東海大学大学院令和４年度博士論文

Study on Caries Detection from Dental Panoramic Radiographs using Deep Learning

(深層学習を用いた歯科パノラマレントゲン画像からのう蝕歯検出に関する研究)

指導　　　　濱本和彦 教授

東海大学大学院総合理工学研究科

総合理工学専攻

BUI HUY TOAN

THESIS TITLE   Study on Caries Detection from Dental Panoramic
Radiographs using Deep Learning

STUDENT NAME   BUI Huy Toan

STUDENT ID   9LTAD006

DEGREE   Doctor of Engineering

PROGRAMME   Course of Science and Technology

ADVISOR   Dr. Kazuhiko HAMAMOTO (Professor)

Abstract

Caries is one of the most well-known diseases that affect billions of lives around the world. Most people experienced at least one or several symptoms such as tooth pain, halitosis, and eventually result in tooth removal. There are many methods to evaluate damages from caries, and among them, dentists prefer staging to C0, C1, C2, C3, C4. While C0 presents a healthy tooth, C1 to C4 presents different stages of caries which lead to different conditions and treatments. Caries treatment is mostly expensive without an early diagnosis and therefore leads to financial burdens; especially in the under-developing country where usually lacks people's awareness of dental hygrines and experienced dentists. As a result, the purpose of this study is to propose a method that provides early caries detection to support a dentist.

Computer-assisted diagnosis (CAD) is a branch of artificial intelligence and computer science. The purpose of CAD is to enable electronic devices to observe, interpret relevant data, and produce a suitable output that can assist humans in making decisions. Machines have improved in terms of convenience and power as technology advances. Therefore, machines may be able to take over a variety of functions previously performed by people. Nowadays, many computer-vision studies prefer to employ image processing techniques and machine learning algorithms to solve different problems, since the combinations between them can partially reduce the complexity by learning without being programmed to perform a specific task but still providing an effective result based on requirements. In machine learning, neural networks are well-known techniques which are widely used to solve pattern recognition problems. Neural networks are simple to construct and often appear to be good ability to generalize and respond to unexpected patterns. For that season, this study applies the image processing and deep learning based on neural network as the main methods in CAD.

Despite the fact that some researchers have sought to remedy this issue. The majority of them are either very complex, needing a huge load of resources and incurring a significant computational burden, or overly simplistic and incapable of producing an adequate outcome. Therefore, my proposed approach streamlined the most complex processes while keeping the benefits of each component. The CAD system mainly focuses on two processes: tooth isolation and caries detection. Firstly, in tooth isolation, the position of each tooth is detected in a panoramic radiograph using

the YOLOv3 model. The original oral images will be resized and fed into the YOLOv3 to split out to several individual tooth images which is useful for centralized diagnosis later. The process can help reducing a huge workload of dentists and nurses, and play an important role in speeding up the diagnosis. On the other hand, the manual isolation of tooth images is fed to a deep convolutional neural network for caries diagnosis. The useful features from the images are extracted using geometry and pre-trained deep learning model, such as Resnet50, Xception, VGG16, ...etc., and fed into other machine learning models such as decision trees, naive Bayes, k-nearest neighbor, and support vector machine for the final diagnosis. On tooth isolation, the final result shows 95.58%, 94.90% for precision and recall, respectively. And on caries detection, the final result shows 91.70%, 90.43%, and 92.67% for accuracy, sensitivity, and specificity. The results are better, or equivalent compared with the previous researches. Finally, the caries detection system is also tested on the automatic isolation tooth for a comprehensive assessment. The final results slightly reduce compared to caries detection using manual isolation and reach to 88.66%, 88.14%, and 89.47% for accuracy, sensitivity, and specificity, respectively. However, an automatic caries detection system which can conduct tooth isolation and caries detection has not been proposed and the reductions are small; therefore, the proposal and evaluation of the automatic caries detection system is important and contribute to the field of dental diagnosis.

The research's objectives were well accomplished in terms of tooth isolation and caries detection. The tooth was found automatically in the oral panoramic radiography utilizing an automated isolation tooth technique. Precision demonstrates a high level of detection accuracy; as a result, the procedure is advantageous and practical for a dentist. All facets of the matter seem to have been considered. However, the method may be greatly enhanced. The outcome of manual tooth isolation has surpassed the prior state-of-the-art in caries detection. Increased specificity demonstrates that the dataset and its results are balanced, consistent, and dependable. On the other side, the automated combination technique results in a modest decrease in total caries detection. However, this automatic system is a new proposal in the field of dental diagnosis and the reduction is small; therefore, the system can contribute to the field of dental diagnosis. The weak may be identified by the automated isolation tooth system's relativity. As a result, it could be recognized as a benefit to get a further understanding of the disease's diagnosis in general.

Acknowledgment

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Research Introduction

1.1 Research Background

1.1.1 Introduction to Oral Health

Oral health is vital in determining the overall health and quality of life. Generally speaking, it refers to a state of good health free of a variety of illnesses and ailments, including dental caries, periodontal gum disease, dental cavities, tooth loss, HIV infection of the mouth, oral cancer, and Noma, in addition to congenital deformities such as cleft lip and palate, among others. The majority of illnesses and disorders have modifiable risk factors with the major noncommunicable diseases, which is why they are classified as such (cancer, cardiovascular diseases, chronic respiratory disease, dental caries, and diabetes). These risk factors include excessive alcohol consumption, cigarette smoking, and poor diets high in free sugar, all of which are on the rise across the world. A relationship has been shown between dental health and overall health [1], according to research. For example, diabetes mellitus has been shown to be associated with the progression and spread of periodontitis. A causal relationship exists between sugar intake and diabetes, obesity, and tooth decay as well,

according to the research. Increases in oral diseases are continuing to be seen in the majority of emerging countries, as living conditions improve and urbanization grows more prevalent. The major reason is an insufficient supply of fluoride (in dental hygiene products such as toothpaste and drinking water) and a costly barrier to community-based oral health treatment. Additionally, as a consequence of increasing marketing of sugary foods and drinks, as well as alcohol and cigarettes, there is an increase in the use of goods that lead to oral health problems and other noncommunicable disorders.

1.1.2 Introduction to Dental Caries

Figure 1.1 Caries on teeth (image credit [2]).

Dental caries (often referred to as tooth decay or dental cavity) is a source of preoccupation for many individuals. Due to caries, most people have experienced a

toothache at least once or multiple times during their lives. Sugar is the primary cause of caries. It is added to food during production, cooking, and consumption, as well as the sugar naturally found in foods such as fruit, bread, and rice. When teeth are healthy, they chop and smash food into little bits to aid in the digesting process in the stomach and intestine. Dental caries occurs when food plaque forms on the surface of the tooth and converts sugar to acid, which corrodes the tooth over time. The earliest sign of caries is a little black (or gray) hole on the tooth's surface (Figure 1.1). However, in certain circumstances, caries develops on the inner surface of the tooth or in the gap between teeth, causing no evident symptoms. Thus, early identification of caries is critical for minimizing tooth affection. Without prompt treatment, caries may cause severe discomfort and in extreme cases, tooth loss.

1.1.3 Dental Care in the Health Care System

According to the World Health Organization (WHO), oral disorders affect over 3.5 billion people globally [3]. Specifically, cavity and lip cancers account for around 180,000 fatalities each year and are consistently ranked among the top 15 most prevalent malignancies globally. When it comes to dental cavities, the majority of instances are left untreated in permanent teeth. Nearly 10% of the population is estimated to have severe periodontal gum disease, and around 2.3 billion persons are estimated to have dental caries of permanent teeth. Additionally, children are considered to be the most vulnerable. Dental caries affects about half of all children globally, with around 530 million children suffering from primary tooth caries. The majority of illnesses are preventable and readily treatable in their early stages. However, therapy is often not covered by universal health insurance. Dental care is often prohibitively costly, accounting for 20% of people's out-of-pocket expenses in

many affluent nations. Dental health service demand exceeds the capacity of the health care system in the majority of developing countries, and is thus unaffordable to the majority of low- and middle-income individuals. To be specific, the direct cost of caries reached around US$ 298 billion worldwide in 2010. Additionally, indirect costs equal around US$ 144 billion and may reach US$ 442 billion in total. Consequently, untreated caries impairs academic and occupational performance and may result in personal or social difficulties. Psychosocial devastation results in a slew of associated difficulties and significantly reduces one's quality of life.

1.1.4   Category of Dental Caries

There are various forms of dental caries [4]. They include caries of the enamel, reversible caries, early childhood, smooth surface, pit and fissure, acute dental caries, and primary and secondary caries. An illustration will be explained in further detail as in Figure 1.2:
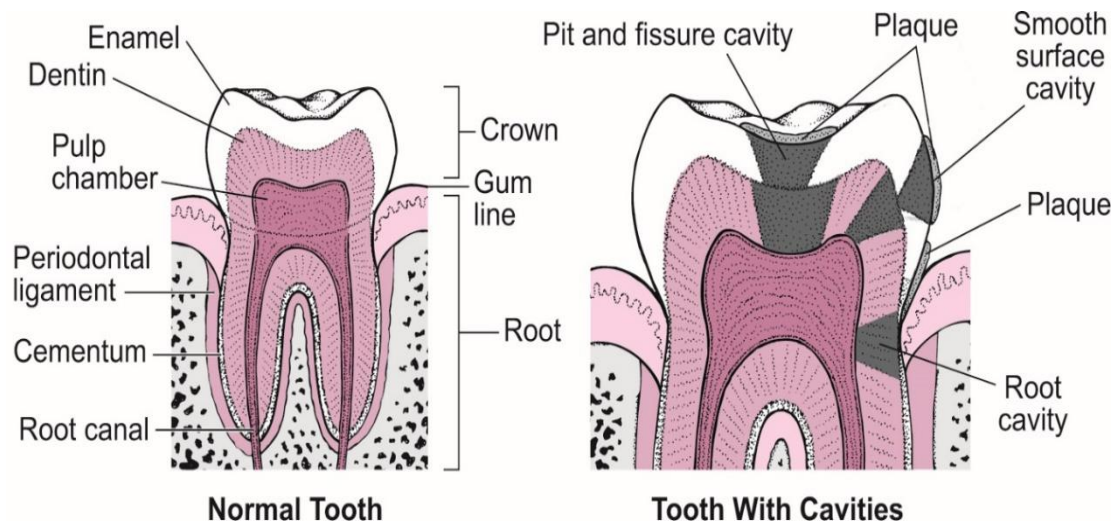
Figure 1.2 Healthy and caries types tooth (image credit [2]). Structure of healthy tooth and a tooth with decay.

1. Caries of the enamel: When bacteria consume food plaques, they produce acids that erode the tooth enamel. Bacteria get energy from sugars found in food and generate lactic acid. Acid lactic acid is one of the components in a chain reaction that demineralizes enamel crystals. Corrosion will continue to eat away at enamel until bacteria reach the surface of the dentin.

2. Smooth surface cavity (Damage to the tooth's side or perimeter): Smooth-surface cavities are more readily addressed due to their moderate growth rate. Often, they will resolve on their own with the assistance of fluoride therapies like dental gels, toothpaste, varnish, or fluoride-enriched water. Cavities often take a long time to propagate through smooth-surface enamel. However, if this occurs, a filling will be required.

3. Root cavity: A root cavity may grow at the base of the tooth, and it is often positioned on the unobservable side. Unlike the crown, which is covered with a tough, protective enamel, the root surface is uncovered and vulnerable. When the gums recede below the enamel line or when they peel away from a tooth, the root surface of the tooth becomes visible. Tooth decay is more prone to develop in an exposed tooth due to its weaker and more brittle structure. Besides, there are several caries types that are usually not listed such as:

4. Acute caries: This condition affects a large number of teeth fast. In contrast to other kinds of caries, lesions in this type of caries are often grey or light brown in color. Due to the poor quality of these caries, it is difficult to locate afflicted teeth. Acute caries is often associated with sensitive teeth and pulp exposure.

5. Caries in early childhood is often caused in tiny infants by the newborn being administered milk at night. For a lengthy period of time, a little amount of milk remains in the mouth, ferments, and produces bacteria nests between the teeth.

Caries in early infancy causes gum inflammation, teeth discoloration, and finally acute pain, impairing communication, eating, and sleeping. The earliest signs appear on the two upper front teeth. The most common prevention technique is to avoid feeding the newborn at night, even if teeth begin to appear.

1.1.5   Stages of Dental Caries

To identify caries, it is vital to know the progression of the disease in the tooth. Caries phases must be accurately identified as a result. There are several methods to explain the stages of caries in practice. The development of sickness is a well-known scale as described in Figure 1.3.
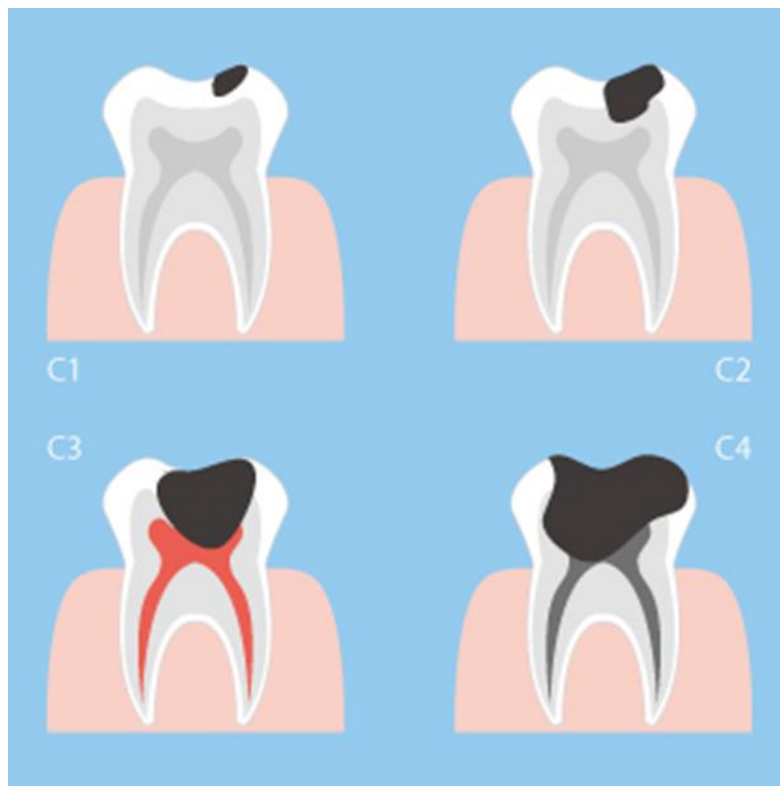
Figure 1.3 Stages of caries, C1 to C4 (image credit [5]).

C0 - At this point, there is no visible damage to the surface of the tooth. There may be some white patches or coloring at this point, so be cautious. In contrast, stains may be eliminated and teeth restored with the proper dental care method.

C1 - This is the initial stage at which the enamel of a tooth is impacted by caries, and it is designated C1 (minor caries). Patients do not experience anything like as pain, discomfort, or annoyance as there are no nerves in the enamel layer (pain, discomfort, or inconvenience). Enamel degradation that results in black spots is irreversible. Caries C1 patients should consult a doctor as soon as possible so that therapy may begin promptly.

C2 - Dentin is now compromised, causing this stage of the sickness C2 (moderate caries). Teeth that have been exposed to the dentin become sensitive. Some persons get a tingling sense in their teeth when they eat or drink anything cold. Accordingly, patients should seek prompt dental treatment so that the affected region may be properly cleaned and the fissure may be permanently sealed.

C3 - at this stage, both enamel and dentin have been perforated and pulp drilled. The infection has now migrated to the tooth core, which contains several nerves and blood arteries. Inflammation and affection are very harmful and cause agonizing pain. If the bacteria penetrate further into the tissue, the diseased area will be suppurated, the face will enlarge, the mouth will smell, and the pulp will be destroyed. In this case, the dying pulp of the tooth must be removed.

C4 (severe caries which is the last phrase) - at this point, just the tooth's root remains; the tooth's whole crown has been destroyed. The body may easily get infected with various bacteria through the death pulp. Tooth is mostly destroyed or showed in black. To prevent infection, the remaining root of the tooth must be extracted and replaced with an implant, bridge, or denture.

1.1.6 Importance of Early Treatment in Dental Caries

Adult patients' teeth are permanent, which implies that they cannot be restored to their pre-existing shape or size. When an acid-damaged tooth is pulled during dental treatment, the tooth's ability to regenerate will be severely limited. As a consequence, caries should be recognized as soon as possible in order to prevent tooth damage from occurring. The earlier caries is discovered, the less probable it is that the tooth will need to be removed. As a result, early detection of caries is crucial for rapid treatment, which results in time, effort, and money savings for both the patient and the dentist. Alternatively, late detection may result in permanent damage, requiring the extraction of a tooth or many teeth. The identification of caries is not difficult, despite the fact that each patient's treatment must be adapted to their specific needs. In light of the fact that, the vast majority cases of caries present symptoms in the form of discoloration, fractured shape, or a hole in the tooth, the practical experience for identifying whether or not a tooth is affected seems to be the same. As people's need for early diagnosis grows, the problem of unskilled dentists, especially in developing countries, becomes even more acute. Because of this, a strong, stable, and scalable solution is necessary.

1.1.7 Computer-Aid Diagnosis (CAD) for Caries

With the advancement of medical imaging technology in recent years, computer-aided diagnostic systems (CADs) have become critical in the early identification of a variety of disorders, including cancer, diabetes, and even caries [6, 7]. Caries may be discovered in a variety of ways and approaches. Several studies recommended that photoacoustic pictures, wavelengths, or ultrasound images be used for detection [8-10]. Other study has described a strategy using an RGB oral endoscope

picture [11, 12]; however, most systems lack the ability to see the detailed structure of the tooth, particularly the tooth root, and so fail to diagnose caries. Besides, dental radiograph is an simple and inexpensive imaging technique which can be conducted at most dental office or hospital; while other imaging technique such as CT-radiograph or near infrared range may be more expensive and not popular [13]. As a result, the dental radiographic image is the most often used method and is preferred for the early identification of caries using CAD.

1.2 Objectives

The purpose of this study is to develop a computer-aided diagnostic system that will assist dentists in the caries detection process by meeting the following criteria:

1. Automatic extraction of a single tooth from a panoramic radiograph - this contribution helps save time and effort for the nurse and dentist when it comes to tooth isolation, which is often time-consuming and exhausting for a human being. The tooth's position should be indicated, along with the bounding box that may be used to slice the tooth for subsequent diagnosis.

2. Caries detection: caries teeth should be identified among normal teeth. This contribution improves the accuracy of the dentist's diagnosis and helps in training new dentists. Additionally, the system might give an early diagnosis for people living in faraway places without access to medical professionals or facilities.

By archiving those mentioned targets, the implement of this study can help dentist in two tasks: caries patients screening and caries diagnosis. While the screening

task can remove a large of certain non-caries patients, the diagnosis task can help in consolidating the dentist's final diagnosis.

1.3   Thesis Structure

There are six sections to this dissertation: an introduction, background information, tooth localization, caries detection, automated system combination, and a conclusion and suggestions for further research. The first chapter examines the broad issue of oral health and the relevance of performing this study. The remainder is as follows:

Chapter 2 contains the foundational knowledge of this study. This may assist readers in gaining a better idea of the suggested strategy prior to delving further in the proposed method.

Chapter 3 discusses the approach for tooth localization using the Yolov3 model. The procedure begins with the original panoramic radiograph as its input. The result is a bounding box that positions each individual tooth, thereby segmenting it for further diagnosis.

Chapter 4 presented a classification system for caries and non-caries teeth. Each tooth, which is divided manually from panoramic radiographs dataset, was diagnosed in this section. For assessment purposes, the final outcome was compared to related studies.

Chapter 5 synthesizes the methods given in chapters three and four. This chapter's outcome is compared to chapter four in order to provide a thorough evaluation.

Chapter 6 summarizes the whole study, provides the author's perspective, evaluates the research's strengths and weaknesses, and concludes with a strategy for future work.

CHAPTER 2

Background Knowledge

Before detailing the explanation of the following chapter, I present the necessary background information. First, the first section provides a broad overview of computer vision and image processing generally. Then, the second section discusses neural networks and the system's unique topologies.

2.1 Computer Vision and Image Processing

Image processing is a subfield of computer science concerned with the analysis and manipulation of digital images [14]; also known as a kind of signal processing that takes a photograph as an input and outputs it in a number of formats and qualities. The purpose of image processing is to improve the quality of a raw image (by smoothing, sharpening, and contrasting it, for illustration) by the use of computer software.

By decomposing the raw picture into a collection of the most relevant data, image processing methods, such as feature extraction, are utilized to tackle the challenge of describing an image [15]. When an image is acquired, a feature set (alternatively called a feature vector) is formed. This feature set comprises a collection of values that define the image's unique characteristics. Due to the attributes retrieved from the picture, our perception of the raw image may be significantly affected. In

practice, the great majority of picture data is often redundant; just a few critical characteristics need to be collected in order to make demand-driven judgments regarding the image. There are characteristics with varying degrees of complexity. The intensity level, the form of the item, and the object's size are all well-known low-level features. Semantic features of an object are high-level properties that express the notion of the thing in human terms. Because the use of features varies significantly by purpose, no one characteristic can guarantee the optimum recognition outcomes. Color is efficient in differentiating between night and day sceneries, for example, but circularity is more successful at differentiating between a ball and a tree picture. In classification and recognition problems, the approach of feature extraction (i.e., the process of creating a collection of features) becomes more significant.

2.2   Machine Learning

Machine learning is a study of computer algorithms that may improve automatically via experience and the usage of data that is referred to as machine learning (ML)[16]. Machine learning algorithms create a model based on sample data, known as training data, in order to make predictions or choices without being explicitly coded [17]. Computer vision, medicine, email filtering, and voice recognition are just a few of the fields where machine learning algorithms are being utilized to solve problems where it is difficult or impossible to create traditional algorithms to accomplish the required tasks [18].

Statistical learning is a subset of machine learning that is closely connected to computational statistics, which is concerned with generating predictions using computers; nevertheless, statistical learning is not all of machine learning. The

discipline of machine learning benefits from the study of mathematical optimization since it provides tools, theory, and application fields. Data mining is a closely connected topic of research that focuses on exploratory data analysis utilizing unsupervised learning techniques to solve problems [19]. Some machine learning implementations make use of data and neural networks in a manner that is intended to simulate the functioning of a biological brain. Machine learning is referred to as predictive analytics when it is used to solve business challenges across a variety of industries.

In deep learning, each level learns to turn its input data into a representation that is gradually more abstract and composite. For example, of face recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face. It is important to note that a deep learning process may perform learning itself which characteristics should be ideally placed in which levels. It is still necessary to fine-tune the results manually; for example, adjusting the number of layers and the size of the layers might yield varying levels of abstraction [20, 21].

The term "deep" in the phrase "deep learning" refers to the number of layers that the data is changed through. Deep learning systems have a significant depth of credit assignment path (CAP). The CAP is the sequence of transformations that takes place from input to output. CAPs are used to represent links between input and output that might be considered possibly causative. In a feedforward neural network, the depth of the CAPs is the same as the depth of the network plus one (as the output layer is also parameterized). The CAP depth is theoretically limitless in recurrent neural networks,

in which a signal may travel through a layer more than once [22]. Although there is no commonly accepted depth threshold that distinguishes shallow learning from deep learning, the majority of scholars believe that deep learning requires CAP depths greater than 2. It has been shown that CAP of depth 2 is a universal approximator in the sense that it can simulate any function. More layers, however, do not improve the network's capacity to approximate functions in any way. Deep models (CAP > 2) are able to extract better features than shallow models, and as a result, the addition of more layers aids in the successful learning of the features.

2.2.1   Decision Tree

Decision trees (DTs) are a subclass of supervised machine learning techniques that recreate the tree-like structure of data by segmenting it according to a specified parameter defined as an attribute's "test" [23]. The root, nodes, and leaves are the three fundamental components of the tree. The root node denotes the start of the dataset as a whole. Each node has an attribute test that categorizes the data as sub-nodes or leaves. The final outcome that dictates a path of action is the leaf.

Finally, and probably most significantly, the bulk of complicated, specialized situations are impossible of being grasped via the use of a simple linear model. They need a non-linear approach to convey topics in a unified fashion. As a result, the DT structure becomes substantially larger and branches to a large number of nodes, often leading to overfitting. To avoid overfitting, a constraint on the number of branching nodes and leaves is given under specific parameters, including 1) the minimum number of training instances in each leaf; and 2) the tree's depth, defined as the maximum number of nodes from the root to a leaf.

Figure 2.1 describes an example of fruits classification using a decision tree model. In the example, the model classifies three types of fruits (apple, cucumber, and banana) using the shape (round or not round) and color (yellow or not yellow) features. In the first round, by using the shape if it is round, the apple has been defined when the answer comes to yes. Otherwise, that should be other fruits, such as banana or cucumber, when the answer comes to no. In the second round, by using the color if it is yellow, banana is defined when the answer comes to yes. Otherwise, cucumber is defined when answer comes to no.
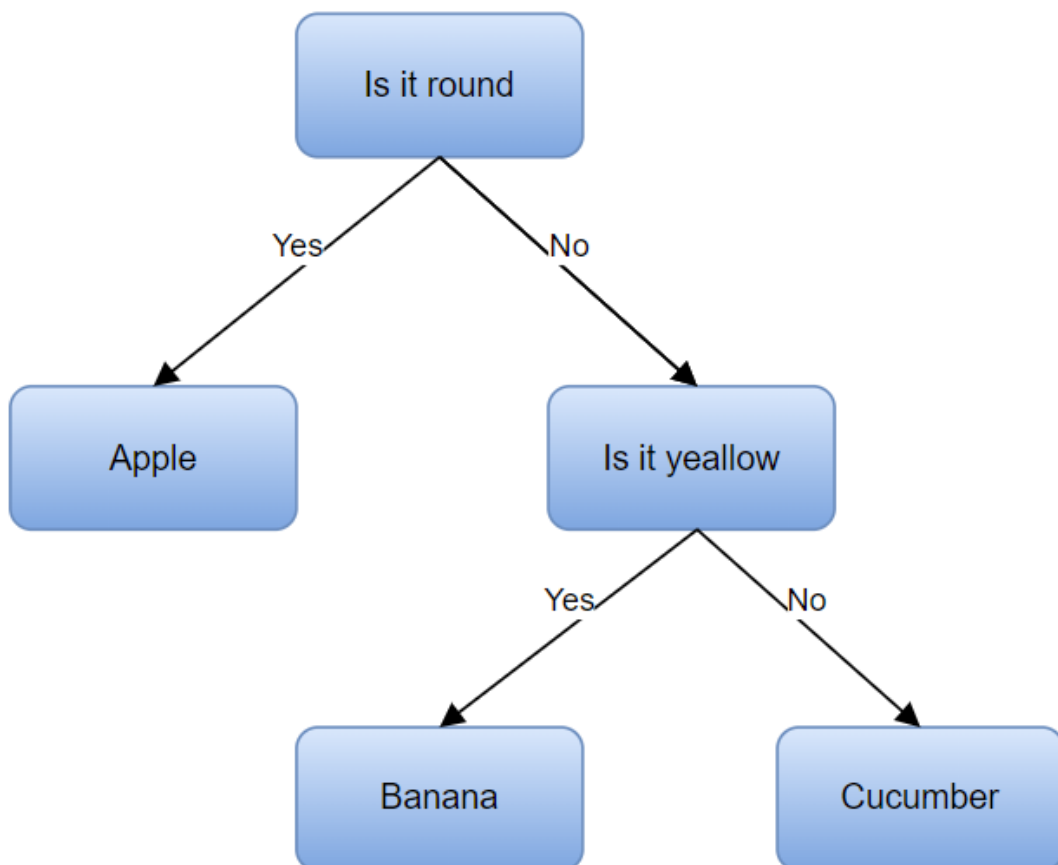
Figure 2.1 Examples of fruits classification using decision tree.

2.2.2 K-nearest Neighbor

K-nearest neighbor (KNN) is a well-known statistic-based machine learning technique that was invented in 1951 by Evelyn Fix and Joseph Hodges [24], and subsequently extended by Thomas Cover [25]. KNN is used to tackle problems involving regression and classification. In both circumstances, the training sample's input is a set of k-values representing the number of nearest samples, and the output is dependent on whether k-NN is used for classification or regression.
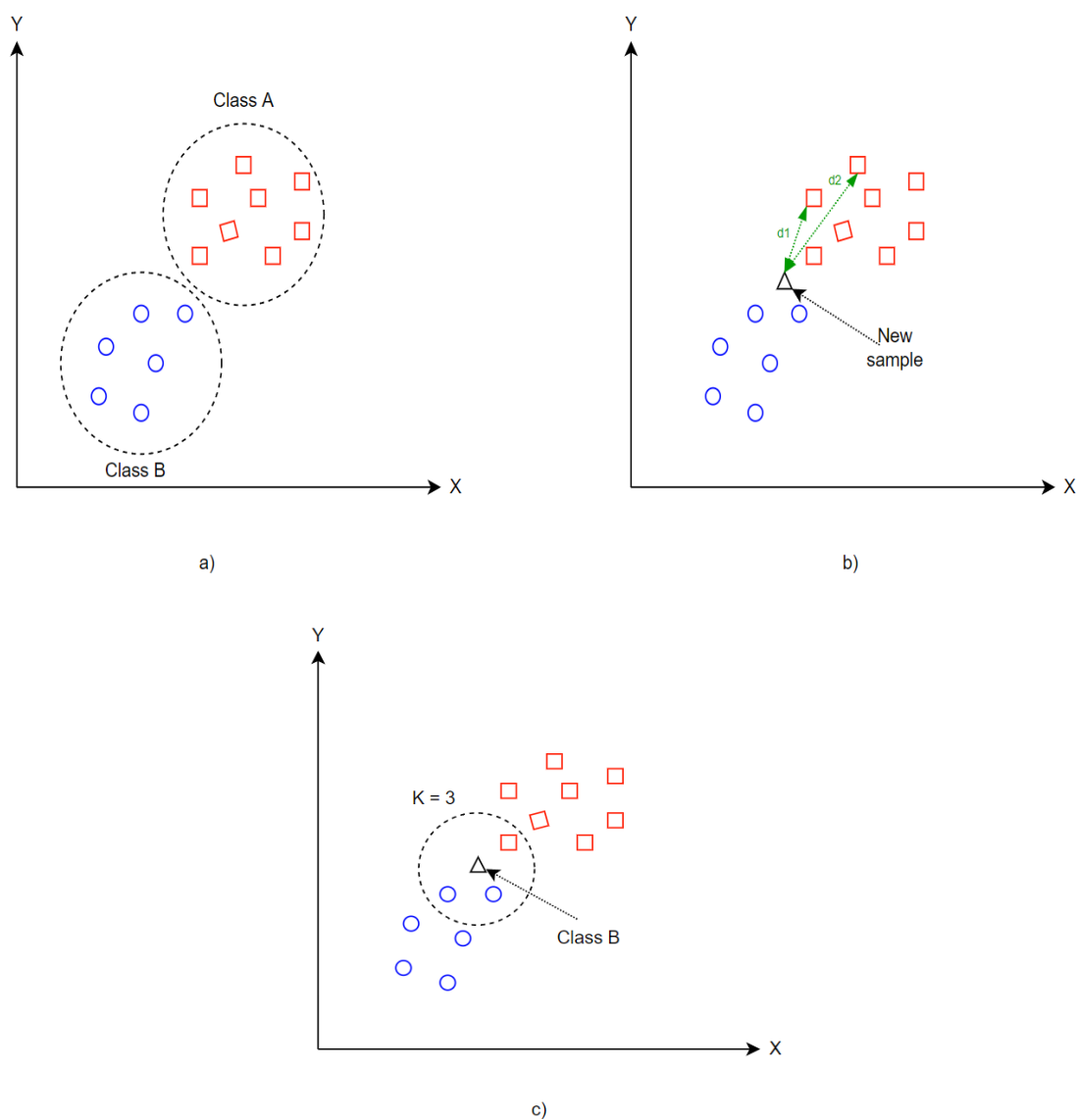
Figure 2.2 Example of KNN classification. a) two known-classes b) adding new sample c) assign the new sample to known-class with k = 3.

Class membership is the result of k-NN categorization. A majority vote of its neighbors classifies an item, with the object being assigned to the most common class among its k closest neighbors (k is a positive integer, typically small). If k equals 1, the item is simply allocated to the class of its closest neighbor.

The value of an object's attribute is returned as the result of a k-NN regression analysis. There are k-nearest neighbors, and this number is the average of their values. When using the k-NN model with the parameter k = 3, the procedure depicted in Figure 2.2 is as follows: As a result, three samples that are closest to the new unknown datapoint are defined. Distance metrics such as Euclidean distance, Manhattan distance, and Chebyshev distance are used to measure the length of a path between an unknown sample and one or more nearby neighbors. The formula of distance metric is shown in formulas (2.1),(2.2), and (2.3).

$$Euclidean\ distance = \sqrt{\sum (x_i - y_i)^2} \qquad (2.1)$$

$$Manhattan\ distance = \sum |(x_i - y_i)| \qquad (2.2)$$

$$Chebyshev\ distance = \max_i |x_i - y_i| \qquad (2.3)$$

Following the calculation of the k number of nearest neighbors, I assign the class of the new unknown point based on the class of the majority of neighbors in the area around the point. Two of the nearest neighbors belong to class B, while one of the nearest neighbors belong to class A, in this case. As a result, class B is assigned to the newly discovered unknown location.

18

2.2.3 Naïve Bayes

Naive Bayes (NB) [26] is one of the well-known probabilistic classifier machine learning techniques based on Bayes' theorem with an independence assumption between features. In basic words, a Naive Bayes classifier asserts that the existence of one feature in a class is independent to the presence of any other feature. The Naive Bayes model is easy to create and is particularly suitable for enormous data sets, and also is recognized to beat even the most powerful classification systems owing to its simplicity.

Bayes theorem aims to compute a posterior probability $P(y|x)$ from $P(y)$, $P(x)$, and $P(x|y)$ as described in Equation (2.4) and (2.5):

$$P(Y_k|x) = \frac{P(x|Y_k)P(Y_k)}{P(x)} \tag{2.4}$$

$$P(Y_k|X) = P(x_1|Y_k) \times P(x_1|Y_k) \times \ldots \times P(x_n|Y) \times P(Y_k) \tag{2.5}$$

where:

$P(y|x)$ is the posterior probability of class c given predictor x.

$P(Y)$ is the prior probability of class.

$P(x|y)$ is the likelihood which is the probability of predictor given class.

$P(y|x)$ is the prior probability of predictor.

2.2.4 Random Forest

In 1995, Tim Kan Ho [27] developed an ensemble learning system for regression, classification, and other tasks that he named random forest (also known as random decision forests). RF is an acronym for Random Forest, which is an

abbreviation for Random Decision Forest. In order to address Eugene Kleinberg's "stochastic discrimination" approach [28-30], Ho devised a technique. However, whereas in classification problems, the output of a random forest is a class that has been selected by the majority of trees, regression tasks provide the mean or average forecast of the individual trees. If overfitting from the training set is avoided, then random choice forests outperform ordinary decision trees in terms of performance. Random forests consistently beat choice trees in the vast majority of situations, despite the fact that data attributes may have an impact on the ultimate performance of the model. Figure 2.3 depicts the process by which random forests generate a result based on the majority vote of tree outcomes.

For further information, in 2006, Leo Breiman and Adele Cutler submitted a "trademark" application for the expansion of random forests. Using Ho's random selection features in conjunction with Breiman's bagging notion, the goal is to generate a collection of decision trees with a controlled variance. Finally, random forests are generally used as a black box technique since they may frequently provide fantastic results with a small amount of input data and configuration.
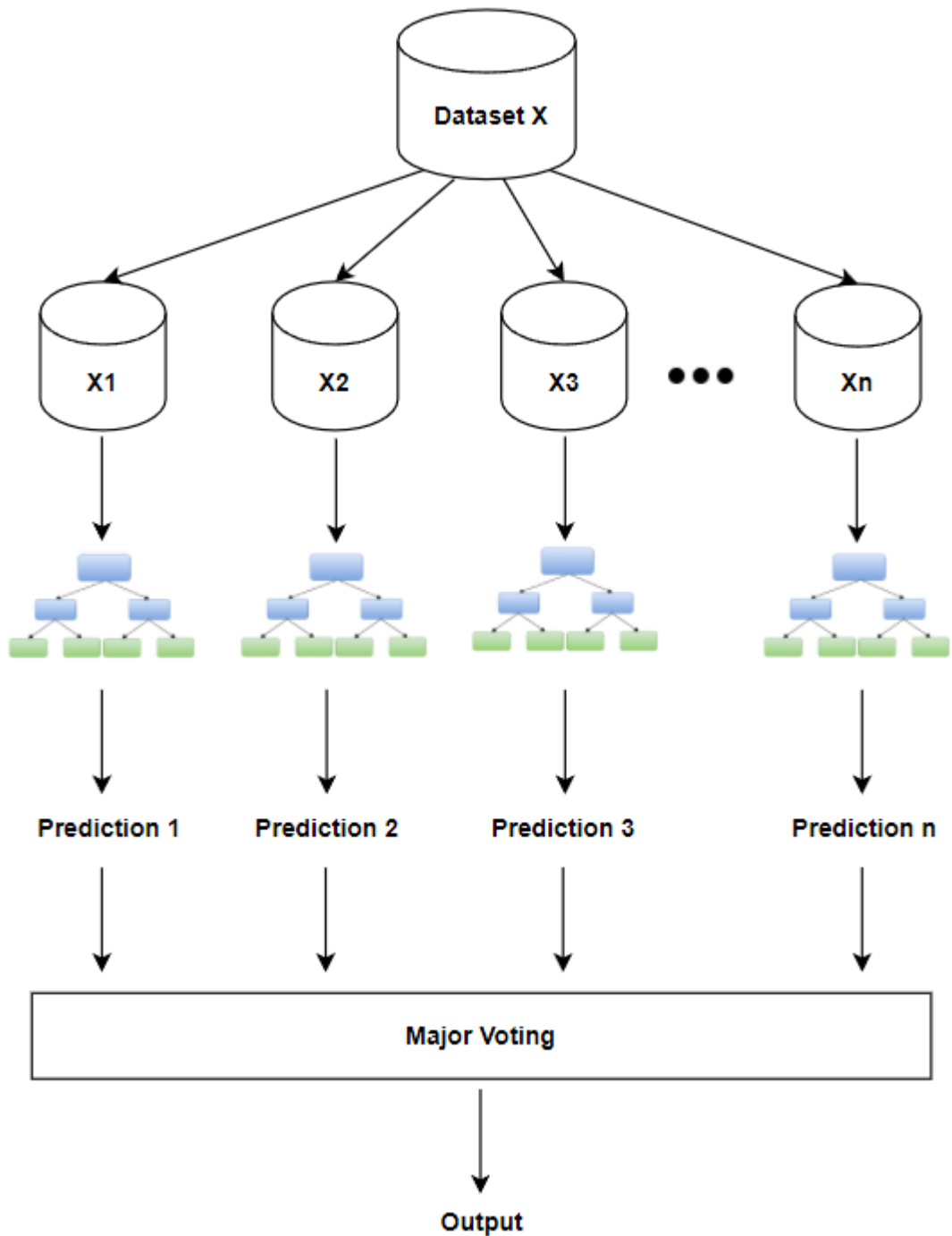
Figure 2.3 Example of random forest.

2.2.5 Support Vector Machine

Support vector machines (SVMs, sometimes referred to as support vector networks) are supervised learning models and related algorithms invented by Vladimir

Vapnik at AT&T Bell Laboratories in 1992 [31]. SVMs are the most powerful approach for classification and regression analysis that is based on statistical learning and Vapnik-Chervonenkis theory.

In this scenario, the SVM model seeks to determine the ideal hyperplane that best describes the difference between data, caries, and non-caries. To keep the training set small, the Gaussian radial basis function is used in the classifier. For a given training data $D = \{x_i, y_i\}$ and $y_i \in \{-1, 1\}$ the SVM classifier and the mapping function of the Gaussian kernel can be described as follows in Equation (2.6) and (2.7) :

$$\min_{\omega, b, \xi} \frac{1}{2} ||W||^2 + C \sum_i \xi_i^2 \; subject \; to \; y_i(W^T X_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \qquad (2.6)$$

where $C > 0$ is the selected parameter and ξ is a set of slack variables.

$$K(X, Y) = e^{\frac{||X-Y||^2}{A}} \qquad (2.7)$$

where K is the kernel function and A is the constant.

2.3 Convolutional Neural Network

Convolutional neural networks (CNNs) are a form of neural networks that are employed mostly for grid-like data, such as sequences and images [32]. The phrase "convolutional" stems from the fact that CNNs undertake convolution operations as opposed to the traditional weighted sum utilized in traditional feedforward neural network.

Convolution operation is a mechanism to calculate weighted sum in which a collection of weights, termed a kernel or a filter, is shared between local patches in the input. Convolution of input I with a kernel W of size H×H is defined as follows in Equation (2.8) for 2D discrete data such as digital images:

$$
\begin{aligned}
O(m,n) &= X(m,n) \times W(m,n) \\
&= \sum_{i=\frac{H-1}{2}}^{\frac{H-1}{2}} \sum_{j=\frac{H-1}{2}}^{\frac{H-1}{2}} W(i,j)X(m-i,n-j)
\end{aligned}
\tag{2.8}
$$

Typically, a CNN is structured as a series of phases, with the initial stages consisting of two types of layers, namely convolutional and pooling layers [21]. Planes that arrange the nodes in a convolutional layer are known as feature maps. In the convolutional layer, inputs are convolved with a set of weights known as a filter to generate an output feature map containing specific features. Various inputs are convolved with various filters. In addition, several filter banks are convolved with an input to extract distinct characteristics.
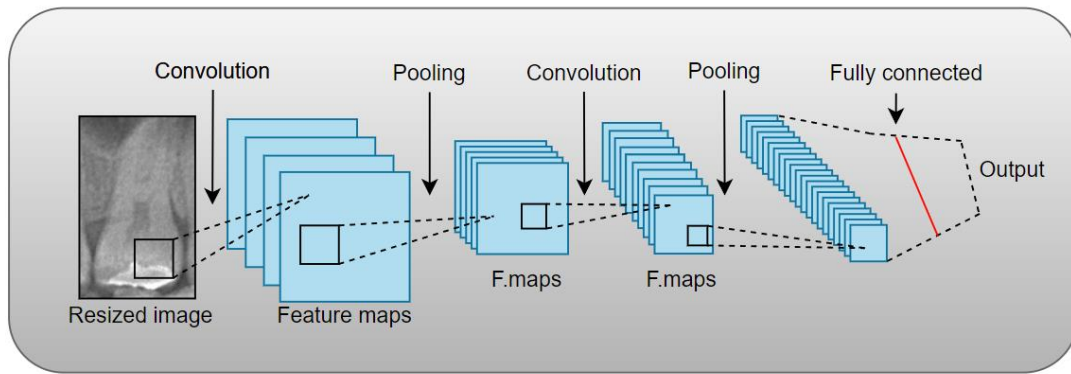
Figure 2.4 A model of convolutional neural network.

The model begins with a convolution layer whose activation function varies. Following an activation function is the pooling layer as illustrated in Figure 2.4. In the pooling layer, the input values of local regions are statistically summed into one output unit. Multiple methods of pooling may be utilized. For instance, max pooling returns the largest value inside the region, whereas average pooling returns the average value. Two to three repetitions of the stage are followed by further convolutional layers. The final stage consists of multiple completely interconnected layers and an output layer that displays forecasts for each class. Several variants of CNNs model have been proposed over the years such as Alexnet, GoogLeNet ...etc. and are generally defined as next sub suction.

2.3.1 AlexNet

In 2012, AlexNet [33], a convolutional neural network design, introduced the use of convolutional layers that were successively layered. Using graphics processing units, AlexNet's designers trained the network (GPUs). AlexNet is well-known for its remarkable result in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2010. In general, the network comprises of three layers of alternating

convolutional and max-pooling, two levels of convolutional processing, and three layers of fully connected processing. Following is a quick description of the model's distinctive attributes:

- ReLU - A utilized activation function, ReLU.

- Local response normalization - Normalization strategy used to improve generalizability. After the first two pooling layers, it is applied to each active output of each node. The normalized response $y_{i,j}^P$ j for each active output a $x_{i,j}^P$ is defined as in Equation (2.9):

$$y_{i,j}^P = x_{i,j}^P / \left(k + \alpha \sum_{j=\max\left(0, i - \frac{N}{2}\right)}^{\min\left(N-1, i + \frac{N}{2}\right)} \binom{P}{i,j}^2 \right)^\beta$$

(2.9)

where n is the number of neighboring filters used to normalize $x_{i,j}^P$, N is the number of total filters in the layer, and k, a, b are specified parameters.

- Overlapping pooling — For pooling layers, each location of a pooling window overlaps

The model also utilizes an overfitting avoidance approach employed dubbed "drop out". If you train using dropout, there is a 50% chance that each hidden node will be allocated with zero as an output in the forward pass when you use dropout. This permits the output provided by the network in each forward pass be produced with different architectures

2.3.2 GoogLeNet

GoogLeNet [34] is a 22-layer (27 layers including pooling layers) deep convolutional neural network that's a variant of the Inception Network, a Deep Convolutional Neural Network constructed by researchers at Google. The GoogLeNet architecture submitted at the ImageNet Large-Scale Visual Recognition Challenge 2014(ILSVRC14) addressed computer vision tasks like as picture categorization and object recognition. There are a number of computer vision tasks that GoogLeNet is presently being deployed.

The emergence of CNN, larger datasets, efficient computing resources, and intuitive CNN topologies accelerated the development of solutions for common computer vision applications that are efficient and performant. Researchers discovered that an increase in the number of network layers and units led to a significant performance advantage. Extending the layers to build more extensive networks, however, came at a cost. Large networks are susceptible to overfitting and either the bursting gradient or the vanishing gradient problem. Utilizing the Inception module in particular, the GoogLeNet architecture was able to resolve the majority of the problems that large networks faced. The Inception module is a neural network architecture that detects features at multiple scales using convolutions with various filters and minimizes the computational burden of training an extended network through dimensional reduction.

2.3.3 Residual Neural Network (Resnet)

Residual Networks, or ResNets, instead of learning unreferenced functions, learn residual functions with reference to the layer inputs. Residual nets let these stacked layers to fit a residual mapping, as opposed to hoping each few piled levels

directly match a desired underlying mapping. Residual blocks are stacked atop one another to build networks; for instance, a ResNet-18, ResNet-50, ResNet-101 consists of 18, 50, and 101 layers of these blocks, respectively.
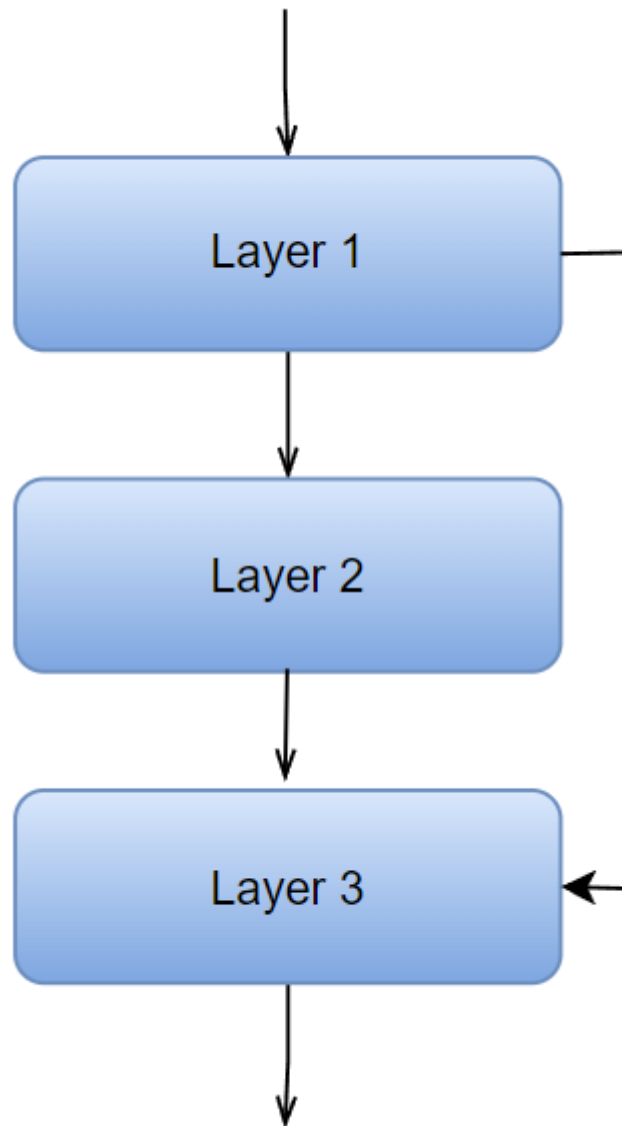
Figure 2.5 A form of residual neural network.

There are two major reasons to add skip connections: to prevent the problem of vanishing gradients and to minimize the Degradation problem, in which adding

additional layers to a sufficiently deep model results in an increase in training error [35]. During training, the weights mute the upstream layer and enhance the layer that was skipped. In the simplest example, just the weights for the next layer's link are modified; the upstream layer does not have explicit weights. This is most effective when only a single nonlinear layer is traversed, or when all intermediate levels are linear. Otherwise, an explicit weight matrix must be learnt for the omitted link. An illustration of residual network model is depicted as in Figure 2.5.

Bypassing significantly optimizes the network by employing fewer layers during the first phases of training. This accelerates learning by decreasing the influence of vanishing gradients, since there are fewer layers through which information must travel. As it learns the feature space, the network then progressively recovers the skipped levels. At the conclusion of training, when all layers are extended, the agent remains closer to the manifold and so acquires knowledge more quickly. A neural network devoid of residual elements examines a larger portion of the feature space. This renders it more susceptible to disturbances that lead it to depart off the manifold, necessitating additional training data for recovery.

2.3.4   Very Deep Convolutional Networks (VGG Net)

VGG Net [36] is the name of a pre-trained convolutional neural network (CNN) developed by Simonyan and Zisserman of the Visual Geometry Group (VGG) at the University of Oxford in 2014; it was able to finish as the runner-up in the classification task of the ILSVRC (ImageNet Large Scale Visual Recognition Competition) 2014. VGG Net was trained using the ImageNet ILSVRC data set, which contains pictures of 1000 classes divided into three sets of 1,300,000 training photos, 100,000 testing

images, and 50,000 validation images [37]. The model achieved 92.7% accuracy in ImageNet.

VGG Net has trained to extract the characteristics (feature extractor) that differentiate items and is used to categorize unseen things. VGG was created with the intention of improving classification accuracy by increasing the CNNs' depth. For object identification, VGG 16 and VGG 19 with 16 and 19 weight layers, respectively, have been employed. VGG Net accepts 224×224 RGB pictures as input and processes them via a stack of convolutional layers with fixed filter size of 33 and stride of 1. Five max pooling filters are placed between convolutional layers to downsample the input representation (image, output matrix of a hidden layer, etc.). Three fully connected layers of 4096, 4096, and 1000 channels, respectively, follow the stack of convolutional layers. The last layer is a soft-max layer.

2.3.5 Xception Net

Google's Xception [38] (Extreme version of Inception) is evaluated. With a modified depth-wise separable convolution, it outperforms Inception-v3 [39] (also by Google, 1st Runner-Up in ILSVRC 2015) on ImageNet ILSVRC and JFT. The modified depth-wise separable convolution consists of a pointwise convolution and a depth-wise convolution.

Two small variations distinguish Xception from Inception: the sequence of operations and the presence or absence of non-linearity. The operation's sequence: As previously stated, the original depth-wise separable convolutions as often implemented (e.g., in TensorFlow) execute channel-wise spatial convolution prior to 11 convolutions, but the modified depth-wise separable convolution performs 11 convolutions prior to channel-wise spatial convolution. It is said that this is

29

insignificant since, when utilized in a stacked environment, only minor variations show at the beginning and conclusion of each linked conception module. The Existence or Nonexistence of Nonlinearity: After the first operation in the original Inception Module, there occurs non-linearity. No intermediary ReLU nonlinearity exists in Xception, the modified depth-wise separable convolution.

2.4 YOLOv3 Model: You Only Look Once

You Only Look Once (YOLO) is a real-time object identification system that uses neural networks [40]. This is a detection and recognition technique for different things in a photograph (in real-time). YOLO performs object detection as a regression problem and outputs the observed photos' class probabilities. The YOLO method detects objects in real-time using convolutional neural networks (CNN).

In addition, YOLO generates predictions with a single network evaluation, as opposed to systems such as R-CNN that need thousands for a single picture and show advantage on:

- Quick response: YOLO enhances the speed of detection since it can anticipate objects in real-time.

- Hight precision: YOLO is a prediction approach that offers accurate findings with little background mistakes.

- Learning capabilities: The method has exceptional learning capabilities, allowing it to pick up on the representations of things and use them in object detection.

This makes it extraordinarily speedy, almost a thousand times faster than R-CNN and one hundred times faster than Fast R-CNN

$$y = (p_c, b_x, b_y, b_h, b_w, c)$$
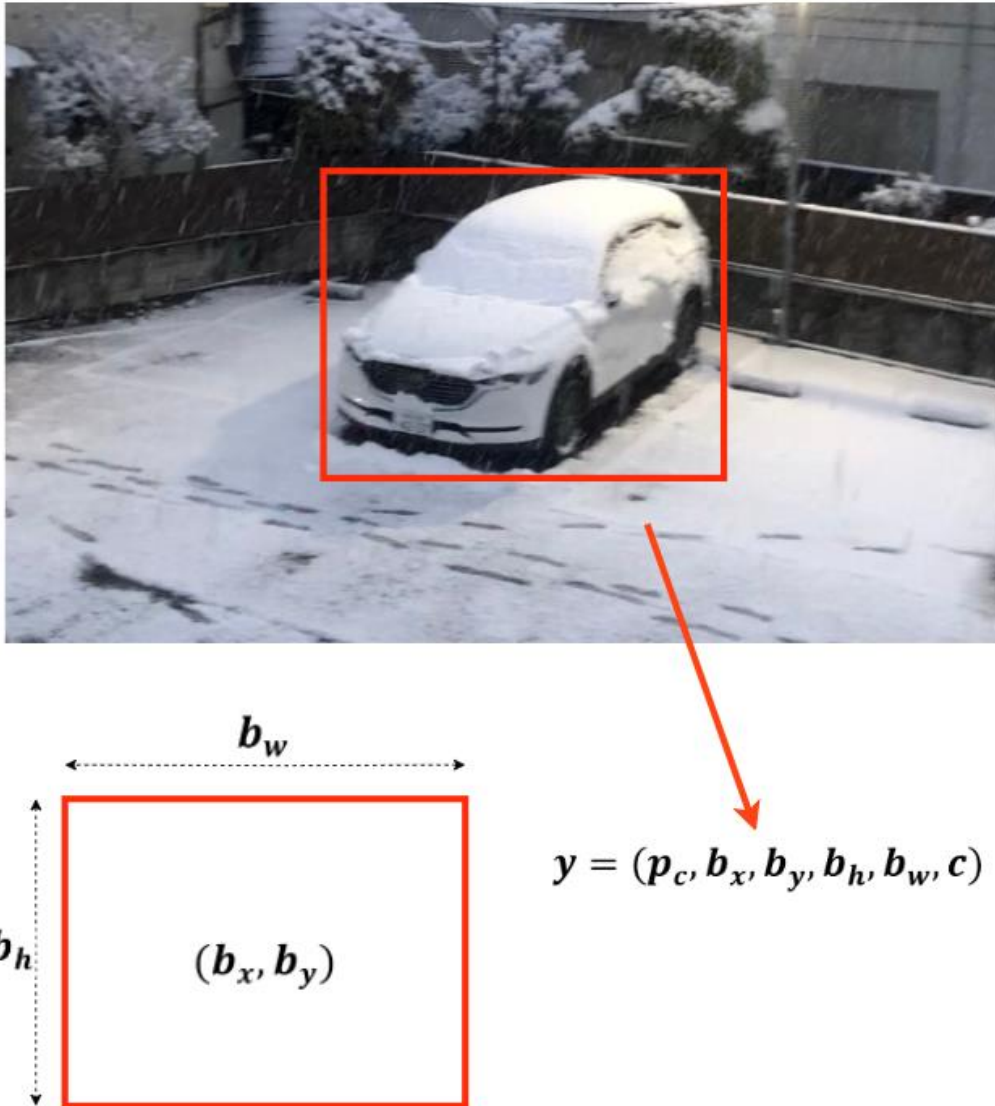
Figure 2.6 Example of car detection using YOLO.

As implied by the name, the method detects objects by performing just one forward propagation across a neural network. This indicates that a single algorithm run is sufficient to anticipate the complete picture. The CNN algorithm is used to concurrently predict multiple class probabilities and bounding boxes. Numerous variations of the YOLO algorithm exist. YOLOv3 and tiny-YOLO are two popular

examples. The benefits of speed, accuracy, and learning capabilities have propelled YOLO to global viral status for a variety of applications such as detecting animation, moving objects, and signals. Figure 2.6 describes a car detection using YOLO with parameters as: $b_x b_y$ is the coordinates of center of bounding box, b_x and b_w is high and width of bounding box, c is class of bounding box, and p_c is the probability of confidence.

CHAPTER 3

Tooth Localization

3.1 Introduction

Caries detection may be performed on bitewing or panoramic radiographs. However, panoramic images give a more thorough perspective of an individual's dental condition, which dentists prefer. On the other hand, tooth localization is critical for centralized diagnosis in every situation of caries detection. When compared to bitewing images, a panoramic radiograph image is substantially more complicated and useful. While the bitewing represents six or seven teeth in a specific region of the mouth, the panoramic image depicts the whole mouth and describes the state of each individual tooth, offering an intuitive view of the patient's dental health (Figure 3.1). Although the teeth in a bitewing picture are often relatively similar in form, a panoramic view analyzes all incisors (or front teeth), canines (or cuspids), premolars (or bicuspids), and molars (or back teeth). Because it is difficult to record the position of the bitewing in the mouth for medical re-examination in practice, bitewing is only acceptable for one-time analysis and not for medical history reference. Finally, but certainly not least, teeth detection on panoramic radiographs is beneficial for various research involving human

identification. Tooth isolation is focused on panoramic radiography for the reasons indicated above.
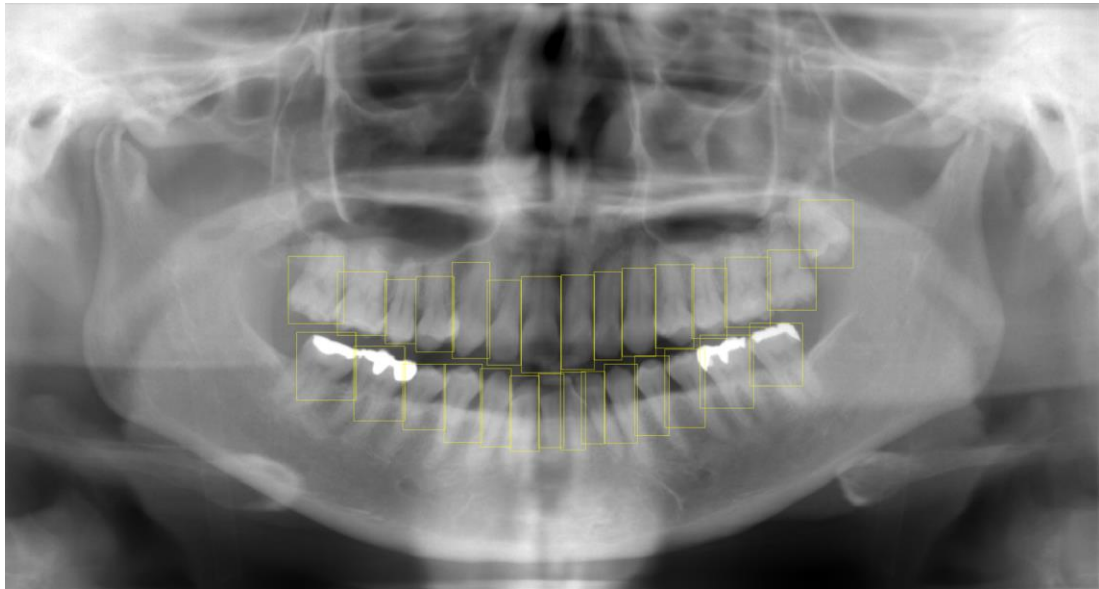
Figure 3.1 Panoramic radiograph and bounding box ground truth (image credit [41]).

3.2   Literature Review

Vijayakumari [42] pioneered a technique for personal identification based on biometric edge detection. To begin, radiographs are preprocessed to improve the contrast using frequency domain image enhancement and edges are detected using an isoperimetric approach. The images are additionally processed using Sobel and Canny edge detection, as well as nodal graph representation. The radiographs are then compared to provide the best result. Due to the absence of an explanation for the dataset and the small number of testing images, this research exposes a dataset issue. Additionally, the final images are fuzzy and incapable of capturing the deed in its entirety. To conclude, the suggested technique is unsatisfactory and has a low detection rate for teeth.

Radhiyah et al. [43] demonstrated a scheme to process panoramic radiographs using Gaussian Filtering and Histogram Equalization. The teeth are manually marked in the image prior to utilizing watershed segmentation. After that, the peak signal-to-noise ratio (PSNR) is calculated for comparison reasons, illustrating the appropriateness of each pre-processing procedure. Gaussian Filtering has a success rate of 80.58% on average, whereas Histogram Equalization has a success rate of 81.88% on average. The research analyzed a small sample of 28 images and provided no method for estimating the success factor. In conclusion, the result seems to provide a negligible contribution due to the low quality of the results and the manual marking of teeth, which requires human input.

Nomir et al. [44] focused on the shape and appearance of the teeth while identifying humans. A vector feature is produced from each tooth using the forcefield energy function of the grayscale picture and Fourier descriptors of the tooth's form. Fourier descriptors and potential energy are used to construct the feature vector. Following that, voting is conducted to determine the best match for the query image based on the results of each individual tooth. The team analyzed a total of 162 antemortem photographs. This research does not provide an explanation for the findings or a conclusive statement. Additionally, the antemortem image does not correspond to the medical history. As a result, there is some doubt about the research's utility.

Huang et al. [45] employed gray-scale integral projection and angle correction to analyze skewed images in order to minimize any information loss. The tooth loss gap valley is then located using an adaptive windowing technique. Additionally, isolation curves are removed, and a missing line is utilized to locate missing teeth. 241 upper jaw teeth received a 95.63% score, while 230 lower jaw teeth received a 98.71%

score. The research employed just bitewing images of the jaw (back teeth), which each include 6-7 teeth; as a consequence, the image is basic and does not fully depict the complete issue of tooth isolation in the oral cavity, which includes a variety of tooth types, sizes, and forms. Additionally, the data set is quite small (60 images) and does not provide a comprehensive picture of the tooth's state. Although the study is fascinating, it is not very useful. The absence of additional teeth, in particular, hampers the method when confronted with real-world scenarios.

Mahdi et al. [46] used an optimization approach with a deep learning system to recognize teeth in dental panoramic radiographs. To begin, candidates are selected using Faster R-CNN using panoramic radiographs. The detected candidates are refined in the most efficient manner possible. To get a better outcome in the second stage of detection, a deep learning-based faster RCNN technique leveraging Resnet is applied. The study is tested in dataset of 1000 images in several quality (1400-3100 × 800-1536 pixels) and the precision and recall rates were 98.8% and 97.8%, respectively. On the other hand, their strategy used a two-stage procedure, requiring the method to be trained twice; as a consequence, the method adds more computational weight, consumes additional labor and time, and causes the system to be less efficient and slow. Unlike the previous research, a unique strategy is offered in this work that is simpler to implement than the two-stage training technique while still producing outstanding outcomes.

The present researches are usually whether providing a simple and poor performance or providing a good performance with an extra high-tech investment requirement in hardware. Recognizing the current problem, this study' section aims to improve tooth isolation outcome while keeping the friendly complexity which help the method much simpler and friendly-use.

3.3 Proposed Method
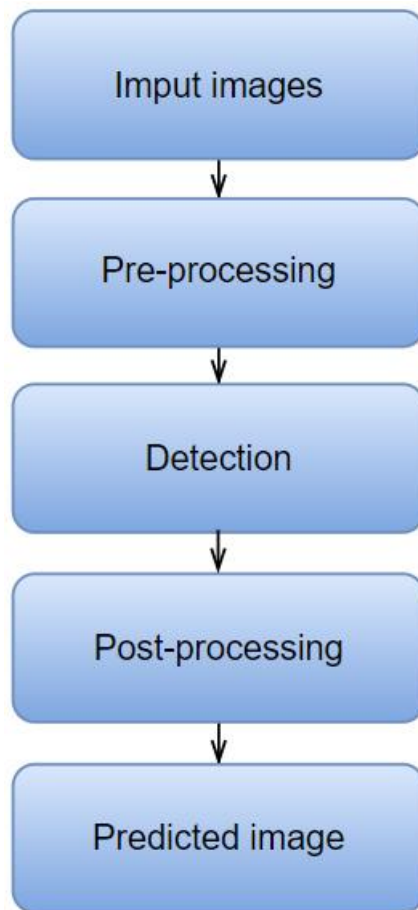
Figure 3.2 Flow chart of the proposed method (image credit [41]).

This section details the recommended approach. This section will explain how to build the technique and offer a detailed description of the structure of the Yolov3 detection model. Three critical components of the proposed approach are preprocessing, teeth detection, and postprocessing (Figure 3.2). Preprocessing will use the augmentation technique to provide more diverse and recent samples for the training dataset. The images will then be resized to fit the input size of the Yolov3 model. The section on tooth detection will cover the creation of the Yolov3 model and how it suits the circumstance. The model generates a set of tooth coordinates in a new image size

that is too small to see; as a consequence, the coordinates are scaled in the post-processing step to match the original image size.

3.3.1 Preprocessing

To begin, the region-of-interest (ROI) is used to the image in order to improve accuracy and decrease computation time. To avoid encroachment on the teeth, an area is picked in the picture's center that was 2000 pixels wide and had a preliminary ratio of 1:1.4 in contrast to the original image (Figure 3.3).
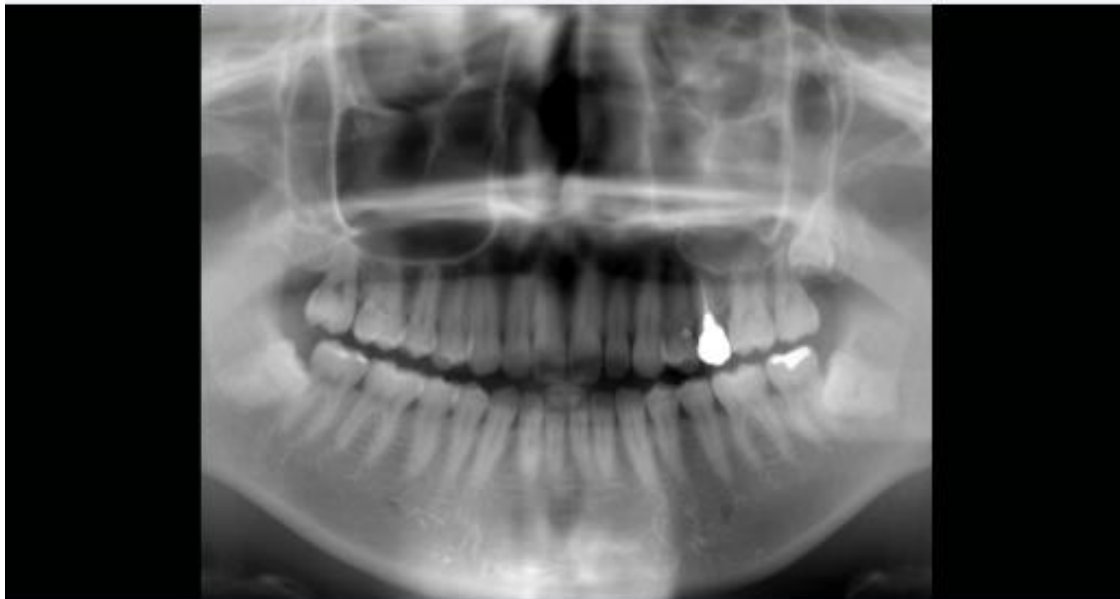
Figure 3.3 Region of interest on panoramic radiograph (image credit [41]).

Following that, as previously stated, a dataset is developed that encompasses as much of the issue as possible in order to address true medical problems. As a consequence, despite the fact that the dataset is already sufficiently big, it is still enhanced using two widely-used techniques: horizontal flipping and scaling [47]. Horizontal flipping improves the model's variety learning by doubling the challenge at both locations. Scaling enables the model to see the issue from a variety of angles,

which may be triggered by images recorded by a variety of medical devices. Additionally, augmentation may often improve the model's performance by making it more accurate and adequate. Images have been reduced in size to fit a particular detection model.

3.3.2 Localization using YOLOv3
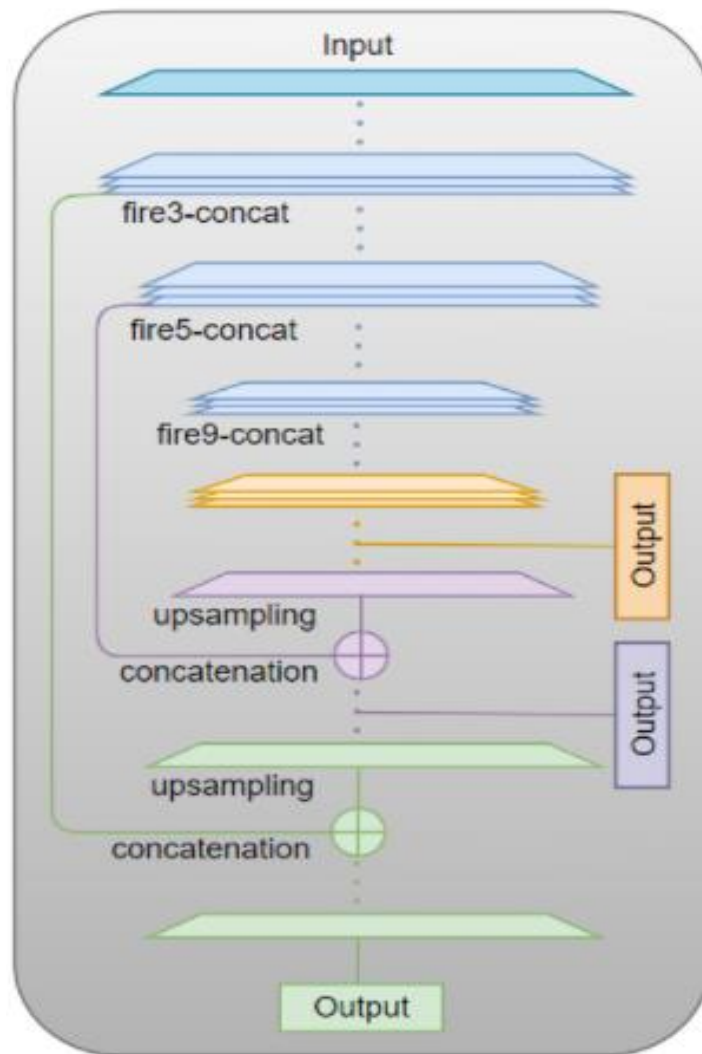
Figure 3.4 Yolo model (image credit [41]).

In this research, a Yolov3 model is used with a "Squeeze net" as the base network [40, 48], often referred to as the "backbone." The default input size for the model is 227×227, which suggests a 1:1 ratio; however, owing to the size of the image (1536×2000), the model input size is adjusted to 227×307. This is a vital step in maintaining the same picture ratio. Because a significant change in image ratio may harm the tooth shape and alter the object issue, particularly for the front teeth and overlapped teeth, a significant change in image ratio can harm the tooth shape and alter the object problem. As a result, increasing the quantity of the inputs will assist in boosting the performance of the deep model and reducing information loss.

As with extract features in Yolov3, a base network may have several outputs, referred to collectively as a detection head. The more detection heads a network has, the more accurately it can recognize small objects, since each higher level of detection head doubles the layer's size. As a result, this study aims to increase the number of detection heads and concatenate the findings from previous detection heads with a proper layer in order to provide a further improved outcome. To minimize overfitting and reduce complexity, the model's size should be considered. As a consequence, three detecting heads are used in our detection model. The first detection head will link to the backbone through the "fire9-concat" layer. To generate the output, the second detection head will up-sample the result of the first detection head and concatenate it with the output of "layer5-concat." Similarly, the third detection head's output will be generated by upsampling the result of the second detection head and concatenating it with the output of "layer3-concat." Figure 3.4 and Table 3.1 explains the detail of connection and input parameters.

Table 3.1 Parameter setting of Yolov3 model (table credit [41]).

Parameter	Value
Mini batch size	32
Number of anchor box	11
Iteration	1000
Initial learning rate	0.001
L2regularization	0.0005

3.3.3 Post-processing

As an output result, the detection model produces a vector of tooth positions and size that matches the magnified image (227 × 307). Due to the proximity or overlap of numerous teeth in oral imaging, a higher resolution image may assist patients and physicians in seeing more clearly. As a consequence, the output vector value is rescaled to match the dimensions of the source image (1536×2876). Additionally, this will offer readers a more complete view of our results.

3.4 Experimental Setup and Result

3.4.1 Dataset

In fact, the tooth comes in a range of sizes, forms, and varieties. Therefore, a large dataset is necessary that is capable of accurately describing all dental conditions or at least covering the greatest amount of diversity. As a consequence, a dataset is adopted by Dr. Kumon Makoto at the Shinjuku East Dental Office. The collection comprises 396 high-resolution panoramic radiographs (1536 × 2876 pixels) (11,707 single teeth totally) from 396 patients. Adults have a greater challenge, since their teeth are no longer milk teeth but permanent teeth, which implies that each tooth extracted

cannot be returned to its original form. Each picture portrays a real patient from the hospital who was 18 years old at the time and was carefully selected with the patient's agreement to contribute to the research's objectives. Finally, the committee at Tokai University evaluated the permission to use and disseminate dataset images ethically. The mouth and a piece of the patient's jaw bone are shown in Figure 3.1 as an example image from the collection. Some of the teeth have cavities, which provides variation to the dataset.

The data set is divided into two parts: a training set that comprises 90% of all images and a testing set that contains 10% of all images. Additionally, k-fold cross-validation is applied on both the training and testing sets to minimize overfitting and to provide a more complete analysis of the final performance result.

3.4.2 Measurements

The measurement is determined using the well-known object identification technique described in the Pascal visual object classes (VOC) challenge [49]. The intersect-over-union, precision, recall, F1-score, and average precision components are all considered in the computation.

3.4.2.1 Intersect Over Union:

The bounding box overlap area is computed, also known as intersect over union (IOU), between the predicted bounding box and the ground truth bounding box to assess the bounding box's accuracy in identifying objects. The overlap should be at least 0.5 (50%) to be regarded as a successful detection, as in Equation (3.1):

$$IOU = \frac{area\ (B_p\ \cap B_t)}{area\ (B_p\ \cup B_t)} \tag{3.1}$$

where B_p denotes the predicted bounding box, B_t the ground truth bounding box, $B_p \cap B_t$ the intersection of the predicted and ground truth bounding boxes, and $B_p \cup B_t$ the union of the predicted and ground truth bounding boxes.

3.4.2.2 Precision and Recall

The performance assessment is calculated using two well-known metrics: accuracy and recall. While accuracy indicates the percentage of properly identified items relative to the total number of detected objects, recall indicates the percentage of correctly detected things relative to the total number of detected objects, as in Equation (3.2) and (3.3).

$$Precision = \frac{TP}{TP + FP}$$
(3.2)

$$Recall = \frac{TP}{TP + FN}$$
(3.3)

where true positive (TP) denotes the number of teeth accurately discovered; false positive (FP) denotes the number of teeth incorrectly detected; and false negative (FN) denotes the number of teeth undetected.

3.4.2.3 F1-score and Average Precision

Simultaneously, studies should emphasize the importance of precision and recall. However, recall and precision are typically inversely connected, which means that increasing precision decreases recall and vice versa. As a consequence, additional evaluation metrics are provided, such as the F1-score and average accuracy, in order to facilitate future comparisons with other investigations. The F1-score equalizes the

weights of precision and memory to get the harmonic mean of precision and recall. The formula below may be used to calculate the F1-score, as in Equation (3.4).

$$F1 - score = 2\frac{precision \; \times recall}{precision \; + recall} \tag{3.4}$$

The accuracy and recall value can be provided for a certain IOU threshold and, as a result, build a curve to represent the relationship between precision and recall. The average precision is shown by the area under the curve (AP). The average accuracy value is determined as in Equation (3.5) and (3.6):

$$AP = \frac{1}{N} \sum_{r \in \{0,...,1\}} p(r) \tag{3.5}$$

where N is the number of points on the precision-recall curve interpolated, and $p(r)$ is the greatest precision recorded for a technique with a recall greater than r.

$$p(r) = \max_{\tilde{r}:\tilde{r} \geq r1} p(\tilde{r}) \tag{3.6}$$

where $p(\tilde{r})$ is the measured precision at recall $\tilde{r}$.

3.4.3 Experimental Result

3.4.3.1 Overall Performance

The findings of the 5-fold cross-validation experiment, as well as the average precision, recall, f1-score, and average precision, are summarized in Table 3.2. The result demonstrates a highly promising performance, since precision and recall values are often near to one another and, in most situations, exceed 90%. The highest result in

fold-4 is 96.97% for precision and 96.16% for recall. Precision and recall both perform at an average of more than 90%, suggesting their dependability.

Table 3.2 Experimental result of k-fold and average of proposed method

(table credit [41]).

Measure	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Avg
Precision	95.96	95.45	94.44	96.97	96.27	95.58
Recall	95.15	94.65	93.65	96.16	94.90	94.90
F1-score	95.55	95.05	94.04	96.56	95.58	95.36
m-AP	97.42	94.33	95.08	97.30	96.61	96.15

Additionally, the f1-score and average precision values are included in the overall assessment. Due to the frequent inversion of accuracy and recall, the f1-score value may be utilized as a referee when comparing following investigations. The method offers a reliable outcome in average accuracy, which is often more than 90%, suggests that it is trustworthy for dentist's references in screening and diagnosis process using panoramic radiographs. Finally, the simple implementation of method can help raising physician's applicability and patient's awareness to improve overall social health.

Figure 3.5 illustrates an example of the detection approach that makes use of the bounding box for each tooth, as well as the confidence rate associated with each bounding box. Additionally, the execution time of the recommended approach is calculated on our system, MATLAB 2020, on a gen-9 core i7 computer equipped with an NVIDIA GeForce RTX 2060 graphics card, and published it in Table 3 to provide a complete view of both performance and computational complexity.
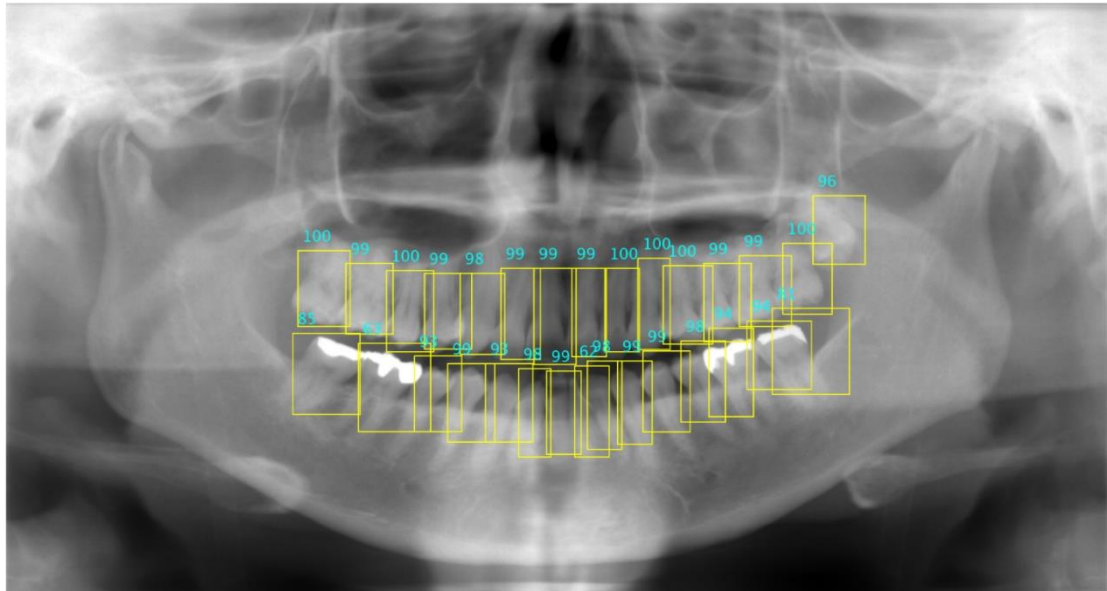
Figure 3.5 Teeth detection and confidence rate (image credit [41]).

3.4.3.2 Comparison to State-of-the-art

Table 3.3 Performance comparison to previous studies (table credit [41]).

Method	Size	Based Technique	Precision	Recall	F1-score	mAP	Iteration
Radhiyah et al.	28 images	Histogram Equalization	-	81.88	-	-	-
Mahdi et al.	1000 images	Fater R-CNN on Resnet101	98.8	97.8	98.20	98.10	9000
Proposed Method	396 images	Yolov3 on Squeezenet	95.58	94.90	95.36	96.15	1000

After evaluating the final performance, the preceding state-of-the-art is again benchmarked (Table 3.3). When compared to the original research, our novel approach exceeds it by about 10% in recall values and delivers more comprehensive evaluation values on panoramic radiographs. In terms of f1-score value and mean average accuracy, the second research exceeds our approach by around 3% and 2%,

46

respectively. On the other hand, the second research utilized a two-stage approach, which implies that one detection mistake might affect another, demanding intensive training to reach a reasonable result. Additionally, their approach demands far more computation time, effort, and high-tech equipment than our alternative. A high level of performance is archived in nine times the time with iteration. For the reason stated above, our proposed approach has a greater chance of universal adoption, especially in developing and impoverished countries.

3.5 Conclusion

This section describes a new mechanism for recognizing teeth in oral images that resolves a problem. Firstly, the radiograph is enhanced in order to provide variation. The tooth is subsequently recognized using an oral image-based Yolov3 detection model. The model is modified to account for the complexities of the circumstance and explored layer by layer. Additionally, the Yolov3 model and the input images are scaled to fit each other while preserving the original image ratio for optimal detection. Later, the bounding box value is adjusted to correspond to the original picture, allowing readers to perceive the result more clearly. The final assessment demonstrates a successful conclusion, with precision and recall rates of 95.58% and 94.90%, respectively, for our approach. Additionally, for a complete conclusion, the outcome is compared to the previous state-of-the-art. The proposed technique is somewhat less efficient (by around 3% on average) than the prior best state-of-the-art, which is a research constraint.

The following are the major results drawn from our involvement in this investigation. To begin, the image must be rescaled to fit the classification model's input; nevertheless, the image's ratio must stay constant or as close to constant as

possible. For more information, the results highly advise against decreasing the width to height ratio of the image and instead recommend lowering the height to width ratio. Because any drop in the width of the image relative to its height may result in information loss, impairing learning and resulting in a poor detection performance as a consequence. Second, our technique resulted in an exceptional conclusion that is remarkably close to the preceding state-of-the-art, while requiring less labor, less time, and less high-tech gadgets, resulting in a wide applicability advantage.

CHAPTER 4

Caries Detection

4.1 Introduction

The primary and ultimate goal of this study is to develop a decision-making system capable of distinguishing between healthy and carious teeth. Caries identification is often a multi-step procedure involving numerous stages; as a result, caries detection using radiographs is very tough for the majority of dentists, particularly inexperienced dentists. Computer-aided diagnostic systems (CADs) have become more significant in the early detection of a number of ailments, including cancer, diabetes, and even dental cavities, as medical imaging technology has advanced [6, 7]. Caries may be detected by a number of different processes and approaches. According to certain research, photoacoustic images, wavelengths, or ultrasound images should be employed for detection [8-10]. Another research suggested an approach based on an RGB oral endoscope image [11, 12]; nevertheless, the majority of systems are unable to provide a complete anatomy of the tooth, especially the tooth root, and so struggle to detect caries. In comparison to oral endoscopic imaging, dental radiographs provide a higher-quality image and show detailed structural deformation inside the tooth [13];

thus, the dental radiographic image is the most often used method and is favored for early detection of caries.

Dental radiographs, which are used to detect tooth problems and monitor oral health, are taken in a clinical environment utilizing a low-dose X-ray to acquire images of the interior of teeth and gums. Radiographs are often shown in grayscale images, but they may also be displayed in color. However, color radiographs require a significant investment that is unaffordable for the majority of institutions, especially those in low-income countries; to compensate, grayscale radiographs are focused. Regrettably, there is no credible dataset that provides high-quality images, descriptions, or ground truth. The majority of data in this field is shared under stringent conditions, such as all researchers publishing in a particular journal or being members of a certain group or event. Certain scholars make their confidential study data available to the public. The picture quality, the volume of data, the absence of description and ground truth, and/or the data's long-term availability are all frequent issues with the data. Dr. Kumon Makoto, director of Shinjuku East Dental Office, contributed the dataset and ground truth for this study as part of a research contract with Tokai University. On May 19, 2003, Dr. Kumon Makoto was certified by the Academy of Clinical Dentistry and registered as a professional dentist with the number 148529. He could supply a reliable dataset because he has 18 years of expertise as a dentist and is responsible for over 200 patients every month. More importantly, all of the patients who took part in the dataset collection were real Dr. Makoto patients who were being treated by him. During treatment, each caries tooth in the dataset was confirmed in the patient's medical history. For the stated reason, the dataset is proved reliable and can be utilized for research and publication for the reasons stated above.

4.2   Literature Review

Caries may be detected during a dental checkup using radiography by a fracture in the tooth, missing portions of a tooth, or tooth loss. Except for the dentist's diagnostic expertise, there is no visual symptom or criteria for the shape, size, or severity of tooth decay, posing a severe hurdle to image-based computer-aided diagnosis systems.

Wei Li et al. [50] developed a strategy for detecting tooth decay by combining a support vector machine (SVM) with a backpropagation neural network (BPNN). The Autocorrelation Coefficient and the Gray Level Co-occurrence Matrix are two features that are employed independently in this approach for feature extraction. Then, for classification purposes, SVM and BPNN models were utilized independently. SVM has a 79% accuracy on the testing set, whereas BPNN has a 75% accuracy. The resulting product is inefficient, and further effort is necessary to improve it. Apart from that, the article does not contain a description of the dataset, which may reflect negatively on the study's validity.

Yang Yu et al. [51] made an effort to enhance the backpropagation neural network layer and feature extraction from the autocorrelation coefficient matrix. The technique was assessed with 80 private tooth images (55 for training and 35 for testing) and achieved an accuracy of 94%; however, as the number of layers in the image increases, the system becomes highly computationally costly. The backpropagation neural network has increased in size. Sensitivity (SEN), specificity (SPEC), precision (PRE), and F-measure are also omitted. Additionally, the research's relatively limited testing data (35 images) without cross-validation exposes inadequacies, showing that the study is unable to completely address the issue of tooth decay.

51

Shashikant Patil et al.[52] demonstrated an intelligent system tailored for dragonflies. Multi-linear principal component analysis was used to extract the feature set (MPCA). After loading the attributes into a neural network classifier, it was trained using the adaptive dragonfly algorithm as an optimization technique (ADA). The proposed MPCA model non-linear Programming with ADA (MNP-ADA) was evaluated using 120 private dental images divided into three test cases. Each test case has a total of 40 images, 28 images are used for training and 12 images are used for testing. Other classifiers and feature sets, such as linear discriminant analysis (LDA) [53], principal component analysis (PCA) [54], and independent component analysis (ICA) [55], as well as fruit fly (FF) [56] and grey-wolf optimization (GWO) [57], were also used in the testing for comparison. According to the final average results, the MNP-ADA model achieves 90% accuracy, 94.67% sensitivity, and 63.33% specificity. The findings indicate that the test has a poor specificity, implying that non-caries patients are misclassified as caries patients. As a result, distinguishing caries from non-caries patients is inefficient, and performance must be improved. Due to the great accuracy but limited specificity of the result, it may raise issues about the data balance between caries and non-caries images. The findings section explores further into other measure values in this research, such as precision and f1-score.

Most researches do not provide an impacting result. Shashikant Patil et al.[52] seems performance the best in the state-of-the-art. However, specificity in the research is low with 63.33% in combination with 90% accuracy and 94.67% sensitivity. That raise on a concern about data balance, between caries and non-caries patients, and the impact of research in real case. Therefore, I would like to experiment a reliable in a large and balance data. The final result can comprehend the whole problem and well as produce a trustworthy opinion for dentist's reference.

4.3 Proposed Method
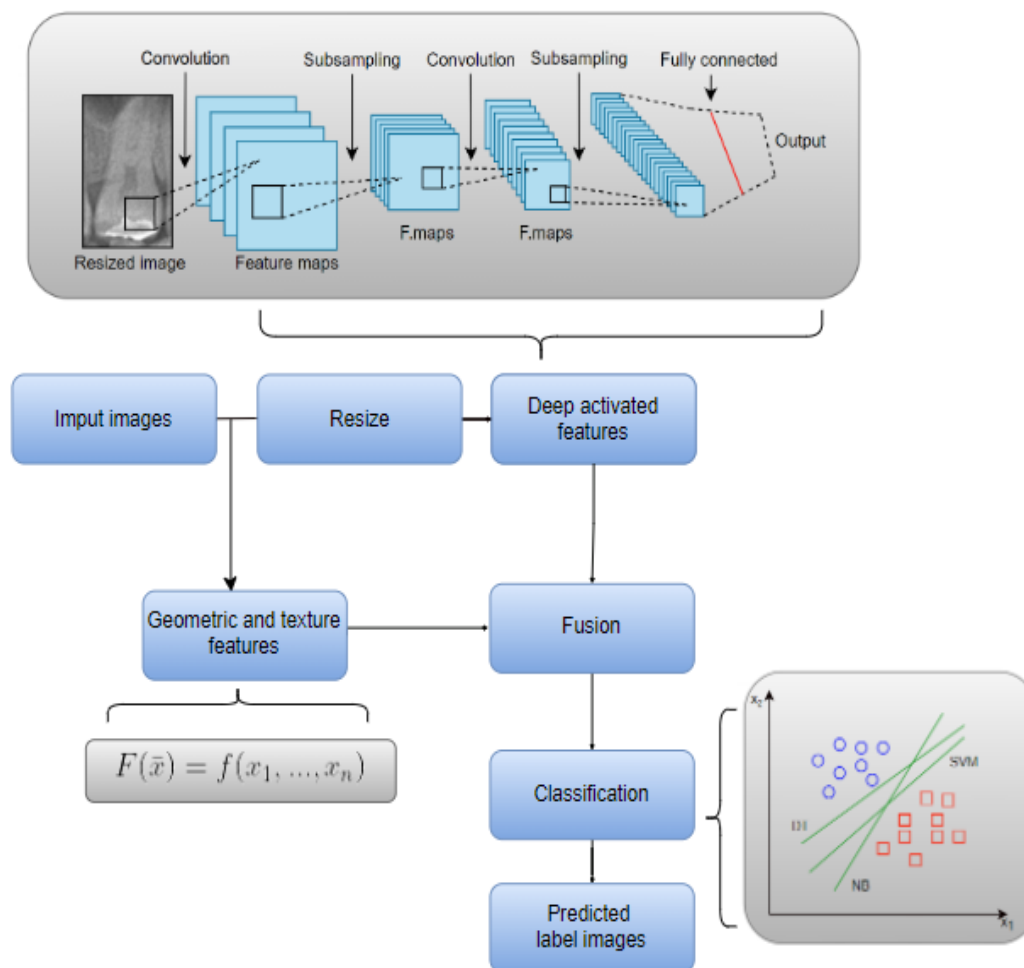
Figure 4.1 Diagram for caries prediction (image credit [58]).

The two fundamental stages of caries detection are feature extraction and classification. To begin, pre-trained models are employed such as Alexnet [33], Googlenet [34], VGG16 [36], VGG19 [36], Resnet18 [35], Resnet50 [35], Resnet101 [35], and Xception [38] to discover which deep activated characteristics best represent radiography. The experiments were conducted at the base levels of each model. Later, to supplement the information included in the feature, mathematic qualities such as

mean and standard deviation were extracted, as well as textural features such as Haralick's features [59]. Subsequently, the two sets of traits are combined to create fusion features. The second step validates the feature set used in classification models such as support vector machine (SVM), Nave Bayes (NB), k-nearest neighbor (KNN), decision tree (DT), and random forest. The complete process is shown in Figure 4.1, as well as various sub-stages.

4.3.1   Features Descriptors using Pre-trained CNN

Table 4.1 Convolutional neural network (CNN) model specification

(table credit [58]).

Network Model	Depth	Size (MB)	Parameter ($\times 10^6$)	Input Size
Alexnet	8	227	61.0	$227 \times 227 \times 3$
Googlenet	22	27	7.0	$224 \times 224 \times 3$
VGG16	23	528	138.4	$224 \times 224 \times 3$
VGG19	26	549	143.7	$224 \times 224 \times 3$
Resnet18	18	45	11.5	$224 \times 224 \times 3$
Resnet50	50	98	25.6	$224 \times 224 \times 3$
Resnet101	101	171	44.7	$224 \times 224 \times 3$
Xception	126	88	22.9	$299 \times 299 \times 3$

A pre-trained CNN is used as a feature descriptor in this study to extract deep activated features. The eight most well-known networks, Alexnet, Googlenet, VGG16, VGG19, Resnet18, Resnet50, Resnet101, and Xception network, were utilized to choose the best descriptor pre-trained networks. The parameters of each pre-trained

model, such as depth, parameter, size, and input size, are listed in Table 4.1. For the highest rate of deep learning, the most often suggested extraction layers are typically the last layer before the "prediction" layer; accordingly, in our preliminary experiment, Multiple levels are examined preceding the "prediction" layer (except the "drop" layer, because the drop layer likely shows the same information as the previous respective layer). Before a picture can be fed into a particular network, it must be scaled to a certain size. When radiographs are grayscale, the network analyzes RGB images; as a result, the grayscale channel is increased to compensate for the image's missing channel. The Results section contains information about the layers and network testing.

4.3.2   Features Descriptor using Geometric Features

Geometrical considerations are critical for discussing any issue. Because the qualities are obtained via a mathematical process, they are comprehensible and explicable. Regardless of the contribution of deep active feature descriptors, geometric features may include important and relevant information that is observable to humans. Additionally, whereas deep activated features investigate data in ways that are opaque to humans, geometric features are often acquired via an expert's domain expertise; hence, geometric features are essential and irreplaceable when addressing a challenging issue.

In practical terms, dentists manually differentiate between caries and non-caries depending on the extent of structural damage to the tooth. The variation in size, shape, contrast, margin, and intensity, among other factors, may account for the structural damage to the tooth. The suspicious aspects that determine the tooth's condition are obtained using their attributes, such as mean, Haralick's features, and gray level co-occurrence matrices (GLCM) features [60, 61]. Table 4.2 describes the name and

55

formula of used features in detail. In the formula, $I(x, y)$ presents the pixel value I at the coordinate point x, y of the candidate image N. $p(i, j)$ presents the $(i, j)th$ entry of GLCM matrix. N_g presents number of distinct gray levels in the image. μ and σ present the mean and standard deviation values.

Table 4.2 Geometric features and formula (table credit [58]).

Features	Name	Formula
F1	Mean	$$\frac{1}{n} \sum_{(x,y) \in N} I(x, y)$$
F2	Entropy	$$-\sum_i \sum_j p(i,j) \log p(i,j)$$
F3	Autocorrelation	$$\sum_i \sum_j (i \cdot j) p(i,j)$$
F4	Contrast	$$\sum_i \sum_j |i - j|^2 p(i,j)$$
F5	Correlation	$$\sum_i \sum_j \frac{(i - u_x)(j - u_y) p(i,j)}{\sigma_x \sigma_y}$$
F6	Cluster prominence	$$\sum_i \sum_j (i + j - u_x - u_y)^4 p(i,j)$$
F7	Cluster shade	$$\sum_i \sum_j (i + j - u_x - u_y)^3 p(i,j)$$
F8	Dissimilarity	$$\sum_i \sum_j |i \cdot j| p(i,j)$$
F9	Maximum Probabilities	$$\max_{i,j} p(i,j)$$
F10	Sum of square variance	$$\sum_i \sum_j |i - \mu^2| p(i,j)$$
F11	Sum of average	$$\sum_{i=2}^{2N_g} i \cdot p_{x+y} i$$

F12	Sum of entropy	$$\sum_{i=2}^{2N_g} P_{x+y}(i) \log p_{x+y}(i)$$
F13	Sum of variance	$$\sum_{i=2}^{2N_g} (i - SumEnt)^2 \cdot p_{x+y}(i)$$
F14	Difference entropy	$$-\sum_{i=0}^{N_g-1} p_{x-j}(i) \log p_{x-y}(i)$$

4.3.3 Fusion Features

This step integrates the deep network features and geometric attributes that were collected before. Each deep active feature is associated with the extracted geometric features as a whole.

The fusion feature is then included in a classification model in the subsequent step. Additionally, the proposed method evaluated the geometric and fusion features' performance relative to deep activated features by feeding each deep activated and fusion feature into the classifier under the identical conditions as fusion features (Figure 4.2). The Results section contains a full discussion of the fusion and deep activated feature comparisons.
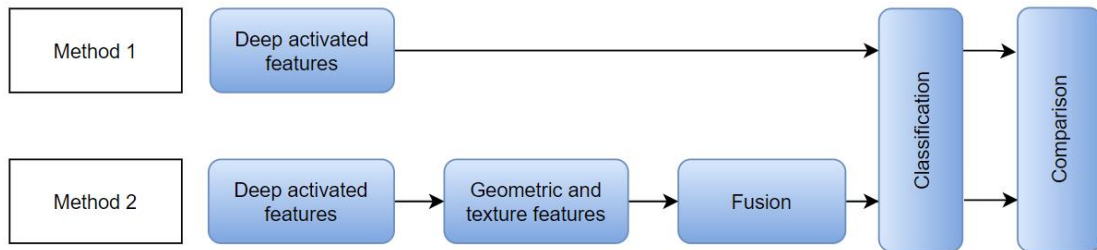
Figure 4.2 Diagram for experiment of deep activated and fusion features (image credit [58]).

4.3.4 Classification Model

Before being given into the classification system, each deep active feature is merged with geometric characteristics. Additionally, the deep activated features are evaluated independently and compared to the fusion features in order to determine the fusion effectiveness of deep activated and geometric features. The bulk of the tests was conducted using an "optimal margin classifier," more often referred to as a support vector machine (SVM).

4.4 Experimental Setup and Result

4.4.1 Dataset

To the best of our knowledge, the tooth is available in a range of sizes, forms, and architectures. Tooth decay features contribute to this variation; hence, the greater the collection, the more accurately it can depict tooth decay challenges. Our data was collected and labeled by Dr. Makoto Kumon from Shinjuku Dental East Office. The Tokai university's committee for the right of use and publication certified the dataset for quality and ethics; nonetheless, the images in the dataset are panoramic oral radiographs of all teeth; however, dental diagnosis and treatment should be performed on each individual tooth. As a consequence, each sub-image of the tooth, which is manually segmented, included the target tooth for diagnosis and its label. Segmentation is simple and may be accomplished in a number of different ways. Due to the ease at which segmentation may well be performed by any dentist or nurse, the statistically significant impact is not anticipated in this study (Figure 4.3). To simulate real-world situations in which the area determined for each tooth varies depending on who performs the segmentation, the area and range of teeth are not considered to be fixed

in any size, but rather extremely flexible depending on the size and position of the tooth, as well as the surrounding space.
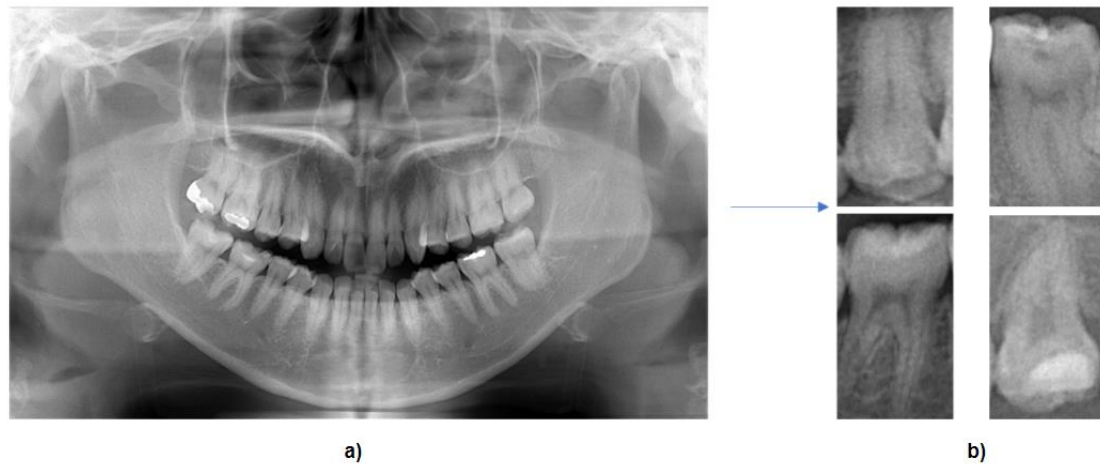
Figure 4.3 Samples of oral and tooth image. (a) oral panoramic radiograph and (b) segmented tooth radiographs (image credit [58]).

After segmentation, the dataset comprised 533 image samples from 95 patients: 229 caries teeth and 304 non-caries teeth. The dataset may be considered balanced because of the little variation in the number of caries and non-caries photographs (caries/non-caries is around 0.43/0.57). Each picture is a two-dimensional grayscale representation of the target tooth and its neighbors, which may include black empty space or a piece of nearby teeth. The images show teeth in their natural form, unaltered in terms of color, size, or angle. Each image is varied in size, which corresponds to the standard segmentation approach, and is subsequently combined into a single layer of the same size for feature extraction.

4.4.2 Measurements

In this study, three well-known metrics were used to evaluate the suggested method's performance: accuracy (ACC), sensitivity (SEN), and specificity (SPEC) (SPEC). In addition, the measurements include precision or positive predictive value

(PPV), negative predictive value (NPV), f1-score, area under the curve (AUC), and processing time to provide a holistic perspective of the suggested method's benefits and for additional research reasons, as in Equation (4.1) to (4.6).

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.1}$$

$$SEN = \frac{TP}{TP + FN} \tag{4.2}$$

$$SPEC = \frac{TN}{TN + FP} \tag{4.3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4.4}$$

$$NPV = \frac{TN}{TN + FN} \tag{4.5}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{4.6}$$

where true positive (TP) indicates the number of caries images correctly classified as caries; true negative (TN) refers to the number of non-caries images correctly classified as non-caries; false-positive (FT) means the number of non-caries images classified wrongly as caries; false-negative (FN) denote the number of caries images classified wrongly as non-caries.

Based on two mentioned factor values: True positive rate (also known as sensitivity or recall) and true negative rate (also known as specificity or selectivity), a ROC curve (receiver operating characteristic curve) is produced to present all model performance at all classification threshold. An ROC curve plots true positive rate

(TPR) as $sensitivity$ and false positive rate (FPR) as $1 - specificity$. Reducing the

classification threshold increases the number of objects classified as positive, hence

increasing the number of False Positives and True Positives. Figure 4.4 depicts a typical

ROC curve.
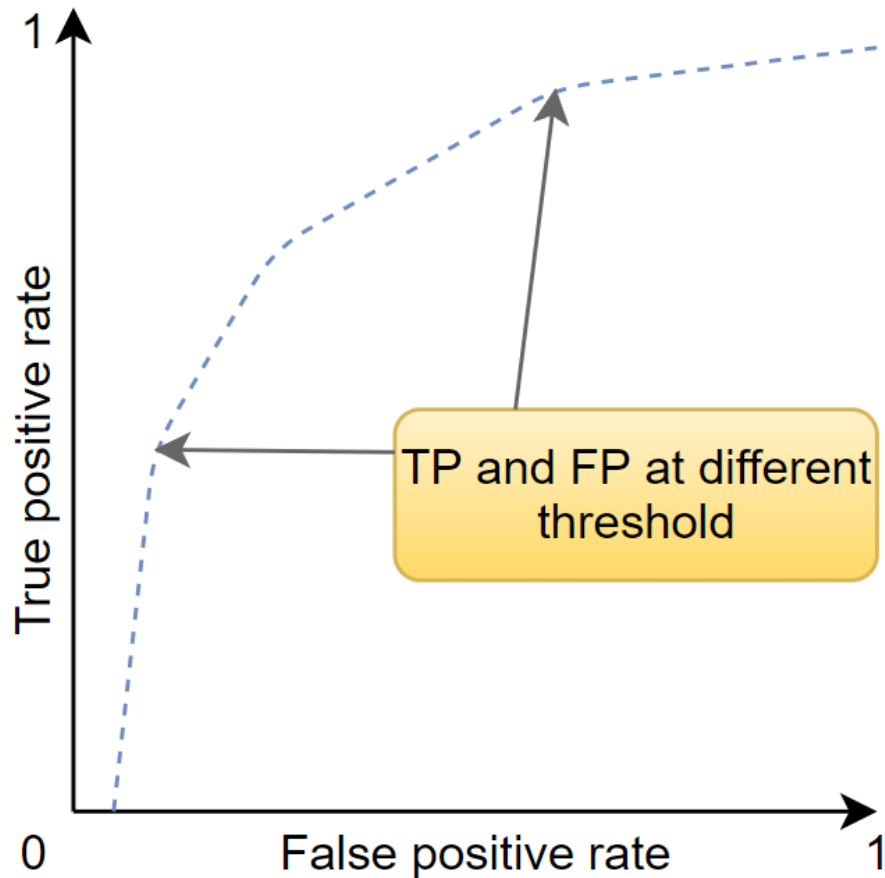
Figure 4.4 Typical ROC curve with TP and FP at different thresholds.

Area under the ROC curves (AUC) quantify the whole two-dimensional area

underneath the entire ROC curve. AUC is an aggregate performance metric over all

potential categorization criteria. The likelihood that the model rates a random positive

example higher than a random negative example is one approach to analyze AUC. A

typical AUC curve is presented in Figure 4.5.
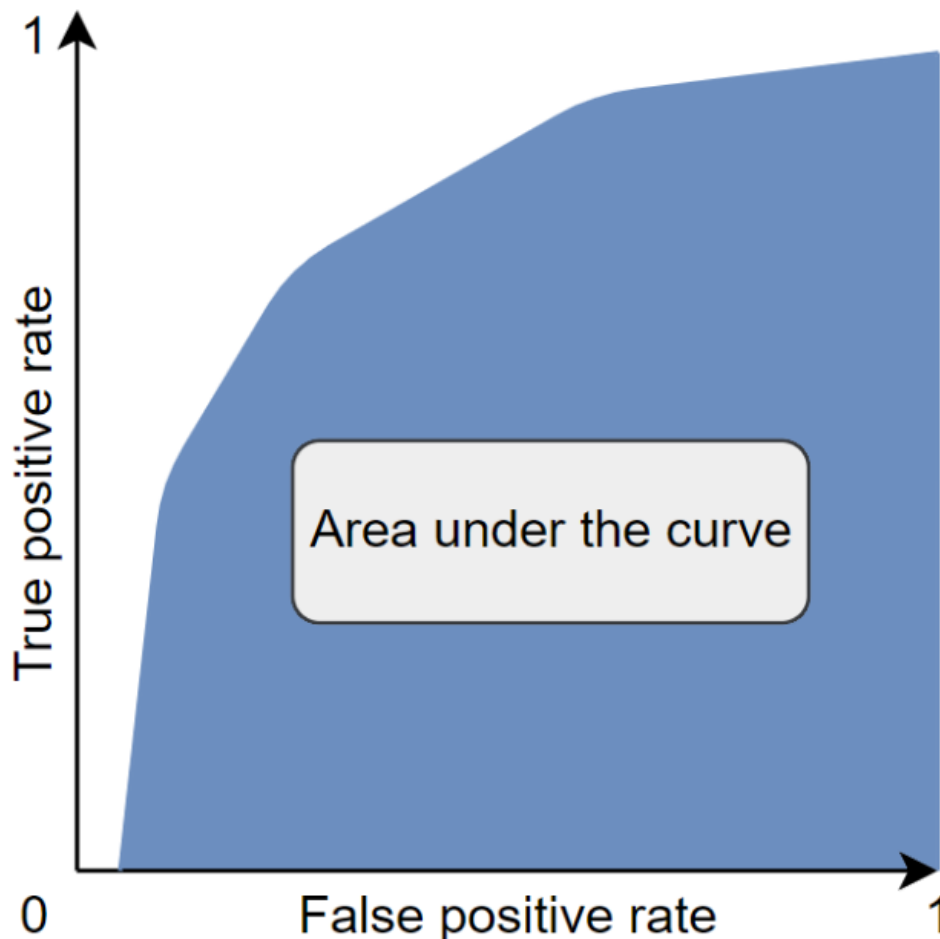
Figure 4.5 Area under the ROC curve.

4.4.3 Experimental Result

4.4.3.1 Deep Activated Features Performance

In the first step of the experimental result, the optimal layers are determined in each deep pre-trained network that best represents the issue. Table 3 summarizes the characteristics extracted from each deep pre-trained network by layer. To get the final classification result, a support vector machine model is used to assess the extracted feature set. Various layers are investigated before deciding on the prediction layer, since there is no standard for determining the layer in each network. Some of them illustrate the superiority of the pooling layer, while others may choose the layer first. With an accuracy of 90.57%, a sensitivity of 91.30%, and a specificity of 90%, the

"fc8" layer in the VGG16 model has the greatest accuracy, sensitivity, and specificity. Additionally, Resnet50, Resnet101, and Xception all achieve an accuracy rate of 88%, which is quite promising. Notably, none of the deep active features has an accuracy lower than 80%, demonstrating their effectiveness.

Table 4.3 Performance of deep activated features corresponding to networks (table credit [58]).

Network	Alex net	Google net	VGG 16	VGG 19	Resnet 18	Resnet 50	Resnet 101	Xception
Layer	fc8	pool5-7x7_s1	fc8	fc8	pool5	avg_pool	pool5	avg_pool
ACC	0.8679	0.8302	0.9057[1]	0.8113	0.8491	0.8868	0.8868	0.8868
SEN	0.7826	0.8261	0.9130	0.7391	0.8261	0.8696	0.8261	0.9130
SPEC	0.9333	0.8333	0.9000	0.8667	0.8667	0.9000	0.9333	0.8667
PPV	0.9000	0.7919	0.8750	0.8095	0.8261	0.8696	0.9048	0.8400
NPV	0.8485	0.8621	0.9310	0.8125	0.8667	0.9000	0.8750	0.9286
F1-score	0.7200	0.6786	0.8077	0.6296	0.7037	0.7692	0.7600	0.7778
AUC	0.9087	0.8333	0.9587	0.8674	0.9014	0.9565	0.9072	0.9464

[1]The highest performance for each measured factor regarding network was highlighted in bold.

4.4.3.2 Fusion Features Performance
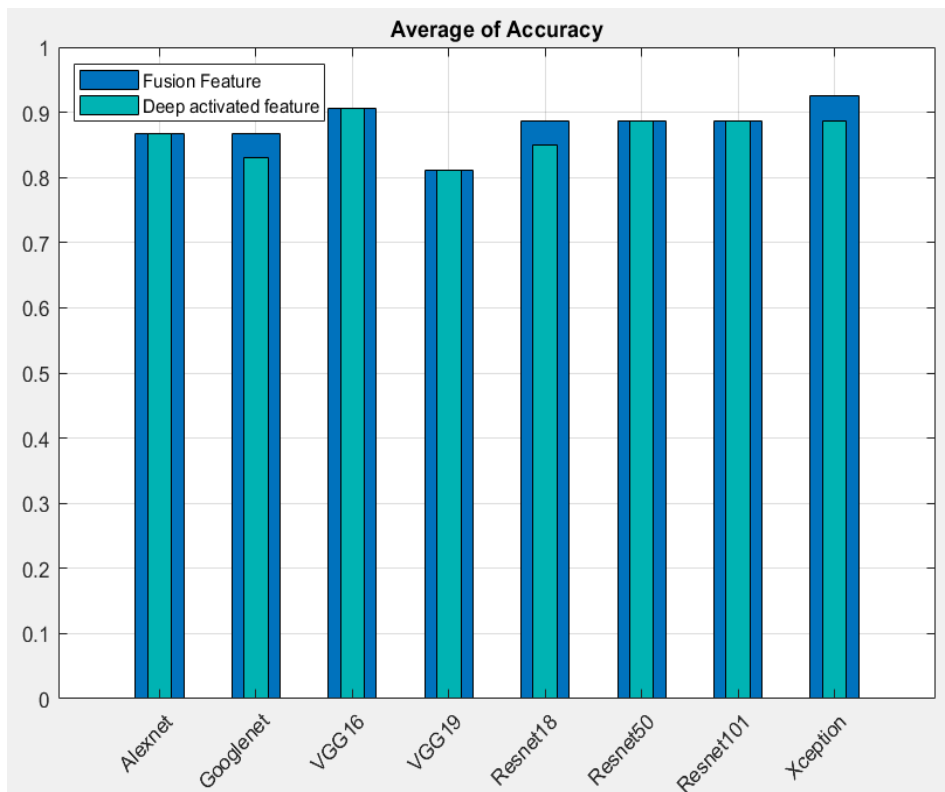
Figure 4.6 Overlay bar graphs distribution of average accuracy between fusion feature
and deep activated feature (image credit [58]).

To increase performance so far, each deep activated feature set is combined
with geometric features and will be fed into the SVM model (Table 4.4). The
conclusion of the fusion feature indicates that the fusion Xception feature has grown.
As seen in Figure 4.6, the Xception network's fusion properties were the most
significant qualities after the combination, increasing accuracy, sensitivity, and
specificity to 90.45%, 100%, and 86.67%, respectively. The most significant change is
the increase in sensitivity from 91% to 100%; hence, the fusion Xception features set
demonstrated geometric contribution when combined with deep activated features.
Resnet18 and Googlenet also exhibit an increase in accuracy of 83.02% to 86.79% and
84.91% to 88.68%, respectively, despite the accuracy that their performance is

consistent with Xception fusion characteristics. Interestingly, none of the fusion feature sets outperform their deep activated counterparts in terms of accuracy. Finally, when comparing fusion features to deep activated features alone, fusion features clearly demonstrate an advantage.

Table 4.4 Performance of fusion features corresponding to networks

(table credit[58]).

Network	Alexnet	Googlenet	VGG 16	VGG 19	Resnet18	Resnet50	Resnet101	Xception
ACC	0.8679	0.8679	0.9057	0.8113	0.8868	0.8868	0.8868	**0.9245[1]**
SEN	0.7826	0.8696	0.9130	0.7826	0.8696	0.8696	0.8261	**1.0000**
SPEC	0.9333	0.8667	0.9000	0.8333	0.9000	0.9000	**0.9333**	0.8667
PPV	0.9000	0.8333	0.8750	0.7826	0.8696	0.8696	**0.9048**	0.8519
NPV	0.8485	0.8966	0.9310	0.8333	0.9000	0.9000	0.8750	**1.0000**
F1-score	0.7200	0.7407	0.8077	0.6429	0.7692	0.7692	0.7600	**0.8519**
AUC	0.9087	0.8949	0.9594	0.8659	0.9123	0.9580	0.9087	**0.9688**

[1] The highest performance for each measured factor regarding network was highlighted in bold

The training and testing sets are randomly divided for cross-validation in order to develop and evaluate the best caries detection technique. The k-fold cross-validation approach is a well-known and trustworthy technique for evaluating a method's robustness. The use of k-fold cross validation indicates the proposed approach's ability to cover the whole issue and adapt to unknown samples; also, this methodology was used to prevent method overfitting on our testing data.
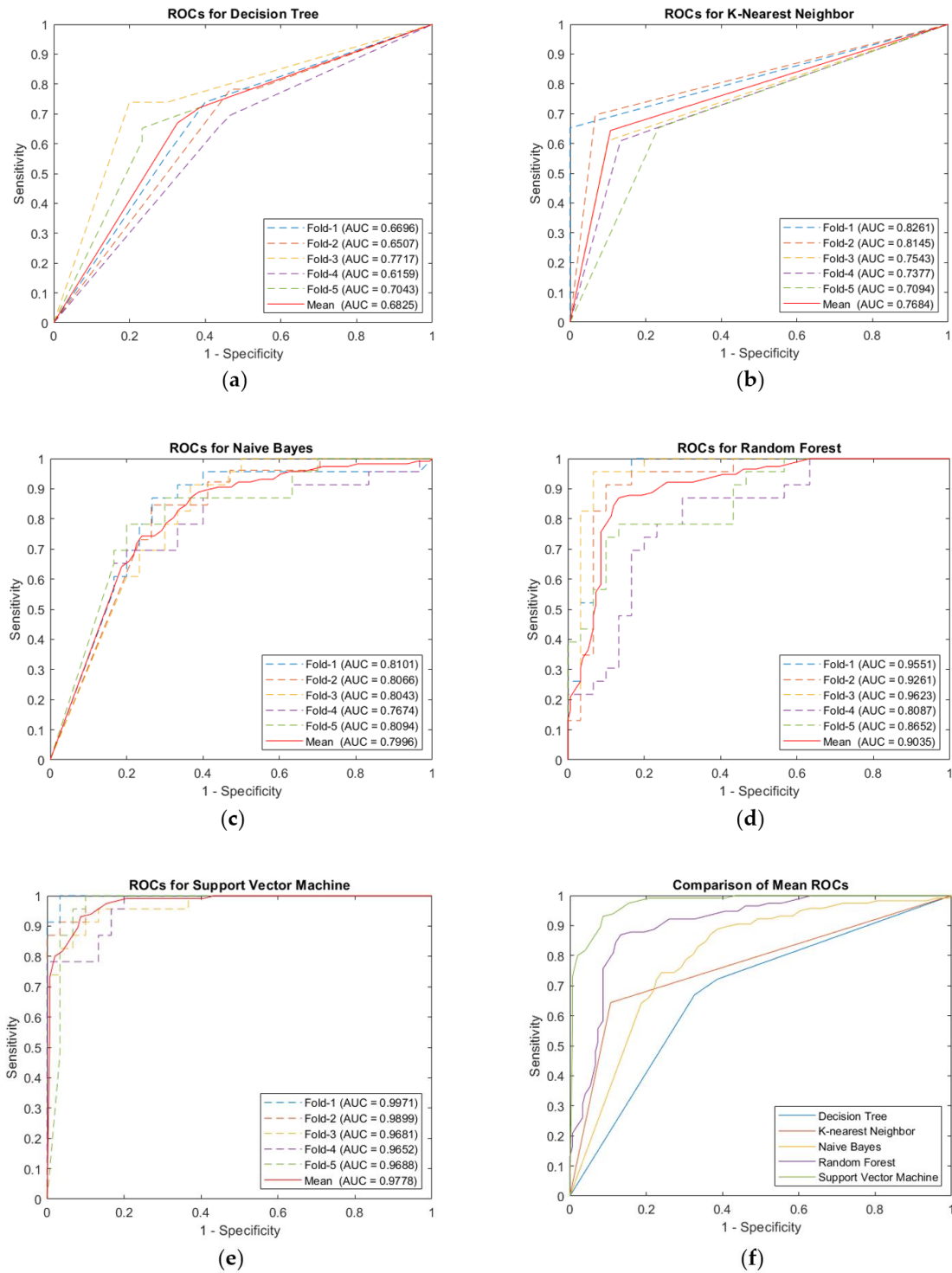
Figure 4.7 Comparison of the ROC curves for five classifiers. (**a**) Decision tree, (**b**) K-nearest neighbor, (**c**) Naïve Bayes, (**d**) Random Forest, (**e**) Support vector machine, and (**f**) Comparison of mean of receiver operating characteristic (ROC) curves for each classifier (image credit [58]).

4.4.3.3 Classification Model Performance

Table 4.5 Performance of fusion features based on classifiers (table credit [58]).

| Classifier | Measure | Five-Fold Cross-Validation | | | | | Mean |
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
Decision Tree	Accuracy	0.6415	0.6038	0.7170	0.6038	0.6981	0.6528
	Sensitivity	0.6522	0.7826	0.7391	0.6957	0.6087	0.6957
	Specificity	0.6333	0.4667	0.7000	0.5333	0.7667	0.6200
	PPV	0.5769	0.5294	0.6538	0.5333	0.6667	0.5920
	NPV	0.7037	0.7368	0.7778	0.6957	0.7188	0.7265
	F1-score	0.4412	0.4615	0.5313	0.4324	0.4667	0.4666
	AUC	0.6696	0.6507	0.7717	0.6159	0.7043	0.6825
K-Nearest Neighbor	Accuracy	0.8491	0.8302	0.7736	0.7547	0.7170	0.7849
	Sensitivity	0.6522	0.6957	0.6087	0.6087	0.6522	0.6435
	Specificity	1.0000	0.9333	0.9000	0.8667	0.7667	0.8933
	PPV	1.0000	0.8889	0.8235	0.7778	0.6818	0.8344
	NPV	0.7895	0.8000	0.7500	0.7429	0.7419	0.7649
	F1-score	0.6522	0.6400	0.5385	0.5185	0.5000	0.5698
	AUC	0.8261	0.8145	0.7543	0.7377	0.7094	0.7684
Naïve Bayes	Accuracy	0.7358	0.7333	0.7170	0.7547	0.7547	0.7391
	Sensitivity	0.6087	0.7308	0.6087	0.6522	0.6522	0.6505
	Specificity	0.8333	0.7353	0.8000	0.8333	0.8333	0.8071
	PPV	0.7368	0.6786	0.7000	0.7500	0.7500	0.7231
	NPV	0.7353	0.7813	0.7273	0.7576	0.7576	0.7518

	F1-score	0.5000	0.5429	0.4828	0.5357	0.5357	0.5194
	AUC	0.8101	0.8066	0.8043	0.7674	0.8094	0.7996
Random Forest	Accuracy	0.9057	0.8679	**0.9245[1]**	0.7736	0.7925	0.8528
	Sensitivity	0.8696	0.9565	0.9565	0.7391	0.6522	0.8348
	Specificity	0.9333	0.8000	0.9000	0.8000	0.9000	0.8667
	PPV	0.9091	0.7857	0.8800	0.7391	0.8333	0.8295
	NPV	0.9032	0.9600	0.9643	0.8000	0.7714	0.8798
	F1-score	0.8000	0.7586	0.8462	0.5862	0.5769	0.7136
	AUC	0.9551	0.9261	0.9623	0.8087	0.8652	0.9035
Support Vector Machine	Accuracy	**0.9623**	**0.9245**	0.8868	**0.8868**	**0.9245**	**0.9170**
	Sensitivity	0.9565	0.8696	0.7391	0.9565	1.0000	0.9043
	Specificity	0.9667	0.9667	1.0000	0.8333	0.8667	0.9267
	PPV	0.9565	0.9524	1.0000	0.8148	0.8519	0.9151
	NPV	0.9667	0.9063	0.8333	0.9615	1.0000	0.9336
	F1-score	0.9167	0.8333	0.7391	0.7857	0.8519	0.8253
	AUC	0.9971	0.9899	0.9681	0.9652	0.9688	0.9778

[1] The highest performance for accuracy regarding network was highlighted in bold

The most significant results from each classification model were then used to determine which classification model was most appropriate for the characteristics set. This analysis made use of decision trees (DT), k-nearest neighbor (KNN), Naive Bayes (NB), random forest (RF), and support vector machine (Table 4.5). Moreover, k-fold cross-validation is performed at this step to prevent overfitting the approach and to achieve the final average assessment. With an accuracy of 91.70%, a sensitivity of 90.43%, and a specificity of 92.67%, the support vector machine is definitely the most

dominant model. Due to the balanced nature of the used dataset and the small difference in the number of caries and non-caries samples, precision (also known as positive predictive value) and recall pair (also known as sensitivity) also show promising values of 91.51% and 90.43%, respectively, as mentioned previously in Section 3.1. Finally, receiver operating characteristic (ROC) curves are constructed for each classifier; the ROC curves indicate how each classifier performed in each fold of the experiment. The mean ROC curve for each classifier is interpolated and compared in Figure 4.7

4.4.3.4 Caries Detection Execution Time

The execution time of the proposed caries detection technique is also assessed to ensure a thorough assessment. The experiments were conducted in a Windows 10 environment using the Matlab2020a environment. The basic method was carried out on a CPU core i7-9750 HF and a GeForce GTX 2060 video card.

Table 4.6 Total execution time for each function of the proposed system

(table credit [58]).

Function Name	Time(s)
Load data	0.37
Deep activated features extraction	9.99
Geometric features extraction	2.52
Fusion features combination	0.01
Training classification model	0.62
Predict and evaluation	0.28
Total	13.79

Each function process is thoroughly analyzed since it is one of the factors that contribute to determining the method's complexity. According to Table 4.6, the whole method takes 13.79 seconds, and the most complex function, deep activated feature extraction, takes less than ten seconds. Additionally, the calculation of the geometric feature took just 2.52 seconds to complete. The recommended approach works well and is extensible to a range of computer systems based on these results. According to the prediction and evaluation timeframes, it will take just 0.28 seconds to determine if a tooth has caries or not (less than 1 second). Dentists will find this to be optimal even in major hospitals with a high patient volume.

4.4.3.5 Comparison to Related Study

The proposed approach was compared to currently available state-of-the-art techniques in Table 4.7. Due to the fact that the various techniques were applied to individual datasets, the size and complexity of each dataset impacted performance. The distinctions and benefits/disadvantages of each strategy are highlighted to provide a fair comparison. Additionally, a reference is given to the relevant article and a description, since some methodologies are not fully discussed but have been tested on other datasets in other works. The comparison table demonstrates that Wei Li et al. and Yang Yu et al. method underperform, while Shashikant Patil et al. outperforms much more; yet, with an accuracy of 90.00%, a sensitivity of 94.67%, and a specificity of 63.33%, both a data imbalance and a bad performance result can be obviously observed. In comparison to earlier methods, our proposed technique obtained 92.67% specificity, a 29.34% increase, while maintaining higher than 90% sensitivity. The drop in sensitivity of 4.24% is a reasonable trade-off for the improvement in specificity.

Table 4.7 Performance comparison of the proposed method and with the previous

methods (table credit [58]).

Reference	Method	Image	ACC %	SEN %	SPEC %	PPV %	NPV %
Wei Li et al. [50, 52]	• Autocorrelation and GLCM features • SVM classification	120	53.33	59.33	06.67	73.67	6.67
Yang Yu et al. [51, 52]	• Autocorrelation coefficients matrix • Neural network classification	120	73.33	77.67	53.33	90.33	53.33
Shashikant Patil et al. [52]	• Multi-linear principal component analysis • Non-linear programming with adaptive dragonfly algorithm • Neural network classification	120	90.00	94.67	63.33	91.00	63.33
Proposed method	• Deep activated features • Geometric features • SVM classification	533	91.70	90.43	92.67	91.51	93.36

4.5   Conclusion

This section described a caries detection technique based on radiography images. Firstly, dentists classified radiography images manually as caries or non-caries. Later in the feature extraction method, dental images were utilized to extract the deep active features. The optimal layer for obtaining deep activated features from each deep pre-trained model was identified throughout the research. Fusion features were created

71

by extracting geometric features and fusing them with deep activated features. The optimal features set was determined via a performance comparison of deep activated features and fusion features. The number of geometric attributes was minimized while retaining the most important information. Following that, the fusion is fed into classification models such as support vector machine (SVM), decision tree (DT), k-nearest neighbor (KNN), Naive Bayes (NB), and random forest to distinguish between caries and non-caries images (RF). The proposed approach obtained 91.70% for accuracy, 90.43% for sensitivity, and 92.67% for specificity, respectively. In comparison to previous state-of-the-art techniques, accuracy is increased by 1.7%, from 90% to 91.70%, and specificity by 29.34%, from 63.33% to 92.67%; sensitivity remained high at 90.43%. The proposed approach provided two key contributions: first, it determined the ideal feature set, which consists of a mix of deep activation and geometric characteristics; and second, it fitted a strong classification model to describe the situation. Second, by raising the specificity factor, performance may be improved. The complexity or size of the model has no effect on the performance of the deep activated feature. While VGG16 has a better deep activated function than Xception, the fusion result is the opposite. Although the selected deep active feature had a considerable influence, analytically calculated attributes also contributed to the result. It is more crucial to determine which deep activated feature features are compatible with analytically computed features than it is to determine which deep activated feature features are the best among all pre-trained models. While the majority of research aims to create the deepest networks possible in order to maximize learning performance, our results demonstrate that network depth is not necessarily a determinant in performance. More crucially, the combination of calculated features may be vital for performance enhancement, making it indispensable for the depth of the pre-trained model. The

processing time of 13.79 seconds for the whole experiment and 0.28 seconds for prediction demonstrates that the technique may be implemented widely in a short period of time on a low-tech machine.

CHAPTER 5

Automated System Combination

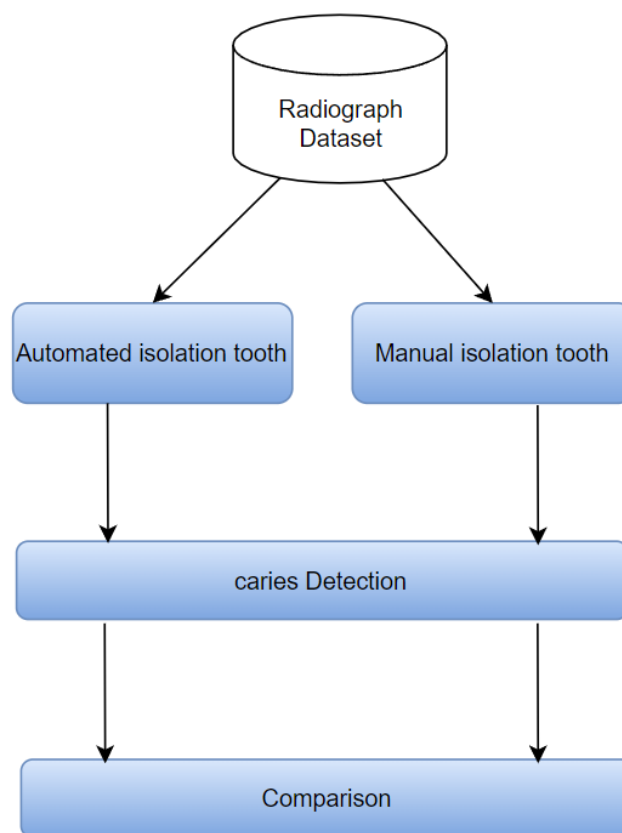## 5.1 Experimental Strategy

Figure 5.1 Caries detection of automated and manual tooth isolation.

This section will integrate the automated tooth separation and caries detection systems discussed in Chapters 3 and 4. The bounding box created by the isolation tooth systems will be utilized to segment the individual tooth for caries detection. The automated system's output will be compared to the results from manual segmentation tooth caries detection, as indicated in Figure 5.1.

5.2 Experimental Result

This section compares the results of the suggested approach to those of manual and automated methods. The measure metric is equivalent to the metric described in section 4.4.2. Therefore, the author prefers to proceed right to the examination of the results.

Table 5.1 Comparison between manual and automated isolation input images.

Method	ACC%	SEN%	SPEC%	PPV%	NPV%	AUC%
Caries detection on manual isolation	91.70	90.43	92.67	91.51	93.36	97.78
Caries detection on automated isolation	88.66	88.14	89.47	92.86	82.93	93.80

According to Table 5.1, the overall performance of caries detection on automated isolation images has been decreased by about 4%, reaching 88.66% for accuracy, 88.14% for sensitivity, and 88.47% for specificity, respectively. However, this is a minor decrease and the results remain encouraging. Additionally, at this stage, the performance is being slowed down due to the automated isolation technique and may be enhanced by boosting the automated isolation tooth method's performance.

Table 5.2 Performance on each stage of caries.

Stage	Number of samples	Detection Performance on each class (%)
C0	218	89.47
C1	67	75.59
C2	189	88.89
C3	43	91.12
C4	16	100.00

Lastly, for each example of caries stage, caries detection on automated isolation teeth is determined. Table 5.2 provides a summary of the number of samples collected and the precision of the results for each stage of caries. The detection performance of C0 is indicative of the overall performance's specificity, which is the classification output for non-caries. The result of the caries classification is determined by separating the values C1 through C4 and averaging them to determine their sensitivity.

The best performance is observed in the C3 and C4 molars, with 91.12% and 100%, respectively, despite the fact that caries is typically rather evident in these teeth. In addition, C0 and C2 perform well with respective scores of 88.89% and 89.47%. In contrast, C1 has a puzzling classification performance of around 75%, indicating that approximately one-third of C1 cases were mistakenly categorized as non-caries. Although C1 value is not particularly promising, it is nevertheless used in caries screening. The primary issue may be the little damage in the C1 stage and the poor quality of radiographs. Consequently, the approach may perform better with additional high-quality data. In addition, due to the fact that individuals may experience discomfort in stages C3 and later, C2 is also effective for screening.

5.3   Conclusion

This section presented the automated caries detection method's implementation and validation. Both the tooth isolation approach and the caries detection method have been merged into a single, integrated system. On the basis of the decrease in the outcome metric, the performance and impact of manually isolated teeth and automatically isolated teeth were examined and explained. The varying contributions of each method to a strategy have been described. This part is a step forward in uniting all systems and making a significant contribution to research and medical practice through implementation.

The result reveals that the accuracy metric, which is used to measure overall performance, has suffered a moderate decline of about 3%. Accordingly, both sensitivity and specificity have experienced a reduction of around 2%, falling from 94.43% to 88.14%, and 3%, falling from 93.67% to 89.47%, respectively. The drop can be directly attributed to the relative ease with which teeth can be isolated. On the other hand, this is a relatively minor adjustment in light of the overall intricacy of the issue. As a consequence of this, the procedure is regarded as having a high potential for further development study and for the capability of being implemented in actual cases at dental hospitals.

Last but not least, the effectiveness of caries detection on each class in order to demonstrate the degree of difficulty associated with detection on each class. C1 and C2 are the very early stages of the disease, which should be given more attention and have been fairly diagnosed; nevertheless, the performance has been getting better in the more recent stages, which are C3 and C4. This provides an explanation for the performance

contribution that each phrase has made to the overall evaluation score, including accuracy, sensitivity, and specificity.

CHAPTER 6

Research Conclusion

6.1 Conclusion

This study aims to demonstrate that a computer can be used to aid in the process of separating individual teeth and finding cavities. Despite the fact that numerous researchers have attempted to address this issue, it remains unsolved. The vast majority of them are either extremely complex, requiring a large number of resources and placing a significant stress on the computer system, or excessively simple, unable to give an effective result. As a result, our proposed approach reduced the most complex processes while preserving the advantages of each component.

The objective of the study was achieved in terms of tooth isolation and the detection of caries. Oral panoramic radiography was utilised in conjunction with an automated tooth isolation technique to locate the tooth automatically. Due to its high degree of precision, which demonstrates a high level of detection accuracy, the method is practical and applicable for use by physicians. It appears that every aspect of the problem has been considered. However, the procedure might be considerably enhanced. It has been demonstrated that the outcomes of manually isolating teeth are superior to the previous state of the art in caries identification. An rise in the specificity

of the dataset indicates that the dataset and its results are well-balanced, consistent, and reliable. The automated combination technique, on the other hand, generates data that leads to a modest reduction in the total detection of caries. A component of the automated isolation dental system, relativity can be used to identify the weak. As a result, it may be useful to have a more comprehensive understanding of the situation in general.

6.2 Future Work

This study has two limitations: the automated tooth isolation reduced overall system performance due to an inaccuracy in tooth border identification, and the C1 stage of caries detection did not function as anticipated. As a result, future studies should investigate other strategies for achieving these two objectives. First, a Yolo-based network model should be examined in greater detail. Additionally, single-shot detection (SSD) [62], a competitor to Yolo, should be utilised to compare performance. Second, further image processing approaches should be researched for a more precise image processing that allows the deep learning model to gain a better knowledge of the issue and, as a result, make a higher improvement in C1 stage caries detection.

Bibliography

[1]   H. C. Gift and M. Redford, "Oral Health and The Quality Of Life," *Clinics in Geriatric Medicine,* vol. 8, no. 3, pp. 673-684, 1992/08/01/ 1992.

[2]   (Jan 26th, 2021). *Cavities*. Available: https://www.msdmanuals.com/en-jp/home/mouth-and-dental-disorders/tooth-disorders/cavities

[3]   (Oct 1st, 2020). *Oral health*. Available: https://www.who.int/health-topics/oral-health/

[4]   M. Rathee and A. Sapra, "Dental caries," *StatPearls [Internet],* 2021.

[5]   (Feb 27th, 2022). *Oral care*. Available: https://www.ufcgroup.jp/english/07oralcare/02.php

[6]   C. Mosquera-Lopez, S. Agaian, A. Velez-Hoyos, and I. Thompson, "Computer-Aided Prostate Cancer Diagnosis From Digitized Histopathology: A Review on Texture-Based Systems," *IEEE Reviews in Biomedical Engineering,* vol. 8, pp. 98-113, 2015.

[7]   R. F. Mansour, "Evolutionary Computing Enriched Computer-Aided Diagnosis System for Diabetic Retinopathy: A Survey," *IEEE Reviews in Biomedical Engineering,* vol. 10, pp. 334-349, 2017.

[8]   A. Sampathkumar, D. A. Hughes, K. J. Kirk, W. Otten, and C. Longbottom, "All-optical photoacoustic imaging and detection of early-stage dental caries," in *2014 IEEE International Ultrasonics Symposium*, 2014, pp. 1269-1272.

[9]   D. A. Hughes, J. M. Girkin, S. Poland, C. Longbottom, and S. Cochran, "Focused ultrasound for early detection of tooth decay," in *2009 IEEE International Ultrasonics Symposium*, 2009, pp. 1-3.

[10]   P. Usenik, M. Bürmen, A. Fidler, F. Pernuš, and B. Likar, "Near-infrared hyperspectral imaging of water evaporation dynamics for early detection of incipient caries," (in eng), *J Dent,* vol. 42, no. 10, pp. 1242-7, Oct 2014.

[11]   S. Li *et al.*, "Low-Shot Learning of Automatic Dental Plaque Segmentation Based on Local-to-Global Feature Fusion," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 664-668.

[12]   E. Maslak, B. Khudanov, D. Krivtsova, and T. Tsoy, "Application of Information Technologies and Quantitative Light-Induced Fluorescence for the Assessment of Early Caries Treatment Outcomes," in *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, 2019, pp. 912-917.

[13]   K. Angelino, D. A. Edlund, and P. Shah, "Near-Infrared Imaging for Detecting Caries and Structural Deformities in Teeth," (in eng), *IEEE journal of translational engineering in health and medicine,* vol. 5, pp. 2300107-2300107, 2017.

[14]   M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. Springer Science+Business Media New York: Springer, Boston, MA, 1993, pp. XIX, 555.

[15]   I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction* (no. Studies in Fuzziness and Soft Computing). Berlin, Heidelberg: Springer, 2006, p. 778.

[16]   T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[17]   J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming," in *Artificial Intelligence in Design '96*, J. S. Gero and F. Sudweeks, Eds. Dordrecht: Springer Netherlands, 1996, pp. 151-170.

[18]   J. Hu, H. Niu, J. Carrasco, B. Lennox, and F. Arvin, "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 12, pp. 14413-14423, 2020.

[19]   C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[20]   Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, no. 8, pp. 1798-1828, 2013.

[21]   Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015/05/01 2015.

[22]   J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks,* vol. 61, pp. 85-117, 2015/01/01/ 2015.

[23]   S. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica (Slovenia),* vol. 31, pp. 249-268, 01/01 2007.

[24]   E. Fix and J. L. Hodges, "Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties," *International Statistical Review,* vol. 57, p. 238, 1989.

[25]   N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician,* vol. 46, no. 3, pp. 175-185, 1992/08/01 1992.

[26]   G. I. Webb, "Naïve Bayes," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 713-714.

[27]   H. Tin Kam, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995, vol. 1, pp. 278-282 vol.1.

[28]   A. Irle and J. Kauschke, "On Kleinberg's Stochastic Discrimination Procedure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 33, pp. 1482-1486, 08/01 2011.

[29]   E. M. Kleinberg, "An overtraining-resistant stochastic modeling method for pattern recognition," *Annals of Statistics,* vol. 24, pp. 2319-2349, 1996.

[30]   E. M. Kleinberg, "On the algorithmic implementation of stochastic discrimination," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 5, pp. 473-490, 2000.

[31]   C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995/09/01 1995.

[32]   I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[33]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, 2012.

[34]   C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[35]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[36]   K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv 1409.1556,* 09/04 2014.

[37]   A. Gulli and S. Pal, *Deep Learning with Keras* (Packt Publishing). Packt Publishing, 2017.

[38]   F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800-1807.

[39]   C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826.

[40]   J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*. 2016, pp. 779-788.

[41] T. H. Bui, K. Hamamoto, and M. P. Paing, "Tooth Localization using YOLOv3 for Dental Diagnosis on Panoramic Radiographs," *IEEJ Transactions on Electronics, Information and Systems,* vol. 142, no. 5, pp. 557-562, 2022.

[42] B. Vijayakumari, R. R. Kirubalini, and C. R. Manisha, "Cadaver identification with dental radiographs using isoperimetric and nodal graph approach," *IET Biometrics,* https://doi.org/10.1049/iet-bmt.2019.0064 vol. 9, no. 1, pp. 38-45, 2020/01/01 2020.

[43] A. Radhiyah, T. Harsono, and R. Sigit, "Comparison study of Gaussian and histogram equalization filter on dental radiograph segmentation for labelling dental radiograph," in *2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC)*, 2016, pp. 253-258.

[44] O. Nomir and M. Abdel-Mottaleb, "Human Identification From Dental X-Ray Images Based on the Shape and Appearance of the Teeth," *IEEE Transactions on Information Forensics and Security,* vol. 2, no. 2, pp. 188-197, 2007.

[45] P.-W. Huang, P.-L. Lin, C.-H. Kuo, and Y. S. Cho, *An effective tooth isolation method for bitewing dental X-ray images*. 2012, pp. 1814-1820.

[46] F. P. Mahdi, K. Motoki, and S. Kobashi, "Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs," *Scientific Reports,* vol. 10, no. 1, p. 19261, 2020/11/06 2020.

[47] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data,* vol. 6, no. 1, p. 60, 2019/07/06 2019.

[48] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 02/23 2016.

[49] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision,* vol. 88, no. 2, pp. 303-338, 2010/06/01 2010.

[50] W. Li, W. Kuang, Y. Li, Y. Li, and W. Ye, "Clinical X-Ray Image Based Tooth Decay Diagnosis using SVM," in *2007 International Conference on Machine Learning and Cybernetics*, 2007, vol. 3, pp. 1616-1619.

[51] Y. Yu, Y. Li, Y.-J. Li, J.-M. Wang, D.-H. Lin, and W.-P. Ye, "Tooth Decay Diagnosis using Back Propagation Neural Network," 2006, vol. 2006, pp. 3956-3959.

[52] S. Patil, V. Kulkarni, and A. Bhise, "Intelligent system with dragonfly optimisation for caries detection," *IET Image Processing,* vol. 13, no. 3, pp. 429-439, 2019.

[53] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion," *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, vol. 26, pp. 732-739, 06/01 2004.

[54]   R. Lazcano *et al.*, "Porting a PCA-based hyperspectral image dimensionality reduction algorithm for brain cancer detection on a manycore architecture," *Journal of Systems Architecture,* vol. 77, pp. 101-111, 2017/06/01/ 2017.

[55]   R. Montefusco-Siegmund, P. E. Maldonado, and C. Devia, "Effects of ocular artifact removal through ICA decomposition on EEG phase," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 1374-1377.

[56]   W.-T. Pan, "A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example," *Knowledge-Based Systems,* vol. 26, pp. 69-74, 2012/02/01/ 2012.

[57]   S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software,* vol. 69, pp. 46-61, 2014/03/01/ 2014.

[58]   T. H. Bui, K. Hamamoto, and M. P. Paing, "Deep Fusion Feature Extraction for Caries Detection on Dental Panoramic Radiographs," *Applied Sciences,* vol. 11, no. 5, p. 2005, 2021.

[59]   R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. SMC-3, no. 6, pp. 610-621, 1973.

[60]   L. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 37, no. 2, pp. 780-795, 1999.

[61]   D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing,* vol. 28, no. 1, pp. 45-62, 2002/01/01 2002.

[62]   W. Liu *et al.*, *SSD: Single Shot MultiBox Detector*. 2016, pp. 21-37.

Author Biography

Author: Mr. BUI Huy Toan

Degree: Doctor of Engineering

Date of Graduation: September, 2022

Date of Birth: June 25th, 1993

Place of Birth: Phutho, Vietnam

Undergraduate and Graduate Education:

Master's Degree of Engineering in Computing in Engineering Systems,

King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 10520, 2018.

Bachelor's Degree in Information Technology,

Hanoi University of Science and Technology, Hanoi, Vietnam, 2016.

Presentations and Publications:

[1]   T. H. Bui, K. Hamamoto, and M. P. Paing, "Deep Fusion Feature Extraction for Caries Detection on Dental Panoramic Radiographs," Applied Sciences, vol. 11, no. 5, p. 2005, 2021.

[2]   T. H. Bui, K. Hamamoto, and M. P. Paing, "Tooth Localization using YOLOv3 for Dental Diagnosis on Panoramic Radiographs," IEEJ Transactions on Electronics, Information and Systems, vol. 142, no. 5, pp. 557-562, 2022.