

UNIVERSITI TEKNOLOGI MARA

**A HYBRID DEEP LEARNING AND
HANDCRAFTED FEATURE
EXTRACTION WITH OUTLIERS
DETECTION FOR IMAGE-BASED
AIR QUALITY INDEX (AQI)
CLASSIFICATION**

UMI NAJIAH BINTI AHMAD RAZIMI

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Information Technology)

Faculty of Computer and Mathematical Sciences

December 2025

ABSTRACT

Air quality remains a critical global concern due to its direct impact on human health and the environment. Traditional air quality monitoring methods have limitations in terms of coverage, cost, and accuracy, especially in extreme environmental conditions. This research introduces an AQI classification model using the same dataset as the Eff-AQI model, which leverages image-based data to assess air quality levels in urban environments. The model combines deep learning-based feature extraction with handcrafted features and image outlier detection techniques to address challenges such as the inability to capture complex pollution-specific patterns effectively and the lack of robust mechanisms for detecting outliers or anomalies in data. These issues often lead to inaccurate classifications, particularly in extreme environmental conditions, like high pollution levels, or due to technical issues in image capture, including poor image quality. The model uses VGG16 to extract high-dimensional features from AQI images, which represent visual data captured from various urban locations with different pollution levels. These images offer a comprehensive view of the environmental conditions affecting air quality. Handcrafted features, such as color (RGB) and texture (LBP), further enhance classification performance by capturing finer details, complementing the high-dimensional features extracted by the deep learning model. For training, a Fully Connected Neural Network (FCNN) is employed to classify the AQI levels based on the extracted features. The key contribution in this research is the incorporation of three outlier detection techniques using DBSCAN, LOF, and Isolation Forest, which enable the model to identify and accurately classify extreme pollution events, improving its robustness in challenging scenarios. The model was validated on two datasets, achieving an average precision, recall, and F1-score of 0.97, and an overall accuracy of 96% on the original dataset. For the unseen dataset, the model achieved an average precision of 0.77, recall of 0.75, and an F1-score of 0.75, with an overall accuracy of 77% demonstrating its generalizability across different data distributions. Confusion matrices were also evaluated, providing additional insight into model performance across AQI categories. These results show that the model can effectively classify AQI levels across diverse environmental conditions and offer a more scalable, reliable solution for real-time air quality monitoring. The scope of this study focused on leveraging image-based data for AQI classification, specifically addressing the challenges of capturing pollution-specific patterns and detecting outliers in urban environmental images. By overcoming the limitations of current image-based techniques, this model shows promise for wider application in environmental monitoring and public health initiatives. Future work will focus on expanding the dataset, improving outlier detection methods, and using temporal and weather information to enhance model accuracy and real-time performance.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful. Alhamdulillah, all praises to Allah for giving me the opportunity and simplifying this path to complete this thesis.

First and foremost, I would like to express my deepest gratitude to my main supervisor Dr Azliza Mohd Ali for her unwavering support, insightful guidance and constant encouragement throughout the course of my research. I am grateful to her for introducing me to such an interesting area of research, and for all the help and encouragement she has given me over the past three years. The comments she has made on initial drafts of this thesis have also been invaluable.

I would also like to extend my sincere thanks to my co-supervisors, Dr. Rozianawayt Osman and Dr. Suhaili Beeran Kutty, for their continuous support and constructive feedback. Their valuable input and dedication have greatly contributed to the successful completion of this project.

Additionally, there have been countless others who have assisted and advised me along the way. I am deeply appreciated of the support from my colleagues and friends whose encouragement and assistance have made this journey possible.

Last but certainly not least, I have to thank my parents Ahmad Razimi Abdul Ghani and ----- . Without my parents' constant love and support I would certainly not be where I am today, and for that I cannot thank them enough. Finally, I must thank my husband, Nik Masrul Hasyimi Bin Nik Muhammad and to our precious daughters and son, Nik Hannani Binti Nik Masrul Hasyimi, Nik Humayrah Binti Nik Masrul Hasyimi and Nik Anas Bin Nik Masrul Hasyimi. This piece of victory is dedicated to all of you. Alhamdulillah.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF SYMBOLS	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER 1 INTRODUCTION	1
1.1 Research Background	1
1.2 Motivation of This Research	2
1.3 Problem Statement	3
1.4 Research Questions	5
1.5 Research Objectives	6
1.6 Scope and Limitation	6
1.7 Research Significant	9
1.8 Thesis Outline	10
CHAPTER 2 LITERATURE REVIEW	12
2.1 Introduction to Air Quality Index (AQI)	13
2.1.1 Definition and Importance of AQI	13
2.1.2 AQI Calculation	14
2.1.3 The Impact of AQI on Public Health	15
2.2 Current methods of AQI monitoring	16
2.2.1 Ground-Based Sensors	17
2.2.2 Manual Sampling	17
2.2.3 Air Quality Monitoring in Malaysia	17
2.3 Image-Based Approaches for Air Quality Monitoring	19

CHAPTER 1

INTRODUCTION

This chapter provides the motivation for the research, describes the problem statement, and presents the research questions and objectives addressed by this research. It also exposes the deficiencies of currently available Air Quality Index (AQI) classification approaches and point out the necessity to address them through an effective implementation of strong image-based techniques combined with outlier detection for enhanced accuracy and scalability. Moreover, it emphasizes the limitations and scope of the research and highlights the significance of improving AQI classification systems and its role in advancing more accurate and scalable air quality monitoring methodologies. By addressing these aspects, the chapter sets the stage for developing innovative methodologies to enhance Air Quality Index (AQI) classification.

1.1 Research Background

Air pollution is the release of pollutants into the air in the vicinity of people and/or the environment that are detrimental to health and the overall environment. These pollutants stem from natural sources like wildfires and dust storms or human activity, including industrial processes, transportation and energy generation. Air pollution is a major cause of respiratory and cardiovascular diseases, neurological impairments and developmental problems in children, as well as having negative environmental impacts on wildlife and ecosystems due to biodiversity loss and disruption of ecological balance (Cozma et al., 2022).

The rapid urbanization on course to affect 68% of the global population by 2050 further complicates air pollution problems, especially in developing countries (Abdulganiyu et al., 2023). Air quality monitoring plays an important role in reducing risks to public health and environmental degradation. However, there are certain limitations in traditional air quality monitoring systems based on remote sensing or sensor methods such as high cost, limited spatial coverage and weather dependency (Ahmed et al., 2022). This has led to increasing interest for alternative methodologies