

UNIVERSITI TEKNOLOGI MARA

**PREDICTOR MODEL FOR *Aedes*
MOSQUITO WITH
MICROCLIMATE EFFECT USING
MACHINE LEARNING
ALGORITHMS**

NUR ATHEN BINTI MOHD HARDY ABDULLAH

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Environmental Health & Safety)

Faculty of Health Sciences

June 2025

ABSTRACT

Dengue remains a major public health threat in Malaysia, with a marked rise in cases and fatalities in 2023. While machine learning has been widely applied for dengue outbreak prediction, fewer studies have focused on predicting adult mosquito abundance, the primary vectors of dengue, particularly at the community level. This study aimed to forecast the abundance of adult female *Aedes* mosquitoes using microclimatic data and machine learning models, including Neural Network (NN), Support Vector Regression (SVR), Random Forest Regression (RFR), and Multiple Linear Regression (MLR). Two study sites were selected: Kolej Jasmine (a dengue hotspot in urban Bukit Raja, Petaling) and Kolej Dahlia (a non-dengue hotspot in suburban Jeram, Kuala Lumpur), both occupied by university staff and students. Environmental (temperature, relative humidity, rainfall), entomological (Adult Index [AI], Dengue Positive Trap Index [DPTI]), and epidemiological (dengue case) data were collected, with 20 gravid oviposit sticky (GOS) traps installed at each site and NS1 kits used to detect dengue virus in trapped mosquitoes. Only *Aedes albopictus* (*Ae. albopictus*) was found at the non-dengue site, while both *Aedes aegypti* (*Ae. aegypti*) and *Ae. albopictus* were present at the dengue hotspot. Models were developed using two approaches: (i) time lagged microclimate only predictors and (ii) combined time lagged microclimate and AI or DPTI, with the latter consistently yielding higher prediction accuracy. The best AI prediction models was SVR for *Ae. albopictus* at the non-dengue site ($R^2 = 0.904$, MAE = 0.125, RMSE = 0.177), MLR for total *Aedes* ($R^2 = 0.944$, MAE = 0.156, RMSE = 0.240) and *Ae. aegypti* ($R^2 = 0.929$, MAE = 0.081, RMSE = 0.108) at the dengue site, and SVR for *Ae. albopictus* at the dengue site ($R^2 = 0.904$, MAE = 0.125, RMSE = 0.177). For DPTI prediction, SVR performed best for total *Aedes* ($R^2 = 0.553$, MAE = 3.594, RMSE = 4.397) and *Ae. aegypti* ($R^2 = 0.557$, MAE = 2.384, RMSE = 3.074), while MLR was best for *Ae. albopictus* ($R^2 = 0.326$, MAE = 2.228, RMSE = 2.635). Key microclimate predictors varied by species and site, with rainfall and temperature being most influential overall. These results demonstrate the potential of machine learning to enhance mosquito surveillance and support proactive vector control by predicting oviposition activity and guiding timely interventions.

ACKNOWLEDGEMENT

Firstly, I wish to thank God for giving me the opportunity to embark on my PhD and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Prof Madya Ts. Dr. Nazri Che Dom and Prof Madya Sr Dr Siti Aekbal Salleh.

My appreciation goes to the staff of Kolej Dahlia and Kolej Jasmine who provided the facilities and assistance during sampling. Special thanks to my colleagues and friends for helping me with this project.

I would also like to express my heartfelt gratitude to my family for their unwavering support, encouragement, and understanding throughout this journey. Their patience and understanding have been a constant source of strength.

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	xii
LIST OF ABBREVIATIONS	xv
 CHAPTER 1 INTRODUCTION	 1
1.1 Research Background	1
1.2 Problem Statement	3
1.3 Hypothesis	6
1.4 Research Questions	6
1.4.1 To profile the distribution of dengue hotspots in Selangor from 2017 to 2021	6
1.4.2 To determine the environmental, entomological, and epidemiological factors at the study sites and explore their association	6
1.4.3 To develop mosquito predictor models for total <i>Aedes</i> , <i>Ae. albopictus</i> and <i>Ae. aegypti</i> using machine learning algorithms microclimate factors	7
1.5 Research Objectives	7
1.5.1 To profile the distribution of dengue hotspots in Selangor from 2017 to 2021	7
1.5.2 To determine the environmental, entomological, and epidemiological factors at the study sites and their association	7

CHAPTER 1

INTRODUCTION

1.1 Research Background

Dengue is an endemic disease in 128 countries throughout the world (Usman *et al.*, 2016; Schmidt *et al.*, 2011). Malaysia has experienced dengue since 1902 and this disease continues to become a public health risk since the 1970's, with its first major outbreak in 1973 (Cheah *et al.*, 2014; Mohd Zaki *et al.*, 2014). This disease is transmitted through the bite of an infected *Aedes* mosquito and is commonly caused by four closely related serotypes of the dengue virus which are DENV- 1, DENV-2, DENV-3, and DENV-4. The fifth serotype of dengue, DENV-5, has been detected in Sarawak, Malaysia during screening of viral samples taken from a 37 year-old farmer admitted in hospital in the year 2007 (Vasilakis *et al.*, 2011). Initially, the infection in the farmer was thought to be a typical case of DENV-4 until virus isolation and complete genetic sequencing were performed (Normile, 2013). This virus was phylogenetically distinct from the three previous forms of sylvatic DENV-4 and bore some similarity with DENV-2 (Normile, 2013). Although the exact cause of DENV-5 transmission remains unknown, several factors may contribute to its emergence, including genetic changes from sylvatic strains to human strains (Vasilakis *et al.*, 2011), the high mutation rate of DENV (Monath, 1994), and extensive deforestation (Rudel *et al.*, 2009). The emergence of DENV-5 serotype indicates the potential for additional viral variants to emerge in the future. Additionally, dengue can present with a wide spectrum of clinical symptoms, from mild fever to severe haemorrhagic manifestations, which can be fatal if not treated properly.

Controlling the mosquito population is widely regarded as one of the most effective methods to prevent the transmission of arboviruses to humans, especially given the limited availability of effective antiviral treatments or vaccines for many of these viruses (Wilson *et al.*, 2020; Roiz *et al.*, 2018). In recent year, dengue predictor model has been extensively developed to assist in dengue surveillance (Sylvestre *et al.*, 2022; Leung *et al.*, 2021) and to help governments in controlling this disease (Batista *et al.*, 2021). A study by Sylvestre *et al.* (2022) on the systematic review of dengue prediction using data-driven method had revealed that 60% of 119 studies