ASSESSMENT OF CONSISTENCY DETECTION IN MANDIBULAR FIRST PREMOLARS ROOT CANAL MORPHOLOGY USING U-NET MODEL: EVALUATION WITH DICE COEFFICIENT AND INTERSECTION OVER UNION – A PILOT STUDY

MELISSA WONG LI ZHENG

FACULTY OF DENTISTRY

UNIVERSITY OF MALAYA

KUALA LUMPUR

2024

ASSESSMENT OF CONSISTENCY DETECTION IN MANDIBULAR FIRST PREMOLARS ROOT CANAL MORPHOLOGY USING U-NET MODEL: EVALUATION WITH DICE COEFFICIENT AND INTERSECTION OVER UNION – A PILOT STUDY

MELISSA WONG LI ZHENG

RESEARCH REPORT SUBMITTED IN

PARTIAL FULFILMENT OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF ORAL SCIENCE

FACULTY OF DENTISTRY

UNIVERSITY OF MALAYA

KUALA LUMPUR

2024

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: MELISSA WONG

Registration/Matric No: 22079982

Name of Degree: Master of Oral Science (Restorative Dentistry)

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"): Assessment of Consistency Detection in Mandibular First Premolars Root Canal Morphology Using U-Net Model: Evaluation with Dice Coefficient and Intersection Over Union – A Pilot Study.

Field of Study: Endodontics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

24/9/2024

Subscribed and solemnly declared before,

Witness's Signature Date

24/9/2024

Name:
Designation:

ASSESSMENT OF CONSISTENCY DETECTION IN MANDIBULAR FIRST PREMOLARS ROOT CANAL MORPHOLOGY USING U-NET MODEL: EVALUATION WITH DICE COEFFICIENT AND INTERSECTION OVER UNION – A PILOT STUDY

ABSTRACT

This study assesses the consistency of the U-Net model in pulp space segmentation of extracted mandibular first premolars (MFPs) using CBCT images. Five training, validation, and testing datasets were randomly generated from a pool of 130 CBCT images. Mean dice coefficient (DC) and Intersection over Union (IoU) scores were measured and compared across these datasets. The CBCT images were loaded into ITK-SNAP software (Version 4.0.2) for semi-automatic segmentation, performed by a postgraduate student (MW) with an ICC of 0.95. The datasets were split into five different training (70%), validation (20%), and testing (10%) sets from the same data pool. Image augmentation was applied, and the files were resized for the U-Net model. The mean DC and IoU scores for the training, validation, and testing datasets were compared using one-way ANOVA. This was followed by a post-hoc LSD test for multiple comparisons between the groups. A p-value < 0.05 was considered statistically significant. The mean DC from datasets 1 and 5 showed no significant difference during training, validation and testing phase with the p value of 0.32, 0.20, and 0.06 respectively. Similarly, the mean IoU from datasets 1 and 5 showed no significant difference during training, validation, and testing phase with the p value of 0.14, 0.23 and 0.12 respectively. In the training phase, the mean DC for dataset 2 (0.39 ± 0.04) was significantly higher than dataset 1, 3, 4 and 5 (p = 0.00). Similarly, during the testing phase, the mean DC for dataset 2 (0.34 ± 0.05) was significantly higher than dataset 1, 4 and 5 (p = 0.00), and dataset 3 (p = 0.02). The similar p-values for datasets 2 compared to datasets 1, 4 and 5 suggest that the performance differences between these datasets are consistent and significant in both the training and testing phases. The difference in p-values between datasets 2 and 3 during

training and testing indicates that the performance gap was more pronounced during training than testing phases. The mean IoU for dataset 2 (0.40 ± 0.07) performed significantly higher than dataset 1 (0.17 ± 0.08, p = 0.00), 3 (0.26 ± 0.10, p = 0.00), 4 (0.32 ± 0.06, p = 0.04) and 5 (0.22 ± 0.09, p = 0.00) during the training phase. Comparably, the mean IoU during the testing phase for dataset 2 (0.35 ± 0.05) was significantly higher than dataset 1 (0.13 ± 0.05, p = 0.00), 3 (0.27 ± 0.07, p = 0.01), 4 (0.17 ± 0.08, p = 0.00) and 5 (0.18 ± 0.06, p = 0.00). The findings indicate that the consistency of the U-Net model in pulpal space segmentation of MFPs was affected despite the five datasets being randomly generated from the same data pool. When the predicted and ground truth images overlap completely, the DC and IoU value is 1. In the present study, the overall mean DC and IoU values across all phases were below the ideal value of 1, which indicates that the U-Net model performance in pulp segmentation is less consistent in the assessment of MFP.

Keywords: Pulp space segmentation, U-Net, dice coefficient, intersection over union, mandibular first premolars

PENILAIAN KETEKALAN PENGESANAN MORFOLOGI SALUR AKAR GIGI GERAHAM KECIL PERTAMA MANDIBULAR MENGGUNAKAN MODEL U-NET: PENILAIAN DENGAN PEKALI DICE DAN PERSILANGAN ATAS PERSATUAN – SATU KAJIAN PERINTIS

ABSTRAK

Kajian ini menilai konsistensi model U-Net dalam segmentasi ruang pulpa pada gigi geraham kecil pertama mandibular (MFP) yang diekstrak menggunakan imej CBCT. Lima set data latihan, validasi, dan ujian telah dijana secara rawak daripada kumpulan 130 imej CBCT. Koefisien dadu min (DC) dan skor Intersection over Union (IoU) diukur dan dibandingkan merentas set data ini. Imej CBCT dimuatkan ke dalam perisian ITK-SNAP (Versi 4.0.2) untuk segmentasi separa automatik, yang dilakukan oleh seorang pelajar pascasiswazah (MW) dengan ICC sebanyak 0.95. Set data dibahagikan kepada lima set latihan (70%), validasi (20%), dan ujian (10%) yang berbeza daripada kumpulan data yang sama. Augmentasi imej digunakan, dan fail diubah saiz untuk model U-Net. Min skor DC dan IoU untuk set data latihan, validasi, dan ujian dibandingkan menggunakan ANOVA satu hala. Ini diikuti oleh ujian LSD post-hoc untuk perbandingan berbilang antara kumpulan. Nilai $p < 0.05$ dianggap signifikan secara statistik. Min DC daripada set data 1 dan 5 menunjukkan tiada perbezaan yang signifikan semasa fasa latihan, validasi dan ujian dengan nilai p masing-masing 0.32, 0.20, dan 0.06. Begitu juga, min IoU daripada set data 1 dan 5 menunjukkan tiada perbezaan yang signifikan semasa fasa latihan, validasi, dan ujian dengan nilai p masing-masing 0.14, 0.23 dan 0.12. Dalam fasa latihan, min DC untuk set data 2 (0.39 ± 0.04) adalah lebih tinggi secara signifikan daripada set data 1, 3, 4 dan 5 ($p = 0.00$). Begitu juga, semasa fasa ujian, min DC untuk set data 2 (0.34 ± 0.05) adalah lebih tinggi secara signifikan daripada set data 1, 4 dan 5 ($p = 0.00$), dan set data 3 ($p = 0.02$). Nilai p yang serupa untuk set data 2 berbanding set data 1, 4 dan 5

mencadangkan bahawa perbezaan prestasi antara set data ini adalah konsisten dan signifikan dalam kedua-dua fasa latihan dan ujian. Perbezaan nilai p antara set data 2 dan 3 semasa latihan dan ujian menunjukkan bahawa jurang prestasi lebih ketara semasa fasa latihan berbanding ujian. Min IoU untuk set data 2 (0.40 ± 0.07) menunjukkan prestasi yang lebih tinggi secara signifikan daripada set data 1 (0.17 ± 0.08, p = 0.00), 3 (0.26 ± 0.10, p = 0.00), 4 (0.32 ± 0.06, p = 0.04) dan 5 (0.22 ± 0.09, p = 0.00) semasa fasa latihan. Sebanding, min IoU semasa fasa ujian untuk set data 2 (0.35 ± 0.05) adalah lebih tinggi secara signifikan daripada set data 1 (0.13 ± 0.05, p = 0.00), 3 (0.27 ± 0.07, p = 0.01), 4 (0.17 ± 0.08, p = 0.00) dan 5 (0.18 ± 0.06, p = 0.00). Penemuan ini menunjukkan bahawa konsistensi model U-Net dalam segmentasi ruang pulpa MFP terjejas walaupun lima set data dijana secara rawak daripada kumpulan data yang sama. Apabila imej ramalan dan ground truth bertindih sepenuhnya, nilai DC dan IoU adalah 1. Dalam kajian ini, nilai min keseluruhan DC dan IoU merentas semua fasa adalah di bawah nilai ideal 1, yang menunjukkan bahawa prestasi model U-Net dalam segmentasi pulpa kurang konsisten dalam penilaian MFP.

Kata kunci: Segmentasi ruang pulpa, U-Net, koefisien dice, intersection over union (IoU), gigi geraham kecil pertama mandibular

ACKNOWLEDGEMENTS

Firstly, I would like to extend my heartfelt gratitude to my supervisors, Associate Professor Dr. Mariam Abdullah, Dr. Nora Sakina Mohd Noor, and Dr. Amirul Fahmi Che Mohd Fadzilah. Their unwavering guidance, encouragement, and invaluable support have been instrumental throughout the various stages of this research. Their expertise and insightful feedback have significantly contributed to this study, and I am deeply appreciative of their dedication and mentorship.

I would like to send my sincere appreciation to Mr. Bryan Wong, whose expertise as a data engineer was invaluable to this project. He played a crucial role in data processing, implementing the U-Net architecture, and overseeing the training, validation, and test processes. His technical skills and dedication significantly contributed to the successful execution of this study.

Special thanks to Dr. Alias for his invaluable insights throughout this study, particularly in the statistical analysis part.

I also wish to thank my peers and friends for their unwavering support, motivation, and insightful guidance throughout my study.

Finally, I would like to convey my deepest gratitude to my family for their support and patience throughout this process. Their constant encouragement and understanding have been a source of strength for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

3D	:	Three dimensional
AI	:	Artificial intelligence
ALARA	:	As low as reasonably achievable
ANN	:	Artificial neural network
BraTS	:	Brain Tumour Segmentation
CBCT	:	Cone beam computed tomography
CNN	:	Convolutional neural network
CT	:	Computed tomography
DC	:	Dice coefficient
DEC	:	Digital equipment corporation
DICOM	:	Digital imaging and communications in medicine
DL	:	Deep learning
GAN	:	Generative adversarial network
GPU	:	Graphic processing unit
IoU	:	Intersection over Union
M	:	Mean
MFPs	:	Mandibular First Premolars
ML	:	Machine learning
MNIST	:	Modified national institute of standards & technology
MONAI	:	Medical open network for artificial intelligence

MRI	:	Magnetic resonance imaging
NIfTI	:	Neuroimaging informatics technology initiative
NNs	:	Neural networks
RAM	:	Random access memory
ReLU	:	Rectified linear unit
ROI	:	Region of interest
SD	:	Standard deviation
YOLOv5x	:	You only look once v5

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Mandibular first premolars (MFPs) pose significant challenges in endodontic treatment due to their diverse anatomical variations, including multiple root canals, and the presence of root canal system bifurcations as indicated by previous research (Bürklein et al., 2017; Liu et al., 2013). These complexities necessitate a thorough understanding and identification subsequently meticulous procedures for root canal preparation and filling. This careful approach is essential to achieve a high success rate in root canal treatments (Aung & Myint, 2021). Misidentification of root canal system variations has been associated to 42% of unsuccessful endodontic treatments as reported in a study, ultimately compromising the success of the procedure (Hoen & Pink, 2002).

Cone beam computed tomography (CBCT) has attained significant popularity as a method for three-dimensional (3D) visualisation since early year 2000 (Orhan et al., 2020). It has the ability to yield detailed 3D images without the presence of artifacts and overlapping of bone and other dental structures which are typically observed in conventional radiographs (Patel et al., 2019). Numerous scientific investigations have evaluated the diagnostic precision of CBCT and have consistently shown that CBCT significantly enhances the identification of root canal systems. (Celikten et al., 2016; Madani et al., 2017; Rahimi et al., 2017; Torres et al., 2015; Zhang et al., 2015). However, reconstructing the dental pulp into a 3-D structure from CBCT images heavily relies on the clinical experience of dentists as the images are in two-dimensional slice-based views that lack the depth, and may appear blur.

Expert manual segmentation with CBCT is commonly considered as the “gold standard” for comparison due to its assurance of high-quality annotations (Wenger et al.,

2014). Nevertheless, this approach is depicted by its high labour demands and vulnerability to variability among different practitioners, stemming from varying interpretations, level of expertise, and segmentation tools employed (Millioni et al., 2010). Therefore, numerous developers in biomedical imaging have endeavoured to integrate automated pulp segmentation with machine learning, recognising its advantages in speed and efficiency over manual segmentation methods (Haque & Neubert, 2020).

Deep learning (DL), a subset of machine learning, has significantly revolutionised the field of image segmentation. This process encompasses splitting an image into numerous segments to refine its representation and enhance analysis. DL achieves human-like thinking by utilizing multi-layer Artificial Neural Networks (ANNs), which is similar to the intricate structures of the human brain (LeCun et al., 2015). An example of multi-layer ANNs is Convolutional Neural Networks (CNNs), a division of deep learning models, which have been notably successful in image segmentation functions (Jena et al., 2018). These models learn to automatically extract features from images and classify each pixel or region into predefined classes (Yamashita et al., 2018). The learning process in DL can be supervised, which is the most prevalent form, or it can operate with partial supervision or no supervision (Haque & Neubert, 2020).

U-Net is a prominent CNN architectures for segmentation, which was introduced by Ronneberger in 2015 (Ronneberger et al., 2015). The U-Net comprises a contracting path (encoder) to capture contextual information and a symmetric expanding path (decoder) that facilitates accurate localisation, and the architecture resembles a U shape, hence its name (Fig 1.1). The contracting path features 3x3 convolutional and pooling layers that systematically lower the spatial resolution and enhance the number of channels to capture the context of the input image. Within the expanding path, up-sampling is utilized to accurately localize the context, progressively recovering spatial resolution, and diminishing the number of channels. Skip connections linking the contracting and

expanding paths help the network maintain high-resolution features from the contracting path, improving segmentation accuracy. In the last layer, 1x1 convolution is utilised to create a segmentation map, where each pixel is assigned to different classes or categories. This architecture is particularly well-suited for biomedical image segmentation where the available dataset is limited, and the segmentation accuracy is of utmost importance (Ronneberger et al., 2015).

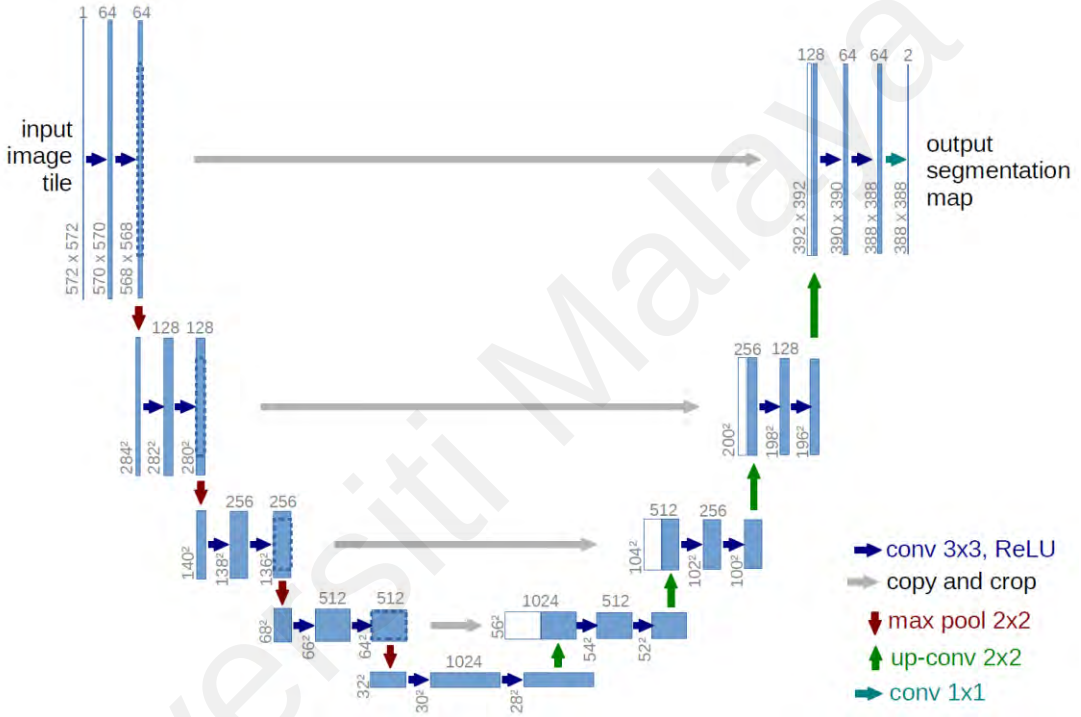

Figure 1. 1: The U-Net architecture

The foundational process for validating the accuracy of U-Net in biomedical segmentation depends on the training datasets that comprise of the original images or data (CBCT image), paired with their segmented target outputs (ground truth). These datasets are crucial in enabling the U-Net model to comprehend the complex relationships throughout the input images and their corresponding segmented ground truth during the training and validation phase. Subsequently, the model's accuracy is tested using the testing datasets. These datasets are typically randomly divided into training, validation and testing datasets, with implementation of augmentation to the training and validation

datasets to prevent overfitting of the model. Overfitting occurs when the model learns the details of a training data too well, resulting in high accuracy on the training dataset but poor performance on a new, unseen data (Jabbar & Khan, 2015). Hence, this makes the model less efficient in real-world applications, as it fails to adapt to variations and complexities in new data, undermining its reliability in clinical settings. The Dice coefficient (DC), commonly employed in image segmentation is a statistical measure used to measure the similarity between the predicted and the ground truth, quantifying the overlap between the predicted and ground truth regions as illustrated in Figure 1.2 (Zou et al., 2004). Intersection over Union (IoU) is defined as the ratio of the intersection to the union of the predicted and ground truth regions, assessing the accuracy of an image segmentation model seen in Figure 1.2 (Rahman & Wang, 2016).

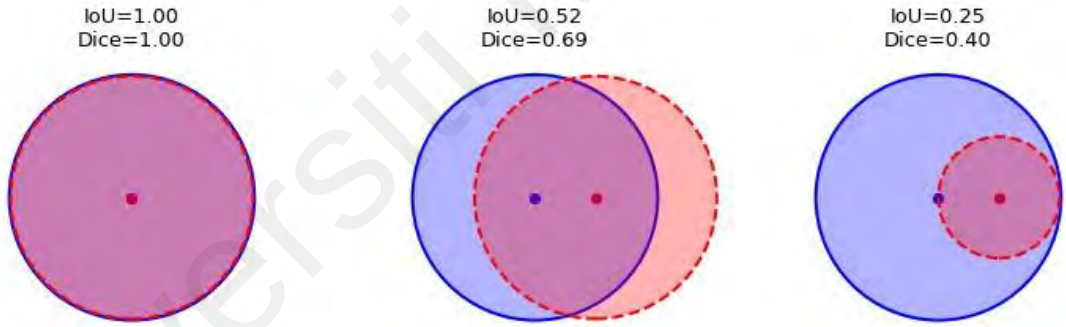

Figure 1. 2: Examples of DC and IoU representation (Monteux, 2019)

Numerous studies have been conducted to investigate on the accuracy of the model by using various type of encoder and decoder pathway in U-Net model in brain tumour segmentation (Aboussaleh et al., 2023; Long et al., 2021) , different variation of U-Net model in detecting and classifying diverse c-shape canals in mandibular first molars (Sherwood et al., 2021) or even the different type of augmentation on the training and validation datasets (Cirillo et al., 2021).

Currently, there is a research gap in applying U-Net to detect pulpal space morphology in different types of teeth. No studies have compared the DC and IOU across different training, validation, and testing datasets, generated randomly from the same data pool. This lack of comparison raises a research question regarding the consistency of the U-Net model when applied to pulpal segmentation.

1.2 Aim of study

The aim of this pilot study is to assess consistency of U-Net model in pulp space segmentation on extracted MFPs using CBCT images with randomly generated training, validation, and testing datasets from the same data pool.

1.3 Objectives

1. To compare DC of U-Net in pulp space segmentation of MFPs with five sets of randomly generated training, validation, and testing datasets from the same data pool during their respective phases.
2. To compare the IoU of U-Net in pulp space segmentation of MFPs with five sets of randomly generated training, validation, and testing datasets from the same data pool during their respective phases.

1.4 Null hypothesis

1. There is no significant difference in DC of U-Net model in pulp space segmentation on MFPs among all five randomly generated training, validation, and testing datasets from the same data pool during their respective phases.

2. There is no significant difference in IoU of U-Net model in pulp space segmentation on MFPs among all five randomly generated training, validation, and testing datasets from the same data pool during their respective phases.

CHAPTER 2: LITERATURE REVIEW

2.1 Anatomical variation in MFPs pulpal morphology

The dental pulp is anatomically bounded by dentine located within the tooth, usually not visible during clinical examination. The pulpal space usually seen as a radiolucent area within the tooth during radiographic imaging. MFPs are generally challenging to manage and have a high rate of complications and failures, largely due to differences in their root canal morphology (England et al., 1991; Slowey, 1979). Multiple studies over the past twenty years have reported that MFPs often have more than one canal (Hülsmann, 1990; Martins et al., 2017; Pan et al., 2019; Pineda & Kuttler, 1972; Rahimi et al., 2007; Vertucci, 1984; Yoshioka et al., 2004; Zillich & Dowson, 1973). Some studies have demonstrated that the prevalence of two canals in mandibular premolars ranges from 12.3% to 34% (Hülsmann, 1990; Martins et al., 2017; Moayedi & Lata, 2004; Pan et al., 2019; Pineda & Kuttler, 1972; Vertucci, 1984; Yoshioka et al., 2004; Zillich & Dowson, 1973), while the prevalence of three canals ranges from 0.5% to 4.3% (Hülsmann, 1990; Moayedi & Lata, 2004; Rahimi et al., 2007; Vertucci, 1984). (Zillich & Dowson, 1973) reported that second or third canals existed in at least 23% of MFPs. Another study has reported that more than one fifth (22.1%) of MFPs showed two or more root canals (Bürklein et al., 2017).

The root canal morphology exhibited a large variation, with Vertucci type I being 21.9% and type V being 55.7% in majority of cases (Bürklein et al., 2017). Another study conducted showed the canal classification of MFPs with Vertucci type I being the most common (65.2%), followed by type V (22.6%), type III (2.6%) and type VII (0.9%) (Liu et al., 2013). They also observed that a single apical foramen was present in 50.4% of its samples, while two or three foramens were observed in 28.7% and 14.8% of cases, respectively (Liu et al., 2013). Many studies have indicated that there appears to be a

racial predisposition for the presence of two or more canals in maxillary and mandibular premolars (Bürklein et al., 2017; Liu et al., 2013; Pan et al., 2019; Rahimi et al., 2007).

2.2 Impact of accurate identification of pulpal anatomy on root canal treatment success

Variation in root canal morphology is now considered common, and it is widely accepted that a primary cause of endodontic failure is the inability to locate, clean, shape and obturate all the root systems' canals in three dimensions. A study done in the University of Washington reported the failure rate of nonsurgical root canal treatment in all teeth and found that MFPs has the highest failure rate of 11.45% (Robertson et al., 1980). Failure to properly identify anatomical variations in root canals can result in endodontic treatment failure (Poorni et al., 2010). Missed roots and/or canals will results in 42% of failed nonsurgical root canal treatment cases (Hoen & Pink, 2002). Therefore, it is crucial to fully understand the morphology of root canal systems and their variations to achieve successful therapy, eliminate infection, and protect the infected tooth from further microbial invasion.

2.3 Limitation of conventional radiographs in evaluating the anatomic pulpal morphology

High quality radiograph is crucial for identifying the internal and external morphology of the root canals. To date 2-dimensional (2D) conventional and digital intraoral radiography are commonly used in the field of endodontics. However, radiographs only produce a 2-dimensional (2D) image of a 3-dimensional (3D) object, which limits the detailed information that can be obtained. The diagnostic value of these images is influenced by factors such as beam angulation, superimposition of anatomical structures and patient-related factors (Shearer et al., 1991). As a result, they have limited

value in cases with complex root canal anatomy. Accurate visualisation of the 3D dental pulp is crucial for assisting in clinical decisions and treatment planning before nonsurgical endodontic treatment.

2.4 CBCT imaging in pulpal segmentation and its limitations

Cone Beam Computed Tomography (CBCT) has gained its popularity in the field of dentistry for 3D imaging. CBCT employs a rotating arm and a cone-shaped ionizing radiation source to capture images using an X-ray source and detector (Durack & Patel, 2012). This technology provides fast and accurate 3D imaging by capturing multiple cross-sectional images, which contributes and facilitates to the accuracy of pulpal segmentation. The American Association of Endodontists recommended the use of CBCT imaging for the initial treatment of teeth with suspected complex root canal morphology and the potential for extra canals, particularly in the mandibular premolars (Fayad & Johnson, 2023). However, it is crucial to adhere to the ALARA (As Low As Reasonably Achievable) principle, which advocates for minimizing radiation exposure during imaging procedures. Hendee & Edwards (1986) concluded that CBCT is valuable in identifying complex root morphologies, including multiple roots, root fusion, and accessory canals, which are frequently missed on conventional radiographs

Image segmentation is defined as the partitioning of an image into non-overlapping, constituent regions that are homogenous with respect to some characteristics such as intensity or texture (Pal & Pal, 1993). Segmentation can be conducted through manual, automatic, and semi-automatic techniques. The manual method, where structures are outlined slice by slice by an operator is the most commonly used approach. Numerous studies have incorporated manual segmentation as a baseline for comparison (Forst et al., 2014; Liu et al., 2010; Wang et al., 2011). Nonetheless, this method is subjective, reliant on user's experience and skills, and is both time-consuming and tedious. Achieving

precise segmentation of individual teeth and their substructures remains challenging due to several factors. Metal artifacts from orthodontic treatments, indirect restorative materials, or calcification can hinder manual segmentation by causing scattering in CBCT images (Salineiro et al., 2019; Sanders et al., 2007). Moreover, the internal boundaries of tooth substructures often display similarities, such as low contrast at the dentine-pulp interface, further complicating the segmentation process (Arifin et al., 2019). Additionally, the diverse and intricate morphology of dental pulp can also affect the accuracy of pulp segmentation (Duan et al., 2021; Kakehbaraei et al., 2018).

Threshold-based segmentation methods appear to be the most common for segmenting tooth pulp cavities in CBCT imaging due to the low contrast or difference in grey levels between the pulp cavity and dentine tissue (Wang et al., 2019). However, this method does not ensure accurate and robust results, as it relies on manually selecting a single tooth region for further processing (Duan et al., 2021). Additionally, it performs poorly when the intensity of the pulp cavity differs significantly from the background or when the pulp cavity region is very small especially in the apical third region (Wang et al., 2019).

2.5 Artificial intelligence (AI)

2.5.1 History of AI

The concept of AI was formally founded around 1950s. Alan Turing issued "Computing Machinery and Intelligence", where it explains whether a machine is able to display intelligent behaviour different from that of a human, naming it the Turing Test (Turing, 1980). During the assessment, a blinded human evaluator will interact with both human and a machine on separate occasions. The conversation is limited to plain text to focus on logical responses rather than speech interpretation. The machine is considered

to demonstrate human-like intelligence if the evaluator is unable to differentiate between machine and human (Turing, 1980).

In 1955, the terminology AI was introduced during a two-month workshop on AI led by a few computer scientist (McCarthy et al., 2006). However, the idea remained theoretical due to several limitations in the 1950s due to the lack of computer capability in storage, which means codes could only be executed and not saved. Additionally, computers during that time were prohibitively expensive. Finally, funding sources were hesitant to invest in this emerging field (Tatnall, 2012).

From 1957 to 1974, AI rapidly advanced due to increased computer power, accessibility, and improved algorithms. A notable example is ELIZA, a chatbot therapist, which was named after a fictional character named Eliza Doolittle, interprets the questions in any language used and answers through words (Weizenbaum, 1966). However, two "AI Winters" occurred during the mid-1970s to the late 1980s due to a lack of practical applications and reduced funding (Hendler, 2008). Despite this drawback, AI saw breakthroughs, particularly in 1980s, through machine learning (ML) and expert systems (Schmidhuber, 2015). ML enables machine to learn through practice while expert systems emulate human's critical thinking for problem solving (Liebowitz, 1995). An example of an expert system is the R1(Xcon) program, which helped Digital Equipment Corporation (DEC), a computer manufacturer to select computer assembly components using about 2500 rules (McDermott, 1980).

With the further advancement of Graphic processing units (GPU) available in the market, the field of computer vision significantly improved leading to the development of an eight layered DL network in 2012 (Krizhevsky et al., 2012). This creation won an award due to its low error margin of 15.3%, which is 10.8% better than the runner up. The model continues to improve and won in 2017 with a further reduction in error to 2.25%, which is half of the human error rate 5.1% (Russakovsky et al., 2015).

2.5.2 Classification of AI

A study has developed various methods to classify AI due to its diverse applications and various types (Sarker, 2022). AI is a broad term that represents all forms of non-human intelligence, which can be divided further into weak AI and strong AI. The majority of AI today is categorized as weak AI or narrow AI, as it involves software designed to address individual or specific tasks. Strong AI is defined as the level of AI intelligence equivalent to human intelligence, possessing self-awareness and behaviour as adaptable as humans (Flowers, 2019). However, there are currently no applications of strong AI due to ethical concerns and the potential dangers it could pose (Evgrafova et al., 2022).

Weak AI can be further divided into ML and expert systems. ML can also be divided into supervised, semi-supervised, and non-supervised learning based on the analytical methods. Supervised learning relies on labeled input for training, allowing the model to learn the algorithm with these labeled input as "supervisors". The model learns by analyzing the labeled input, identifying common features, and using this knowledge to predict outcomes for unlabeled input (Hastie et al., 2009). In contrast, non-supervised learning independently identifies various features in unlabeled data (Hastie et al., 2009). Semi-supervised learning falls between the two, using a lesser amount of labeled features and an extensive amount of unlabeled features during training phase (Zhu & Goldberg, 2022). A novel approach known as weakly-supervised learning has gained popularity since 2018 to reduce labeling expenses. Specifically, for object segmentation task, it uses image-level labels (i.e., knowing what objects are present in the images) instead of detailed object boundary or location information for training (Zhou, 2018). ML can also be classified into supervised, semi-supervised, and unsupervised learning based on the theoretical methods. DL is currently a subset of ML and is a highly prominent research area (Sarker, 2021). It can be applied in both supervised and unsupervised contexts. A

deep neural network can extract features from imported data without human intervention. It learns these features from large datasets. On the contrary, expert systems require human intervention to learn, hence less data is required compared to deep neural networks.

Neural networks (NNs) are computational models inspired by the human neural structure and function. They consist of interconnected layers of nodes, or "neurons", which process and transmit information. These networks learn from data by adjusting the connections (weights) between neurons based on the input they receive, enabling them to recognize patterns, make decisions, and perform complex tasks. NN from the basis of many DL algorithms, with variations such as artificial neural networks (ANN), convolutional neural networks (CNN), and generative adversarial networks (GAN) being used for different applications, from image recognition to natural language processing and more.

2.5.2.1 Artificial Neural Network (ANN)

This model is a fundamental framework for deep learning, consisting of at least three layers: an input layer, one or more hidden layers, and an output layer. In this model, inputs are processed only in the forward direction. Input neurons extract features from the input data and send it to the hidden layers. The data passes through each hidden layer successively, where all hidden layers weigh the data received from previous layers and adjust before sending it to the next layer. Finally, the results are compiled and presented in the output layer. Each hidden layer acts as both an input and output layer, enabling the ANN to understand more complex features (Agatonovic-Kustrin & Beresford, 2000).

2.5.2.2 Convolutional Neural Network (CNN)

CNN is a DL model primarily used for image recognition and generation. The main difference between ANN and CNN is that CNN includes convolutional layers, along with pooling layers and fully connected layers within its hidden layers, making it widely used in the field computer vision. Convolution layers create feature maps from input data using convolutional kernels. These kernels move across the input image, effectively folding it (Wang et al., 2021). This process simplifies the images by sharing weights during the convolution operation. The pooling layer typically follows each set of convolution layers, reducing the dimensions of the feature maps to facilitate further extraction (Wang et al., 2021). The fully connected layer, which comes after the convolution and pooling layers, connects to all activated neurons in the previous layer. It transforms the 2D feature maps into a one dimensional format, which is then associated with nodes representing different categories for classification (LeCun et al., 2015; Nam, 2020). Due to its intricate architecture, CNN is known for having higher efficiency and accuracy in image recognition compared to ANN.

2.5.2.3 Generative Adversarial Network (GAN)

Developed by Goodfellow and colleagues in 2014, through an unsupervised learning approach that aims to automatically identify patterns in input data and generate new data with similar features or patterns as the input data (Goodfellow et al., 2014). In GAN, there are two key components: the generator and the discriminator. The generator's objective is to create data that is indistinguishable from the original data, confusing the discriminator. Conversely, the discriminator purpose is to correctly differentiate between the real data and the artificially generated data by the generator. This adversarial setup

forces both networks to improve, resulting in the generator producing more realistic data over time (Ravanbakhsh et al., 2019).

GAN technology has rapidly expanded its presence in AI applications, particularly in tasks like image-to-image translation and the generation of realistic images of objects, scenes, and people (Aggarwal et al., 2021; Gui et al., 2020). A study introduced a novel 3D-GAN framework, which builds upon traditional GAN but is tailored for generating 3D objects within a 3D space (Wu et al., 2016). This approach combines advancements in GAN and volumetric convolutional networks, enabling the direct generation of 3D objects or their transformation from 2D images. This capability significantly broadens the scope of applications for processing 3D data compared to conventional 2D GAN models (Han et al., 2019).

2.6 AI in enhancing biomedical image analysis

AI has revolutionized biomedical image segmentation by offering advanced algorithms and techniques to improve accuracy and efficiency. Image segmentation is a critical step in medical image analysis, enabling the delineation of structures for diagnosis, treatment planning, and research. AI techniques, particularly DL models like CNN, have shown remarkable success in segmenting biomedical images, including those from magnetic resonance imaging (MRI), computed tomography (CT), and histopathology samples (Almeida & Tavares, 2021; Roy et al., 2019). This concept forms the foundation of many DL algorithms where the models/networks comprise of numerous layers that process input data, i.e., images to produce outputs i.e., presence or absence of a disease while learning progressively more complex features. CNN have been the most successful type of model for image analysis thus far (Wang et al., 2021).

CNN works have been in existence since the 1980s (Fukushima, 1980), and started in biomedical image analysis in 1995 through work by Lo et al. (1995). The first

successful real-life application of CNN was demonstrated through a model developed by Lecun et al. (1998) called LeNet in the 1990s. It was designed for handwritten digit recognition tasks, specifically for recognizing digits in the Modified National Institute of Standards and Technology (MNIST) dataset (Lecun et al., 1998). While CNNs showed promising results initially, their widespread adoption was hindered until various new techniques were developed for effectively training deep networks and significant advancements were made in computing systems. This progress led to creation of ImageNet and its improved version, AlexNet during the early 2010s, which played a crucial role in popularizing CNNs for image analysis tasks making it the technique of choice in computer vision (Krizhevsky et al., 2012). Litjens et al. (2017) have reported that CNNs have a wide range of applications in image analysis, including lesion detection, organ segmentation, and lesion segmentation. Recent studies have highlighted advancements in diagnostic accuracy, including the early detection of Alzheimer's disease using MRI data (Pan et al., 2020; Sarraf & Tofighi, 2016; Suk et al., 2017), analysing tumours with multiscale techniques (Zhao & Jia, 2016), segmenting the striatum (Choi & Jin, 2016), and detecting and classifying tumours and lesions (Ghafoorian et al., 2017).

2.7 AI in dentistry

Inspired by the revolutionary impact of AI in various industries, its growth in dentistry has been substantial in recent years. In the dental sector, AI applications can be divided into several categories: diagnosis, decision-making, treatment planning, and predicting treatment outcomes. Among these, the use of AI for diagnostic purposes is the most prevalent. Many reports indicate that AI can provide more accurate and efficient diagnoses, thereby reducing the workload of dental practitioners (Ding et al., 2023). This efficiency enables AI research to extend across all dental specialties along with the

improvement in computing technology. To achieve this goal, interdisciplinary cooperation between computer scientists and clinicians is essential. Clinicians are responsible for manually labelling the regions of interest (ROI) on radiographic images, while computer scientist prepare the dataset and develop the DL algorithm (Sherwood et al., 2021).

2.7.1 AI in operative dentistry

Dental caries is typically diagnosed by dentists via visual and tactile examination or by radiographic imaging. However, detecting early carious lesions can be challenging, and they may go undetected, resulting in more severe stages of dental caries. For dental caries to be visible on a radiograph, 30-40% demineralization must occur (Langland et al., 2002). This shortcoming can lead to more complex and expensive treatments, such as root canal therapy, indirect dental restorations or even extraction followed by implants. In a 2D radiograph, each pixel in the grayscale image has an intensity, or brightness, that represents the object's density. By analysing these characteristics, an AI algorithm can learn the patterns and provide predictions for tooth segmentation and caries detection. A study conducted has successfully developed a CNN algorithm for detecting dental caries with periapical radiographs improving diagnostic consistency (Lee et al., 2018). Another study suggested a CNN algorithm for detecting caries in intraoral images (Kühnisch et al., 2022). A comparison of the cost effectiveness of AI for detecting interproximal caries with dentist's diagnoses showed that AI yielded significantly superior results (Schwendicke et al., 2021). With its accuracy comparable to or even surpassing that of dentists in detecting early carious lesions, AI plays an important role in preventive dentistry by enabling early intervention.

2.7.2 AI in endodontics

The rapid advancement in AI in the field of endodontics highlight its potential to significantly enhance patient care. In particular, the development of DL has led to notable improvements in both diagnostic accuracy and treatment planning (Khanagar et al., 2023). For example, DL have shown outstanding results in tasks like detection of periapical lesions (Ekert et al., 2019; Endres et al., 2020; Pauwels et al., 2021), vertical root fractures (Fukuda et al., 2020; Talwar et al., 2016) and external root resorption (Reduwan et al., 2024) via CNN architectures, as well as determining the working length of the tooth (Saghiri, Asgar, et al., 2012; Saghiri, Garcia-Godoy, et al., 2012) by ANN. In addition, previous studies also demonstrated that AI has been employed in predicting the tooth prognosis and cases that may require retreatment (Campo et al., 2016; Lee et al., 2022), assessing stem cell vitality (Bindal et al., 2017), and detecting root canal morphologies (Hiraiwa et al., 2019).

A study has demonstrated that the system achieved high level of accuracy of 87% when diagnosing between single and multiple roots in the distal roots of mandibular first molars by analysing panoramic radiographs (Hiraiwa et al., 2019). The CNN model in another reported study has exhibited high accuracy in determining the presence of single or multiple root canal on mandibular first premolar using DPT radiographs (Fu et al., 2023). A study on 3D tooth segmentation reported that CNN model exhibited comparable accuracy and higher efficiency than human assessors in identifying root canal morphology (Lahoud et al., 2021). In Sherwood et al. (2021) demonstrated that CNN algorithms serve as a valuable tool in aiding dental professionals detect C-shaped root canal anatomy in mandibular second molars using CBCT images. A paper published presented a two-phase DL solution for accurate tooth and pulp cavity segmentation in CBCT images, yielding favourable results (Duan et al., 2021).

2.8 U-Net

The U-Net is a CNN architecture specifically developed for image segmentation by Olaf Ronnerberger and colleagues in 2015 (Ronneberger et al., 2015). Originally designed for biomedical image segmentation, its distinctive U-shape structure, comprise of two paths: an encoder and a decoder path. The encoder path, also known as the contracting or the analysis path, which resembles a regular convolutional network and provides classification information. This path captures context by progressively reducing the spatial dimensions of the input image and extracting feature maps through convolution and max-pooling layers. The second path is the decoder path, also known as the expansion or the synthesis path, which includes up-convolutions and matches the features from the encoder path. This design helps the network learn localised classification information. Furthermore, the decoder path enhances the resolution of the output, which is then fed into a final convolutional layer to produce a fully segmented image (Ronneberger et al., 2015). With classification networks often struggle to offer pixel-level contextual information, the U-Net has the ability to classify entire images into a single label, which is essential during medical image analysis.

U-Net is renowned for its ability to generate highly detailed segmentation maps even when limited data is available. This capability is particularly significant in medical imaging, where obtaining properly labelled images can be challenging and time-consuming. The U-Net utilizes random elastic deformation on the training data, allowing the network to learn variations without needing additional labelled data (Ronneberger et al., 2015). Moreover, to separate touching objects from the same class, U-Net applies a weighted loss function that penalises the model for failing to distinguish between them. Additionally, U-Net's context-based learning enables faster training compared to many other segmentation models.

Since its introduction, U-Net has experienced a surge in usage in medical imaging. Consequently, there have been numerous advancements in the U-Net architecture, with researchers implementing new techniques or integrating other imaging methods into U-Net. For instance, Çiçek et al. (2016) extended the original U-net architecture from 2D to 3D space. They applied this 3D U-Net to segment dense volumetric images based on sparse annotations. Additionally, Milletari et al. (2016) introduced another volumetric CNN named V-Net for segmenting prostate volumes. Building on the 2D U-Net architecture, a few studies have enhanced the 3D U-Net for segmenting brain tumours by incorporating multi-subregion architectures (Isensee, Kickingereder, et al., 2018; Isensee, Petersen, et al., 2018). Recent studies have demonstrated that methods based on the U-Net architecture outperform others on six publicly available segmentation tasks and can adapt automatically to various datasets (Isensee et al., 2019). Due to its superior performance and elegant design, U-Net have emerged as the benchmark in medical image segmentation, particularly excelling in segmenting medical images (Falk et al., 2019).

With numerous reports highlighting the advantages of U-Net in biomedical imaging, it is no surprise that the dental field is quickly adopting this technology. A study reported to have successfully developed a U-Net model in segmenting dental restorations on panoramic radiographs (Rohrer et al., 2022). Additionally, a study demonstrated that the 3D U-Net architecture achieved promising results in automatic tooth segmentation and classification without the need for expert adjustment (Shaheen et al., 2021). Another study proposed a two-phase DL solution for fast and accurate segmentation on any single tooth from a CBCT scan and performing accurate tooth and pulp cavity segmentation (Duan et al., 2021). Sherwood et al. (2021) reported that U-Net and its various modified models successfully segmented and classified C-shaped canal morphologies in mandibular second molars using CBCT.

2.9 Consistency of U-Net

The general workflow for validating the accuracy of U-Net architecture in biomedical segmentation relies on training datasets, which consists of the original input image and the corresponding segmented target (ground truth). These datasets are vital for the model to understand the complex relationships between input images and their corresponding segmented ground truth during training and validation (Wang et al., 2020). Parameters such as the DC and IoU are employed to evaluate the model's performance in learning these relationships and during the testing stage as well. To prevent overfitting, these datasets are typically randomly divided into training, validation, and testing datasets.

Several variations of the U-Net architecture, such as 2D U-Net (Duan et al., 2021; Li et al., 2020), 2.5D U-Net (Zhou et al., 2022), 3D U-Net (Shaheen et al., 2021), and 3.5D U-Net (Hsu et al., 2022), have been suggested for segmenting CBCT images. The accuracy on the 3.5D U-Net model has been shown to outperform the other mentioned variations (Hsu et al., 2022). In another investigation, the U-Net model was compared with the Xception U-Net and the residual U-Net for segmenting and classifying c-shaped canal morphologies in mandibular second molars using CBCT. The study revealed that the Xception U-Net achieved the best performance (Sherwood et al., 2021).

Other studies have also modified the encoder and/or the decoder phase on the U-Net to overcome with its limitations (Aboussaleh et al., 2023; Long et al., 2021). Long et al. (2021) introduced a multi-branch pooling (MP) method in the encoder and multi-branch dense prediction (MDP) method in the decoder, demonstrating high accuracy in segmenting MRI brain tissues. The U-NetPlus model integrates VGG-11 and VGG-16 as encoders with batch-normalized pre-trained weights and replaces the transposed convolution in the decoder layer with nearest-neighbor interpolation (Kamrul Hasan &

Linte, 2019). This pre-trained encoder (Iglovikov & Shvets, 2018) accelerates convergence, resulting in improved results (He et al., 2019). Additionally, the use of nearest-neighbor interpolation in the decoder eliminates artifacts produced by the transposed convolution. Aboussaleh et al. (2023) have developed three types of encoders: VGG-19, ResNet50, and MobileNetV2. Their results show a good performance in brain tumour segmentation, as measured by the Dice coefficient (Aboussaleh et al., 2023).

Data augmentation have been continuously emphasized to achieve better segmentation results, especially when working with small training datasets. A study applied two data augmentation techniques: vertical flipping and random intensity variation to MRI scans (Benson et al., 2019). The network was trained both with and without data augmentation, and the results indicated that data augmentation led to a slight increase in accuracy in dice coefficient (Benson et al., 2019) (Cirillo et al., 2021) investigated the impact of augmentation techniques: flipping, rotation, scaling, brightness adjustment, and elastic deformation on the training of a standard 3D U-Net (Ronneberger et al., 2015) on the Brain Tumour Segmentation (BraTS) dataset (Spyridon Bakas et al., 2017; S. Bakas et al., 2017; Menze et al., 2015). Cirillo et al. (2021) findings showed that data augmentation significantly improved validation performance in many instances. They speculated that data augmentation had not been extensively explored for brain tumour segmentation because the BraTS training dataset was already quite large, and several studies had suggested that data augmentation might not provide substantial benefits (Havaei et al., 2017; Lyksborg et al., 2015). As per our knowledge, there ia a lack of studies that investigates the consistency of the U-Net model in pulpal segmentation on MFPs on five sets of randomly generated training, validation, and testing dataset from the same data pool.

2.10 Evaluation metrics

2.10.1 Dice coefficient (DC)

DC is a statistical measure used to gauge the similarity between two sets of data (Müller et al., 2022). In the context of image segmentation, it quantifies the overlap between the predicted segmentation and the ground truth, and it is represented as:

$$DC = \frac{2(G \cap P)}{G + P}$$

where G represents the ground truth, P represents the predicted segmentation result and ∩ represents union. DC is a metric of area overlap between the predicted segmentation results P and the ground truth G. A higher DC value signifies better segmentation accuracy due to greater overlap between P and G, where the value ranges from 0 to 1.

2.10.2 Intersection over Union (IoU)

IoU also known as Jaccard index, is a metric used to calculate mean average precision in image segmentation tasks (Müller et al., 2022). It quantifies the overlap between the predicted segmentation and the ground truth, indication how well the predicted areas match the actual areas. It is represented as:

$$IoU = \frac{TP\ (area\ of\ overlap)}{TP + FP + FN\ (area\ of\ union)}$$

where TP represents true positive, which is the area of overlap, FP is false positive, which is the wrong detection and FN is false negative where the area of the ground truth was not detected. The value ranges from 0 to 1, with a higher value indicates better overlap and thus a more accurate segmentation.

CHAPTER 3: MATERIALS AND METHOD

The study was conducted in the Faculty of Dentistry, University of Malaya approved by the Medical Ethics Committee, Faculty of Dentistry, University of Malaya (DF RD2314/0089).

3.1 Sample size calculation

The sample size was calculated using G-Power (Version 3.1.9.6), based on an alpha error of 0.05, 80% power and effect size f = 0.333, using the results from a previous study by Sherwood et al. (2021). The total sample size determined in this study was N = 130 of extracted human MFPs.

3.2 Preparation of extracted MFPs

The extracted human MFPs with intact apices were collected from a previous postgraduate research project conducted in the University of Malaya, which were utilized in this study. The MFPs were distinguished from mandibular second premolars by evaluating their anatomical characteristics, such as longer crown, more crown taper from contact to cervix, more pointed cusp, more prominent buccal ridge, and shorter root with pointed apex. The MFPs usually exhibit two-cusp type with pointed tip buccal cusp and lingual cusp is shorter and small reaching 2/3 of the length of the buccal cusp, however the mandibular premolar occasionally possesses a two-cusp type with one buccal and one lingual cusp and a slightly more common three cusp-type with one buccal and two lingual cusps (Gutmann & Fan, 2020).

All teeth samples in the existing study were obtained from a previous postgraduate project. The protocol includes, all selected teeth were washed thoroughly with tap water and debrided with ultrasonic scaler tips and were immersed in a 5.25% sodium

hypochlorite solution for 24 hours to effectively disinfect the teeth by eliminating organic substances, followed by storing them in chloroxylenol solution (Dettol®) until CBCT image acquisition stage.

The inclusion criteria of the samples are MFPs that are sound with intact clinical crown or minimal caries limited to ICDAS score 4, fully developed roots, minimal dental restorations, and absence of root resorptions. The exclusion criteria encompass teeth that are fractured, aberrant shape teeth, grossly carious teeth (ICDAS 5 & 6), teeth that undergone root canal treatment and full coverage restorations.

3.3 CBCT image acquisition

The CBCT images of the selected extracted MFPs were acquired using a Kodak CS 9000 3D CBCT machine (Kodak Carestream Health, Inc, Rochester, NY) operating at 60kVp, 6.3mA, with a field of view of 5 x 5 cm, a voxel size of 0.076mm and an exposure time of 10.8 seconds. The investigator in the previous postgraduate research project mounted ten MFPs on a vinyl polysiloxilane putty material (Imflex Putty, Meta Biomed Co., Chungcheongnam-do, Korea) to maintain their stable position in relation to the horizontal plane during CBCT imaging. This process was repeated to capture all 130 MFPs and the CBCT images were stored in 13 CD-ROMs in Digital Imaging and Communications in Medicine (DICOM) format. Figure 3.1 showed an example of the ten MFPs CBCT image.

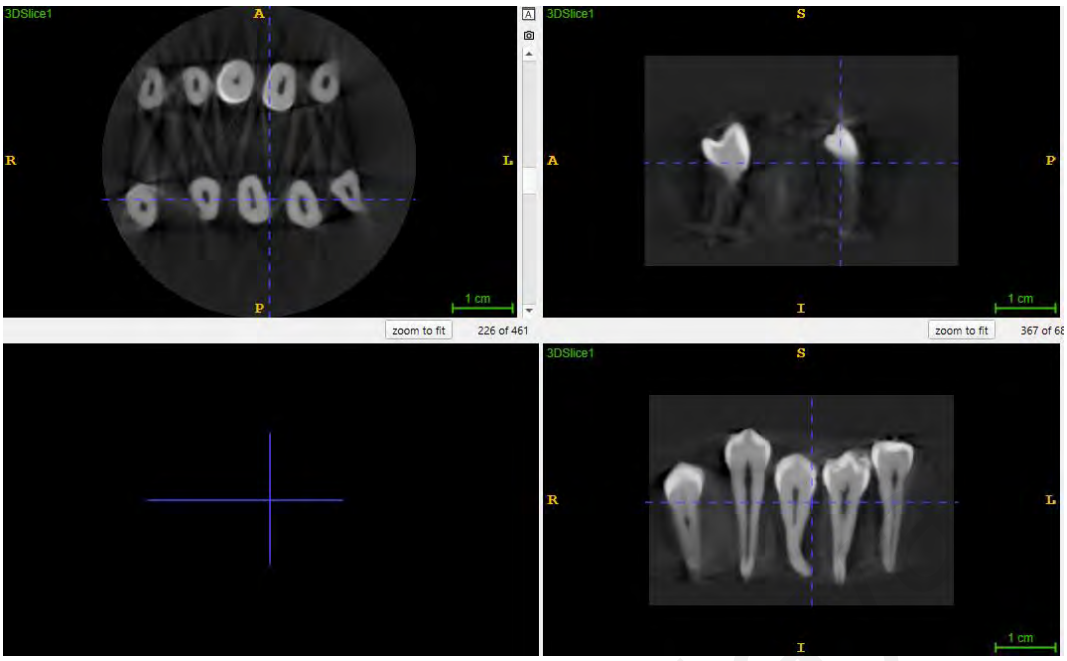

Figure 3. 1: Example of CBCT image acquired from previous research project

3.4 Segmentation of the pulp space

The DICOM files were imported to 3D Slicer software (Version 5.6.1), and the individual MFPs were cropped from the original file utilizing the crop volume module within the software. To ensure proper cropping of the MFPs, the CBCT images were viewed in axial, sagittal and transverse planes to accurately identify the region of interest (ROI) as shown in Figure 3.2 This procedure was repeated for all 130 MFP CBCT images, and the data for each individual MFP were saved and labelled in separate DICOM files.

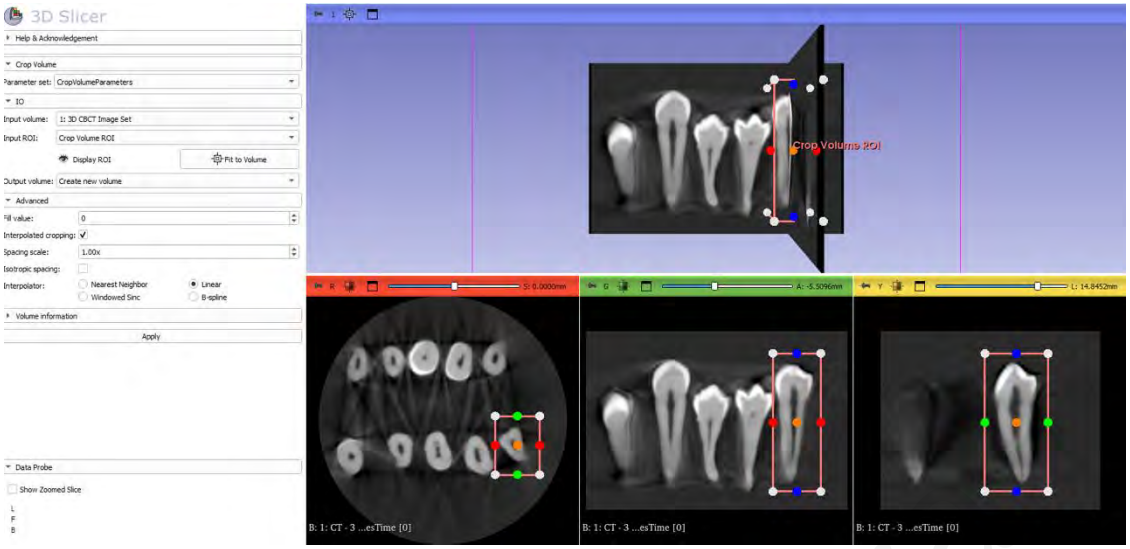

Figure 3. 2: Cropping of individual MFPs with the crop volume module on 3D Slicer (Version 5.6.1) software.

All CBCT images were viewed in a dimly lit room on a monitor (BenQ GL2460, BenQ Corporation, Taipei, Taiwan) and segmented by a postgraduate student (MW) using ITK-SNAP (Version 4.0.2) (open source; Penn Image Computing and Science Laboratory (PICSL) at the University of Pennsylvania, Philadelphia, PA and Scientific Computing and Imaging Institute (SCI) at the University of Utah, Salt Lake City, UT) (Yushkevich et al., 2006). ITK-SNAP software allows for semiautomatic segmentation of the pulpal space. Intraclass correlation coefficient (ICC) was recorded to ensure reliability of the data throughout the study using the test-retest method over a three-week interval. The ICC value was 0.95, which showed an excellent level of reliability (Koo & Li, 2016).

Automatic segmentation was performed using the active contour segmentation mode, also known as the snake mode, to capture the ROI for segmentation viewed in the axial, sagittal, and transverse planes, as shown in Figure 3.3. In the second step, the operator identifies the optimal threshold range for each CBCT image through visual analysis, considering the anatomical boundaries between the hard structures of the tooth and the pulp space (Figure 3.4). In the third step, bubble-like formations or seed-like

geometries were added throughout the entire pulp extension in multiplane reconstructions, ensuring that the cavity corresponding to the pulp space was filled according to the predetermined threshold range as shown in Figure 3.5. After the automatic segmentation was completed, the operator used the paintbrush tool to refine and add any missing pulpal space in the apical third of the root manually, as illustrated in Figure 3.6. Finally, the segmentation process was completed in all three planes, resulting in the formation of a 3D reconstruction image of the pulp space (Figure 3.7). The segmentation images were exported and saved in the Neuroimaging Informatics Technology Initiative (NIfTI) format serving as the ground truth.

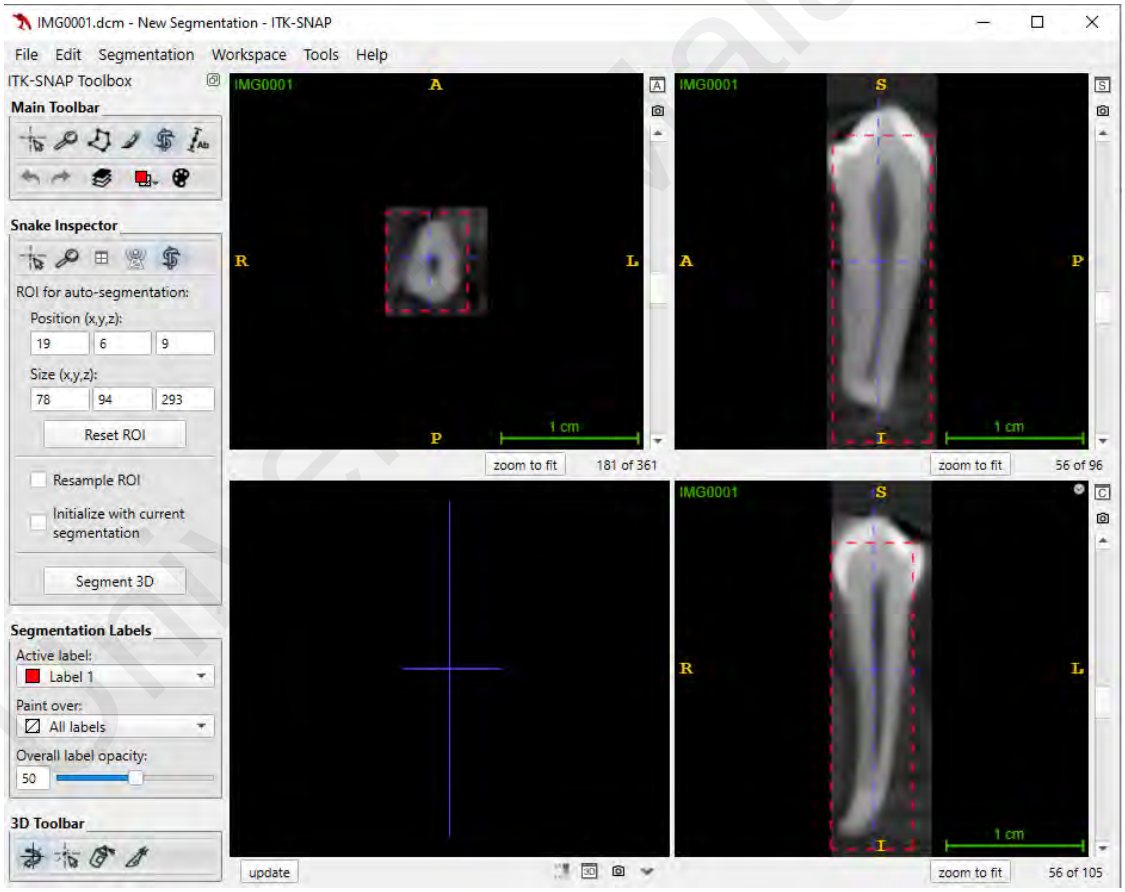

Figure 3. 3: Selecting the region of interest (ROI).

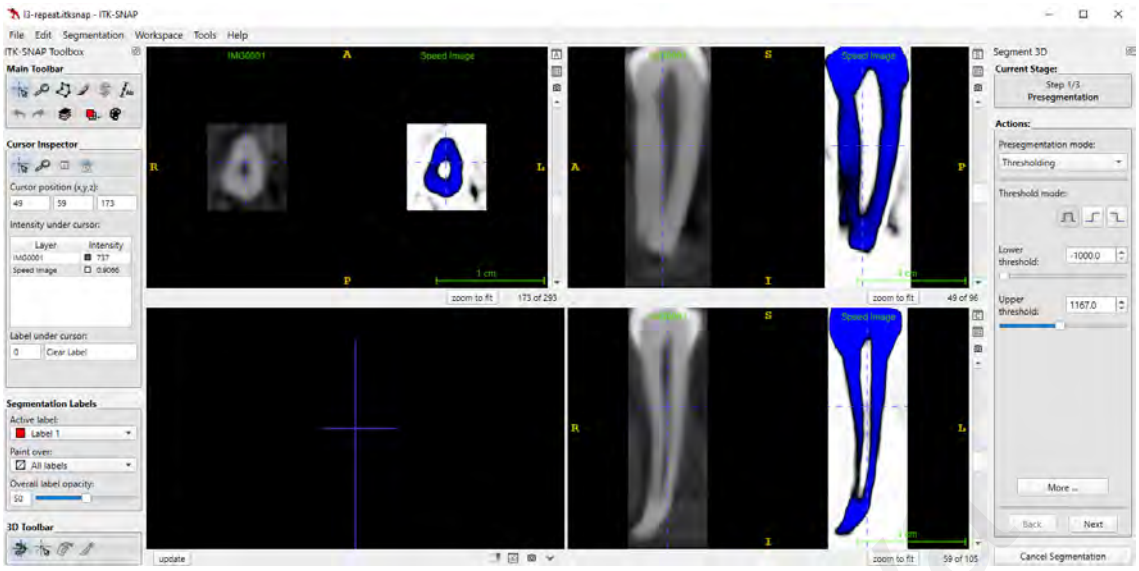

Figure 3. 4: Adjusting the optimal threshold range to segment the pulpal space from the dental hard tissues.

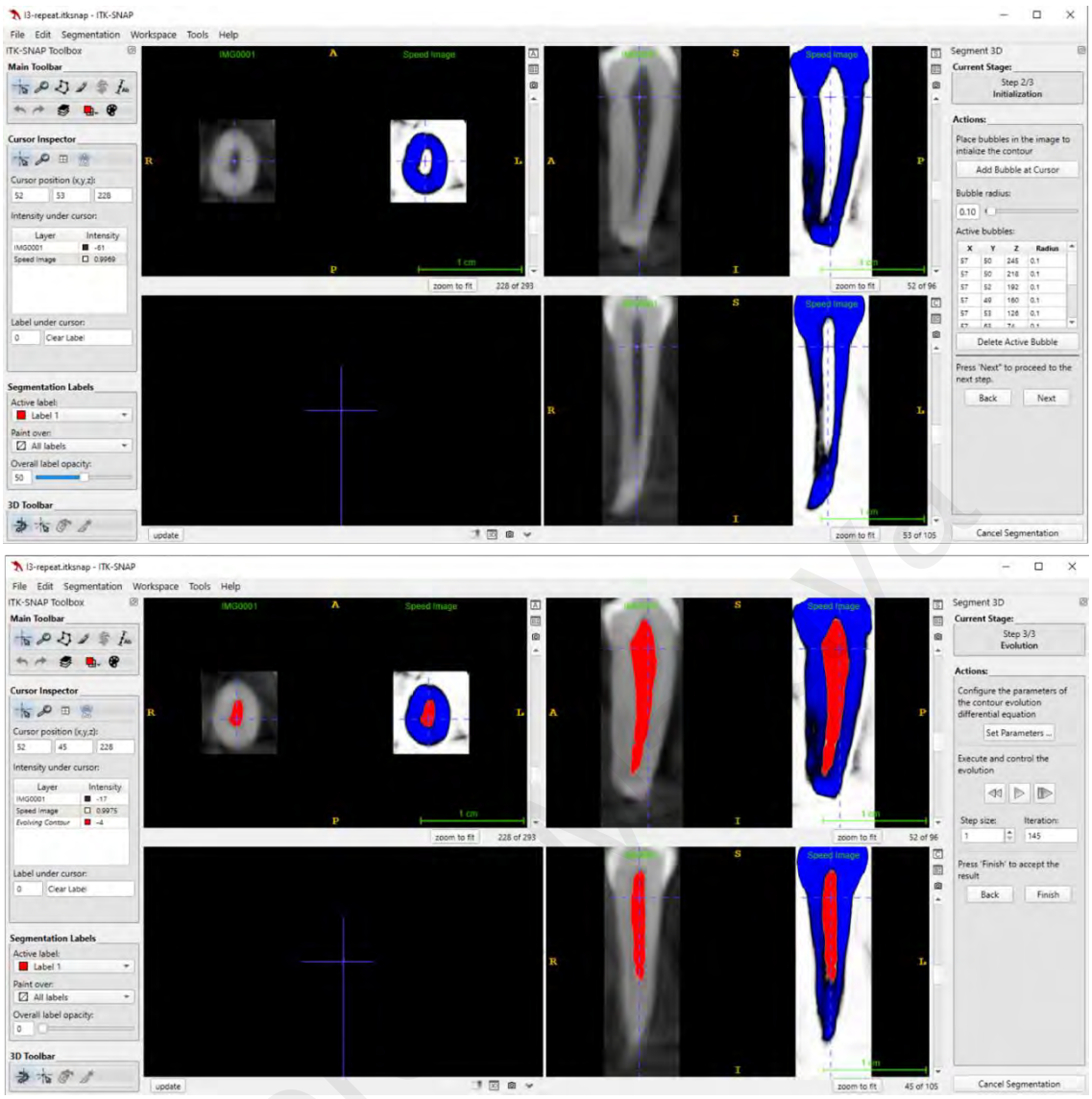

Figure 3. 5: Adding the bubbles into the entire pulpal space and segmentation filled up the entire space through combination of all the bubbles.

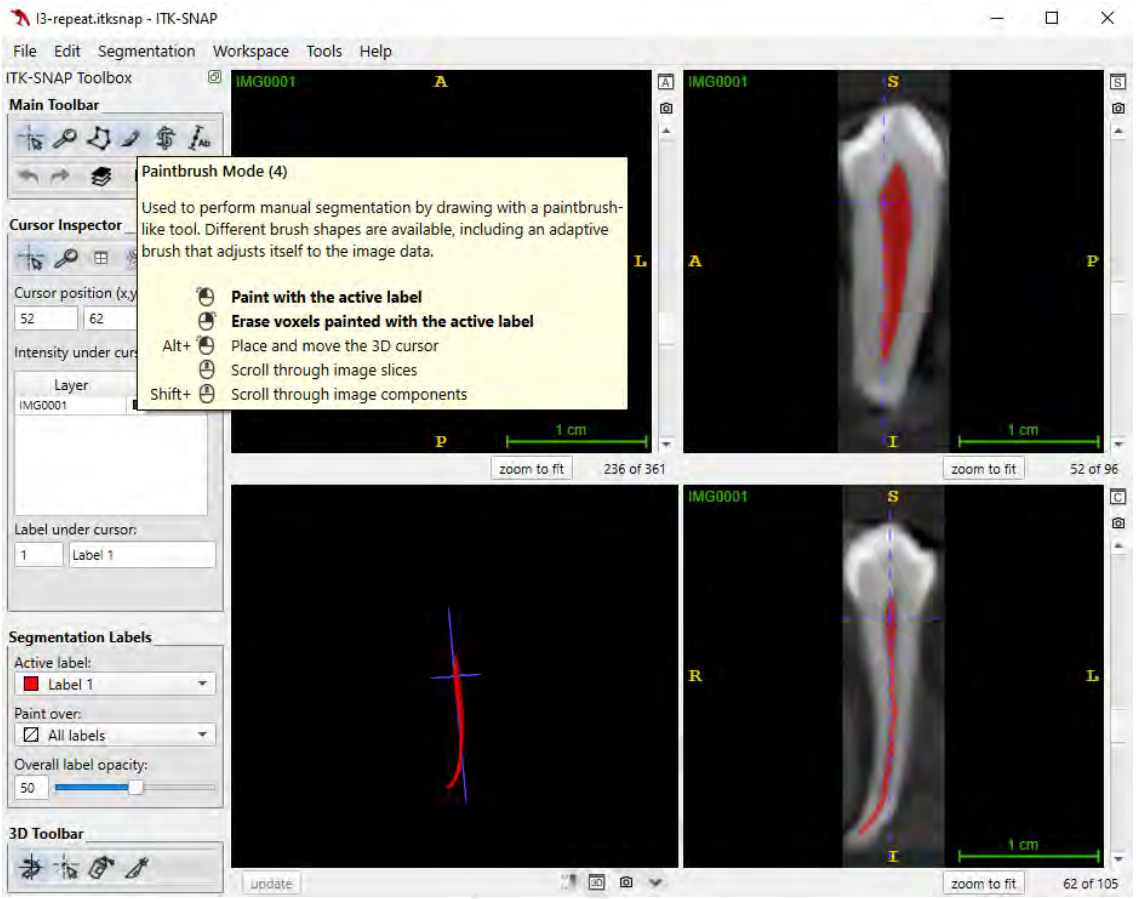

Figure 3. 6: Refinement of the segmentation with the paintbrush mode.

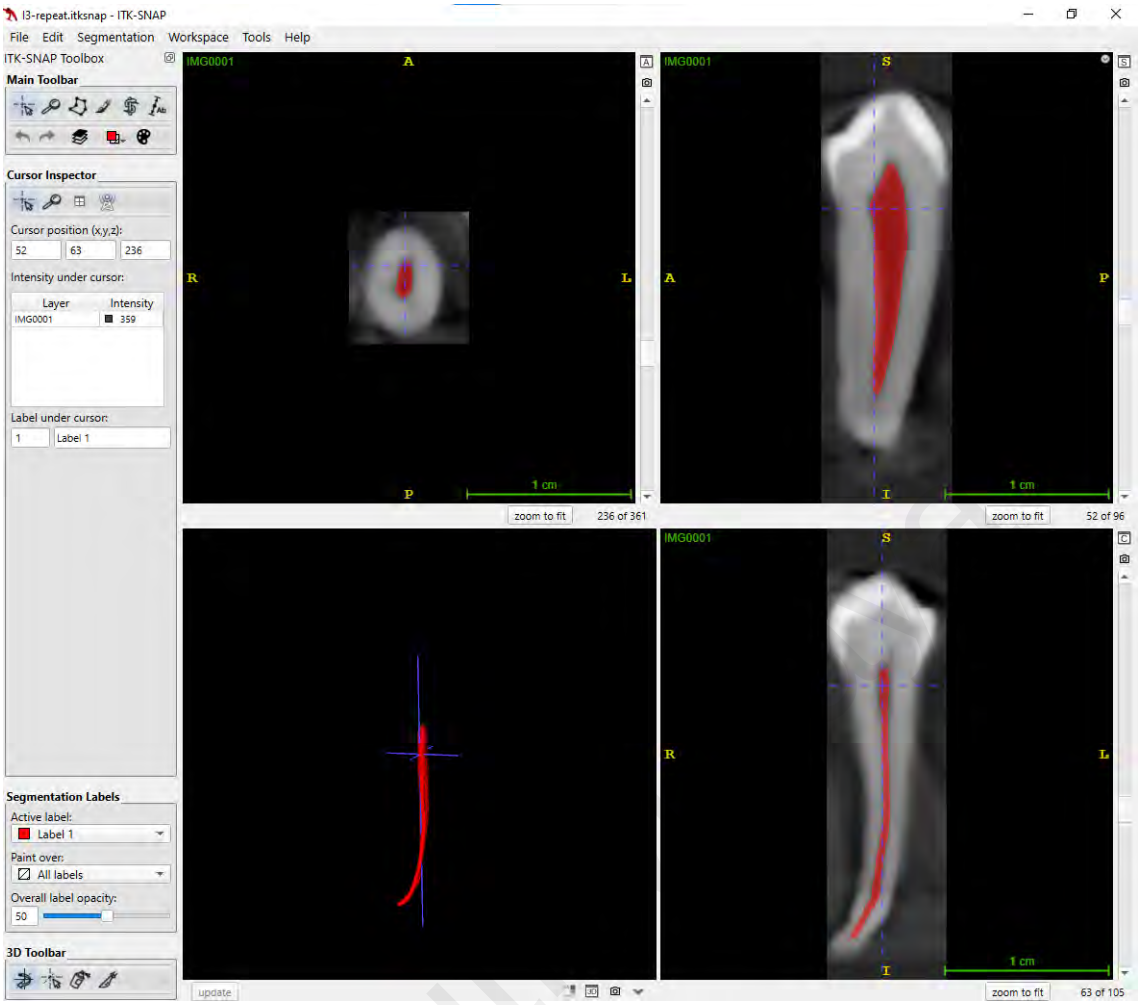

Figure 3. 7: Completion of segmentation process of the pulpal space.

3.5 Generating random dataset

The total dataset was randomly divided into training, validation, and testing datasets, following a 70%, 20%, and 10% split, respectively. The Python Random library was implemented to generate training dataset (n=91), validation dataset (n=26), and testing dataset (n=13). This process was repeated five times to create five different randomized sets of the training, validation, and testing datasets (Figure 3.8).

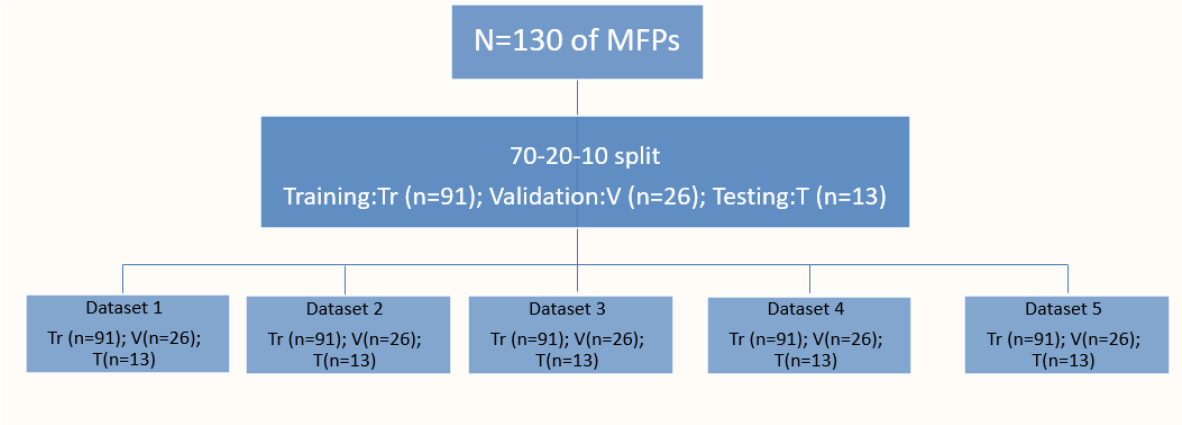

Figure 3. 8: Outline of generating 5 random datasets

3.6 Data pre-processing

The original DICOM files were converted into NIfTI file using the dicom2nifti Python library. The Medical Open Network for Artificial Intelligence (MONAI) framework, an open-source framework built on the PyTorch platform specialised in medical imaging, was employed for pre-processing, and augmenting the datasets. The dataset's intensity values were scaled to ensure consistent normalization across all datasets, facilitated by the MONAI framework. To perform U-Net segmentation, it is necessary for all datasets to have dimensions that are powers of 2. Therefore, all datasets were resized to (128,128,128) for width, height, and length, ensuring that each dimension corresponds to 2^7.

3.7 Data augmentation

To enhance the model robustness and improve its ability to generalize, several data augmentation techniques were applied to the images. These data augmentation was done using the MONAI framework where random rotations were applied to make the model invariant to orientation changes, which is crucial for 3D medical scans that can be captured from various angles. Additionally, random horizontal and vertical flips were

performed to increase the diversity of the training and validation datasets. To simulate different imaging conditions and equipment variations, random alteration in intensity such as brightness and contrast adjustments were implemented (Figure 3.9). These augmentations help create a more comprehensive training and validation datasets, ultimately leading to better model performance.

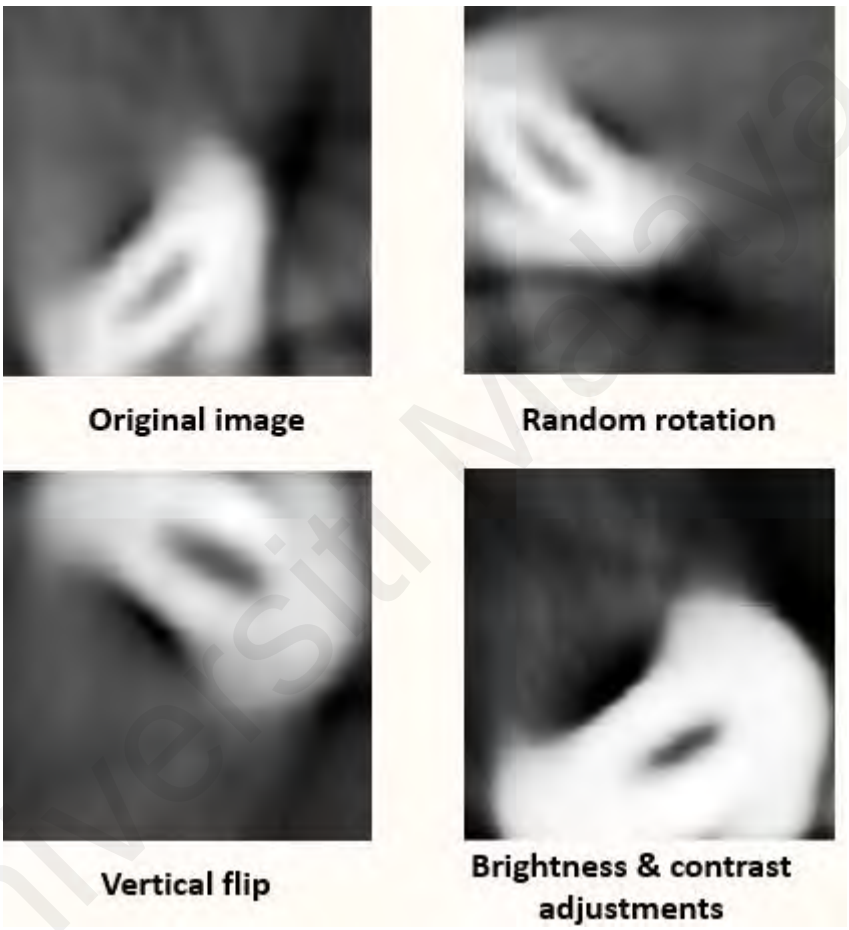

Figure 3. 9: Implementation of various types of augmentation into the training and validation datasets.

3.8 U-Net architecture

The U-Net architecture was adopted from the MONAI library, which was initialized with parameters optimized for 3D imaging, a slight modification from the original U-Net proposed by (Ronneberger et al., 2015). The channels were defined to take in 16, 32, 64, 128 and 256 feature maps, respectively shown in Figure 3.10. These values specify the number of feature maps at each level of the U-Net model, resulting in five levels in both the contracting (encoder) path and the symmetrical expanding (decoder) path. Each subsequent level in the encoder path doubles the number of feature maps, allowing the model to capture increasingly complex features as the spatial resolution is reduced. This is achieved through the application of 3x3 convolutions, followed by a rectified linear unit (ReLU) activation function. A 2x2 max pooling operation with stride of 2 is then applied for down-sampling and at each down-sampling the number of feature channels were doubled. This combination of operations enables the model to effectively learn hierarchical features while progressively reducing the spatial dimensions of the input image.

In each level of the expansive (decoder) path, the feature map was up-sampled and subjected to a 2x2 convolution (up-convolution) that reduced the number of feature channels by half. This was followed by concatenation with the correspondingly cropped feature map from the contracting (encoder) path. Two 3x3 convolutions were performed, each followed by a ReLU. Cropping was necessary to compensate for the loss of border pixels during each convolution. In the final layer, a 1x1 convolution was used to map each component feature vector to the desired number of output classes. This approach ensured accurate reconstruction of spatial dimensions and details while preserving the learned features.

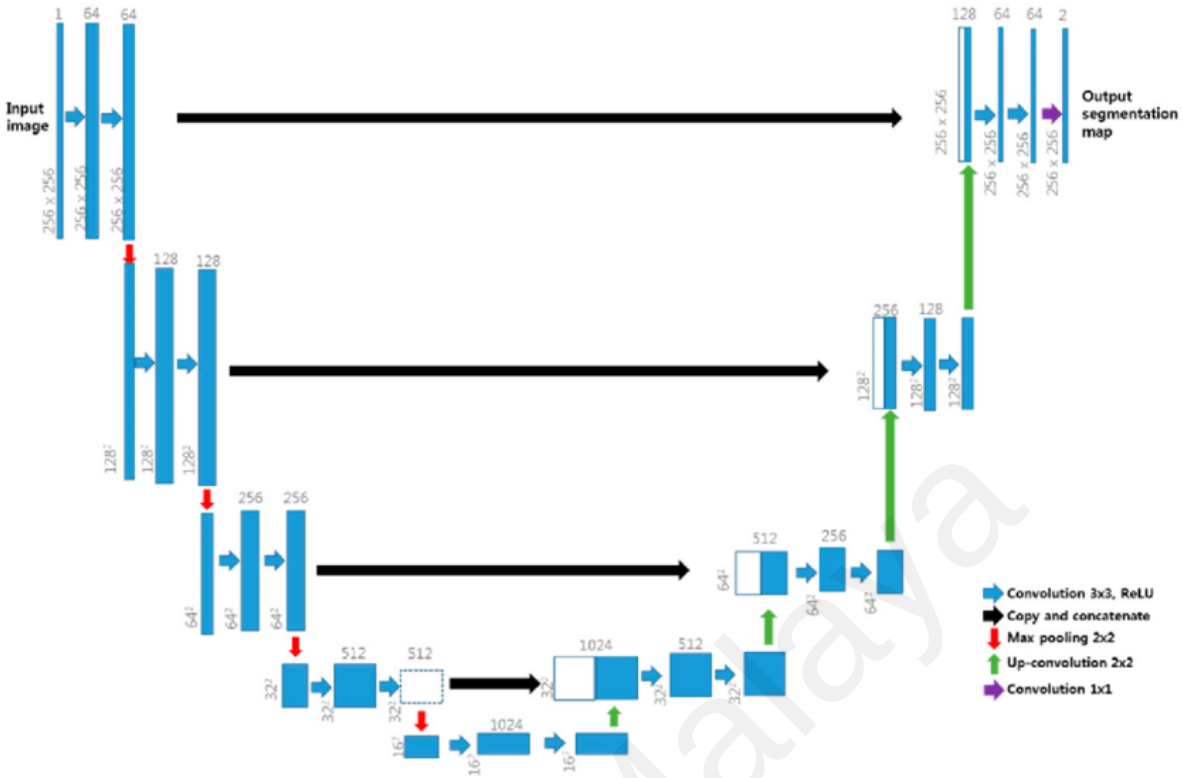

Figure 3. 10: The U-Net architecture employed in the present study.

3.9 Training, validation and testing stage

Google Colab, a robust cloud-based infrastructure, was used to run the code throughout the study. Python was chosen as the primary programming language due to its extensive libraries and frameworks tailored for machine learning in medical imaging. The cloud consists of Python 3 Google Compute Engine Backend as the graphic processing unit (GPU), a cloud RAM of 12 GB, of which 9.57 GB were utilised during the execution of the study and an allocated disk size of 78.2 GB, with 31.69 GB utilised during the study.

The model was implemented by first running Dataset 1 (Training n=91; Validation n=26; Testing n=13) through the training, validation, and testing phases, recording the DC and IoU after each phase. This process was then repeated for Datasets

2,3,4, and 5, following the same procedure as for Dataset 1, with the DC and IoU recorded during their respective phases. The model's parameters were refined through iterations based on validation performance in Dataset 1 and 2. Iterations refers to the process of repeating the training phase multiple times until an optimal performance is achieved. Initially, iterations were not introduced during the training phase for the first two datasets, leading to poor segmentation performance as illustrated in Figure 3.11 after the testing phase. Once iteration was added in the training phase for Dataset 1 and 2, the respective model's performance was then validated and tested with their respective datasets, recording the DC and IoU and hence iterations were then added into the training phases for Dataset 3,4 and 5.

Figure 3. 11: Segmented image without iteration.

3.10 Statistical analysis

The data collected and recorded were analysed with Statistical Program for Social Sciences software (SPSS, Version 29.0, IBM, NY, USA). The normality of the data distribution was assessed using the Kolmogorov-Smirnov test, indicated suitability for parametric testing. The means of DC and IOU score among training, validation, and testing datasets were compared using one-way ANOVA followed by post-hoc LSD for multiple comparison test between the groups. P-value, $p < 0.05$ was set as statistically

significant. Intraclass correlation coefficient (ICC) was implemented to measure the intra-observer reliability on the pulpal space segmentation.

CHAPTER 4: RESULTS

4.1 Dice coefficient (DC) during the training phase

The raw data can be found in Appendix A, and it is represented through Figure 4.1. Epoch refers to one complete pass through the entire training dataset by the U-Net. Figure 4.1 demonstrated the DC increases with epoch, which indicates that U-Net is learning. Training Dataset 2 demonstrated the highest DC with epoch, while Training Dataset 1, 4 and 5 exhibited similar DC with epoch.

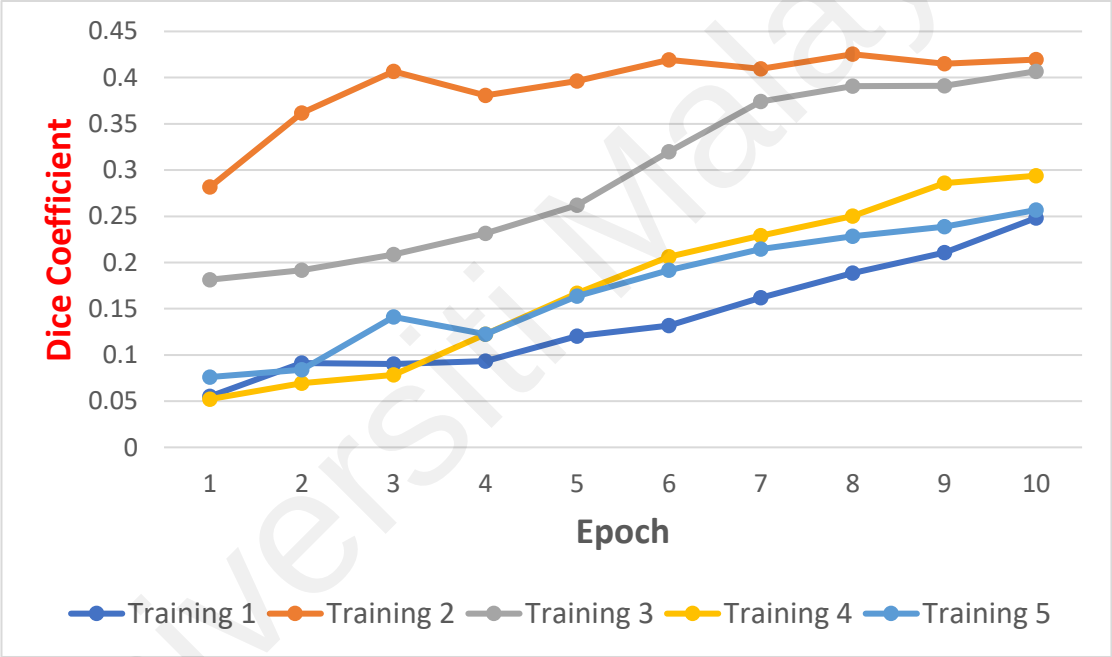

Figure 4. 1: Dice coefficient (DC) over epoch of different training dataset

Descriptive statistics were conducted on the mean and standard deviation of DC among the different training datasets during training phase presented in Table 4.1 and Table 4.2.

Table 4. 1: Mean DC among the different training datasets during training phase

Training Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	91	0.13	0.06	0.09	0.18
2	91	0.39	0.04	0.36	0.42
3	91	0.29	0.09	0.23	0.36
4	91	0.17	0.09	0.11	0.24
5	91	0.17	0.06	0.12	0.21

Table 4. 2: Fisher's LSD post hoc test on mean DC among the training datasets

(I) Training Datasets	(J) Training Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.25*	0.00	-0.31	-0.18
	3	-0.15*	0.00	-0.22	-0.09
	4	-0.03	0.26	-0.10	0.02
	5	0.03	0.32	-0.09	0.03
2	1	0.25*	0.00	0.18	0.31
	3	0.09*	0.00	0.03	0.16
	4	0.21*	0.00	0.15	0.28
	5	0.21*	0.00	0.15	0.28
3	1	0.15*	0.00	0.09	0.22
	2	-0.09*	0.00	-0.16	-0.03
	4	0.12*	0.00	0.05	0.18
	5	0.12*	0.00	0.05	0.18
4	1	0.03	0.26	-0.02	0.10
	2	-0.21*	0.00	-0.28	-0.15
	3	-0.12*	0.00	-0.18	-0.05
	5	0.00	0.91	-0.06	0.06
5	1	0.03	0.32	-0.03	0.09
	2	-0.21*	0.00	-0.28	-0.15
	3	-0.12*	0.00	-0.18	-0.05
	4	-0.00	0.91	-0.06	0.06

* The mean difference showed significant difference ($p < 0.05$).

One-way ANOVA was used to compare the different training datasets to the mean DC. There is an overall significant difference among the five datasets with the mean DC,

F = 21.35 (p = 0.00). Post hoc test with Fisher's Least Significant Difference (LSD) was done for multiple comparisons. The results tabulated in Table 4.1 and 4.2, showed the mean DC for training Dataset 2 (0.39 ± 0.04) and 3 (0.29 ± 0.09) were significantly higher than the other training datasets (p = 0.00). The mean DC for training Dataset 1 (0.13 ± 0.06), 4 (0.17 ± 0.09) and 5 (0.17 ± 0.06) showed no significant difference during the training phase.

4.2 Intersection over Union (IoU) during the training phase

The raw data can be obtained in Appendix B, and it is represented through Figure 4.2. Figure 4.2 demonstrated as epoch increase, IoU increases with all the training datasets. Training Dataset 2 demonstrated the highest IoU with epoch compared to the other training datasets. However, IoU of Dataset 1, 3 and 5 performed similarly among each other with epoch.

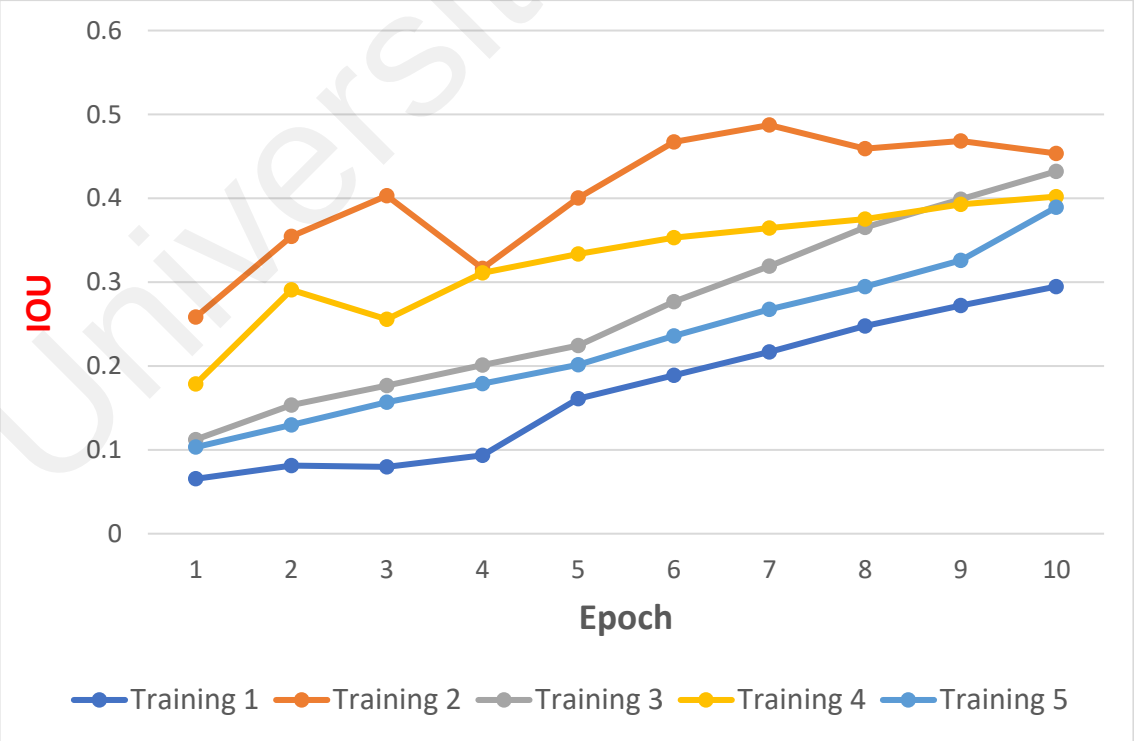

Figure 4. 2: Intersection over Union (IoU) over epoch of different training datasets

Descriptive statistics were conducted on the mean and standard deviation of IoU among the different training datasets during training phase presented in Table 4.3 and Table 4.4.

Table 4. 3: Mean IoU among the different training datasets during training phase

Training Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	91	0.17	0.08	0.10	0.23
2	91	0.40	0.07	0.35	0.46
3	91	0.26	0.10	0.18	0.34
4	91	0.32	0.06	0.27	0.37
5	91	0.22	0.09	0.16	0.29

Table 4. 4: Fisher's LSD post hoc test on mean IoU among the training datasets

(I) Training Datasets	(J) Training Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.23*	0.00	-0.31	-0.15
	3	-0.09*	0.01	-0.17	-0.01
	4	-0.15*	0.00	-0.23	-0.07
	5	-0.05	0.14	-0.13	0.20
2	1	0.23*	0.00	0.15	0.31
	3	0.14*	0.00	0.06	0.21
	4	0.81*	0.04	0.00	0.16
	5	0.17*	0.00	0.09	0.25
3	1	0.09*	0.01	0.01	0.17
	2	-0.14*	0.00	-0.21	-0.06
	4	-0.05	0.13	-0.13	0.19
	5	0.03	0.34	-0.04	0.11
4	1	0.15*	0.00	0.07	0.23
	2	-0.81*	0.04	-0.16	-0.00
	3	0.05	0.13	-0.19	0.13
	5	0.09*	0.01	0.01	0.17
5	1	0.05	0.14	-0.02	0.13
	2	-0.17*	0.00	-0.25	-0.09
	3	-0.03	0.34	-0.11	0.04
	4	-0.09*	0.01	-0.17	-0.01

* The mean difference showed significant difference ($p < 0.05$)

One-way ANOVA was used to compare the mean IoU among the different training datasets. There is an overall significant difference among the five datasets with the mean IoU, F = 10.76 (p = 0.00). Post hoc Fisher's LSD test was done for multiple comparisons. The results as shown in Table 4.3 and 4.4, indicated that the mean IoU in training Dataset 2 (0.40 ± 0.07) was significantly higher than all the other datasets. The mean IoU between Dataset 1 (0.17 ± 0.08) and Dataset 5 (0.22 ± 0.09) showed no significant difference (p = 0.14). Similarly, during the training phase, the mean IoU for Dataset 3 (0.26 ± 0.10) showed no significant difference when compared to Dataset 4 (0.32 ± 0.06) and Dataset 5 (0.22 ± 0.09).

4.3 Dice coefficient (DC) during the validation phase

The raw data can be found in Appendix A, and it is represented through Figure 4.3. Overall DC for Dataset 3 was the highest, but at the 4th and 5th epoch, Dataset 2 was higher than Dataset 3. Datasets 1, 4 and 5 have similar increase in DC over epoch, with Dataset 4 having the highest DC during the 10th epoch.

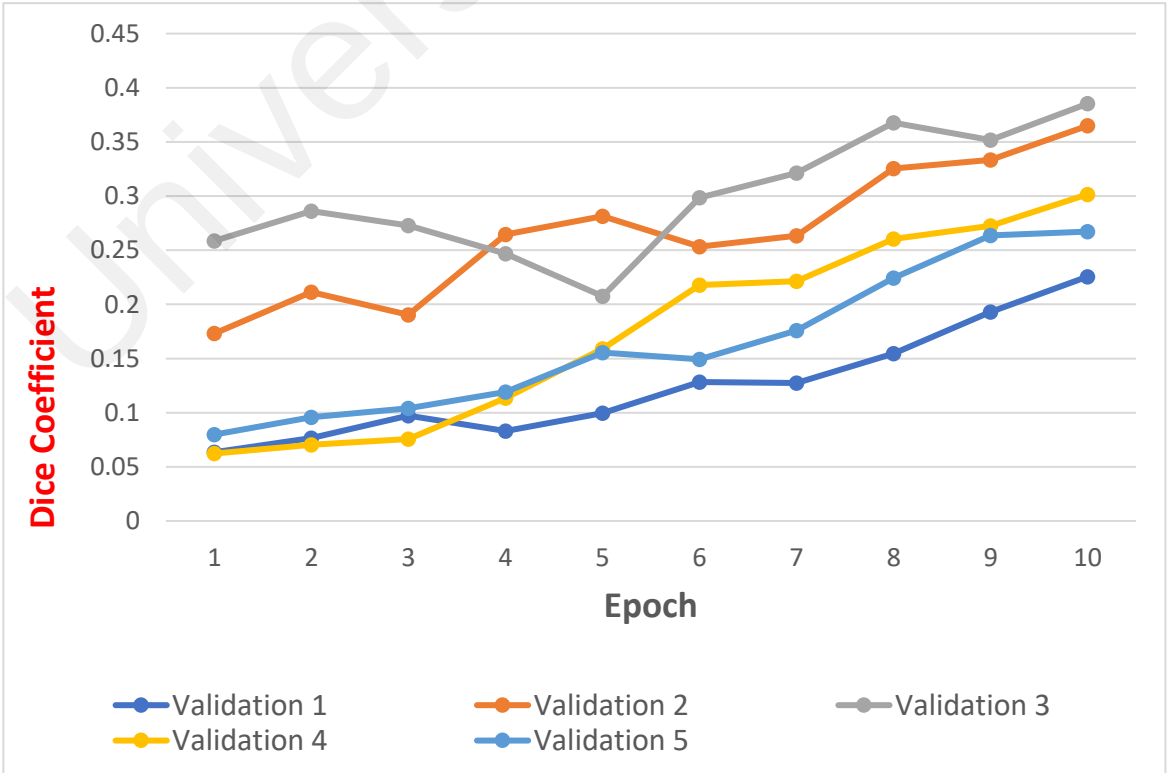

Figure 4. 3: Dice coefficient (DC) over epoch of different validation datasets

Descriptive statistics were conducted on the mean and standard deviation of DC among the different validation datasets during validation phase presented in Table 4.5 and Table 4.6.

Table 4. 5: Mean DC among the different validation datasets during validation phase

Validation Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	26	0.12	0.05	0.08	0.16
2	26	0.26	0.06	0.22	0.31
3	26	0.29	0.05	0.25	0.34
4	26	0.17	0.09	0.11	0.24
5	26	0.16	0.06	0.11	0.21

Table 4. 6: Fisher's LSD post hoc test on mean DC among the validation datasets

(I) Validation Datasets	(J) Validation Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.14*	0.00	-0.20	-0.08
	3	-0.17*	0.00	-0.23	-0.11
	4	-0.05	0.10	-0.11	0.01
	5	-0.03	0.20	-0.09	0.02
2	1	0.14*	0.00	0.08	0.20
	3	-0.03	0.27	-0.09	0.02
	4	0.09*	0.00	0.02	0.15
	5	0.10*	0.00	0.04	0.16
3	1	0.17*	0.00	0.11	0.23
	2	0.03	0.27	-0.02	0.09
	4	0.12*	0.00	0.06	0.18
	5	0.13*	0.00	0.07	0.19
4	1	0.05	0.10	-0.01	0.11
	2	-0.09*	0.00	-0.15	-0.02
	3	-0.12*	0.00	-0.18	-0.06
	5	0.01	0.69	-0.04	0.07
5	1	0.03	0.20	-0.02	0.09
	2	-0.10*	0.00	-0.16	-0.04
	3	-0.13*	0.00	-0.19	-0.07
	4	-0.01	0.69	-0.07	0.04

* The mean difference showed significant difference ($p < 0.05$).

One-way ANOVA followed by post hoc Fisher's LSD were done to compare the means DC among the different validation datasets. The analysis reported of an overall

significant difference among the five datasets with the mean DC, F = 11.85 (p = 0.00). The results from the multiple comparisons are shown in Table 4.5 and 4.6, which showed no significant difference in the mean DC between validation Dataset 3 (0.29 ± 0.05) and Dataset 2 (0.26 ± 0.06) with a p = 0.27 and between validation Dataset 1 (0.12 ± 0.05), Dataset 4 (0.17 ± 0.09) and Dataset 5 (0.16 ± 0.06).

4.4 Intersection over Union (IoU) during the validation phase

The raw data can be obtained in Appendix B. Figure 4.4 demonstrated the trend in the changes in IoU whereby there is an increase in IoU with epochs across all the validation datasets. Dataset 3 had the highest overall IoU, except during the 4th and 5th epochs when Dataset 2 showed higher values. The increase in IoU was similar among Datasets 1, 4 and 5 up to the 5th epoch.

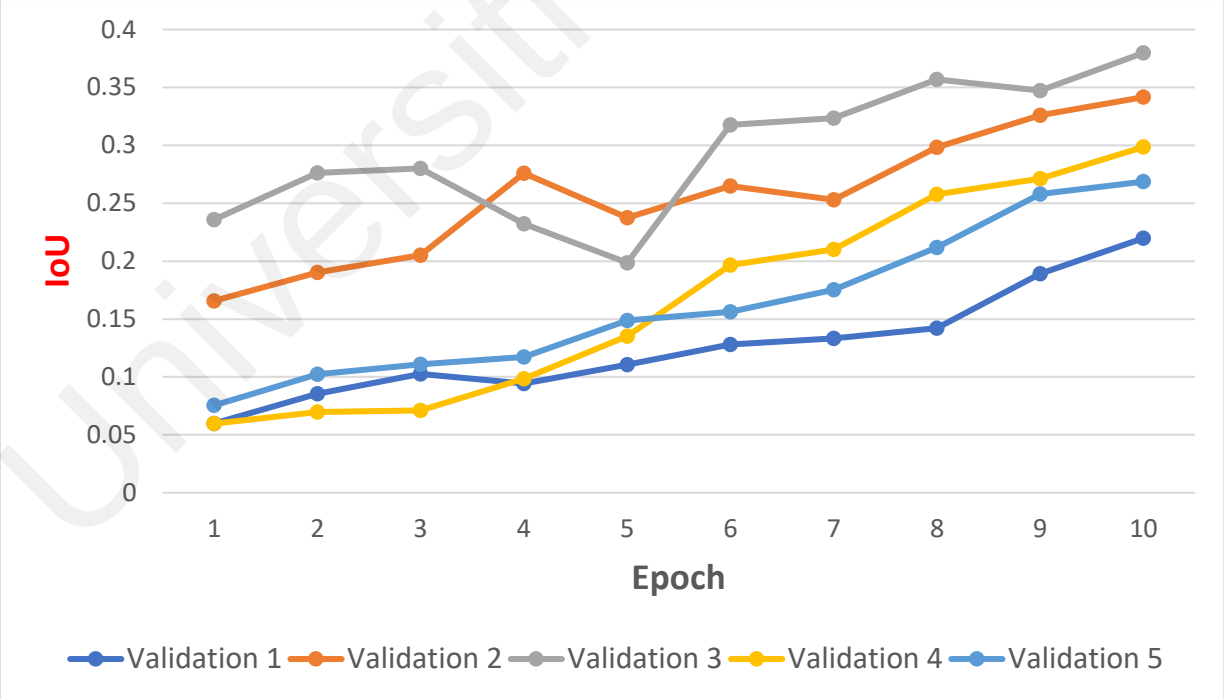

Figure 4. 4: Intersection over Union (IoU) over epoch of different validation datasets

Descriptive statistics was done on the mean and standard deviation of IoU among the different validation datasets during validation phase presented in Table 4.7 and Table 4.8.

Table 4. 7: Mean IoU among the different validation datasets during validation phase

Validation Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	26	0.12	0.04	0.09	0.16
2	26	0.25	0.05	0.21	0.29
3	26	0.29	0.06	0.25	0.33
4	26	0.16	0.09	0.10	0.16
5	26	0.16	0.06	0.11	0.20

Table 4. 8: Fisher's LSD post hoc test on mean IoU among the validation datasets

(I) Validation Datasets	(J) Validation Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.12*	0.00	-0.18	-0.06
	3	-0.16*	0.00	-0.22	-0.10
	4	-0.04	0.18	-0.1	0.01
	5	-0.03	0.23	-0.09	0.02
2	1	0.12*	0.00	0.06	0.18
	3	-0.03	0.19	-0.09	0.02
	4	0.08*	0.00	0.02	0.14
	5	0.09*	0.00	0.03	0.15
3	1	0.16*	0.00	0.10	0.22
	2	0.03	0.19	-0.02	0.09
	4	0.12*	0.00	0.06	0.18
	5	0.13*	0.00	0.07	0.19
4	1	0.04	0.18	-0.01	0.10
	2	-0.08*	0.00	-0.14	-0.02
	3	-0.12*	0.00	-0.18	-0.06
	5	0.00	0.88	-0.05	0.06
5	1	0.03	0.23	-0.02	0.09
	2	-0.09*	0.00	-0.15	-0.03
	3	-0.13*	0.00	-0.19	-0.07
	4	-0.00	0.88	-0.06	0.05

* The mean difference showed significant difference ($p < 0.05$).

To compare the mean IoU values across different validation datasets, a one-way ANOVA was performed, followed by Fisher's LSD post hoc analysis. The analysis indicated a significant overall difference among the five validation datasets regarding the mean IoU, with an F-value of 11.40 (p = 0.00). Table 4.7 and 4.8 demonstrated that the mean IoU for validation Dataset 3 (0.29 ± 0.06) and Dataset 2 (0.25 ± 0.05) showed no significant difference, with p = 0.19. Additionally, the mean IoU for validation Dataset 1 (0.12 ± 0.04), Dataset 4 (0.16 ± 0.09) and Dataset 5 (0.16 ± 0.06) also showed no significant difference among them.

4.5 Dice coefficient (DC) during the testing phase

The collected data can be located in Appendix A. Figure 4.5 demonstrates that the DC increases as the epoch number increases across all testing datasets.Dataset 2 consistently exhibited a higher DC throughout the epochs compared to the other datasets. At the 10th epoch, Datasets 2 and 3 had similar DC values, whereas Datasets 4 and 5 displayed similar increases in DC over the epochs.

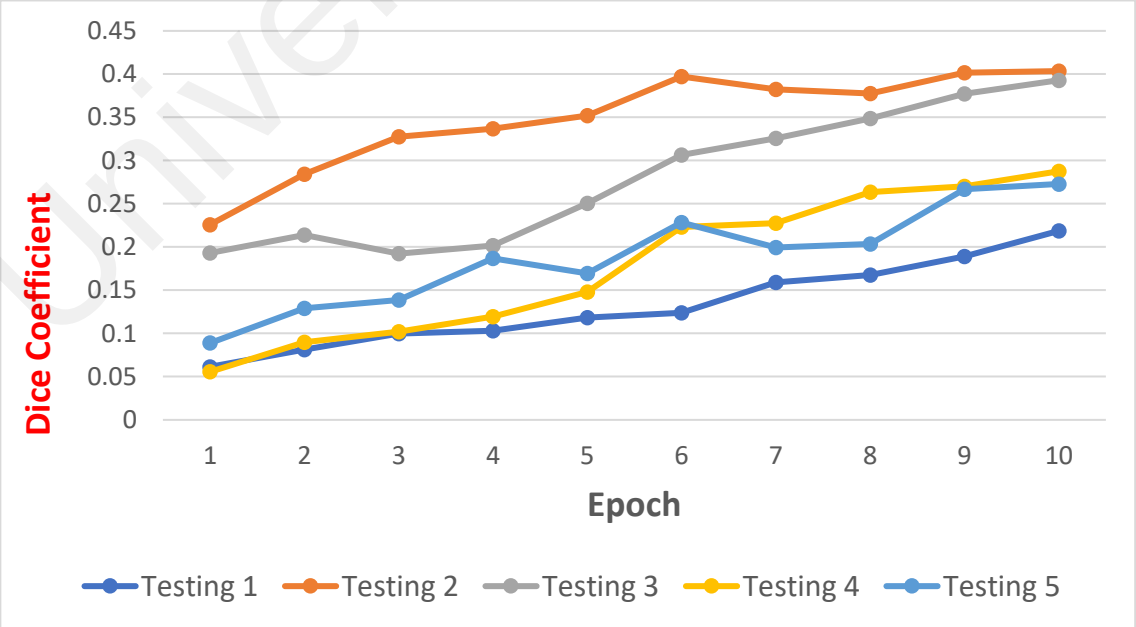

Figure 4. 5: Dice coefficient (DC) over epoch of different testing datasets

Descriptive analysis was implemented on the mean and standard deviation of DC for all testing datasets. The findings are represented in Table 4.9 and Table 4.10.

Table 4. 9: Mean DC among the different testing datasets during testing phase

Testing Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	13	0.13	0.04	0.09	0.16
2	13	0.34	0.05	0.30	0.39
3	13	0.28	0.07	0.22	0.33
4	13	0.17	0.08	0.11	0.23
5	13	0.18	0.05	0.14	0.23

Table 4. 10: Fisher's LSD post hoc test on mean DC among the testing datasets

(I) Testing Datasets	(J) Testing Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.21*	0.00	-0.27	-0.15
	3	-0.14*	0.00	-0.20	-0.08
	4	-0.04	0.13	-0.10	0.01
	5	-0.05	0.06	-0.11	0.00
2	1	0.21*	0.00	0.15	0.27
	3	0.06*	0.02	0.00	0.12
	4	0.17*	0.00	0.10	0.23
	5	0.16*	0.00	0.09	0.22
3	1	0.14*	0.00	0.08	0.20
	2	-0.06*	0.02	-0.12	-0.00
	4	0.10*	0.00	0.04	0.16
	5	0.09*	0.00	0.03	0.15
4	1	0.04	0.13	-0.01	0.10
	2	-0.17*	0.00	-0.23	-0.10
	3	-0.10*	0.00	-0.16	-0.04
	5	-0.00	0.74	-0.07	0.05
5	1	0.05	0.06	-0.00	0.11
	2	-0.16*	0.00	-0.22	-0.09
	3	-0.09*	0.00	-0.15	-0.03
	4	0.00	0.74	-0.05	0.07

* The mean difference showed significant difference ($p < 0.05$).

A one-way ANOVA followed by post hoc Fisher's LSD test was conducted to compare the mean DC among the different testing datasets. The analysis revealed a significant overall difference in the mean DC among the five datasets, $F = 16.70$, $p = 0.00$.

The result from the multiple comparisons with the mean DC is shown in Table 4.9 and 4.10, indicating that Dataset 2 (0.34 ± 0.05) was the highest compared to the other testing datasets. Dataset 1 (0.13 ± 0.04), Dataset 4 (0.17 ± 0.08) and Dataset 5 (0.18 ± 0.05) showed no significant difference in mean DC during testing phase.

4.6 Intersection over Union (IoU) during the testing phase

The collected data is available in Appendix B. Figure 4.6 shows the IoU trend among the testing datasets across epochs. Dataset 2 consistently had a higher IoU throughout the epochs compared to the other datasets. Datasets 1, 4 and 5 exhibited similar increase in IoU over the epochs. By the 10th epoch, Datasets 4 and 5 had a close to similar IoU values.

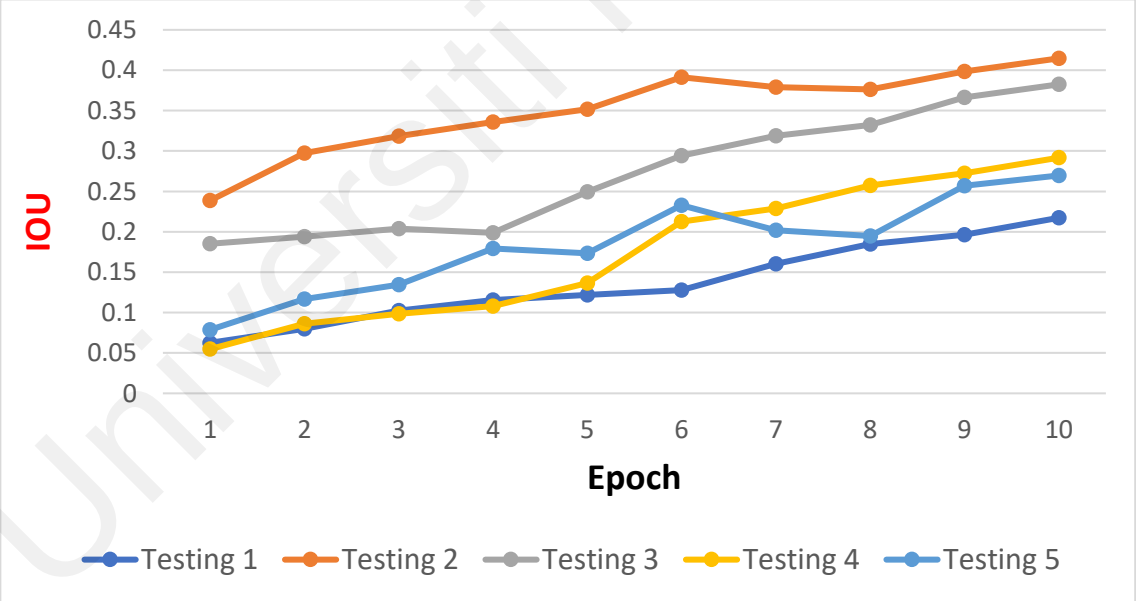

Figure 4. 6: Intersection over Union (IoU) over epoch of different testing datasets

Descriptive analysis was conducted on the mean and standard deviation of IoU for all testing datasets, with the results displayed in Table 4.11 and Table 4.12.

Table 4. 11: Mean IoU among the different testing datasets during testing phase

Testing Dataset	Sample size, n	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Minimum	Maximum
1	13	0.13	0.05	0.10	0.17
2	13	0.35	0.05	0.31	0.38
3	13	0.27	0.07	0.21	0.32
4	13	0.17	0.08	0.11	0.23
5	13	0.18	0.06	0.14	0.22

Table 4. 12: Fisher's LSD post hoc test on mean IoU among the testing datasets

(I) Testing Datasets	(J) Testing Datasets	Mean Difference (I-J)	p – value	95% Confidence Interval	
				Lower Bound	Upper Bound
1	2	-0.21*	0.00	-0.27	-0.15
	3	-0.13*	0.00	-0.19	-0.07
	4	-0.03	0.21	-0.09	0.02
	5	-0.04	0.12	-0.10	0.01
2	1	0.21*	0.00	0.15	0.27
	3	0.07*	0.01	0.01	0.13
	4	0.17*	0.00	0.11	0.23
	5	0.16*	0.00	0.10	0.22
3	1	0.13*	0.00	0.07	0.19
	2	-0.07*	0.01	-0.13	-0.01
	4	0.09*	0.00	0.03	0.15
	5	0.08*	0.00	0.02	0.14
4	1	0.03	0.21	-0.02	0.09
	2	-0.17*	0.00	-0.23	-0.11
	3	-0.09*	0.00	-0.15	-0.03
	5	-0.00	0.76	-0.06	0.05
5	1	0.04	0.12	-0.01	0.10
	2	-0.16*	0.00	-0.22	-0.10
	3	-0.08*	0.00	-0.14	-0.02
	4	0.00	0.76	-0.05	0.06

* The mean difference showed significant difference ($p < 0.05$).

A one-way ANOVA was performed to compare the mean IoU values across different testing datasets, followed by Fisher's LSD post hoc analysis. The analysis demonstrated a significant overall difference among the five testing datasets regarding

the mean IoU, with an F-value of 16.54 and p – value of 0.00. Table 4.11 and 4.12 showed a significant difference in the mean IoU in testing Dataset 2 (0.35 ± 0.05), which was the highest among all the other testing datasets. Dataset 1 (0.13 ± 0.05), Dataset 4 (0.17 ± 0.08) and Dataset 5 (0.18 ± 0.06) showed no significant difference in mean IoU during testing phase.

CHAPTER 5: DISCUSSION

5.1 Methodology

5.1.1 Dataset preparation

In the current study, pulp segmentation was performed by the same observer on all 130 extracted MFPs using ITK-SNAP software (Version 4.0.2). This software allows for semi-automatic segmentation, which streamlines the pulp segmentation process. The reliability of the pulp segmentation process was evaluated using the intraclass correlation coefficient (ICC). In this study, an ICC value of 0.95 was obtained, indicating an excellent level of reliability in the segmentation results (Koo & Li, 2016). Extracted MFPs was selected in the present study as to eliminate interference from surrounding tissues, facilitating accurate segmentation of the pulpal space (Timme et al., 2021).

ITK-SNAP is a popular software tool designed for interactive and semi-automatic segmentation of 3D medical images, including those from MRI, CT, PET, and 3D ultrasound. The semi-automatic approach depends on the areas of interest being distinguished from surrounding tissues by different image intensities, textures, or location (Yushkevich et al., 2016). A study comparing semi-automatic and manual segmentation of MRI images reported that the semi-automatic approach reduced segmentation time and enhanced accuracy (Yushkevich et al., 2018). Researchers from earlier studies have employed this open-source software for 3D pulp segmentation and pulp volume reconstruction to estimate the age, gender, and ethnicity (Fardim et al., 2020; Koçak & Okkesim, 2021; Marroquin Penaloza et al., 2016; Merdietio Boedi et al., 2022).

Within our methodological framework, ITK-SNAP software was implemented during the image preparation stage before feeding them into the U-Net model. This facilitates the creation of precise ground truth labels essential for training the model,

which is crucial for supervised machine learning. This procedure has supported the evolution of various machine learning algorithms for tasks in the field of dentistry such as, dental CBCT segmentation and lesion detection (Zheng et al., 2021), tooth segmentation and classification (Lahoud et al., 2021; Shaheen et al., 2021), and tooth and pulp segmentation (Duan et al., 2021).

5.1.2 Generating random datasets

In the present study, the Python random library, a tool for generating random numbers, was employed to create five distinct datasets. These datasets were used to assess the consistency of U-Net model in pulpal segmentation of MFPs. Multiple randomised datasets help in evaluating the reliability and generalisability of the segmentation outcomes provided by the model. Similar to a previous study, the authors evaluated the model's performance on five datasets for classification, finding it sufficient and appropriate for the task (Ahad et al., 2020).

5.1.3 70: 20: 10 learning model

The training, validation, and testing dataset in the present study follows the 70: 20: 10 split which emphasizes the need for comprehensive analysis and evaluation of machine learning models (Nguyen et al., 2021; Sekeroglu et al., 2022). This approach was supported by earlier studies Nguyen et al. (2021) and Pham et al. (2018), which demonstrated that using 70% of the data for training and validating models was most effective. Tan et al. (2024) employed a data split of 70% for training, 10% for validation, and 20% for testing, while Duman et al. (2024) utilised an 80% split for training and a 10% split for both validation and testing datasets. Both studies reported high accuracy in the specific tasks assigned, which are tooth substructure segmentation and detection of a second mesio-buccal canal in maxillary molars respectively (Duman et al., 2024; Tan et al., 2024). However, the methodology in our current research was in conflict with

previous study, where the authors found that a 60 : 40 ratio outperformed the ratios of 70 : 30 and 80 : 20 in the training datasets, although it did not significantly affect the model's overall performance (Sekeroglu et al., 2022). Nonetheless, the study also pointed out that comparing ML models is a difficult and intricate endeavour due to the wide range of models employed, the diversity of datasets used, and the various training strategies implemented (Sekeroglu et al., 2022). This is evident in a research paper where ML models were used to predict cancer incidence rates between males and females (Sekeroglu & Tuncal, 2021). The findings indicated that the female group required lower training ratios (60%) compared to the male group (80%) to achieve optimal results (Sekeroglu & Tuncal, 2021).

5.1.4 U-Net architecture in MONAI framework and data augmentation

The model deployed in the present study consists of five layers, with channels designed to process input data of sizes 16, 32, 64, 128, and 256. This slightly differs from the original U-Net model, which had a minimum resolution of 32 pixels (Ronneberger et al., 2015). After multiple trials and errors, we obtained the optimal sizes for our datasets to ensure compatibility and performance with the U-Net model, resulting in a minimum resolution of 16 pixels for our model. In the current study, the MONAI framework, an open-source platform based on PyTorch for DL in medical imaging, has been used to deploy the U-Net model. MONAI helps streamline and standardise the current research workflow, making it more efficient, simpler, and more consistent. Reproducing computational studies involves executing pre-existing code in different context, which can introduce variations between codes and potentially complicate and affect the outcomes (Cardoso et al., 2022). Several publications have demonstrated the effectiveness of U-Net model for internal organs and head and neck structures segmentation, using the MONAI framework (LaBella et al., 2022; Popa et al., 2022).

Data augmentation was performed using the MONAI framework. The types of augmentation deployed to the training and validation dataset in the present study were random rotations, horizontal and vertical flips, and adjustments to the brightness and contrast. The advantages of employing various data augmentation techniques during training improves the model's performance and robustness (Mikołajczyk & Grochowski, 2018). Numerous studies have reported that applying multiple augmentation techniques per image consistently enhances test accuracy leading to an improved generalisation (Ethiraj & Bolla, 2022; Fort et al., 2021).

5.1.5 Dice coefficient (DC)

The DC is pivotal in assessing model performance, serving as a statistical measure that quantifies the similarity between two datasets (Zou et al., 2004). It's value ranges from 0 to 1, where 0 indicates no similarity and 1 indicates a perfect matching. In this study, it specifically evaluates the resemblance between the U-Net model's predicted pulp segmentation of extracted MFPs and the ground truth. Many reported studies have utilized the DC to benchmark the accuracy of segmentation models in dental imaging. Lahoud et al. (2021) have used DC to compare between fully automated AI-driven and refined AI-driven tooth segmentation with the reference image, while Duan et al. (2021) uses DC to measure the accuracy for segmentation of tooth and pulp cavity via a two-phase DL approach. Similarly, various studies have employed the DC to evaluate the accuracy of pulpal segmentation by DL models (Hatvani et al., 2019; Sherwood et al., 2021; Wang et al., 2023). The widespread use of the DC highlights its importance in advancing the reliability and clinical relevance of segmentation models in dental research.

5.1.6 Intersection over Union (IoU)

IoU is an evaluation metric in image segmentation that calculates the overlap ratio between predicted segmentation masks and ground truth data (Müller et al., 2022). Parallel to the DC, it's value ranges from 0 to 1, with 0 showing no overlap between the predicted image and the ground truth regions, and 1 showing precise fit between both images. Higher IoU values indicate better segmentation performance. Several studies have incorporated IoU as an evaluation metric for automatic tooth segmentation (Chen et al., 2020; Dou et al., 2022; Lahoud et al., 2021; Li et al., 2020; Shaheen et al., 2021; Yang et al., 2020). Specifically, in the context of pulpal space segmentation, Duman et al. (2024) employed a You Only Look Once v5 (YOLOv5x) architecture, a different DL model to detect second mesio-buccal canals in maxillary molars, utilising IoU for performance evaluation. The findings from the present study will further highlight the importance of IoU as an evaluation metric in the investigation of pulpal segmentation in MFPs.

5.2 Results

5.2.1 DC in training, validation and testing phase

The primary objective of this study was to compare the DC of U-Net in pulp space segmentation of MFPs across five sets of randomly generated training, validation, and testing datasets from the same data pool. Based on our findings, we rejected the null hypothesis that there is no significant difference in the DC of U-Net model in pulp space segmentation of MFPs among all five randomly generated training, validation, and testing datasets from the same data pool during their respective phases. In the present study, there was no significant difference in the mean DC of the U-Net model between datasets 1, 4, and 5 during all phases. Additionally, datasets 2 and 3 did not show any significant

difference during the validation stage only. As per our knowledge, there were no studies that compares the consistency of the U-Net based on difference datasets generated from the same data pool, the lack of significant difference in the mean DC of the U-Net model between datasets 1,4, and 5 across all phases and between datasets 2 and 3 during the validation stage, could be due to the similar pattern in the distribution of the root canal morphology in their respective training, validation and testing dataset, leading to comparable segmentation performance by the U-Net model.

The mean DC for all five datasets during the training, validation and testing phase were below than the ideal value of 1 in predicting the segmentation image of the pulpal space of MFPs. In contrast to other reported literature, Hatvani et al. (2019) compared the U-Net model's performance in enhancing the resolution of 2D CBCT dental images, using the micro-CT images as ground truth. The authors found that U-Net with subpixel image enhancement achieved a high DC value of 0.91, while U-Net alone without any image enhancement had a DC value of 0.89. The difference in DC compared to our study lies in the reported methodology, where the authors trained the model with 5680 axial view slices from 13 teeth, rather than using individual teeth. This approach increased the training dataset size compared to our study. Additionally, their model was trained for 20 epochs, allowing it to learn more from the training data and thereby improving its performance. Similar to this study, one paper employed a loss function technique with and without adding a smooth term to the CBCT images to improve accuracy. This resulted in a DC of 0.87 without smooth term and 0.88 with the smooth term for single-rooted teeth, and 0.86 without the smooth term and 0.87 with the smooth term for multi-rooted teeth (Duan et al., 2021). Addition of smooth term will allow the predicted image to have smoother boundaries, reducing zig-zag edges and isolated errors, resulting in a more accurate predicted image and a high DC value. Duan et al. (2021) employed a smooth term by calculating the difference between pixels across the entire image. This approach

made the predicted boundaries more unified, reduced the number of isolated pixels, and resulted in a smoother overall image. Sherwood et al. (2021) reported that U-Net achieved a mean DC of 0.76 in detecting and classifying C-shaped canals in mandibular second molars during the training phase, and a mean DC of 0.66 during the testing phase. These results were still higher than those in the current study because the number of training and validation datasets was increased through data augmentation, resulting in 3780 axial image slices for the training dataset and 1620 axial image slices for the validation dataset. Additionally, the classification of the C-shaped canals was identified first, and the model was exposed to various type of C-shaped canals during the training stage, allowing the model to be able to identify them during testing stage. In contrast with the present study, the training and validation datasets were not increased through augmentation.

According to our results, datasets 1, 4, and 5 showed no significant difference in the mean DC across all phases. This may be attributed to the similar pattern of distribution of the types of MFPs pulpal configuration present in their training, validation, and testing datasets, resulting in comparable performance. Supporting this observation, a study reported by Sherwood et al. (2021) compared the accuracy of segmentation and classification of C-shaped canals in mandibular second molars using three types of DL models: U-Net, Residual U-Net, and Xception U-Net. The study found no significant difference in the mean DC during the training and validation stages. This outcome was because the models were trained separately with the five different types of C-shaped canals based on Fan et al. (2004) classification, and their performances were tested using datasets that contained all five types of C-shaped canals. The consistent performance across different models and datasets highlights the importance of having a representative distribution of morphological variations in the training dataset to achieve stable and reliable segmentation outcomes.

5.2.2 IoU in training, validation, and testing phase

The second objective of this study was to compare the mean IoU of U-Net in pulp space segmentation of MFPs across five sets of randomly generated training, validation, and testing datasets from the same data pool. As indicated by our findings, the null hypothesis was rejected, indicating a significant difference in mean IoU of U-Net model in pulp space segmentation among all five randomly generated training, validation, and testing datasets from the same data pool during their respective phases. However, there was no significant difference in the mean IoU between datasets 1 and 5, and datasets 3 with 4 and 5 during training phase. Additionally, datasets 2 and 3 did not show any significant difference during the validation phase only and datasets 1, 4, and 5 showed no significant difference in mean IoU during validation and testing phase.

A study demonstrated a high IoU value of 0.88 in detecting a second mesio-buccal canal in non-endodontically treated maxillary molars employing a different type on DL model called, YOLOv5x (Duman et al., 2024). In contrast to our findings, their model was trained for 500 epochs using 922 axial image slices from 153 patients, compared to the parameters used in our study, whereby the U-Net model was trained over 10 epochs on 91 training datasets only. Currently, there are few to no studies that utilise IoU to measure the consistency of the U-Net model in pulpal segmentation (Duman et al, 2024). In our findings, the mean IoU for dataset 1 and 5 remained consistent throughout the phases, as we hypothesise that it could be due to the similar distribution of MFPs pulpal configurations in their training datasets.

5.3 Limitations of study

One of the limitations of the study was that the CBCT images were obtained from a previously conducted postgraduate research, and the mounting of the extracted MFPs

on the vinyl polysiloxilane putty material was not standardised, leading to variation in the CBCT slices for each tooth. Additionally, the variation in the types of pulpal configurations of the extracted MFPs was not separated and was randomly distributed into the datasets. Another limitation of this study was the lack of computational power. This constraint significantly impacted our ability to increase the model complexity and fully optimise the U-Net architecture for pulpal space segmentation. Limited computational resources restricted the size of the datasets the model could process simultaneously, for example we could split the types of pulpal configuration into different datasets, training and validating them simultaneously. Furthermore, the lack of computational power resulted in longer training times, which further limited the number of experiments and iterations the model could perform. In the present study, the total time taken to run 1 complete dataset from training to testing phase was two hours. Consequently, this may have affected the overall DC and IoU values in the current findings in pulpal space segmentation as the model does not have sufficient training iterations resulting in a model that is not fully trained and optimised adequately. A lack of sufficient training data can lead to overfitting, reduced generalisation, and poor performance on diverse cases, ultimately lowering the current model's accuracy in pulpal space segmentation. Addressing all these limitations should be a priority for future research.

CHAPTER 6: CONCLUSION

6.1 Conclusion

Given the limitations of this study, the following conclusions can be derived from the findings:

1. Datasets 1, 4, and 5 demonstrated consistency in mean DC across all three phases, and the mean IoU remained consistent during the validation and testing phases.
2. Datasets 2 and 3 exhibited consistencies in both the mean DC and mean IoU during the validation phase.
3. Datasets 1 and 5 and datasets 3, with 4 and 5 demonstrated consistency in mean IoU during the training phase.
4. The mean DC and IoU for all five datasets during the training, validation, and testing phases were below the ideal value of 1, indicating that the model demonstrates some accuracy but falls short of achieving perfect segmentation, thus lacking optimal accuracy and reliability.
5. The consistency of the U-Net model is affected even though the five datasets were randomly generated from the same data pool.
6. The U-Net model demonstrated the lowest performance in mean DC and IoU with dataset 1 across all phases, indicating that dataset 1 presented greater challenges for the model, possibly due to differences in the data or more complex segmentation tasks.

6.2 Recommendations

Future research should aim to enhance computational power. Increasing computational resources will allow for more complex model architectures. By increasing computational resources, it will be possible to divide the extracted MFPs into separate training channels based on their respective Vertucci classifications (Vertucci, 1984). This approach will facilitate the handling of larger datasets and allow for extensive hyperparameter tuning. Ultimately, these improvements will lead to greater accuracy and robustness in pulpal space segmentation results. Another recommendation is by increasing the sample size of the training and validation datasets through augmentation techniques. Adding data augmentation can artificially expand the dataset, helping to improve the model's generalisation in pulpal segmentation. Moreover, future research should consider the types of pulpal configurations in the datasets and ensure they are evenly distributed in the training datasets to enhance the model's learning algorithm.

REFERENCES

Aboussaleh, I., Riffi, J., Fazazy, K. E., Mahraz, M. A., & Tairi, H. (2023). Efficient U-Net Architecture with Multiple Encoders and Attention Mechanism Decoders for Brain Tumor Segmentation. *Diagnostics (Basel)*, 13(5), 1-19.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, 22(5), 717-727.

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1), 1-9.

Ahad, A., Yimer, M., & Bhatti, Z. (2020). Identifying the Machine Learning Techniques for Classification of Target Datasets. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 4, 45-52.

Almeida, G., & Tavares, J. M. R. S. (2021). Versatile Convolutional Networks Applied to Computed Tomography and Magnetic Resonance Image Segmentation. *Journal of Medical Systems*, 45(8), 79-89.

Arifin, A. Z., Tanuwijaya, E., Nugroho, B. A., Priyatno, A. M., Indraswari, R., Astuti, E. R., & Navastara, D. A. (2019). Automatic Image Slice Marking Propagation on Segmentation of Dental CBCT. *TELKOMNIKA Telecommunication Computing Electronics and Control*, 17, 3218-3225.

Aung, N. M., & Myint, K. K. (2021). Diagnostic Accuracy of CBCT for Detection of Second Canal of Permanent Teeth: A Systematic Review and Meta-Analysis. *International Journal of Dentistry*, 2021, 1-18.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., & Davatzikos, C. (2017). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The cancer imaging archive*, 286.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., & Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, *4*, 1-13.

Benson, E., Pound, M. P., French, A. P., Jackson, A. S., & Pridmore, T. P. (2019). Deep hourglass for brain tumor segmentation. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4.

Bindal, P., Bindal, U., Lin, C. W., Kasim, N. H. A., Ramasamy, T., Dabbagh, A., Salwana, E., & Shamshirband, S. (2017). Neuro-fuzzy method for predicting the viability of stem cells treated at different time-concentration conditions. *Technology and Health Care*, 25(6), 1041-1051.

Bürklein, S., Heck, R., & Schäfer, E. (2017). Evaluation of the Root Canal Anatomy of Maxillary and Mandibular Premolars in a Selected German Population Using Cone-beam Computed Tomographic Data. *Journal of Endodontics*, 43(9), 1448-1452.

Campo, L., Aliaga, I. J., De Paz, J. F., García, A. E., Bajo, J., Villarubia, G., & Corchado, J. M. (2016). Retreatment Predictions in Odontology by means of CBR Systems. *Computational Intelligence and Neuroscience*, *2016*, 7485250, 1-11.

Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., Nath, V., He, Y., Xu, Z., Hatamizadeh, A., Zhu, W., Liu, Y., Zheng, M., Tang, Y., Yang, I., Feng, A. (2022). *MONAI: An open-source framework for deep learning in healthcare.* Accessibility Forum 2024. *https://arxiv.org/abs/2211.02701*

Celikten, B., Tufenkci, P., Aksoy, U., Kalender, A., Kermeoglu, F., Dabaj, P., & Orhan, K. (2016). Cone beam CT evaluation of mandibular molar root canal morphology in a Turkish Cypriot population. *Clinical Oral Investigations*, 20(8), 2221-2226.

Chen, Y., Du, H., Yun, Z., Yang, S., Dai, Z., Zhong, L., Feng, Q., & Yang, W. (2020). Automatic Segmentation of Individual Tooth in Dental CBCT Images from Tooth Surface Map by a Multi-Task FCN. *IEEE Access*, 8, 1-14.

Choi, H., & Jin, K. H. (2016). Fast and robust segmentation of the striatum using deep convolutional neural networks. *Journal of Neuroscience Methods*, 274, 146-153.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016.

Cirillo, M., Abramian, D., & Eklund, A. (2021). *What is The Best Data Augmentation For 3D Brain Tumor Segmentation?* https://doi.org/10.1109/ICIP42928.2021.9506328

Ding, H., Wu, J., Zhao, W., Matinlinna, J. P., Burrow, M. F., & Tsoi, J. K. (2023). Artificial intelligence in dentistry—A review. *Frontiers in Dental Medicine*, *4*, 1-13.

Dou, W., Gao, S., Mao, D., Dai, H., Zhang, C., & Zhou, Y. (2022). Tooth instance segmentation based on capturing dependencies and receptive field adjustment in cone beam computed tomography. *Computer Animation and Virtual Worlds*, 33(5), 1-16.

Duan, W., Chen, Y., Zhang, Q., Lin, X., & Yang, X. (2021). Refined tooth and pulp segmentation using U-Net in CBCT image. *Dento maxillo facial radiology*, 50(6), 1-8. https://doi.org/10.1259/dmfr.20200251

Duman, Ş. B., Çelik Özen, D., Bayrakdar, I. Ş., Baydar, O., Alhaija, E. S. A., Helvacioğlu Yiğit, D., Çelik, Ö., Jagtap, R., Pileggi, R., & Orhan, K. (2024). Second mesiobuccal canal segmentation with YOLOv5 architecture using cone beam computed tomography images. *Odontology*, 112(2), 552-561.

Durack, C., & Patel, S. (2012). Cone beam computed tomography in endodontics. *Brazilian Dental Journal*, 23(3), 179-191. https://doi.org/10.1590/s0103-64402012000300001

Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., & Schwendicke, F. (2019). Deep Learning for the Radiographic Detection of Apical Lesions. *Journal of Endodontics*, 45(7), 917-922.e915.

Endres, M. G., Hillen, F., Salloumis, M., Sedaghat, A. R., Niehues, S. M., Quatela, O., Hanken, H., Smeets, R., Beck-Broichsitter, B., Rendenbach, C., Lakhani, K., Heiland, M., & Gaudin, R. A. (2020). Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs. *Diagnostics (Basel)*, 10(6), 1-21.

England, M. C., Jr., Hartwell, G. R., & Lance, J. R. (1991). Detection and treatment of multiple canals in mandibular premolars. *Journal of Endodontics*, 17(4), 174-178.

Ethiraj, S., & Bolla, B. K. (2022). Augmentations: An Insight into their Effectiveness on Convolution Neural Networks. *Communications in Computer and Information Science Advances in Computing and Data Science*, 309-322.

Evgrafova, I., Egorova, L., Marchenko, A., & Tarasov, A. (2022). Ethical problems of practical interaction between strong artificial intelligence and humans in the urban transport environment and legal proceedings. *Transportation Research Procedia*, 63, 2094-2098.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., & Seiwald, K. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1), 67-70.

Fan, B., Cheung, G. S., Fan, M., Gutmann, J. L., & Fan, W. (2004). C-shaped canal system in mandibular second molars: Part II--Radiographic features. *Journal of Endodontics*, 30(12), 904-908.

Fardim, K., Junior, E., Rodrigues, R., Araújo, E., Gomes, J., Costa, A., Silva, P., & Lopes, S. (2020). Volume measurement of mandibular teeth pulp chamber as a prediction tool of gender and ethnicity in a Brazilian population. *Brazilian Dental Science*, 24(1), 1-6.

Fayad, M. I., & Johnson, B. R. (2023). The Impact of Cone Beam Computed Tomography in Nonsurgical and Surgical Treatment Planning. *3D Imaging in Endodontics: A New Era in Diagnosis and Treatment* (pp. 83-107). Springer.

Flowers, J. C. (2019). Strong and Weak AI: Deweyan Considerations. In *AAAI Spring Symposium: Towards Conscious AI Systems*, (Vol. 2287, pp: 1-7).

Forst, D., Nijjar, S., Flores-Mir, C., Carey, J., Secanell, M., & Lagravere, M. (2014). Comparison of in vivo 3D cone-beam computed tomography tooth volume measurement protocols. *Progress in Orthodontics*, *15*(1), 69-81.

Fort, S., Brock, A., Pascanu, R., De, S., & Smith, S. L. (2021). Drawing Multiple Augmentation Samples Per Image During Training Efficiently Decreases Test Error. ArXiv, abs/2105.13343. https://doi.org/10.48550/arXiv.2105.13343

Fu, D., Shi, K., Wu, F., Zhu, F., & Zhu, H. (2023). An artificial intelligence system for detecting the root canal morphology of the mandibular first premolar on panoramic radiographs. *Research Square,* 1-13.

Fukuda, M., Inamoto, K., Shibata, N., Ariji, Y., Yanashita, Y., Kutsuna, S., Nakata, K., Katsumata, A., Fujita, H., & Ariji, E. (2020). Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiology*, *36*(4), 337-343.

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193-202.

Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., Keizer, K., de Leeuw, F. E., Ginneken, B. V., Marchiori, E., & Platel, B. (2017). Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. *Neuroimage Clinical*, *14*, 391-399.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27, 2672-2680.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE transactions on knowledge and data engineering,* 35(4), 3313-3332.

Gutmann, J. L., & Fan, B. (2020). Tooth morphology and pulpal access cavitites. In H. K. Berman LH (Ed.), *Cohen's Pathway of the Pulp* (pp. 192-235). Elsevier Inc., St. Louis Missouri.

Han, X.-F., Laga, H., & Bennamoun, M. (2019). Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE transactions on pattern analysis and machine intelligence*, 43(5), 1578-1604.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* (2nd ed.). Springer.

Hatvani, J., Horváth, A., Michetti, J., Basarab, A., Kouamé, D., & Gyöngy, M. (2019). Deep Learning-Based Super-Resolution Applied to Dental Computed Tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3, 120-128.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., & Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35, 18-31.

Haque, I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18, 1-12.

He, K., Girshick, R., & Dollár, P. (2019). Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4918-4927).

Hendee, W. R., & Edwards, F. M. (1986). ALARA and an integrated approach to radiation protection. In *Seminars in nuclear medicine* (Vol. 16, No. 2, pp. 142-150). WB Saunders.

Hendler, J. (2008). Avoiding another AI winter. *IEEE Intelligent Systems*, 23(02), 2-4.

Hiraiwa, T., Ariji, Y., Fukuda, M., Kise, Y., Nakata, K., Katsumata, A., Fujita, H., & Ariji, E. (2019). A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac Radiology*, 48(3), 1-7.

Hoen, M. M., & Pink, F. E. (2002). Contemporary endodontic retreatments: an analysis based on clinical treatment findings. *Journal of Endodontics*, 28(12), 834-836.

Hsu, K., Yuh, D.Y., Lin, S.C., Lyu, P.S., Pan, G.X., Zhuang, Y.C., Chang, C.C., Peng, H.H., Lee, T.Y., Juan, C.H., Juan, C.E., Liu, Y.J., & Juan, C.J. (2022). Improving performance of deep learning models using 3.5D U-Net via majority voting for tooth segmentation on cone beam computed tomography. *Scientific Reports*, 12(1), 1-15.

Hülsmann, M. (1990). Mandibular first premolar with three root canals. *Endodontics and Dental Traumatology*, 6(4), 189-191.

Iglovikov, V., & Shvets, A. (2018). Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*. https://doi.org/10.48550/arXiv.1801.05746

Isensee, F., Jäger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2019). Automated design of deep learning methods for biomedical image segmentation. *arXiv preprint arXiv:1904.08128*. https://doi.org/10.1038/s41592-020-01008-z

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2018). Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017*, Revised Selected Papers 3, (pp. 287-297). Springer International Publishing.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., & Wirkert, S. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*. https://doi.org/10.48550/arXiv.1809.10486

Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 70(10.3850), 978-981.

Jena, M., Mishra, S. P., & Mishra, D. (2018). A survey on applications of machine learning techniques for medical image segmentation. *Internationa Journal of Engineering & Technology*, 7(4), 4489-4495.

Kakehbaraei, S., Seyedarabi, H., & Zenouz, A. T. (2018). Dental Segmentation in Cone-beam Computed Tomography Images Using Watershed and Morphology Operators. *Journal of Medical Signals and Sensors*, 8, 119 - 124.

Kamrul Hasan, S. M., & Linte, C. A. (2019). U-NetPlus: A Modified Encoder-Decoder U-Net Architecture for Semantic and Instance Segmentation of Surgical Instruments from Laparoscopic Images. *Annual International Conference IEEE Engineering in Medicine and Biology Society*, 2019, 7205-7211.

Khanagar, S. B., Alfadley, A., Alfouzan, K., Awawdeh, M., Alaqla, A., & Jamleh, A. (2023). Developments and Performance of Artificial Intelligence Models Designed for Application in Endodontics: A Systematic Review. *Diagnostics (Basel)*, 13(3).

Koçak, N., & Okkesim, A. (2021). Usability of first molar teeth measurements for age and sex estimation in cone-beam computed tomography images. *Mersin Üniversitesi Tıp Fakültesi Lokman Hekim Tıp Tarihi ve Folklorik Tıp Dergisi*, 11(3), 576-587.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84 - 90.

Kühnisch, J., Meyer, O., Hesenius, M., Hickel, R., & Gruhn, V. (2022). Caries Detection on Intraoral Images Using Artificial Intelligence. *Journal of Dental Research*, 101(2), 158-165.

LaBella, D., Noah-Vermillion, L. K., Kincaid, R., Sanford, T., Sey, O., Hahn, S. S., LaCombe, M. A., & Rice, S. R. (2022). Efficacy of Medical Image Transformations within an Open Sourced Head and Neck Deep Learning Auto-Contouring Framework. *International Journal of Radiation Oncology*Biology*Physics*, 114(3), e99-e100.

Lahoud, P., EzEldeen, M., Beznik, T., Willems, H., Leite, A., Van Gerven, A., & Jacobs, R. (2021). Artificial Intelligence for Fast and Accurate 3-Dimensional Tooth Segmentation on Cone-beam Computed Tomography. *Journal of Endodontics*, 47(5), 827-835.

Langland, O. E., Langlais, R. P., & Preece, J. W. (2002). *Principles of dental imaging*. Lippincott Williams & Wilkins.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2018). Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry*, 77, 106-111.

Lee, S. J., Chung, D., Asano, A., Sasaki, D., Maeno, M., Ishida, Y., Kobayashi, T., Kuwajima, Y., Da Silva, J. D., & Nagai, S. (2022). Diagnosis of Tooth Prognosis Using Artificial Intelligence. *Diagnostics (Basel)*, 12(6).

Li, Q., Chen, K., Han, L., Zhuang, Y., Li, J., & Lin, J. (2020). Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN. *Journal of Xray Science and Technology*, 28(5), 905-922.

Liebowitz, J. (1995). Expert systems: A short introduction. *Engineering Fracture Mechanics*, 50(5-6), 601-607.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88.

Liu, N., Li, X., Liu, N., Ye, L., An, J., Nie, X., Liu, L., & Deng, M. (2013). A micro-computed tomography study of the root canal morphology of the mandibular first premolar in a population from southwestern China. *Clinical Oral Investigations*, 17(3), 999-1007.

Liu, Y., Olszewski, R., Alexandroni, E. S., Enciso, R., Xu, T., & Mah, J. K. (2010). The validity of in vivo tooth volume determinations from cone-beam computed tomography. *Angle Orthodontist*, 80(1), 160-166.

Lo, S. B., Lou, S. A., Lin, J. S., Freedman, M. T., Chien, M. V., & Mun, S. K. (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging*, 14(4), 711-718.

Long, J.-S., Ma, G.-Z., Song, E.-M., & Jin, R.-C. (2021). Learning U-Net Based Multi-Scale Features in Encoding-Decoding for MR Image Brain Tissue Segmentation. *Sensors*, 21(9), 1-20.

Lyksborg, M., Puonti, O., Agn, M., & Larsen, R. (2015). An ensemble of 2D convolutional neural networks for tumor segmentation. In *Image Analysis: 19th Scandinavian Conference, SCIA 2015, Copenhagen, Denmark, June 15-17, 2015. Proceedings 19,* (pp. 201-211). Springer International Publishing.

Madani, Z. S., Mehraban, N., Moudi, E., & Bijani, A. (2017). Root and Canal Morphology of Mandibular Molars in a Selected Iranian Population Using Cone-Beam Computed Tomography. *Iran Endodontic Journal*, 12(2), 143-148.

Marroquin Penaloza, T. Y., Karkhanis, S., Kvaal, S. I., Vasudavan, S., Castelblanco, E., Kruger, E., & Tennant, M. (2016). Reliability and repeatability of pulp volume reconstruction through three different volume calculations. *Journal of Forensic Odontostomatology*, 34(2), 35-46.

Martins, J. N. R., Francisco, H., & Ordinola-Zapata, R. (2017). Prevalence of C-shaped Configurations in the Mandibular First and Second Premolars: A Cone-beam Computed Tomographic In Vivo Study. *Journal of Endodontics*, 43(6), 890-895.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.

McDermott, J. P. (1980). RI: an Expert in the Computer Systems Domain. In *AAAI Conference on Artificial Intelligence* (Vol. 1, pp. 269-271).

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., . . . Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993-2024.

Merdietio Boedi, R., Shepherd, S., Oscandar, F., Mânica, S., & Franco, A. (2022). 3D segmentation of dental crown for volumetric age estimation with CBCT imaging. *International Journal of Legal Medicine*, 137(1), 123-130.

Mikołajczyk, A., & Grochowski, M. (2018, 9-12 May 2018). Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)* (pp. 117-122). IEEE.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV*) (pp. 565-571). IEEE.

Millioni, R., Sbrignadello, S., Tura, A., Iori, E., Murphy, E., & Tessari, P. (2010). The inter-and intra-operator variability in manual spot segmentation and its effect on spot quantitation in two-dimensional electrophoresis analysis. *Electrophoresis*, *31*(10), 1739-1742.

Moayedi, S., & Lata, D. (2004). 2. Mandibular first premolar with three canals. *Endodontology*, 16(1), 26-29.

Monteux, A. (2019). *Metrics for semantic segmentation*. https://ilmonteux.github.io/2019/05/10/segmentation-metrics.html

Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1), 210.

Nam, C. (2020). *Neuroergonomics Principles and Practice: Principles and Practice*. New York: Springer.

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, 2021, 1-15.

Orhan, K., Bayrakdar, I. S., Ezhov, M., Kravtsov, A., & Özyürek, T. (2020). Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *International Endodontic Journal*, 53(5), 680-689.

Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9), 1277-1294.

Pan, D., Zeng, A., Jia, L., Huang, Y., Frizzell, T., & Song, X. (2020). Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. *Frontiers in Neuroscience*, 14, 1-19.

Pan, J. Y. Y., Parolia, A., Chuah, S. R., Bhatia, S., Mutalik, S., & Pau, A. (2019). Root canal morphology of permanent teeth in a Malaysian subpopulation using cone-beam computed tomography. *BMC Oral Health*, 19(1), 1-15.

Patel, S., Brown, J., Semper, M., Abella, F., & Mannocci, F. (2019). European Society of Endodontology position statement: Use of cone beam computed tomography in Endodontics: European Society of Endodontology (ESE) developed by *International Endodontic Journal*, 52(12), 1675-1678.

Pauwels, R., Brasil, D. M., Yamasaki, M. C., Jacobs, R., Bosmans, H., Freitas, D. Q., & Haiter-Neto, F. (2021). Artificial intelligence for detection of periapical lesions on intraoral radiographs: Comparison between convolutional neural networks and human observers. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 131(5), 610-616.

Pham, B. T., Prakash, I., Jaafari, A., & Bui, D. T. (2018). Spatial Prediction of Rainfall-Induced Landslides Using Aggregating One-Dependence Estimators Classifier. *Journal of the Indian Society of Remote Sensing*, 46, 1457-1470.

Pineda, F., & Kuttler, Y. (1972). Mesiodistal and buccolingual roentgenographic investigation of 7,275 root canals. *Oral Surgergy, Oral Medicine, Oral Pathology*, 33(1), 101-110.

Poorni, S., Karumaran, C. S., & Indira, R. (2010). Mandibular first premolar with two roots and three canals. *Australian Endodontic Journal*, 36(1), 32-34.

Popa, D. L., Popa, R. T., Barbulescu, L., Ivanescu, R. C., & Mocanu, M. L. (2022). Segmentation of Different Human Organs on 3D Computer Tomography and Magnetic Resonance Imaging using an Open Source 3D U-Net Framework. *2022 23rd International Carpathian Control Conference (ICCC)*, 54-57.

Rahimi, S., Mokhtari, H., Ranjkesh, B., Johari, M., Frough Reyhani, M., Shahi, S., & Seif Reyhani, S. (2017). Prevalence of Extra Roots in Permanent Mandibular First Molars in Iranian Population: A CBCT Analysis. *Iranian Endodontic Journal*, 12(1), 70-73.

Rahimi, S., Shahi, S., Yavari, H. R., Manafi, H., & Eskandarzadeh, N. (2007). Root canal configuration of mandibular first and second premolars in an Iranian population. *Journal of Dental Research, Dental Clinics, Dental Prospects*, 1(2), 59-64.

Rahman, M., & Wang, Y. (2016). *Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation* (Vol. 10072, pp. 234-244). Springer, Cham.

Ravanbakhsh, M., Sangineto, E., Nabi, M., & Sebe, N. (2019, January). Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1896-1904). IEEE.

Reduwan, N. H., Abdul Aziz, A. A., Mohd Razi, R., Abdullah, E. R. M. F., Mazloom Nezhad, S. M., Gohain, M., & Ibrahim, N. (2024). Application of deep learning and feature selection technique on external root resorption identification on CBCT images. *BMC Oral Health*, 24(1), 1-10.

Robertson, D., Leeb, I. J., McKee, M., & Brewer, E. (1980). A clearing technique for the study of root canal systems. *Journal of Endodontics*, 6(1), 421-424.

Rohrer, C., Krois, J., Patel, J., Meyer-Lueckel, H., Rodrigues, J. A., & Schwendicke, F. (2022). Segmentation of dental restorations on panoramic radiographs using deep learning. *Diagnostics*, 12(6), 1-8.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.

Roy, S. B., Panda, A., & Naskar, R. (2019). An Automated Ensembled Deep Neural Network Approach towards Accurate Segmentation of Biomedical Images. *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, 240-245.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., & Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115, 211-252.

Saghiri, M. A., Asgar, K., Boukani, K. K., Lotfi, M., Aghili, H., Delvarani, A., Karamifar, K., Saghiri, A. M., Mehrvarzfar, P., & Garcia-Godoy, F. (2012). A new approach for locating the minor apical foramen using an artificial neural network. *International Endodontic Journal*, 45(3), 257-265.

Saghiri, M. A., Garcia-Godoy, F., Gutmann, J. L., Lotfi, M., & Asgar, K. (2012). The reliability of artificial neural network in locating minor apical foramen: a cadaver study. *Journal of Endodontics*, 38(8), 1130-1134.

Salineiro, F. C., Talamoni, I. P., Velasco, S. K., Barros, F. M., & Cavalcanti, M. G. P. (2019). Artifact induction by endodontic materials. *Clinical and Laboratorial Research in Dentistry*, 1-10.

Sanders, M. A., Hoyjberg, C., Chu, C. B., Leggitt, V. L., & Kim, J. S. (2007). Common orthodontic appliances cause artifacts that degrade the diagnostic quality of CBCT images. *Journal of the California Dental Association*, 35(12), 850-857.

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 1-20.

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, *3*(2), 1-20.

Sarraf, S., & Tofighi, G. (2016). Classification of Alzheimer's Disease Structural MRI Data by Deep Learning Convolutional Neural Networks. *ArXiv*, *abs/1607.06583*.

Schmidhuber, J. (2015). Deep learning. *Scholarpedia*, 10(11), 32832.

Schwendicke, F., Rossi, J. G., Göstemeyer, G., Elhennawy, K., Cantu, A. G., Gaudin, R., Chaurasia, A., Gehrung, S., & Krois, J. (2021). Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection. *Journal of Dental Research*, 100(4), 369-376.

Sekeroglu, B., Ever, Y. K., Dimililer, K., & Al-Turjman, F. (2022). Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems. *Data Intelligence*, 4(3), 620-652.

Sekeroglu, B., & Tuncal, K. (2021). Prediction of cancer incidence rates for the European continent using machine learning models. *Health Informatics Journal*, *27*(1), 1-15.

Shaheen, E., Leite, A., Alqahtani, K. A., Smolders, A., Van Gerven, A., Willems, H., & Jacobs, R. (2021). A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study. *Journal of Dentistry*, 115, 1-7.

Shearer, A. C., Horner, K., & Wilson, N. H. (1991). Radiovisiography for length estimation in root canal treatment: an in-vitro comparison with conventional radiography. *International Endodontic Journal*, 24(5), 233-239.

Sherwood, A. A., Sherwood, A. I., Setzer, F. C., Shamili, J. V., John, C., & Schwendicke, F. (2021). A Deep Learning Approach to Segment and Classify C-Shaped Canal Morphologies in Mandibular Second Molars Using Cone-beam Computed Tomography. *Journal of Endodontics*, 47(12), 1907-1916.

Slowey, R. R. (1979). Root canal anatomy. Road map to successful endodontics. *Dent Clinics of North America*, 23(4), 555-573.

Suk, H. I., Lee, S. W., & Shen, D. (2017). Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical Image Analysis*, 37, 101-113.

Talwar, S., Utneja, S., Nawal, R. R., Kaushik, A., Srivastava, D., & Oberoy, S. S. (2016). Role of Cone-beam Computed Tomography in Diagnosis of Vertical Root Fractures: A Systematic Review and Meta-analysis. *Journal of Endodontics*, 42(1), 12-24.

Tan, M., Cui, Z., Zhong, T., Fang, Y., Zhang, Y., & Shen, D. (2024). A progressive framework for tooth and substructure segmentation from cone-beam CT images. *Computers in Biology and Medicine*, 169, 1-9.

Tatnall, A. (2012). History of computers: hardware and software development. *Encyclopedia of Life Support Systems*.

Timme, M., Borkert, J., Nagelmann, N., Streeter, A., Karch, A., & Schmeling, A. (2021). Age-dependent decrease in dental pulp cavity volume as a feature for age assessment: a comparative in vitro study using 9.4-T UTE-MRI and CBCT 3D imaging. *International Journal of Legal Medicine*, 135(4), 1599-1609.

Torres, A., Jacobs, R., Lambrechts, P., Brizuela, C., Cabrera, C., Concha, G., & Pedemonte, M. E. (2015). Characterization of mandibular molar root and canal morphology using cone beam computed tomography and its variability in Belgian and Chilean population samples. *Imaging Science in Dentistry*, 45(2), 95-101.

Turing, A. M. (1980). Computing Machinery and Intelligence. *Creative Computing*, 6(1), 44-53.

Vertucci, F. J. (1984). Root canal anatomy of the human permanent teeth. *Oral Surgery, Oral Medicine, Oral Pathology*, 58(5), 589-599.

Wang, J., Zhu, H., Wang, S.-H., & Zhang, Y.-D. (2021). A Review of Deep Learning on Medical Image Analysis. *Mobile Networks and Applications*, 26(1), 351-380.

Wang, L., Li, J. P., Ge, Z. P., & Li, G. (2019). CBCT image based segmentation method for tooth pulp cavity region extraction. *Dentomaxillofacial Radiology*, 48(2), 1-11.

Wang, S., Chen, W., Xie, S. M., Azzari, G., & Lobell, D. B. (2020). Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2), 207-232.

Wang, Y., He, S., Yu, L., Li, J., & Chen, S. (2011). Accuracy of volumetric measurement of teeth in vivo based on cone beam computer tomography. *Orthodontics and Craniofacial Research*, 14(4), 206-212.

Wang, Y., Xia, W., Yan, Z., Zhao, L., Bian, X., Liu, C., Qi, Z., Zhang, S., & Tang, Z. (2023). Root canal treatment planning by automatic tooth and root canal segmentation in dental CBCT with deep multi-task feature learning. *Medical Image Analysis*, 85, 1-13. https://doi.org/10.1016/j.media.2023.102750

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M., & Chau, D. H. P. (2021). CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1396-1406.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

Wenger, E., Mårtensson, J., Noack, H., Bodammer, N. C., Kühn, S., Schaefer, S., Heinze, H. J., Düzel, E., Bäckman, L., Lindenberger, U., & Lövdén, M. (2014). Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. *Human Brain Mapping*, 35(8), 4236-4248.

Wu, J., Zhang, C., Xue, T., Freeman, W., & Tenenbaum, J. (2016). Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *Advances in neural information processing systems,* 29, 1486-1494.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4), 611-629.

Yang, Y., Xie, R., Jia, W., Chen, Z., Yang, Y., Xie, L., & Benxiang, J. (2020). Accurate and automatic tooth image segmentation model with deep convolutional neural networks and level set method. *Neurocomputing*, 419, 108-125.

Yoshioka, T., Villegas, J. C., Kobayashi, C., & Suda, H. (2004). Radiographic evaluation of root canal multiplicity in mandibular first premolars. *Journal of Endodontics*, 30(2), 73-74.

Yushkevich, P., Pashchinskiy, A., Oguz, I., Mohan, S., Schmitt, J. E., Stein, J. M., Zukić, D., Vicory, J., McCormick, M., Yushkevich, N., Schwartz, N., Gao, Y., & Gerig, G. (2018). User-Guided Segmentation of Multi-modality Medical Imaging Datasets with ITK-SNAP. *Neuroinformatics*, 17, 83-102.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3), 1116-1128.

Yushkevich, P. A., Yang, G., & Gerig, G. (2016). ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. *Annu Int Conf IEEE Engineering in Medicine and Biology Society*, 2016, 3342-3345.

Zhang, X., Xiong, S., Ma, Y., Han, T., Chen, X., Wan, F., Lu, Y., Yan, S., & Wang, Y. (2015). A Cone-Beam Computed Tomographic Study on Mandibular First Molars in a Chinese Subpopulation. *PLoS One*, 10(8), 1-9.

Zhao, L., & Jia, K. (2016). Multiscale CNNs for Brain Tumor Segmentation and Diagnosis. *Computational and Mathematical Methods in Medicine*, *2016*, 1-17.

Zheng, Z.-W., Yan, H., Setzer, F. C., Shi, K. J., Mupparapu, M., & Li, J. (2021). Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection. *IEEE Transactions on Automation Science and Engineering*, 18, 603-614.

Zhou, H., Cao, M., Min, Y., Yoon, S., Kishan, A., & Ruan, D. (2022). Ensemble learning and tensor regularization for cone-beam computed tomography-based pelvic organ segmentation. *Medical Physics*, 49(3), 1660-1672.

Zhou, Z.H. (2018). A brief introduction to weakly supervised learning. *National science review*, *5*(1), 44-53.

Zhu, X., & Goldberg, A. B. (2022). *Introduction to semi-supervised learning*.Springer Nature.

Zillich, R., & Dowson, J. (1973). Root canal morphology of mandibular first and second premolars. *Oral Surgery, Oral Medicine and Oral Pathology*, 36(5), 738-744.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., 3rd, Jolesz, F. A., & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology*, 11(2), 178-189.