Development of an Intelligent Model for Prediction of Periodontitis Stage and Grade

by

Nazila Ameli

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Medical Sciences- Dentistry

University of Alberta

ABSTRACT

The published and unpublished data in this thesis address a crucial gap in the utilization of comprehensive automated methods for diagnosing periodontitis. This research focuses on the development of a clinical decision support system (CDSS) for analyzing text and image data to diagnose periodontitis, aiming to enhance accuracy and efficiency in periodontitis stage and grade classification.

Periodontitis, a complex inflammatory disease affecting dental supporting tissues, poses significant risks to oral and systemic health, including tooth loss and associations with conditions like diabetes and coronary artery disease. Early detection is crucial to mitigate its impact and avoid invasive treatments. My thesis aims to predict periodontal disease stage and grade by integrating novel methods for measuring bone loss (BL) with natural language processing (NLP) analysis of patients' chart data.

In Chapter 2, utilizing Bidirectional Encoder Representations from Transformers (BERT), an automated methodology was developed to extract critical information from textual clinical notes for the classification of periodontitis stages and grades. This process involved fine-tuning the BERT model on a dataset of 309 clinical notes, which encompassed a diverse range of unstructured and narrative texts containing patient histories, medical and dental histories, possible treatment plans, and diagnostic observations.

Through meticulous pre-processing, including tokenization and normalization, the clinical notes were adequately prepared for the BERT model. The fine-tuned BERT model demonstrated a remarkable ability to accurately identify and classify periodontitis stages and grades based on the extracted information. The evaluation of the model's performance

indicated high accuracy, showcasing its potential to streamline the diagnostic process by providing precise classifications from unstructured clinical notes.

In Chapter 3, advanced deep learning (DL) models were utilized to analyze periapical (PA) radiographs, automating the process of BL segmentation and quantifying BL percentage relative to root length for predicting the stage and grade of periodontal disease. This approach involved the integration of U-Net and YOLO-V9 models to address specific aspects of the diagnostic process.

The U-Net model was trained to perform precise segmentation of BL areas within the radiographs. By leveraging its convolutional neural network (CNN) architecture, U-Net demonstrated a high level of accuracy in delineating the boundaries of BL regions. This automated segmentation provided a reliable foundation for further analysis. Concurrently, the YOLO-V9 model was utilized to detect and localize the coordinates of the apex of teeth within the radiographs. The YOLO-V9 model's performance indicated that its detection capabilities were comparable to those of the specialists. This ensured that the model could accurately identify critical anatomical landmarks necessary for subsequent measurements. By combining the outputs of these models, the maximum BL was measured as a percentage of root length. This quantitative assessment was then used to classify the stage and grade of periodontal disease. Through this chapter, it became evident that the U-Net model was successful in the correct segmentation of BL, and the YOLO-V9 model effectively detected the coordinates of the tooth apex. The integration of these models facilitated accurate classification of periodontal disease stages and grades, highlighting the potential of DL techniques in enhancing periodontal diagnostics and treatment planning.

Finally, in Chapter 4, by integrating these approaches through a multimodal transformer model, a comprehensive automated diagnosis system is proposed that enhances diagnostic accuracy and enables timely interventions to prevent adverse outcomes and costly treatments. This innovative system combines the strengths of textual clinical note analysis and PA radiograph examination through a multimodal transformer model.

By combining these modalities, the multimodal transformer model synthesizes textual and visual data, enabling an accurate assessment of each patient's condition. The model focuses on patients with the highest amount of BL, as identified through radiographic analysis, and correlates this data with the contextual information from clinical notes to classify the stage and grade of periodontal disease more accurately.

The findings demonstrated that this integrated approach significantly outperforms previous single-modality models. While BERT was highly effective for text processing and the U-Net and YOLO-V9 models excelled in image processing, the model achieved a higher classification accuracy by combining these methodologies. The multimodal transformer model's ability to analyze and integrate diverse data types leads to a more precise and comprehensive diagnosis, facilitating timely and effective interventions in periodontal disease management. This advanced system not only improves diagnostic outcomes but also has the potential to reduce the need for extensive treatments by enabling early and accurate identification of periodontal disease stages and grades.

In conclusion, this work demonstrates the potential for developing CDSS tools to improve periodontal disease management and patient outcomes.

PREFACE

This thesis is an original work by Nazila Ameli. The research projects, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Development of an intelligent reminder algorithm to reduce periodontal risks”, No. Pro00107743, on 02/18/2021

Chapter 2 of this thesis has been submitted as Ameli N, Firoozi T, Gibson M, Lai H. “Classification and Prediction of Periodontitis Stage and Grade using Natural Language Processing Techniques” in the Journal of PLoS Digital Health.

Chapter 3 has been submitted as two manuscripts as:

1- Ameli N, Gibson M, Ghaedsharaf S, Lai H. “Evaluating two approaches to automate segmentation of alveolar bone loss from periapical radiographs” in the Journal of Clinical and Experimental Dental Research, and

2- Ameli N, Gibson M, Kornerup I, Lagravere M, Gierl M, Lai H. "Automating Bone Loss Measurement on Periapical Radiographs for Predicting the Periodontitis Stage and Grade" in the Journal of Frontiers in Oral Health

Chapter 4 of the thesis has also been submitted as Ameli N, Gibson M, Kornerup I, Lagravere M, Gierl M, Lai H. “Multimodal Transformer Model for Automated Classification of Periodontitis Stage and Grade” in Journal of Clinical Periodontology.

DEDICATION

This work is dedicated to my sons, Amirsam and Arianik, the Sun and Moon of my life, whose presence brightens my world and fills it with endless love and joy. May you always be strong, happy, and fearless in the pursuit of your dreams!

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my beloved Mehdi for the countless ways you enrich my life every day. Reaching this point would have been impossible without your love, support, and care. Thank you for making me smile even in the darkest times and for holding my hands tightly, even when you faced your own challenges. Supporting me through my third thesis defense in my academic journey is just one of the many ways you have shown your constant commitment and love. You are my promise of a soulmate forever.

I am profoundly grateful to my parents for their endless care, sacrifices, and support. Despite the thousands of miles between us, you both offered unwavering support and encouragement, helping me complete this work. I cannot thank you enough for your love.

I would also like to express my sincere gratitude to my supervisor, Professor Hollis Lai, for his thoughtful encouragement and guidance throughout this journey. He has taught me the value of hard work in pursuing my dreams. Gaining admission to this rigorous program and completing it would not have been possible without his support and inspiration.

I extend my appreciation to my thesis committee members, Professors Manuel Lagravere, Mark Gierl, and Monica Gibson, and my dearest, Dr. Ida Kornerup for their expertise and kind guidance in developing and completing this research.

Finally, I gratefully acknowledge the financial support for this dissertation provided by the Network for Canadian Oral Health Research (NCOHR) Frontiers Seed Grant and the Alberta Innovates Graduate Student Scholarship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COMMON ABBREVIATIONS

Abbreviation	Definition
AI	artificial intelligence
ML	machine learning
CNN	convolutional neural network
RNN	recurrent neural network
DL	deep learning
DNN	deep neural network
NN	neural network
BL	bone loss
PA	periapical
CEJ	cemento-enamel junction
CAL	clinical attachment loss
RegEx	regular expression
LLM	large language model
FCN	fully connected network
BERT	bidirectional encoder representations from transformers

Abbreviation	Definition
LIME	local interpretable model-agnostic explanations
ReLU	rectified linear unit
LSTM	long short-term memory
SVM	support vector machine
MLP	multi-layer perceptron
NLP	natural language processing
IoU	intersection over union
PNG	portable network graphics
YOLO	you only look once
mAP	mean average precision
MAE	mean absolute error
RMSE	root mean square error
ICC	intraclass correlation coefficient
ROC	receiver operating characteristic
AUC	area under curve

Abbreviation	Definition
SSD	single-shot multibox detector
GELAN	generalized efficient layer aggregation network
PGI	programmable gradient information
ViT	vision transformer
MML	multimodal learning
CDSS	clinical decision support system
CHMS	Canadian health measures survey
OHRQoL	oral health-related quality of life
PPD	periodontal pocket depth
BOP	bleeding on probing
EDR	electronic dental records
SRP	scaling and root planing

Chapter 1- Introduction

The dissertation aims to address the area of machine learning (ML) applications in dentistry, with a particular focus on periodontology. The overarching objective is to bridge existing gaps in knowledge by demonstrating the potential and efficacy of artificial intelligence (AI) in dentistry, highlighting its role in improving decision-making processes and patient care. Through comprehensive reviews and empirical studies, the thesis endeavours for broader adoption and integration of AI methodologies for clinical classification tasks in periodontal practice, which will ultimately advance the field and enhance treatment outcomes for patients.

Periodontitis is a multifactorial and microbiome-associated inflammatory disease that occurs in the dental supporting tissues (periodontium, which includes gingiva, periodontal ligament, cementum, and alveolar bone presenting as bone loss (BL)). Progression of the disease can adversely affect oral and systemic health and result in tooth loss, reduction of masticatory performance as well as having association with diabetes, coronary artery disease and rheumatoid arthritis. Thus, periodontitis and its complications will impose substantially negative effects on oral health-related quality of life. Moreover, early detection and diagnosis of periodontitis can help in preventing the consequent costly and invasive dental treatment.

The purpose of my thesis is to predict the stage and grade of periodontal disease by: 1) introducing a novel method of measuring BL and 2) integrating the outcomes with the output of a natural language processing (NLP) model for extracting information from

patients' chart data (including medical/dental history and clinical findings obtained through examination). To achieve this purpose, I developed and implemented automated methods to:

1) Extract the important latent information in patients' charts (including the stage and grade of the disease), which are collected as text files in clinicians' notes to classify patients with periodontitis through an NLP model using bidirectional encoder representations from transformers (BERT), and utilize the outcome for predicting the periodontitis stage and grade, 2) Analyze patients' radiographs and measure the percentage of BL using the deep learning (DL) models to make predictions on periodontitis' stage and grade, and

3) Integrate the results obtained from the two steps above through a multimodal model to develop an automated approach for diagnosis of patients with periodontitis and predict the stage and grade of disease.

My work demonstrates potential in the steps required for developing a clinical decision support system (CDSS) tool by automating and integrating the review of patient notes and conducting BL measurements. This approach enhances the accuracy of periodontal disease diagnosis within the CDSS framework. By incorporating these advancements, timely interventions can be facilitated to prevent complications such as tooth loss, systemic health risks, and the need for costly and invasive treatments.

1.1 PERIODONTITIS

1.1.1 Definition

Periodontitis or periodontal disease encompasses a spectrum of conditions affecting the periodontium, which includes the gingiva, periodontal ligament, cementum, and alveolar bone. This range of diseases varies from gingivitis to viral infections and tumors. Consequently, the causes can vary from straightforward single-factor agents like the herpes simplex virus to more intricate, multifactorial dysbiosis mediated by bacteria-host immune system interactions (1).

1.1.2 Prevalence

Periodontitis, the sixth most common type of inflammatory disease of the human body, affects the supportive tissues of the teeth and may cause extensive tooth loss if left untreated (2,3). Worldwide, it is the most common oral disease after dental caries (https://www.cda-adc.ca/stateoforalhealth/global/). Several meta-analyses have indicated that a significant proportion of adults are affected by mild to moderate periodontal disease, with the severest form impacting approximately 5–20% of any given population (4-6). Mild periodontal disease is characterized by its early onset, while severe cases typically arise from a chronic state of the condition (7).

It has been reported that approximately 6% of the general adult population in Canada suffers from severe periodontal disease (8). Between March 2007 and February 2009, Statistics Canada conducted the Canadian Health Measures Survey (CHMS)(https://www.canada.ca/en/health-canada/services/healthy-living/reports-publications/oral-health/canadian-health-measures-survey.html), gathering data from approximately 6,000 individuals residing in 15 randomly chosen communities across

Canada. This sample reflects 97% of the Canadian population aged six to seventy-nine years. The results showed that 21% of adults with teeth have, or have had, a moderate or a severe periodontal (gum) problem. In sum, periodontal issues are pervasive in the population and require accurate identification and treatment.

1.1.3 Systemic Disease Connection

Over the past few decades, multiple studies have investigated the connections between periodontitis and systemic diseases as potential risk factors (9). While certain diseases such as diabetes (10), obesity (11), and cardiovascular conditions (12) have established associations with periodontitis, many of these links are not particularly strong. It's crucial to recognize the importance of managing periodontal disease independently due to the significant impact on oral health-related quality of life (OHRQoL), which is mainly associated with tooth loss (13).

A recent review by Fischer et al. (14) reported that the treatment of periodontitis could significantly improve OHRQoL and contribute to lowering systemic inflammation markers, including certain cytokines linked to cardiovascular diseases. Some included studies also indicated that periodontal treatment might lead to better glycemic control among patients with diabetes mellitus (15-17).

1.1.4 Etiopathogenesis

Dental plaque and calculus are recognized initiators of gingival inflammation, known as gingivitis, which can potentially progress to the more severe condition of periodontitis (18-20). Notably, while the presence of these local factors is consistent, individual variations in the host response play a significant role in determining disease progression (21). Central to understanding periodontal disease development is the concept

of dysbiosis, where an imbalance between host and bacterial factors disrupts the oral microbiome, leading to disease (22). Among the pathogens implicated in this process, Porphyromonas gingivalis stands out as a keystone species capable of inducing dysbiosis even at low levels (22).

Furthermore, herpes viruses, particularly herpes simplex virus type 1 and human cytomegalovirus, have been associated with aggressive forms of periodontitis (23,24). However, the role of other herpes viruses, such as Epstein-Barr virus and herpes simplex virus type 2, remains less clear, with ongoing research attempting to elucidate their significance in periodontal disease (24). Besides, genetic and epigenetic influences, along with systemic conditions like diabetes and lifestyle factors such as smoking, have been identified as significant contributors to both the onset and progression of periodontal disease (25-28). Thus, understanding these multifaceted interactions is essential for developing effective strategies for prevention and management.

1.1.5 Risk Factors

1.1.5.1 Smoking

Smoking is identified as the primary environmental risk factor for periodontitis. Studies have demonstrated that compared to non-smokers or former smokers, smokers exhibit a significantly higher prevalence of red-complex periodontal pathogens in their subgingival biofilm (29-31). Moreover, smoking has been linked to a potential negative impact on host immune cells, particularly neutrophils, rendering the host more susceptible to periodontitis (32-34).

Correspondingly, both light and heavy smokers face an increased risk of developing alveolar BL, with odds ratios of 3.25 and 7.28 respectively, compared to non-smokers. Similarly, light and heavy smokers have higher odds ratios of 2.05 and 4.07 respectively, for developing periodontal attachment loss compared to non-smokers (35). Additionally, smoking adversely affects the outcomes of both active periodontal therapy and long-term maintenance therapy (28,36). Hence, patients should be consistently advised on the significance of smoking cessation for the effective management of periodontitis (37).

1.1.5.2 Diabetes

Patients with uncontrolled diabetes face an elevated risk of developing periodontitis compared to systemically healthy individuals or those with well-controlled diabetes (38,39). The prevalence of periodontitis is notably higher among Canadians with diabetes, with around 18% experiencing moderate to severe periodontal disease, highlighting the importance of periodontal health in diabetic individuals (40).

Plausible biological mechanisms supporting this association have been scientifically validated (39). This connection is partly attributed to alterations in the immune system of individuals with uncontrolled diabetes, leading to impaired neutrophil function or hyper-responsive macrophages producing pro-inflammatory cytokines (39). Furthermore, patients with uncontrolled diabetes demonstrate changes in connective tissue metabolism, influencing the resorptive and formative processes in the periodontium (39).

1.1.6 Diagnosis

Before conducting a periodontal assessment, it's essential to gather a patient's medical history. This step enables the identification of systemic or environmental risk

factors for periodontitis, such as diabetes and smoking (41). A thorough periodontal evaluation comprises various clinical parameters, including the biofilm index, periodontal probing depth, presence of bleeding on probing, gingival recession, mucogingival deformity, furcation involvement, tooth mobility, and occlusal trauma. Additionally, an initial periodontal evaluation should include a comprehensive radiographic assessment to ascertain the extent of horizontal and vertical alveolar BL (42).

The 2017 World Workshop on the classification of Periodontal and Peri-implant Diseases and Conditions (2) defined the current method for classifying periodontitis based on staging and grading. Staging is determined through the severity of the disease and complexity of its management, while grade is an indicator of the rate of periodontitis progression assessed according to the history and the presence of risk factors for the disease (42).

1.1.6.1 Periodontitis Staging

Staging relies on the standard dimensions of severity and extent of periodontitis at presentation but introduces the dimension of complexity of managing the individual patient (Figure 1.1-A). Clinical attachment loss (CAL) and radiographic BL are the primary stage determinants. The goals of staging are to: 1) Classify severity and extent of an individual based on the currently measurable extent of destroyed and damaged tissue attributable to periodontitis, and 2) Assess complexity and specific factors that may determine complexity of controlling current disease and managing long-term function and esthetics of the patient's dentition. Periodontitis staging is categorized into four including stage I (a condition between gingivitis and periodontitis), stage II (established periodontitis in which

a carefully performed clinical periodontal examination identifies the characteristic damages that periodontitis has caused to tooth support), stage III (in which periodontitis has produced significant damage to the attachment apparatus and, in the absence of advanced treatment, tooth loss may occur), and stage IV (with considerable damage to the periodontal support that cause significant tooth loss) (2).

	Periodontitis	Stage I	Stage II	Stage III	Stage IV
Severity	Interdental CAL *(at site of greatest loss)*	1 – 2 mm	3 – 4 mm	≥5 mm	≥5 mm
	RBL	Coronal third (<15%)	Coronal third (15% - 33%)	Extending to middle third of root and beyond	Extending to middle third of root and beyond
	Tooth loss *(due to periodontitis)*	No tooth loss		≤4 teeth	≥5 teeth
Complexity	Local	• Max. probing depth ≤4 mm • Mostly horizontal bone loss	• Max. probing depth ≤5 mm • Mostly horizontal bone loss	In addition to Stage II complexity: • Probing depths ≥6 mm • Vertical bone loss ≥3 mm • Furcation involvement Class II or III • Moderate ridge defects	In addition to Stage III complexity: • Need for complex rehabilitation due to: – Masticatory dysfunction – Secondary occlusal trauma (tooth mobility degree ≥2) – Severe ridge defects – Bite collapse, drifting, flaring – < 20 remaining teeth (10 opposing pairs)
Extent and distribution	Add to stage as descriptor	For each stage, describe extent as: • Localized (<30% of teeth involved); • Generalized; or • Molar/incisor pattern			

Figure 1.1- A. Periodontitis staging: to classify the severity and extent of a patient's disease based on the measurable amount of destroyed and/or damaged tissue as a result of periodontitis

1.1.6.2 Periodontitis Grading

Irrespective of the stage at diagnosis, periodontitis may progress at different rates with individuals, may respond less predictably to treatment in some patients, and may or may not influence general health or systemic disease (43). This information is critical for precision medicine but has been an elusive objective to achieve in clinical practice. In recent years, validated risk assessment tools, and presence of individually validated risk

factors have been associated with tooth loss (44), indicating that it is possible to estimate risk of periodontitis progression and tooth loss. Grading adds another dimension and allows the rate of progression to be considered. Figure 1.1-B illustrates periodontitis grading based on primary criteria represented by the availability of direct or indirect evidence of periodontitis progression. Direct evidence is based on longitudinal observation available for example in the form of older diagnostic quality radiographs. Indirect evidence is based on the assessment of BL at the worst affected tooth in the dentition as a function of age (measured as radiographic BL in percentage of root length divided by the age of the subject). Periodontitis grade can then be modified by the presence of risk factors. Grade A, B, or C periodontitis is determined with direct or indirect evidence of progression rate in three categories: slow, moderate, and rapid progression. Also, risk factor analysis is used as a grade modifier (2).

	Progression		**Grade A: Slow rate**	**Grade B: Moderate rate**	**Grade C: Rapid rate**
Primary criteria *Whenever available, direct evidence should be used.*	Direct evidence of progression	Radiographic bone loss or CAL	No loss over 5 years	<2 mm over 5 years	≥2 mm over 5 years
	Indirect evidence of progression	% bone loss / age	<0.25	0.25 to 1.0	>1.0
		Case phenotype	Heavy biofilm deposits with low levels of destruction	Destruction commensurate with biofilm deposits	Destruction exceeds expectations given biofilm deposits; specific clinical patterns suggestive of periods of rapid progression and/or early onset disease
Grade modifiers	Risk factors	Smoking	Non-smoker	<10 cigarettes/day	≥10 cigarettes/day
		Diabetes	Normoglycemic/no diagnosis of diabetes	HbA1c <7.0% in patients with diabetes	HbA1c ≥7.0% in patients with diabetes

Figure 1.1- B. Periodontitis grading: the rate of periodontitis progression, responsiveness to standard therapy, and potential impact on systemic health

1.1.6.3 Assessment of Periodontitis Stage and Grade

Periodontitis diagnosis can be established from a) Clinical examination by gently probing the gingiva using a standardized periodontal probe with reference to the cemento-enamel junction (CEJ), or b) determine the extent of alveolar BL around the teeth during the radiographic examination (45). Alveolar BL, is defined as any distance from CEJ to the alveolar bone crest that is greater than 2mm (46).

In recent decades, electronic dental records (EDR) have been introduced for collection of patient data, which are found to have superiority over paper charts in terms of higher storage capacity, time efficiency, ease of information retrieval and accessibility (47). However, manual evaluation of a huge amount of information collected from patients is prone to errors particularly by low-experienced clinicians as while there are drastic differences between initial/moderate stages of periodontitis (stage I/II) and severe/advanced forms (stage III/IV), there are often gray areas and overlapping criteria when clinicians are trying to narrow down a case to either stage III or IV category. Thus, the use of sound clinical judgment is crucial when classifying the disease according to its stage and grade (48).

1.1.7 Prevention and Management

1.1.7.1 Home Care

Ensuring satisfactory home care is a crucial aspect of preventing periodontal disease, facilitating successful periodontal therapy, and maintaining the dentition over the long term (49-51). It is imperative for clinicians to educate patients about the significance of efficiently eliminating dental biofilm at home, particularly before embarking on active

periodontal therapy (52). The importance of maintaining adequate home care should be emphasized consistently during both the initial and subsequent stages of periodontal treatment.

1.1.7.2 Scaling and Root Planing (SRP)

Once adequate home care or biofilm control has been achieved, scaling and root planing should be conducted at sites exhibiting periodontal probing depths of 5 mm or greater (41). This treatment phase should be accompanied by addressing local contributing factors, extracting hopeless teeth, and treating active carious lesions (53). Automated instruments, such as piezoelectric or ultrasonic scalers, may be utilized alongside manual instruments (53). In areas where access is challenging, automated instruments might be more effective than curettes for eliminating subgingival biofilm and calculus (54). Clinically, the removal of subgingival calculus can be assessed using a periodontal explorer. Additionally, post-operative intraoral radiographs can aid in evaluating the removal of subgingival calculus, which was visible on pre-operative intraoral radiographs (41).

1.1.7.3 Systemic Antibiotics

Systemic antibiotics have demonstrated effectiveness as adjuncts to SRP in managing chronic periodontitis. The short-term use of a combination of amoxicillin and metronidazole, particularly in patients with periodontal pocket depths of 6 mm or more, has shown clinical significance (55). This combination has also proven useful in treating aggressive periodontitis (56). Azithromycin has shown improved clinical outcomes when

used alongside SRP (57,58), and it has been particularly effective in managing cyclosporine-induced gingival overgrowth (59).

However, a recent systematic review conducted by Khattri et al. (60) concluded that there is limited certainty regarding the long-term effectiveness of adjunctive systemic antimicrobials in non-surgical treatment of periodontitis. Also, insufficient evidence exists to determine the superiority of specific antibiotics when used with SRP.

1.1.7.4 Periodontal Re-evaluation

Between 4-6 weeks following the completion of SRP, a re-evaluation should be undertaken (50). During this phase, a comprehensive periodontal charting should be updated, and the findings should be compared with the initial charting to assess the extent of improvement. Additionally, patient compliance, as determined by adherence to the prescribed home care regimen, should be meticulously assessed (50). Typically, for areas with relatively shallow probing depths (i.e., 1–5 mm), non-surgical management strategies could be considered, including repeated root planing if necessary, frequent periodontal maintenance therapy, and continuous reinforcement of home care (50). It is noted that the efficacy of subgingival calculus removal diminishes with increasing probing depth (61,62). Consequently, for areas exhibiting persistently deep periodontal probing depths (i.e., 6 mm or deeper), surgical periodontal therapy may be warranted. However, surgical therapy should be postponed until the patient demonstrates adequate biofilm removal (50).

1.1.7.5 Periodontal Surgical Therapy

If initial nonsurgical treatments like SRP along with other adjuncts prove ineffective in resolving periodontitis upon reassessment, periodontal surgical therapy becomes necessary (63). Osseous surgery has been found to enhance CAL gain compared to SRP in deeper periodontal pockets measuring 7 mm or more (64). Additionally, regenerative periodontal therapy can be advantageous in specific cases, such as intrabony periodontal defects with well-contained osseous walls (65).

1.1.7.6 Personalized Maintenance Protocols

Patients with a history of periodontal disease require regular and recurrent periodontal maintenance, typically scheduled at intervals ranging from 2 to 6 months (66,67). However, the appropriate frequency should be determined after completing active periodontal therapy and should be adjusted based on continuously evaluating an individual's risk for periodontitis (67). A consistent recall schedule enables prompt detection and intervention if disease recurrence or reactivation occurs in patients previously treated for periodontitis (66). During maintenance therapy, periodontal charting should be updated, and radiographs should be obtained as necessary. Additionally, home care practices should be thoroughly reviewed.

Various factors, including dysbiosis, genetic predispositions, environmental influences, and behavioral habits, contribute to the diverse onset, progression, and response to therapy in periodontal disease. The complexity of these factors makes it challenging to predict the prognosis of periodontally compromised teeth. Therefore, it is essential to establish personalized maintenance protocols based on individual risk factors (68).

1.2 Artificial Intelligence

Artificial intelligence (AI) is a rapidly emerging field in computer science, which refers to the application of machines capable of performing tasks normally requiring human intelligence. ML, a subset of AI, uses algorithms to predict unseen data based on the learnings obtained from intrinsic statistical patterns and structures in data (69-72). Neural networks (NN) are a popular type of ML structure, which have been proven to outperform predictive algorithms, particularly on complex data structures including images and texts. An artificial neuron –a mathematical non-linear model inspired by the human neuron- constitutes any type of NN. Stacking and binding these neurons in layers through mathematical functions is the structure of NN. Through this process, a network is developed for a defined specific task like image or text analysis (70,73).

Deep learning (DL) and the term deep neural network (DNN) refers to NN architectures with more than one hidden layer that are capable of analyzing complex data structures such as images due to their ability to represent an image and its related features including edges, corners and shapes (70,74). The main aim of DNN is to obtain features from the input dataset through supervised, unsupervised or semi-supervised training approaches and transform the data in multiple layers. The output from each layer acts as the input for the subsequent layer and this iteration goes on until the final layer, which is used to fine-tune and train the whole network (75,76). Strengthening the ML algorithm through self-improvement and reinforced learning by evaluating more and more data, is one of the main advantages of DNN (69,74).

1.2.1 Natural Language Processing (NLP)

Natural language processing (NLP) represents a technological advancement facilitating interaction between computers and humans using natural languages (77). NLP encompasses a wide range of methods enabling computers to engage with human text. This engagement spans various activities such as treating human text as data, understanding the significance and associations among words in the text, and employing this acquired knowledge to execute diverse text-oriented operations (78). At its core, NLP involves segmenting text corpora for analysis, utilizing tools like ontology dictionaries, word frequency analysis, and contextual semantic analysis to dissect the smallest units of meaning (79).

The essence of NLP lies in enabling seamless communication between humans and computers, where computers can not only comprehend the meaning conveyed in textual language but also articulate intentions and ideas similarly (80). This dual capability is delineated into 'natural language understanding' or ‘linguistics’ and 'natural language generation,' which evolves the task to understand and generate the text (79).

1.2.1.1 NLP Approaches

1.2.1.1.1 Rule-Based Methods

In the rule-based approach to NLP, extensive sets of sentences are often abstracted into rules tailored for specific human-computer interactions, incorporating essential information markers (81). These rules are then used to generate finite state automata, which convert linguistic input into a parameter sequence. This sequence guides subsequent

information processing methods, thereby enhancing the efficiency of natural language understanding and emphasizing the scalability of the rule set (82,83).

Rule-based NLP models offer several advantages, such as transparency and interpretability, as rules are explicit and easily understandable by humans, facilitating quality control and modifications. They also require less training data compared to data-driven models, saving time, resources, and computational power. Additionally, they can be robust and adaptable to new scenarios as long as the rules cover them, enhancing reliability across various contexts (84,85).

However, these models also have drawbacks. Scaling and maintenance pose challenges due to complex, numerous, and overlapping rules, making them difficult to write, test, and debug. Rules may become outdated or irrelevant with language evolution or domain changes. Furthermore, they struggle with handling natural language's ambiguity, variability, and creativity, leading to errors and exceptions. Moreover, they can exhibit bias or inaccuracy, reflecting the assumptions and judgments of rule-makers (86,87).

Regular expression (RegEx) is an example of a powerful tool for pattern matching and text analysis. By defining a sequence of characters in a string using literal text or special characters with specific meaning, the tool is able to search, match, and manipulate text data (88).

1.2.1.1.2 Machine Learning-Based Methods

ML-based approaches rely on annotated data to build systems, utilizing manually defined features. These systems employ learning techniques to determine parameters used during runtime for data processing and output generation. Notable successes have been

observed in applications like machine translation and search engines using statistical methods (77,89).

NLP tasks encompass various subtasks, traditionally addressed using methodologies like support vector machines (SVM) (90). ML-based NLP techniques can be expressed as a model that is then applied to other texts, also known as supervised ML models. It also could be a set of algorithms that work across large sets of data to extract meaning, which is known as unsupervised ML (91,92).

Despite the rapid advancement of DL algorithms in the modern era, traditional ML algorithms retain their distinct advantages. Traditional methods typically require fewer data points and offer greater interpretability. Moreover, numerous researchers are integrating traditional ML models with DL networks to enhance the interpretability and resilience of neural networks (93). However, practical applications have limitations, including high dependency on the quality of training data, the necessity for data labeling, domain-specific variations, and limitations in capturing nuanced features beyond predefined rules (94,95).

1.2.1.1.3 Deep Learning-Based Methods

Deep learning models are increasingly employed in NLP tasks, utilizing architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) (96). When applying a fully connected network to NLP tasks, several challenges arise. Firstly, the variable input length of different samples makes it impossible to fix the number of neurons in the input and output layers. Secondly, the inability to share features learned from different positions in the input text leads to inefficiencies. Lastly, the

model complexity tends to increase significantly, resulting in a large number of parameters and requiring extensive computations (96,97).

To address the above challenges, RNNs are employed. RNNs scan input data, facilitating parameter sharing across all time steps. They not only receive input from the current moment but also incorporate information from previous steps, allowing past information to influence current decisions effectively. However, traditional RNNs have limitations. They often pass all learned knowledge to the next time step without processing, potentially overwriting early knowledge with more recent information and struggling to capture long-range dependencies (96,98). Long Short-Term Memory (LSTM) models address this by introducing a gating mechanism to overcome the vanishing gradient problem during training with long sequences (99).

DL, by learning word embeddings, enables natural language classification and understanding. Compared to traditional ML, DL-based NLP offers several advantages (93,100,101). Firstly, it facilitates continuous learning by continuously acquiring language features based on word or sentence vectorization, grasping higher-level and more abstract language features to accommodate various NLP tasks. Secondly, it automates feature learning, eliminating the need for manual definition of training sets by automatically acquiring high-level features through neural networks.

1.2.1.1.4 Transformer-Based Models

The significant advancement in NLP models was sparked by the publication of the paper "Attention is All You Need" in 2017 (102). This paper introduced the Transformer model for machine translation, pioneered by the Google Brain team. Transformer, as the fourth major type of DL model following multi-layer perceptron (MLP), CNN, and RNN,

is distinguished by its self-attention mechanism. It broke new ground by abandoning CNN and RNN architectures in the sequence-to-sequence domain, relying solely on a straightforward network structure with attention mechanisms. This led to a considerable acceleration in the training process for sequence tasks and enabled the development of even larger models (103).

The introduction of the BERT model marked the application of pre-training techniques in NLP (104). BERT and subsequent models like GPT-3, RoBERTa, among others, have achieved notable success across various NLP tasks, and revolutionized large language models (LLM), surpassing the previously prevalent RNN-based methods. They demonstrate superior performance compared to sequential DL architectures such as LSTMs and CNNs across a wide range of natural language tasks, as well as in various non-linguistic applications like image recognition (105). But what makes transformer-based models better for NLP tasks?

1.2.1.1.4.1. Self-Attention Mechanism

The self-attention mechanism is an attention mechanism utilized for modelling sequential data. It operates by computing attention weights, which gauge the similarity between various positions within the input sequence. When provided with a position in the input sequence, the self-attention mechanism calculates similarity scores between this position and others, normalizes these scores through the Softmax function, and derives attention weights associated with the other positions. Subsequently, these weights are employed on the corresponding Value vectors, typically feature vectors from the input sequence, to generate the ultimate attention output (102,106).

For instance, consider the scenario where a wolf refuses to eat a rabbit due to its slimness. Identifying whether the pronoun "it" refers to the wolf or the rabbit is straightforward for humans but presents a challenge for machines. The self-attention mechanism enables the machine to associate "it" with the rabbit.

1.2.1.1.4.2 Multi-Head Attention

Multi-head attention refers to the fusion of various distinct self-attentions. In the transformer model, 8 self-attentions are employed for amalgamation. This approach augments the richness of each attention layer without altering the parameter count. Consequently, parallel computations are facilitated, leading to increased computational efficiency (102). Multi-head attention enables models to acquire varied representations and features, thereby bolstering their expressive capacities and demonstrating superior performance in handling complicated tasks and lengthy sequences (107).

1.2.1.1.4.3 Positional Encoding

Positional Embedding plays a crucial role in the transformer model. While self-attention effectively captures word dependency, it lacks the ability to extract absolute or relative word positions. Even if the key and value order are altered, the attention result remains constant. Given the significance of word order in NLP tasks, the transformer model integrates positional embedding to retain word information. Each position in the sequence receives a unique numerical identifier, corresponding to a specific vector. These vectors are then added to word embeddings, enriching each word representation with distinct positional information (102,104).

1.2.1.1.4.4 Advantages of Transformer-Based Models

According to the above characteristics, transformer-based models offer several advantages. In RNNs, the hidden state at a specific time step relies on the output of the previous time step's hidden state, hindering parallel processing. Conversely, the Transformer model concurrently processes information from all contextual positions, thereby minimizing the loss of information propagation (103).

Moreover, the transformer model effectively mitigates information propagation loss. RNN models, including advanced variants like LSTM, may encounter challenges such as gradient explosion, vanishing gradients, or difficulty in capturing long-range dependencies when handling extensive sequences. In contrast, the Transformer model efficiently accesses information from all positions within a sequence, effectively addressing these issues by maintaining a constant distance of one (105).

Lastly, the transformer model adeptly integrates information from all positions. Unlike convolutional models, which typically consider smaller windows and necessitate multiple convolution layers to incorporate information from distant elements, the transformer model seamlessly gathers information from all positions within a distance of one, thereby enhancing information integration across the entire sequence (79).

1.2.2 Automated Radiographic Image Analysis

Despite the improvements in the dental imaging in terms of the quality of image and resolution over the past decade, interpreting them is primarily and subjectively conducted by the trained dentist based on the individual's judgement and experience (108-110). This subjectivity can lead to low consistency among different raters in measuring

BL, as demonstrated in a study conducted by Lanning et al. (111) where twenty-four clinical instructors rated BL for twenty teeth using a view box. The instructor group consisted of eight dental hygiene faculty members, four graduate students, and twelve periodontal faculty members. All of the dental hygiene faculty and most of the periodontal faculty had ten or more years of clinical experience, whereas all but one of the graduate students had less than five years of clinical experience. The study found varying levels of accuracy among the instructors. For teeth with no BL, 74.2 percent (89 of 120) of the clinical instructors' ratings were accurate. For categories of <15 percent, 15–30 percent, and >30 percent BL, the accuracy of ratings was 50 percent, 52.1 percent, and 94.4 percent, respectively. Overall, the clinical instructors' agreement with the correct choice was 70.0 percent, and when corrected for chance agreement, this agreement was Kappa= 0.591 (SE=2.8 percent). These results highlight the variability in measurements based on the experience and subjective judgment of the raters.

This may result in misdiagnosis and, in the process of periodontitis evaluation, may lead to wrong measurement of BL (112). Currently, DL techniques are being widely applied for quick evaluation of dental radiographs, without subjective interpretations. The intelligent systems can perform these tasks through the process of medical image understanding (71). The process consists of 4 stages:

1- Medical image detection with the purpose of classifying and localizing regions of interest by drawing bounding boxes around multiple regions of interest and labeling them. This helps in determining the exact locations of different organs and their orientation.

2- Medical image localization, which predicts the object in an image, draws a bounding box around the object and labels the object.

3- Medical image segmentation that helps in image understanding, feature extraction and recognition, and quantitative assessment of lesions or other abnormalities.

4- Medical image classification, which involves determining and assigning labels to medical images from a fixed set.

Traditional methods of automated medical image analysis include large amounts of rule-based algorithms or manual preprocessing methods that are time-consuming with low quality and poor generalization capability (113-115). This underwent changes with the introduction of CNN algorithms using DL, which allow for direct interference, recognition and classification of medical images (116,117). The term "convolution" refers to the process of computing as well as the result function (118). CNNs are types of DL methods consisting of a minimum of three layers: input, hidden and output layers (119). They apply a supervised learning technique called "backpropogation" and have been utilized for various image analysis tasks such as classification, segmentation and landmark detection (118,120,121).

CNN has been utilized in various aspects of science including speech recognition, detecting objects, analyzing emotions and face recognition (122). However, its great breakthroughs in major image competitions have made it a popular technique for medical image analysis and computer visual tasks (123,124). In the field of dentistry, CNNs have performed tasks such as caries, BL and apical lesions detection as well as classifying, segmenting and detecting anatomic hard- and soft-tissue landmarks (125).

The process of classification in a CNN model involve (71,126): 1) Convolutional layers (the first step) with the purpose of extracting features such as gradients or edges

from the input image using the mathematical transformations, 2) Non-linear activation functions, which is placed between any two layers and guides the input signals into output signals required for the NN to act, 3) Pooling layer, which reduces the number of parameters to learn and the amount of computation to summarize the features generated by the convolution layer, 4) Fully connected layers that are responsible for the interpretation of the feature representations learned by preceding layers.

Due to its demonstrated efficiency in the field of image understanding (image segmentation, classification, localization and detection) through feature extraction of an input data, it is a widely used technique for solving medical image understanding (71). One main advantage of CNN over traditional NN models is in its architecture (127). Depending on the task to be performed, various architectures of CNN models have been proposed so far. For instance, Long et al. (121) proposed a full convolutional network (FCN) in which the last fully connected layers are replaced by fully convolutional layers to obtain pixel-wise predictions. Since then, many novel methods of image segmentation (such as U-Net) based on FCN have been utilized for medical image analysis (128-131).

1.2.2.1 Object Detection and Localization

The DL object detection and localization technique refers to determining where the object locations are within images (132). DL-based object detection using CNNs offers several advantages over traditional methods such as: 1. Hierarchical feature representation, which enables them to represent objects in a more detailed and meaningful way. 2. Increased learning and expressing capability, leading to more accurate and sophisticated object detection (133). There are two main structures for object detection models: 1)

Region proposal-based techniques that first generate region proposals in the initial stage and then classify each proposal into different object classes. Examples of such techniques are R-CNN, Fast R-CNN, and Faster R-CNN (134,135), and 2) Regression-based algorithms, which predict object classes and locations directly such as you only look once (YOLO), single shot multibox detector (SSD), and RetinaNet (136-138).

1.2.2.1.1 You Only Look Once (YOLO) Series

YOLO is an algorithm employing CNN principles, designed to precisely identify objects in real-time (139). Real-time object detection is a necessary component in computer vision systems for tasks like multi-object tracking. It is currently one of the state-of-the-art real-time object detectors (136,140). The requirements for being a state-of-the-art real-time object detector are as follows: (1) a more robust network architecture (consistent reliability and well-peroformance); (2) a more effective feature integration method (141-143); (3) a more accurate detection method (144-145); (4) a more robust loss function (146,147); (5) a more efficient label assignment method; and (6) a more efficient training method (148,149).

YOLO transforms the detection problem into a regression problem. This transformation is fast since a single regression problem does not need a complex pipeline (150). Additionally, YOLO is faster to process than the R-CNN family because it looks through the entire image, unlike the R-CNN family. Furthermore, YOLO has fewer background errors compared to the R-CNN family (151). This is because the R-CNN family cannot process the surrounding information of the object. Moreover, one of the

characteristics of YOLO is to extract features from the entire image and predict bounding boxes (152). For each object that presents grid cells on the image, it divides the image into $S \times S$ grid, and for each grid cell, it predicts the bounding box's location and class probabilities.

In the literature, older versions of YOLO have been widely used in the field of dental radiology to detect mandibular fractures in panoramic radiographs (153), primary and permanent tooth detection on pediatric dental radiographs (154), detection of cyst and tumors of the jaw (155), and detection of impacted mandibular third molar teeth (133). These deep learning techniques focus on how to align the model's predictions as closely as possible with the ground truth. Additionally, they aim to devise architectures conducive to gathering sufficient information for accurate predictions. However, current approaches overlook the substantial information loss incurred during layer-by-layer feature extraction and spatial transformation of input data.

The newly introduced YOLOv9 by Chien-Yao Wang et al. on February 21st, 2024, marked a fresh installment in the YOLO series. This latest model takes a deeper look into data loss in deep network transmission, specifically tackling the challenge of the information bottleneck, a concern overlooked in prior iterations of the YOLO series (156). They proposed the use of programmable gradient information (PGI) to solve the information bottleneck problem and the problem that the deep supervision mechanism is not suitable for lightweight neural networks. PGI can provide complete input details for the target task to calculate objective function, so that reliable gradient information can be obtained for updating network weights (156).

In addition, they introduced a novel lightweight network architecture – Generalized Efficient Layer Aggregation Network (GELAN), based on gradient path planning. The results of the proposed GELAN and PGI on MS COCO dataset-based object detection demonstrated that GELAN, which is a highly efficient and lightweight neural network, has strong and stable performance at different computational blocks and depth settings. It can indeed be widely expanded into a model suitable for various inference devices. Moreover, PGI can be used for a variety of models from lightweight to deep to achieve considerable improvements in accuracy (156).

Combining PGI and GELAN, the YOLOv9 has demonstrated remarkable competitiveness. Its ingenious design enables the deep model to reduce parameter count by 49% and calculation load by 43% compared to YOLOv8. Despite this, it manages to achieve a 0.6% increase in Average Precision (AP) on the MS COCO dataset (156).

1.2.2.2 Image Segmentation

Image segmentation -the process of identifying key components of an image and separating the image into individual sections or regions- is a fundamental task in medical image processing (157). In the field of dentistry, different segmentation methods exist:

1.2.2.2.1 Manual Image Segmentation

Manual segmentation involves the operator manually tracing or outlining the region of interest using a tool such as a mouse or stylus (158). Following this, a segmentation model is generated. This model allows for the visualization of the patient's dental anatomy, providing an accurate assessment of the position, size, and shape of individual teeth, arch

alignment, and inter-arch distances. Additionally, it enables the visualization of various tooth movements, assisting clinicians in formulating an appropriate treatment plan. However; the process is time-consuming, and it can be susceptible to fatigue and errors, particularly when the operator must perform repetitive tasks over an extended period. This can result in inaccuracies and inconsistencies in the segmentation outcomes, leading to inter- and intra-operator variability and low reproducibility (159).

1.2.2.2.2 Semi-Automatic Segmentation

Semiautomatic segmentation is a process that combines automatic segmentation with manual checking and editing of segment boundaries. It begins with the manual initialization of the model, followed by automated boundary detection by the software. The operator then manually corrects and refines these boundaries iteratively until the final segmentation result is achieved after multiple repetitions of both automated and manual refinement.

1- Watershed Method: This technique segments objects or regions of interest in an image based on the concept of watersheds. The operator first selects an image slice containing all teeth and enhances the image, as the watershed algorithm relies on image enhancement to separate tooth regions correctly. The enhanced image is then processed by the watershed algorithm to produce the desired segmentation. The success of this method depends on the quality of the input image, preprocessing techniques, and accurate marker placement (160).

2- Seed Pixels (Region Growing): This method involves selecting seed points or regions of interest within the image and growing these regions by adding adjacent pixels that meet specific criteria, such as color or intensity similarity (161). It is commonly used

in medical imaging, for example, in tumor segmentation, where a radiologist selects seed points within a tumor to initiate the segmentation process.

3- Cluster-Based Segmentation: This technique segments an image into regions or clusters based on similarities between pixels or groups of pixels. It can detect small variations in intensity values and allows for the use of various grouping criteria. Common clustering algorithms used in tooth segmentation include K-means, Fuzzy C-means, and hierarchical clustering, with the choice of algorithm and feature extraction techniques varying depending on the dental image dataset and problem complexity (162).

1.2.2.2.3 Automatic Segmentation

Automatic segmentation is a process in which segment boundaries are assigned automatically by a program, without the need for user interaction. This method has the advantage of enabling rapid segmentation once the approach has been developed. Automatic segmentation can be categorized into two main types:

1. Semantic Segmentation: Semantic segmentation involves assigning a specific label or category to each pixel within an image. The primary goal is to identify and differentiate groups of pixels that represent distinct and meaningful object categories. This approach focuses on distinguishing between categories, such as separating the teeth, jaws, and background, without differentiating between individual entities within each category (163, 164).

2. Instance Segmentation: Instance segmentation combines semantic segmentation with object detection by identifying all instances of a particular category within an image. This method not only segments objects but also distinguishes individual instances of the same object class (165). The task of instance segmentation emphasizes the differences

between individuals within a category, requiring both the category label and an instance label (within the class), to distinguish between individuals in each category.

Over the past few decades, the role of ML in medical applications has expanded significantly. In ML, mathematical models are utilized to allow computers to discover underlying patterns in data and apply this learned knowledge to make predictions on new, unseen data. In recent years, DL-based segmentation algorithms have gained popularity among researchers. These methods have demonstrated exceptional performance in image segmentation, offering greater flexibility and more powerful capabilities compared to traditional ML approaches, while also reducing the need for extensive expert analysis (158, 166).

Several DL models have been introduced for medical image segmentation like the U-Net and mask R-CNN (167). Most image segmentation DL models are based on CNNs (168). Among these models, Mask R-CNN stands out as a potent instance segmentation approach, combining semantic segmentation with the delineation of individual objects. It benefits from extensive computing resources, leveraging hundreds of thousands of images in its training. Despite its strengths, Mask R-CNN is a large model, which impacts its speed. Additionally, retraining the model can be time-consuming, and it does not guarantee that detected objects are free of overlapping. On the other hand, U-Net, which is a famous pixel-based image segmentation model, offers a specialized CNN architecture tailored for medical imaging segmentation. It was employed by Ronneberger to analyze dental structure segmentation on bitewing radiographs (128,169), and is praised for its simplicity, ease of implementation, and control. Moreover, it excels in both training and detecting phases, being significantly faster than Mask R-CNN.

1.2.2.2.3.1 U-Net

U-Net architecture has been thoroughly studied for the biomedical image segmentation due to its ability to produce highly accurate segmented images using limited training data (170,171) The popularity of this algorithm is evident from its widespread adoption across major primary imaging techniques, including computed tomography scans, magnetic resonance imaging, X-rays, and microscopy (171)

The traditional U-Net architecture, designed to handle volumetric input, consists of two phases: the encoder section, which learns representational features while incorporating scale- and location-specific information, and the decoder section, which extracts knowledge from the recognized patterns and previously learned features. By employing jump links between corresponding encoder and decoder layers, deep parts of the network can be effectively trained, enabling a comparison of similar characteristics across different regions (128,169). Researchers have been actively investigating various aspects of the U-Net structure, including enhancements to the encoder and decoder, improvements to convolutional layers, and advancements in skip connections (172).

DL-based computer-aided diagnosis has effectively tackled numerous complex issues in dental radiography. However, there is a limited number of studies that use DL algorithms to perform comprehensive evaluation and measurement of bone level across all points on the tooth. More specifically, previous research utilizing deep CNN architecture for identifying and detecting BL had limitations due to the use of limited sample images, resulting in accuracy levels that did not match those of dentists' diagnoses (173-177).

1.2.2.3 Image Classification

Image classification involves separating pixels into distinct groups based on their data values (178). For a pixel to be categorized into a specific class, it must meet certain criteria that define that class (179). These classes can either be predetermined if the user has identified them based on training data, or they remain unknown (180,181). Typically, image classification begins with extracting features from the image and then categorizing these extracted features. Thus, the key-point of image classification lies in how image features are extracted and analyzed (182).

Traditional classification approaches utilize low-level or mid-level features to represent an image. Low-level features are often derived from grayscale density, color, texture, shape, and spatial information, which are defined by humans and are commonly referred to as hand-crafted features. Mid-level features, as well as learning-based features, are frequently distilled using bag-of-visual-word (BoVW) algorithms, which have been effective and popular in image classification or retrieval frameworks in recent years (183).

In a traditional image classification model, after feature extraction, a classifier such as SVM or random forest is typically employed to assign labels to different types of objects (184). In contrast, the DL approach integrates the processes of image feature extraction and classification into a single network. The advanced representation of high-level features in DL has demonstrated superiority over manually crafted low-level and mid-level features, yielding commendable results in image recognition and classification. Image classification methodologies are categorized into two branches: supervised and unsupervised techniques, depending on the prior knowledge of the classes (185).

1.2.3 Multi-Modality Approaches for Integrating Texts and Images

The current state-of-the-art DL models applied in radiology consider only pixel-value information without data collected in patients' charts. However, in practice, accurate and relevant non-imaging data based on the clinical history and examination enable dentists to interpret imaging findings in the appropriate clinical context, which will lead to a higher diagnostic accuracy, informative clinical decision making, and improved patient outcomes (122).

Fusion models are ML techniques that join data from multiple modalities with the aim of extracting complementary and more complete information for better prediction. (186,187). Thus, in the third phase of my study, I used a multimodal model to combine the radiographic data with the information latent in patients' clinical notes to predict and diagnose the stage and grade of the periodontal disease.

Approaches to multimodal data integration in healthcare can vary depending on the specific clinical application and the available modalities (188). Generally, there are three main categories for fusing information from multiple modalities:

1- Early fusion (feature-level fusion): This method involves combining features extracted from different modalities to create a unified representation (189). Features are first extracted from each modality separately, capturing their unique characteristics, and then combined to form a joint feature representation (122). This fusion approach aims to capture complementary information from different modalities and enhance the overall representation (190). Early fusion can be divided into two groups, type I, which fuses the original features and type II, which joins extracted features, either from manual extraction, imaging analysis software or learned representation from another neural network.

Predicted probabilities are considered as extracted features, thus fusing these types of features from different modalities is also early fusion type II) (189,191).

2- Joint fusion: Joint fusion model involves combining feature representations learned from intermediate layers of neural networks with features from other modalities as input to a final model. The key distinction from early fusion lies in how the loss is back-propagated to the feature-extracting neural networks during training, leading to improved feature representations with each iteration of training.

3- Late fusion (decision-level fusion): In this method, decisions or predictions made independently by each modality are combined to reach a final decision. Each modality provides its own decision based on its specific analysis, and these decisions are aggregated using various techniques such as voting, averaging, or weighted combinations (192). Decision-level fusion aims to leverage the diversity of information provided by different modalities and improve the overall decision-making process. This approach allows flexibility in choosing modalities without the need for end-to-end training but does not model interactions between modalities as feature-level fusion does (190).

Ingesting pixel data along with other data types through the application of multimodal DL models has been successful in applications outside of medicine and dentistry, such as autonomous driving and video classification or social media video classification that combines both visual and textual features (193,194). The improvements in performance for these efforts justify the dental applications in addition to leveraging fusion strategies for medical imaging, which is also primarily motivated by the desire to integrate complementary contextual information and overcome the limitation of image-only models (122).

1.2.3.1 Multimodal Transformer Model

The primary inspiration behind AI is to replicate human sensory perception, including seeing, hearing, touching, and smelling. Typically, each modality is linked to a specific sensor that establishes a unique communication channel, such as vision and language (195). In humans, a crucial mechanism in sensory perception is the ability to integrate multiple modalities of sensory data collectively, allowing effective interaction with the world in dynamic and unconstrained environments. Each modality serves as a distinct source of information with unique statistical characteristics. Essentially, a multimodal AI system must ingest, interpret, and reason about multimodal information sources to achieve human-like perception capabilities. Multimodal learning (MML) is a comprehensive approach to developing AI models that can extract and relate information from multimodal data (195,196).

In the era of DL, deep neural networks have significantly advanced the development of MML. The transformer architecture (102), in particular, has emerged as a highly competitive family, presenting new challenges and opportunities for MML. The recent success of large language models and their multimodal derivatives further underscores the potential of transformers in the creation of foundational multimodal models (150, 197-200).

1.2.3.1.1 Transformers

Transformers are emerging as promising learners. The Vanilla transformer (102), which benefits from a self-attention mechanism, represents a breakthrough model for sequence-specific representation learning initially proposed for NLP. It has achieved state-of-the-art performance on various NLP tasks. Following the success of the Vanilla

transformer, numerous derivative models have been developed, including BERT and GPT (104,201).

Transformers currently hold a dominant position in the NLP domain, motivating researchers to explore their application in other modalities, such as the visual domain. Early attempts in the visual domain often follow the pipeline of combining CNN features with a standard transformer encoder. Researchers have achieved BERT-style pretraining by preprocessing raw images, resizing them to a low resolution, and reshaping them into a 1-D sequence 202).

The Vanilla transformer model, which serves as the foundation of the transformer-based research field, features an encoder-decoder structure. It processes tokenized input. Both its encoder and decoder are composed of stacked transformer layers or blocks. Each block consists of two sub-layers: a multi-head self-attention (MHSA) layer and a position-wise fully-connected feed-forward network (FFN). In some transformer literature, the FFN is also referred to as a MLP (102).

1.2.3.1.2 Vision Transformer (ViT)

The Vision Transformer (ViT) (203) is a seminal work that provides an end-to-end solution by applying the transformer encoder to image data. Both ViT and its variants have been widely utilized in various computer vision tasks, including recognition, detection, and segmentation (204-206). The ViT employs an image-specific input pipeline, where the input image is divided into fixed-size patches (e.g., 16 × 16 or 32 × 32 pixels). These patches are then linearly embedded, and position embeddings are added to form patch-wise sequences, which are subsequently encoded by a standard transformer encoder (196,203) (Figure 1.2).

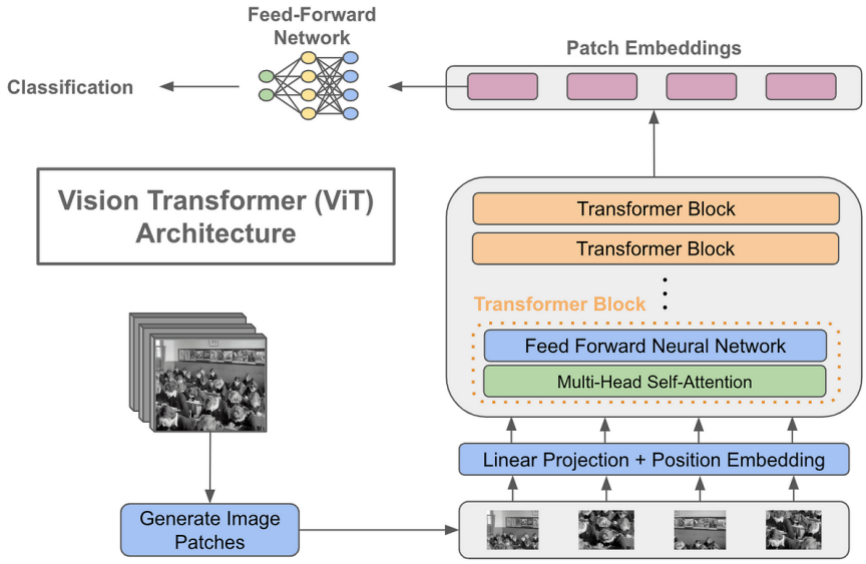


Figure 1.2- Vision transformer (ViT) architecture

1.2.3.1.3 Multimodal Transformers

Recently, a significant number of transformer models have been extensively studied for various multimodal tasks, demonstrating their compatibility with diverse modalities in both discriminative and generative applications (196). The key requirements/designs of the existing multimodal transformer models are as below:

Multimodal input: Given an input from any arbitrary modality, there is a need to perform two primary steps before inputting the data into transformers: (1) tokenize the input and (2) select an embedding space to represent the tokens. In practice, both tokenizing the input and selecting an embedding for the tokens are crucial for the performance of transformers but offer a high degree of flexibility, with numerous alternatives available for implementation.

Self-attention variants in multimodal context: In multimodal transformers, cross-modal interactions are essentially processed by self-attention and its variants. The main

multimodal modelling practices of transformers, from a perspective of self-attention designs, include (1) early summation (token-wise, weighted), (2) early concatenation, (3) hierarchical attention (multi-stream to one-stream), (4) hierarchical attention (one-stream to multi-stream), (5) cross-attention, and (6) cross-attention to concatenation.

1.3 Research Hypothesis and Objectives

1.3.1 Hypothesis

In this dissertation, I propose that similar to healthcare providers, automated detection and classification systems that effectively utilize both dental imaging data and clinical data from EDR — including patient demographics, previous diagnoses, clinical examination findings, and laboratory values — could achieve high diagnostic accuracy in dentistry. This integrated approach holds the potential to serve as a comprehensive CDSS tool, offering enhanced decision-making support for dental professionals in diagnosing and treating patients.

1.3.2 Objective 1

To extract the important latent information in patients' charts, which are collected as text files in clinicians' notes to classify patients with periodontitis through a NLP model using BERT, and utilize the outcome for predicting the stage and grade of periodontitis.

1.3.2.1 Hypothesis 1

Utilizing NLP techniques, particularly BERT, to analyze clinicians' notes text files can effectively extract latent information crucial for accurate classification of patients with periodontitis. Furthermore, leveraging the outcomes derived from this classification can aid in predicting both the stage and grade of periodontitis in patients

1.3.2.2 Rationale

The rationale behind the proposed objective lies in demonstrating the vast potential in the application of NLP, particularly utilizing models like BERT, to unlock valuable insights from unstructured textual data. Clinicians' notes, often comprising rich narratives of patient encounters, serve as a treasure trove of information beyond structured data. These notes contain nuanced details about patient history, symptoms, diagnostic observations, and treatment plans, which can offer valuable context for understanding disease states such as periodontitis.

BERT, being a contemporary NLP model, possesses advanced capabilities in contextual understanding and semantic representation. Its pre-trained embeddings enable the model to comprehend intricate nuances and context within dental narratives, allowing for more precise extraction and representation of latent information embedded within clinicians' notes.

Furthermore, accurate classification of patients with periodontitis is crucial for effective treatment planning and patient management. By employing NLP techniques to analyze clinicians' notes and classify patients with periodontitis, the objective seeks to enhance the accuracy of patient stratification. This improved accuracy not only aids in identifying patients at risk or currently experiencing periodontitis but also serves as a foundation for predicting disease progression.

1.3.3. Objective 2

To analyze patients' radiographs using DL methods (image segmentation and object detection algorithms) to automate the measurement of BL percentage for predicting the stage and grade of periodontitis.

1.3.3.1 Hypothesis 2

Employing DL methods, specifically image segmentation and object detection algorithms, to analyze patients' radiographs can automate the measurement of BL percentage. This automated measurement has the potential to serve as an accurate predictor for determining the stage and grade of periodontitis.

1.3.3.2 Rationale

The rationale behind the proposed objective lies in the transformative capabilities of DL techniques in automating complex image analysis tasks. Radiographs provide crucial visual data for assessing bone level changes, a key indicator in diagnosing and monitoring periodontitis. By applying DL methods such as image segmentation and object detection, the objective aims to automate the measurement of BL percentage, which traditionally requires manual interpretation and measurement by clinicians.

Automation of BL percentage measurement not only streamlines the diagnostic process but also enhances accuracy and consistency, mitigating the potential for human error. Additionally, by accurately quantifying BL percentage, DL-enabled automation can serve as a reliable predictor for determining the stage and grade of periodontitis. This predictive capability will facilitate early detection and intervention while enabling tailored

treatment strategies based on disease severity, ultimately improving patient outcomes in periodontal care.

1.3.4 Objective 3

To integrate the features obtained from image (through ViT) and text (through BERT) using a multimodal model to develop a comprehensive automated approach for predicting the stage and grade of disease.

1.3.4.1 Hypothesis

Integration of features obtained from both image and text processing methods through a multimodal model will lead to the development of a comprehensive automated approach for predicting the stage and grade of disease. This integrated approach is expected to enhance predictive accuracy and provide a more accurate understanding of disease progression.

1.3.4.2 Rationale

The rationale behind the proposed objective stems from the recognition that diseases often manifest through multifaceted symptoms and diagnostic markers, captured in both visual imagery and textual data such as clinical notes. By integrating information obtained from image and text processing methods, the objective seeks to capitalize on the complementary strengths of these modalities to create a more robust predictive model.

The fusion of image and text data through a unified model is anticipated to yield synergistic benefits, enabling a more comprehensive understanding of disease progression. This integrated approach leverages the nuanced insights offered by both visual and textual

information, allowing for a more precise assessment of disease severity and progression. Ultimately, the development of a comprehensive automated approach holds promise for improving diagnostic accuracy, facilitating earlier intervention, and optimizing patient care outcomes.

1.4 Scope of the Thesis

Three hypotheses and objectives are tested in this study as outlined in three chapters (chapters 2-4), with two additional chapters for the introduction (chapter 1) and a final chapter (chapter 5) for general discussion and conclusion.

Chapter 2 tested the first hypothesis and focused on the application of NLP method through BERT analysis to extract the information collected in patients' textual notes for predicting the stage and grade of the periodontitis. The results related to this chapter were recently submitted to the Journal of PLOS Digital Health.

Chapter 3 tested the second hypothesis and focused on the application of different DL models for analyzing PA radiographs to measure the percentage of BL. This measurement was used for predicting the stage and grade of the periodontal disease. The results related to this chapter were recently submitted to the Journals of Clinical and Experimental Dental research and Frontiers in Dental Medicine.

Chapter 4 tested the third hypothesis and integrated the text as well as image data for introducing a comprehensive approach in determining the stage and grade of periodontitis. This study revealed that integrating the text data with their corresponding PA radiographs can improve the classification performance in determining the stage and grade of periodontitis. The results related to this chapter were recently submitted to the Journal of Clinical Periodontology.

Lastly, Chapter 5 discusses the outcome of these studies, their interrelationship, and their significance in the dental setting. Study limitations and future directions are highlighted as well.

Chapter 2- Classification of Periodontitis Stage and Grade Using Bidirectional Encoder Representations from Transformers, A Natural Language Processing Technique

2.1 BACKGROUND

Periodontitis is a multifactorial and microbiome-associated inflammatory disease that occurs in the dental supporting tissues. Progression of the disease can adversely affect oral and systemic health and result in tooth loss, reduction of masticatory performance (207) as well as having association with diabetes (208), and rheumatoid arthritis (209). Thus, periodontitis and its complications will impose substantially negative effects on oral health related quality of life (OHRQoL), while successful and timely diagnosis and management of the disease may improve patients' OHRQoL (207,210). Moreover, early detection and diagnosis of periodontitis can help in preventing the consequent costly and invasive dental treatment (211).

According to the 2017 World Workshop on the classification of Periodontal and Peri-implant Diseases and Conditions (2), the recent method for classifying periodontitis is based on staging and grading. Staging is determined through the severity of the disease and complexity of its management, while grade is an indicator of the rate of periodontitis progression assessed according to the history and the presence of risk factors for the disease. According to the criteria defined for stage and grade of the periodontal disease (2), clinicians traditionally analyze patients' systemic, clinical, and radiographic data collected in periodontal charts over time to determine the stage and grade of periodontitis. In recent decades, electronic dental records (EDR) have been introduced for collection of patient data, which are found to have superiority over paper charts in terms of higher storage

capacity, time efficiency, ease of information retrieval and accessibility (9). In addition, new computerized techniques known as artificial intelligence (AI) and its subsets like machine learning (ML) provide opportunities to extract valuable information from complex data and analyze various relationships to benefit patients (212). AI is an evolving field that seeks to automate tasks that would require intelligence if it was done by humans. ML also has the potential to monitor and detect patterns of patient presentations and risk factors (105, 213).

Recently, neural networks (NN) models- a powerful tool in ML- are gaining attraction and are being extensively used in the area of diagnosis (214), prognosis (215), classification (216), and predictions (217,218) due to their ability of modelling non-linear relationships among the hidden variables in different formats of data including images and texts. Moreover, advancement in text mining and natural language processing (NLP) techniques - also a branch of AI techniques- such as Bidirectional Encoder Representations from Transformers (BERT), significantly improved the accuracy of text classification systems using deep learning models (DL) (219).

NLP methods have been widely used for extracting information from the patient's electronic records in medicine and dentistry (220-223). Chen et al., applied NLP techniques by implementing Sentence2vec and Word2vec approaches to learn sentence vectors and word vectors, and to extract information from Chinese EDR. They reported that their NLP workflow can efficiently structure narrative text from EDR (224). Patel et al. in a recent study developed and applied two automated NLP algorithms (approximate string-matching function and Levestein Distant Function) to extract information from clinical notes to track periodontal disease change over time using longitudinal EDR. They concluded that

utilizing longitudinal EDR data to track disease changes over 15 years was a feasible method that could be applied for studying clinical courses using AI and ML methods (225).

The application of transformers in NLP tasks was accelerated with the advent of a specific transformer-based language model called the BERT (104). BERT is a transformer-based encoder model for language representation that uses a multi-head attention mechanism and a bidirectional approach to learn the contextual relations between words and sentences in a text for an accurate representation of the entire text (226). BERT is a language model which uses transformers for text representation, and supports the pretraining and fine tuning of the model in the case of textual data. Application of BERT and related architectures have resulted in considerable improvements in multiple medical applications, including processing of electronic health records (227), outcome prediction (122) identification of medical terms and concepts (228), and others. Despite its success in the medical field, there are no studies that assess the BERT application in predicting the stage and grade of the periodontal disease using patients' textual notes.

It has been shown that AI techniques such as text mining and DL algorithms can be successfully applied for extracting semantic information from narrative patients' charts in dentistry, classifying the periodontal disease and predicting the factors influencing the periodontitis occurrence (229). Compared to the rule-based approaches using NLP tools, DL models do not depend on the grammatical accuracy of sentences, so can extract implicit aspects following their identification (230). However, the use of AI and DL techniques to review narrative patients' charts and determine the contribution of its information to the stage and grade of periodontitis has not yet been studied to the authors' knowledge. Thus,

in this study, we applied the BERT model to automate the process of classifying patients with periodontitis through unstructured patients' notes.

2.2 MATERIALS AND METHODS

2.2.1 Study Population

To demonstrate the applicability of the proposed method for accurate and timely classification of the stage and grade of periodontitis using the possible contributing factors mentioned in clinicians' notes, we conducted a secondary data analysis on the intake patient textual notes from 2017 to 2021. The University Human Research Ethics Board approved this study (Pro00107743). A total of 1513 patients' charts were extracted from which, 210 textual datasets (3 samples with stage I, 26 samples with stage II, and 181 samples with stage III/IV. For grades, the distribution was 8 samples with grade A, 142 samples with grade B, and 59 samples with grade C), each corresponding to patients with available PA radiographs for subsequent image processing, were reserved for the final model performance analysis to test the generalizability of the model. This selection facilitated the evaluation of the model's performance and enabled a direct comparison between the outcomes derived from text and subsequent image mining methodologies, utilizing an identical set of test samples.

For training the BERT model, three hundred and nine patient clinical notes belonging to 309 patients referred to the University Periodontal Graduate Clinic were selected among the pool of 1513 patients' charts. These clinical notes met the inclusion criteria, which involved having the stage and grade of the disease determined by a graduate student and confirmed by a periodontist. The prevalence of different periodontitis stage

and grades in the collected notes were as follows: stage I= 6, stage II= 36, stage III= 206, stage IV= 61, and grade A= 17, grade B= 202, grade C= 87

2.2.2 Data Pre-processing

The BERT training process for periodontitis staging was completed in two distinct levels. The objective of the first level was to train a model capable of categorizing periodontitis stages into four distinct groups, representative of the nuanced clinical presentation. The second level aimed to categorize these stages into three discrete groups denoted as I, II, and III/IV, aligning with a simplified classification schema intended for integration with the outcomes of image analysis. To achieve this objective, the gathered data pertaining to the stage of periodontal disease underwent re-coding, combining stage III and stage IV into a singular distinct group.

To complete the BERT analysis, we first imported the 309 textual patients' notes coupled with their stage and grade levels (as ground truth labels) into the Google Colaboratory environment (from https://colab.research.google.com/) and preprocessed the texts for computational analysis. In preprocessing, notes were tokenized at both word and sub-word levels to interpolate between word-based and character-based tokenization using the `tensorflow_text` package (from https://www.tensorflow.org/text) in Python. The hardware used for the experiments includes an NVIDIA Tesla T4 GPU, which features 15,360 MiB (15 GB) of memory. The system was running CUDA Version 12.2, supported by NVIDIA Driver Version 535.104.05.

Then, transformer-based special tokens, such as [CLS] and [SEP], were added to make each sentence's beginning and ending understandable to the transformer system. Furthermore, in processing sequence data, it is prevalent for each input sentence to have

different lengths. However, the input data for a DL model must have tensors with a specific shape. Therefore, padding and truncation strategies were used to make the shape of the sentences uniform to the tensor by adding zero values to sentences with fewer tokens (231). Finally, the dataset was split into a training set (70% of the data), a validation set (20% of the data), and a test set (10% of data).

2.2.3 Data Augmentation

The process of data augmentation involved enhancing the existing text data to enrich the training dataset for the BERT model. The aim of this approach was to improve the robustness and generalization capability of the BERT model by exposing it to a wider range of linguistic variations during training, ultimately enhancing its performance in various NLP tasks. Data augmentation was done after splitting the dataset to ensure there will be no data leakage into the test dataset.

To augment the textual patients' information, we utilized a method called synonym replacement, wherein each word in the text was substituted with one of its synonyms, obtained from WordNet, a lexical database of English. This augmentation process was applied selectively to rows associated with specific labels (e.g., "stage I" and "grade A") as there was less data related to these labels, ensuring relevance to the intended target and some balance among the classes. By iteratively replacing words with synonyms, augmented variations of the original text were generated, effectively diversifying the dataset. These augmented texts were then paired with their corresponding labels, forming a new dataset with increased volume and variability. The final result was 1416 texts belonging to 4 stages (stage I= 144, stage II= 360, stage III= 429, and stage IV= 484) and 1447 texts belonging to 3 stages (stage I= 175, stage II= 630, and stage III/IV= 643) for

the 4-categories and 3-categories BERT models, respectively. Also, a total of 1431 texts belonging to 3 grades (grade A= 102, grade B= 581, grade C= 748) were obtained by applying text augmentation methods.

2.2.4 Model Architecture

2.2.4.1 BERT Model for Predicting Periodontitis Stage (4-Categories)

After preprocessing and augmenting the data, the necessary libraries including the "class_weight" module from scikit-learn (232), which aids in addressing class imbalance issues in classification tasks, were imported. The model architecture employed in this chapter combines elements of BERT and LSTM (Long Short-Term Memory) layers to address the classification task. BERT, a pre-trained language model, serves as the foundation for understanding contextual information within textual data (226). Specifically, the "bert-base-uncased" model was utilized, known for its effectiveness in various NLP tasks.

The input sequences were passed through the BERT model, and a recurrent network was used to aggregate the multiple embeddings into a single embedding after the transformer window swipes over the long text. The LSTM model was used as the recurrent layer to aggregate the BERT output at each time step. This model combines the data into a sequence of vectors having the same length relative to their temporal position and temporal dependency with respect to the features in the essays (233). LSTM was selected over other potential operations because the LSTM layers tend to produce more accurate modelling results of deep connections between sequential features that can be used to improve score prediction for text classification tasks (233). By combining BERT's contextual

understanding with LSTM's ability to model sequential data, the model aimed to capture both local and global dependencies within the input data, enhancing its predictive capabilities.

The output of the LSTM layer was subsequently passed through additional dense and dropout layers with rectified linear unit (ReLU) activation functions. These layers enabled the model to learn higher-level features and prevent overfitting by introducing regularization. The final layer consisted of a fully connected neural network layer with softmax activation, producing classification probabilities for each class. For each input text, the selected text stage or text grade was the class with the highest probability in the neural network's output. The number of units in this layer corresponds to the number of unique classes present in the training data (4 stages).

For model compilation, the Adam optimizer was employed with a predefined learning rate of 5e-6. The choice of sparse categorical cross-entropy as the loss function was appropriate for multi-class classification tasks, facilitating the comparison between predicted and actual class labels. During model training, accuracy was used as the evaluation metric to provide insight into the model's performance across different epochs.

To address potential class imbalance issues, class weights were computed and incorporated into the training process. These weights ensure that the model assigns appropriate importance to each class during optimization, thereby mitigating the impact of imbalanced data distributions.

Additionally, early stopping was implemented as a regularization technique to prevent overfitting based on validation loss. Through monitoring the validation loss, the early stopping mechanism terminates training when the model's performance on unseen

data begins to deteriorate. By restoring the best weights observed during training, this approach helps to optimize the model's generalization ability and improve its performance on unseen data.

The model was trained over 40 epochs using the `fit` method, which involves iterating over the training data in batches. The training process included data shuffling, batching, application of class weights, and monitoring of the validation set. By iteratively adjusting the model's parameters based on the optimization objective and evaluation metrics, the training phase aims to optimize the model's performance and enable it to generalize well to unseen data. Finally, the performance of the model was evaluated using evaluation metrics on an unseen new dataset comprising 210 new patients' charts. This assessment provided insights into the model's predictive accuracy and generalization ability when applied to previously unseen patient data.

2.2.4.2 BERT Model for Predicting Periodontitis Stage (3-Categories)

To train the BERT model for categorizing the stage of periodontal disease into 3 groups, the same architecture of combined BERT and LSTM layers was implemented. Like the previous architecture, BERT remains pivotal for contextual understanding within textual data, utilizing the "bert-base-uncased" model pre-trained on large-scale corpora. However, a notable departure lies in the fine-tuning of BERT layers, denoted by setting the “trainable” parameter to “True”. This adjustment enables the model to adapt BERT's representations to the specific nuances of the classification task, enhancing its predictive capabilities.

Similar to the previous model, the output from the BERT layer underwent further processing through dense layers, albeit with the introduction of regularization techniques to mitigate overfitting. In this instance, each dense layer was accompanied by kernel regularization (“tf.keras.regularizers.l2”) with a specified regularization strength (0.01). By penalizing large weights in the model, regularization aids in preventing overfitting and improving the model's generalization ability.

The model was then compiled using the Adam optimizer with a learning rate of 5e-6, while sparse categorical cross-entropy served as the loss function for multi-class classification. The model's performance was evaluated based on accuracy during training. Similar to the previous model, class weights were incorporated into the training process to ensure equal consideration of each class. Furthermore, early stopping was employed as a regularization strategy to prevent overfitting by monitoring the validation loss. The model was trained over 8 epochs and the generalizability of the model was evaluated by testing on the same 210 new unseen patients’ charts (stage I= 3, stage II= 27, and stage III/IV= 180) using the evaluation metrics.

2.2.4.3 BERT Model for Predicting Periodontitis Grade

To train the model for the classification of periodontitis grade, the "bert-base-uncased" variant of a BERT model, which is pre-trained on a large corpus of text data, was utilized. This model was then adapted for the classification task, with the number of output labels set to match the unique labels present in the training dataset (3). The input layers consisted of two tensors: “input_ids” and “attention_masks”, both of shape (311,), which encoded sequences of text data.

The BERT model generated contextualized embeddings for each token in the input sequences. The output from the BERT model was further processed, with only the second output tensor retained. This tensor captures the pooled representation of the input sequence, which is then passed through a dense layer with 64 units and ReLU activation function, followed by a dropout layer with a dropout rate of 0.2 to reduce overfitting.

Subsequently, the output was fed into another dense layer with a softmax activation function, generating probabilities for each class label in the dataset. This architecture formed the basis of a neural network model that learns to classify input sequences into different classes based on their contextual representations.

The model was then compiled using the Adam optimizer with a learning rate of 5e-6 and the sparse categorical cross-entropy loss function. The decision to use a relatively slow learning rate was made to carefully balance the training process, ensuring that the model converges effectively while avoiding potential challenges associated with higher learning rates. With a slow learning rate, the model can fine-tune its parameters more precisely, which is particularly beneficial in complex models, like BERT architectures, where fine-tuning pre-trained weights is crucial.

Accuracy was monitored as a metric during training. Class weights were also computed similarly to the previous models to address any class imbalance present in the training dataset, ensuring that the model learns from all classes equally effectively.

During training, the model was trained over 7 epochs, with a batch size of 20. Early stopping was applied as a regularization technique to prevent overfitting, monitoring the validation loss with a patience of 3 epochs. Finally, the generalizability of the model was

evaluated by testing on the same 210 new unseen patients' charts (8 samples with grade A, 142 samples with grade B, and 59 samples with grade C) using the evaluation metrics.

Figure 2.1 represents a flowchart of the whole process from chart selection to the BERT application for feature extraction and classification.

2.2.5 Evaluation of Model Performance

Evaluation metrics included accuracy, which is the ratio of correctly predicted observations to the total observations; recall, which is the ratio of true positives to the sum of true positives and false negatives; precision, which is the ratio of true positives to the sum of true positives and false positives; and the F1-score, which is the harmonic mean of precision and recall (234). These metrics range from 0 to 1, with 1 indicating perfect performance (235).

2.2.6 Feature Understanding

The Local Interpretable Model-Agnostic Explanations (LIME) method was also utilized to interpret and understand the key words and features that our BERT model identified as most probable for distinguishing between different stages and grades of the disease.

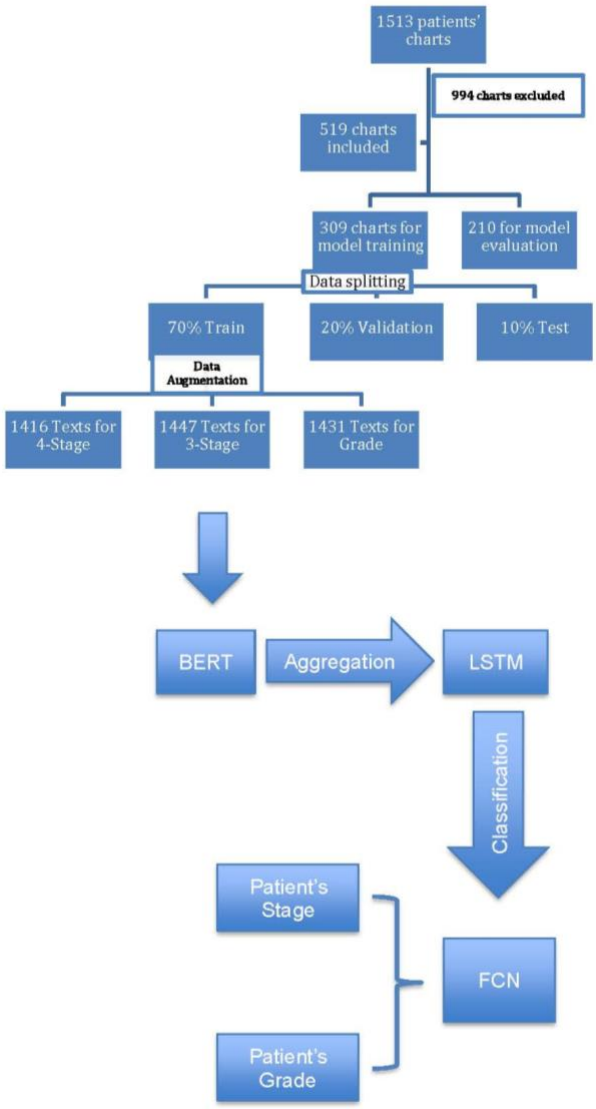
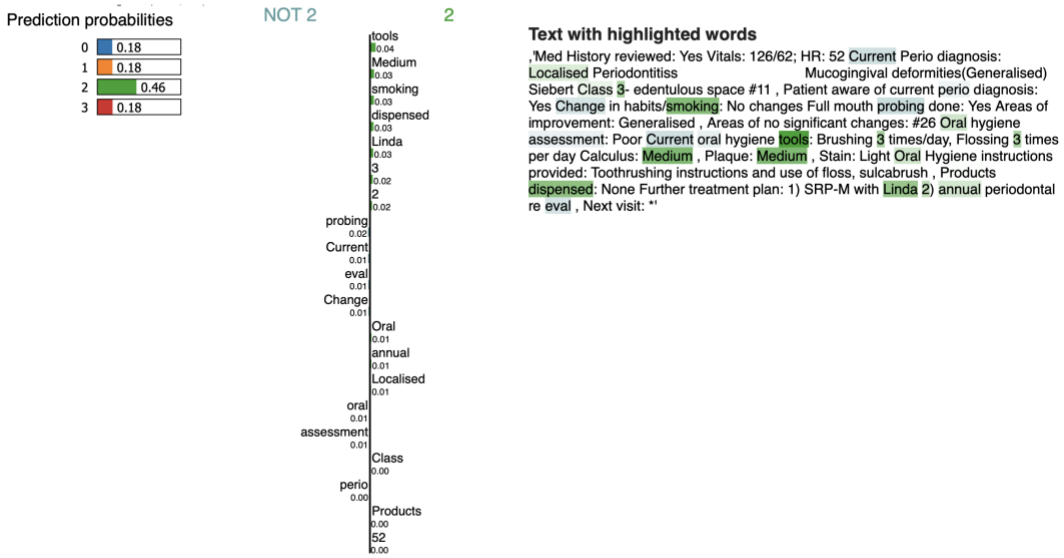


Figure 2.1- Flowchart of the text processing method for classifying stage and grade of periodontitis

2.3 RESULTS

Figure 2.2 (A and B) illustrates the most influential words and features contributing to the model's predictions for stage III and grade B as an example.


Prediction probabilities
0 0.18
1 0.18
2 0.46
3 0.18
NOT 2
2
tools 0.04
Medium 0.03
smoking 0.03
dispensed 0.03
Linda 0.03
3 0.02
2 0.02
probing 0.02
Current 0.01
eval 0.01
Change 0.01
Oral 0.01
annual 0.01
Localised 0.01
oral 0.01
assessment 0.01
Class 0.00
perio 0.00
Products 0.00
52 0.00
Text with highlighted words
,'Med History reviewed: Yes Vitals: 126/62; HR: 52 Current Perio diagnosis: Localised Periodontitiss Mucogingival deformities(Generalised) Siebert Class 3- edentulous space #11 , Patient aware of current perio diagnosis: Yes Change in habits/smoking: No changes Full mouth probing done: Yes Areas of improvement: Generalised , Areas of no significant changes: #26 Oral hygiene assessment: Poor Current oral hygiene tools: Brushing 3 times/day, Flossing 3 times per day Calculus: Medium , Plaque: Medium , Stain: Light Oral Hygiene instructions provided: Toothrushing instructions and use of floss, sulcabrush , Products dispensed: None Further treatment plan: 1) SRP-M with Linda 2) annual periodontal re eval , Next visit: *'

(A)


Prediction probabilities
0 0.23
1 0.52
2 0.25
NOT 1
1
smoking 0.08
re 0.05
or 0.05
texture 0.04
oral 0.04
Good 0.04
significant 0.03
long 0.03
Areas 0.03
provided 0.02
treatment 0.02
perio 0.02
purulent 0.02
Flossing 0.02
46 0.02
N 0.02
per 0.01
probing 0.01
Generalized 0.00
term 0.00
Text with highlighted words
,'Med History reviewed: Yes Vitals: 113/70 HR: 82 Current Perio diagnosis: Generalized periodontal health with localized Periodontitis w.r.t 46 , Patient aware of current perio diagnosis: Yes Change in habits/smoking: N/A Full mouth probing done: Yes Areas of improvement: Overall improvement in gingival texture and appearence. , Areas of no significant changes: 46 Oral hygiene assessment: Good Current oral hygiene tools: Brushing 2x times/day, Flossing 2x times per day Calculus: None , Plaque: None , Stain: Light Oral Hygiene instructions provided: Advised to use water-pik to mainatin the site 46 and continue using interdental brush to maintain the hygiene. , Products dispensed: X-large Interdental brush sample. Further treatment plan: 1. Periodontally tooth #46 has hopeless long term prognosis, as such any resective or regenerative surgery is contraindicated. 2. On our re-evaluation, there was no active infection (no purulent exudate), albeit deep pocket'

(B)

Figure 2.2- LIME-based interpretation of BERT model predictions. (A) Key features for classifying stage III vs. non-stage III. (B) Significant words and features for distinguishing grade B vs. non-grade B

2.3.1 Prediction of Periodontitis Stage (4-Categories)

To predict the stage of periodontal disease, the preprocessed and augmented training dataset was shuffled to introduce randomness, with a batch size of 20 utilized for efficient training. Over the course of 40 epochs, the model was trained on the shuffled dataset, with validation data used to monitor performance and prevent overfitting. The results showed a high accuracy in predicting the stage of the disease using patients' clinical notes (Figure 2.3). The model could predict the patients' stage with 68.01% training, 69.01% validation, and 66.2% test accuracy. Additionally, the model performance was evaluated on a set of 210 new patients' charts (stage I= 3, stage II= 27, stage III= 139, and stage IV= 41) and showed higher accuracy in classifying stage III compared to other stages (64%). The precision column indicated that our model was successful in assigning 70% of the patients in stage III to the right class. The F1-Score column, which is usually used to judge the overall performance of the model and is defined as the average of precision and recall values in the previous columns, showed that our proposed model performed well in predicting the patients with stage III (F1-score=79%) compared to other stages (Table 2.1).

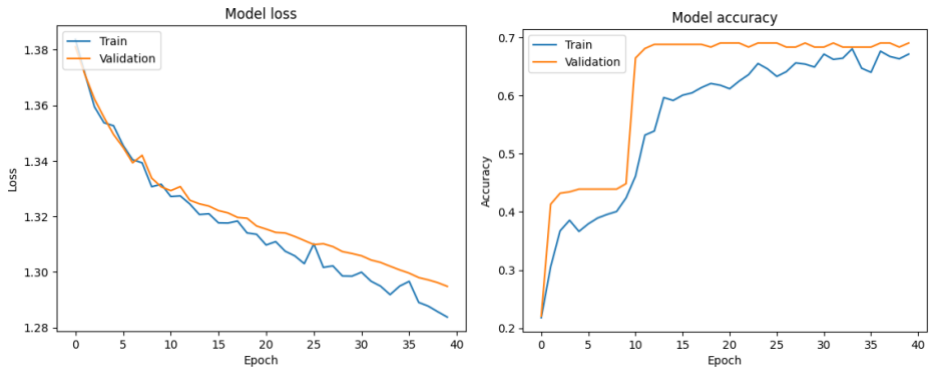


Figure 2.3- The results of the BERT model for predicting the 4-class stage of periodontitis

Table 2.1- Model performance for predicting the stage of periodontitis on a new unseen dataset

Stage	Precision	Recall	F1-score
I	0.00	0.00	0.00
II	0.30	0.30	0.30
III	0.70	0.91	0.79
IV	0.00	0.00	0.00

2.3.2 Prediction of Periodontitis Stage (3-Categories)

To predict the stage of periodontal disease into 3 distinct categories for concatenating with image processing results, the preprocessed and augmented training

dataset was utilized for training a BERT model over the course of 8 epochs, with validation data used to monitor performance and prevent overfitting.

The results showed a high accuracy in predicting the stage of the disease using patients' clinical notes (Figure 2.4). The model could predict the patients' stage with 96.45% training, 98.62% validation, and 79.25% test accuracy. The model performance was evaluated on the same set of 210 new patients' charts and it showed a high accuracy in classifying stage III/IV compared to other stages, and a higher accuracy compared to 4-categories predictions (78% vs. 64%). The precision column indicates that the model was successful in assigning 87% of the patients in stage III/IV to the right class. The F1-Score column showed that our proposed model performed highly accurately in predicting the patients with stage III/IV (87%) compared to other stages (Table 2.2).

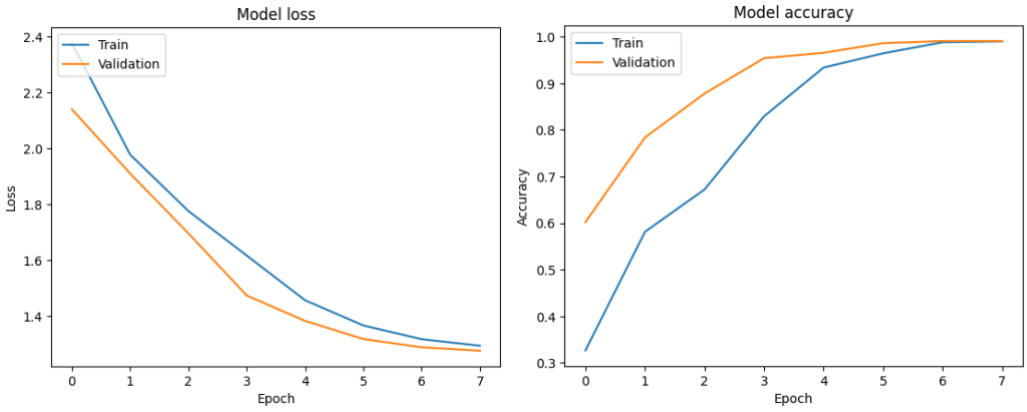


Figure 2.4- The results of the BERT model for predicting the 3-class stage of periodontitis

Table 2.2- Model performance for predicting the stage of periodontitis on new unseen dataset

Stage	Precision	Recall	F1-score
I	0.00	0.00	0.00
II	0.22	0.22	0.22
III/IV	0.87	0.87	0.87

2.3.3 Prediction of Periodontitis Grade

To predict the grade of periodontal disease, the training dataset was preprocessed and enriched by shuffling to ensure randomness. A batch size of 20 was utilized to enhance training efficiency. The model underwent 7 training epochs on the shuffled dataset, with validation data employed to monitor performance and counteract overfitting.

The outcomes revealed promising results in predicting periodontitis grade using patients' clinical notes (Figure 2.5). The model achieved 89.7% training, 88.9% validation, and 72.3% test accuracy in predicting patients' periodontitis grades. Particularly, the model demonstrated high precision in classifying grade B compared to other grades, successfully assigning 72% of patients in grade B to the correct class.

The performance of the model was evaluated on the same 210 patients' chart, which had the image data available. One patient's chart was excluded as the grade of the disease

was not determined by the periodontist (grade A= 8, grade B= 142, and grade C= 59). The proposed BERT model effectively predicted the grades of these new, unseen patients with an accuracy of 65%. Table 2.3 illustrates the model's performance on the new dataset, categorized according to the individual grades of the disease. The F1-Score column showed that the proposed model performed highly accurately in predicting the patients with grade B (77%) compared to other grades.

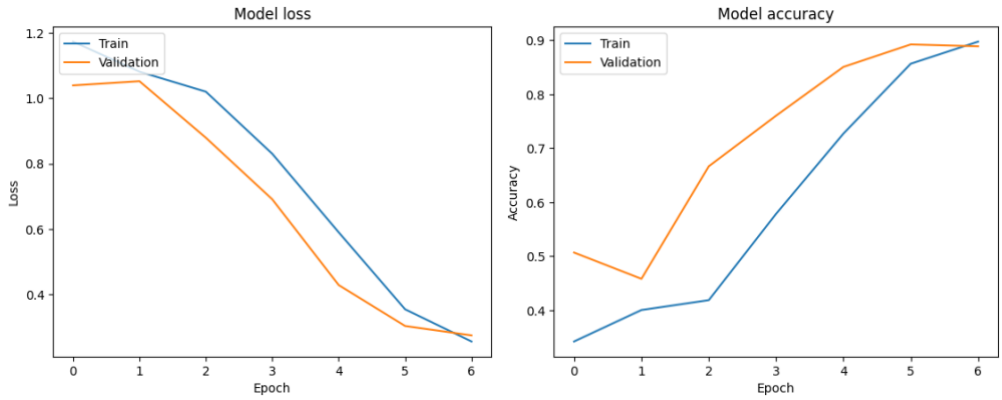


Figure 2.5- The results of the BERT model for predicting the grade of periodontitis

Table 2.3- Model performance for predicting the grade of periodontitis on new unseen dataset

Grade	Precision	Recall	F1-score
A	0.14	0.12	0.13
B	0.72	0.83	0.77
C	0.45	0.29	0.35

2.4 DISCUSSION

The increasing popularity of computerized dental records provides the opportunity to utilize AI-based technologies such as ML and DL models to improve patient care; however, a recent review has reported dentistry's clinical integration of such techniques has lagged (70). To the authors' knowledge, there have only been a few studies that have investigated the use of ML and DL techniques in the diagnosis of periodontitis. Most of these studies have focused on analysing panoramic radiographs/and or clinical examinations or the biomarkers/bacteria extracted from the saliva (211), which is costly and time-consuming. In this dissertation, an NLP approach was evaluated for automatically extracting patient data from textual notes and demonstrated its ability to classify the stage and grade of the periodontitis accurately.

Currently, digital dental and medical records are gaining popularity for storing patient information. Data in electronic medical/dental records can be divided into three kinds: structured data, semi-structured data, and unstructured data. Unstructured text is defined as narrative data, consisting of clinical notes, discharge records, and radiology reports. Unstructured texts store a lot of valuable information however, common structural frameworks are usually lacking, and many errors including improper grammatical use, spelling errors, local dialects, and semantic ambiguities exist, which result in the complexity of data processing and analysis. Retrieval and application of such valuable information is possible through the application of text mining methods (222).

In this chapter of the dissertation, the BERT model was applied to predict the stage and grade of the periodontal disease using the patients' textual notes. The results

demonstrated that the BERT model was able to classify the correct stage and grade of the patients' periodontal disease accurately. To date, no previous studies have used patients' dental records collected as text files for analysis via BERT algorithms. In medicine, Haulcy and Glass compared the performance of five classifiers, as well as CNNs and LSTM networks on classification of Alzheimer's disease using audio features and text features (236). They reported that the top-performing classification models were the SVM and random forest classifiers trained on BERT embeddings, which both achieved an accuracy of 85.4% on the test set. Recently, new domain-specific BERT models pretrained on large-scale biomedical (BioBERT) and medical dictionaries (Med-BERT) have been introduced. These models have shown promising results in extracting valuable information from biomedical and medical literature, respectively, and outperformed the state-of-the-art methods on medical and biomedical tasks (237,238). However, we used a base architecture in the study's BERT model, as the dental patients' notes mostly include a detailed document of the dental history, physical examination, diagnosis, and treatment planning, which is different from the biomedical or medical records. Moreover, as the study is the very first application of the BERT model on patients' dental notes, we decided to use a more generic architecture.

As the current study aimed to examine several different health-related medical and dental possible contributing factors using patient data, NLP methods were best suited to address the complexity and multi-dimensional nature of each due to their known ability to detect complex relationships and classify the outcome accurately (70). NLP models outperform the traditional statistical methods as they can consider a broader range of

features without strict, predetermined predictor and outcome parameters and classify the output with impressive accuracy (239).

In spite of its promising results in medical research, utilization of DL and NN in dental research generally and in periodontal field specifically, is relatively scarce. A recent review by Ossowska et al. (240) has shown that new technologies are developing very quickly in the field of dentistry and AI is spreading into periodontology with the greatest focus on evaluating the periodontal BL, peri-implant BL, and to predict the development of periodontitis due to the psychological features.

The potential utility of applying ML and DL models to automatically classify the stage and grade of the periodontitis using patients' notes have been found in the current study. The BERT method presented, relying on intelligent text mining, yielded promising results. However, these findings must be interpreted with caution according to the preliminary nature of this research as it is a novel method applied on unstructured patient notes recruited from a university periodontal clinic. Therefore, the DL technique used in the present study requires validation to confirm its efficiency on different types of medical/dental records from other centers. Due to the imbalanced input data, including more patients with stage III and grade B compared to stages I or II and grade A, the accuracy of the model was shown to be higher in these classes. This can be attributed to the higher number of referrals among patients in more severe periodontal conditions. Future studies should analyze larger and more balanced datasets using these ML/DL techniques to provide additional validation. However, while our method is novel, our findings showed a relatively high accuracy in correct classification of the stage and grade of the periodontitis, which suggests that our proposed method was successful in extracting

the relevant information from patient notes and classifying the stage and grade of the periodontal disease.

2.5 CONCLUSION

The new proposed model presented in this chapter is the preliminary application of NLP techniques to assign the stage and grade of the periodontal disease using patients' notes compared to the previous models of utilizing radiographic images and salivary samples. This model can identify the stage and grade of the periodontitis without the need for a comprehensive clinical and radiographic examination, which is routinely required for disease diagnosis thus, hope to increase enrollment in clinical trials of new therapies, and improve patient outcomes by enabling periodontists to diagnose periodontal disease using the information collected in patients' notes in an accurate and timely manner. This technique can be also applied to other health-care providers in need of using patients' information as large amounts of unstructured texts to classify stages of a disease. It is also suggested that future studies compare various BERT models to identify the best performing one for extracting information from EDR.

Chapter 3- Automating Bone Loss Measurement on Periapical Radiographs for Predicting the Periodontitis Stage and Grade

3.1 BACKGROUND

Periodontitis is a multifactorial and microbiome-associated inflammatory disease that occurs in the dental supporting tissues (periodontium, which includes gingiva, periodontal ligament, cementum, and alveolar bone presenting as bone loss (BL)) (2,3,211). Progression of the disease can adversely affect oral and systemic health and result in tooth loss, reduction of masticatory performance (207) as well as having association with diabetes (208) and coronary artery disease (241). Thus, periodontitis and its complications will impose substantially negative effects on oral health-related quality of life (OHRQoL) (207,210). Moreover, early detection and diagnosis of periodontitis can help in preventing the consequent costly and invasive dental treatment (211).

In clinical settings, the assessment of periodontal conditions involves visual and tactile examinations. The gold standard for examination includes measurements of periodontal pocket depth (PPD), bleeding on probing (BOP), and clinical attachment loss (CAL) (45). Radiographs are employed to confirm diagnosis and treatment plans. However, variations in probe tip diameter, angulation, probing force, and intra-examiner differences can result in divergent outcomes. Additionally, in instances of mild attachment loss or when determining the subgingival localization of the cementoenamel junction (CEJ), accurately determining CAL is challenging due to the difficulty in ascertaining CEJ location. In such cases, precise and reliable assessment relies on the interpretation of

radiographic bone levels radiographically (45,242). Alveolar BL, is defined as any distance from CEJ to the alveolar bone crest that is greater than 2mm (46).

Despite the improvements in the quality of image and resolution over the past decade, interpreting dental imaging is primarily and subjectively conducted by the trained dentist based on the individual's judgement and experience (69,108,110). Inconsistencies in their interpretation may result in misdiagnosis and, in the process of periodontitis evaluation, may lead to wrong measurement of BL (112). Currently, deep learning (DL) techniques are being widely applied for quick evaluation of dental radiographs, without subjective interpretations.

Traditional methods of automated medical image analysis include large amounts of rule-based algorithms or manual preprocessing methods that are time-consuming with low quality and poor generalization capability (114,115,243). This underwent changes with the introduction of convolutional neural networks (CNN) algorithms using DL, which allow for direct interference, recognition and classification of medical images (116,117). Due to its demonstrated efficiency in the field of image understanding (image segmentation, classification, localization and detection) through feature extraction of input data, it is a widely used technique for solving medical image understanding (71).

Image segmentation is the process of identifying key components of an image and separating the image into individual sections or regions- is a fundamental task in medical image processing (157). Several DL models have been introduced for medical image segmentation like the U-Net and mask-RCNN (167). Most image segmentation DL models are based on CNNs (168). Specifically, the U-Net architecture has been thoroughly studied for biomedical image segmentation due to its ability to produce highly accurate segmented

images using limited training data (170,171). The popularity of this algorithm is evident from its widespread adoption across major primary imaging techniques, including computed tomography scans, magnetic resonance imaging, X-rays, and microscopy (171).

An alternative method that employs CNN principles is You Only Look Once (YOLO), which is mainly designed to precisely identify objects in real time (139). A complex generalized approach toward object identification, YOLO combines concurrent processing of information at different convolutional layers with up-and-down sampling approaches similar to U-Net. One of the characteristics of YOLO is to extract features from the entire image and predict bounding boxes (152). In the literature, older versions of YOLO have been widely used in the field of dental radiology to detect mandibular fractures in panoramic radiographs (153), primary and permanent tooth detection on pediatric dental radiographs (154), detection of cyst and tumors of the jaw (155), periodontal BL detection (244), and detection of impacted mandibular third molar teeth (245). However, new versions of YOLO-v9, released in January 2024, demonstrate superior performance with respect to throughput and computational load requirements (156), and provide a network architecture that requires lower computing and training requirements, hence providing a more effective feature integration method, more accurate object detection performance, a more robust loss function, and an increased label assignment and model training efficiency (156).

Several studies have utilized image mining approaches to detect the amount of BL for classifying periodontitis. While accurate identification and delineation of BL is crucial for periodontal treatment planning and monitoring disease progression, most studies have focused on measuring the BL at two points (mesial and distal) (167,234,246,247). Also,

panoramic radiographs have been used in most studies (173,244), despite it has been shown that the periapical (PA) or bitewing radiographs are the choice images for an accurate evaluation of BL (248,249).

Given the importance of accurate measurement of the amount of BL, and considering that the amount of BL at any point between the mesial and distal aspects of the tooth may be greater than at these two specific sites, which will have the potential for underestimating the extent of BL, the objective of this chapter was to introduce automated approaches for segmenting BL for each tooth across its entire width and detect teeth apexes for measuring the root length in PA radiographs, and use above information to predict the stage and grade of periodontitis by computing the percentage of BL (BL ratio to the root length).

3.2 MATERIALS AND METHODS

3.2.1 Study Population

After obtaining the University Human Research Ethics Board approval under the code Pro00107743, PA radiographs belonging to the same patients whose clinical notes were used in Chapter 2 were collected from the Periodontal Graduate Program in the School of Dentistry between 2017-2021. Two hundred and ten patients had PA radiographs in their records and a total of 1582 images were included in the study for testing the model performance. To obtain predictions from the model for images associated with 210 patients, whose charts were utilized to assess the performance of the text processing model in the previous chapter, and to seamlessly integrate the predictions from both text and image modalities, training was conducted for both BL segmentation and apex detection

models utilizing a dataset comprising 1000 PA images sourced from a private dental radiology office in Iran.

3.2.2 Bone Loss Segmentation

3.2.2.1 Data Preprocessing

To detect the amount of BL, the images were processed using an open-source tool called Roboflow (250) for annotation using polygons. Each radiograph was read by a dentist and a periodontist with more than 5 years of clinical experience. To manually determine the BL (the area between the CEJ of the teeth and bone level in tooth-bearing areas), a polygon passing through the CEJ of the teeth and the corresponding bone levels was drawn (Figure 3.1). The final label was determined based on consensus between the dentist and periodontist.

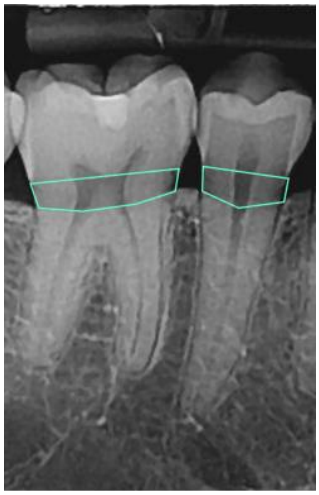


Figure 3.1- A polygon indicating the BL area (between CEJ and bone level)

The annotated area was then changed to a binary mask (BL as the mask area or region of interest and the background structures in black). The images and their corresponding binary masks were exported as Portable Network Graphics (PNG) format

for further analysis (Figure 3.2). The image dataset was preprocessed for future analysis, which included resizing to 160*320 pixels and enhancement using histogram equalization technique.

Figure 3.2- An example of a PA image with its corresponding mask

3.2.2.2 Data Augmentation

Image augmentation continued after data splitting to ensure that data leakage would not occur. First, the total of 1000 PA images was randomly split into three parts: training, validation, and testing in an 80:10:10 ratio (251,252), respectively using scikit-learn library (232), then the augmentation was done through horizontal flipping and/or rotating at different angles (253) to increase out the dataset by two folds. The final result was a total of 1600 PA images and their corresponding masks for training and 200 images and masks for validation.

3.2.2.3 U-Net Architecture for Bone Loss Segmentation

Following data augmentation, the image dataset was normalized by dividing by 255 to scale down the pixel values of the images and their corresponding masks to a range between 0 and 1. The set of 1582 images obtained from 210 patients referred to the university periodontal clinic were finally used to evaluate the model performance and its generalizability.

Different variations of U-Net (128) were trained and evaluated to find the best architecture and hyperparameter settings for the segmentation model. Finally, a U-Net model featuring 4 encoder blocks with increasing filters (64, 128, 256, 528), a bridge of a convolutional block with 1024 filters, and 4 corresponding decoder blocks with decreasing filters (512, 256, 128, 64) was utilized for training. The model parameters are listed in Table 3.1, and the model architecture is provided as Appendix 1.

Table 3.1- Parameters for U-Net model training

Property	Value
Batch size	8
Epochs	15
Learning rate	0.0001
Optimizer	Adam
Drop-out rate	0.2
Loss function	Dice coefficient
Activation	ReLU Sigmoid (final layer)

3.2.2.4 Evaluation of Diagnostic Outcome

Evaluation metrics include accuracy (the ratio of the correctly predicted observations to the total observations), recall (the ratio of true positives to the sum of true positives and false negatives), precision (the ratio of true positives to the sum of true positives and false positives), and F1-score (the harmonic mean of precision and recall

(234). Also, the Intersection over Union (IoU) and area under receiver operating characteristics (ROC) curve were used to evaluate the model performance on the validation dataset. The magnitude of these metrics ranges from 0 to 1 indicating a perfect performance (235).

Additionally, the comparison was made between the maximum amount of BL predicted by the model for each tooth and the measurements obtained by the dentist and periodontist using several metrics that are the standard metrics in understanding the efficiency of the model in terms of error rate, including the mean absolute error (MAE) and root mean square error (RMSE). Also, the intraclass correlation coefficient (ICC) (254,255), which is a desirable measure of reliability that reflects both degree of correlation and agreement between measurements was used to assess the consistency and agreement among three raters (dentist, periodontist, and model predictions) concerning numerical measurements. For the MAE and RMSE the best value equals 0 while the worst value can be infinity (256). For ICC, values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability (257).

3.2.3 Apex detection

3.2.3.1 Data Preprocessing

A total of 1000 PA images were imported into Roboflow (250) for annotation. Annotation was performed by delineating bounding boxes around the apices, ensuring that the center of each box corresponded to the apex's anatomical point. The annotations were

carried out by a general dentist and subsequently reviewed and approved by a periodontist with over 10 years of experience (Figure 3.3).

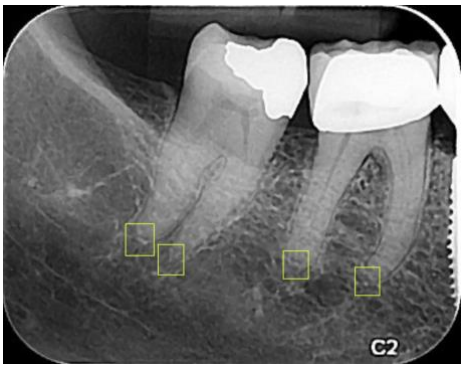


Figure 3.3- Annotating the apexes on PA images using bounding boxes

Subsequently, the annotated images were exported in PNG format for subsequent analysis. The final determination of the bounding box was made through consensus between the dentist and the periodontist. Furthermore, the image dataset underwent preprocessing steps in preparation for future analysis. This preprocessing included resizing all images to dimensions of 160 x 320 pixels. This standardized format ensures consistency and facilitates subsequent computational procedures, ensuring the compatibility of the dataset for further analysis and model training.

3.2.3.2 Data Augmentation

The data was split using Roboflow (250) splitting tool into 70% training, 20% validation and 10% test datasets. Image augmentation was done afterwards during model training to ensure that data leakage would not occur.

3.2.3.3 YOLO-v9 Architecture for Apex Detection

To detect the apexes of the teeth, the YOLOv9 model was utilized. Hyperparameters were defined to optimize training, with a learning rate set at 0.01. The model was trained for 75 epochs with a batch size of 16. During training, various techniques were employed to enhance model performance and robustness. These included: data augmentation using techniques such as blur, median blur, and Contrast Limited Adaptive Histogram Equalization (CLAHE) to introduce diversity and variability in the training dataset. Additionally, optimization was performed using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.01. The training process was monitored using metrics such as box loss, class loss, and DFL (distribution focal loss), which were computed and evaluated across multiple epochs.

The training dataset consisted of annotated images and labels, which were scanned and preprocessed before training commenced. The dataset was divided into training and validation sets, with appropriate caching mechanisms implemented to optimize data loading and processing efficiency. Furthermore, performance evaluation was conducted periodically during training using metrics such as precision, recall, and mean Average Precision (mAP) calculated at various IoU thresholds. The model parameters are listed in Table 3.2.

Table 3.2- Parameters for YOLO-v9 model training

Property	Value
Batch size	16
Epochs	75
Learning rate	0.01
Optimizer	Stochastic Gradient Descent (SGD)

Loss function	Box loss, Class loss, and DFL (distribution focal loss) loss
Activation	Leaky ReLU Sigmoid (final layer)

3.2.3.4 Evaluation of Diagnostic Outcome

To evaluate the model performance and generalizability in the correct detection of root apexes, precision, recall, and mAP were used on the validation dataset. Additionally, the model's performance on the new test dataset (1582 images) was evaluated using MAE (the average magnitude of the errors between the predicted values and the actual values), RMSE (the square root of the mean of the squared differences between the predicted and actual values), and ICC (two-way random effect, single rater, absolute agreement) (255, 258) to compare the agreement in the coordinates of the apexes in the ground truth with those predicted by the model.

3.2.4 Measurement of Maximum Bone Loss Percentage

3.2.4.1 Finding the Maximum BL Height

The maximum BL for each tooth on 1582 new PA radiographs was independently measured by a general dentist and a periodontist, both with over five years of experience. These measurements were conducted using Adobe Photoshop 24.1.1, where the cursor was dragged to draw a line connecting the CEJ to the deepest bone level via the ruler tool. BLs less than 2 mm from the CEJ were not considered normal, as they were not detectable by the model.

To determine the maximum height of BL in every individual PA radiograph first, a custom Python function was developed to automate the process of quantifying BL height within the images. This function utilized established image processing techniques, including thresholding and connected component analysis, to identify regions of interest corresponding to areas of BL.

Upon thresholding the input image, connected component analysis was applied to identify distinct regions representing areas of BL (white areas in the masks). Each identified region was then enclosed within a bounding box, facilitating the extraction of its maximum height. This process ensures that the height measurement accurately reflects the extent of BL within the region of interest.

To evaluate the performance of the method in accurately quantifying maximum BL height across the dataset, the measured maximum BL heights by three raters were compared using the MAE, RMSE, and ICC utilizing the “pingouin” package in Python (259) to assess the reliability or consistency of measurements taken by computer, dentist, and the periodontist.

3.2.4.2 Finding the Root Length

3.2.4.2.1 Finding the Coordinates of the Segmented CEJ Center

To accurately calculate the root length of lower and upper teeth from radiographs, a method was developed to locate the center point of the CEJ using OpenCV for image processing. Firstly, the mask images were preprocessed by thresholding to obtain binary images. The thresholding technique converts grayscale images into binary images, simplifying subsequent image analysis.

Subsequently, contours representing the segmented mask in the binary images were identified using the contour detection algorithm provided by OpenCV. Contours are continuous lines or curves that represent the boundaries of objects in an image. In this context, contours outlined the edges of the BL area in the mask images.

For each identified contour, the bounding rectangle enclosing the contour was calculated. From the bounding rectangle, the center point of the upper border of the segmented area for lower teeth, and lower border of the segmented area for the upper teeth were determined. This center point corresponds to the approximate location of the CEJ center, a critical anatomical landmark for measuring root length. The coordinates were then recorded for further analysis.

To evaluate the performance of the method in accurately locating the CEJ center points across the dataset, the values were compared to their ground truth using the MAE, RMSE, and ICC.

3.2.4.2.1 Measuring the Distance Between CEJ and Detected Apex

To calculate the root length for finding the percentage of BL, the Euclidean distance formula - enables the calculation of the straight-line distance between two points on the image grid solely based on their pixel coordinates- was utilized. The formula encapsulates these computations:

Distance= $\sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$

in which, (x1, y1) and (x2, y2) denote the pixel coordinates of the two points respectively. (x1) and (x2) represent the x-coordinates and (y1) and (y2) represent the y-coordinates of the first and second points. By applying this formula, the exact distance between the

obtained CEJ center point and the root apex were determined. Finally, using the maximum BL height and the root length for each individual tooth, the BL percentage was calculated.

Also, the BL percentage for each tooth on PA radiographs was independently measured by the same dentist and periodontist. These measurements were performed using Adobe Photoshop 24.1.1, where the cursor was used to draw a line connecting the CEJ to the root apexes (root length) with the ruler tool. The BL percentage was calculated by dividing the maximum BL by the measured root length. The BL percentages obtained from all three raters were then compared to evaluate consistency using RMSE, MAE, and ICC metrics to find the consistency in the BL percentage measurements.

3.2.4.4 Determining the Stage and Grade of Periodontitis

The calculated percentage of the maximum BL was then used to categorize the patients' periodontitis stage as stage I (<15%), stage II (15-33%), and stage III/IV (>33%). Also, to assign the grade of the disease, the same value was divided by the patient's age. The outcome values less than 0.25 were assigned to grade A, between 0.25 and 1 were assigned to grade B, and values greater than 1 were assigned to grade C.

To evaluate the final results, evaluation metrics (accuracy, precision, recall, and F1-score) were used to compare the output of this chapter with the ground truth stages and grades. Figure 3.4 represents a flowchart illustrating the process of bone loss measurement and the classification of periodontitis stage and grade.

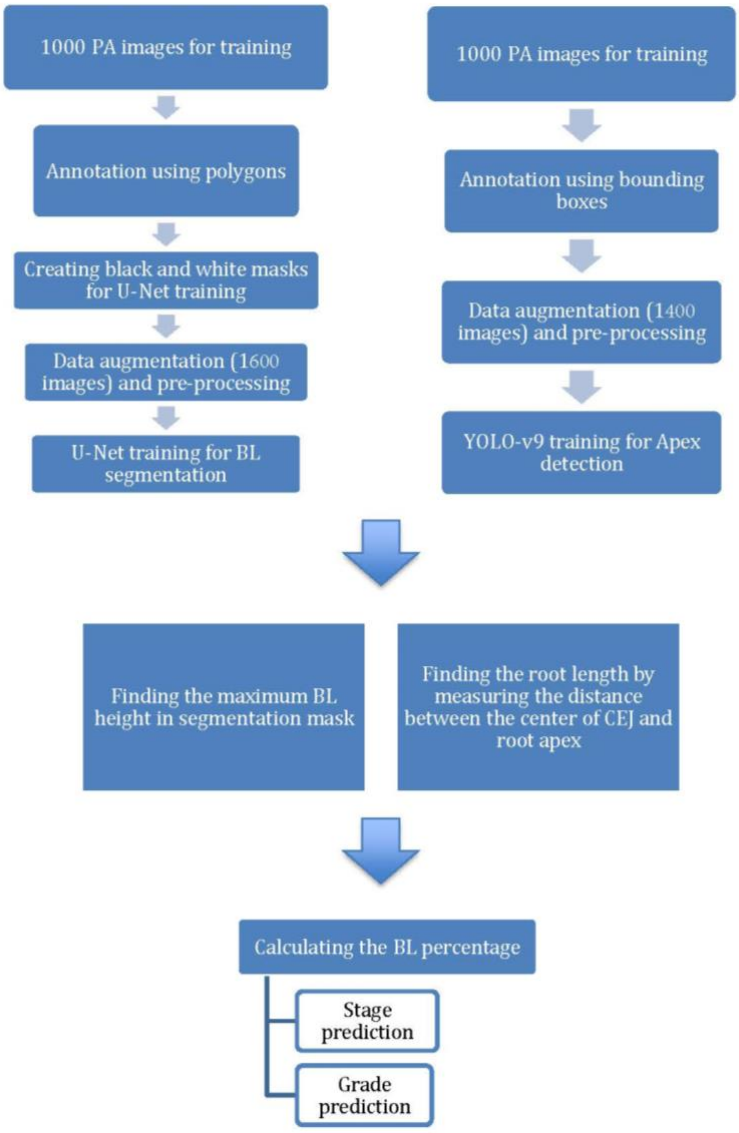


Figure 3.4- Flowchart of bone loss measurement and periodontitis stage and grade classification process

3.3 RESULTS

3.3.1 Bone Loss Segmentation

3.3.1.1 U-Net Performance in Segmenting BL

The U-Net model achieved an accuracy of 94.5%, 92.9%, and 95.5% on training, validation, and test datasets, respectively for segmentation of BL. Over the course of 15 epochs, the model demonstrated a consistent improvement in both training and validation performance, as indicated by the metrics of loss and accuracy. The learning rate remained constant at 0.0001 throughout the training, ensuring a stable and gradual optimization process without drastic changes that could destabilize the learning dynamics. The results of the training process are shown in Figure 3.5.

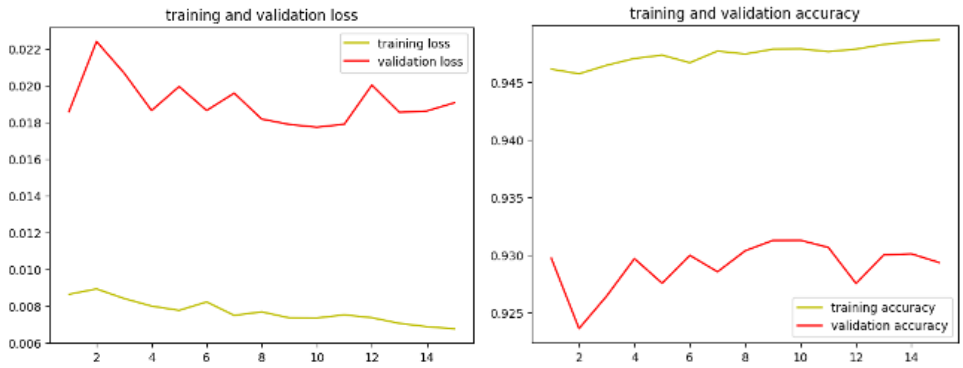


Figure 3.5- Training results of the U-Net model for segmentation of BL on PA radiographs

In addition to the training and validation metrics recorded over the 15 epochs, the model's performance was further evaluated using several other metrics to provide a comprehensive understanding of its effectiveness. The ROC curve and Area Under the Curve (AUC) were used to assess the model's diagnostic ability for both validation and new test images.

The ROC curve plots (Figure 3.6-A and B) the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings, offering a graphical representation of the model's performance. The AUC provides an aggregate measure, with an AUC of 1 indicating a perfect model and an AUC of 0.5 indicating no discriminatory power (259). For the validation dataset, a high AUC value of 0.95 demonstrated the model's strong ability to segment the BL area correctly. Similarly, evaluating the ROC curve and AUC (0.91) on the test dataset confirmed the model's generalization capabilities. Consistent high AUC values for both datasets indicated that the model performed well without overfitting to the training data.

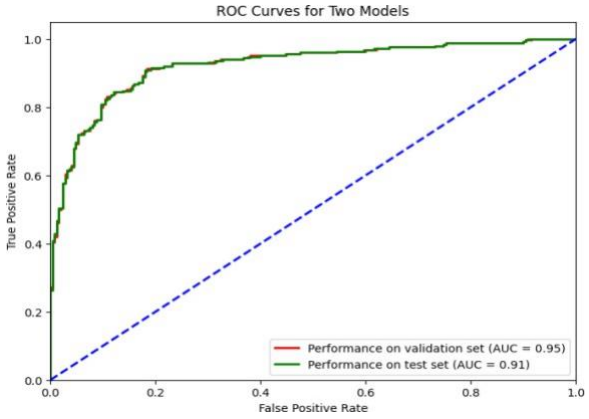


Figure 3.6- ROC curve and AUC, evaluating the model performance on the validation (red) and test (green) datasets

The IoU, also known as the Jaccard Index, was used to further evaluate the segmentation performance. IoU measures the overlap between the predicted and true segmentation masks by dividing the intersection of these masks by their union. IoU values also range from 0 to 1, with 1 indicating perfect segmentation where the predicted mask exactly matches the ground truth. This metric is particularly valuable for assessing how

well the model predicts the boundaries and shapes of objects within an image. High IoU values confirmed the model's ability to accurately capture the details and contours of the segmented objects.

The proposed model showcased a great performance across various metrics on the test dataset. Figure 3.7 represents a sample of the prediction made by the model for segmenting the BL area in a PA image. The overall accuracy of the model on the test dataset was 95.5%, indicating a high level of correctness in its segmentation task.

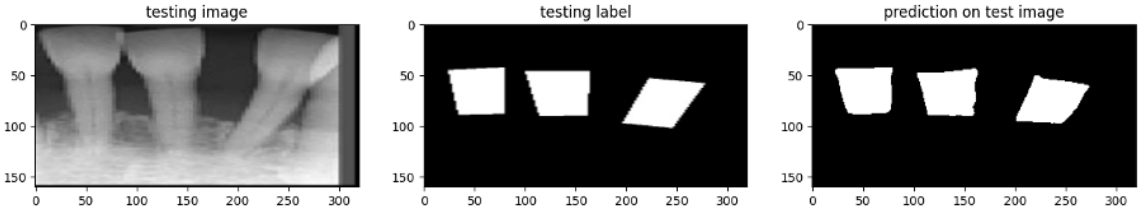


Figure 3.7- Model segmentation performance on the test dataset compared to the ground truth label

Table 2 depicts the evaluation metrics used for assessment of model performance on a set of 1582 new PA images. According to Table 3.3, precision, measuring the proportion of correctly predicted positive pixels among all predicted positive pixels, was notably high at 97.26%. This suggests that the model exhibits a strong capability to accurately segment the regions of interest while minimizing false positives (260). However, the recall metric was comparatively lower at 63.90%. This indicates that while the model accurately identifies many positive pixels, it misses slightly more than 1/3 of the actual positive pixels present in the images. In segmentation tasks, recall reflects the model's ability to capture the entirety of the target objects, highlighting areas where the model may need improvement.

Table 3.3- Performance metrics of the proposed U-Net model on 1582 new set of PA images for segmenting BL

Metric	Value
Accuracy	95.51
Precision	97.26
Recall	63.90
F1-score	77.12
Intersection over Union (IoU)	81.69

The F1-score was recorded at 77.12%. This metric provides a balanced measure of the model's segmentation performance, considering both false positives and false negatives. The F1-score underscores the need for a trade-off between precision and recall, indicating areas where the model may excel in delineating object boundaries while still needing enhancement in capturing entire objects. The IoU was measured at 81.69%, indicative of the effectiveness of the model in capturing the relevant regions of interest within the images.

3.3.1.2 CEJ Centre Localization

In assessing the consistency of locating the CEJ center of the teeth for measuring the BL percentage, a thorough analysis was conducted using the same set of 1582 new test

images. The CEJ center coordinates obtained through the application of OpenCV library on the segmented masks were compared against ground truth annotations, which were verified by a consensus between a dentist and a periodontist.

To quantitatively evaluate the agreement between the predicted and ground truth coordinates, RMSE and MAE metrics were used for both the X and Y coordinates (Table 3.4). Moreover, the ICC was utilized to assess the inter-rater reliability between the model predictions and the ground truth annotations. The ICC values for CEJ_X and CEJ_Y coordinates were calculated as 0.9468 and 0.9629, respectively, indicating excellent agreement between the model predictions and the ground truth (257).

Table 3.4- Assessment of the agreement between the predicted and ground truth coordinates of CEJ center points on the test PA images

Metric	CEJ-X coordinates	CEJ-Y coordinates
RMSE	12.64 (1.07 mm)	10.93 (0.92 mm)
MAE	4.55 (0.38 mm)	4.25 (0.36 mm)

3.3.1.3 Maximum Bone Loss Measurement

The results revealed that the dentist and periodontist exhibited the highest RMSE (16.8, errors in pixel values equal to 1.4 mm) and MAE (13.68 equal to 1.15 mm), indicating the greatest discrepancies between these raters. In contrast, the periodontist and model measurements showed the lowest RMSE (14.80 equal to 1.2 mm) and a moderate MAE (12.17 equal to 1.03 mm), reflecting the best agreement. Measurements between the dentist and the model demonstrated intermediate values, with an RMSE of 15.29 (equal to

1.29 mm) and the lowest MAE (10.70 equal to 0.90 mm), indicating good but slightly better agreement than that between the periodontist and the model.

Also, ICC analysis was done to have a better understanding of agreement among 3 raters. The results indicated that there is a very high level of consistency among the three raters for the maximum BL measurement. Both individual and average raters' scores showed strong agreement, with ICC values consistently above 0.94. This suggests that the ratings are reliable and that the maximum BL measurement done by model can be confidently used for further analysis or decision-making processes.

Pairwise ICC comparison was done to find the agreement between every two raters (Table 3.5). According to the table, the best agreement was observed between the periodontist and the model measurements with the highest ICC1 (0.945) and F-value (35.71), and the 95% confidence interval (CI) of [0.92, 0.96]. A p-value of less than 0.001 indicates that the observed correlation between the raters is highly statistically significant and unlikely to have occurred by chance. In other words, the null hypothesis (that there is no correlation between the raters) is rejected with high confidence. The extremely low p-values support the reliability of the ICC values, reinforcing that the model's predictions are consistent with both the dentist and periodontist's measurements.

Table 3.5- Inter-examiner reliability in maximum BL measurements among periodontist, GP, and the model

Raters	ICC	F-value	95% CI	p-value
Dentist vs. model	0.944	34.73	0.92-0.96	< 0.001

Dentist vs. periodontist	0.937	30.88	0.91- 0.96	< 0.001
Periodontist vs. model	.945	35.71	0.92- 0.96	< 0.001

3.3.2 Apex Detection

The YOLOv9 model, employed for detecting the apexes of teeth, was trained for 75 epochs, leveraging a range of hyperparameters to optimize its performance. The training results are shown in Figure 3.8.

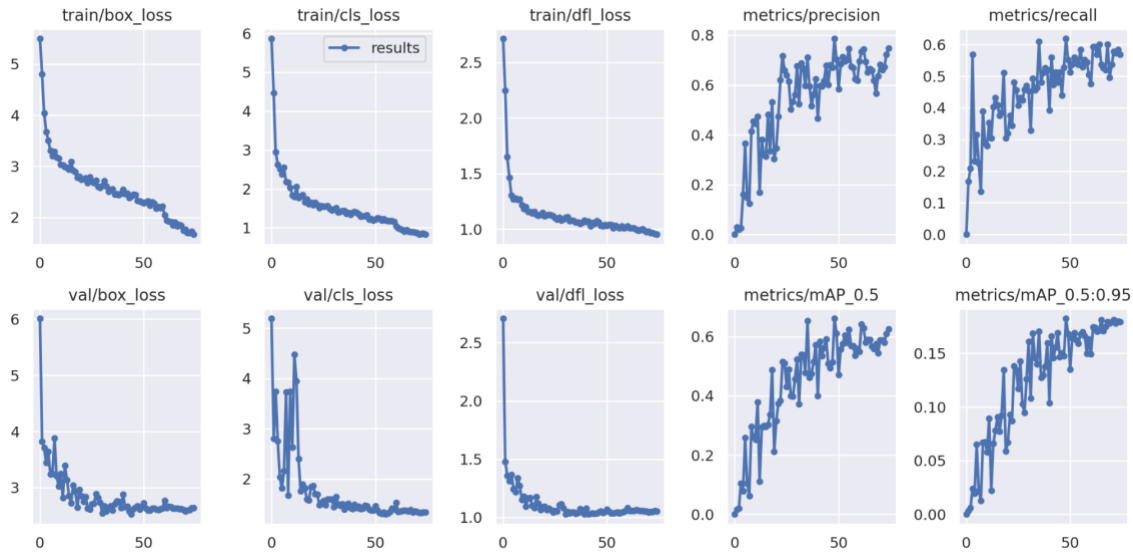


Figure 3.8- Training results of the YOLO-v9 model for apex detection

A sample of the model performance in detection of root apexes has been shown in Figure 3.9.

Figure 3.9- An example of the model performance in detecting apexes on PA images

The evaluation revealed that the model achieved a mAP of 66.7%. Additionally, the precision (P) metric stood at 79.6%, indicating the model's accuracy in correctly identifying apexes. The recall (R) metric, which measures the model's ability to capture all relevant instances, was recorded at 62.4%, showcasing less than two third of the true positives being correctly identified.

Figure 3.10 demonstrates the P-R and F1-curves indicating the performance of the model in detecting root apexes.

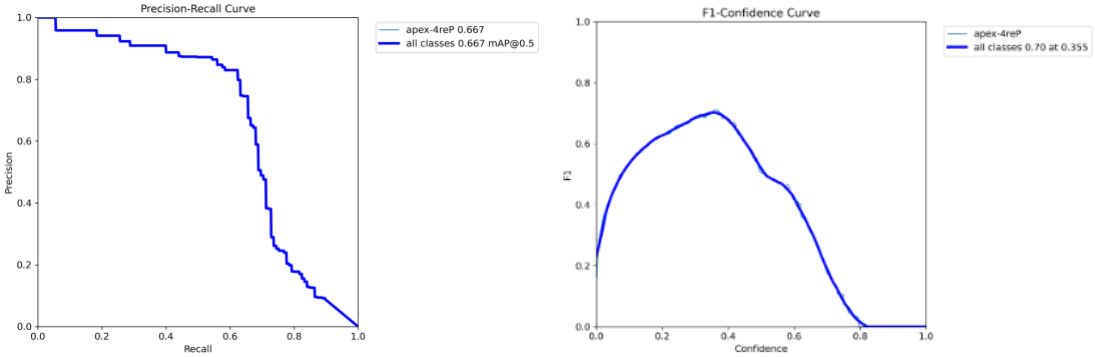


Figure 3.10- Performance analysis of YOLO-v9 model for apex detection: (A) PR and (B) F1-Score curves

To evaluate the performance of the YOLO-v9 model in accurately detecting tooth apexes, a comprehensive analysis was conducted using a new set of 1582 images sourced from patients whose periodontal charts had been previously analyzed through BERT analysis. The coordinates of the detected apexes by the model were compared against ground truth annotations, which were meticulously curated by a qualified dentist and subsequently confirmed by a periodontist to ensure accuracy and reliability.

To quantitatively assess the agreement between the predicted and ground truth coordinates, RMSE and MAE for both the X and Y coordinates were used (Table 3.6).

Furthermore, the ICC measure for absolute agreement was utilized to measure the inter-rater reliability between the model predictions and the ground truth. For the Apex_X coordinate, the ICC value was calculated as 0.9832 with a 95% confidence interval of [0.98, 0.99], indicating nearly perfect agreement between the model predictions and the ground truth. Similarly, for the Apex_Y coordinate, the ICC value was found to be 0.9936 with a 95% confidence interval of [0.99, 1.00], demonstrating excellent agreement between the model predictions and the ground truth.

Table 3.6- Assessment of the agreement between the predicted and ground truth coordinates of detected apexes on the test PA images

Metric	Apex-X coordinates	Apex-Y coordinates
RMSE	1.25 (0.1 mm)	4.8 (0.4 mm)
MAE	4.8 (0.4 mm)	1.6 (0.13 mm)

3.3.3 Bone Loss Ratio

The results indicated that the model has higher consistency with the periodontist (lower RMSE values) compared to the dentist. Specifically, the RMSE (3.98, 0.34 mm) values for the periodontist and measurements made by the model comparison were lower than those for the dentist and model comparison (4.64, 0.39 mm), suggesting that the model's predictions align more closely with those of the periodontist. Additionally, the comparison between the periodontist and the dentist revealed that there is a moderate level of consistency between their measurements (RMSE= 4.28, 0.36 mm). However, the slightly higher MAE and RMSE values in the dentist and periodontist comparison indicated some variability in their BL percentage evaluations.

Additionally, the ICC analysis indicated an excellent consistency among the three raters for the BL percentage measurement. Both individual and average raters' scores demonstrated excellent agreement, with ICC values consistently exceeding 0.94.

Pairwise ICC comparison was done to find the agreement between every two raters (Table 3.7). According to the table, the pair with the best agreement was the model vs. periodontist, with the highest ICC (0.9588) and F-value (47.5692), and a CI of [0.94, 0.97]. A p-value below 0.001 demonstrates that the observed agreement between the raters is statistically significant and not a result of random chance. This allows us to confidently reject the null hypothesis, which asserts that there is no correlation between the raters. The remarkably low p-values affirm the dependability of the ICC values, underscoring the consistency of the model’s predictions with the assessments made by both the dentist and the periodontist.

Table 3.7- Inter-examiner reliability in BL percentage measurements among periodontist, GP, and the model

Raters	ICC	F-value	95% CI	p-value
Dentist vs. model	0.943	34.33	0.92- 0.96	< 0.001
Dentist vs. periodontist	0.954	42.54	0.93- 0.97	< 0.001
Periodontist vs. model	0.959	47.57	0.94- 0.97	< 0.001

3.3.4 Prediction of Periodontitis Stage

The performance of the proposed method for predicting the stage of periodontitis was evaluated on a new test dataset comprising 210 samples (181 stage III/IV, 27 stage II, and 3 stage I). The overall metrics for stage prediction demonstrated promising results, as shown in Table 3.8. The method achieved a precision of 0.7805, an accuracy of 0.7952, a recall of 0.7952, and an F1-score of 0.7878. These metrics indicate a robust performance in identifying the correct stage of periodontitis across the test dataset.

The detailed analysis of the individual stage categories revealed varying degrees of prediction accuracy, as summarized in Table 3.7. For stage III/IV, the method exhibited excellent performance with a precision, recall, and accuracy of 1.0, 0.8950, and 0.8950, respectively, resulting in a high F1-score of 0.9446. This indicates that the model was particularly effective in identifying advanced stages of periodontitis.

Conversely, the method's performance for stage II was significantly lower, with precision, recall, and accuracy of 1.0, 0.1923, and 0.1923, respectively, and an F1-score of

0.3226. This suggests difficulties in correctly identifying stage II cases, potentially due to overlapping features with other stages.

For stage I, the method failed to correctly identify any cases, resulting in precision, recall, accuracy, and F1-score all being 0.0. This highlights a significant limitation in the model's ability to predict early-stage periodontitis, likely due to the insufficient number of training samples for stage I, with only three samples available. To enhance the model's performance in this category, further refinement and the inclusion of a larger, more representative dataset for early-stage periodontitis are necessary.

Table 3.8- Metrics for individual stage categories

Stage	Precision	Accuracy	Recall	F1-Score
I	0.00	0.00	0.00	0.00
II	1.00	0.19	0.19	0.32
III/IV	1.00	0.89	0.89	0.94

3.3.5 Prediction of Periodontitis Grade

The performance of the proposed method for predicting the grade of periodontitis was assessed on a separate new test dataset of 209 samples (8 with grade A, 58 grade B, and 144 grade C). The overall metrics for grade prediction showed a precision of 0.67, an accuracy of 0.66, a recall of 0.66, and an F1-score of 0.66. These results indicate a

reasonable ability to predict the grade of periodontitis, although there is room for improvement.

An in-depth examination of the individual grade categories provided further insights, as detailed in Table 3.9. For grade B, the method performed excellently with precision, recall, and accuracy of 1.0, 0.7083, and 0.7083, respectively, and an F1-score of 0.8293. This reflects a strong capability in identifying grade B periodontitis cases.

For grade C, the method also showed good performance with a precision, recall, and accuracy of 1.0, 0.6207, and 0.6207, respectively, and an F1-score of 0.7660. This suggests that the model is fairly accurate in predicting more advanced grades of periodontitis. However, the method struggled significantly with grade A predictions. The precision, recall, accuracy, and F1-score for grade A were all 0.0, indicating a complete inability to correctly identify any cases in this category. This underscores the need for further increase in the size of data by utilizing a larger, more diverse training dataset to capture the nuances of early-grade periodontitis.

Table 3.9- Metrics for individual grade categories

Garde	Precision	Accuracy	Recall	F1-Score
A	0.00	0.00	0.00	0.00
B	1.00	0.71	0.71	0.77
C	1.00	0.62	0.62	0.83

3.4 DISCUSSION

This chapter aimed to enhance the accuracy and efficiency of diagnosing and grading periodontal disease using advanced ML models. Specifically, a U-Net model for segmenting BL on PA images and a YOLO-v9 model for detecting the apexes of teeth were employed. These methods were chosen for their recent advancements and high performance in medical imaging tasks (156,161,162). Comparing our results with those from existing literature highlights both our achievements and the areas where our approach offers distinct advantages.

3.4.1 Bone Loss Segmentation

The proposed U-Net model achieved high performance in segmenting BL, with an accuracy of 95.62%, precision of 94.67%, recall of 66.06%, F1-score of 78.16%, and IoU of 80.74%. These metrics are competitive with those reported in similar studies, such as the work by Shon et al. (261), where the U-Net model was used to segment the boundary of periodontal BL (PBL) and CEJ on panoramic radiographs. In their approach, the U-Net algorithm was applied separately to detect PBL and CEJ boundaries by creating segmentation masks for each structure. The model was trained independently for PBL and CEJ segmentation, with training conducted for 30 epochs and a batch size of 16 on a dataset of 1044 panoramic radiographs. Shon et al. reported training and validation accuracies of 98.6% and 98.9% for PBL and CEJ segmentation, respectively.

In this chapter, PA radiographs, which are the standard for detecting periodontal disease (262-264), were utilized. The method for segmenting the boundaries of the BL and CEJ involved a single U-Net model, applied to segment the boundaries of both structures.

This approach simplifies the segmentation process and is tailored to the specific radiographic modality used in periodontal diagnosis.

Chang et al. (105) also applied a hybrid DL approach for detecting the radiographic bone level on panoramic radiographs. The aim of their study was to develop an automated method for diagnosing PBL (of individual teeth) for staging the periodontitis according to the 2017 World Workshop criteria (2) on dental panoramic radiographs using the DL hybrid method. They applied a modified CNN from the Mask R-CNN based on a feature pyramid network (FPN) and a ResNet101 backbone to detect the PBL and CEJ on 330 panoramic radiographs which were increased by 64 times using the augmentation techniques. Similar to the approach by Shon et al., the model was applied twice to segment the boundaries of PBL and CEJ separately. Their results demonstrated a Jaccard index of 0.92, pixel accuracy of 0.93, and DC of 0.88 for PBL detection. For CEJ level segmentation, the Jaccard index, pixel accuracy, and DC values were 0.87, 0.91, and 0.84, respectively. These results are comparable to those obtained in our study, highlighting the effectiveness of automated models for detecting PBL on radiographs. However, in this study, PA radiographs were used instead, and a single U-Net model was trained to segment both the PBL and CEJ. This approach represents a more efficient method for determining the extent of PBL on dental radiographs.

Ezhov et al. (266), applied U-Net with CNN architecture to segment the alveolar BL on 1135 CBCT scans. The PBL sensitivity and specificity were 0.95 and 0.97, respectively with the highest values obtained for severe PBL (0.93 and 0.99, respectively). Although CBCT can provide three-dimensional information, there are still some

limitations caused by artifacts, noise and poor soft tissue contrast (267). Additionally, CBCT is not routinely prescribed for the detection of periodontitis due to its relatively high radiation dose compared to PA radiographs.

In a recent study conducted by Li et al. (268), the authors systematically reviewed papers to identify the application of DL for the classification of periodontitis and assess the accuracy of this approach. In terms of dental imaging modalities, the studies included primarily utilized PA images, panoramic images, and CBCT images for periodontitis classification. PA radiograph images capture the teeth and surrounding alveolar bone, thus providing comprehensive information on PBL. However, this modality has a limited view, typically showing only three to four teeth per image (269). Regarding the task of classification using DL models, classical models such as U-Net and YOLO were often utilized in the included studies (128,270), regardless of the specific diagnosis task chosen. The authors reported that U-Net has been proven to quickly and accurately identify targets in medical images and generate high-quality segmentation results. Additionally, the structure of U-Net can be flexibly adjusted according to the specific needs of the task (268,271).

Dujic et al. evaluated several transformer models for the automated detection of periodontal BL using a dataset of 21,819 anonymized PA radiographs from both the upper/lower and anterior/posterior regions. The areas were initially assessed by calibrated dentists based on periodontal BL and categorized into minor, moderate, and severe BL regions. Five vision transformer networks (ViT-base/ViT-large from Google, BEiT-base/BEiT-large from Microsoft, and DeiT-base from Facebook/Meta) were employed and

assessed. The overall diagnostic accuracy ranged from 83.4% to 85.2%, with the ViT-base model achieving the highest accuracy. The authors noted only minor differences in BL detection across the evaluated transformer networks (272).

Several studies have utilized image mining approaches to detect the amount of BL for classifying periodontitis. While accurate identification and delineation of BL is crucial for periodontal treatment planning and monitoring disease progression, most studies have focused on measuring the BL at two points (mesial and distal) (112, 234, 246, 247). The novelty of this research lies in segmenting the BL across each individual tooth rather than focusing solely on mesial and distal points. This approach mitigates the potential for underestimating the extent of BL, providing a more comprehensive and accurate assessment.

3.4.2 Apex Detection

A fundamental component of computer vision, object detection drives advances by leveraging a variety of DL and ML models. Historically, two-stage object detectors have been popular and effective in the field of object detection. However, recent advancements in single-stage object identification methods and their underlying algorithms have led to performance improvements, surpassing many traditional two-stage detectors. The introduction of the YOLO models has further revolutionized the field, with applications across diverse contexts demonstrating remarkable performance relative to their two-stage counterparts (273).

For detecting the apexes of teeth, the proposed YOLO-v9 model exhibited minimal errors. This precision is critical for accurate root length measurements. The YOLO-v9 algorithm represents a significant advancement in the field of object detection, providing real-time performance with remarkable accuracy (156, 273). The recent YOLO-v9 builds upon its predecessors by incorporating enhancements that further improve both accuracy and speed. Notable improvements in YOLOv9 include the integration of advanced backbone networks, which facilitate richer feature representation and enhanced context perception. Additionally, the introduction of PANet (Path Aggregation Network) enables effective feature fusion across different scales, leading to improved localization accuracy. Comparative analysis with previous YOLO versions and other object detection algorithms demonstrates the efficacy of YOLOv9 in achieving a favorable balance between accuracy and speed (156, 274, 275).

A systematic review conducted by Li et al. (268), has shown that various versions of YOLO, from YOLOv3 to YOLOv5, have been utilized for detecting the PBL and classifying the periodontitis stage on various dental images. Consistently, Uzun Saylan et al. (235), utilized a YOLO-v5 algorithm to detect the BL across 685 panoramic radiographs. In addition to general evaluation, models were grouped according to subregions (incisors, canines, premolars, and molars) to provide a targeted assessment. Their findings revealed that the lowest sensitivity and F1 score values were associated with total alveolar BL (0.75 and 0.76, respectively), while the highest values were observed in the maxillary incisor region (1 and 0.95, respectively). They concluded that YOLO architecture has significant potential in analytical studies detecting PBL conditions.

3.4.3 Bone Loss Measurement and Validation

The BL ratio was calculated by dividing the maximum BL by the length of the root, yielding high agreement with clinician assessments. The ICC values were 0.943 for dentist vs. model, 0.954 for dentist vs. periodontist, and 0.959 for periodontist vs. model, indicating strong reliability of the proposed model's measurements. These results underscore the reliability of these models in providing accurate BL measurements, comparable to those of experienced clinicians.

Additionally, the strong ICC values for the coordinates of detected apexes and segmented CEJs, indicating a high level of agreement between both models' predictions and the ground truth values. This demonstrates the accuracy and reliability of the proposed integrated model in detecting these key anatomical landmarks, further validating its potential utility in clinical applications.

3.4.4 Stage and Grade Prediction

The stage and grade prediction results showed varying performance across different categories. For stage III/IV, the model achieved a precision of 1.0 and an F1-score of 0.945, indicating high accuracy for severe cases. However, performance for stage I was poor, with an F1-score of 0.0. This discrepancy is similar to challenges reported in other studies, such as the work by Chang et al. (105), who also found lower accuracy in classifying early stages of periodontitis. However, the results of this chapter emphasize the need for more robust models and larger, more diverse datasets to improve early-stage detection.

Chen et al. (276) also utilized U-net and mask-RCNN models to stage periodontitis using 336 dental PA radiographs. They reported a loss degree deviation of 6.5% between their proposed method and the ground truth drawn by the three periodontists. In addition,

the total diagnostic accuracy of their proposed method was 72.8 % with the highest accuracy for stage III (97.0 %).

In grade prediction, the model performed well for grades B and C, with F1-scores of 0.829 and 0.766, respectively. However, it failed to accurately predict grade A. In this study, the distribution of the test dataset was significantly imbalanced, with the highest frequency of teeth classified as stage III and grade B. While the imbalance in the training dataset was partially mitigated by augmenting the images of minority stages and grades, this issue needs further attention through additional data collection in future research.

The overall metrics for grade prediction were found to be lower compared to those for stage prediction. This finding aligns with the results of Ertas et al. (234), who reported lower clinical accuracy values for classifying the grade of periodontitis using panoramic images (64.5% vs. 88.2%). This could be explained by the fact that the progression of periodontal disease (grade) is influenced by multiple factors. While some individuals may develop severe periodontitis rapidly, others may maintain a mild stage of the disease throughout their lives. Additionally, the progression of periodontitis is less predictable in certain patients, necessitating diverse treatment plans. Established risk factors that accelerate BL include smoking and poorly controlled diabetes, alongside obesity, genetics, physical activity, and nutrition (5, 240).

In contrast to the staging of periodontitis, there are a few studies that determine the grade of periodontal disease solely based on PBL percentage. Shon et al. (261), selected a YOLO-v5 architecture for detecting the teeth objects and classifying them according to the teeth numbering system. The detected tooth numbers were then integrated with the boundaries identified using U-Net to determine the length of the roots and stage of

periodontal disease for each tooth. They reported that the integrated framework had an accuracy of 0.929, with a recall and precision of 0.807 and 0.724, respectively, in average across all four stages.

Similarly, Jiang et al. (246), applied a YOLO-v4 model integrated with U-Net to determine the stages of periodontitis according to the percentage of PBL. After segmenting each tooth using the U-Net model, YOLO-v4 was utilized to detect the 6 key points for each posterior tooth (CEJ mesial and distal, bone level mesial and distal, and root apexes mesial and distal), based on which the PBL% was mathematically calculated on a set of 2560 panoramic radiographs. They concluded that the performance of the model was entirely acceptable, with an overall accuracy of 0.77, although it varied across different teeth.

Comparing our findings with existing literature, several differences and advantages of our approach become evident. In the present chapter, the latest iterations of U-Net and YOLO models (YOLO-v9), were utilized, which have demonstrated superior performance in recent benchmarks. This choice likely contributed to higher accuracy and precision observed in our results, particularly in segmentation and apex detection tasks. The high ICC values and minimal errors in apex detection highlight the precision and reliability of the proposed models, making them suitable for clinical application. These metrics are consistent with or superior to those reported in studies using older model versions or alternative approaches.

Additionally, by integrating U-Net for segmentation and YOLO-v9 for detection, we leveraged the strengths of both models, resulting in a comprehensive approach to diagnosing and grading periodontal disease. This hybrid approach is supported by studies

that have shown improved performance using combined models (105, 267). A novel hybrid framework combining DL architecture demonstrated high accuracy and excellent reliability in the automatic diagnosis of PBL percentage for individual teeth.

The proposed methodology, which aligns well with clinical practices and provides high agreement with clinician assessments, underscores the practical applicability of our models. By using the most recent versions of U-Net and YOLO, this study benefits from the latest advancements in DL, resulting in higher accuracy and precision. In addition to evaluating segmentation and detection performance, the models were validated against clinician assessments, thereby providing a comprehensive evaluation of their clinical utility. This method utilized the percentage rate of PBL to automatically stage and grade periodontitis, following the new criteria proposed at the 2017 World Workshop (2).

3.5 CONCLUSION

This chapter demonstrates the significant potential of utilizing advanced ML models, specifically U-Net and YOLO-v9, for the accurate segmentation and detection of periodontal disease markers on PA images. The results indicate that these models can achieve high precision and reliability in identifying PBL and detecting apexes, crucial for calculating PBL percentages and subsequent disease staging and grading. The comparative analysis with existing literature highlights the superior performance of this approach, particularly due to the use of the latest model iterations.

The high concordance between the proposed model's measurements and clinician assessments underscores the practical utility of these AI tools in real-world clinical settings. By leveraging the strengths of U-Net and YOLO-v9, a comprehensive approach that enhances the accuracy and efficiency of periodontal disease diagnosis and grading is

developed. This method can systematically and precisely assist dental professionals in diagnosing and monitoring periodontitis using PA radiographs. Consequently, it has the potential to significantly enhance dental professionals' performance in the diagnosis and treatment of periodontitis. Future research should aim to expand the dataset size and diversity and refine the models further to improve early-stage detection capabilities.

Chapter 4- Multimodal Deep Learning for Automated Classification of Periodontitis Stage and Grade using Patients' Notes and Periapical Radiographs

4.1 BACKGROUND

Traditionally, healthcare has depended on single-modality approaches, where medical decisions are derived from analyzing one type of data, such as radiology images or clinical records (278). Nevertheless, advancements in technology and the increased availability of diverse data sources have made the integration of multiple modalities increasingly feasible and promising (279, 280). Medical scanners are now capable of producing higher resolution digital images across various modalities, such as MRI and CT (281). Additionally, electronic health records compile a wide range of clinical data in structured formats (282). Furthermore, advanced analytical methods, including DL, are adept at modeling complex multi-modal relationships (283).

Multi-modality approaches involve the fusion and analysis of various data types, encompassing medical images, clinical records, and other relevant sources, to achieve a more comprehensive understanding of patients' conditions (284). Each modality reveals a particular aspect of physiology and pathology, and the effective aggregation and analysis of multimodal data present both unique opportunities and challenges (285). For instance, in diagnosing Alzheimer's disease, relying exclusively on structural MRI scans or speech analysis yields approximately 80% detection accuracy (286). However, by integrating complementary modalities such as audio features, speech transcripts, genomic data, and clinical assessments, multi-modality models have attained diagnostic accuracy exceeding 90% (287, 288).

In the field of dentistry, a recent study conducted by Pfander et al. (289), the authors demonstrated how multimodal learning can assist assembling PA radiographs into an anatomically correct status. Multi-modal DL outperformed both unimodal image-based learning ($p<0.001$) and time-based learning ($p<0.05$). Similarly, Ertas et al. (234) applied traditional ML models to integrate clinical data with features extracted from panoramic radiographs to classify the stage and grade of periodontitis. They achieved the highest clinical accuracy of 0.645 for grade classification using a combination of VGG16 and a support vector machine (SVM), and an accuracy of 0.882 for stage classification using a combination of ResNet50 and SVM.

The current DL models applied in radiology consider only pixel-value information without data collected in patients' charts. However, in practice, accurate and relevant non-imaging data based on the clinical history and examination enable dentists to interpret imaging findings in the appropriate clinical context, which will lead to a higher diagnostic accuracy, informative clinical decision making, and improved patient outcomes (283). Multimodal learning (MML) is a comprehensive approach to developing AI models that can extract and relate information from multimodal data.

In the era of DL, deep neural networks have significantly advanced the development of MML. The transformer architecture, in particular, has emerged as a highly competitive family, presenting new challenges and opportunities for MML. Bidirectional encoder representations from transformers (BERT) (219) and vision transformers (ViT) (203) are state-of-the-art transformer-based models that have shown promising results in the fields of text and image understanding, respectively. BERT (a transformer-based encoder model for language representation) uses a multi-head attention mechanism and a

bidirectional approach to learn the contextual relations between words and sentences in a text, providing an accurate representation of the entire text, and ViT is a seminal work that applies the transformer encoder to image data, offering an end-to-end solution for image understanding (104, 196, 203, 226).

To the authors' knowledge, there have been no studies that integrate PA images of patients with their clinical unstructured notes using the multimodal transformer models for classifying the stage and grade of periodontal disease. Therefore, this chapter aims to introduce a novel method for combining the information derived from PA radiographs with the latent information from clinical notes, including patients' history and clinical findings obtained through examination, to predict and diagnose the stage and grade of periodontitis.

We propose that similar to health-care providers, automated detection and classification systems that can successfully apply both dental imaging data together with clinical data collected as electronic dental records, such as patient demographics, previous diagnoses, clinical examination findings, and laboratory values, may result in better performing and more clinically relevant models.

4.2 MATERIALS AND METHODS

4.2.1 Study population

Two hundred and ten patients' data, including the unstructured clinical notes alongside their corresponding PA radiographs indicating the maximum BL (obtained from the previous chapter) were collected from the Periodontal Graduate Program in the School of Dentistry between 2017-2021. The distribution of periodontitis stages was as follows: 3 samples with stage I, 26 samples with stage II, and 181 samples with stage III/IV. For

grades, the distribution was 8 samples with grade A, 142 samples with grade B, and 59 samples with grade C.

4.2.2 Multimodal Transformer Model

This model was designed to combine text and image data using transformers to classify periodontitis stages and grades into three categories (I, II, and, III/IV or A, B, and C, respectively). It integrated the BERT model for processing text data and the ViT for handling image data.

4.2.2.1 Vision Transformers (ViT) for Image Processing

4.2.2.1.1 Data Preprocessing

To prepare the images for input into the ViT model, a detailed preprocessing pipeline was implemented using the ViTFeatureExtractor from the "transformers" library. Instead of defining a custom preprocess_image function, the feature extractor was employed to ensure consistency with the ViT model's requirements. This extractor performed several essential transformations on the images, including resizing them to 224x224 pixels, which is the input size expected by the ViT model. The images were then normalized to match the distribution of pixel values the model was trained on, ensuring that the inputs are scaled appropriately for optimal performance.

The images were loaded from paths generated based on patient IDs stored in the dataset, ensuring that each image corresponded accurately to its respective patient. To address potential class imbalance in the dataset, class weights were computed using the compute_class_weight function. These class weights, calculated based on the frequency of each class in the dataset, were used to balance the importance of underrepresented classes

during the model's processing, ensuring that the model does not become biased towards more frequent classes. To ensure data integrity, any entries without corresponding images were filtered out before proceeding. The preprocessed images were converted into tensors, stacked into a single tensor, and made ready for input into the ViT model.

4.2.2.1.2 ViT Model

First, the ViT configuration with the correct number of input channels was loaded. This is done using the "ViTConfig.from_pretrained" ('google/vit-base-patch16-224') command, which loads the pre-trained configuration for the ViT model. The ViT model itself was then loaded with the modified configuration using "ViTModel.from_pretrained" ('google/vit-base-patch16-224', config=vit_config). This architecture is pre-trained on ImageNet1K (IN1k) (124) with an input image size of 224 × 224 pixels and output size of Ds = 768D representations. This pre-trained model is capable of understanding and processing image data by leveraging the transformer architecture.

Image paths were generated based on patient IDs, which were stored in the dataset. The images were selected based on the highest amount of BL among all PA radiographs that a patient had. The preprocessing function then loaded each image from the generated paths, applied the transformations, and converted the images into tensors. The preprocessed images were stacked into a single tensor, making them suitable for input into the ViT model.

The ViT model processes images by dividing them into small patches of 16x16 pixels and treats each patch as a token, similar to how words are processed in natural

language models. These patches are passed through multiple layers of the ViT model, where the model learns to focus on different aspects of the image at different layers. The output from the model is a feature representation associated with the [CLS] token, which serves as a summary of the entire image.

These extracted features, which are 768-dimensional vectors, were then converted into a NumPy array and subsequently into a PyTorch tensor, making them ready for further processing. This approach allowed the features from the ViT model to be integrated seamlessly with features extracted from patients' clinical notes, facilitating a comprehensive multimodal analysis. Unlike traditional training processes, the ViT model was used purely for feature extraction, and no classification or further training was performed. This ensures that the extracted features remain consistent with the pre-trained model's capabilities and can be effectively used in combination with other data modalities.

4.2.2.1.3 Feature Visualization

To visualize the feature that the model attended to most, activations from the fifth encoder layer were extracted for analysis, as this intermediate layer captures more nuanced patterns, such as early signs of BL or subtle periodontal changes. Although this was not the primary focus of this thesis, this approach was used to gain a visual understanding of the features the model attended to most during its internal processing.

Principal component analysis (PCA) was employed to reduce the dimensionality of the activation vectors from this layer, condensing the most significant components into a form that could be visually represented as a heatmap. The first principal component was

then mapped back onto the original radiograph to create a heatmap overlay, where the color intensity represents the relative importance of each image region in the model's decision-making process.

4.2.2.2 BERT for Text Processing

4.2.2.2.1 Data Preprocessing

After importing 210 textual patients' notes coupled with their stage and grade levels (as ground truth labels) into the Google Colaboratory (from https://colab.research.google.com/), the texts were preprocessed for computational analysis. In the preprocessing stage, notes were tokenized at both word and sub-word levels using the function, preprocess_text from the "tensorflow_text" package (from: https://www.tensorflow.org/text) in Python. The dataset was carefully cleaned by dropping any rows with missing values to ensure data integrity. Special tokens like [CLS] and [SEP] were added to mark the beginning and end of sentences for the transformer model. Since sentences often vary in length, padding and truncation were applied to standardize the input shape by adding zeros to shorter sentences. It returns the processed data as tensors, which are the input format required by the BERT model.

To address potential class imbalance, class weights were computed using the compute_class_weight function from the "scikit learn" library (232), ensuring that underrepresented classes received appropriate emphasis during training. These weights were later incorporated into the model's loss function.

4.2.2.2.2 BERT Model

After data preprocessing, the BERT model ('bert-base-uncased') was loaded with TFBertModel.from_pretrained. This model, pre-trained on a large corpus of text, is capable of understanding the contextual relationships between words in a sentence. BERT operates on the transformer architecture, which includes 12 layers of self-attention mechanisms and feedforward networks in the “bert-base-uncased” variant. These layers allow BERT to process text bidirectionally, capturing meanings by considering the context of each word from both preceding and following words.

Once the text data is tokenized, it was passed through the BERT model to extract features. This was done using the model's call method, which takes the input IDs and attention masks produced by the tokenizer. The model outputs a tensor representing the hidden states of the tokens in the input sequence. Specifically, the [CLS] token's hidden state (which serves as a summary representation of the input text) was extracted using last_hidden_state [:, 0, :].

The [CLS] token is a special token added at the beginning of the input sequence. The vector corresponding to this [CLS] token is a 768-dimensional vector that contains the aggregated information from the entire sequence. This vector is often used as the summary representation for classification tasks, as it captures the essence of the input text in a compact form.

To address the class imbalance in the dataset, class weights were calculated using the compute_class_weight function from sklearn.utils.class_weight (232). These weights

are inversely proportional to the frequency of each class, ensuring that underrepresented classes are given appropriate importance during training. The class weights were incorporated into the cross-entropy loss function, which was used to train the model.

The BERT model processes text by first tokenizing it into smaller units called tokens (which could be words, subwords, or characters). After tokenization, each token is represented by a vector, and as the tokens pass through the BERT model's layers, their representations are updated based on both the token itself and its surrounding context. Each of these tokens is represented by a hidden state vector of size 768, where 768 is the hidden dimension of the "bert-base" model. These vectors capture various linguistic and contextual features learned by the model, making them rich representations of the text.

The extracted features were initially in a NumPy array format. To prepare these features for further processing and concatenating with those of images, they were converted into PyTorch tensors. This involves first converting the features to a NumPy array with text_features.numpy () and then to a PyTorch tensor with torch.tensor (text_features_np). These tensors, each containing rich semantic information from the input text, are 768-dimensional vectors (one for each token, with the [CLS] token being of primary interest for classification tasks). These vectors are then ready to be combined with the image features extracted by the ViT model.

4.2.2.3 Feature Concatenation and Periodontitis Classification

The ViT model processed the images and extracted features from the CLS token of the final hidden state. Similarly, the BERT model processed the preprocessed text data and

extracted features from the CLS token of its final hidden state. These extracted features represent the high-level information from both modalities. Since both the ViT and BERT models in their base configurations output 768-dimensional vectors, the process of concatenating the features was straightforward. The matching dimensions of 768 from both models allowed for a simple and direct concatenation of the features along the feature dimension.

After extracting features from both text and image data using the BERT and ViT models, the next step involved combining these features and using a neural network for periodontitis stage and grade classification. This process integrates the rich, contextual features from the text with the detailed visual features from the images to create a comprehensive representation for each sample, which is then used for predicting the class labels.

The text and image features, which have been converted into PyTorch tensors, were concatenated along the feature dimension to form a single combined feature tensor. This concatenation integrated the information from both modalities, allowing the model to leverage both textual and visual data for more accurate predictions. The combined features were then converted to a floating-point tensor to ensure compatibility with the subsequent neural network layers.

In neural network (NN) computations, especially within DL frameworks like PyTorch, tensors are typically represented in floating-point format. This format allows the representation of a wide range of values with a high degree of precision, which is essential for the accurate processing and learning capabilities of NN.

A NN was defined to process the combined features extracted from both text and image data. The architecture of this network consisted of several key layers designed to transform, normalize, and ultimately classify the input data. The first layer in the network was a fully connected layer, also known as a dense layer, which contained 256 neurons. This layer's primary function was to transform the high-dimensional combined input features into a more manageable size, facilitating further processing in the subsequent layers.

To prevent overfitting, a dropout layer with a dropout rate of 0.5 was applied. This layer randomly sets a fraction of the neurons to zero during each training step, which helps enhance the model's generalization capability by ensuring that it does not become too reliant on any particular subset of neurons. Following the dropout layer, a batch normalization layer was included to stabilize and accelerate the training process. Batch normalization normalizes the output of the previous layer, reducing internal covariate shift and making the network more robust and efficient.

The final layer of the network was a fully connected output layer with 3 neurons, corresponding to the three classes for classification (A, B, C for periodontitis grade, and I, II, III/IV for periodontitis stage). This layer produces the final classification logits, which are then used to predict the class labels for the input data.

The model was then trained using a stratified 10-fold cross-validation approach to ensure robustness and generalization. Cross-validation is a technique used to maximize the use of limited datasets by iteratively training multiple models on different data portions. K-fold cross-validation is a common approach where the data is divided into K parts. A model is trained on K-1 parts and validated on the remaining part. This process repeats

until all data has been used for validation. The overall model performance is averaged across all folds (290). Because there can be significant variation in the accuracy obtained with different train/test splits, this method provides a more generalizable estimate of model performance than taking just a single split (291).

In this chapter, instead of splitting the data into training and testing groups, the stratified 10-fold cross-validation method was applied. The dataset was split into 10 folds, and the model was trained and validated on each fold. This process was repeated for each fold, and the results were averaged to obtain an overall performance measure. Figure 4.1 provides a detailed illustration of the 10-fold cross-validation method.

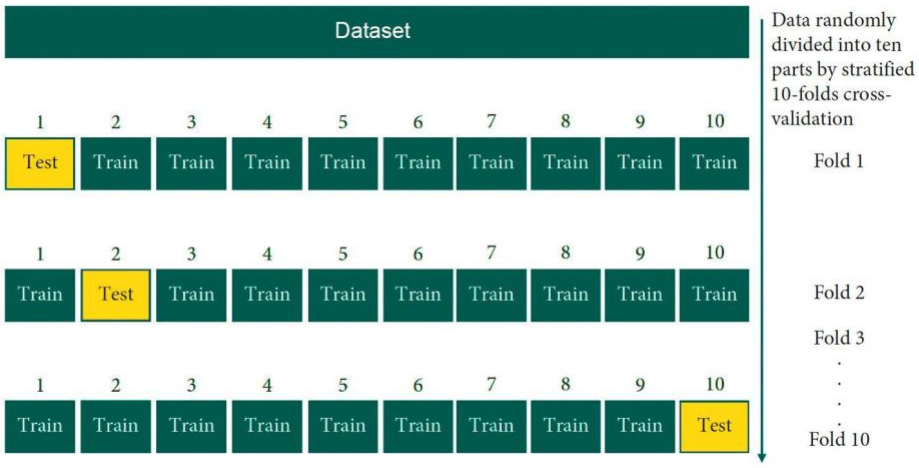


Figure 4.1- Schematic of 10-fold cross-validation method to evaluate the model performance

The loss function used was cross-entropy loss, which is well-suited for multi-class classification tasks. The Adam optimizer was employed, with an adjusted learning rate to balance convergence speed and stability. When the validation loss plateaus, the scheduler reduces the learning rate, helping the model to fine-tune its parameters and achieve better

convergence. To further prevent overfitting, L2 regularization was applied, which adds a penalty term to the loss function based on the magnitude of the model's weights.

The training process was organized into epochs, where each epoch involves training the model on the training data and evaluating its performance on the validation data. To store and analyze the results, the training and validation losses and accuracies for each fold were recorded. Additionally, classification reports and confusion matrices were generated for each fold to provide detailed performance metrics. These metrics help in understanding the model's strengths and weaknesses and guide further improvements.

Finally, the model performance was tested on a set of 21 new, unseen datasets (including texts and images). These datasets included 18 samples in stage III/IV, 2 samples in stage II, and 1 sample in stage I for stage classification. For grade classification, the datasets consisted of 6 samples in grade C, 14 samples in grade B, and 1 sample in grade A.

Figure 4.2 depicts a flowchart outlining the process for extracting features from texts and PA images using transformer models (BERT and ViT, respectively) to classify periodontitis stage and grade. Each box represents a 768-dimensional feature vector corresponding to a specific patch of the PA radiograph, capturing visual characteristics for the ViT model, and a 768-dimensional feature vector corresponding to a specific token (or group of tokens) in the clinical notes, capturing semantic and contextual information for the BERT model.

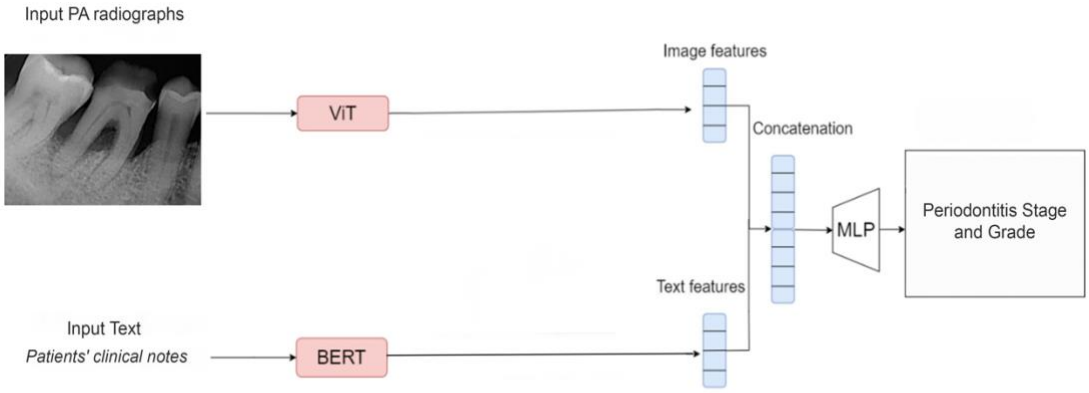


Figure 4.2- Flowchart of the multimodal transformer model used for classifying the stage and grade of periodontitis

4.3 RESULTS

In this chapter, clinical notes obtained from 210 patients, along with their corresponding PA radiographs with the highest amount of BL, were used to train a transformer-based multimodal model to classify the stage and grade of periodontal disease.

Figure 4.3 represents a PA radiograph with a heatmap overlay, generated using PCA applied to the activation vectors from the ViT model, specifically from the output of its fifth encoder layer. This heatmap visually highlights the regions of the radiograph that the model identified as most significant during its internal processing.

In the heatmap overlay, different colors represent varying levels of importance or attention that the ViT model assigned to specific regions of the radiograph. These colors reflect how much weight the model placed on each patch of the image during its analysis. The areas of high activation, depicted in red or orange on the heatmap, represent the regions where the model's focus was strongest, indicating that these are the areas the model deemed most significant for its analysis. In contrast, the yellow or green areas show moderate

activation, meaning the model still finds these regions important but to a lesser extent. These could be indicative of earlier stages of periodontitis or regions adjacent to more affected areas. Lastly, areas with low activation, displayed in blue, are deemed less important by the model for classifying the severity of periodontitis.

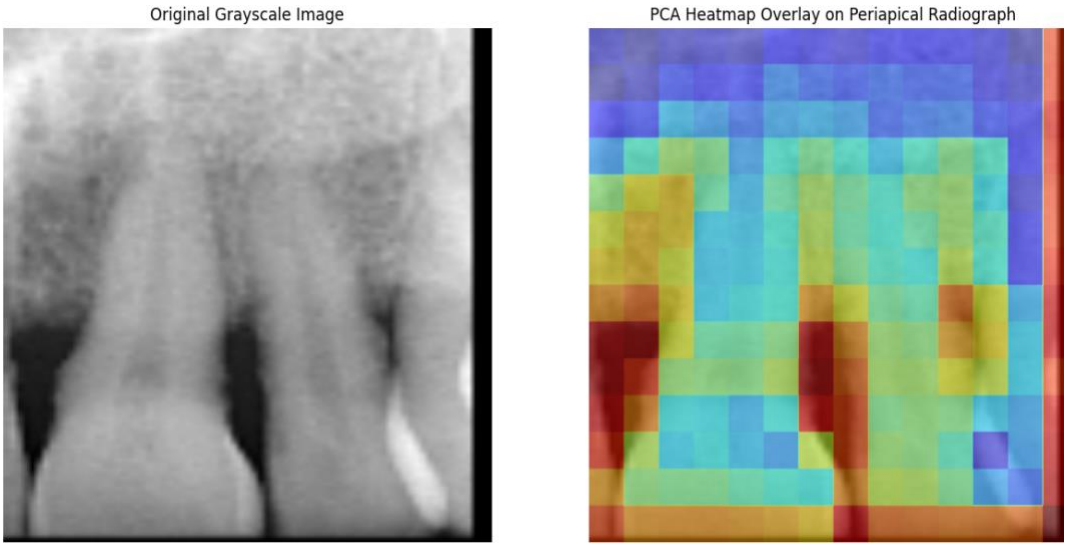


Figure 4.3- Sample PA radiograph and the heatmap showing key regions identified by the ViT model

4.3.1 Multimodal Transformer Model for Prediction of Stage

The text and image features extracted from the ViT and BERT models were concatenated along the specified dimension into a single tensor, which is then used as the input for the subsequent neural network layers for classification. In terms of training accuracy, the model showed a general improvement over the 20 epochs (Figure 4.4). The average accuracy for training and validation was found to be 87.88% and 73.33%, respectively across all 10 folds.

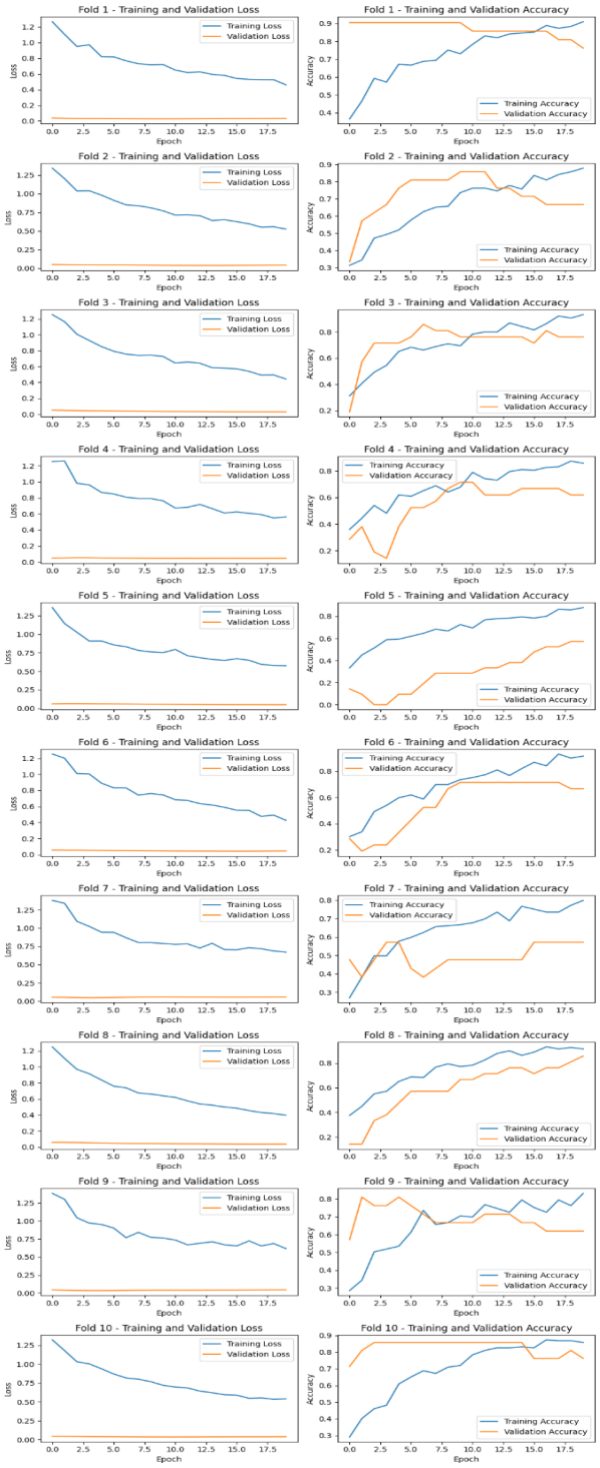


Figure 4.4- Training results for the 10-fold multimodal transformer model for classifying the periodontitis stage

The performance of the final model on a set of 21 new test data (stage I: 1, stage II: 2, and stage III/IV:18) are shown in Table 4.1. Overall, the model demonstrated an accuracy of 0.86, indicating that 86% of the total instances were correctly classified. For stage I, the precision, recall, and F1-score were all 0.00, indicating that the model failed to correctly identify any instances of this stage. This result is likely due to the very small number of instances (only 1) in this class in both training and test dataset, which can make it challenging for the model to learn distinguishing features.

In contrast, the model's performance for stage II was moderately better, achieving a precision and recall of 0.50, resulting in an F1-score of 0.50. This suggests that the model correctly identified half of the instances in this stage. However, the most notable performance was observed in stage III/IV (severe periodontitis), where the model achieved a precision of 0.89, recall of 0.94, and an F1-score of 0.92. This indicates that the model was effective at identifying instances of more severe periodontal disease.

Table 4.1- Multimodal transformer model performance on the test dataset for predicting the periodontitis stage

Stage	Precision	Recall	F1-Score
I	0.00	0.00	0.00
II	0.50	0.50	0.50
III/IV	0.89	0.94	0.92

4.3.2 Multimodal Transformer Model for Prediction of Grade

Similar to the previous step, the text and image features extracted from the ViT and BERT models were concatenated along the specified dimension into a single tensor. This combined tensor was then used as the input for the subsequent neural network layers for classification. Over the course of 20 epochs, the model demonstrated a general improvement in training accuracy (Figure 4.5). The average accuracy for training and validation across all 10 folds was found to be 84.75% and 65.61%, respectively.

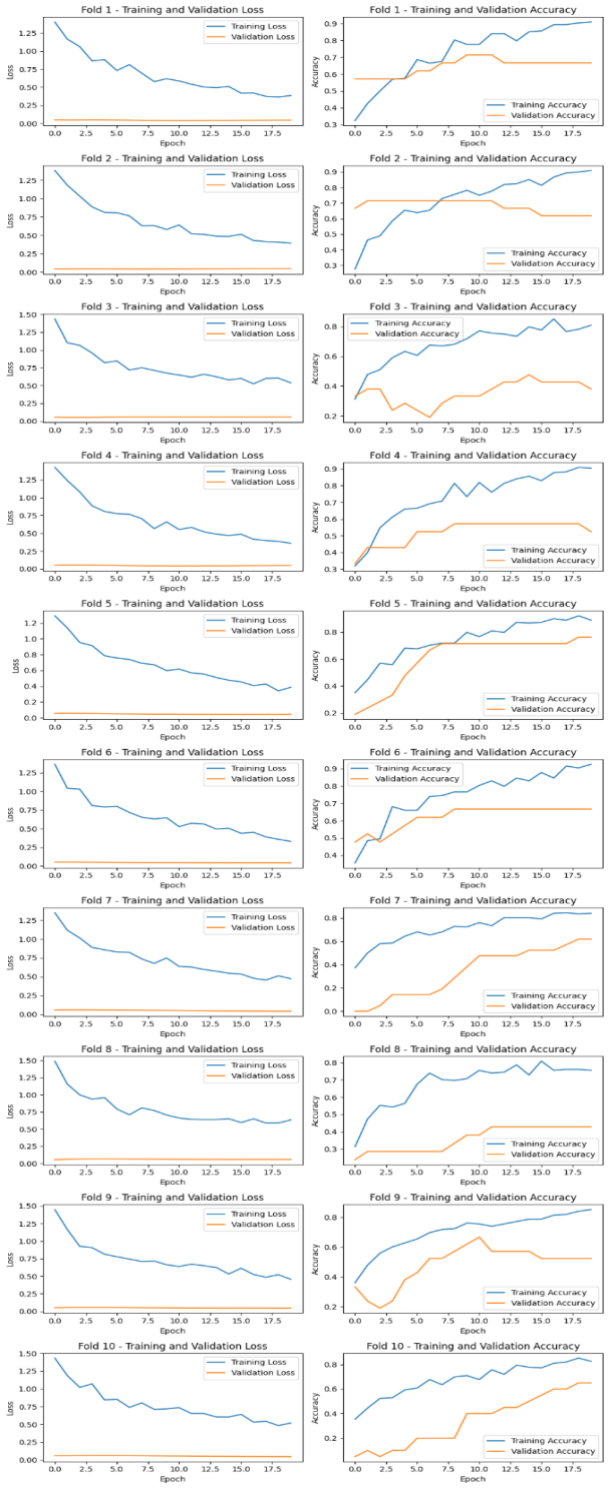


Figure 4.5- Training results for the 10-fold multimodal transformer model for classifying the periodontitis grade

The performance of the final model on a set of 21 new test data (grade A: 1, grade B: 14, and grade C: 6) is shown in Table 4.2. Overall, the model achieved an accuracy of 0.76, meaning it correctly classified 76% of the instances. For grade A, the precision, recall, and F1-score were all 0.00, indicating that the model failed to correctly identify any instances of this grade. This outcome is likely due to the very small number of instances (only 1) in this class in both the training and test datasets, making it difficult for the model to learn distinguishing features.

In contrast, the model's performance for grade C was moderately better, with a precision and recall of 0.67, resulting in an F1-score of 0.67. This indicates that the model correctly identified two third of the instances in this grade. However, the most notable performance was observed in grade B, where the model achieved a precision of 0.80, a recall of 0.86, and an F1-score of 0.83.

Figures 4-4 and 4-5 show consistently low validation loss from the start across all 10 cross-validation folds for both stage and grade classification tasks, which may seem unusual but is likely due to several factors. The small validation set size (only 40 samples per fold) likely causes the model to quickly memorize the validation data, leading to artificially low loss values that do not represent the broader data distribution. Additionally, the validation set may contain simpler or more homogeneous samples than the training data, making it easier for the model to achieve low loss early on, even as training performance improves more slowly. This pattern across all folds suggests the model finds it easier to predict the validation samples, possibly due to an imbalance in data complexity or early overfitting caused by the limited validation set.

Table 4.2- Multimodal transformer model performance on the test dataset for predicting the periodontitis grade

Grade	Precision	Recall	F1-Score
A	0.00	0.00	0.00
B	0.80	0.87	0.83
C	0.67	0.67	0.67

4.4 DISCUSSION

In clinical practice, neither imaging nor clinical data alone are sufficient for the diagnosis of many clinical diseases (292-294). This underscores the importance and utility of multimodal fusion techniques in clinical applications. Multimodal approaches in even less complex clinical applications have also demonstrated improved performance over single-modality models. This includes areas where single-modality models have already achieved high performance, such as pixel-based models for automated skin cancer detection (216). A systematic review by Salvi et al., (190) also highlighted the importance of multimodality approaches in healthcare. Researchers have asserted that integrating multiple data modalities enhances medical decision-making, and that personalized medicine significantly benefits from the combined consideration of imaging, and clinical data.

Multimodal transformers are becoming the leading option in multimodal learning for various tasks (203), such as classification (295, 296), segmentation (297), and cross-

modal retrieval (298). They are proving to be highly effective in enhancing performance on these tasks by utilizing a pre-train-and-transfer approach (298). These models provide several advantages over conventional backbones in terms of flexibility and training load (300). In this chapter, multimodal transformers were used for feature-level fusion to integrate PA radiographs with patients' clinical notes, aiming to predict the stage and grade of periodontitis.

Feature-level fusion offers several significant benefits. This architecture allows for the early integration of information obtained from radiographic images with clinical notes, including patient history, and other relevant clinical indicators. This early fusion enables the model to learn complex correlations between visual features and clinical metrics, leading to more accurate predictions of the stage and grade of periodontitis (301). A systematic review by Huang et al. (283) reported that in the healthcare field, the majority of studies evaluating multimodal approaches (11 out of 17) used early fusion to combine the multimodal input.

In this chapter, a ViT-base architecture was applied for extracting image features from PA radiographs. ViTs divide an image into a series of non-overlapping small blocks, which are analyzed as elements similar to words (203). It has been shown that ViTs outperform other models in certain X-ray, computed tomography (CT), and MRI classification tasks (302-304). Dujic et al. (305) compared the diagnostic performance of five different transformer networks for automated detection of the periodontal BL on 21,819 anonymized PA radiographs. The main differences between the transformer networks were in their size, training strategy and fine-tuning approach. They reported only

minor differences among the tested networks, with overall accuracy ranging from 83.4% to 85.2%. The ViT-base model demonstrated the highest accuracy.

The results showed that integrating patients' clinical notes with their PA radiograph that has the highest percentage of BL from the full-mouth PA series through a multimodal transformer model can achieve an accuracy of 86% and 76% for predicting the stage and grade of periodontitis, respectively. To our knowledge, this is the first study to concatenate patients' clinical notes with their PA radiographs to classify the periodontitis stage and grade using a multimodal transformer (ViT and BERT) model.

Abiyev et al. (306) applied the text and image embedding state-of-the-art models (ViT and BERT) to assess their efficacy in extracting the hidden and distinct features from the surgery video frames of laparoscopy surgeries. These features were then used as inputs for convolution-free transformer architectures to extract comprehensive multidimensional representations. They concluded that the utilized multimodal model represents remarkable performance regarding the recognition of laparoscopic surgical videos.

Lyu et al. (307) proposed a multimodal transformer to fuse time series data from structured electronic health records (EHR) with textual information from clinical notes to improve performance of in-hospital mortality prediction. They applied Clinical BERT (308) as their pre-trained language model to learn clinical specific contextual embeddings because it is a more appropriate domain-specific model, trained on a comprehensive set of medical clinical notes. The results demonstrated that their model outperforms other methods applied for prediction of in-hospital mortality rate with the F1-score of 0.490.

In the oral health field, Zhao et al. (309), developed a multimodal DL method to predict systemic diseases and disorders based on oral health conditions. In the first phase,

a dual-loss autoencoder was employed to extract features related to periodontal disease from 1,188 panoramic radiographs. In the second phase, these image features were fused with demographic data and clinical information from EHR to predict systemic diseases. They also made a comparison among the panoramic-only, EHR data-only, and the combination features to further evaluate whether combining images and EHR data features could enhance model performance. The AUC results showed that the model trained with the combined features outperformed those trained with only one type of feature.

Hsieh and Cheng (310) also evaluated the effectiveness of DL methods and multimodal feature fusion techniques in improving automated dental condition classification. The research utilized a dataset of 11,653 clinically sourced images representing six common dental conditions: caries, calculus, gingivitis, tooth discoloration, ulcers, and hypodontia. Features were extracted using five CNN models and fused into a matrix. Classification models were then constructed using Support Vector Machines (SVM) and Naive Bayes classifiers. The results showed that the SVM classifier integrated with feature fusion demonstrated superior performance, achieving a Kappa index of 0.909 and an accuracy of 0.925. This significantly surpassed the performance of individual CNN models, which achieved a Kappa index of 0.814 and an accuracy of 0.847. The authors concluded that the amalgamation of feature fusion with advanced ML algorithms can significantly bolster the precision and robustness of dental condition classification systems.

The proposed model also demonstrated promising accuracy in predicting the stage and grade of periodontitis. Despite utilizing a sample size of only 210, which is considerably smaller than the sample sizes in previous studies employing multimodal approaches (289, 311), the results were comparable. However, these findings must be

interpreted with caution due to the preliminary nature of this research. However, obtaining a large number of samples is not always feasible in dental studies due to various constraints, including patient availability and the nature of referral practices (70). So, future studies should analyze larger and more balanced datasets using different multimodal techniques to provide additional validation. Additionally, collaborations across multiple dental centers to pool data can help in achieving larger, more balanced datasets.

The accuracy of the model was higher for stages III and grade B compared to stages I or II and grade A, which can be attributed to the imbalanced input data, including more patients with more severe periodontal conditions. This imbalance likely results from the higher number of referrals among patients with severe conditions. Future studies should analyze larger and more balanced datasets using different multimodal techniques to provide additional validation. Similarly, the study by Alawaji et al. (312) found that the majority of patients referred to periodontal clinics (78.4%) have moderate to severe periodontitis, which can explain the fewer number of patients with mild stages of periodontal disease recruited in this study.

4.5 CONCLUSION

The newly proposed model represents a preliminary application of a multimodal transformer model to determine the stage and grade of periodontal disease by combining clinical notes with PA radiographs. Unlike previous models that utilized only radiographic images, text data, and salivary samples, this approach integrates multiple data sources for a more comprehensive analysis. This model can identify the stage and grade of periodontitis without requiring a comprehensive clinical and radiographic examination, which is typically necessary for diagnosis. By streamlining the diagnostic process, this

approach offers a valuable clinical support decision tool for dental professionals, enhancing diagnostic accuracy of periodontitis diagnosis, and ultimately improving patient outcomes. This technique can also be applied by other healthcare providers who need to use large amounts of unstructured text and images to classify disease stages. Future studies should compare various multimodal models to identify the best-performing one for extracting and integrating information from EDR.

Chapter 5- General Discussion and Conclusion

5.1 DISCUSSION

Periodontitis, a multifactorial inflammatory disease affecting the supporting structures of the teeth, presents a significant challenge in dental healthcare due to its potential for causing tooth loss and systemic health complications (2,3,212). Up to 47% of North American adults experience some form of periodontal disease (313,314) and, as the population ages, this prevalence is predicted to increase (315). Accurate and timely diagnosis is critical in mitigating its adverse effects on oral health-related quality of life (OHRQoL) (207,210). Traditional diagnostic methods, which rely heavily on clinical measurements and radiographic analysis, often suffer from variability and interpretation inconsistencies, underscoring the need for advanced diagnostic tools (48).

The increasing adoption of electronic dental records (EDR) has provided an opportunity to utilize artificial intelligence (AI) and deep learning (DL) techniques for improving periodontal diagnostics (69-71, 298). In this dissertation, AI techniques, particularly those involving natural language processing (NLP) and image analysis, have shown promising results in automating the classification of periodontitis stages and grades. This chapter aims to discuss the applied methodologies, results, and implications of using AI to enhance periodontal disease management.

The second chapter focus was on using Bidirectional Encoder Representations from Transformers (BERT), an advanced NLP model (219), to classify periodontitis stages and grades from unstructured clinical notes. BERT's ability to understand and interpret contextual information within text data makes it highly effective for extracting relevant details from medical applications (104,122,227), where patient records often consist of

narrative notes. By extracting meaningful information from these notes, BERT can help clinicians make informed decisions about diseases.

In this dissertation, 309 clinical notes obtained from a cohort of patients were preprocessed and augmented using various augmentation techniques, including synonym replacement, to balance the dataset, ensuring the model was trained on a diverse and representative sample. The BERT model was trained to categorize periodontitis into three and four stages as well as three grades. The results demonstrated that BERT could accurately classify the stage and grade of periodontitis, with training and validation accuracies reaching high percentages. The model's performance was evaluated using new patient charts as a test dataset, showing particularly strong accuracy in classifying more advanced stages of the disease. This highlights the potential of NLP techniques in dental diagnostics, offering a method to leverage the rich information contained in clinical notes to aid in disease classification. A significant strength of Chapter 2 lies in the high accuracy achieved by the BERT model in classifying advanced stages of periodontitis. However, the model showed lower accuracy in classifying earlier stages of the disease, which could be attributed to the fewer samples available for these stages. Furthermore, while BERT effectively extracted critical information from clinical notes, these notes alone may not include essential details about key radiographic changes, such as BL, which are crucial for a comprehensive diagnosis.

Complementing the textual analysis, image analysis plays a crucial role in periodontal diagnostics (316). Automated measurement of periodontal bone loss (BL) using DL models including U-Net and YOLO-v9 in chapter 3 has shown significant potential. Radiographs, specifically peri-apical (PA) images, which are known to be the

standard radiographs for detecting periodontal disease, provide critical information about the extent of disease (317), and accurate interpretation of these images is essential for proper diagnosis and treatment planning. The use of U-Net for segmenting BL on PA radiographs and YOLO-v9 for detecting apexes represents a significant advancement in the automation of image analysis in dentistry.

U-Net, a well-known architecture for biomedical image segmentation (128,318), was employed to segment BL on PA radiographs. This model's architecture, which includes encoder and decoder blocks with convolutional layers and batch normalization (169-171), was optimized to enhance its segmentation accuracy. YOLO-v9, a real-time object detection model (139), was utilized to detect the root apexes essential for calculating BL percentages. The integration of these models provided a comprehensive approach to diagnosing and grading periodontitis. The U-Net model achieved high accuracy in segmenting BL, while the YOLO-v9 model demonstrated precision in detecting apexes, crucial for accurate root length measurements. The results indicated high agreement between the model's measurements and those of experienced clinicians, underscoring the models' reliability in providing consistent BL assessments.

Chapter 3 demonstrated significant advancements in the automation of dental image analysis through the successful application of U-Net and YOLO-v9 models. The models achieved high accuracy in both segmenting BL and detecting root apexes, with results closely matching those of experienced clinicians, which underscores their reliability. However, while radiographic analysis is essential, relying solely on radiographs can be insufficient for assessing the full progression of periodontitis. Radiographs may not fully capture the dynamic progression of the disease or reflect the impact of systemic

factors and patient history that are critical for grading the disease. This limitation highlights the need for integrating radiographic data with clinical notes to achieve a clear view of the patient's condition.

The final chapter was the integration of textual and image data using a multimodal transformer model, which represents a novel approach to periodontal diagnostics. This method utilized both BERT for processing text data and Vision Transformer (ViT) for analyzing image data, offering a comprehensive approach capable of learning complex correlations between different data types, resulting in more accurate and reliable diagnostic outcomes. Multimodal learning (MML) allows for the integration of various data types, providing a thorough understanding of complex health conditions (309,310). By integrating clinical notes with PA radiographs, the multimodal model enhances the accuracy of periodontitis classification.

Chapter 4's integration of BERT and ViT in a multimodal framework represents a cutting-edge approach to periodontal diagnostics. This chapter's strength lies in the significantly higher accuracy achieved through the combination of textual and image data, outperforming the individual modalities. The multimodal approach closely mirrors the comprehensive evaluation process used by periodontists, who typically rely on both clinical notes and radiographic evidence to assess the stage and grade of periodontitis. By combining these data types, the model can more accurately reflect the real-world clinical setting, providing a more complete diagnostic picture. This integration also addresses the limitations identified in the previous chapters, particularly the challenges of relying solely on text or radiographic data. The multimodal model offers a more detailed and precise diagnosis, which is crucial for effective treatment planning.

Feature-level fusion was employed to merge the text and image features, allowing the model to learn complex correlations between clinical and radiographic data (321). This integrated approach achieved high accuracy in classifying the stage and grade of periodontitis, highlighting the importance of combining diverse data types in medical diagnostics. Consistent with previous studies utilizing multimodal approaches in the healthcare area, the multimodal model's performance, evaluated using a dataset of patients with both clinical notes and PA radiographs, demonstrated significant improvements in diagnostic accuracy compared to the two single-modality approaches.

In Chapter 2, the accuracy of predicting periodontitis stage and grade using clinical notes alone was determined to be 79.25% for the three-category stage classification, 66.2% for the four-category stage classification, and 72.3% for grade classification. When using BL percentage and radiographic images alone, the accuracy for stage and grade predictions was 79.5% and 66%, respectively. Both modalities—text analysis and image processing techniques—demonstrated similar accuracies when used independently for stage and grade predictions. The lower accuracy of radiographs in predicting the grade of the disease, compared to patient notes, can be attributed to several factors. The grade of the disease is primarily related to its progression, which may not always be visible on radiographs (2). Additionally, the grade is often influenced by risk factors such as systemic diseases or habits like smoking, which are typically documented in patient notes rather than detectable through imaging (2). Similarly, Machado et al., in their study comparing the accuracy of panoramic radiographs with clinical examinations and periodontal charts for diagnosing periodontitis, reported that imaging modalities do not replace clinical periodontal

evaluation. Instead, they serve as a screening tool to guide patients towards a definitive diagnosis of periodontitis (322).

However, the integration of clinical notes and PA images through a multimodal transformer model demonstrated higher accuracy rates of 86% for stage prediction and 76% for grade prediction. This indicates that the multimodal transformer model, which combines clinical notes and radiographs, is the most effective approach for diagnosing periodontitis. These findings confirm that accurate diagnosis of health-related conditions benefits from a multimodal technique that integrates information from various modalities, as opposed to relying on a single modality (190, 323-326).

Oh et al. (327) conducted a study to identify discrepancies in periodontitis classification among dental practitioners with different educational backgrounds. The study included two cohorts: dental practitioners with periodontal backgrounds (n1 = 31) and those with other educational backgrounds (n2 = 33). The survey instrument featured three periodontitis cases (one with stage III grade C, one with stage II grade B, and one with stage IV grade B), along with guidelines for classification and a questionnaire comprising both closed and open-ended questions. The study evaluated the accuracy of correct classification and the agreement between the two cohorts. The findings revealed a fair level of agreement in periodontitis classification among practitioners from different educational backgrounds. The periodontal cohort demonstrated better accuracy in classifying the stage (71.33%) and grade (64%) compared to the non-periodontal cohort (61.67% and 49.33%, respectively).

In comparison, the results of the single modality approaches (image and text) and the multimodal approach show higher accuracy in determining the stage and grade of

periodontitis than those observed in Oh et al.'s study (327) involving human evaluators. The primary objective in Chapter 3 was to measure BL by segmenting the affected areas on radiographs using U-Net and detecting critical anatomical landmarks like apexes with YOLO-v9. This precise measurement was crucial for understanding the extent of the disease, but the approach remained focused on radiographic data alone. In contrast, Chapter 4 aimed to take the analysis a step further by utilizing ViT to extract the most relevant features from the radiographs with the highest amount of BL. The extracted features were then concatenated with the BERT output from the clinical notes to classify the stage and grade of periodontitis. This shift from U-Net/YOLO to ViT reflects an evolution from simple segmentation and detection tasks to a more sophisticated feature extraction process that integrates both text and image data, ultimately providing a more comprehensive diagnostic tool.

The findings of this dissertation have significant implications for the field of dental diagnostics. The application of AI to automate the classification of periodontitis stages and grades can greatly enhance diagnostic accuracy and efficiency, ultimately leading to improved patient outcomes. By minimizing dependence on subjective interpretations and increasing the consistency of diagnostic assessments, these AI models hold the potential to revolutionize periodontal disease management. Moreover, integrating these techniques into Clinical Decision Support Systems (CDSS) can further augment clinical workflows and decision-making processes, providing real-time, evidence-based recommendations and enhancing overall patient care.

The implications of these findings for clinical practice are substantial. By automating the classification of periodontitis stages and grades, AI models can reduce the

workload on clinicians and minimize the potential for diagnostic errors. The increased accuracy and consistency of these models can lead to more effective treatment planning and better patient outcomes. Furthermore, the integration of AI technology in dental diagnostics can help streamline workflows, making the diagnostic process more efficient and accessible.

The discoveries made in the three chapters of this dissertation underscore the transformative potential of AI in dental diagnostics. Methodologically, the use of BERT for NLP and U-Net and YOLO-v9 for image analysis, followed by the integration of these models in a multimodal transformer framework, represents a comprehensive and innovative approach to periodontitis classification. The results from each chapter highlight the effectiveness of these methodologies, with high accuracies achieved in both single and multimodal approaches.

Despite the challenges posed by data imbalance, the results of this research are clinically useful. The high accuracies observed, especially in advanced stages of periodontitis, suggest that these AI models can provide reliable diagnostic support, even when faced with unbalanced datasets. Clinically, this means that AI can assist in identifying severe cases that require immediate attention, thereby improving patient management and outcomes.

This work can be applied to different contexts by adapting the AI models to other dental or medical conditions that require accurate classification and diagnosis. For instance, the methodologies used in this dissertation can be extended to other inflammatory diseases or conditions detectable through clinical notes and radiographic images. Moreover, the integration of NLP and image analysis in a multimodal framework can be utilized in

broader healthcare settings, providing comprehensive diagnostic support for various diseases, thereby enhancing the overall quality of patient care. By incorporating these insights, the dissertation not only contributes to the field of periodontal diagnostics but also opens new venues for the application of AI in healthcare, demonstrating the value of multimodal approaches in improving diagnostic accuracy and patient outcomes.

5.2 LIMITATIONS

5.2.1 Study 1

The potential limitations of the second chapter, which utilized the BERT-base model for predicting the stage and grade of periodontitis using patients' clinical notes, include several key factors. One major issue is the quality, quantity, and consistency of the data. Clinical notes, written by various residents and confirmed by different periodontists, exhibited significant variation in quality and detail. Inconsistent documentation, use of abbreviations, and variations in terminology could affect the model's performance. However, robust preprocessing pipelines were applied to address many of these issues.

Another limitation is related to the domain-specific language. BERT-base is pretrained on general language corpora and may not fully capture the nuances of dental terminology and specific medical jargon related to periodontitis, potentially impacting prediction accuracy. Clinical notes often contain complex contextual information that can be difficult for the model to fully comprehend, particularly when references to patient history or other implicit information are present. Fine-tuning the BERT model on a more domain-specific corpus could partially mitigate this limitation.

Data imbalance also posed a significant challenge. The dataset was imbalanced, with the majority of training data pertaining to patients in stage III and grade B. An

imbalanced dataset can cause the model to become biased towards predicting more common categories, reducing its effectiveness for rarer conditions. To address this, data was selectively augmented with a higher augmentation rate for minority classes, and class weights were computed during BERT training. Additionally, model performance was analyzed on 210 new patients' notes to assess generalizability, and regularization techniques, such as dropout, were utilized during training to prevent overfitting.

Interpretability of the BERT models is another concern. These models are complex and often considered black boxes, making it difficult to interpret specific predictions. This lack of interpretability can be a limitation in clinical settings, where understanding the rationale behind a prediction is important for decision-making.

Finally, the computational resources required for training and deploying BERT models are significant. This high demand for computational power and resources might not be feasible for all institutions, potentially limiting the widespread application of this approach in clinical practice.

5.2.2 Study 2

In the third chapter of this study, BL was segmented using a U-Net model, and apex coordinates were determined using a YOLO-v9 model, demonstrating excellent consistency with the ground truth. Using this information, the BL percentage was calculated to determine the stage and grade of periodontal disease.

A significant challenge encountered was obtaining an adequate and diverse set of images for training and testing. A robust ML model requires a large dataset that captures the variability in patient anatomy and disease presentation. Without sufficient images, the models may not generalize well to new data, leading to inaccurate predictions and

potentially undermining the study's findings. For this chapter, 1000 PA images were acquired from a private practice in Iran to train the models, and data augmentation techniques were employed to artificially increase the dataset size. The final models were validated on a set of 1582 images obtained from another center, enhancing the generalizability of the results.

Another limitation involves the use of the latest versions of ML models, such as YOLO-v9. While leveraging state-of-the-art models can enhance performance, these models are often computationally intensive and require significant resources for training and deployment. Furthermore, the most recent versions of these models may not have been extensively validated in clinical settings, leading to potential uncertainties in their application to medical imaging tasks.

The study would also benefit from employing cross-validation methods and comparing different segmentation models. Cross-validation ensures that the model's performance is consistently reliable across various subsets of the data, reducing the risk of overfitting. By comparing different segmentation models, researchers can identify the most effective approach for their specific application. However, implementing cross-validation and model comparison requires additional computational resources and time, which may not always be feasible within the study's constraints.

Lastly, the quality of the images and annotations plays a crucial role in the study's success. High-quality images with precise annotations are necessary for training accurate models. Any inconsistencies or errors in the annotations can lead to incorrect predictions and misclassification of periodontal disease stages and grades. To mitigate this limitation, multiple annotators, including dental professionals with expertise in periodontal disease,

were engaged to ensure accurate image annotations. Additionally, preprocessing methods such as Contrast Limited Adaptive Histogram Equalization (CLAHE) were used to standardize image quality and address diversity in image quality.

5.2.3 Study 3

In chapter four, clinical notes alongside PA images with the highest percentage of BL were used to train a transformer multimodal model to predict the stage and grade of periodontal disease. Several potential limitations of this approach need to be considered.

First, the quality and quantity of data are critical factors. Same as the second chapter, clinical notes, written by various residents and confirmed by different periodontists, exhibited significant variation in quality and detail. The limited availability of high-quality, well-annotated datasets can prevent the model from generalizing effectively to new data. Robust preprocessing pipelines were implemented to address many of these issues, but inconsistencies in the data can still impact the model's performance.

Another important aspect is the comparison of various multimodal techniques. While transformer models are powerful, their performance can vary depending on how well they integrate and process different types of data, such as text and images. Exploring and comparing different multimodal techniques could provide insights into the most effective methods for combining clinical notes and PA images. This process requires significant computational resources and time, which may not always be feasible.

Data imbalance is also a significant challenge. The dataset was imbalanced, with a disproportionate number of samples representing certain stages and grades of periodontal disease (majority with stage III and grade B). However, the study by Alawaji et al. (302) found that the majority of patients referred to periodontal clinics (78.4%) have moderate

to severe periodontitis, which can explain the fewer number of patients with mild stages of periodontal disease recruited in this study. This imbalance can lead to biased predictions, where the model is more likely to predict the more common categories, reducing its effectiveness for rarer conditions.

Generalizability is another critical concern. Testing the model on a new, large dataset obtained from other clinics is essential to assess its robustness and applicability in different clinical settings. Differences in patient demographics, clinical note styles, and imaging techniques across various clinics can impact the model's performance. Ensuring that the model generalizes well to these new datasets is crucial for its practical application. However, this requires access to diverse datasets from multiple sources, which can be challenging to obtain.

5.3 RECOMMENDATIONS

Future research should focus on expanding the dataset to include a more diverse patient population and different clinical settings. This would help validate the generalizability of the models and refine their performance further. Additionally, addressing the challenge of imbalanced datasets is crucial for improving the model's early-stage detection capabilities. Techniques such as oversampling, synthetic data generation, and advanced data augmentation methods can help address these issues (328-330).

Furthermore, validation studies across different healthcare centers are necessary to confirm the robustness and applicability of these AI models in varied clinical environments. Collaboration between academic institutions, healthcare providers, and AI researchers can facilitate the development of standardized protocols and datasets, enhancing the reliability and adoption of these technologies in clinical practice.

It is also important to compare various models for both text and image processing to identify the best-performing algorithms. Evaluating different transformer architectures, CNNs, and multimodal fusion techniques can provide insights into which approaches are most effective for integrating clinical notes and PA images. This comparative analysis is essential for optimizing model performance and ensuring the highest accuracy in predicting periodontal disease stages and grades.

5.4 CONCLUSIONS

The integration of AI and DL models in periodontal diagnostics presents a promising avenue for improving the accuracy and efficiency of disease classification. The use of advanced NLP techniques like BERT to analyze clinical notes, coupled with sophisticated image analysis models like U-Net and YOLO-v9, provides a comprehensive approach to understanding and managing periodontitis. The multimodal fusion of textual and image data further enhances diagnostic capabilities, offering a robust framework for future research and clinical application. By leveraging these technologies, the field of dentistry can achieve significant advancements in patient care, ultimately leading to improved oral health outcomes.

The transformative potential of AI in dental diagnostics is evident from the high accuracy and consistency achieved by these models in classifying periodontitis stages and grades. As the field continues to evolve, it is imperative to keep pace with technological advancements and integrate these innovations into routine clinical practice. The success of these AI models in periodontal diagnostics sets the stage for broader applications in other areas of dentistry and healthcare, paving the way for a new era of precision medicine. By embracing these advancements, dental professionals can enhance their diagnostic

capabilities, improve patient outcomes, and contribute to the overall advancement of the field.

References

1- Kinane D.F., Stathopoulou P.G., Papapanou P.N.: Periodontal diseases. Nat Rev Dis Primers 2017; 3: 17038.

2- Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework andproposal of a new classification and case definition. J Periodontol. 2018; 89 Suppl 1:S159-S172.doi: 10.1002/JPER.18-0006.

3- Pihlstrom B. L., Michalowicz B. S., Johnson N. W. Periodontal diseases. Lancet 2005; 366:1809–1820. doi: 10.1016/S0140-6736(05)67728-8

4- Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJL, Marcenes W. Global burden of severe periodontitis in 1990–2010. J Dent Res [Internet]. 2014. November 26 [cited 2018 Oct 11];93(11):1045–53.

5- Tonetti MS, Jepsen S, Jin L, Otomo-Corgel J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. J Clin Periodontol [Internet]. 2017. [cited 2018 Aug 22];44:456–62

6- Kassebaum NJ, Smith AGC, Bernabé E, Fleming TD, Reynolds AE, Vos T, et al. Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990–2015: A systematic analysis for the global burden of diseases, injuries, and risk factors. J Dent Res [Internet]. 2017. April 1 [cited 2018 Aug 14];96(4):380–7.

7- Kaura Parbhakar K, Rosella LC, Singhal S, Quiñonez CR. Risk of complications among diabetics self-reporting oral health status in Canada: A population-based cohort study. PLoS One. 2020;15(1):e0218056. doi: 10.1371/journal.pone.0218056.

8- Summary report on the findings of the oral health component of the Canadian health measures survey 2007–2009 [Internet]. 2010.

9- Genco R.J., Borgnakke W.S.: Risk factors for periodontal disease. Periodontol 2000 2013; 62: pp. 59-94.

10- Simpson T.C., Weldon J.C., Worthington H.V., et. al.: Treatment of periodontal disease for glycaemic control in people with diabetes mellitus. Cochrane Database Syst Rev 2015; CD004714

11- Martens L., De Smet S., Yusof M.Y., et. al.: Association between overweight/obesity and periodontal disease in children and adolescents: a systematic review and meta-analysis. Eur Arch Paediatr Dent 2017; 18: pp. 69-82.

12- Dietrich T., Webb I., Stenhouse L., et. al. Evidence summary: the relationship between oral and cardiovascular disease. Br Dent J 2017; 222: 381-385.

13- Cullinan M.P., Seymour G.J.: Periodontal disease and systemic illness: will the evidence ever be enough?. Periodontol 2000 2013; 62: 271-286.

14- Fischer RG, Lira Junior R, Retamal-Valdes B, Figueiredo LC, Malheiros Z, Stewart B, Feres M. Periodontal disease and its impact on general health in Latin America. Section V: Treatment of periodontitis. Braz Oral Res. 2020; 34(supp1 1):e026. doi: 10.1590/1807-3107bor-2020.vol34.0026.

15- Graziani F, Gennai S, Solini A, Petrini M. A systematic review and meta-analysis of epidemiologic observational evidence on the effect of periodontitis on diabetes. An update of the EFP-AAP review. J Clin Periodontol. 2018;45(2):167-87.

16- Teeuw WJ, Gerdes VE, Loos BG. Effect of periodontal treatment on glycemic control of diabetic patients: a systematic review and meta-analysis. Diabetes Care. 2010;33(2):421-7.

17- Quintero AJ, Chaparro A, Quirynen M, Ramirez V, Prieto D, Morales H, et al. Effect of two periodontal treatment modalities in patients with uncontrolled type 2 diabetes mellitus: A randomized clinical trial. J Clin Periodontol. 2018;45(9):1098-106.

18- Loe H., Theilade E., Jensen S.B.: Experimental gingivitis in man. J Periodontol 1965; 36: pp. 177-187.

19- Schätzle M., Löe H., Lang N.P., et. al.: The clinical course of chronic periodontitis. J Clin Periodontol 2004; 31: 1122-1127.

20- Lang N.P., Schätzle M.A., Löe H.: Gingivitis as a risk factor in periodontal disease. J Clin Periodontol 2009; 36: 3-8.

21- Anerud A., Löe H., Boysen H.: The natural history and clinical course of calculus formation in man. J Clin Periodontol 1991; 18: 160-170.

22- Hajishengallis G., Lamont R.J.: Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. Mol Oral Microbiol 2012; 27: 409-419.

23- Slots J.: Herpesviral-bacterial interactions in periodontal diseases. Periodontol 2000 2010; 52: 117-140.

24- Li F., Zhu C., Deng F.Y., et. al.: Herpesviruses in etiopathogenesis of aggressive periodontitis: a meta-analysis based on case-control studies. PLoS One 2017; 12: e0186373.

25- Vieira A.R., Albandar J.M.: Role of genetic factors in the pathogenesis of aggressive periodontitis. Periodontol 2000 2014; 65: 92-106.

26- Larsson L., Castilho R.M., Giannobile W.V.: Epigenetics and its role in periodontal diseases: a state-of-the-art review. J Periodontol 2015; 86: 556-568.

27- Sonnenschein S.K., Meyle J.: Local inflammatory reactions in patients with diabetes and periodontitis. Periodontol 2000 2015; 69: 221-254.

28- Nociti F.H., Casati M.Z., Duarte P.M.: Current perspective of the impact of smoking on the progression and treatment of periodontitis. Periodontol 2000 2015; 67: 187-210.

29- Camelo-Castillo AJ, Mira A, Pico A, et al. Subgingival microbiota in health compared to periodontitis and the influence of smoking. Front Microbiol. 2015;6:119.

30- Haffajee AD, Socransky SS. Relationship of cigarette smoking to the subgingival microbiota. J Clin Periodontol. 2001;28:377–388.

31- Chigasaki O, Takeuchi Y, Aoki A, et al. A cross-sectional study on the periodontal status and prevalence of red complex periodontal pathogens in a Japanese population. J Oral Sci. 2018;60:293–303.

32- Persson L, Bergström J, Ito H, et al. Tobacco smoking and neutrophil activity in patients with periodontal disease. J Periodontol. 2001;72:90–95.

33- Shivanaikar SS, Faizuddin M, Bhat K. Effect of smoking on neutrophil apoptosis in chronic periodontitis: an immunohistochemical study. Indian J Dent Res. 2013;24:147.

34- White PC, Hirschfeld J, Milward MR, et al. Cigarette smoke modifies neutrophil chemotaxis, neutrophil extracellular trap formation and inflammatory response-related gene expression. J Periodontal Res. 2018;53:525–535.

35- Grossi SG, Genco RJ, Machtei EE, Ho AW, Koch G, Dunford R, Zambon JJ, Hausmann E. Assessment of risk for periodontal disease. II. Risk indicators for alveolar bone loss. J Periodontol. 1995;66(1):23-9. doi: 10.1902/jop.1995.66.1.23. PMID: 7891246.

36- Müller Campanile V, Megally A, Campanile G, et al. Risk factors for recurrence of periodontal disease in patients in maintenance care in a private practice. J Clin Periodontol. 2019;46:918–926.

37- Ryder MI, Couch ET, Chaffee BW. Personalized periodontal treatment for the tobacco- and alcohol-using patient. Periodontol 2000. 2018;78:30–46

38- Emrich LJ, Shlossman M, Genco RJ. Periodontal disease in non-insulin-dependent diabetes mellitus. J Periodontol. 1991;62:123–131.

39- Mealey BL, Oates TW. Diabetes mellitus and periodontal diseases. J Periodontol. 2006;77:1289–1303

40- Zuk A, Quiñonez C, Lebenbaum M, Rosella LC. The association between undiagnosed glycaemic abnormalities and cardiometabolic risk factors with periodontitis: results from 2007–2009 Canadian Health Measures Survey. J Clin Periodontol [Internet]. 2017. February 1 [cited 2019 Apr 30];44(2):132–41.

41- Kwon T, Lamster IB, Levin L. Current Concepts in the Management of Periodontitis. Int Dent J. 2021;71(6):462-476. doi: 10.1111/idj.12630. Epub 2021 Feb 19. PMID: 34839889; PMCID: PMC9275292.

42- Papapanou PN, Sanz M, Buduneli N, et al. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J Periodontol. 2018;89(Suppl 1):S173–S182.

43- Morelli T, Moss KL, Preisser JS, et al. Periodontal profile classes predict periodontal disease progression and tooth loss. J Periodontol. 2018; 89: 148– 156.

44- Lang NP, Suvan JE, Tonetti MS. Risk factor assessment tools for the prevention of periodontitis progression a systematic review. J Clin Periodontol. 2015; 42(Suppl 16): S59– 70

45- Lin PL, Huang PY, Huang PW. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. Comput Methods Programs Biomed. 2017;148:1-11. doi: 10.1016/j.cmpb.2017.06.012.

46- Carranza FA, Newman MG. Carranza's clinical periodontology .9th Ed. Sounders Co; 2002. Chap: 1,2,4,27,45

47- D'Cruz L, Rattan R. Electronic clinical dental records: unintended consequences. Br Dent J. 2018 Apr 27;224(8):582-3. doi: 10.1038/sj.bdj.2018.311.

48- Siqueira R, Anh drade N, Yu SH, Kornman KS, Wang HL. Challenges and decision making for the classification of two complex periodontal cases. Clin Adv Periodontics. 2021 Jun;11(2):103-110. doi: 10.1002/cap.10126. Epub 2020 Nov 4.

49- Kwon T, Levin L. Cause-related therapy: a review and suggested guidelines. Quintessence Int. 2014;45:585–591.

50- Kwon T, Salem DM, Levin L. Nonsurgical periodontal therapy based on the principles of cause-related therapy: rationale and case series. Quintessence Int. 2019;50:370–376

51- Kwon T, Kim DM, Levin L. Successful nonsurgical management of post-orthodontic gingival enlargement with intensive cause-related periodontal therapy. N Y State Dent J. 2015;81:21–23.

52- Kwon T, Wang JCW, Levin L. Home care is therapeutic. Should we use the term "Home-care Therapy" Instead of "Instructions"? Oral Health Prev Dent. 2020;18:397–398

53- Oosterwaal PJ, Matee MI, Mikx FH, et al. The effect of subgingival debridement with hand and ultrasonic instruments on the subgingival microflora. J Clin Periodontol. 1987;14:528–533.

54- Matia JI, Bissada NF, Maybury JE, et al. Efficiency of scaling of the molar furcation area with and without surgical access. Int J Periodontics Restorative Dent. 1986;6:24–35.

55- Zandbergen D, Slot DE, Niederman R, et al. The concomitant administration ofsystemic amoxicillin and metronidazole compared to scaling and root planingalone in treating periodontitis: a systematic review. BMC Oral Health 2016;16:27.

56- Keestra JA, Grosjean I, Coucke W, et al. Non-surgical periodontal therapy withsystemic antibiotics in patients with untreated aggressive periodontitis: a systematicreview and meta-analysis. J Periodontal Res 2015;50(6):689–706.

57- Zhang Z, Zheng Y, Bian X. Clinical effect of azithromycin as an adjunct to nonsurgicaltreatment of chronic periodontitis: a meta-analysis of randomizedcontrolled clinical trials. J Periodontal Res 2016;51(3):275–83.74.

58- Buset SL, Zitzmann NU, Weiger R, et al. Non-surgical periodontal therapy supplementedwith systemically administered azithromycin: a systematic review ofRCTs. Clin Oral Investig 2015;19(8):1763–75.

59- Ramalho VL, Ramalho HJ, Cipullo JP, et al. Comparison of azithromycin and oralhygiene program in the treatment of cyclosporine-induced gingival hyperplasia.Ren Fail 2007;29:265–70.

60- Khattri S, Kumbargere Nagraj S, Arora A, Eachempati P, Kusum CK, Bhat KG, Johnson TM, Lodi G. Adjunctive systemic antimicrobials for the non-surgical treatment of periodontitis. Cochrane Database Syst Rev. 2020 Nov 16;11(11):CD012568. doi: 10.1002/14651858.CD012568.pub2.

61- Rabbani GM, Ash MM, Caffesse RG. The effectiveness of subgingival scaling and root planing in calculus removal. J Periodontol. 1981;52:119–123.

62- Stambaugh RV, Dragoo M, Smith DM, et al. The limits of subgingival scaling. Int J Periodontics Restorative Dent. 1981;1:30–41.

63- Heitz-Mayfield LJ, Lang NP. Surgical and nonsurgical periodontal therapy. Learned and unlearned concepts. Periodontol 2000 2013;62(1):218–31.

64- Mailoa J, Lin GH, Khoshkam V, et al. Long-term effect of four surgical periodontaltherapies and one non-surgical therapy: a systematic review and meta-analysis.J Periodontol 2015;86(10):1150–8.

65- Matarasso M, Iorio-Siciliano V, Blasi A, et al. Enamel matrix derivative and bonegrafts for periodontal regeneration of intra-bony defects. A systematic reviewand meta-analysis. Clin Oral Investig 2015;19(7):1581–93.

66- Cohen RE. Research, Science and Therapy Committee, American Academy of Periodontology. Position paper: periodontal maintenance. J Periodontol. 2003;74:1395–1401.

67- Surveillance Report 2018 – Dental Checks. 2018. Intervals between Oral Health Reviews (2004) NICE Guideline CG19. London: National Institute for Health and Care Excellence (UK)http://www.ncbi.nlm.nih.gov/books/NBK551810/Available from: Accessed January 4, 2020.

68- Kumar S. Evidence-based update on diagnosis and management of gingivitis and periodontitis. Dent Clin North Am. 2019 Jan;63(1):69-81. doi: 10.1016/j.cden.2018.08.005.

69- Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Guly.s B. 3D Deep learning on medical images: A review. Sensors (Basel). 2020;20(18):5097.

70- Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769-774.

71- Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evol Intell. 2021:1-22. doi: 10.1007/s12065-020-00540-3.

72- Kim EG, Oh IS, So JE, Kang J, Le VNT, Tak MK, Lee DW. Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network. J Clin Med. 2021;10(22):5400.

73- Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 2017, 19, 221–248, doi:10.1146/annurev-bioeng-071516-044442.

74- Mehta S, Suhail Y, Nelson J, Upadhyay M. Artificial intelligence for radiographic image analysis. Seminars in Orthodontics 2021. 27. 10.1053/j.sodo.2021.05.007

75- Cohen AM, Ip HH, Linney AD. A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. Br J Orthod. 1984;11:143–154

76- Pohlheim H. Evolutionary algorithms: overview, methods and operators. GEATbx Version. 2006;3(8).

77- Khurana, D., Koli, A., Khatter, K. et al. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl 82, 3713–3744 (2023). https://doi.org/10.1007/s11042-022-13428-4

78- Cambria E, White B. Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag 2014;9(2):48–57.

79- Zhang H, Shafiq MO. Survey of transformers and towards ensemble learning using transformers for natural language processing. J Big Data. 2024;11(1):25. doi: 10.1186/s40537-023-00842-0. Epub 2024 Feb 4.

80- Bobba PS, Sailer A, Pruneski JA, Beck S, Mozayan A, Mozayan S, Arango J, Cohan A, Chheang S. Natural language processing in radiology: Clinical applications and future directions. Clin Imaging. 2023;97:55-61. doi: 10.1016/j.clinimag.2023.02.014.

81- Cheng LT, Zheng J, Savova GK, Erickson BJ. Discerning tumor status from unstructured MRI reports--completeness of information in existing reports and

utility of automated natural language processing. J Digit Imaging. 2010 Apr;23(2):119-32. doi: 10.1007/s10278-009-9215-7.

82- Sommer KK, Marschollek M, Gietzelt M. Programming techniques for improving rule readability for rule-based information extraction natural language processing pipelines of unstructured and semi-structured medical texts. Health Informatics J. 2023 Apr-Jun;29(2):14604582231164696. doi: 10.1177/14604582231164696.

83- Bao XY, Huang WJ, Zhang K, Jin M, Li Y, Niu CZ. [A customized method for information extraction from unstructured text data in the electronic medical records]. Beijing Da Xue Xue Bao Yi Xue Ban. 2018 Apr 18;50(2):256-263. Chinese.

84- Yao L, Mao C, Luo Y. Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. BMC Med Inform Decis Mak 19 (Suppl 3), 71 (2019). https://doi.org/10.1186/s12911-019-0781-4

85- Berge GT, Granmo OC, Tveit TO, Ruthjersen AL, Sharma J. Combining unsupervised, supervised and rule-based learning: the case of detecting patient allergies in electronic health records. BMC Med Inform Decis Mak. 2023 Sep 18;23(1):188. doi: 10.1186/s12911-023-02271-8.

86- Alfattni G, Peek N, Nenadic G. Attention-based bidirectional long short-term memory networks for extracting temporal relationships from clinical discharge summaries. J Biomed Inform. 2021; 123:103915. doi: 10.1016/j.jbi.2021.103915.

87- Xu X, Cai H, Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure, Advanced

Engineering Informatics 2021; 48, 101288, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2021.101288.

88- Erwig, Martin & Gopinath, Rahul. Explanations for regular expressions. 2012; 7212. 394-408. 10.1007/978-3-642-28872-2_27.

89- Montejo-Ráez A, Jiménez-Zafra SM. Current Approaches and Applications in Natural Language Processing. Applied Sciences. 2022; 12(10):4859. https://doi.org/10.3390/app12104859

90- Wright A, McCoy AB, Henkin S, Kale A, Sittig DF. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. J Am Med Inform Assoc. 2013;20(5):887-90. doi: 10.1136/amiajnl-2012-001576. Epub 2013 Mar 30.

91- Sarker, I.H. Machine learning: algorithms, real-world applications and research directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

92- Ozgur A. Supervised and unsupervised machine learning techniques for text document categorization. 2004.

93- Li C, Weng Y, Zhang Y, Wang B. A systematic review of application progress on machine learning-based natural language processing in breast cancer over the past 5 years. Diagnostics (Basel). 2023;13(3):537. doi: 10.3390/diagnostics13030537.

94- Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVylder J, Walter M, Berrouiguet S, Lemey C. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res.

2021;23(5):e15708. doi: 10.2196/15708. PMID: 33944788; PMCID: PMC8132982.

95- Pillai, A.S., Tedesco, R. (Eds.). Machine learning and deep learning in natural language processing (1st ed.). CRC Press. 2023 https://doi.org/10.1201/9781003296126

96- Yin W, Kann K, Yu M, Schütze H. Comparative study of CNN and RNN for Natural Language Processing. 2017

97- Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference. 1. 10.3115/v1/P14-1062.

98- Wen Y, Zhang W, Luo R, Wang J. Learning text representation using recurrent convolutional neural network with highway layers. 2016.

99- Lu Y, Salem F. Simplified gating in long short-term memory (LSTM) recurrent neural networks. 2017: 1601-1604. 10.1109/MWSCAS.2017.8053244.

100- Lu, H., Ehwerhemuepha, L., Rakovski, C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. BMC Med Res Methodol 22, 181 (2022). https://doi.org/10.1186/s12874-022-01665-y

101- Rezaeenour J, Ahmadi M, Jelodar H, Shahrooei R. Systematic review of content analysis algorithms based on deep neural networks. Multimed Tools Appl. 2023;82(12):17879-17903. doi: 10.1007/s11042-022-14043-z.

102- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems, 2017; 30.

103- Li Jiayin. The evolution, applications, and future prospects of large language models: An in-depth overview. Applied and Computational Engineering. 2024; 35. 234-244. 10.54254/2755-2721/35/20230399.

104- Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018 arXiv preprint arXiv:1810.04805.

105- Lappin, S. Assessing the Strengths and Weaknesses of Large Language Models. J of Log Lang and Inf 33, 9–20 (2024). https://doi.org/10.1007/s10849-023-09409-x

106- Lei S, Yi W, Ying C, Ruibin W. Review of attention mechanism in natural language processing. Data analysis and knowledge discovery, 2020; 4(5), 1-14.

107- Shazeer N. Fast transformer decoding: One write-head is all you need. arXiv preprintar 2019 Xiv: 1911.02150.

108- Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 2017;8(1):171–182. doi: 10.1007/s13244-016-0534-1.

109- Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014;23(9):727–731. doi: 10.1136/bmjqs-2013-002627.

110- Lee Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106–111. doi: 10.1016/j.jdent.2018.07.015.

111- Lanning SK, Best AM, Temple HJ, Richards PS, Carey A, McCauley LK. Accuracy and consistency of radiographic interpretation among clinical instructors using two viewing systems. Journal of Dental Education, 2006; 70: 149-159. https://doi.org/10.1002/j.0022-0337.2006.70.2.tb04071.x

112- Tsoromokos N, Parinussa S, Claessen F, Moin DA, Loos BG. Estimation of alveolar bone loss in periodontitis using machine learning. Int Dent J. 2022;72(5):621-627. doi: 10.1016/j.identj.2022.02.009.

113- Hoi SCH, Jin R., Zhu J, Lyu M.R. Batch mode active learning and its application to medical image classification. In Proceedings of the ACM International Conference Proceeding Series; New York, 2006; 148:417–424.

114- Rahman MM, Bhattacharya P, Desai BC. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans. Inf. Technol. Biomed. 2007, 11, 78–69, doi:10.1109/TITB.2006.884364.

115- Wernick M, Yang Y, Brankov J, Yourganov G, Strother S. Machine learning in medical imaging. IEEE Signal Process. Mag. 2010, 27, 25–38, doi:10.1109/MSP.2010.936730.

116- LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–444.

117- Szeliski R. Computer Vision: Algorithms and applications. Springer Science & Business Media; 2010. 824 p.

118- Nebauer C. Evaluation of convolutional neural networks for visual recognition. IEEE Trans Neural Netw. 1998;9:685–696.

119- Arik SO, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging. 2017;4: 014501.

120- Yang X, Wu N, Cheng G, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and neck radiation therapy. Int J Radiat Oncol Biol Phys. 2014;90:1225–1233.

121- Long JL, Zhang N, Darrell T. Do convnets learn correspondence? Adv Neural Inf Process Syst.2014.

122- Huang M, Huang S, Zhang Y, Bhatti U. Medical Image Segmentation Using Deep Learning with Feature Enhancement. IET Image Processing. 2020. 14. 10.1049/iet-ipr.2019.0772

123- Schlempe J, Oktay O, Schaap M, Heinrich M, Kainz M, Glocker B, et al. Attention gated networks: learning to leverage salient regions in medical images. Med Image Anal 2019; 12 (53): 197–207

124- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015; 10 (115): 211–252

125- Schwendicke F, Golla, T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: A scoping review. J. Dent. 2019; 91, 103226.

126- Mutasa S, Sun S, Ha R. Understanding artificial intelligence-based radiology studies: CNN architecture. Clin Imaging. 2021; 80: 72-76. doi: 10.1016/j.clinimag.2021.06.033.

127- Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput. 2017; 29(9): 2352-2449. doi: 10.1162/NECO_a_00990.

128- Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Int. Conf. on Medical Image Computing & Computer-assisted Intervention, Munich, Germany 2015; 3(9351): 234–241

129- Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol. 2017;208(4):754–760.

130- Khened M, Kollerathu VA, Krishnamurthi, G. Fully convolutional multiscale residual dense nets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med Image Anal 2018; 2(51): 21–45

131- Drozdzal M, Chartrand G, Vorontsov E, et al. Learning normalized inputs for iterative estimation in medical image segmentation. Med Image Anal 2017; 10(44): 1–13

132- Ariji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y, et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. Oral Surg Oral Med Oral Pathol Oral Radiol 2019; 128: 424–30. doi: 10.1016/j.oooo.2019.05.014

133- Celik B, Celik ME. Automated detection of dental restorations using deep learning on panoramic radiographs. Dentomaxillofac Radiol. 2022;51(8):20220244. doi: 10.1259/dmfr.20220244. Epub 2022 Sep 12.

134- Girshick R, Donahue J, Darrell T, Malik J. Rich feature Hierarchies for accurate object detection and semantic segmentation. In: Malik J, ed. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Columbus, OH, USA. 2016. doi: https://doiorg. login.ezproxy.library.ualberta.ca/10.1109/CVPR.2014.81

135- Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017; 39: 1137–49.doi: https://doiorg. login. ezproxy. library. ualberta. ca/ 10.1109 /TPAMI.2016.2577031

136- Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. 2018.

137- Leibe B, Matas J, Sebe N, Welling M. Computer Vision – ECCV 2016. In: SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016. Cham: Springer International Publishing; 2016. https://doi-org. login. ezproxy. library. ualberta. ca/10.1007/978-3-319-46448-0

138- Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. In: Doll.r P, ed. 2017 IEEE International Conference on Computer Vision (ICCV); Venice 2016. https://doi-org.login.ezproxy. library.ualberta.ca/ 10.1109/ICCV.2017.324

139- Astuti ER, Putra RH, Putri DK, Ramadhani NF, Noor TNEBT. A., Putra BR, et al. The sensitivity and specificity of YOLO V4 for tooth detection on panoramic radiographs. J Int Dent Med Res 2023; 16(1): 442-446.

140- Redmon, J, Farhadi, A. YOLO9000: Better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 21-26 July 2017, 7263-7271. https://doi.org/10.1109/CVPR.2017.690

141- Liang T, Chu X, Liu Y, Wang Y, Tang Z, Chu W, et al. CB- NetV2: A composite backbone network architecture for object detection. arXiv preprint arXiv:2107.00420, 2021.

142- Li Y, Mao H, Girshick R, He K. Exploring plain vision transformer backbones for object detection. arXiv preprint arXiv:2203.16527, 2022

143- Tan M, Pang R, Le QV. Efficient- Det: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pat- tern Recognition (CVPR), pages 10781–10790, 2020.

144- Tian Z, Shen C, Chen H, He T. FCOS: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 9627–9636, 2019

145- Tian Z, Shen C, Chen H, He T. FCOS: A simple and strong anchor-free object detector. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 44(4):1922–1933, 2022.

146- Oksuz K, Cam BC, Akbas E, Kalkan S. A ranking-based, balanced loss function unifying classification and localisation in object detection. Advances in Neural Information Processing Systems (NeurIPS), 33:15534–15545, 2020.

147- Oksuz K, Cam BC, Akbas E, Kalkan S. Rank & sort loss for object detection and in- stance segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 3009–3018, 2021.

148- Wang J, Song L, Li Z, Sun H, Sun J, Zheng N. End-to-end object detection with a fully convolutional network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15849–15858, 2021

149- Zhu B, Wang J, Jiang Z, Zong F, Liu S, Li Z, Sun J. AutoAssign: Differ- entiable label assignment for dense object detection. arXiv preprint arXiv: 2007.03496, 2020.

150- Wang CY, Bochkovskiy A, Liao H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2022; 10.48550/arXiv.2207.02696.

151- Du J. 2018 J. Phys.: Conf. Ser. 1004 012029

152- Demetriou D, Mavromatidis P, Robert PM, Papadopoulos H, Petrou MF, Nicolaides D. Realtime construction demolition waste detection using state-of-the-

art deep learning methods; single-stage vs two-stage detectors. Waste Manag. 2023;167:194-203. doi: 10.1016/j.wasman.2023.05.039.

153- Son DM, Yoon YA, Kwon HJ, An CH, Lee SH. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. Diagnostics (Basel) 2021;11(6):933. doi: 10.3390/diagnostics11060933.

154- Kaya E, Gunec HG, Gokyay SS, Kutal S, Gulum S, Ates HF. Proposing a CNN method for primary and permanent tooth detection and enumeration on pediatric dental radiographs. J Clin Pediatr Dent 2022; 46(4):293-298. doi: 10.22514/1053-4625-46.4.6.

155- Yang H, Jo E, Kim HJ, Cha IH, Jung YS, Nam W, Kim JY, Kim JK, Kim YH, Oh TG, Han SS, Kim H, Kim D. Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. J Clin Med. 2020;9(6):1839. doi: 10.3390/jcm9061839.

156- Wang CY, Yeh IH, Liao HYM., 2024. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv:2402.13616.

157- Suetens P, Bellon E, Vandermeulen D, Smet M, Marchal G, Nuyts J, Mortelmans L. Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. Eur J Radiol. 1993;17:14–21

158- Mishra P, Nikhil V, Rathi S, Tyagi S. Tooth segmentation in dentistry: a review. International Journal of Research and Review. 2023; 10(11): 319-324. DOI: https://doi.org/10.52403/ijrr.20231137

159- Verykokou S, loannidis C, Angelopoulos C. Evaluation of 3D modeling workflows using dental CBCT data for periodontal regenerative treatment. J Pers Med 2022;12(9):1355-66.

160- Na SD, Lee G, Lee JH, Kim MN. Individual tooth region segmentation using modified watershed algorithm with morphological characteristic. Mater Eng 2014;24(6):303-9.

161- Kumar AS.Analysis of dental segmentation technique for dental radiographs. Elementary education online;202120(4):3868-75.

162- Alsheghri A, Ghadiri F, Zhang Y, Lessard O, Keren J, Cheriet F, Guibault F. Semi- supervised segmentation of tooth from 3D scanned dental arches. Int Med Imag 2022; 12032:766-71

163- Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille A. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell. 2017;40(4):834–48.

164- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440. 2015.

165- He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969. 2017.

166- Qiu B, Wel H, Kareema J, Glas HH. Automatic segmentation of mandible from conventional methods to deep learning—a review. J Pers Med 2021;11(7):629-38.

167- Widyaningrum R, Candradewi I, Aji NRAS, Aulianisa R. Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. Imaging Sci Dent. 2022;52(4):383-391. doi: 10.5624/isd.20220105.

168- Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D. DRINet for Medical Image Segmentation. IEEE Trans Med Imaging. 2018;37(11):2453-2462. doi: 10.1109/TMI.2018.2835303.

169- Bayrakdar IS, Orhan K, .elik ., Bilgir E, Sağlam H, Kaplan FA, G.rür SA, Odabaş A, Aslan AF, R.żyło-Kalinowska I. A U-Net approach to apical lesion segmentation on panoramic radiographs. Biomed Res Int. 2022;2022:7035367. doi: 10.1155/2022/7035367.

170- Gadosey PK, Li Y, Adjei Agyekum E, Zhang T, Liu Z, Yamak PT, et al. SD-UNet: stripping down U-Net for segmentation of biomedical images on platforms with low computational budgets. Diagnostics (Basel) 2020; 10: 110.

171- Zhou Z, Siddiquee MM, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 2020; 39: 1856-67.

172- Hou S, Zhou T, Liu Y, Dang P, Lu H, Shi H. Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement. Comput Biol Med. 2023;152:106296. doi: 10.1016/j.compbiomed.2022.106296.

173- Kim J, Lee HS, Song IS, et al. DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. Sci Rep 2019;9:1e9.

174- Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. Sci Rep 2020;10:1e8.

175- Lee JH, Kim Dh, Jeong SN, et al. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci 2018; 48: 114e23.

176- Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, D.rfer C, Schwendicke F. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 2019 Jun 11;9(1):8495. doi: 10.1038/s41598-019-44839-3.

177- Chen CC, Wu YF, Aung LM, Lin JC, Ngo ST, Su JN, Lin YM, Chang WJ. Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. J Dent Sci. 2023; 18(3) :1301-1309. doi: 10.1016/j.jds.2023.03.020

178- Huang Y, Wu Z, Wang L, Member S, Tan T. Feature coding in image classification: A comprehensive study. 2013: 1–15

179- Sarraf A, Azhdari M, Sarraf S. A comprehensive review of deep learning architectures for computer vision applications. Technol. Sci. Am. Sci. Res. J. Eng. 2021, 77(1): 1–29 [Online]. Available: http://asrjetsjournal.org/.

180- Chithra K, Santhanam T. A novel denoising technique for mixed noise removal from grayscale and color images. J. Theor. Appl. Inf. Technol., 2018; 96(3): 626–642, 2018.

181- Helms RM. Introduction to image technology. IBM Syst. J. 2010; 29(3): 313–332, doi: 10.1147/sj.293.0313.

182- Lorente O, Riera I, Rana A. Image classification with classic and deep learning techniques. 2021, [Online]. Available: http://arxiv.org/abs/2105.04895.

183- Cheng G, Guo L, Zhao T, Han J, Li H, Fang J. Automatic landslide detection from remote-sensing imagery using a scene classification method based on boVW and pLSA. Int. J. Remote Sens. 2013; 34(1): 45–59, 2013, doi: 10.1080/01431161.2012.705443.

184- Aya M, Ashraf H, Moatamad H. Image classification based deep learning: A Review. Aswan University Journal of Sciences and Technology 2022. 2. 10.21608/aujst.2022.259887.

185- Lashari SA, Ibrahim R. A framework for medical images classification using soft set. 2013 Procedia Technol;11: 548–556, doi: 10.1016/j.protcy.2013.12.227.

186- Person M, Jensen M, Smith AO, Gutierrez H. Multimodal fusion object detection system for autonomous vehicles. J. Dyn. Syst. Meas. Control. 2019;141:071017. doi: 10.1115/1.4043222.

187- Trzcinski T. Multimodal social media video classification with deep neural networks. In Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018 (eds. Romaniuk, R. S. & Linczuk, M.) (SPIE, 2018)

188- Cui C, et al. Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review. Progress Biomed. Eng. 2023, 5 (2), 10.1088/2516-1091/acc2fe

189- Dhanesh R, Graham T. Deep Multimodal Learning: A Survey on Recent Advances and Trends. IEEE Signal Processing Magazine. 2017; 34. 96-108. 10.1109/MSP.2017.2738401.

190- Salvi M, Loh HW, Seoni S, Barua PD, García S, Molinari F, Acharya UR. Multi-modality approaches for medical support systems: A systematic review of the last decade, Information Fusion 2024;103:102134, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.102134.

191- Kiela D, Grave E, Joulin, A, Mikolov T. Efficient large-scale multi-modal classification. In The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), (2018).

192- Kini AS, et al. Ensemble deep learning and internet of things-based automated COVID-19 diagnosis framework Contrast Media Mol. Imaging, 2022 (2022), p. 7377502, 10.1155/2022/7377502

193- ur Rehman A, Belhaouari SB, Kabir MA, Khan A. On the use of deep learning for video classification. Applied Sciences. 2023; 13(3):2007. https://doi.org/10.3390/app13032007

194- Muhammad K, Ullah A, Lloret J, Ser JD, de Albuquerque VHC. Deep learning for safe autonomous driving: current challenges and future directions, in IEEE Transactions on Intelligent Transportation Systems 2021; 22(7): 4316-4336, doi: 10.1109/TITS.2020.3032227.

195- Baltruˇsaitis T, Ahuja C, Morency LP. Multimodal machinelearning: A survey and taxonomy," TPAMI, 2018.

196- Xu P, Zhu X, Clifton D. Multimodal learning with transformers: A survey. 2022. 10.48550/arXiv.2206.06488.

197- Tsimpoukelli M, Menick J, Cabi S, Eslami S, Vinyals O, Hill F. Multimodal few-shot learning with frozen languagemodels, NeurIPS, 2021.

198- Sung YL, Cho J, Bansal M. Vl-adapter: Parameter efficient transfer learning for vision-and-language tasks," inCVPR, 2022.

199- Alayrac JB, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Lenc K, Mensch A, Millican K, Reynolds M, et al. Flamingo: avisual language model for few-shot learning, NeurIPS, 2022.

200- Chen X, Wang X, Changpinyo S, Piergiovanni A, Padlewski P, D Salz S. Goodman A. Grycner B. Mustafa, et al. Pali:A jointly-scaled multilingual language-image model," arXiv,2022.

201- Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018.

202- Chen M, Radford A, Child R, Wu J, Jun H, Luan D, Sutskever I. Generative pretraining from pixels. in ICML, 2020.

203- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani, M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale," arXiv, 2020.

204- Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, J´egou H. Training data-efficient image transformers & distillation through attention. in ICML, 2021.

205- Beal J, Kim E, Tzeng E, Park DH, Zhai A, Kislyuk D. Toward transformer-based object detection. arXiv, 2020.

206- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows," arXiv, 2021.

207- Borges T de F, Regalo SC, Taba Jr M, Siéssere S, Mestriner W Jr, Semprini M. Changes in masticatory performance and quality of life in individuals with chronic periodontitis. J Periodontol; 2013 (84): 325-331. doi: 10.1902/jop.2012.120069. Epub 2012 May 1. PMID: 22548588

208- Preshaw P, Alba A, Herrera D, Jepsen S, Konstantinidis A, Makrilakis K, et al. Periodontitis and diabetes: a two-way relationship. Diabetologia; 2012 (55): 21–31. doi: 10.1007/s00125-011-2342-y. PMID: 22057194

209- Araújo VMA, Melo IM, Lima V. Relationship between periodontitis and rheumatoid arthritis: review of the literature. Mediators Inflamm; 2015; 2015:259074. doi: 10.1155/2015/259074. PMID: 26347200

210- Graziani F, Music L, Bozic D, Tsakos G. Is periodontitis and its treatment capable of changing the quality of life of a patient? Br Dent J; 2019 (227): 621-625. PMID: 31605074

211- Kim EH, Kim S, Kim HJ, Jeong HO, Lee J, Jang J, et al. Prediction of chronic periodontitis severity using machine learning models based on salivary bacterial copy number. Front Cell Infect Microbiol; 2020 (10): 571515. doi: 10.3389/fcimb.2020.571515. PMCID: PMC7701273

212- Monsarrat P, Bernard D, Marty M, Cecchin-Albertoni C, Doumard E, Gez L, et al. Systemic periodontal risk score using an innovative machine learning strategy: An observational study. J Pers Med; 2022 12(2): 217. doi: 10.3390/jpm12020217. PMID: 35207705

213- Bertoldi C, Forabosco A, Lalla M, Generali L, Zaffe D, Cortellini P. How intraday index changes influence periodontal assessment: A preliminary study. International Journal of Dentistry; 2017: 1-10. Retrieved from https://doi.org/10.1155/2017/7912158. PMID: 28828006

214- Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med; 2021 13(1): 152. doi: 10.1186/s13073-021-00968-x. PMID: 34579788

215- Walczak S., Velanovich V. Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks. Decision Support Systems; 2018 (106): 110-118

216- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature; 2017 542(7639): 115-118. doi: 10.1038/nature21056. PMID: 28117445

217- Fei Y, Li WQ. Improve artificial neural network for medical analysis, diagnosis and prediction. J Crit Care; 2017 (40): 293. doi: 10.1016/j.jcrc.2017.06.012. PMID: 28647140.

218- Ameli N, Gibson MP, Khanna A, Howey M and Lai H. An application of machine learning techniques to analyze patient information to improve oral health outcomes. Front. Dent. Med; 2022 (3): 833191. doi: 10.3389/fdmed.2022.833191

219- Mitchell JR, Szepietowski P, Howard R, Reisman P, Jones JD, Lewis P, et al. A question-and-answer system to extract data from free-text oncological pathology reports (CancerBERT Network): Development Study. J Med Internet Res; 2022 24(3): e27210. doi: 10.2196/27210. PMID: 35319481

220- Patel JS, Su C, Tellez M, Albandar JM, Rao R, Iyer V, Shi E, Wu H. Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data. Front Artif Intell; 2022 (5): 979525. doi: 10.3389/frai.2022.979525. PMID: 36311550

221- Sun W, Cai Z, Li Y, Liu F, Fang S, Wang G. Data processing and text mining technologies on electronic medical records: A Review. J Healthc Eng; 2018; 2018: 4302425. doi: 10.1155/2018/4302425. PMID: 29849998

222- Pethani F, Dunn AG. Natural language processing for clinical notes in dentistry: A systematic review. Journal of Biomedical Informatics; 2023 (138): 104282. doi: 10.1016/j.jbi.2023.104282. PMID: 36623780.

223- Benicio DHP, Xavier-Júnior JC, Paiva KRS, Camargo JDDAS. Applying text mining and natural language processing to electronic medical records for extracting and transforming texts into structured data. 2022, Available at http://dx.doi.org/10.2139/ssrn.3991515

224- Chen Q, Zhou X, Wu J, Zhou Y. Structuring electronic dental records through deep learning for a clinical decision support system. Health Informatics J; 2021 27(1):1460458220980036. doi: 10.1177/1460458220980036. PMID: 33446032

225- Patel JS, Kumar K, Zai A, Shin D, Willis L, Thyvalikakath TP. Developing automated computer algorithms to track periodontal disease change from longitudinal electronic dental records. Diagnostics (Basel); 2023 13(6):1028. doi: 10.3390/diagnostics13061028. PMID: 36980336

226- Wei Q, Ji Z, Si Y, Du J, Wang J, Tiryaki F, et al. Relation extraction from clinical narratives using pre-trained Language Models. AMIA Annu Symp Proc; 2020 (2019): 1236-1245. PMID: 32308921

227- Li F, Jin Y, Liu W, Rawat BP, Cai P, Yu H. Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)-based models on large-scale electronic health record notes: an empirical study. JMIR Med Inform. 2019; 7(3): e14830. doi: 10.2196/14830. PMID: 31516126

228- Xu D, Gopale M, Zhang J, Brown K, Begoli E, Bethard S. Unified medical language system resources improve sieve-based generation and Bidirectional Encoder Representations from Transformers (BERT)-based ranking for concept normalization. J Am Med Inform Assoc; 2020; 27(10):1510–19. doi: 10.1093/jamia/ocaa080. PMID: 32719838

229- Kim J, Amar S. Periodontal disease and systemic conditions: A bidirectional relationship. Odontology; 2006; 94(1): 10-21. doi: 10.1007/s10266-006-0060-6. PMID: 16998613

230- Ray P, Chakrabarti A. A mixed approach of deep learning method and rule-based method to improve aspect aevel sentiment analysis. Applied Computing and Informatics; 2022 18(1/2): 163 178. https://doi.org/10.1016/j.aci.2019.02.002

231- Lagouvardos S, Dolby J, Grech N, Antoniadis A, Smaragdakis Y. Static analysis of shape in TensorFlow programs, in 34th European Conference on Object-Oriented Programming (ECOOP 2020), LIPIcs, Vol. 166, pp. 15:1–15:29, 2020. https://doi.org/10.4230/LIPIcs.ECOOP.2020.15

232- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, WeissR, Dubourg V, Vanderplas J, Passos A, CournapeauD, Brucher M, Perrot M, and Duchesnay E. Scikit-learn: Machine learning in Python. Journal of MachineLearning Research 2011; 12:2825–30

233- Qin H. Comparison of deep learning models on time series forecasting: a case study of dissolved oxygen prediction. preprint [2019]. Available at: https://www.researchgate.net/publication/337386775

234- Ertaş K, Pence I, Cesmeli MS, Ay ZY. Determination of the stage and grade of periodontitis according to the current classification of periodontal and peri-implant diseases and conditions (2018) using machine learning algorithms. J Periodontal Implant Sci. 2023;53(1):38-53. doi: 10.5051/jpis.2201060053.

235- Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022; 12(1):5979. doi: 10.1038/s41598-022-09954-8. PMID: 35395867; PMCID: PMC8993826.

236- Haulcy R, Glass J. Classifying alzeimer's disease using audio and text-based representations of speech. Front Psychol; 2021; (11): 624137. doi: 10.3389/fpsyg.2020.624137. PMID: 33519651

237- Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics; 2020; 36(4): 1234-1240. doi: 10.1093/bioinformatics/btz682. PMID: 31501885

238- Liu, N, Hu Q, Xu H, Xu X, Chen M. Med-BERT: A pretraining framework for medical records named entity recognition. IEEE Transactions on Industrial Informatics.2021; 18(8), 5600-5608.

239- Bichu YM, Hansa I, Bichu AY, Premjani P, Flores-Mir C, Vaid NR. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. Prog Orthod; 2021; (22): 18. doi: 10.1186/ s40510- 021-00361-9. PMID: 34219198

240- Ossowska A, Kusiak A, Świetlik D. Artificial intelligence in dentistry-narrative review. Int J Environ Res Public Health; 2022; 19(6):3449. Published 2022 Mar 15. doi:10.3390/ijerph19063449. PMID: 35329136

241- Humphrey LL, Fu R, Buckley DI, Freeman M, Helfand M. Periodontal disease and coronary heart disease incidence: a systematic review and meta-analysis. J. Gen. Internal Med. 2008; 23: 2079. doi: 10.1007/s11606-008-0787-6

242- Trombelli L., Farina R., Silva C.O., et al. Plaque-induced gingivitis: case definition and diagnostic considerations. J Clin Periodontol. 2018;45:S44–S67

243- Hoi SCH, Jin R, Zhu J, Lyu MR. Batch mode active learning and its application to medical image classification. In Proceedings of the ACM International Conference Proceeding Series; New York, 2006; 148: 417–424.

244- Uzun Saylan BC, Baydar O, Yeşilova E, Kurt Bayrakdar S, Bilgir E, Bayrakdar İŞ, Çelik Ö, Orhan K. Assessing the effectiveness of artificial intelligence models for detecting alveolar bone loss in periodontal disease: A panoramic radiograph study. Diagnostics (Basel). 2023;13(10):1800. doi: 10.3390/diagnostics13101800.

245- Çelik B, Çelik ME. Automated detection of dental restorations using deep learning on panoramic radiographs. Dentomaxillofac Radiol. 2022;51(8):20220244. doi: 10.1259/dmfr.20220244. Epub 2022 Sep 12.

246- Jiang L, Chen D, Cao Z, Wu F, Zhu H, Zhu F. A two-stage deep learning architecture for radiographic staging of periodontal bone loss. BMC Oral Health. 2022;22(1):106. doi: 10.1186/s12903-022-02119-z.

247- Lee CT, Kabir T, Nelson J, Sheng S, Meng HW, Van Dyke TE, Walji MF, Jiang X, Shams S. Use of the deep learning approach to measure alveolar bone level. J Clin Periodontol. 2022;49(3):260-269. doi: 10.1111/jcpe.13574. Epub 2021 Dec 31.

248- Kim TS, Obst C, Zehaczek S, Geenen C. Detection of bone loss with different X-ray techniques in periodontal patients. J Periodontol. 2008;79(7):1141-9. doi: 10.1902/jop.2008.070578. PMID: 18597595.

249- Albandar JM, Abbas DK, Waerhaug M, Gjermo P. Comparison between standardized periapical and bitewing radiographs in assessing alveolar bone loss. Community Dent Oral Epidemiol. 1985;13(4):222-5. doi: 10.1111/j.1600-0528.1985.tb01908.x. PMID: 3862504.

250- Dwyer B, Nelson J, Hansen T, et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.

251- Joseph VR. Optimal ratio for data splitting, statistical analysis and data mining: The ASA Data Science Journal. 2022; 15(4): 531–538, https://doi.org/10.1002/sam.11583.

252- Mirza A, Rajak RK. Segmentation of Polyp Instruments using UNet based deep learning model. Nordic Machine Intelligence. 2021

253- Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

254- Bruton A, Conway JH, Holgate ST. Reliability: what is it, and how is it measured? Physiotherapy. 2000; 86: 94–99

255- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012.

256- Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021;7:e623. doi: 10.7717/peerj-cs.623. PMID: 34307865; PMCID: PMC8279135.

257- Portney LG, Watkins MP. Prentice Hall; New Jersey: 2000. Foundations of clinical research: applications to practice.

258- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30–46

259- Vallat R. (2018). Pingouin: statistics in Python. Journal of Open Source Software, 3(31), 1026. https://doi.org/10.21105/joss.01026

260- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143: 29–36.

261- Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432. doi: 10.1371/journal.pone.0118432. PMID: 25738806; PMCID: PMC4349800.

262- Shon HS, Kong V, Park JS, Jang W, Cha EJ, Kim S-Y, Lee E-Y, Kang T-G, Kim KA. Deep learning model for classifying periodontitis stages on dental panoramic radiography. Applied Sciences. 2022; 12(17):8500. https://doi.org/10.3390/app12178500

263- Berghuis G, Cosyn J, De Bruyn H, Hommez G, Dierens M, Christiaens V. A controlled study on the diagnostic accuracy of panoramic and peri-apical radiography for detecting furcation involvement. BMC Oral. Health. 2021; 21:115. doi: 10.1186/s12903-021-01460-z

264- Papapanou PN, Tonetti MS. Diagnosis and epidemiology of periodontal osseous lesions. Periodontology 2000. 2000; 22: 8–21. doi: 10.1034/j.1600-0757.2000. 2220102.x.

265- Pepelassi EA, Diamanti-Kipioti A. Selection of the most accurate method of conventional radiography for the assessment of periodontal osseous destruction. J Clin Periodontol 1997; 24(8) :557–567. https://doi.org/10.1111/j.1600-051X.1997.tb00229.x

266- Ezhov M, Gusarev M, Golitsyna M, et al. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. Sci Rep 11, 15006 (2021). https://doi.org/10.1038/s41598-021-94093-9

267- Venkatesh E, Elluru SV. Cone beam computed tomography: basics and applications in dentistry. J Istanbul Univ Fac Dentistry 2017;51(3 Suppl1):S102–s121.

268- Li X, Zhao D, Xie J, Wen H, Liu C, Li Y, Li W, Wang S. Deep learning for classifying the stages of periodontitis on dental images: a systematic review and

meta-analysis. BMC Oral Health. 2023;23(1):1017. doi: 10.1186/s12903-023-03751-z. PMID: 38114946; PMCID: PMC10729340.

269- Kumar A, Bhadauria HS, Singh A. Descriptive analysis of dental X-ray images using various practical methods: a review. Peer J Comput Sci. 2021;7:e620.

270- Redmon J SD, Girshick R, Farhadi A. You Only Look Once: Unified, real-time object detection. In: arxiv:150602640[csCV] 2015.58.

271- Yin XX, Sun L, Fu Y, Lu R, Zhang Y. U-Net-Based Medical Image Segmentation. Journal of healthcare engineering 2022; 2022: 4189781.

272- Dujic H, Meyer O, Hoss P, Wölfle UC, Wülk A, Meusburger T, Meier L, Gruhn V, Hesenius M, Hickel R, et al. Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks. Diagnostics. 2023; 13(23):3562. https://doi.org/10.3390/diagnostics13233562

273- Patil S, Waghule S, Waje S, Pawar P, Domb S. Efficient object detection with YOLO: A comprehensive guide. International Journal of Advanced Research in Science, Communication and Technology. 2024; 519-531. 10.48175/IJARSCT-18483.

274- Sirisha U, Praveen SP, Srinivasu PN, et al. Statistical analysis of design aspects of various YOLO-based deep learning models for object detection. Int J Comput Intell Syst 16, 126 (2023). https://doi.org/10.1007/s44196-023-00302-w

275- Flores-Calero M, Astudillo, C, Guevara B, Diego E, Maza, J, Lita, B, Defaz, B, Ante, J, Zabala-Blanco, D, Armingol, JM. Traffic sign detection and recognition using YOLO object detection algorithm: A systematic review. Mathematics. 2024; 12. 297. 10.3390/math12020297.

276- Chen IH, Lin CH, Lee MK, Chen TE, Lan TH, Chang CM, Tseng TY, Wang T, Du JK. Convolutional-neural-network-based radiographs evaluation assisting in early diagnosis of the periodontal bone loss via periapical radiograph. J Dent Sci. 2024;19(1):550-559. doi: 10.1016/j.jds.2023.09.032.

277- Sunnetci KM, Ulukaya S, Alkan A. Periodontal bone loss detection based on hybrid deep learning and machine learning models with a user-friendly application. Biomed Signal Process Control 2022;77.

278- Zhang F, Li Z, Zhang B, Du H, Wang B, Zhang X. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease Neurocomputing, 361 (2019), pp. 185-195, 10.1016/j.neucom.2019.04.093

279- Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review Comput. Methods Programs Biomed., 161 (2018): 1-13, 10.1016/J.CMPB.2018.04.005

280- Nweke HF, The YW, Mujtaba G, Al-garadi MA. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions Aktuel. Aspekte Kernfusionsforsch., Informationstag., 46 (2019): 147-170, 10.1016/J.INFFUS.2018.06.002

281- Doi K. Diagnostic imaging over the last 50 years: Research and development in medical imaging science and technology Phys. Med. Biol., 51 (13) (2006), 10.1088/0031-9155/51/13/R02

282- Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review J. Med. Syst.2016; 40 (12), 10.1007/s10916-016-0628-9

283- Huang C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digital Med.2020; 3 (1), 10.1038/s41746-020-00341-z

284- Jiang F, et al. Artificial intelligence in healthcare: past, present and future Stroke Vasc. Neurol., 2 (4) (2017), pp. 230-243, 10.1136/SVN-2017-000101

285- Jiang Y, Li W, Hossain MS, Chen M, Alelaiwi A, Al-Hammadi M. A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition Aktuel. Aspekte Kernfusionsforsch., Informationstag.2020; 53: 209-221, 10.1016/J.INFFUS.2019.06.019

286- Searle T, Ibrahim Z, Dobson R. Comparing natural language processing techniques for Alzheimer's dementia prediction in spontaneous speech. 2020, [Online]. Available: http://arxiv.org/abs/2006.07358

287- Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. PLoS Comput. Biol.2018; 14 (9), 10.1371/journal.pcbi.1006376

288- Syed ZS, Syed MSS, Lech M, Pirogova E. Automated recognition of Alzheimer's dementia using bag-of-deep-features and model ensembling IEEE Access.2021; 9: 88377-88390, 10.1109/ACCESS.2021.3090321

289- Pfänder L, Schneider L, Büttner M, Krois J, Meyer-Lueckel H, Schwendicke F. Multi-modal deep learning for automated assembly of periapical radiographs, Journal of Dentistry. 2023; 135, 104588, ISSN 0300-5712, https://doi.org/10.1016/j.jdent.2023.104588.

290- Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365. doi: 10.1371/journal.pone.0224365. PMID: 31697686; PMCID: PMC6837442.

291- White J, Power SD. k-Fold cross-validation can significantly over-estimate true classification accuracy in common EEG-based passive BCI experimental designs: An empirical investigation. Sensors (Basel). 2023;23(13):6077. doi: 10.3390/s23136077. PMID: 37447926; PMCID: PMC10346713.

292- Thung KH, Yap PT, Shen D. Multi-stage diagnosis of Alzheimer's disease with incomplete multimodal data via multi-task deep learning. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (eds. Cardoso, M. J. et al.) vol. 10553, 160–168 (Springer International Publishing, 2017).

293- Li H, Fan Y. Early prediction of Alzheimer's disease dementia based on baseline hippocampal MRI and 1-year follow-up cognitive measures using deep recurrent neural networks. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) 368–371 (IEEE, 2019)

294- Qiu, S. et al. Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimers Dement. Diagn. Assess. Dis. Monit. 10, 737–749 (2018).

295- Nagrani A, Yang S, Arnab A, Jansen A, Schmid C, Sun C. Attention bottlenecks for multimodal fusion. In NeurIPS, 2021.

296- Lee S, Yu Y, Kim G, Breuel T, Kautz J, Song Y. Parameter efficient multimodal transformers for video representation learning. In ICLR, 2021.

297- Strudel R, Garcia R, Laptev I, Schmid C. Segmenter: Transformer for semantic segmentation. In ICCV, 2021.

298- Kim W, Son B, Kim I. ViLT: Vision-and-language transformer without convolution or region supervision. In ICML: 5583–5594, 2021.

299- Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, et al. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258, 2021.

300- Ma M, Ren J, Zhao L, Testuggine D, Peng X. Are Multimodal Transformers Robust to Missing Modality? 2022 18156-18165. 10.1109/CVPR52688.2022.01764.

301- Lian Z, Tao J, Liu B, Huang J, Yang Z, Li R. (2020). Context-dependent domain adversarial neural network for multimodal emotion recognition. in Interspeech (Cary, NC), 394–398.

302- Niu C, Wang G. Unsupervised contrastive learning-based transformer for lung nodule detection. Phys. Med. Biol. 2022; 67: 204001. doi: 10.1088/1361-6560/ac92ba.

303- Wu Y, Qi S, Sun Y, Xia S, Yao Y, Qian W. A vision transformer for emphysema classification using CT images. Phys. Med. Biol. 2021; 66: 245016. doi: 10.1088/1361-6560/ac3dc8.

304- Park S, Kim G, Oh Y, Seo JB, Lee SM, Kim JH, Moon S, Lim JK, Park CM, Ye JC. Self-evolving vision transformer for chest X-ray diagnosis through knowledge distillation. Nat. Commun. 2022; 13: 3848. doi: 10.1038/s41467-022-31514-x.

305- Dujic H, Meyer O, Hoss P, Wölfle U, Wülk A, Meusburger T, et al. Automatized detection of periodontal bone loss on periapical radiographs by Vision Transformer Networks. Diagnostics. 2023 13. 3562. 10.3390/diagnostics13233562.

306- Abiyev RH, Altabel MZ, Darwish M, Helwan A. A multimodal transformer model for recognition of images from complex laparoscopic surgical videos. Diagnostics. 2024; 14(7):681. https://doi.org/10.3390/diagnostics14070681

307- Lyu W, Dong X, Wong R, Zheng S, Abell-Hart K, Wang F, Chen C. A multimodal transformer: fusing clinical notes with structured EHR data for interpretable in-hospital mortality prediction. AMIA Annu Symp Proc. 2023; 2022: 719-728. PMID: 37128451; PMCID: PMC10148371.

308- Huang K, Altosaar J, Ranganath R. Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342. 2019 Apr 10

309- Zhao D, Homayounfar M, Zhen Z, Wu M-Z, Yu SY, Yiu K-H, et al. A multimodal deep learning approach to predicting systemic diseases from oral conditions. Diagnostics. 2022; 12(12):3192. https://doi.org/10.3390/diagnostics12123192

310- Hsieh ST, Cheng, YA. Multimodal feature fusion in deep learning for comprehensive dental condition classification. 2024: 303 – 321.

311- Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Brief Bioinform. 2022;23(2): bbab569. doi: 10.1093/bib/bbab569. PMID: 35089332; PMCID: PMC8921642.

312- Alawaji YN, Alshammari A, Mostafa N, Carvalho RM, Aleksejuniene J. Periodontal disease prevalence, extent, and risk associations in untreated individuals.

Clin Exp Dent Res. 2022;8(1):380-394. doi: 10.1002/cre2.526. Epub 2022 Jan 10. PMID: 35015383; PMCID: PMC8874091.

313- Eke PI, Dye B, Wei L, Slade G, Thornton-Evans, G, Borgnakke W et al. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. J Periodontol 2015;86(5):611-22

314- Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ, CDC Periodontal Disease Surveillance workgroup. Prevalence of periodontitis in adults in the United States: 2009 and 2010. J Dent Res. 2012;91(10):914-20.

315- Brown LJ, John BA, Wall TP. The economics of periodontal disease. Periodontol 2000. 2002;29:223-34.

316- Ansari S, deo V, Ansari S, Gupta S. Radiology in Periodontal Disease Diagnosis. International Journal of Multidisciplinary Health Sciences. 2014; 1. 2-9.

317- Corbet EF, Ho DKL, Lai SML. Radiographs in periodontal disease diagnosis and management. Australian Dental Journal 2009;54:(1 Suppl): S27-43.

318- Ronneberger O, Fischer P, Brox T. Dental X-ray image segmentation using a U-shaped deep convolutional network. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), Vol. 9351 234–243 (ISBI, 2015).

319- Bouchey B, Castek J, Thygeson J. Multimodal learning. In: Ryoo, J., Winkelmann, K. (eds) Innovative Learning Environments in STEM Higher Education. Springer Briefs in Statistics. 2021; Springer, Cham. https://doi.org/10.1007/978-3-030-58948-6_3

320- Luo H. Editorial: Advances in multimodal learning: pedagogies, technologies, and analytics. Front Psychol. 2023; 14: 1286092. doi: 10.3389/fpsyg.2023.1286092. PMID: 38022937; PMCID: PMC10644768.

321- Ross A. Fusion, feature-level. In: Li, S.Z., Jain, A. (eds) Encyclopedia of Biometrics. 2009; Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_157

322- Machado V, Proença L, Morgado M, Mendes JJ, Botelho J. Accuracy of panoramic radiograph for diagnosing periodontitis comparing to clinical examination. Journal of Clinical Medicine. 2020; 9(7):2313. https://doi.org/10.3390/jcm9072313

323- Meskó B. The impact of multimodal large language models on health care's future. J Med Internet Res. 2023; 25: e52865. doi: 10.2196/52865. PMID: 37917126; PMCID: PMC10654899.

324- Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: A scoping review. npj Digit. Med. 5, 171 (2022). https://doi.org/10.1038/s41746-022-00712-8

325- Chen X, Xie H, Tao X, et al. Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics. Artif Intell Rev 57, 91 (2024). https://doi.org/10.1007/s10462-024-10712-7

326- Amal S, Safarnejad L, Omiye JA, Ghanzouri I, Cabot JH, Ross EG. Use of multi-modal data and machine learning to improve cardiovascular disease care. Front Cardiovasc Med. 2022; 9: 840262. doi: 10.3389/fcvm.2022.840262. PMID: 35571171; PMCID: PMC9091962.

327- Oh SL, Yang JS, Kim YJ. Discrepancies in periodontitis classification among dental practitioners with different educational backgrounds. BMC Oral Health 2021; 21, 39. https://doi.org/10.1186/s12903-020-01371-5

328- Leevy JL, Khoshgoftaar TM, Bauder RA, et al. A survey on addressing high-class imbalance in big data. J Big Data. 2018; 5: 42. https://doi.org/10.1186/s40537-018-0151-6

329- Sánchez-Gutiérrez ME, González-Pérez PP. Addressing the class imbalance in tabular datasets from a generative adversarial network approach in supervised machine learning. Journal of Algorithms & Computational Technology. 2023; 17. doi:10.1177/17483026231215186

330- Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering. 2005; 30: 25-36.

Appendix

APPENDIX 1. U-NET MODEL ARCHITECTURE FOR SEGMENTING BONE LOSS ON PERIAPICAL

Layer (type)	**Output Shape**	**Param #**	**Connected to**
input_8 (InputLayer)	(None, 160, 320, 1)	0	[]
conv2d_133 (Conv2D)	(None, 160, 320, 64)	640	['input_8[0][0]']
batch_normalization_126 (BatchNormalization)	(None, 160, 320, 64)	256	['conv2d_133[0][0]']
activation_126 (Activation)	(None, 160, 320, 64)	0	['batch_normalization_126[0][0]']
conv2d_134 (Conv2D)	(None, 160, 320, 64)	36928	['activation_126[0][0]']
batch_normalization_127 (BatchNormalization)	(None, 160, 320, 64)	256	['conv2d_134[0][0]']
activation_127 (Activation)	(None, 160, 320, 64)	0	['batch_normalization_127[0][0]']
dropout_63 (Dropout)	(None, 160, 320, 64)	0	['activation_127[0][0]']
max_pooling2d_28 (MaxPooling2D)	(None, 80, 160, 64)	0	['dropout_63[0][0]']

conv2d_135 (Conv2D)	(None, 80, 160, 128)	73856	['max_pooling2d_28[0][0]']
batch_normalization_128 (BatchNormalization)	(None, 80, 160, 128)	512	['conv2d_135[0][0]']
activation_128 (Activation)	(None, 80, 160, 128)	0	['batch_normalization_128[0][0]']
conv2d_136 (Conv2D)	(None, 80, 160, 128)	147584	['activation_128[0][0]']
batch_normalization_129 (BatchNormalization)	(None, 80, 160, 128)	512	['conv2d_136[0][0]']
activation_129 (Activation)	(None, 80, 160, 128)	0	['batch_normalization_129[0][0]']
dropout_64 (Dropout)	(None, 80, 160, 128)	0	['activation_129[0][0]']
max_pooling2d_29 (MaxPooling2D)	(None, 40, 80, 128)	0	['dropout_64[0][0]']
conv2d_137 (Conv2D)	(None, 40, 80, 256)	295168	['max_pooling2d_29[0][0]']
batch_normalization_130 (BatchNormalization)	(None, 40, 80, 256)	1024	['conv2d_137[0][0]']
activation_130 (Activation)	(None, 40, 80, 256)	0	['batch_normalization_130[0][0]']
conv2d_138 (Conv2D)	(None, 40, 80, 256)	590080	['activation_130[0][0]']

batch_normalization_131 (BatchNormalization)	(None, 40, 80, 256)	1024	['conv2d_138[0][0]']
activation_131 (Activation)	(None, 40, 80, 256)	0	['batch_normalization_131[0][0]']
dropout_65 (Dropout)	(None, 40, 80, 256)	0	['activation_131[0][0]']
max_pooling2d_30 (MaxPooling2D)	(None, 20, 40, 256)	0	['dropout_65[0][0]']
conv2d_139 (Conv2D)	(None, 20, 40, 528)	1217040	['max_pooling2d_30[0][0]']
batch_normalization_132 (BatchNormalization)	(None, 20, 40, 528)	2112	['conv2d_139[0][0]']
activation_132 (Activation)	(None, 20, 40, 528)	0	['batch_normalization_132[0][0]']
conv2d_140 (Conv2D)	(None, 20, 40, 528)	2509584	['activation_132[0][0]']
batch_normalization_133 (BatchNormalization)	(None, 20, 40, 528)	2112	['conv2d_140[0][0]']
activation_133 (Activation)	(None, 20, 40, 528)	0	['batch_normalization_133[0][0]']
dropout_66 (Dropout)	(None, 20, 40, 528)	0	['activation_133[0][0]']
max_pooling2d_31 (MaxPooling2D)	(None, 10, 20, 528)	0	['dropout_66[0][0]']
conv2d_141 (Conv2D)	(None, 10, 20, 1024)	4867072	['max_pooling2d_31[0][0]']

batch_normalization_134 (BatchNormalization)	(None, 10, 20, 1024)	4096	['conv2d_141[0][0]']
activation_134 (Activation)	(None, 10, 20, 1024)	0	['batch_normalization_134[0][0]']
conv2d_142 (Conv2D)	(None, 10, 20, 1024)	9438208	['activation_134[0][0]']
batch_normalization_135 (BatchNormalization)	(None, 10, 20, 1024)	4096	['conv2d_142[0][0]']
activation_135 (Activation)	(None, 10, 20, 1024)	0	['batch_normalization_135[0][0]']
dropout_67 (Dropout)	(None, 10, 20, 1024)	0	['activation_135[0][0]']
conv2d_transpose_28 (Conv2DTranspose)	(None, 20, 40, 512)	2097664	['dropout_67[0][0]']
concatenate_28 (Concatenate)	(None, 20, 40, 1040)	0	['conv2d_transpose_28[0][0]', 'dropout_66[0][0]']
conv2d_143 (Conv2D)	(None, 20, 40, 512)	4792832	['concatenate_28[0][0]']
batch_normalization_136 (BatchNormalization)	(None, 20, 40, 512)	2048	['conv2d_143[0][0]']
activation_136 (Activation)	(None, 20, 40, 512)	0	['batch_normalization_136[0][0]']
conv2d_144 (Conv2D)	(None, 20, 40, 512)	2359808	['activation_136[0][0]']

batch_normalization_137 (BatchNormalization)	(None, 20, 40, 512)	2048	['conv2d_144[0][0]']
activation_137 (Activation)	(None, 20, 40, 512)	0	['batch_normalization_137[0][0]']
dropout_68 (Dropout)	(None, 20, 40, 512)	0	['activation_137[0][0]']
conv2d_transpose_29 (Conv2DTranspose)	(None, 40, 80, 256)	524544	['dropout_68[0][0]']
concatenate_29 (Concatenate)	(None, 40, 80, 512)	0	['conv2d_transpose_29[0][0]', 'dropout_65[0][0]']
conv2d_145 (Conv2D)	(None, 40, 80, 256)	1179904	['concatenate_29[0][0]']
batch_normalization_138 (BatchNormalization)	(None, 40, 80, 256)	1024	['conv2d_145[0][0]']
activation_138 (Activation)	(None, 40, 80, 256)	0	['batch_normalization_138[0][0]']
conv2d_146 (Conv2D)	(None, 40, 80, 256)	590080	['activation_138[0][0]']
batch_normalization_139 (BatchNormalization)	(None, 40, 80, 256)	1024	['conv2d_146[0][0]']
activation_139 (Activation)	(None, 40, 80, 256)	0	['batch_normalization_139[0][0]']
dropout_69 (Dropout)	(None, 40, 80, 256)	0	['activation_139[0][0]']
conv2d_transpose_30 (Conv2DTranspose)	(None, 80, 160, 128)	131200	['dropout_69[0][0]']

concatenate_30 (Concatenate)	(None, 80, 160, 256)	0	['conv2d_transpose_30[0][0]', 'dropout_64[0][0]']
conv2d_147 (Conv2D)	(None, 80, 160, 128)	295040	['concatenate_30[0][0]']
batch_normalization_140 (BatchNormalization)	(None, 80, 160, 128)	512	['conv2d_147[0][0]']
activation_140 (Activation)	(None, 80, 160, 128)	0	['batch_normalization_140[0][0]']
conv2d_148 (Conv2D)	(None, 80, 160, 128)	147584	['activation_140[0][0]']
batch_normalization_141 (BatchNormalization)	(None, 80, 160, 128)	512	['conv2d_148[0][0]']
activation_141 (Activation)	(None, 80, 160, 128)	0	['batch_normalization_141[0][0]']
dropout_70 (Dropout)	(None, 80, 160, 128)	0	['activation_141[0][0]']
conv2d_transpose_31 (Conv2DTranspose)	(None, 160, 320, 64)	32832	['dropout_70[0][0]']
concatenate_31 (Concatenate)	(None, 160, 320, 128)	0	['conv2d_transpose_31[0][0]', 'dropout_63[0][0]']
conv2d_149 (Conv2D)	(None, 160, 320, 64)	73792	['concatenate_31[0][0]']
batch_normalization_142 (BatchNormalization)	(None, 160, 320, 64)	256	['conv2d_149[0][0]']
activation_142 (Activation)	(None, 160, 320, 64)	0	['batch_normalization_142[0][0]']

conv2d_150 (Conv2D)	(None, 160, 320, 64)	36928	['activation_142[0][0]']
batch_normalization_143 (BatchNormalization)	(None, 160, 320, 64)	256	['conv2d_150[0][0]']
activation_143 (Activation)	(None, 160, 320, 64)	0	['batch_normalization_143[0][0]']
dropout_71 (Dropout)	(None, 160, 320, 64)	0	['activation_143[0][0]']
conv2d_151 (Conv2D)	(None, 160, 320, 1)	65	['dropout_71[0][0]']