CHAPTER 1

INTRODUCTION

1.1 Research gap

At present, the prediction models concerning the Malaysian region are scarce. Currently, most studies mainly focus on the Northwest Pacific, while the models for the sea areas around Malaysia (0°–10°N, 95°–120°E) are nearly nonexistent, and the research in this region is also limited.

Meanwhile, traditional prediction methods have limitations. Traditional numerical models are computationally complex and more sensitive to initial conditions, failing to achieve real-time early warning.

Finally, there is insufficient modeling. Traditional statistical methods such as linear regression are unable to capture the complex nonlinear relationship between typhoon paths and environmental variables (such as sea temperature gradients and terrain obstruction).

1.2 Problem Background

Tropical cyclones (TCs) are regarded as extreme weather events, along with gales, rainstorms, and storm surges, which can cause huge losses in coastal areas worldwide.(Chen, R., Zhang, W., & Wang, X. 2020) It will exert a considerable

influence on the residents' housing, people's property, urban construction, road traffic and other economic constructions in coastal areas. In Southeast Asia, particularly in the surrounding sea areas of Malaysia, the generation and trajectory of tropical cyclones are characterized by complexity and uncertainty, which renders predicting the landing point of tropical cyclones an extremely challenging task. *The prediction of the intensity, location and time of the landfall of a tropical cyclone well advance in time and with high accuracy can reduce human and material loss immensely.(Kumar, S., Biswas, K., & Pandey, A. K. 2021)*

Traditional tropical cyclone trajectory prediction primarily depends on conventional numerical models (e.g., WRF, ECMWF), yet their computational complexity is considerable and they are highly sensitive to initial conditions, thereby failing to satisfy the timeliness demands of disaster emergency responses. Moreover, studies on tropical cyclones in the surrounding waters of Malaysia are scarce. The existing models mostly concentrate on the Northwest Pacific or the North Atlantic, leading to insufficient analysis of the crucial factors in the waters around Malaysia and inadequate consideration of the specific geographical and climatic conditions of the Malaysian sea area.

1.3 Research Questions

How to construct a prediction model for the landing points of tropical cyclones in the waters adjacent to Malaysia based on the random forest algorithm and with the utilization of IBTrACS data (outputting longitude and latitude coordinates)?

Which factors affect the prediction of tropical cyclone landfall points in Malaysia?

1.4 Research Objectives

Construct a random forest model: Based on the tropical cyclone trajectories and environmental variables in the IBTrACS dataset, develop a prediction model for tropical cyclone landing points in the surrounding sea areas of Malaysia (with the output of longitude and latitude coordinates).

Analysis of Key Influencing Factors: Through feature importance analysis and nonlinear modeling, quantify the dynamic influence of sea surface temperature, topography and atmospheric conditions on the landing locations of tropical cyclones.

1.5 Research Scope

1、In terms of the geographical range, the tropical cyclone data from the surrounding sea areas of Malaysia (latitude and longitude range: 0° - 10°N, 95° - 120°E) are employed, with the tropical cyclone data from other regions excluded.

2、Regarding the definition of landfall, a tropical cyclone is regarded as having landed if the distance between its center and the coastline of Malaysia is no more than 50 kilometers. The potential landfall points on the east coast (the side facing the South China Sea) are of key concern.

3、Concerning the prediction duration, with the tropical cyclone entering the study area (latitude 0°–20°N, longitude 100°–120°E) as the starting point, the landing point within the next 72 hours is predicted.

4、Employing the relevant data in the IBTrACS dataset, a tropical cyclone landing point prediction model is established.

5、The impacts of key factors, such as climate and terrain, on the landing points of tropical cyclones will be analyzed.

Geographical Range	Malaysian waters (0° - 10°N, 95° - 120°E).
Landfall Definition	Tropical cyclones landing within 50 km of the Malaysian coastline
Forecast time	Landfall predictions up to 72 hours in advance
Dataset	IBTrACS dataset https://www.ncei.noaa.gov/products/international-best-track-archive
Tools	Python programming language with Anaconda environment and IDEs like Jupyter Notebook

1.6 Research Contribution

Theoretical Contribution: It augments the studies on tropical cyclones in the Malaysian context.

Methodological Contribution: A novel model of the random forest algorithm for tropical cyclone prediction has been constructed.

Enhancing disaster early warning capabilities: The model can assist the Meteorological Department of Malaysia in optimizing the prediction of tropical cyclone landing points, locking high-risk areas 48 hours in advance, and reducing economic losses in coastal communities due to heavy rainfall and tropical cyclone landings.

Data/Tool Contribution: It provides a reusable dataset and model framework for future studies.

1.7 Thesis Organization

Chapter1: Introduction. The background and significance of tropical cyclone landfall points, the methods hypothesized in this article, and the issues to be addressed, etc.

Chapter2: Literature Review. An analysis of current prediction methods and their deficiencies, as well as the assistance offered by the literature to this paper.

Chapter3: Construction of Random Forest Model. Comprising dataset screening, data selection, data preprocessing, etc., and detailed descriptions of the construction of the random forest model.

Chapter4: Analysis of Influencing Factors. An analysis of the influence of various climatic factors on landfall points.

Chapter5: Results, Discussion and Conclusion. Perform visualization of the results, expound the insights on the significance and limitations of the obtained results, summarize and present suggestions for subsequent research.