UNIVERSITI TEKNOLOGI MARA

AQUILARIA ESSENTIAL OIL SPECIES DISCRIMINATION USING K-NN MODEL APPLIED TO GC-MS COUPLED WITH GC-FID DATA

NOOR AIDA SYAKIRA BINTI AHMAD SABRI

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)

Faculty of Electrical Engineering

December 2025

ABSTRACT

The classification of *Aquilaria* essential oil, widely referred to as agarwood oil, presents a challenge due to its chemically diverse composition and the subjectivity associated with traditional species identification methods. This study introduces a machine learning approach for species discrimination by analyzing chemical compounds obtained through gas chromatography-mass spectrometry (GC-MS) coupled with gas chromatography-flame ionization detection (GC-FID). The research focuses on four *Aquilaria* species, namely *Aquilaria beccariana*, *Aquilaria malaccensis*, *Aquilaria crassna* and *Aquilaria subintegra*, with the aim of developing a reliable classification framework using the k-Nearest Neighbour (k-NN) algorithm. A total of thirty-eight chemical compounds was identified across all samples, with six consistently detected in all species. These six were subsequently narrowed to four significant compounds using a structured pre-processing workflow comprising cross-correlation analysis to assess compound consistency, boxplot analysis to visualize interspecies variation and selection frequency to prioritise compounds appearing repeatedly across species. Based on statistical strength and visual consistency, four significant compounds, namely dihydro-β-agarofuran, δ-guaiene, 10-epi-γ-eudesmol and γ-eudesmol, were selected for model development due to their high discriminatory power. These compounds were tested in four pairings using the k-NN model under four distance metrics (Euclidean, Minkowski, Correlation and Spearman). The Euclidean and Minkowski distance measures achieved 100% discrimination accuracy, with the pairings of δ-guaiene with 10-epi-γ-eudesmol and 10-epi-γ-eudesmol with γ-eudesmol performing most consistently. The findings showed that *Aquilaria subintegra* and *Aquilaria beccariana* were most clearly distinguished, while *Aquilaria malaccensis* and *Aquilaria crassna* exhibited partial similarity but were still reliably identified through optimal pairings. The model was rigorously evaluated using classification metrics including accuracy, sensitivity, specificity, precision and confusion matrices, with Euclidean and Minkowski achieving 100% accuracy while Correlation and Spearman recorded below 63%. This research presents a validated framework for *Aquilaria* oil species discrimination based on chemical profiling and intelligent modelling, offering practical potential for automated authentication systems and demonstrating applications of electrical engineering in signal analysis, pattern recognition and smart classification technology.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful. All praise belongs to Allah SWT, whose mercy and guidance have brought me to the completion of this PhD journey. Through every trial and moment of doubt, His presence was my source of strength. May peace and blessings be upon Prophet Muhammad (peace be upon him), his family and his companions.

This journey has been a transformative experience, both intellectually and personally. My deepest appreciation goes to my supervisor, Assoc. Prof. Ir. Ts. Dr. Nurlaila binti Ismail, for her tireless guidance, wisdom and continuous support throughout this work. I am also sincerely thankful for the insights and advice of my co-supervisors, Prof. Ir. Ts. Dr. Hj. Mohd Nasir bin Taib and Assoc. Prof. Ts. Dr. Zakiah binti Mohd Yusoff, which have been invaluable to the development of this thesis.

Gratitude is extended to all members of the Advanced Signal Processing Research Interest Group (ASP RIG) for their helpful discussions, technical guidance and moral support throughout the research process. Warm thanks also go to all friends and fellow researchers who stood by me with kindness and encouragement. Your support made this journey lighter and more meaningful.

Endless love and thanks go to my dearest parents, Ahmad Sabri bin Ibrahim and , whose prayers, sacrifices and encouragement have carried me through every stage of this academic pursuit. Their unwavering belief in my abilities gave me the courage to push forward, even in the most difficult times. A heartfelt thank you goes to my younger brother, Azamin whose quiet support and motivation served as a reminder of the importance of perseverance and inner strength.

I am also deeply thankful to the one who has walked this journey alongside me, Muhammad Ikhsan bin Roslan, for providing emotional strength, patience and understanding when I needed it most. Your belief in me never wavered, and your presence made all the difference.

Lastly, I acknowledge my own perseverance. For choosing resilience over doubt, progress over fear and purpose over pressure. This achievement is a collective reflection of faith, love and unwavering commitment. Alhamdulillah.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Research Background

Agarwood, derived from trees of the *Aquilaria* genus, is one of the world's most valuable natural products, renowned for its aromatic resin and essential oil [1–8]. The essential oil extracted from agarwood has long held significant cultural, medicinal and commercial value, particularly across regions in Southeast Asia, the Middle East and beyond [2], [9]. However, the complex formation process of agarwood and the high chemical variability among *Aquilaria* species make consistent identification and classification of its essential oil a persistent challenge in both scientific and commercial domains [2], [3].

The quality and identity of *Aquilaria* oil are largely determined by its chemical composition, which varies significantly across species due to differences in genetics, geography, harvesting methods and resin formation stages [10–12]. Traditionally, agarwood oil discrimination has relied heavily on expert sensory evaluation, including assessments of aroma, colour and viscosity [13–16]. While widely practiced, these subjective methods are prone to inconsistency, bias and limited scalability [16]. The absence of a standardized, data-driven framework for species-level discrimination has become a major bottleneck for quality control, regulatory enforcement and fair market valuation [17–19].

To address these limitations, analytical techniques such as Gas Chromatography-Mass Spectrometry (GC-MS) and Gas Chromatography-Flame Ionization Detector (GC-FID) have been employed to objectively profile chemical compounds in *Aquilaria* oil [20], [21]. When used in a coupled configuration, GC-MS and GC-FID provide a robust platform for qualitative and quantitative analysis with high sensitivity and resolution [22], [23]. These compound profiles, expressed through peak area (%) values, serve as a rich source of information for computational analysis [21].

Recent advancements in machine learning have enabled intelligent discrimination models for natural product analysis [24–29]. Among these, the k-Nearest Neighbour (k-NN) algorithm stands out for its simplicity, interpretability and