Predicting the Western Pacific tropical cyclone trajectories using various deep learning approaches

MAN HIM LI

School of Computer Science

University of Adelaide

A thesis submitted for the degree of

Master of Data Science

Supervised by:

Dr. Huangying Zhan

Dr. Rafael Felix Alves

June 2022

This thesis paper is dedicated to my institutional supervisor, Dr. Huangying Zhan for constantly giving me valuable advice along my final year of the postgraduate degree. Thanks for enlightening me with all data science and machine learning knowledge. It has been a tough year for all of us being optimistic to learn and grow during the pandemic.

Acknowledgements

Personal

First of all, I would like to express my sincere gratitude to my lovely and devoted wife, Maggie Zhou for taking care of me and supporting me during my final period of the postgraduate degree. Without you, I would not be able to accomplish the thesis.

Institutional

School of Computer Science, the Faculty of Sciences, Engineering and Technology, The University of Adelaide, Australia

Abstract

In recent years, utilising deep learning algorithms in typhoon track predictions has become a very popular task due to more online public data accessible nowadays in different weather organisation repositories. The development of typhoons and their moving pattern tendencies are considered as a complex meteorological process that makes the weather agencies hardly optimise both prediction efficiency and accuracy together at the same time. The ensemble models utilised by the worldwide weather agencies for typhoon prediction are calculated by a supercomputer which requires a few hours to deliver a prediction in the next 24 hours timeframe. The implementation of a deep learning algorithm for typhoon prediction is still at the developing stage that it is yet to replace the position of a supercomputer on typhoon track prediction in terms of accuracy. However, the prediction efficiency from a deep learning model can be a contribution to the task as it only requires limited training time and prediction results can be obtained instantly depending on the real-time data availability. This project proposed different deep learning frameworks on real historical typhoon data extracted from the NOAA archive and the Kitamoto laboratory data repository. Various deep learning approaches were conducted to predict the next 3 hours and 6 hours of typhoon tracks in the Western Pacific Region. The experimental result showed that the LSTM regression model and the CNN-LSTM ensemble model delivered a similar prediction result toward the 3-hour prediction, however, the CNN-LSTM ensemble model performed better on 6 hours prediction compared to other deep learning approaches conducted in this project. The CNN-LSTM ensemble model could be a potential deep learning approach to deliver a more robust prediction result if diverse image data type was used to extract relevant track feature in the model.

Contents

List of Figures

List of Tables

1 Introduction

Contents

1.1 Introduction

Tropical cyclone is one of the natural disasters bringing destructive impacts to the coastal areas. It usually develops in the subtropical region such as the Western Pacific Region, Australia, India and the Gulf Coast of North America. The severe weather brought by typhoons has imposed threats to millions of lives. The presence of the El Niño effect has a significant influence on the typhoon intensity. During the El Niño year, the sea surface temperature of the Pacific Ocean rises abnormally and the Pacific Walker Circulation is pushed eastward towards the American continent. When a typhoon forms over the Pacific Ocean during the El Niño year, it usually has a longer travel distance over the ocean before it reaches a continent L. Wu et al., 2019. Therefore, the intensity of the Western Pacific's typhoons during the El Niño year is relatively stronger. The last El Niño year was 2018-2019. There were 26 typhoons formed in the Western Pacific Ocean in 2018 which was the year exceeding the mean occurrence of typhoons in a year L. Wu et al., 2019.

In 2018, Super typhoon Mangkhut imposed catastrophic damage to the Western North Pacific Region, mainly the Philippines and Southern China. It was developed in the western North Pacific Ocean on 7 September 2018 when it was initially classified as a category 1 tropical storm C. W. Choy, Lau, and He, 2020 at the early stage. However, it quickly became a category 5 super typhoon with an average wind speed reaching 287 km/hr over the next few days due to environmental conditions favouring typhoon development C. W. Choy, Lau, and He, 2020. It gradually headed in a northwestern direction towards Luzon of the Philippines and thereafter the Pearl River Delta. Eventually, more than 4 million people in 21 cities were directly suffering from the destructive damage caused by this super typhoon in the western North Pacific region Niu et al., 2020.

Even though the substantial damage caused by Mangkhut was inevitable, there were no reported fatalities in Hong Kong C.-w. Choy, M.-c. Wu, and Lee, 2020. Hong Kong was sufficiently prepared for the incoming super typhoon that a Standby Signal No. 1 was issued to the public by the Hong Kong Observatory in advance to enhance awareness*Super Typhoon Mangkhut(1822)* 2018. The prevention was contributed by the sophisticated track prediction models C.-w. Choy, M.-c. Wu, and Lee, 2020.

The Hong Kong Observatory utilized its existing ensemble prediction system which had successfully predicted the track of Mangkhut a week before cutting across Luzon C. W. Choy, Lau, and He, 2020. It suggested that the intensity of Mangkhut would be reduced after passing across the terrain of Luzon and it would then move towards the western part of the Pearl River Delta instead of directly landing over Hong Kong. Figure 1.1 and Figure 1.2 illustrated the forecast track of Mangkhut predicted by the ensemble models and the actual track of Mangkhut over the western North Pacific region respectively. Therefore, the accuracy of tropical cyclones tracks prediction has been playing a very important role to minimise damage from the impacted regions.

The National Hurricane Centre provided a summary of different types of models *NHC Track and Intensity Models* 2019 being used nowadays to predict tropical cyclone tracks and intensities. There are three categories including statistical models,

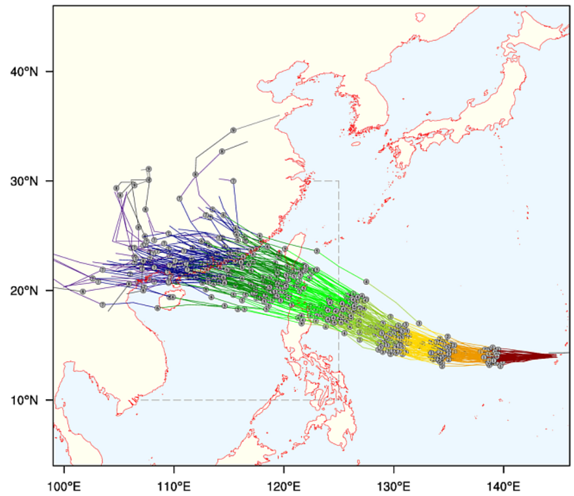


Figure 1.1: Predicted track of tropical cyclone Mangkhut by various ensemble models utilised by the Hong Kong Observatory on 11 September 2018. AU YEUNG, 2018

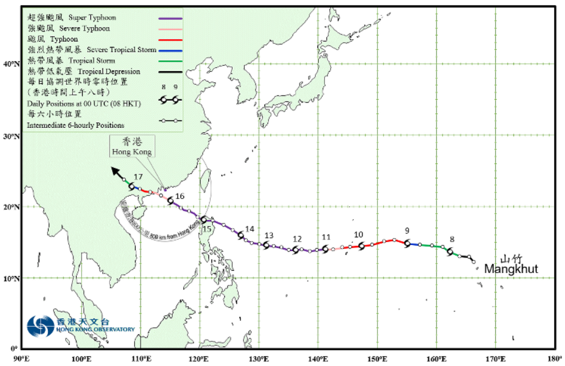


Figure 1.2: Actual track of tropical cyclone Mangkhut over South China Region C.-w. Choy, M.-c. Wu, and Lee, 2020

dynamic models and ensemble models. A dynamic model is the most accurate model with the input of a large number of current and historical meteorological variables in a supercomputer to perform high-level real-time predictions. This model usually requires several hours of running on a supercomputer. A statistical model, however, simply runs on an ordinary computer to study the relationship behaviour of the tropical cyclones based on the historical data. An ensemble model is a collection of multiple statistical and dynamic models which performs high-level prediction of meteorological events.

The ensemble prediction system used by the Hong Kong Observatory is a combination of different global dynamic models delivering a set of possible outcomes of typhoon tracks and intensities C. W. Choy, Lau, and He, 2020. The system includes the European Centre for Medium-Range Weather Forecasts (ECMWF), the UK Met Office (EGRR) and the National Centers for Environmental Prediction (NCEP) C. W. Choy, Lau, and He, 2020. These models are run by supercomputers to generate a list of possible typhoon tracks and intensities predictions. Different models usually have different predicted tracks as shown in figure 1.

As more historical data is available over the internet nowadays, deep learning models have become more popular to make time-series predictions without using dynamic models in a computationally expensive supercomputer Rui, Zhang, and Wang, 2020. The deep learning approach has the capability of making a single prediction within a few seconds in an ordinary computer for a complicated non-linear event such as predicting tropical cyclone trajectories Alemany et al., 2018. The concept of a time series regression model in deep learning is to learn and predict the moving pattern in a time series event instead of just calculating statistical relationships from a conventional statistical model Alemany et al., 2018.

Apart from implementing a single LSTM model in a regression task, ensemble learning of deep learning is another approach that combines multiple deep learning models to deliver a better generalisation prediction performance. This project has implemented a simple LSTM regression model to predict sequential typhoon trajectories, a fused CNN-LSTM model to learn features from the satellite images and meteorological data for track prediction and lastly using a sequence classification approach with an LSTM network.

2

Motivations

Contents

2.1 Availability of data

Nowadays, the availability of massive historical data on the internet benefits the utilization of deep learning algorithms. There are various meteorological data publicly available online on different data repository websites. For instance, the International Best Track Archive for Climate Stewardship project from the National Oceanic and Atmospheric Administration provided a collection of tropical cyclones data collected by different worldwide weather agencies Hennon, 2021. In addition, satellite typhoon image data is available online from the Kitamoto Laboratory developed by Dr Kitamoto from the Graduate University for Advanced Studies (Sokendai)Asanobu, 2022.

2.2 Nowcasting and Forecasting

The frequency and intensity of tropical cyclones over the Western North Pacific region have been increasing in the 21st century Ying et al., 2012. Even though the presence of tropical cyclones is inevitable, the improving accuracy of track prediction technology allows weather organizations to dispatch warning messages to the public in advance to prepare for incoming natural disasters. The ensemble models have been widely used by various weather organizations to deliver a high-level prediction of tropical cyclones trajectories. However, it requires a supercomputer running for several hours to make a prediction. For instance, Japan Meteorological Agency updates tropical cyclone activity and forecasts every 3 to 6 hours to cover the next 24 hours track information Agency, 2021. The prediction accuracy of ensemble models is not replaceable in that the tracking error predicted by the supercomputer is less than 80 km for 24 hours track forecast. However, the deep learning model can contribute to the typhoon prediction task at the time the supercomputer is calculating the next predictions. The most significant advantage of using a deep learning model is its prediction efficiency which only requires a few seconds to summarise a track prediction in an ordinary computer Alemany et al., 2018. Since the moving pattern of a tropical cyclone is non-linear and fast-changing, implementing a deep learning algorithm can become a useful approach to nowcast the typhoon moving patterns within a short period. Nowcasting and forecasting typhoon track in deep learning model and ensemble model can become a useful combined tool to help track typhoon moving pattern effectively.

2.3 LSTM Regression model

This project implemented various deep learning approaches to predict typhoon moving patterns. The long short-term memory neural network (LSTM) was the primary neural network to be used differently to compare which approach was the best fit for the current research. The traditional recurrent neural network (RNN) model has been widely used in predicting a sequential event, however, the

RNN model has limited long-term memory dependency that far early neurons have no connections to the final output which is caused by an issue of vanishing gradientsLian et al., 2020. An LSTM model contains a cell state and hidden state to store useful long-term and short-term memory respectively which helps resolve the vanishing issue. As predicting tropical cyclones trajectories has a consistent dependency on both short-term and long-term historical data, utilising the LSTM model in this project was more suitable.

2.4 CNN-LSTM ensemble learning

A convolutional neural network has been widely used for image recognition. A CNN model consists of a convolutional layer, pooling layers and a fully connected layer. The last layer is usually associated with a softmax function to deliver a set of probability for classification prediction. The biggest advantage of using the CNN model is its capability of learning spatial patterns from the input image Giffard-Roisin et al., 2020. A set of feature vectors can be extracted from the last fully connected layer to represent the input image. This project used the pretrained Resnet18 model to extract features from a typhoon satellite image to increase feature dimension in an ensemble learning approach.

2.5 LSTM Classification model

LSTM neural network can also be used for classification in a time series prediction task Pham, 2021. An example of using an LSTM model in a classification task is sentiment analysis in natural language processing. The architecture combines a sequential learning structure from the LSTM and the spatial pattern learning capability from a CNN model to deliver a set of probabilities based on prediction. This project used an LSTM classification model as one of the deep learning approaches to forecasting future typhoon tracks. Instead of predicting latitude and longitude directly, a bin index was used to represent a specific degree of latitude and longitude. Further implementation details will be discussed in the section of the proposed methodology.

2.6 Target variables

The latitude and longitude of tropical cyclones were used in this project as the target variable. The biggest motivation for using longitude and latitude as output was because of its direct representation of typhoon track coordination on Earth. Some other researchers, for example, Alemany et al., 2018 utilized a grid LSTM model to predict tropical cyclone tracks in terms of grid blocks which represented 1 degree of latitude times 1 degree of longitude. However, a conversion marginal error of 50 km could be made when the final grid locations were converted back to latitude and longitude. In addition, latitude and longitude are also used as input features as they have indirect meaning to the strength of Coriolis force on Earth and topography which have significant impacts on the motion of tropical cyclones ITO et al., 2020. Therefore, this project used latitude, and longitude with other features as a set of input to various deep learning models to predict typhoon moving patterns.

2.7 Useful tool for future research

This project implemented various deep learning models trained in an ordinary computer to make typhoon track predictions. The meteorological and satellite data and the link to the data archive and repository will be uploaded on Github as a reference for future researchers to continue optimising deep learning models in the typhoon track prediction task. Everyone can scrap the data and train their deep learning models without using a supercomputer for typhoon analysis.

3 Literature Review

The recent El Niño year was in 2018 that was regarded as one of the most active typhoon seasons in the decade. The El Niño effect was the main contributor to bringing a favourable environment for typhoon formation. The monsoon trough over the Pacific Ocean moved eastwards which favoured a stronger development of tropical cyclone L. Wu et al., 2019. In 2018, there were 26 official typhoons formed over the Western Pacific Ocean L. Wu et al., 2019. Between September and November 2018, there were five Category 5 super typhoons formed in this region. They were super typhoon Mangkhut, Kong-Rey, Yutu and Trami and this typhoon dataset was used for final testing in this project for track predicting visualisation purposes. Especially the super typhoon Mangkhut in 2018 brought the most destructive damage and economic loss to the Philippines and Hong Kong for the last three decades Choy, M.-c. Wu, and Lee, 2020. However, there were no reported fatalities in Hong Kong that primarily contributed to the significant improvement of the current typhoon track prediction systems.

The accuracy of predicting tropical cyclone tracks has been improving for decades due to the advancement of statistical-dynamic models. However, this traditional prediction methodology usually requires high computational hardware and a long-running time. More importantly, these models normally make track predictions without analysing whether the input meteorological features are correlated to the moving motion of a tropical cyclone which highly enhances unnecessary calculating time Lian et al., 2020. As more advanced data collection technologies are available today, such as data collected from satellites, ocean stations and ground stations, predicting tropical cyclone trajectories has become a big data analysis task Lian et al., 2020. The next important step was to collect the right typhoon features for track predictions.

Recurvature of a tropical cyclone is defined as the original track of a storm system veering from its original westward direction to the northward and eastward in the Northern Hemisphere PAN, 2011. The effect of curvature could be linked to various meteorological factors. ITO *et al.* ITO et al., 2020 studied various factors affecting the motion of tropical cyclones. The beta gyres effect generated a steering effect for the tropical cyclone due to the interaction between the pressure gradient force on Earth and the surrounding wind circulation. Researchers ITO et al., 2020 also mentioned sea level temperature, locations and sizes of tropical cyclones had direct causation to their motion. Moreover, the topographical factor such as the presence of terrain had a direct impact on the curvature and deformation of tropical cyclones. However, some of the meteorological features might not have a direct relationship to the displacement of latitude and longitude along the typhoon lifecycle.

Sun *et al.* Youqiang Sun et al., 2014 delivered research on evaluating the granger causality test in a multivariate time series forecast analysis. The result showed that the granger causality test was able to select essential features with appropriate lagged values in a time series forecast. This paper compared the prediction performance between the granger causality test with lagged values and other proposed methods such as PCA-based and causality test without lagged values. It showed that the granger causality test with lagged values delivered the best predicting performance with the minimum value of root mean square error implemented in different regression models.

The Granger causality test has also been used in different scenarios to predict various meteorological events. Cermak and Bodri Cermak and Bodri, 2016 proposed a granger causality analysis to look for a causality relationship between ground air surface temperature and the possibility of precipitation in different time scales. The test successfully delivered promising evidence that there was a direct causal relationship between the ground air temperature difference on an hourly time scale and the possibility of precipitation.

Therefore, a granger causality test will be the first step in this paper to study the causal relationship between different meteorological features and tracks of tropical cyclones before implementing a deep learning algorithm for prediction.

Researchers have been working on predicting tropical cyclone trajectories using lightweight deep learning algorithms compared to dynamic models used by weather organizations in a supercomputer. Song *et al*. Gao et al., 2018 utilized an LSTM neural network to predict tropical cyclone tracks based on the historical data from the North China Sea Marine Forecasting Center of the State Oceanic Administration of China. The LSTM model provided a satisfactory performance in predicting typhoon track in the next 6 to 24 hours with the prediction error up to 45 km to 105 km respectively. Another experiment from this paper emphasized the importance of using a large size of the training dataset to deliver a high-quality prediction performance. Since only track information was utilized as input in this paper, Song *et al*. Gao et al., 2018 suggested that other features such as typhoon centre pressure, wind speed and surrounding temperature could be included as other input features in the LSTM model to enhance prediction accuracy.

Apart from using a single deep learning model for typhoon track prediction, Lian *et al*. Lian et al., 2020 proposed a combination of CNN and GRU models to predict tropical cyclone tracks. A granger causality test was also implemented to study the causality relationship between features and tropical cyclone tracks. The multiple layers CNN model selected essential features which had a direct impact on the motion of typhoons. The extracted features were sent to the GRU model to predict a non-linear trajectory in the next 12 to 72 hours time step. The overall accuracy of this model architecture in the paper outperformed other traditional deep learning approaches.

Yang *et al*. Giffard-Roisin et al., 2020 proposed another fused deep learning methodology of using the Convolutional and LSTM neural networks to combine computer vision and time series forecast algorithms to predict the tropical cyclone tracks in the next 24 hours time step. Two 3D CNN models were used for capturing wind and wind pressure features from moving frame typhoon images to predict latitude and longitude displacements. The fused ensemble learning model in this paper delivered a single prediction within a few seconds. The performance of this fusion model performed better than the existing statistical model used by National Hurricane Center.

Li Yongjiao Sun et al., 2021 implemented a multitask learning and feature weighting framework for typhoon track prediction. The dataset was 6 years of the Western Pacific's typhoon dataset from 2001 to 2005 obtained from the Joint Typhoon Warning Center data archive. A ResNet and LSTM model were used to extract features from heatmap images which represented different meteorological data, such as humidity and surface sea temperature. The feature vector extracted from the ResNet model was manually weighted based on the correlation of the feature to the typhoon moving pattern. The combined vector with all data was sent to the LSTM model to predict typhoon intensity and trajectory. Eventually, the mean position error in the 6-hour prediction from the proposed model was 38.75 km.

Alemany *et al*. Alemany et al., 2018 implemented a grid-based recurrent neural network model which studied the tropical cyclone moving behaviour from grid to grid. This paper utilized different features such as latitude, longitude, wind speed and pressure as input to predict typhoon trajectories. The data in this research was derived from the National Hurricane Center and the result of 6 hours interval of predicted tracks was delivered in the result. However, the grid-based model induced a significant marginal error of as much as 50 km when the grid location was converted back to latitude and longitude in the final step. Gustav and Xuechen Edholm and Zuo, 2018 compared the overall performance between a grid-based LSTM model and a conventional LSTM model. Both models delivered similar prediction performance as no one model had a significant statistical advantage over the other. Alemany *et al.* Alemany et al., 2018 also emphasized the biggest advantage of using a deep learning approach to predict tropical cyclone trajectories over the ensemble models is time efficiency which could only take a few seconds to make on prediction. Apart from just using a recurrent neural network, some other researchers utilized a fused deep learning approach to predict tropical cyclone tracks.

A different network was used by Ruttgers et al., 2022 for typhoon track and intensity prediction. Jeon Ruttgers et al., 2022 implemented a Generative Adversarial Network and utilised satellite images and reanalysis of meteorological data for typhoon prediction. The meteorological data such as wind velocity, surface pressure, sea surface temperature, humidity and cloud cover were included in the dataset. The final model delivered a 44.5 km average distance error for 6-hour predictions

Chan *et al.* Tong et al., 2022 and Li *et al.* Wang et al., 2022 both utilised satellite images to explore the typhoon intensities. Chan *et al.* Tong et al., 2022 extracted features, such as cloud cover, typhoon eye, eye walls and area of rain bands to predict the intensities and strength of typhoons in the North Pacific Region. Li *et al.* Wang et al., 2022 used infrared satellite images with different colour channels representing cloud cover, sea surface temperature, air temperature etc. to extract different features to predict typhoon intensity.

Apart from tropical cyclone predictions, the LSTM model has been used in predicting different meteorological events. Qing and Niu Qing and Niu, 2018 imposed an LSTM model to predict solar irradiance in the next 10 hours. The selected features were date, temperature and dew points etc. The proposed LSTM model was considered the most outstanding prediction model compared to other regression models in terms of the least loss function and less overfitting issue. Therefore, the LSTM model has been widely used by many scholars for weather predictions. Before the implementation of a deep learning algorithm, a features selection technique is important to select useful meteorological variables which efficiently reduce the complexity of our proposed LSTM model to make predictions.

4
Data

Contents

4.1 Choosing dataset

Reliability and accuracy of the typhoon dataset back to the 20$^{\text{th}}$ century contained uncertainties when surveillance technology had not been well developed which might lead to data bias. As more advanced technologies have been continuously invented in the last two decades, the measuring ability starts to become more accurate. Therefore, this project is going to utilise the dataset from the year 2000 onwards which contains 548 tropical cyclones from the Western Pacific Region. The selected data was divided into three different sets. 80% of the data was used for training, the other 10% was used for validation and the remaining 10% of the dataset was used as a testing dataset for final evaluation associated with the predefined performance metric.

Basin	Western Pacific	Western Pacific	Western Pacific	Western Pacific	Western Pacific	...
ISO TIME	6/9/2018 18:00	6/9/2018 21:00	7/9/2018 00:00	7/9/2018 03:00	7/9/2018 06:00	...
NATURE	TS	TS	TS	TS	TS	...
LATITUDE	12.00N	12.10N	12.20N	12.28N	12.40N	...
LONGITUDE	169.3E	168.6E	168.0E	167.3E	166.6E	...
DISTANCE TO LAND	2467km	2443km	2427km	2402km	2380km	...
WIND SPEED	20kts	20kts	20kts	22kts	25kts	...
PRESSURE	1007mb	1007mb	1007mb	1005mb	1004mb	...
R34NE	0kts	0kts	0kts	0kts	0kts	...
R34SE	0kts	0kts	0kts	0kts	0kts	...
R34SW	0kts	0kts	0kts	0kts	0kts	...
....	...	...	...	...	...	...
R64NW	0kts	0kts	0kts	0kts	0kts	...

Table 4.1: An example of partially recorded data of Super Typhoon Mangkhut

4.2 Meteorological Data

The meteorological data of historical tropical cyclones in the Western Pacific regions are publicly available on the repository of the International Best Track Archive for Climate Stewardship affiliated with the National Oceanic and Atmospheric Administration(NOAA). An example of a Western Pacific's tropical cyclone data extracted from the IBTrACS database Hennon, 2021 is shown in Table 9.1. Each column of data illustrated the typhoon status 3 hours after the previous timestamp. The feature includes time occurrence in UTC, typhoon nature, latitude, longitude, distance to land, wind speed, wind pressure, and maximum radius wind in four quadrants in 34, 50 and 64 miles away from the centre of the typhoon which measures a radius distance between the centre of a tropical cyclone and the location of the strongest wind. The example in Table 9.1 indicates zero in all quadrants of the maximum radius wind due to this example showing an initial development timeline of typhoon Manghkut. At the later timestamp, the maximum radius of wind gradually increases as the typhoon starts to develop into a supertyphoon.

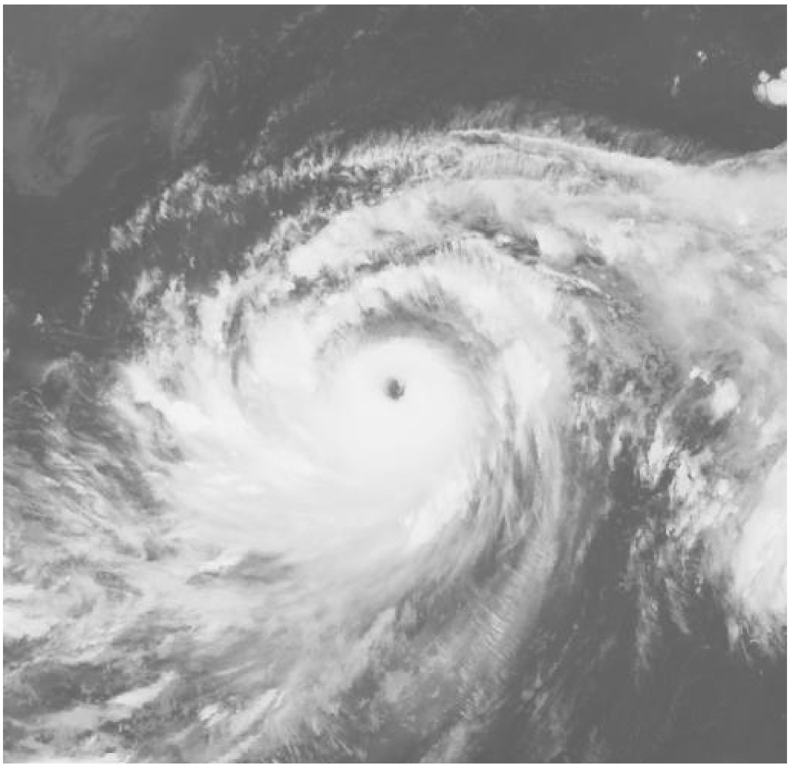


Figure 4.1: An example of a typhoon image scrapped from the Kitamoto Laboratory

4.3 Image Data

Typhoon satellite image data will be used in a pretrained convolutionary neural network ResNet-18 in this project to extract image features for track prediction. The typhoon satellite image data was scrapped from the Kitamoto Laboratory which was established by a Japanese university professor, Dr KITAMOTO, Asanobu Asanobu, 2022. The image data repository of the Kitamoto Laboratory contained satellite images of the Western Pacific Region's typhoons since 1951. One satellite image in the dataset represents the feature in typhoon size, the existence of eye and cloud cover appearing in one specific timestamp. This project scrapped 30,850 grayscale satellite images from the repository to match the timestamps of an entire dataset containing 548 typhoons for model training. However, the repository does not fully cover all the satellite images for all typhoons with all timestamps. A zero NumPy array will be assigned to the dataset with no satellite images. An example of a typhoon satellite image was shown in Figure 4.1.

5 Proposed Methodology

Contents

5.1 Introduction

This project utilised three different deep learning approaches to predict typhoon trajectories. The first model was a simple LSTM model using meteorological data to predict typhoon moving patterns in terms of latitude and longitude. The next model was a combination of CNN and LSTM models trained by a set of meteorological data and satellite image data for track prediction. The final model also utilised the LSTM network which was similar to the first model, but the output layer of this final model was a set of bin indexes instead of predicting latitude and longitude directly. The final model could be seen as a classification model predicting bin location representing a specific latitude and longitude. Before moving on to the model implementation, deciding loss function, an exploratory data analysis, statistical test and data pre-processing were needed to be done in a machine learning pipeline for data cleansing.

5.2 Deciding Loss function and Evaluation Metric

Mean square error(MSE) was used in this project as a loss function for model optimisation based on the validation loss. The lower the validation loss, the better a model could generalise the prediction track in this project. MSE measured the total sum of the difference between the actual and predicted values. However, MSE becomes biased when the difference value is high. Therefore, root means square error becomes a more effective final evaluation metric to prevent the influence of a large error value. In this project, RMSE measured the square root of the difference between the actual observed track and the predicted track of tropical cyclones in a degree of latitude and longitude. Since every degree of latitude and longitude is equal to 111 kilometres on Earth, the final value of RMSE was converted into mean position error by multiplying 111 by the RMSE value.

5.3 Exploratory Data Analysis

Exploratory data analysis is considered as a critical step to help researchers better understand the collected data, such as detecting missing data, data anomalies and recognising data distribution. Graphs and tables were used in this session to visualise the statistical information of the dataset. The statistical summary and graph representation in the exploratory data analysis helped us to have a basic idea of whether the dataset was ready for further implementation.

5.4 Statistical test - Granger causality test

The Granger causality test is a hypothesis test to mathematically determine whether one time series of variables is applicable to predict the other time-series data. This project utilised the granger causality test to study whether a set of meteorological features had a direct causality related to the typhoon movement in latitude and longitude. One of the common purposes of using the Granger causality test in previous research was for feature selection, which had been conducted by Youqiang Sun et al., 2014. Since data implemented to a granger causality test had to be stationary, checking data stationarity was the first step to be executed before the granger causality test. The procedure for implementing data to the test in this project is shown below:

1. Inspection of any missing values in the time series dataset.
2. Setting null hypothesis, a lagged time series variables x_{t} did not have granger causality to the tropical cyclone tracks y_{t}.
3. Applied Augmented Dickey-Fuller test to check whether the time series data is stationary.
4. If two or more time-series data were not stationary, a cointegration Johansen test was used to test whether these non-stationary time series data had constant equilibrium in the long term which is considered stationary.
5. If time series data either passed the Augmented Dickey-Fuller test or cointegration Johansen test, the next step was to determine lagged values by using the Autocorrelation function to check for correlation of time series data with different time steps.
6. Applied F test to obtain a p-value for lagged time series variables.

7. Whichever the p-value of a time series data was lower than 0.05, the null hypothesis was rejected and confirmed the granger causality relationship.

8. Selected this time series variable as one of the features to predict tropical cyclone track in an LSTM model.

5.5 Data Preprocessing

Data Preprocessing was another essential pipeline in machine learning to remove data inconsistency and made the input data ready for model implementation. The general procedure of data preprocessing is shown in the list below.

1. Replaced all missing values with -1

2. Data normalisation feature by feature with the application of MinMaxScaler() imported from the sklearn library

3. Reshaped the data dimension and data structure for implementation in the Pytorch framework.

4. Grouped the data sequentially in a sliding window for predicting future typhoon track in 3 hours and 6 hours. Data from 5 timestamps in the past were grouped to predict the next timestamp and two timestamps in the future for 3 hours and 6 hours prediction respectively. That could be seen as using the last 15 hours of typhoon data to predict the next 3 and 6 hours of the location of the typhoon track.

5. Created a dictionary to store meteorological data, image data and ground truth target variables.

6. Read the image data and assign it to a dictionary. A zero NumPy array was assigned to the dataset with no satellite images.

7. Split data into training 80%, validation 10%, and test dataset 10% accordingly based on the timeline. That means the first 16 years of the typhoon from 2000 to 2016 were used for training, the next 2 years 2017-2019 for validation and finally 2 years 2020-2021 for testing.

8. Extracted four individual typhoon datasets as extra testing data for visualisation purposes. The typhoon dataset included super typhoon Manghkurt in 2018, KONGREY in 2018, YUTU in 2018 and WUTIP in 2019. Since these four typhoons were all category 5 supertyphoons that occurred in the El Nino year in 2018-2019, a prolonged life cycle of these four typhoon datasets benefited the track visualisation result on the map.

5.6 Model 1 - LSTM Regression model

5.6.1 Concept and Definition

Long Short Term Memory neural network is an advanced version of a conventional recurrent neural network model in a time series forecasting task. The gradient vanishing and exploring problem are resolved in an LSTM model due to its sophisticated architecture to regulate long-term and short-term dependencies. There are three main components in the LSTM architecture which include the input gate, forget gate and output gate. The input gate is responsible for deciding which new relevant value should be added to the memory. Forget gate is responsible for filtering what information is irrelevant from the previous state. The output gate is responsible for conglomerating all information from the input gate and forget gate in the current timestamp to decide an output for the next timestamp. LSTM model consists of hidden state and cell state. The hidden state is responsible for regulating short-term memory while the cell state is responsible for conveying long-term memory in each LSTM layer.

Forget Gate

$$f_t = \sigma(W_f \times x_t + U_f \times h_{t\text{-}1} + b_f) \tag{5.1}$$

Forget gate in the LSTM determines what time-series data from the previous timestamp should be transferred to the current timestamp. The function of forget gate is shown in equation 5.1. The sigmoid function is used in this gate to make the f_t output value between 0 and 1. Every memory will be forgotten if f_t is zero and every memory will be kept if f_t is 1. The output f_t will then multiply against the cell state $C_{t\text{-}1}$ in the previous timestamp to obtain a new cell state in the current timestamp which is shown in equation 5.4.

Input Gate

$$i_t = \sigma(W_i \times x_t + U_i \times h_{t\text{-}1} + b_i) \tag{5.2}$$

$$\overline{C_t} = tanh(W_c \times x_t + U_c \times h_{t\text{-}1} + b_c) \tag{5.3}$$

$$C_t = f_t \times C_{t\text{-}1} + i_t \times \overline{C_t} \tag{5.4}$$

Input gate is to decide what new important data should be added as a new memory. The *tanh* function is used in the input gate to gather information from the previous hidden state $h_{t\text{-}1}$ and new input x_t at the current timestamp t as shown in equation 5.3. The output C_t will be positive when the new input is relevant to be added as a new memory and it will be negative when it is irrelevant. The output C_t from equation 5.3 will then combine with the cell state from the previous timestamp $C_{t\text{-}1}$ to obtain a new cell state C_t which is demonstrated in equation 5.4.

Output Gate

$$o_t = \sigma(W_o \times x_t + U_o \times h_{t\text{-}1} + b_o) \tag{5.5}$$

$$h_t = o_t \times tanh(c_t) \tag{5.6}$$

$$Output = Softmax(h_t) \tag{5.7}$$

Output gate is responsible to gather all information from the input gate and forget gate to deliver a relevant output. Similar to forget gate, the sigma function from equation 5.5 is used in this gate to make the output value o_t between 0 and 1. The value delivered from the output gate o_t will then multiply against the *tanh* function of long term memory C_t. Therefore, the long term memory C_t and the recent output o_t are included in the hidden state at current time t. The final output will be associated with the Softmax function applied to hidden state which is shown in equation 5.7.

5.6.2 Prediction Methodology

Data sequence Given a multiple time series data representing historical tracks and meteorological features of different tropical cyclones as input $X = [T_1, T_2, T_3, \dots T_n]$, where each tropical cyclone included a time series of latitude, longitude and other selected meteorological variables. This project was going to train all the variables in the previous timestamp as input in an LSTM model to predict the latitude and longitude of the next timestamp as output. The idea was to make the LSTM model learn the patterns of tropical cyclones' moving motion based on the historical data and to forecast the future track of prevailing tropical cyclones in terms of spatial coordinates on Earth. A multidimensional matrix representing one tropical cyclone is shown in a matrix equation 5.8 below where each row represents input features including latitude, longitude and various meteorological variables and each column represents data sequence in ordered timestamps from 1 to n.

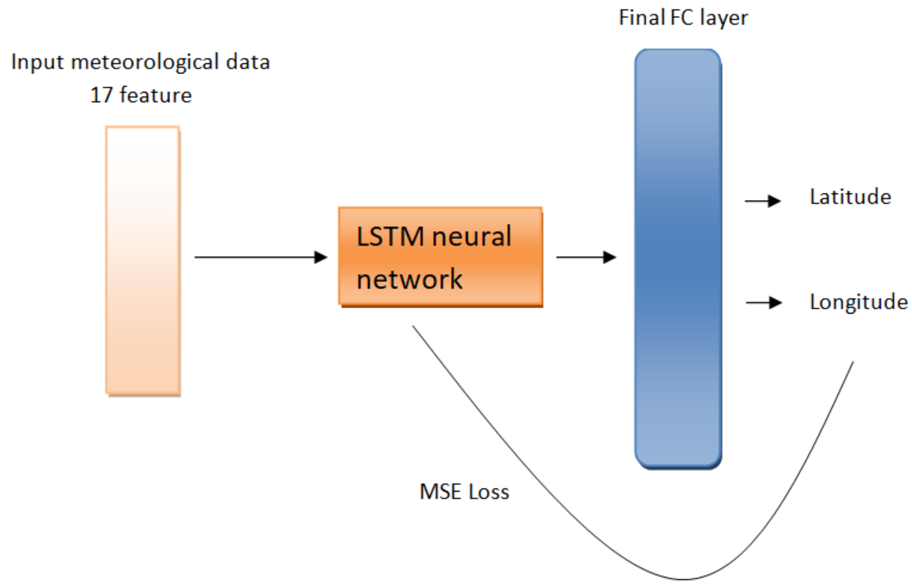


Figure 5.1: The LSTM Regression model architecture

Figure 5.1 visualises the general architecture of the LSTM regression model to predict latitude and longitude in this project.

$$T_{\mathrm{n}} = \begin{bmatrix} lat_1 & long_1 & MetA_1 & ... \\ lat_2 & long_2 & MetA_2 & ... \\ lat_3 & long_3 & MetA_3 & ... \\ ... & ... & ... & ... \\ lat_{\mathrm{n}} & long_{\mathrm{n}} & MetA_{\mathrm{n}} & ... \\ \end{bmatrix} \tag{5.8}$$

5.7 Model 2 - CNN-LSTM ensemble model

5.7.1 Concept and Definition

The Convolutional Neural network (CNN) is one of the reputed deep learning models specialised in image feature recognition. A convolution layer in the CNN model is capable of capturing spatial features to differentiate a unique image from the others. A filter used in the convolution layer to capture image features is also called a kernel. The kernel shifts side way along the input image based on the parameter setting of stride deciding the shifting distance. The RGB value from the image captured by the kernel then performed a matrix multiplication to deliver a set of feature outputs which can differentiate from the other non-identical images. Apart from using for image classification, the CNN model can also be implemented for delivering the image feature vector from the last fully connected layer before applying the softmax function. A pretrained ResNet-18 CNN model was used in this project to capture typhoon satellite images to expand the feature dimensions for sequential learning in the LSTM model. Instead of extracting the final output from a softmax function layer, a vector is extracted from the last fully connected layer of the pretrained model to represent part of the typhoon features implemented in an LSTM regression model. Sets of grayscale typhoon satellite images were used as the input image data in this project.

5.7.2 Prediction Methodology

The extracted image feature from the pretrained model delivered a vector containing 512 features with some numerical values. Since the original meteorological features had only 17 features, the feature dimension imbalance in this situation would affect the prediction capability in the LSTM model as the image feature dominated the sequential feature to predict latitude and longitude in the next timestamp. Therefore, the dimension of the original meteorological features was firstly expanded from 17 to 512 in a fully connected layer constructed in the LSTM neural network and then combined with the image vector to form an extended feature vector having a dimension of 1,024 features to represent the current feature in a specific timestamp. The pretrained ResNet-18 model and the LSTM model were trained simultaneously with the optimisation function to enhance the generalisation performance for typhoon track prediction. Figure 5.2 visualises the general architecture of the CNN-LSTM ensemble model to predict latitude and longitude in this project.

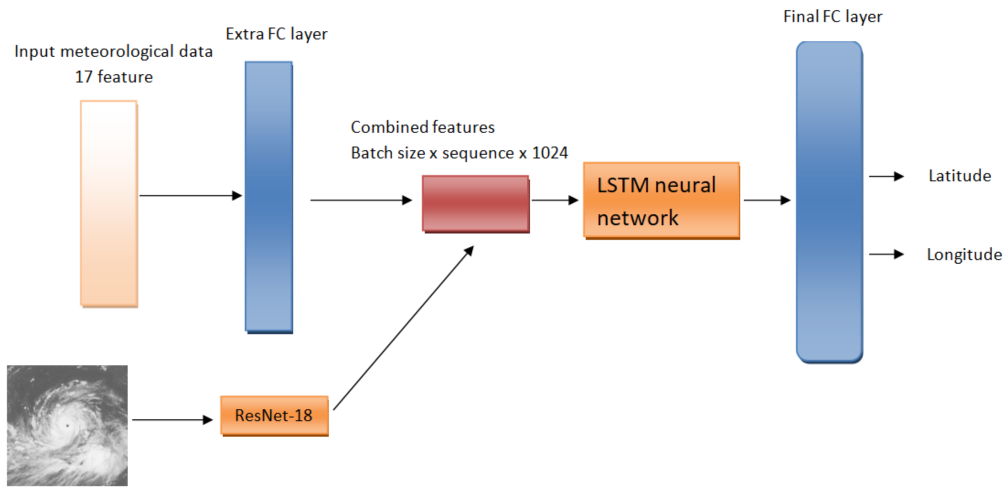


Figure 5.2: The CNN-LSTM model architecture

5.8 Model 3 - LSTM Classification model

5.8.1 Concept and Definition

Forget gate and input gate in the LSTM neural network have significant contributions to memorising important sequential data. This memorising property can also be applied in a classification task. The last output layer in this LSTM classification model was modified to deliver an unnormalised vector indicating the score for each predicted class. A Cross-Entropy loss function is used here to calculate the difference between two probability distributions, the actual and the predicted distribution. The loss function delivers training loss and validation loss in each epoch and gradually decreases towards the global minima associated with the optimiser.

5.8.2 Prediction Methodology

The LSTM classification model predicted the bin index of latitude and longitude instead of predicting the value of latitude and longitude directly. Each bin index represented a range of latitude and longitude. For instance, the maximum latitude and the minimum latitude from the dataset of this project were 51.3 and 1.3 respectively. If this range of latitude was divided into 1000 bins each bin representing 0.05 degrees. For example, 23 degrees of latitude would be represented as 460^{th} bin index. The predicted output from the LSTM classification model was a set of

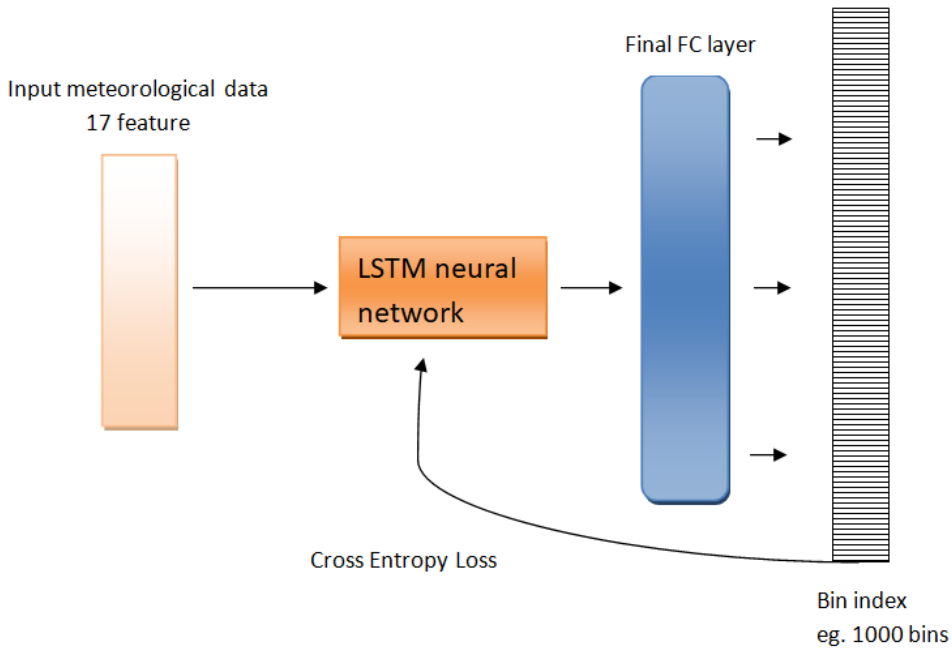


Figure 5.3: The LSTM classification model architecture

raw and unnormalised vectors indicating scores for each class. This output from the LSTM model became the input in the Cross-Entropy loss function to compare with the label of latitude and longitude bin index which was converted from the original latitude and longitude degree. The model was trained until the validation loss has reached the global minima in different hyperparameters setting. RMSE is still the primary evaluation metric for the classification model to calculate the mean position error between the actual and predicted track. The predicted score distribution from the classification model is then converted back to the degree of latitude and longitude to measure the difference between the actual latitude and longitude. Figure 5.3 visualises the general architecture of the LSTM classification model to predict bin index for latitude and longitude.

5.9 3 hours and 6 hours track predictions

This project was going to utilise three different models to predict typhoon tracks in the next 3 and 6 hours. The models' architectures for predicting 3 hours and 6 hours were the same except for the sliding window setting. The sliding window used for predicting 3 hours typhoon track was set as using 15 hours of data in the past to predict 3 hours in the future, and the other window used 15 hours of data in the past to predict 6 hours in the future. Three models with two sliding window settings were trained separately to enhance generalisability performance for predicting the future 3 hours and 6 hours of the typhoon track.

5.10 Hyperparameters optimisation

The objective of hyperparameter searching for deep learning models was to resolve the overfitting issues and minimise validation loss to reach global minima. The searching hyperparameters included batch size, number of hidden layers and learning rate of the optimiser in this project. Optimiser 'adam' was used consistently in the project because of its capability of handling sparse gradients to resolve vanishing gradient issues.

5.11 Models Evaluation

RMSE was the primary performance metric in this project for evaluating the generalising ability of a model in typhoon track predictions. An optimised model was evaluated with the 10% testing dataset and four individual typhoon testing dataset which was extracted initially for track visualisation in the next section.

5.12 Track visualisation

Four individual typhoon datasets of super typhoon Manghkurt in 2018, KONGREY in 2018, YUTU in 2018 and WUTIP in 2019 were initially extracted as extra testing data for visualisation in this section. The predicted tracks from four individual typhoon testing datasets predicted by all three models were visualised in a map through the Matplotlib Basemap toolkit in Python. The predicted track and the actual track were plotted on the map for visualisation which was easier to capture pattern differences.

6
Experimental Result

Contents

6.1 Exploratory Data Analysis

Exploratory data analysis was a preliminary step to capture data inconsistency and abnormality. It helped detect errors that lead to producing adverse prediction performance in a deep learning model. The dataset used in this project contained 548 Western Pacific tropical cyclones having an official name given by the Japan

Meteorological Agency. There were a total of 31,699 timestamps and each timestamp represented one set of typhoon data. The time difference between the two timestamps was separated by 3 hours apart. The overall distribution of each meteorological feature in the dataset was represented in the graph below.

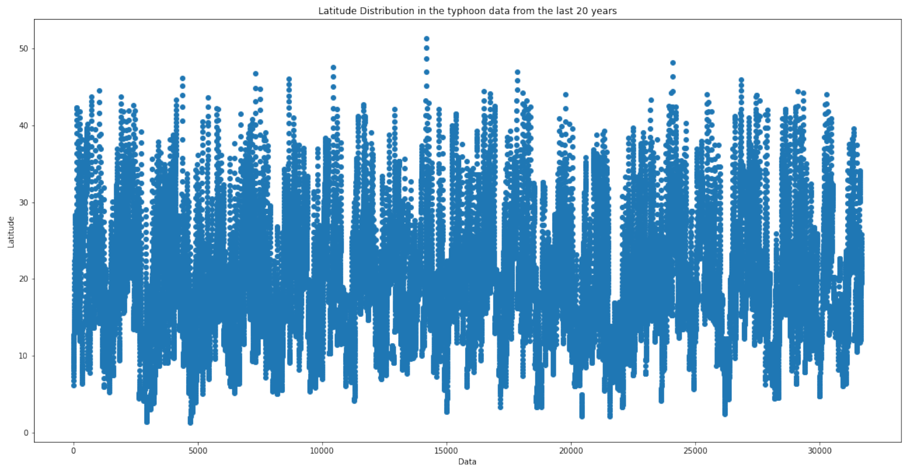


Figure 6.1: Latitude distribution in the entire dataset

Figure 6.1 illustrated the overall latitude distribution for all Western Pacific typhoons in the last 21 years. Most of the tropical cyclones in this region lied at the latitude between 10° to 30° North of the equator. The maximum latitude degree obtained in this dataset was 51.3° North of the equator. It was very rare to have a Western Pacific typhoon moving beyond 50° North of the equator where the sea temperature in this area is not favourable for typhoon development.

Figure 6.2 visualised the longitude distribution along with the typhoons' timestamps in the last 21 years. Some of the data appeared as a negative value but the majority of data was clustered within 100° to 180° East of the meridian. The presence of a negative longitude degree in this dataset was because of the international date line on the projection map. When a typhoon moves beyond to the east of the international date line, the longitude becomes a negative value after passing the dateline at 180°. That was the reason why the maximum value in the current cluster in Figure 6.2 was 180° and some values were appearing as negative values in the

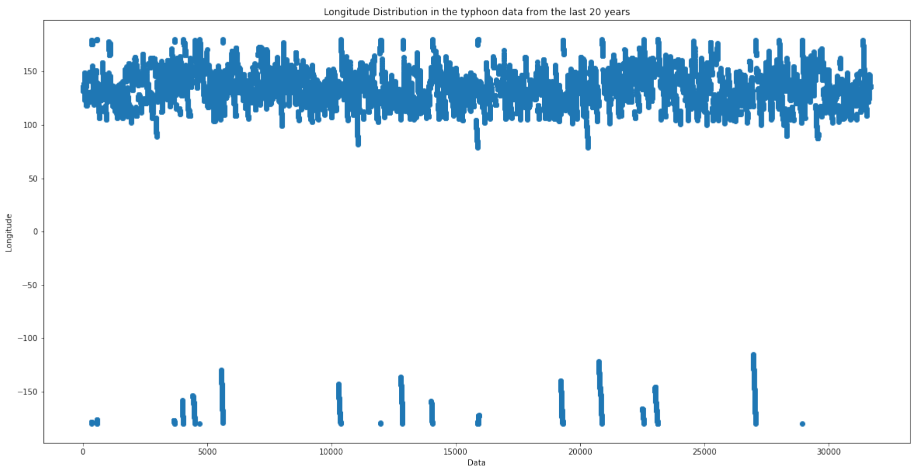


Figure 6.2: Longitude distribution in the entire dataset

dataset. The solution to convert the negative values back to positive values was to add 360° to the negative longitude data. Otherwise, the prediction performance of longitude from deep learning models would be adversely affected.

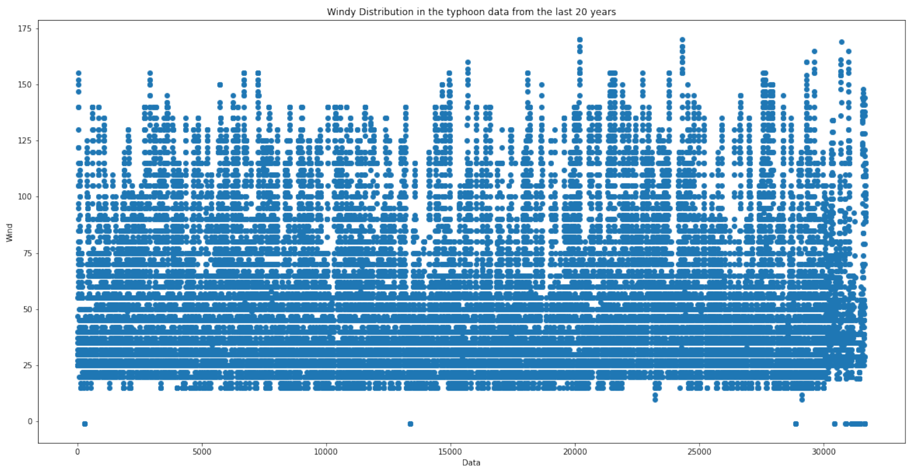


Figure 6.3: Wind distribution in the entire dataset

Figure 6.3 illustrated the wind distribution of the Western Pacific's typhoons from 2000 to 2021. The overall wind speed distribution of typhoons usually fell within the range of 25 to 100 knots. Some occasional existences of stronger typhoons might reach beyond 125 knots. The decision line to classify a category 5 super typhoon is beyond 157 knots.

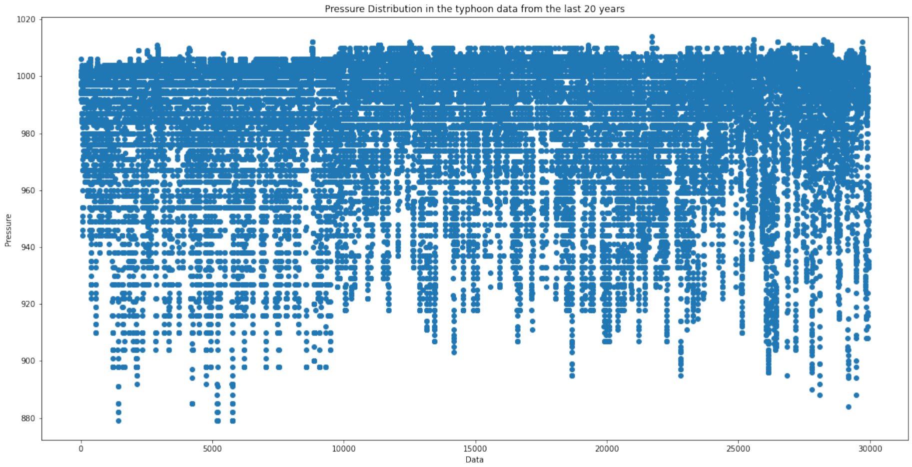


Figure 6.4: Pressure distribution in the entire dataset

Figure 6.4 showed how the pressure of typhoons distributed over the last 21 years of the typhoon in the Western Pacific Region. The stronger the typhoon, the lower the pressure was recorded. It was uncommon to record the wind pressure of a typhoon as less than 900 millibars unless the occurrence of a supertyphoon. The decision line of classifying a category 5 super typhoon is below 920 millibars. Therefore data in figure 6.4 mostly clusters within the range of 900 - 1000 millibars.

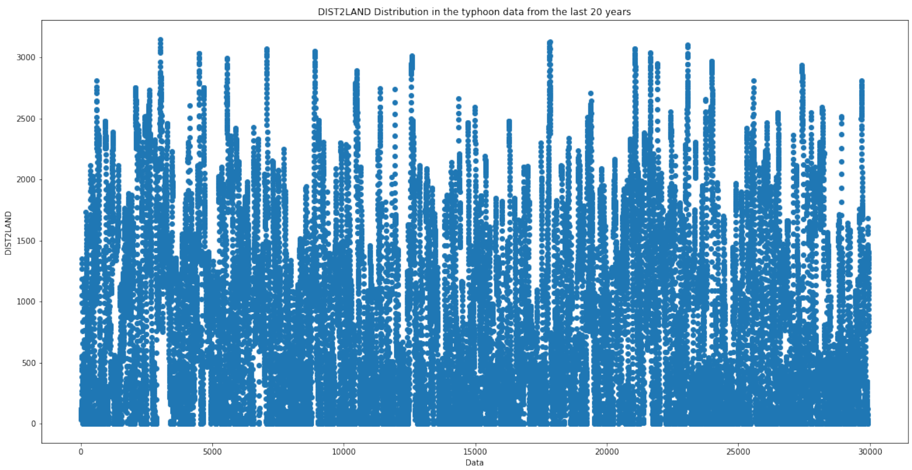


Figure 6.5: Distance to Land distribution in the entire dataset

Distance to land was a numerical data indicating how far was the location of the typhoon away from the closest continent. It was an important feature having a direct influence on a typhoon's moving pattern. When a typhoon moved across a continent, its original track would be deflected due to the absence of water vapour to contain its energy. Most of the typhoons in the Western Pacific Region were usually surrounded by continents, such as China, Japan, Malaysia and the Philippines. Figure 6.5 illustrated it was not common to have a distance greater than 2000 km away from any continent in the Pacific Region. As the typhoon moved towards any continent, the distance away from a continent was also decreasing until it becomes zero when it landed on a continent.

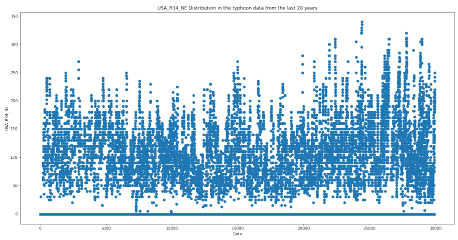


Figure 6.6: MRW N/E 34 miles

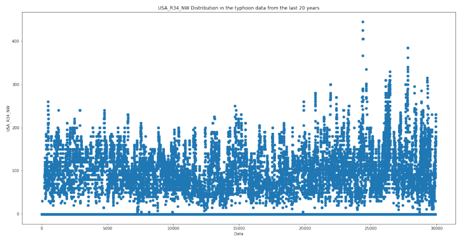


Figure 6.7: MRW N/W 34 miles

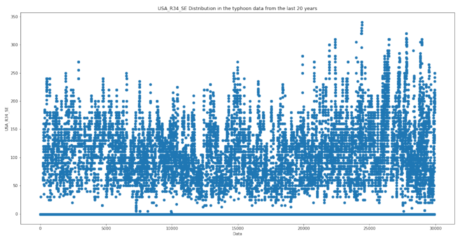


Figure 6.8: MRW S/E 34 miles

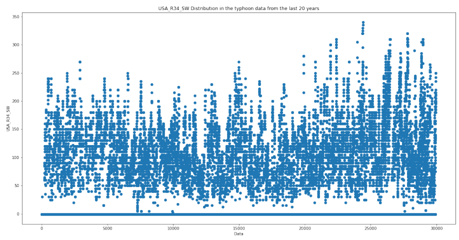


Figure 6.9: MRW S/W 34 miles

Figure 6.6, 6.7, 6.8 ,6.9 ,6.10, 6.11, 6.12 ,6.13,6.14, 6.15, 6.16, 6.17 distributed the maximum radius of wind (MRW) in four quadrants of direction, Northeast, Northwest, Southeast and Southwest having a distance of 34, 50 and 64 miles away from the center of typhoon eye. MRW data could be seen as a feature representing typhoon strength. The stronger a typhoon, the larger radius of wind would be recorded at a distance away from the typhoon eye due to its bigger typhoon size.

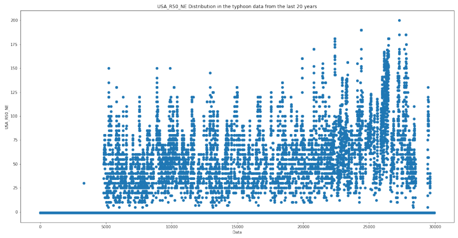


Figure 6.10: MRW N/E 50 miles

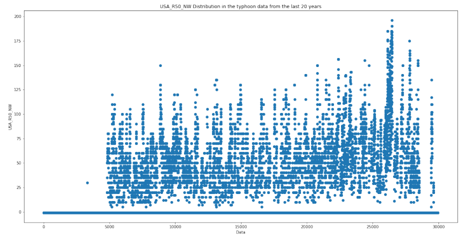


Figure 6.11: MRW N/W 50 miles

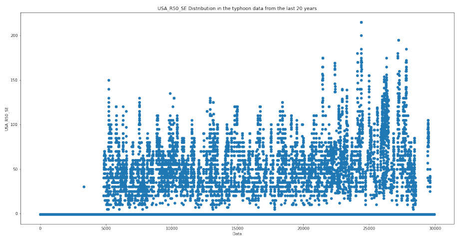


Figure 6.12: MRW S/E 50 miles

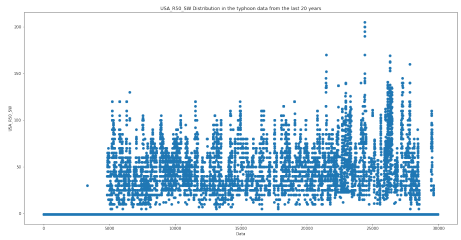


Figure 6.13: MRW S/W 50 miles

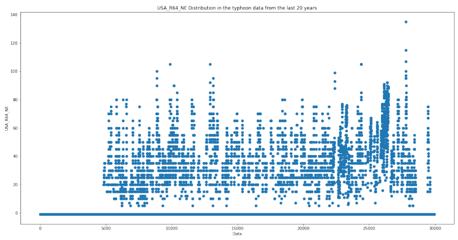


Figure 6.14: MRW N/E 64 miles

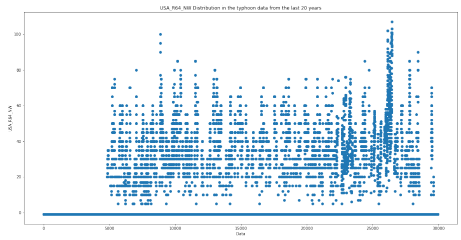


Figure 6.15: MRW N/W 64 miles

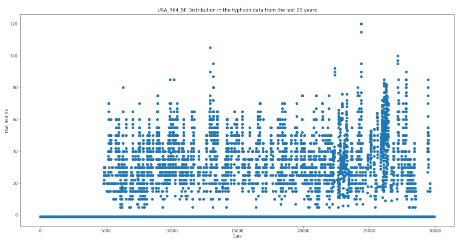


Figure 6.16: MRW S/E 64 miles

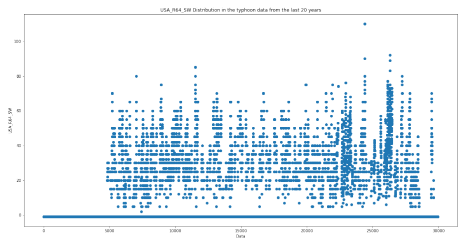


Figure 6.17: MRW S/W 64 miles

Some patterns in the graphs between different radius distances were very similar to each other. For instance, the MRW pattern shown in Figure 6.10 and Figure 6.14 were similar as the maximum wind 50 and 64 miles away from the typhoon centre was recorded consistently in a presence of stronger typhoons across the last 21 years in the Western Pacific Region.

6.2 Statistical test

Feature	Average P value
Latitude	0.768
Longitude	0.554
Distance to land	0.433
Wind Speed	0.392
Pressure	0.388
MRW 34 N/E	0.42
MRW 34 N/W	0.432
MRW 34 S/E	0.419
MRW 34 S/W	0.417
MRW 50 N/E	0.263
MRW 50 N/W	0.271
MRW 50 S/E	0.263
MRW 50 S/W	0.265
MRW 64 N/E	0.193
MRW 64 N/W	0.195
MRW 64 S/E	0.194
MRW 64 S/W	0.197

Table 6.1: Augmented Dickey–Fuller test result

Data stationarity was a criterion of conducting the granger causality test, therefore, the ADF and the Cointegration test were executed to test whether the given time series of data are stationary. Since features in different typhoons might have different stationary properties, this project took the average of each P-value obtained from each typhoon dataset and the average P-value became the representative figure for a specific feature. The statistical results of the augmented Dickey-Fuller test (ADF), the Cointegration test and the granger causality test were shown in Table 6.1, 6.2 and 6.5.

The P values in Figure 6.1 represented a result from the ADF test of each meteorological feature. The figures showed that all P values are greater than the significance level of 0.05. It indicated that none of the individual meteorological features is stationary. Even though all individual features were non-stationary, the

Table 6.2: Cointegration test

Table 6.3: Test with Latitude

Feature	Average P value
Latitude	0.000
Longitude	0.603
Distance to land	0.598
Wind Speed	0.827
Pressure	0.824
MRW 34 N/E	0.845
MRW 34 N/W	0.853
MRW 34 S/E	0.841
MRW 34 S/W	0.843
MRW 50 N/E	0.842
MRW 50 N/W	0.840
MRW 50 S/E	0.841
MRW 50 S/W	0.840
MRW 64 N/E	0.840
MRW 64 N/W	0.840
MRW 64 S/E	0.840
MRW 64 S/W	0.839

Table 6.4: Test with Longitude

Feature	Average P value
Latitude	0.604
Longitude	0.000
Distance to land	0.599
Wind Speed	0.753
Pressure	0.740
MRW 34 N/E	0.767
MRW 34 N/W	0.781
MRW 34 S/E	0.785
MRW 34 S/W	0.785
MRW 50 N/E	0.752
MRW 50 N/W	0.754
MRW 50 S/E	0.757
MRW 50 S/W	0.752
MRW 64 N/E	0.733
MRW 64 N/W	0.735
MRW 64 S/E	0.733
MRW 64 S/W	0.732

Cointegration test came into play to test whether two non-stationary time series of data are correlated with each other. However, the result from Table 6.2 showed that none of the features correlated with latitude and longitude since the average P values from the Cointegration test was all greater than 0.05 as well.

The result from two previous tests summarised that none of the individual and paired features had stationary properties. Each feature did not meet the criteria to perform the granger causality test as the features have no granger causality relationship to the latitude and longitude associated with any lagged values.

Even though the features did not meet the criteria to conduct the granger causality test, this project still delivered the test result to clarify whether a given meteorological feauture did not granger cause the target variables in a given timestamp. The result was shown in table 6.5. The table concluded that the average P-value of each feature paired with latitude and longitude was greater than 0.05. The result from the statistical tests taken in this project summarised that all typhoon feature data used in this project were not stationary in time series. The

Table 6.5: Granger causality test

Table 6.6: Test with Latitude in 5 lags

Feature	Average P value
Latitude	1.000
Longitude	0.161
Distance to land	0.266
Wind Speed	0.349
Pressure	0.362
MRW 34 N/E	0.401
MRW 34 N/W	0.402
MRW 34 S/E	0.402
MRW 34 S/W	0.401
MRW 50 N/E	0.384
MRW 50 N/W	0.393
MRW 50 S/E	0.399
MRW 50 S/W	0.394
MRW 64 N/E	0.411
MRW 64 N/W	0.411
MRW 64 S/E	0.414
MRW 64 S/W	0.405

Table 6.7: Test with Longitude in 5 lags

Feature	Average P value
Latitude	0.168
Longitude	1.000
Distance to land	0.241
Wind Speed	0.300
Pressure	0.335
MRW 34 N/E	0.397
MRW 34 N/W	0.388
MRW 34 S/E	0.386
MRW 34 S/W	0.400
MRW 50 N/E	0.351
MRW 50 N/W	0.339
MRW 50 S/E	0.352
MRW 50 S/W	0.361
MRW 64 N/E	0.347
MRW 64 N/W	0.337
MRW 64 S/E	0.357
MRW 64 S/W	0.340

summary could be supported by the research delivered by Huang, Xie, and Gu, 2020. Gu *et al.* Huang, Xie, and Gu, 2020 conducted an experiment studying wind properties during the typhoon event. The experiment came up with a conclusion showing typhoon-induced wind was mostly non-stationary.

6.3 Model 1 - LSTM Regression model

6.3.1 Experimental Result

The results of validation loss against training loss from the LSTM Regression models were shown in Figure 6.18 and 6.19. The learning curves from the figures showed that both models delivered an outstanding generalisability for predicting typhoon trajectories in the next 3 hours and 6hours as the gap between training losses and validation losses from the two models was insignificant without overfitting. Table 6.8 and 6.9 delivered a setting summary after performing hyperparameter searching on the LSTM regression model for 3 hours and 6 hours prediction. The optimised LSTM models for predicting latitude and longitude were then tested with the testing dataset. The testing dataset was the last 10% of typhoon data in the entire data. The RMSE for predicting latitude and longitude prediction in the next 3 hours and 6 hours from the LSTM Regression model was 0.6 and 0.95 respectively. Since one degree of latitude and longitude is equal to 111 km, the mean positional error is 66km and 105.5km for 3 hours and 6 hours prediction. Four individual supertyphoon datasets were extracted initially to act as an extra testing dataset for track visualisation. Table 6.10 summarised the overall mean position errors in 3 hours and 6 hours prediction for all testing datasets.

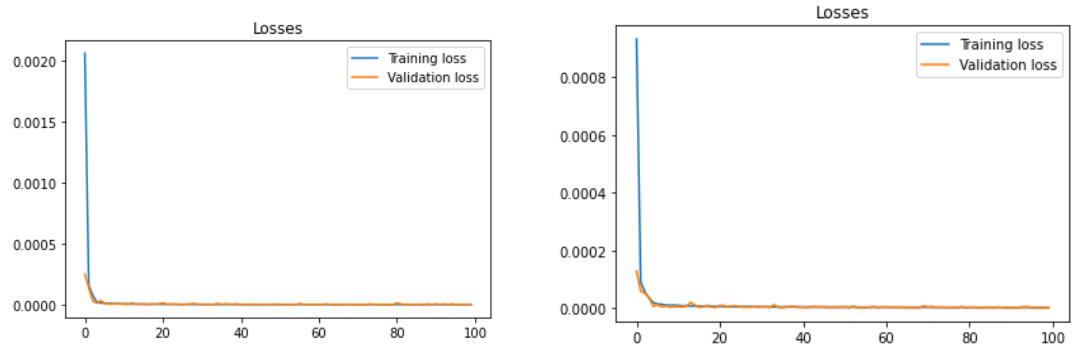


Figure 6.18: Learning curve for latitude (Left) and longitude (Right) model - 3hours

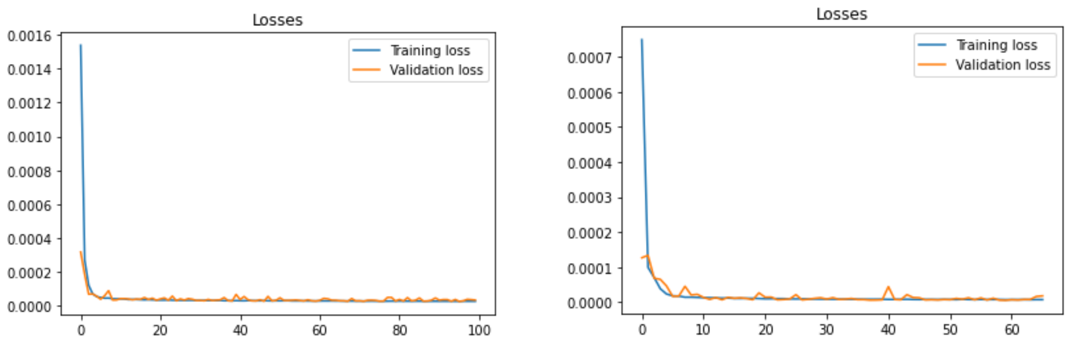


Figure 6.19: Learning curve for latitude (Left) and longitude (Right) model - 6hours

Parameter setting	Latitude (3 hrs)	Longitude (3 hrs)
Hidden dimension	16	128
Optimisers	Adam	Adam
Learning rate	0.001	0.001
Batch size	8	16
Epoch	100	100
Sliding window	5:1	5:1

Table 6.8: The LSTM Regression model hyperparameters setting for 3 hours prediction

Parameter setting	Latitude (6 hrs)	Longitude (6 hrs)
Hidden dimension	32	64
Optimisers	Adam	Adam
Learning rate	0.001	0.001
Batch size	8	8
Epoch	100	100
Sliding window	5:2	5:2

Table 6.9: The LSTM Regression model hyperparameters setting for 6 hours prediction

Testing dataset	MPE (3 hrs)	MPE (6 hrs)
10% testing data	66 km	105.5 km
MANGKHURT	16.6 km	28.9 km
KONGREY	19.9 km	40.7 km
YUTU	14.1 km	23.7 km
WUTIP	11.8 km	26.64 km

Table 6.10: Prediction performance on testing dataset from the LSTM Regression model

6.3.2 Track visualisation

Figure 6.20, 6.21, 6.22 and 6.23 visualised the predicted tracks and actual tracks of four supertyphoons on Basemap from python.

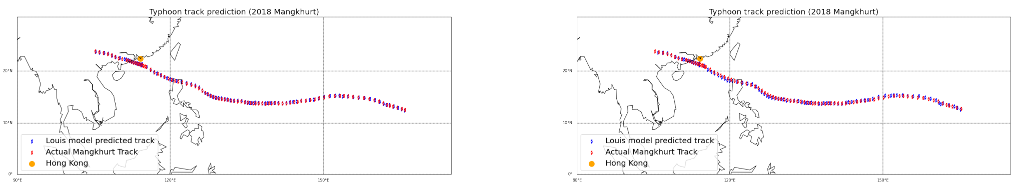


Figure 6.20: Supertyphoon MANGKHURT track predictions

Figure 6.21: Supertyphoon KONGREY track predictions

Figure 6.22: Supertyphoon YUTU track predictions

Figure 6.23: Supertyphoon WUTIP track predictions

6.3.3 Model Evaluation

In general, the 3 hours and 6 hours mean position errors derived from the LSTM Regression model were not significantly large as most of the distance errors in the individual typhoon dataset were less than 20 km and 50km for 3 hours and 6 hours prediction respectively. In addition, the predicted track and the actual track shown on the maps were mostly overlapped indicating an outstanding prediction performance from the LSTM Regression model.

6.4 Model 2 - CNN-LSTM ensemble model

6.4.1 Experimental Result

The CNN-LSTM ensemble models combined the meteorological features and the typhoon satellite image features to predict the typhoon tracks in the next 3 hours and 6 hours. The satellite grey-scale image features extracted from the pretrained ResNet-18 model were combined with the meteorological features as an input to the LSTM Regression model. The results of validation loss against training loss from the CNN-LSTM ensemble models predicting 3 hours and 6 hours typhoon track were shown in Figure 6.24 and Figure 6.25. The learning curves from the figures showed that the validation losses and training losses on both models predicting 3 hours and 6 hours of typhoons can converge towards the global minima without any sense of overfitting and underfitting. However, there was some minor fluctuation in the validation curves over time. The reason behind the fluctuation was due to possible missing satellite images in some timestamps in the validation dataset that affected the general learning performance for generalising typhoon track. Table 6.12 listed the hyperparameter setting of an optimised CNN-LSTM model for 3 hours and 6 hours track predictions. The track errors in 3 and 6 hours from the optimised model were 0.59 and 0.65, which were equivalent to the mean position error of 65.5 km and 72.1 km respectively. Table 6.13 delivered a summary of the mean position errors for predicting testing typhoon dataset in the next 3 and 6 hours from the ensemble model.

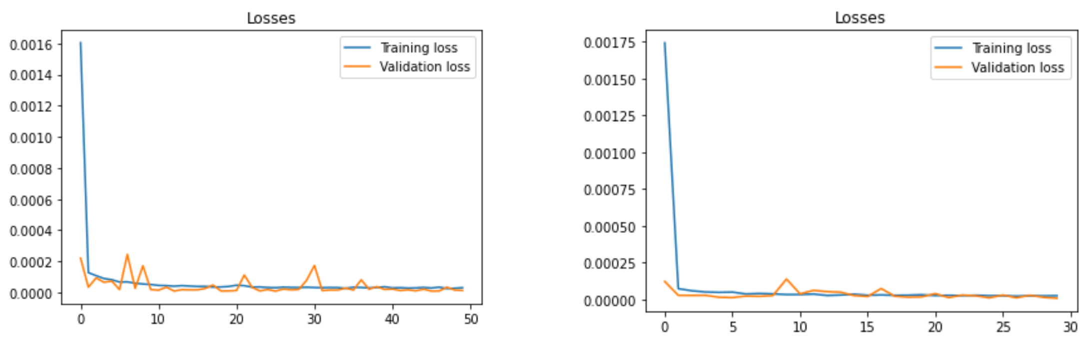


Figure 6.24: Learning curve graph for latitude (Left) and longitude (Right) model - 3hours

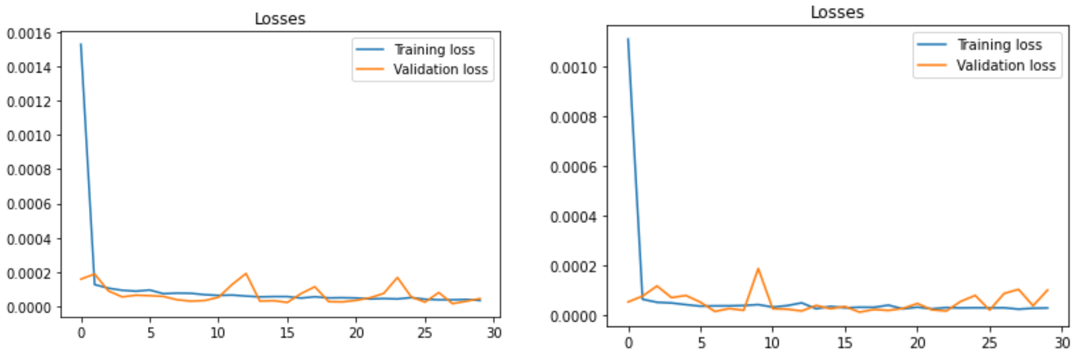


Figure 6.25: Learning curve graph for latitude (Left) and longitude (Right) model - 6hours

Parameter setting	Latitude (3 hrs)	Longitude (3 hrs)
Hidden dimension	128	64
Optimisers	Adam	Adam
Learning rate	0.001	0.001
Batch size	16	16
Epoch	100	100
Sliding window	5:1	5:1

Table 6.11: The hyperparameters setting of the CNN-LSTM ensemble model for 3 hours prediction

Parameter setting	Latitude (6 hrs)	Longitude (6 hrs)
Hidden dimension	32	32
Optimisers	Adam	Adam
Learning rate	0.001	0.001
Batch size	16	16
Epoch	100	100
Sliding window	5:2	5:2

Table 6.12: The hyperparameters setting of the CNN-LSTM ensemble model for 6 hours prediction

Testing data	MPE (3 hrs)	MPE (6 hrs)
10% testing data	65.5 km	72.1 km
MANGKHURT	27.8 km	39.0 km
KONGREY	38.1 km	35.2 km
YUTU	55.1 km	39.96 km
WUTIP	38.4 km	43.3 km

Table 6.13: Prediction performance on testing data in the CNN-LSTM ensemble model

6.4.2 Track visualisation

Figure 6.26, 6.27, 6.28 and 6.29 visualised the overall predicted track pattern of four supertyphoons against their actual track on the map.

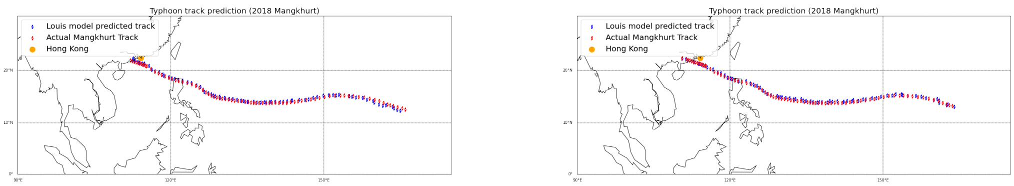


Figure 6.26: Supertyphoon MANGKHURT predictions

Figure 6.27: Supertyphoon KONGREY predictions

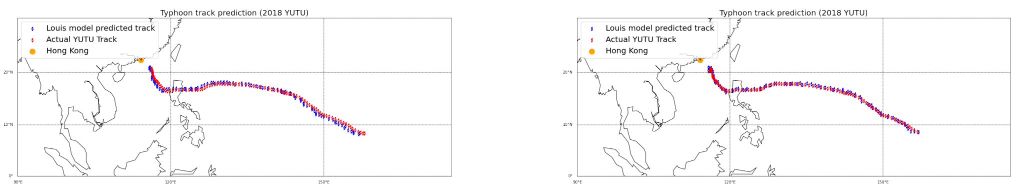


Figure 6.28: Supertyphoon YUTU predictions

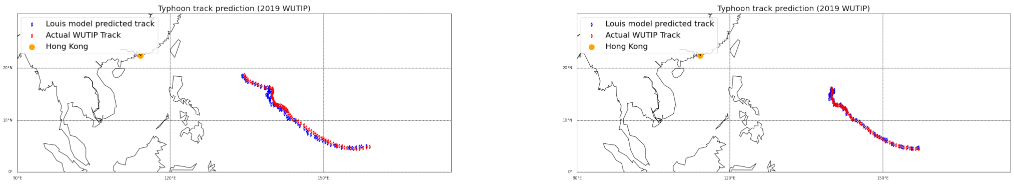


Figure 6.29: Supertyphoon WUTIP predictions

6.4.3 Model Evaluation

In general, track prediction of the ensemble model from the testing datasets delivered a slight improving result compared to the previous LSTM model, especially in 6 hours prediction. Extracting different meteorological features of changing cloud covers, and the presence of typhoon eye from the additional satellite image feature might help the ensemble model to predict the typhoon track accurately in a longer timestamp. Therefore, the CNN-LSTM ensemble model performed better than the LSTM regression model in the 6 hours prediction. However, the typhoon orientation in the image was always in the same position as no obvious track information could be extracted between a short timestamp. Therefore, 3 hours prediction from the ensemble model and the previous LSTM regression model were very similar. Utilising moving frame of typhoon satellite images might significantly improve the overall prediction performance of this model.

6.5 Model 3 - LSTM Classification model

6.5.1 Experimental Result

The prediction outcome from the classification model is a set of a normalised vector containing a score for each class. The vector index position with the highest score belongs to the predicted class. Figure 6.30 and Figure 6.31 visualised the result of validation loss against training loss from the LSTM classification models predicting typhoon track in the next 3 hours and 6 hours respectively. Compared to the ensemble model and LSTM regression model, all validation losses were larger than training losses which were regarded as underfitting. In addition, validation losses had been fluctuating over time converging to the global minima. One of the reasons behind this situation was that the validation dataset was 10% of the total dataset which contained typhoon data from 2017 to 2019. These two years of typhoon dataset might not be able to represent the overall bin index distributions of the entire dataset as it caused a lot of noise to the classification model. In general, the classification approach was not capable of identifying the true bin index based on the given sequential meteorological features. However, this project was going to use the RMSE to measure the position error of the predicted track instead of using traditional performance metrics in classification, such as F1, precision and recall. The final predicted bin index was converted back to latitude and longitude and compared with the actual track in terms of distance. Similar to the previous models, Table 6.14 and Table 6.15 delivered a summary of hyperparameter searching on the LSTM classification model for 3 hours and 6 hours prediction. The optimised LSTM classification models were then tested with the testing dataset. The final RMSE, mean positional error for predicting typhoon tracks in the next 3 hours and 6 hours from the classification model was 1.14, 127 km and 1.65, 183 km respectively. Table 6.16 delivered a summary of predicting four supertyphoons in the next 3 and 6 hours from the classification model.

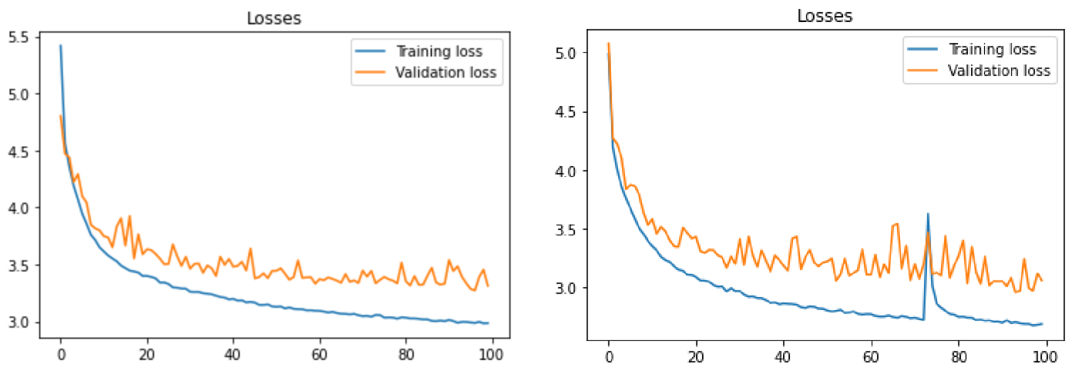
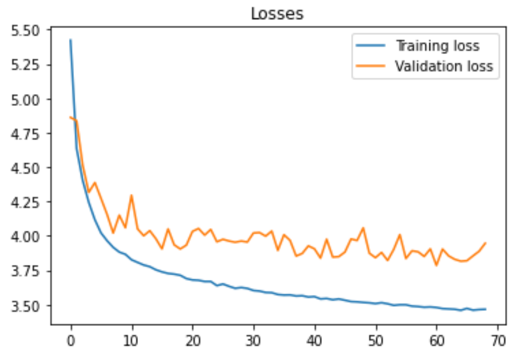


Figure 6.30: Learning curve graph for latitude (Left) and longitude (Right) model - 3hours



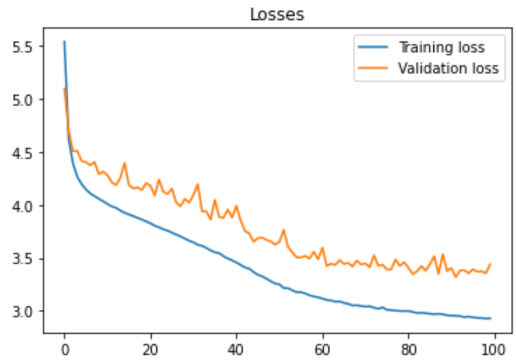


Figure 6.31: Learning curve graph for latitude (Left) and longitude (Right) model - 6hours

Parameter setting	Latitude (3 hrs)	Longitude (3 hrs)
Hidden dimension	16	16
Optimisers	Adam	Adam
Learning rate	0.01	0.01
Batch size	16	8
Epoch	100	100
Sliding window	5:1	5:1

Table 6.14: The LSTM Regression model hyperparameters setting for 3 hours prediction

Parameter setting	Latitude (6 hrs)	Longitude (6 hrs)
Hidden dimension	16	32
Optimisers	Adam	Adam
Learning rate	0.01	0.001
Batch size	8	8
Epoch	100	100
Sliding window	5:2	5:2

Table 6.15: The LSTM Regression model hyperparameters setting for 6 hours prediction

Supertyphoon	MPE (3 hrs)	MPE (6 hrs)
MANGKHURT	76.6 km	71.0 km
KONGREY	50.0 km	118.7 km
YUTU	41.1 km	90.2 km
WUTIP	55.5 km	62.7 km

Table 6.16: Prediction performance on testing dataset from the classification model

6.5.2 Track visualisation

All visualised results of predicted track against actual tracks were plotted in Figure 6.32, 6.33, 6.34 and 6.35 accordingly.

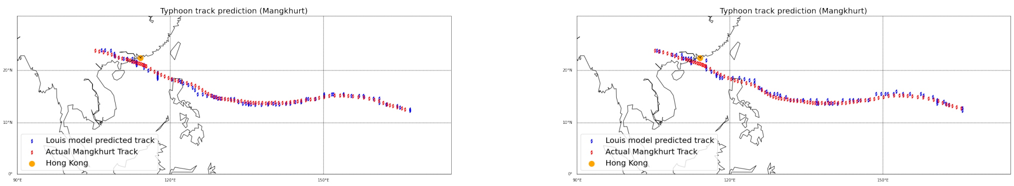


Figure 6.32: Supertyphoon MANGKHURT track predictions

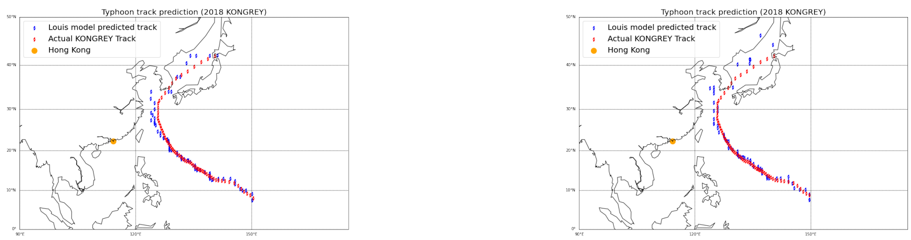


Figure 6.33: Supertyphoon KONGREY track predictions

Figure 6.34: Supertyphoon YUTU track predictions

Figure 6.35: Supertyphoon WUTIP track predictions

6.5.3 Model Evaluation

In general, the prediction results from the classification model on the testing dataset were not as good as the prediction results from the previous models, especially for 6 hours predictions. The predicted tracks in blue colour did not overlap a lot with the actual tracks in red colour on the maps along the typhoon life cycle. The reason behind it was mainly due to the conversion error converting the bin index back to the latitude and longitude degree which would eventually magnify the error in RMSE measurements. Alemany *et al.* Alemany et al., 2018 performed a similar approach of using a grid-based LSTM model to perform track prediction for the Atlantic hurricanes. Alemany *et al.* Alemany et al., 2018 emphasised that the conversion error from grid coordination back to latitude and longitude degree was as much as 50 km and it was inevitable. According to the learning curves shown in Figure 6.30 and 6.31, this model is facing an underfitting problem of not capable of capturing the relationship between the input data and the ground truth bin index. One of the possible solutions to improve the accuracy of this classification model was to increase data size and model complexity.

6.6 Overall prediction summary

Models	Track errors (3 hrs)	Track errors (6 hrs)
LSTM Regression model	66 km	105.5 km
CNN-LSTM ensemble model	65.5 km	72.1 km
LSTM Classification model	127 km	183 km

Table 6.17: Model summary between various deep learning approaches

Table 6.17 summarised the 3-hours and 6-hours track errors predicted by three deep learning approaches conducted in this project. All result were based on the usage of 10% testing dataset. The LSTM regression model and the CNN-LSTM ensemble model both had a very similar 3-hours prediction performance. However, the CNN-LSTM model delivered the least track errors on 6-hours predictions compared to the other two deep learning approaches. The ensemble approach had the potential to perform even better if the input typhoon images had more related information regarding typhoon position. For instance, Yang *et al.* Giffard-Roisin et al., 2020 performed a similar ensemble model architecture to construct a typhoon tracking model by using moving frames of typhoon images. The typhoon centres were in a different position from the fixed size images which could extract better features regarding typhoon moving pattern. Apart from using satellite images, a heatmap image could also be an input image having an informative feature for predicting typhoon tracks. For instance, Li Yongjiao Sun et al., 2021 collected sea surface temperature data from a heatmap image in different timestamps to enhance data diversity for the typhoon track prediction task. And finally, the LSTM classification model did not perform well on both 3-hours and 6-hours track prediction as the conversion process of converting bin index back to latitude and longitude magnified the track errors in final performance measurements.

7
Future work

Contents

7.1 Implementation of transformer

A more versatile backbone of the transformer can be an alternative approach for future research to deliver a more robust time series forecasting task Wen et al., 2022. The transformer has been widely used in different areas, such as natural language processing and computer vision. The transformer has been proven to work very well on a multi-modality dataset, however, the diversity of the dataset in the current project might not be sufficient to optimise the usage of the transformer. Future researchers can increase the data diversity by using different data types, such as collecting features from the heatmap images or 3D images.

7.2 Image data diversity

The typhoon satellite images used in this project had no significant related track features which were beneficial to the 3 hours of typhoon track predictions. Future researchers can conduct different experiments on different types of images, such as a heatmap image to identify surrounding sea surface temperature and a moving frame of a satellite image locating the typhoon position in different timestamps. These images may lead to a different prediction performance with more typhoon track-related features.

7.3 Reference dataset

Future researchers can download the dataset from this project as a starting point for future typhoon prediction analysis. Using the same dataset in future research can deliver a more sophisticated evaluation with the implementation of different algorithms. The scrapping python code will be posted on Github so that everyone can be used it as a scrapping tool for collecting typhoon satellite image data. The meteorological dataset will also be provided on Github with a hyperlink linking to the NOAA data archive repository.

8 Conclusion

This project proposed three deep learning approaches to typhoon track prediction. The last 21 years of the Western Pacific typhoon dataset were used to train the models. The prediction performance of the models were evaluated based on the predicted mean position errors. Our LSTM Regression model and the ensemble model delivered a very similar prediction performance in 3-hour track prediction, however, the CNN-LSTM model had the best performance in 6-hour prediction among other deep learning approaches. The CNN-LSTM ensemble model extracted feature from the satellite images model and combined it with the meteorological features in the regression model to predict latitude and longitude in the next timestamp. The extra satellite image feature benefited the prediction from getting more meteorological features, such as cloud cover, and the presence of a typhoon eye. These additional features helped the model to predict typhoon track better in a longer timestamp. This ensemble deep learning approach had significant potential to outperform the traditional regression model if the input image had more relevant track information, for instance, previous researchers used moving frames of satellite images in a deep learning model to make a prediction. Even though the conversion errors from the LSTM classification model were inevitable, the prediction performance of the classification model could be enhanced by using more data and increasing model complexity to avoid underfitting. Deep learning approaches to typhoon track prediction are yet to replace the contributions made by the supercomputers used by the weather organisation nowadays. However, the deep learning models can also be contributed to the typhoon track forecasting task based on their outstanding prediction efficiency. Future researchers can use this research as a reference or starting point to improve the typhoon track prediction accuracy with other various sophisticated deep learning algorithms.

9
Plan vs Progress

Timeline	Original plan from last year	Progress of this year
February 2022	Started looking at Pytorch implementation	Implemented Pytorch framework for the LSTM Regression model Developed a new idea of using CNN-LSTM model for typhoon track prediction
March 2022	Implemented statistical test to select useful features	Discovered no granger causality relationship between the meteorological features Scrapped typhoon satellite images from the Kitamoto laboratory website
April 2022	Used diverse deep learning models and compared their prediction performance	Implemented three different approaches and performed hyperparameter searching in each model 1.LSTM Regression model 2.CNN-LSTM ensemble model 3.LSTM Classification model
May 2022	Optimised and evaluated model performance Performed track visualisation on a selected testing dataset	Delivered and compared 3 hours and 6 hours track prediction from each model Evaluated model performance from testing dataset Visualised track prediction in Basemap
June 2022	Thesis writing and presentation	Accomplished presentation on 2 June 2022 and submitted thesis on 13 June 2022

Table 9.1: Plan vs Progress

Bibliography

Hennon, Christopher (2021). *IBTrACS data*. URL: http://ibtracs.unca.edu/index.php?name=introduction.

Asanobu, KITAMOTO (2022). *Digital Typhoon - Typhoon Images and Information*. URL: http://agora.ex.nii.ac.jp/~kitamoto/index.html.en.

Ying, Ming et al. (May 2012). "Impacts of Climate Change on Tropical Cyclones in the Western North Pacific Basin. Part II: Late Twenty-First Century Projections". In: *Tropical Cyclone Research and Review* 1, pp. 231–241.

Agency, Japan Meteorological (2021). *Typhoon Analysis and Forecasts*. URL: https://www.jma.go.jp/jma/en/Activities/forecast.html.

Alemany, Sheila et al. (Feb. 2018). "Predicting Hurricane Trajectories Using a Recurrent Neural Network". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.

Lian, Jie et al. (2020). "A Novel Data-Driven Tropical Cyclone Track Prediction Model Based on CNN and GRU With Multi-Dimensional Feature Selection". In: *IEEE Access* 8, pp. 97114–97128.

Giffard-Roisin, Sophie et al. (2020). "Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data". In: *Frontiers in Big Data* 3, p. 1. ISSN: 2624-909X. URL: https://www.frontiersin.org/article/10.3389/fdata.2020.00001.

Pham, Tuan D. (2021). "Time–frequency time–space LSTM for robust classification of physiological signals". eng. In: *Scientific reports* 11.1, pp. 6936–6936. ISSN: 2045-2322.

ITO, Kosuke et al. (2020). "Recent Progress in the Fundamental Understanding of Tropical Cyclone Motion". In: *Journal of the Meteorological Society of Japan. Ser. II* advpub.

Wu, Liang et al. (2019). "Tropical cyclones and multiscale climate variability: The active western North Pacific Typhoon season of 2018". eng. In: *Science China. Earth sciences* 63.1, pp. 1–11. ISSN: 1674-7313.

Choy, Chun-wing, Man-chi Wu, and Tsz-cheung Lee (2020). "Assessment of the damages and direct economic loss in Hong Kong due to Super Typhoon Mangkhut in 2018". eng. In: *Tropical Cyclone Research and Review* 9.4, pp. 193–205. ISSN: 2225-6032.

PAN, Chi-kin (Sept. 2011). *Why Tropical Cyclone Recurves?* URL: https://www.hko.gov.hk/en/education/tropical-cyclone/tracking/00155-why-tropical-cyclone-recurves.html.

Sun, Youqiang et al. (July 2014). "Using causal discovery for feature selection in multivariate numerical time series". In: *Machine Learning* 101.

Cermak, Vladimir and Louise Bodri (June 2016). “Attribution of precipitation changes on ground–air temperature offset: Granger causality analysis”. In: *International Journal of Earth Sciences* 107.

Gao, Song et al. (May 2018). “A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network”. In: *Acta Oceanologica Sinica* 37, pp. 8–12.

Sun, Yongjiao et al. (2021). “Distributed Typhoon Track Prediction Based on Complex Features and Multitask Learning”. eng. In: *Complexity (New York, N.Y.)* 2021, pp. 1–12. ISSN: 1076-2787.

Edholm, Gustav and Xuechen Zuo (2018). “A comparison between a conventional LSTM network and a grid LSTM network applied on speech recognition”. In.

Ruttgers, Mario et al. (2022). “Prediction of Typhoon Track and Intensity Using a Generative Adversarial Network With Observational and Meteorological Data”. eng. In: *IEEE access* 10, pp. 48434–48446. ISSN: 2169-3536.

Tong, Biao et al. (2022). “Identification of tropical cyclones via deep convolutional neural network based on satellite cloud images”. eng. In: *Atmospheric measurement techniques* 15.6, pp. 1829–1848. ISSN: 1867-1381.

Wang, Chong et al. (2022). “Tropical Cyclone Intensity Estimation From Geostationary Satellite Imagery Using Deep Convolutional Neural Networks”. eng. In: *IEEE transactions on geoscience and remote sensing* 60, pp. 1–16. ISSN: 0196-2892.

Qing, Xiangyun and Yugang Niu (2018). “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM”. In: *Energy* 148, pp. 461–468. ISSN: 0360-5442. URL: https://www.sciencedirect.com/science/article/pii/S0360544218302056.

Huang, Peng, Wen Xie, and Ming Gu (2020). “A comparative study of the wind characteristics of three typhoons based on stationary and nonstationary models”. eng. In: *Natural hazards (Dordrecht)* 101.3, pp. 785–815. ISSN: 0921-030X.

Wen, Qingsong et al. (2022). “Transformers in Time Series: A Survey”. eng. In.