UNIVERSITI TEKNOLOGI MARA

PREDICTION OF CARIES RISK IN ADULTS VIA SUPERVISED MACHINE LEARNING ALGORITHM

MOHD HIDIR BIN MOHD ATNI

Dissertation submitted in partial fulfillment
of the requirements for the degree of
Dr in Dental Public Health

Faculty of Dentistry

August 2025

ACKNOWLEDGEMENT

Alhamdulillah, all praise and thanks are due to Allah SWT. He has granted me the strength, patience, and clarity needed to complete this dissertation. Each milestone in this journey has been possible only through His mercy and guidance.

I am profoundly grateful to my supervisory team, Associate Professor Dr. Budi Aslinie Md Sabri, Associate Professor Dr. Nik Mohd Mazuan Nik Mohd Rosdy, and Mr. Mohd Azrul Amir Muhammad Tajuddin. Your thoughtful guidance, constructive feedback, and consistent encouragement have shaped both this research and my growth as a scholar. Your mentorship has shown me the value of perseverance, intellectual curiosity, and humility in academic work.

I owe immeasurable gratitude to my beloved mother, whose prayers and unconditional love have sustained me throughout this process. I also honour the memory of my late father, whose wisdom and encouragement continue to guide my steps. To my wife and our children, your patience, sacrifices, and unwavering belief in me have been my greatest source of strength, and I extend heartfelt thanks to every member of our extended family for their constant support and encouragement.

My warmest thanks go to my classmates who have stood at my side since the first day. Together we crossed seas and borders for meetings, attachments and postgraduate courses, gathering lessons from lecture halls and from the many lands that welcomed us. The fellowship we forged on those roads and the late-night talks that turned every obstacle into new understanding have given this journey a depth I will carry always.

Finally, I wish to acknowledge the many colleagues, mentors and friends whose names are not listed here yet whose contributions have quietly shaped this work. Every shared insight, word of encouragement and act of kindness has moved me closer to this goal. May this dissertation stand as a testament to the truth that knowledge grows best when nurtured by a community of generous hearts and minds.

ABSTRACT

Background: Dental caries is a prevalent global health issue, requiring effective tools for early detection and prevention. Traditional methods of caries risk assessment rely on subjective clinical judgment, which can vary significantly among practitioners. Machine learning offers a promising solution by providing objective and reproducible predictions, potentially improving screening accuracy and facilitating early intervention. **Objectives**: The study aimed to address five key objectives: to identify relevant caries risk factors, to model the influence of identified caries risk factors on the caries risk, develop a machine learning prediction model for caries risk assessment, assess performance of machine learning models in predicting caries risk by comparing them against an expert-based benchmark and evaluate the agreement and consistency between human raters and the machine learning prediction model. **Methods**: A systematic review was conducted to identify evidence-based caries risk factors, leading to the selection of 12 predictors. A balanced dataset of 3,000 samples was constructed, categorized into low, moderate, and high-risk groups. Seven machine learning algorithms, including Decision Tree, XGBoost, k-Nearest Neighbours, Logistic Regression, Multi-Layer Perceptron, Random Forest, and Support Vector Machines, were trained and optimised using K-fold cross-validation. Ensemble learning methods were employed to enhance model performance. Agreement between human raters and machine predictions was assessed using Fleiss' and Cohen's Kappa. Additionally, machine learning models and human performance were compared to benchmark via Accuracy, F1-score, and ROC-AUC. **Results**: The machine learning models demonstrated superior performance compared to human raters. Ensemble learning models, particularly those combining k-Nearest Neighbours and Logistic Regression, achieved the highest ROC-AUC (0.97) exceeding human rater performance. Cohen's Kappa values indicated moderate to substantial agreement between machine predictions and certain raters, while inter-rater agreement among humans was inconsistent. These findings highlighted the reliability and consistency of machine learning predictions over human judgment in caries risk assessment. **Conclusion**: This study established a robust computational framework for caries risk prediction, leveraging machine learning to complement clinical decision-making. The results underscored the potential of predictive models to enhance the accuracy and consistency of caries risk assessment.

TABLE OF CONTENTS

Contents

CHAPTER 1
INTRODUCTION

1.1 Research Background

Dental caries is a widespread oral health issue that affects people worldwide, regardless of age or socioeconomic status (Bernabe et al., 2020). Despite advancements in preventive and therapeutic strategies, it continues to be a significant public health challenge, contributing to pain, tooth loss, and a reduced quality of life (T. T. Wu et al., 2021). Effective identification of risk factors is essential for developing targeted interventions to reduce the prevalence of this disease (S.-A. Sadegh-Zadeh et al., 2022).

However, traditional methods of caries risk assessment (CRA) often depend on clinical expertise and subjective judgment, resulting in inconsistencies in accuracy (Kühnisch et al., 2022). These limitations are particularly problematic for population-based screening programs, where scalability, efficiency, and standardised risk stratification are critical to addressing oral health disparities at scale (Mejàre et al., 2013). This inconsistency underscores the urgent need for objective, data-driven approaches to ensure standardised, precise, and reliable risk predictions, especially in light of increasing global oral health disparities (Anil et al., 2023).

Recent advancements in machine learning have transformed predictions in healthcare by enabling the analysis of complex datasets to uncover intricate patterns and associations (W. Li et al., 2021). Machine learning holds great potential for CRA by integrating various factors such as oral hygiene practices, dietary habits, biological indicators, and socioeconomic conditions (Anil et al., 2023). These models can provide highly accurate, personalized predictions, allowing for the effective identification of high-risk individuals compared to traditional methods (Fontana et al., 2020). Additionally, machine learning can highlight critical risk factors, facilitating the design of tailored preventive strategies (Wongvibulsin et al., 2019). This technological evolution promises to revolutionize caries risk prediction and management, enhancing the precision and efficiency of oral healthcare (Anil et al., 2023).

Despite the significant promise of machine learning in healthcare, its application in caries risk prediction faces challenges related to validation and implementation (Fontana et al., 2020). Many existing models have shown potential in research settings