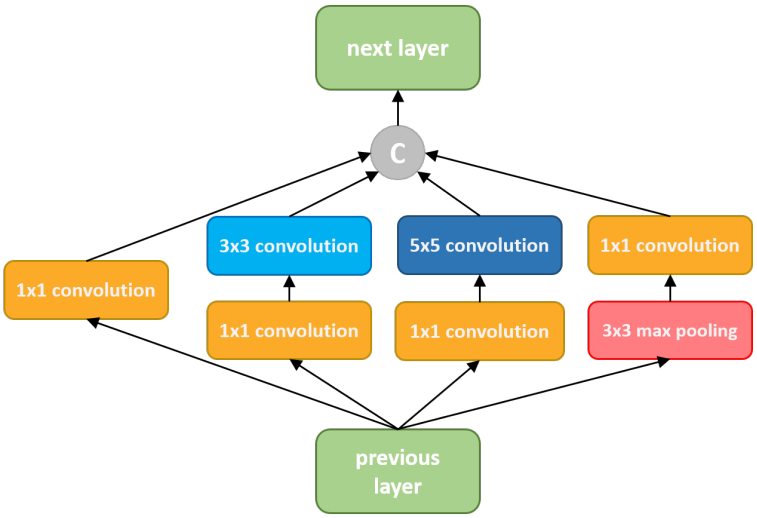


Figure 4.2: The architecture of the Inception module with dimension reductions.

There are 1×1, 3×3, and 5×5 kernel sizes in a typical Inception block. The 1×1 kernel performs as a dimension reduction layer, which is used to merge different dimensions of the input tensor to decrease the computational load of the network. Max pooling layer has also been used in its own path along with a dimension reduction kernel. Finally, the convolution results of all paths are concatenated to be passed into the next layer.

Despite being wider and employing different kernel sizes, the InceptionNet model has only 9 Inception modules stacked one over the other. Hence, it only has 22 layers (or 27 layers including pooling layers). At the end of the last Inception block, a global average pooling layer is used, followed by a linear layer with SoftMax loss as the activation function. A schematic view of the network is depicted in Fig. 4.3.

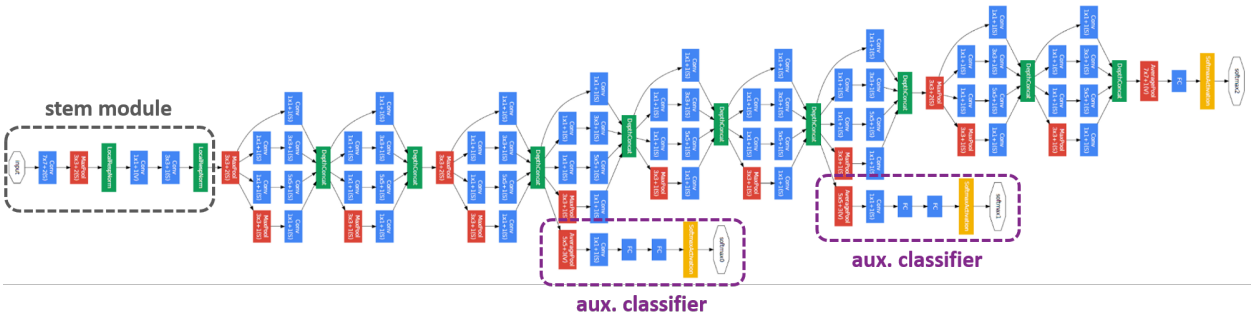


Figure 4.3: The architecture of the InceptionNet model. The black box is the stem, and purple boxes are auxiliary classifiers.

The black box in Fig. 4.3 is the stem, which includes some preliminary convolution layers. Purple boxes are the auxiliary classifiers. Since the model is a deep classifier, the vanishing gradient problem still is an important issue. To prevent the middle layers from dying out, the authors have devised auxiliary classifiers as the SoftMax layers being applied on top of some Inception blocks in the middle of the network to compute the auxiliary loss. The final loss function is a weighted summation of the main loss (related to the last layer) and auxiliary loss values.

The above-mentioned network architecture is the original version of the Inception networks. In the next versions, the authors proposed a number of upgrades to increase the accuracy score, and at the same time, reduce the computational complexity. The upgrades

include converting convolutions with larger kernel sizes into smaller ones by utilizing smart factorization, including a factorized 7×7 convolution, label smoothing as a routine regularizing component to hinder the overfitting, etc. While there exist models with higher accuracy scores on ImageNet, such as ResNet and EfficientNet, InceptionNet is relatively small and consists of a few layers. This key advantage makes it a perfect choice when dealing with small datasets, like the case of the current research study.

4.1.3 CheXNet

Transfer learning is an emerging solution to deal with problems with a lack of sufficient data to train a neural network from scratch. It simply means to benefit from a pretrained model in a new classification task, where the target patterns are expected to overlap with the original dataset that the pretrained model is previously trained on. Most pretrained models are the ones trained on ImageNet dataset, since ImageNet has more than a million images from 1000 different classes. Classes in ImageNet are common objects, where the higher level features are different from objects that usually appear in a radiography image. Therefore, using a pretrained model from a similar type of image might usually result in a better accuracy score. As mentioned in previous sections, there are few available datasets related to dental radiography. Hence, no pretrained models are available being trained with dental radiography. In spite of having a lack of models in this area, there are a few models with different types of x-ray images. CheXNet is considered as the most robust pretrained model publicly available for academic use.

CheXNet is a robust model for lung disease detection based on chest x-ray images [26]. It has been trained on CXR-14, one of the largest publicly available datasets of chest x-rays from adult cases with 14 different disease labels, including pneumothorax, pneumonia, etc. CheXNet is a 121-layer convolutional neural network trained on the dataset mentioned above with over $100,000$ frontal view x-ray images. The architecture of CheXNet is the same as a 121-layered DenseNet model, one of the well-known CNNs that has previously achieved SoTA on ImageNet dataset. The architecture of CheXNet is presented in Fig. 4.4.

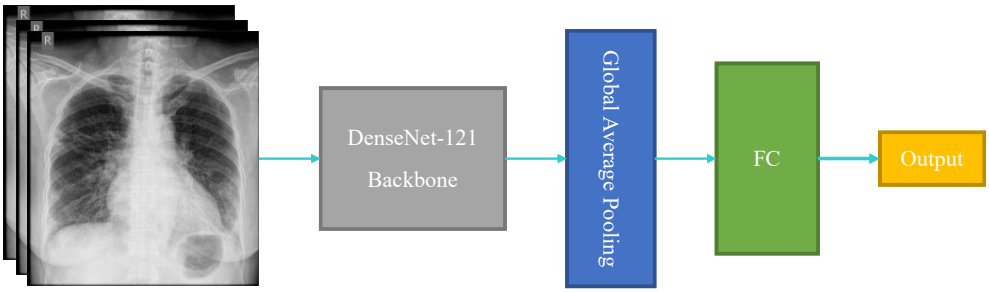


Figure 4.4: The architecture of CheXNet.

The input is passed to a DenseNet-121 backbone, followed by a global average pooling, and a Fully Connected (FC) layer to generate the output. The architecture of the DenseNet itself is shown in Fig. 4.5.

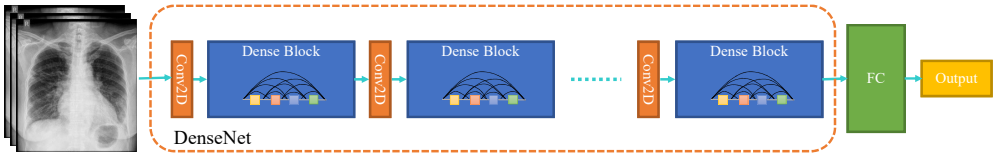


Figure 4.5: The architecture of a 121-layered version of DenseNet model.

As seen above, DenseNet consists of multiple dense blocks. Each dense block has four layers, and the output of each layer is connected to all the subsequent layers.

CheXNet is trained on many images for many epochs, and the authors have claimed that it outperforms general radiologists in terms of disease detection f1-score. The frontal chest x-ray is indeed different from dental Panoramic x-ray, however, the basic image features extracted by the very first layers could be very informative. Thus, it is considered to be helpful to include the pretrained CheXNet in our proposed model as well.

4.2 Material

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 (either deeper or wider), 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.

4.2.1 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 repositories. 280 images are obtained from the Diagnostic Imaging Center of the Southwest State University of Bahia (UESB) [4]. 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. 4.6.

(a) (b)

Figure 4.6: a) An example image from the UESB dataset with b) the related tooth annotation mask [4]

Besides, we also collected 120 images from a local dentistry clinic taken with x-ray camera model Cranex 3D from Soredex. The radiographs are anonymized and contrast enhancement is applied on them by the source practices before they are shared with the research team. 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 tooth segmentation 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.

4.2.2 Labeling Process

UESB images have tooth masks manually prepared along with the dataset, which is used to segment 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, as explained with details in previous chapter. To perform labeling, an oral and maxillofacial radiologist commented on Panoramic images one by one. Segmented 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 pulp. Pulpitis caries lesions result in a collapsed tooth. It is worth mentioning that we have benefited from an expert oral radiologist with +20 years of experience as the label provider. However, there is a lack of external ground truth, which requires each patient to have multiple Panoramic and Intra-oral radiography in different time intervals. Therefore, our performance is limited to the radiologist's annotations.

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

[1] `https://radiopaedia.org`

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 [76].

4.3 Model Architecture

In this section, the proposed model and its components are investigated. Details of the composing parts, such as Capsule Network and CheXNet, are already explained in the previous section.

4.3.1 Proposed Network: PaXNet

The proposed network for caries detection in Panoramic dental x-rays (PaXNet) consists of two main modules; feature extraction and classification. 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, 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 set as non-trainable, and their top dense layers are removed. InceptionNet, as mentioned in the previous section, is a powerful model trained on the ImageNet dataset [75]. Since caries appears in different sizes, this model's ability to learn features with multiple sizes, thanks to its multiple-sized convolution kernels, is beneficial for caries detection. However, the model is trained on RGB images of common objects that share no high-level similarities with Panoramic dental x-rays. This is why CheXNet is included in feature extraction block as well. CheXNet is a robust model for thoracic disease detection based on chest x-ray images [26]. Although Panoramic dental x-rays are different from chest x-rays, since they are both x-rays, they share many similar features of a 2D grey-colored radiography. There might be some features specific to a tooth that is not covered by the two models mentioned above. To solve this potential problem, a

pretrained encoder is used to benefit from these types of image features.

The number of labeled tooth images is limited and expanding the dataset requires huge amount of time and effort for annotating specialists. By contrast, there is a considerable number of unlabeled images available both on the web and in our own dataset. To benefit from this vast dataset, a new approach is advised. An unsupervised side-task is designed, and a model is trained using a large 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 also embedded in PaXNet so that the model could be able to learn task-related specific features as well. Table 4.1 presents a comparison between these four feature extraction networks.

Table 4.1: Comparison between feature extractor networks

Model	**Num. Layers**	**Num. Params**	**Trainable**	**Dataset**	**Dataset Size**
InceptionNet	42	451,160	False	ImageNet	1,281,167
CheXNet	140	1,444,928	False	CXR-14	112,000
Encoder	4	14,085	False	Unlabeled teeth	11,769
CNN	8	35,808	True	-	-

As the similarity of the training dataset to the caries detection dataset increases, the number of available samples is reduced, to almost one tenth. 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 [77]. The swish activation function is formulated as follows:

$$f(x) = \frac{x}{1 + e^{-x}} \tag{4.2}$$

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 the Leaky-ReLU.

2) Classification

In PaXNet, all extracted features are concatenated, and higher-level features are created based on them using a CNN. Then, the last convolution 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.

SoftMax, or normalized exponential function, is a generalized logistic function applicable to multiple dimensions. It can be considered the same as Sigmoid function, being used when there are multiple classes to be predicted. SoftMax ensures that sum of the values for all classes are always set to 1. The standard SoftMax function is defined by the formula 4.3 below. While this case is a binary classification, we detached the two classes into two different neuron. Thus, Sigmoid is no longer applicable as the last layer here.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{4.3}$$

where $\vec{Z}$ is the input vector, K is the number of classes set to 2 in this case, e^{z_i} is the standard exponential function for input vector, and e^{z_j} is the same as the latter one for output vector.

Furthermore, the capsule corresponds to the class with a higher probability 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 before. It is suggested by [67] 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 the proposed PaXNet model. The high-level architecture of the proposed model is illustrated in Fig. 4.7.

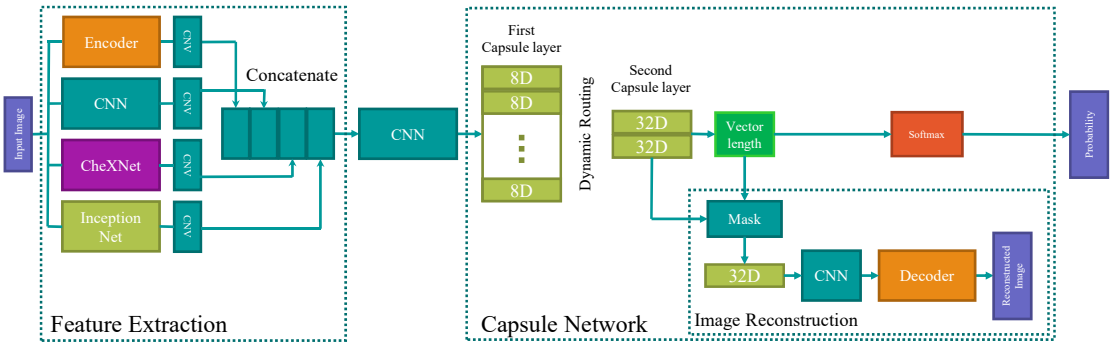


Figure 4.7: Architecture of the proposed PaXNet, consisting of four networks in feature extraction module and a 2-layered capsule network to generate probabilities, connected with a CNN.

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 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 [67].

4.4 Results and Discussions

In this section, 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.

4.4.1 Experimental Results

PaXNet is trained using 319 samples with caries and 1519 healthy samples. In order to deal with the class imbalance problem, the smaller class is re-sampled. It is worth mentioning that datasets of OPG images will always have class imbalance regarding dental caries. The reason is that tooth cavitation is more likely to be detected on molars and pre-molars with a higher probability than the anterior teeth [5]. Thus, even in worst cases, tooth decay could be seen in less than half of the teeth in a full mouth, including in our introduced dataset. While there are several methods to alleviate imbalance, re-sampling is employed in this case in order to increase the overall data size.

After performing re-sampling, 80% of data is used for training and 20% is excluded as for the test set. Moreover, data augmentation is also applied to the samples in the training process. Since the dataset is relatively small, data augmentation with different approaches becomes essential. Each sample in the dataset is rotated randomly and a random zoom and shift are applied as well, as listed in Table 4.2.

Table 4.2: 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 [78]. Fig. 4.8 illustrates the loss versus learning rate during 10 epochs of training while the learning rate is changing exponentially. The best learning rate is where the graph has suddenly decreased, or scientifically-speaking, the largest negative skew value in the chart during changes of the learning rate.

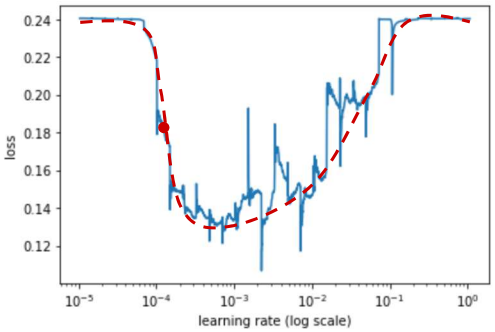


Figure 4.8: loss versus learning rate, and the optimal loss value highlighted in red.

According to Fig. 4.8 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. Tables 4.3 and 4.4 respectively present the statistical results and the confusion matrix of the PaXNet over a dataset of 5948 segmented tooth images

Table 4.3: Statistical performance of the proposed PaXNet

Dataset	**Accuracy**	**Loss**	**Precision**	**Recall**	**F0.5-score**
Training	91.23%	0.13	-	-	-
Test	86.05%	0.15	89.35%	50.73%	0.78

Table 4.4: Confusion matrix of the proposed PaXNet on a test set of 1786 images

n=1786	**Predicted Negative**	**Predicted Positive**
Actually Negative	839	54
Actually Positive	440	453

While the accuracy score is high, considering the relatively small 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. 4.9 shows the above-mentioned map.

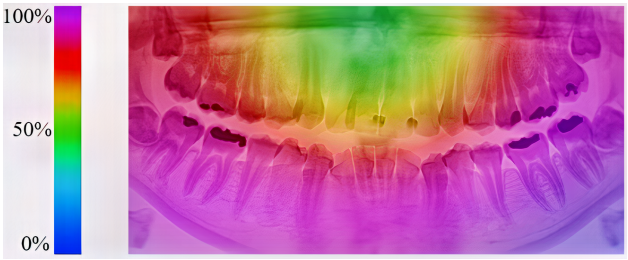


Figure 4.9: Location-based accuracy map of PaXNet. The average accuracy of the model in different parts of the jaw is plotted. Purple is interpreted as having the highest score, while blue is considered the lowest.

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.

4.4.2 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 4.5.

Table 4.5: 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 as well on other samples. Such phenomenon, also known as *"shortcut learning"*, happens when the network finds shortcuts as decision rules to classify input data, and while it performs well on standard benchmarks, it fails to generalize over external testing data [79]. 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) [80]. 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. 4.10 exhibits Grad-CAMs of five samples with caries.

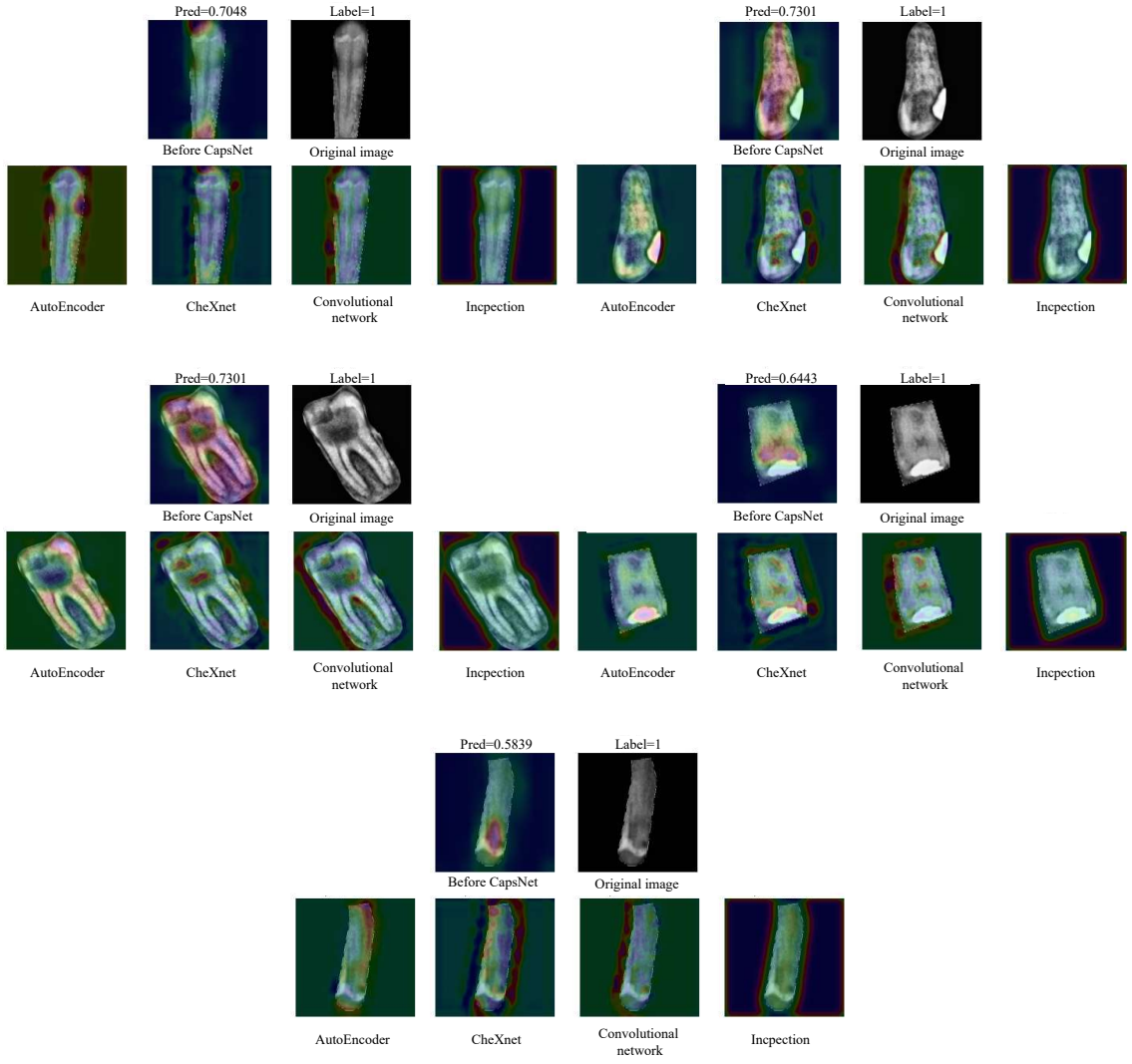


Figure 4.10: Some examples of Grad-CAM after each feature extraction network and before the capsule classifier.

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 4.6 presents the test accuracy of PaXNet with different feature extraction units.

Table 4.6: 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. 4.11 depicts the Grad-CAM of a sample with caries in various transformation.

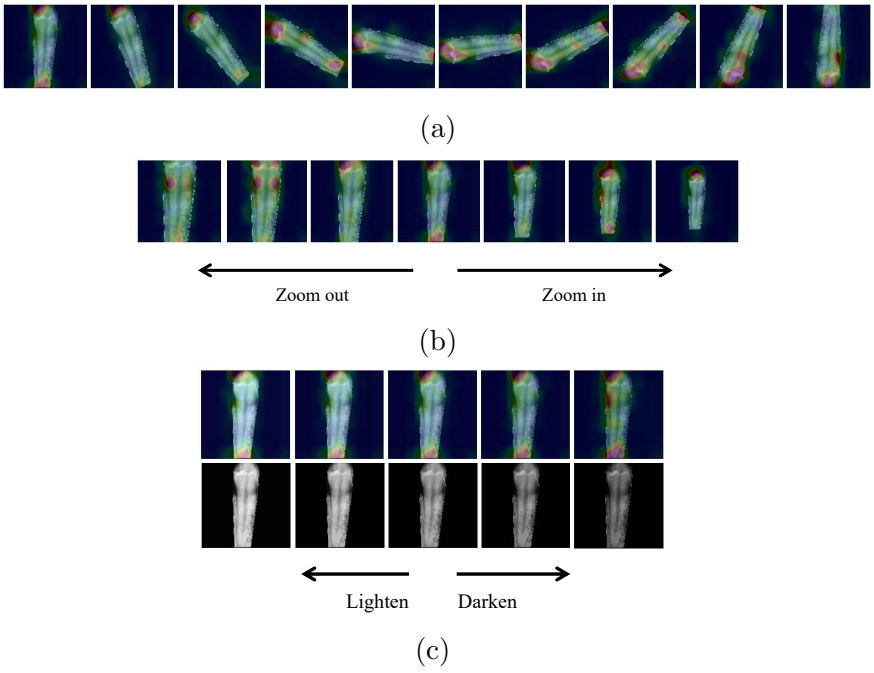


Figure 4.11: The Grad-CAM visualization of an infected sample after transformation with a) rotation, b) zoom, and c) brightness

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 ±30% using a quadratic function. Fig. 4.11 illustrates that by applying various transformations, the proposed network is still classifying the sample accurately using correctly distinguished image features.

4.5 Summary

Dental decision support systems can help dentists by providing high-throughput diagnostic assistance to better detect dental diseases, such as caries. While capturing Panoramic radiography could be very helpful to see full patient dentition in a single image, detection of dental caries using Panoramic radiography is a very challenging task due to the low image quality. There have been a number of studies to detect dental diseases using radiography, but no comprehensive approach has been introduced to implement an end-to-end system to detect dental caries from dental Panoramic x-ray.

To help radiologists with dental caries diagnosis, we propose an automatic caries detection model using several feature extractors and capsule network for feature selection. The feature extraction part is an ensemble of four models, each of which has been built with a different purpose in mind. There are three pretrained models to utilize transfer learning advantages: InceptionNet to bring basic object detection features learned from a very large dataset, CheXNet to bring common patterns mutual between x-ray images, encoder to bring higher-level tooth appearance features learned from unlabeled portion of our dataset. A two-layered capsule network comes after the features are extracted, to select the best combinations based on the input feature vector, with respect to the order of changes in the vector. This order of numbers essentially represents the geometrical relationship between lower-level features. This unique ability of capsule networks helps offering a more robust performance towards dental caries detection.

To improve the explainability of our approach, Grad-CAM visualization is used to validate our extracted features in terms of correct localization. Developed model could successfully reach 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. Besides, PaXNet is also capable of categorizing caries 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 in Panoramic radiography, the achieved results are acceptable, although not outstanding in comparison with previous works on Bitewing/Periapical.

5. Conclusion and Future Works

5.1 Conclusion

With the recent advancements in artificial intelligence, it has dominated the development of computer-aided diagnosis systems for decision support tasks. One of the most applicable fields to develop decision support systems is the field of radiology. Machine learning algorithms, and more specifically deep learning models, have demonstrated remarkable progress in image processing tasks. These models have excelled at automatic detection of diseases in biomedical images. They have even outperformed the domain experts, i.e., radiologists, in certain tasks. Recently, researchers have noticed the capability of machine learning in healthcare, and interests have grown towards developing these techniques specifically for biomedical image processing tasks. Despite their constantly growing interests, there are quite a few number of studies using artificial intelligence techniques for disease classification in dental x-ray images. Due to the sparsity of research studies in the field of dentistry, there is room for improvement.

In this study, we introduced a dataset of dental Panoramic radiography for academic use. Moreover, an end-to-end dental disease diagnosis system is developed based on using genetic algorithm and pure image processing techniques to extract teeth from the raw Panoramic image, as discussed in Chapter 3. Then in Chapter 4, the extracted teeth are fed to an ensemble model utilizing different pretrained models and capsule network to find patterns and give decision on whether each tooth has caries lesion or not. The feature extraction part of the model consists of InceptionNet, CheXNet, a pretrained Encoder on the unlabeled images of teeth in the aforementioned dataset, and a trainable basic convolutional neural network to learn the features from the input dataset. In the feature selection part, a two-

layered capsule network is employed to decide on the presence of caries lesion in each tooth image.

To better evaluate the performance of our proposed model, not only a comparison with previous studies is discussed, but also some robustness tests have been devised to visualize the output of the model and its performance on different subsets of the dataset. Teeth extraction result comparison demonstrates that the achieved accuracy score is comparable with other studies, while they have benefited from higher quality dental radiography, i.e., Bitewing and Periapical x-rays. There have been no previous studies to perform teeth segmentation and caries detection on Panoramic radiography using evolutionary algorithms and deep learning. Regarding the performance of the proposed caries detection model, metric scores are showing a good performance. Location-based accuracy map confirms the model is looking through relevant regions of the Panoramic image. Gradient class activation maps also validate the correctness of the extracted features by visualizing convolution outputs from each network. Finally, dividing the dataset into two parts of those with mild caries and those with severe caries, show that the model performs better on detecting severe caries lesions with a relatively notable margin ($\approx 21\%$). While the results can be enhanced using more data and more complex models, the obtained results demonstrate the applicability of the proposed method. This marks as a step towards the development of an end-to-end decision support system for dental disease detection to perform comparable to a domain expert that can be clinically used in health centers in the future.

5.2 Future works

5.2.1 Dataset

Since a limited number of research studies have been conducted to develop deep learning-based models for dental disease detection, there is a good opportunity to improve the models for better results in the future. The main reason behind the lack of sufficient research, which is also applicable to many other similar approaches in biomedical data processing, is the lack of curated datasets with confident labeling. There are only two public datasets with dental radiography. First one with 120 Periapical x-rays introduced by the authors of [15],

and second one with 120 Bitewing x-rays provided by the organizers of ISBI2015 challenges[1]. Both datasets are for 2015 and very small for being used to train deep neural network.

In this research, efforts have been made to provide the first public dataset of dental Panoramic x-ray images with caries labels. More images with better quality, and from different institutions will provide a better possibility of developing a more sophisticated neural network. With larger datasets, latest architectures, such as EfficientNet-based ones, can also be trained to increase the performance metrics as well as the robustness.

5.2.2 Segmentation

Another critical improvement is to provide images with annotations by radiologists for caries lesions. While there are some other studies with annotation for dental pathologies, no studies have provided caries annotations for Panoramic images so far. Due to the toughness of annotating procedure for dental caries in Panoramics, it needs multiple expert radiologists put a considerable amount of time into meticulously creating masks. By having annotation masks, segmentation models based on U-Net architectures can be applied to segment caries lesions. Segmentation will result in a more explainable model that can be used as an actual clinical decision support system that can outperform radiologists in the future.

5.2.3 Prognosis

Another worth-mentioning point is the fact that tooth decay is a dynamic disease. It starts with a small lesion on the tooth surface and develops dental caries after some time. Caries prognosis requires detection of initial caries lesions before they go through advanced stages. A degree of uncertainty exists when diagnosing initial lesions that have not cavitated dental hard tissues. This is because early lesions may be active (progressing), inactive (arrested), or regressing (remineralizing) [10]. To overcome such a challenge, a patient must go under clinical assessment in specific time intervals. With a database of several x-rays per patient in different time points, it is possible to develop a prognosis model that can identify patients at risk of caries development before the disease process progresses to the point of

[1] `http://www-o.ntust.edu.tw/~cweiwang/ISBI2015/challenge2/`

clinical expression. To do so, a segmentation model can segment lesions in each image and compare the area to classify a lesion into three categories of active, inactive, or regressing. And finally, assign a progression rate for active lesions to anticipate emerging dental caries in the future.

References

[1] S. Corbella, S. Srinivas, and F. Cabitza, "Applications of deep learning in dentistry," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2020.

[2] 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.

[3] 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, 2018.

[4] 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.

[5] A. Ferreira Zandoná, E. Santiago, G. Eckert, B. Katz, S. Pereira de Oliveira, O. Capin, M. Mau, and D. Zero, "The natural history of dental caries lesions: a 4-year observational study," *Journal of dental research*, vol. 91, no. 9, pp. 841–846, 2012.

[6] J. D. Ruby, C. F. Cox, N. Akimoto, N. Meada, and Y. Momoi, "The caries phenomenon: a timeline from witchcraft and superstition to opinions of the 1500s to today's science," *International journal of dentistry*, vol. 2010, 2010.

[7] A. W. Lufkin, *A history of dentistry*. Lea & Febiger, 1948.

[8] M. P. Richards, "A brief review of the archaeological evidence for palaeolithic and neolithic subsistence," *European journal of clinical nutrition*, vol. 56, no. 12, pp. 1270–1278, 2002.

[9] S. C. White and M. J. Pharoah, *Oral radiology-E-Book: Principles and interpretation*. Elsevier Health Sciences, 2014.

[10] D. T. Zero, A. F. Zandona, M. M. Vail, and K. J. Spolnik, "Dental caries and pulpal disease," *Dental Clinics*, vol. 55, no. 1, pp. 29–46, 2011.

[11] 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.

[12] 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.

[13] 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.

[14] R. B. Ali, R. Ejbali, and M. Zaied, "Detection and classification of dental caries in x-ray images using deep neural networks," in *Int. Conf. on Software Engineering Advances (ICSEA)*, 2016, p. 236.

[15] 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.

[16] J. Zhou and M. Abdel-Mottaleb, "A content-based system for human identification based on bitewing dental X-ray images," *Pattern Recognition*, vol. 38, no. 11, pp. 2132–2142, 2005.

[17] A. K. Jain and H. Chen, "Matching of dental X-ray images for human identification," *Pattern Recognition*, vol. 37, no. 7, pp. 1519–1532, 2004.

[18] R. G. Birdal, E. Gumus, A. Sertbas, and I. S. Birdal, "Automated lesion detection in panoramic dental radiographs," *Oral Radiology*, vol. 32, no. 2, pp. 111–118, 2016.

[19] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.

[20] 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.

[21] J. Aps, "Extraoral radiography in pediatric dental practice," in *Imaging in Pediatric Dental Practice*. Springer, 2019, pp. 31–49.

[22] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep Instance Segmentation of Teeth in Panoramic X-Ray Images," in *Proceedings - 31st Conference on Graphics, Patterns and Images, SIBGRAPI 2018*, 2019, pp. 400–407.

[23] M. M. Hasan, W. Ismail, R. Hassan, and A. Yoshitaka, "Automatic segmentation of jaw from panoramic dental X-ray images using GVF snakes," *World Automation Congress Proceedings*, vol. 2016-Octob, no. July, 2016.

[24] P. Moynihan, "Sugars and dental caries: evidence for setting a recommended threshold for intake," *Advances in nutrition*, vol. 7, no. 1, pp. 149–156, 2016.

[25] E. B. Shokouhi, M. Razani, A. Gupta, and N. Tabatabaei, "Comparative study on the detection of early dental caries using thermo-photonic lock-in imaging and optical coherence tomography," *Biomedical optics express*, vol. 9, no. 9, pp. 3983–3997, 2018.

[26] 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.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[29] D. V. Tuzoff, L. N. Tuzova, M. M. Bornstein, A. S. Krasnov, M. A. Kharchenko, S. I. Nikolenko, M. M. Sveshnikov, and G. B. Bednenko, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiology*, vol. 48, no. 4, p. 20180051, 2019.

[30] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[31] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, and H. Fujita, "Classification of teeth in cone-beam ct using deep convolutional neural network," *Computers in biology and medicine*, vol. 80, pp. 24–29, 2017.

[32] T. Hiraiwa, Y. Ariji, M. Fukuda, Y. Kise, K. Nakata, A. Katsumata, H. Fujita, and E. Ariji, "A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography," *Dentomaxillofacial Radiology*, vol. 48, no. 3, p. 20180218, 2019.

[33] S. Vinayahalingam, T. Xi, S. Bergé, T. Maal, and G. de Jong, "Automated detection of third molars and mandibular nerve by deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019.

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

[35] 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.

[36] 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.

[37] T. Ekert, J. Krois, L. Meinhold, K. Elhennawy, R. Emara, T. Golla, and F. Schwendicke,

"Deep learning for the radiographic detection of apical lesions," *Journal of endodontics*, vol. 45, no. 7, pp. 917–922, 2019.

[38] J. Krois, T. Ekert, L. Meinhold, T. Golla, B. Kharbot, A. Wittemeier, C. Dörfer, and F. Schwendicke, "Deep learning for the radiographic detection of periodontal bone loss," *Scientific reports*, vol. 9, no. 1, pp. 1–6, 2019.

[39] M. Murata, Y. Ariji, Y. Ohashi, T. Kawai, M. Fukuda, T. Funakoshi, Y. Kise, M. Nozawa, A. Katsumata, H. Fujita *et al.*, "Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography," *Oral radiology*, vol. 35, no. 3, pp. 301–307, 2019.

[40] D. W. Kim, S. Lee, S. Kwon, W. Nam, I.-H. Cha, and H. J. Kim, "Deep learning-based survival prediction of oral cancer patients," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[41] V. Rushton and K. Horner, "The use of panoramic radiology in dental practice," *Journal of dentistry*, vol. 24, no. 3, pp. 185–201, 1996.

[42] Y. Lai, F. Fan, Q. Wu, W. Ke, P. Liao, Z. Deng, H. Chen, and Y. Zhang, "Lcanet: Learnable connected attention network for human identification using dental images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 905–915, 2020.

[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] P.-L. Lin, P.-W. Huang, Y. Cho, and C.-H. Kuo, "An automatic and effective tooth isolation method for dental radiographs," *Opto-Electronics Review*, vol. 21, no. 1, pp. 126–136, 2013.

[45] A. Wirtz, S. G. Mirashi, and S. Wesarg, "Automatic teeth segmentation in panoramic x-ray images using a coupled shape model in combination with a neural network," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 712–719.

[46] P. Pandey, A. Bhan, M. K. Dutta, and C. M. Travieso, "Automatic image processing based dental image analysis using automatic gaussian fitting energy and level sets," in *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. IEEE, 2017, pp. 1–5.

[47] 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.

[48] A. Yuniarti, A. S. Nugroho, B. Amaliah, and A. Z. Arifin, "Classification and numbering of dental radiographs for an automated human identification system," *Telkomnika*, vol. 10, no. 1, p. 137, 2012.

[49] S. Dighe and R. Shriram, "Preprocessing, segmentation and matching of dental radiographs used in dental biometrics," *International Journal of Science and Applied Information Technology*, vol. 1, no. 2, 2012.

[50] G. Dibeh, A. Hilal, and J. Charara, "A novel approach for dental panoramic radiograph segmentation," in *2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*. IEEE, 2018, pp. 1–6.

[51] 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.

[52] 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.

[53] 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.

[54] K. Moutselos, E. Berdouses, C. Oulis, and I. Maglogiannis, "Recognizing occlusal caries in dental intraoral images using deep learning," in *2019 41st Annual International Con-*

ference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 1617–1620.

[55] 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.

[56] 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*, vol. 131, no. 6, pp. 711–720, 2021.

[57] 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.

[58] J. Oliveira and H. Proença, "Caries detection in panoramic dental x-ray images," in *Computational Vision and Medical Image Processing*. Springer, 2011, pp. 175–190.

[59] J. Holland, "Adaptation in natural and artificial systems, univ. of mich. press," *Ann Arbor*, 1975.

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

[61] A. F. Sheta and K. De Jong, "Parameter estimation of nonlinear systems in noisy environments using genetic algorithms," in *Proceedings of the 1996 IEEE international symposium on intelligent control*. IEEE, 1996, pp. 360–365.

[62] 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.

[63] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern recognition*, vol. 33, no. 2, pp. 225–236, 2000.

[64] B. Bataineh, S. N. Abdullah, K. Omar, and M. Faidzul, "Adaptive thresholding methods for documents image binarization," in *Mexican Conference on Pattern Recognition*. Springer, 2011, pp. 230–239.

[65] W. Mustafa, A. S. Abdul-Nasir, and Z. Mohamed, "Malaria parasites segmentation based on sauvola algorithm modification," *Malaysian Appl. Biol.*, vol. 47, no. 2, pp. 71–76, 2018.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

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

[73] 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.

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

[75] 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.

[76] 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.

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

[78] 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.

[79] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut learning in deep neural networks," *Nature Machine Intelligence*, vol. 2, no. 11, pp. 665–673, 2020.

[80] 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.