CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Research Conclusion

This research endeavors to leverage machine learning techniques to develop a prediction model for the landfall points of tropical cyclones in Malaysia and its adjacent seas. Through an in - depth analysis of the spatial distribution characteristics of tropical cyclones and the trends of environmental variable changes, this paper presents a dual - output modeling framework grounded in the Random Forest Regression model. This framework is designed to predict the latitude and longitude coordinates of cyclone landfalls separately.

Regarding data, this study relies on the IBTrACS tropical cyclone track database. From this database, nearly 15,000 valid observation records spanning from 2000 to 2025 were extracted. In the aspect of feature selection, this paper primarily focuses on two core variables: wind speed (Wind_kts) and central pressure (Pressure_mb), which are used to characterize the intensity of tropical cyclones. To safeguard the generalization ability of the model, a training/test data ratio of 80% : 20% was employed for model construction and evaluation.

The experimental results demonstrate that the model achieves relatively satisfactory performance in latitude prediction. The mean absolute error (MAE) is 5.51°, and the root - mean - square error (RMSE) is 6.09°. Conversely, in longitude prediction, the errors are relatively larger, with an MAE of 10.75° and an RMSE of 14.22°. These findings suggest that the model is more robust in latitude - direction prediction, with a relatively rapid error convergence. In contrast, due to the stronger spatial heterogeneity in the longitude direction, the prediction exhibits greater volatility.

The key contributions of this research can be summarized as follows:

1. Propose and implement a dual - output cyclone landfall prediction model, which still attains favorable results even under the scenario of limited data.

2.Establish a prediction workflow grounded in IBTrACS data, encompassing data cleaning, exploratory data analysis (EDA), model construction, evaluation, and visualization.

3.Furnish a fundamental modeling paradigm for meteorological forecasting and disaster management in the Malaysian region, demonstrating significant potential for generalization and engineering applications.

The conclusions drawn in this chapter lay a practical foundation for the improvement suggestions and future extensions presented in the subsequent chapters.

5.2 Limitations of the Study

Although the random forest model developed in this study has achieved certain results in predicting the landfall points of tropical cyclones, there are still several limitations in the following aspects. These factors may potentially impact the model's generalization ability and prediction accuracy:

1. Limited data dimensions and feature selection

The feature variables employed in this research are confined to wind speed (Wind_kts) and air pressure (Pressure_mb). While these two variables are closely associated with the intensity of cyclones, they are insufficient in characterizing the environmental context. Variables such as sea surface temperature, vertical wind shear, and relative humidity, which play pivotal roles in the formation and path evolution of cyclones, are not incorporated.

In future research, the integration of ERA5 reanalysis data can be considered to further enhance the feature representation capabilities.

2. Absence of high - resolution spatial information

The current model input does not differentiate specific geographical location features. As a result, it fails to account for the micro - regulatory effects of factors such as terrain influence and coastal line morphology on the cyclone's landing path. The model's prediction results are relatively coarse - grained, rendering it challenging to achieve precise predictions for small - scale regions.

Introducing geographical coordinate grids or spatial raster coding techniques could be explored to improve the model's spatial discrimination ability.

3. Relatively simple model structure with no consideration of time dependency

Although the random forest model demonstrates strong capabilities in modeling static input variables, it is unable to capture the time - series characteristics of cyclone movement, such as trends in path changes and variations in speed.

In the future, time - series modeling approaches based on architectures such as Long Short - Term Memory (LSTM) or Transformer can be investigated to supplement the model's dynamic prediction capabilities.

4. Potential bias in the definition method of landfall points

In this study, the combined thresholds of wind speed and latitude - longitude are used to define the first "landfall point" of a cyclone. However, this method may not comprehensively capture actual geographical landfall events, leading to labeling errors that can affect the quality of model training.

Integrating on - site reports or typhoon bulletin data can be considered to optimize the criteria for determining landfall.

5.3 Future Research Directions

Building upon the practical insights gained from this study and considering the limitations elaborated previously, future research endeavors can be extended and refined in the following directions to enhance both the performance of the model and the depth of the research:

1. Integration of multi - source meteorological data for enhanced feature representation

This research has predominantly relied on IBTrACS data. In future investigations, the incorporation of high - resolution reanalysis datasets such as ERA5 and MERRA2 is advisable. By integrating these with crucial environmental variables including sea surface temperature, wind shear, humidity, and sea - level pressure, a more comprehensive feature set can be formulated.

This approach facilitates the model's learning of the physical mechanisms driving the development paths of tropical cyclones.

2. Exploration of the performance advantages of deep learning and ensemble models

Although random forests exhibit stability in scenarios with limited sample sizes, in the context of larger - scale datasets, the introduction of neural network architectures such as Convolutional Neural Networks (CNNs), Long Short - Term Memory networks (LSTMs), or Transformer models can be contemplated to extract spatial and temporal information. Moreover, the construction of ensemble models (e.g., a combination of Random Forests, XGBoost, and Artificial Neural Networks) can be pursued to enhance the model's robustness and generalization capabilities.

This would augment the model's proficiency in fitting non - linear and temporal relationships. 3. Development of a high - spatiotemporal resolution prediction system

In future research, the scope of the prediction output can be extended from a single landfall point to a landfall path or area probability. Leveraging rasterization techniques in conjunction with Geographic Information Systems (GIS), dynamic modeling of the cyclone path and landing zone can be achieved, thereby enabling more actionable disaster prevention and mitigation strategies.

This aligns more closely with the practical requirements of real - world early warning systems.

4. Establishing a link with real - world disaster loss data for practical model evaluation

Future studies can correlate the model's output with real - world impact data, such as flood occurrences, building damage, and population evacuation in Malaysia and its neighboring regions. This exploration aims to uncover the practical significance of cyclone landfall point prediction in emergency response and resource allocation.

This strengthens the model's practical guiding significance for society.

5. Creation of a user - friendly visualization platform

To facilitate the practical application of research findings, the development of a web - based interactive visualization platform is envisioned. This platform can dynamically present prediction results, error margins, and path trend diagrams, offering valuable support to government meteorological agencies and disaster response organizations.

This transition enables the model to progress from theoretical research to practical implementation.

5.4 Summary

This chapter offers a comprehensive synthesis of the principal findings, limitations, and prospective research directions of this study.

This research established a regression model centered on the random forest algorithm, leveraging the IBTrACS dataset to predict the landfall locations of tropical cyclones in the coastal areas of Malaysia. The findings indicate that, even under the constraints of a small sample size and limited feature dimensions, the random forest model can effectively discern the non - linear relationships among cyclone intensity, position, and environmental variables, thereby yielding relatively satisfactory prediction outcomes. The low mean squared error and stability exhibited by the model validate its practical feasibility.

Nonetheless, the study confronts several challenges, such as restricted data dimensionality and the absence of external variables (e.g., sea surface temperature, wind shear). Moreover, the model's output represents a single landfall point, failing to account for the dynamic evolution of tropical cyclone trajectories.

In light of these discoveries, this chapter puts forward several forward - looking avenues for future research. These include integrating multi - source meteorological data, exploring deep learning models, developing spatio - temporal prediction systems, correlating with real - world disaster data, and creating visualization platforms.

These endeavors will lay a solid foundation for enhancing the accuracy, adaptability, and practical utility of the model, and will also furnish data and algorithmic support for the establishment of meteorological disaster warning systems.

In summary, this study not only demonstrates the potential of machine learning applications in tropical cyclone prediction but also provides a theoretical framework and viable research pathways for future related investigations