FEATURE EXTRACTION AND TROPICAL CYCLONE PREDICTION SYSTEM USING CORRELATION COEFFICIENT AND LOGISTIC REGRESSION

[1]Thu Zar Hsan

Faculty of Computer Systems and Technologies,
University of Computer Studies, Yangon, Myanmar
thuzarhsan@ucsy.edu.mm, thuzarhsan@gmail.com

[2]Thin Lai Lai Thein

Faculty of Information Science,
University of Computer Studies, Yangon, Myanmar
tllthein@ucsy.edu.mm

Abstract

Since tropical storms (TS) are among the most hazardous natural disasters, timely and precise track forecasts are essential for preventing and mitigating disasters. In this study, we proposed a feature extraction method based on the correlation between two similar cyclones. The cyclone track is predicted using a multiple logistic regression technique that compares three sigmoid functions. We were able to estimate the latitude and longitude of a cyclone's location 24 hours ahead of time by using the last twelve hours of observations (two positions, at six hourly intervals, and the current position). We investigate the real-world transient circulation of TCs in the Northern Indian Ocean between 1945 and 2022. There is a mean error of 137.5 km in estimating the area between the historical track and the forecasted points. The tropical cyclone track was most significantly influenced by the minimum sea level pressure and maximum sustained wind speed for short-term forecasts.

***Keywords*: Tropical cyclone track; logistic regression; correlation; feature extraction.**

1. Introduction

One of the most extreme climatic phenomena that greatly affects social and economic development and causes catastrophic natural disasters is the tropical cyclone. Every year, powerful cyclones destroy coastal communities, taking lives and causing significant property damage. For instance, when Cyclone Nargis struck Myanmar in 2008, 2.4 million people were impacted, 84,500 people died, and 53,800 went missing [Ali, MM.; Kishtawal, C. (2007)]. For this reason, forecasting a tropical cyclone's path is crucial.

However, a number of parameters, including as the thermodynamics and kinetics of the tropical storm structure and the surrounding meteorological conditions, influence the development of a tropical cyclone. The greatest sustained wind speed, the lowest sea level pressure, the wind intensity, and the eye diameter are only a few of the different factors that can affect a cyclone's path. Forecasting tropical cyclone trajectories is a huge difficulty due to the interplay of these elements. Researching new tropical cyclone track forecasting techniques is essential because of the tropical storms have an enormous effects on people and the challenges involved in predicting them.

At the moment, there are two main kinds of forecasting techniques for TCs: numerical weather prediction (NWP) is the dominant method. When the initial situation and atmospheric boundary situation are known, NWP computes the approximate result of partial differential equations, including atmospheric state variables. A statistical model, which often makes use of multiple regression, is the alternative forecasting technique. The relationship between the TC's mobility and its unique historical features forms the basis of the statistical model [Wang, L. et al., (2023)]. Nowadays, numerous deep learning models are used to forecast cyclone tracks; however, they have certain limitations, such as high computational cost, overfitting, lack of interpretability, and dependency on data accuracy. When utilizing deep learning to solve a problem, several limits must be considered.

The major goal of this system is to develop a new forecasting model for the TC track using the multidimensional logistic regression method. The proposed model uses only a simple multiple logistic regression method, which uses the position of the past 18 hours in terms of latitude and longitude to Magnitude and Direction by using a mathematical equation, maximum sustained wind speed, and minimum sea level pressure for new cyclone track and two most correlated cyclones that happened in the same basin (Northern Indian Ocean) to extract features for model. This process can reduce time complexity and improve the prediction accuracy of the next timestamp for 6-hour short term prediction. Features of direction and magnitude are passing through the multiple logistic regression method with sigmoid function, predicted results are not labeled as binary form as “0” or “1”, “yes” or “no”, and “true” or “false” which is assumed as the probability value and changing as the direction and magnitude of the next point. This research shows less powerful and time consuming for prediction model can be developed by using mathematical equations with a logistic regression model changing the nature of the input variable. Finally, this proposed system is developed in a real-world tropical cyclone dataset from the Joint Typhoon Warning Center (JTWC), which supports the storms’ geographical factor at 6-hour intervals and this model performs better than some existing traditional methods, statical method, and deep learning technique for short term prediction.

The remaining portions of the paper are arranged as follows: Section 2 reviews the related work of the system. In this section, prediction cyclone track is not the modern issue and many researchers have still solved this article by using various techniques. Various forms of input historical data such as Satellite images and statistical data, and various learning methods. The main issue is to forecast short-term and long-term prediction with less time consumption and high rate of accuracy. Section 3 introduced type of dataset and proposed system. Section 4 carries out the experimental results and discusses step by step of the proposed model. The last section 5 recaps this research and offers suggestions for future plans.

2. Related Work

The traditional approaches for predicting tropical storm paths generally rely on integrated, statistical, dynamic, and numerical models. In order to manipulate complex dynamical equations, a numerical model needs a significant amount of computing power. In order to replicate the internal organization of tropical storms in real-time, the model must produce a grid system. This approach is less time-efficient, nevertheless, a statistical model is more efficient than a numerical model since it simply computes the behavior pattern of tropical cyclones using previous data [Lian, J.; Dong, P.; Zhang, Y.; Pan, J. (2020)]. It is also important to note that the integrated structure is a predictive structure that integrates various models, physical values, and initial structure conditions; this model typically produces better-forecasted results than a single model [Aleman, S. et. al. (2018)] [Moradi Kordmahalleh, M et.al. (2016)]. However, as ground observation stations, meteorological satellites and ocean observation stations are gradually established, the observation data system is improved so that an increasing amount of data is computed. Finding a way to increase the effectiveness and precision of tropical cyclone track prediction in large-scale spatial-temporal data remains a major problem, and machine learning techniques are still becoming more sophisticated these days, with more integrating another method to apply them to increase tropical cyclones’ track prediction.

In [Nakamura, J. et.al. (2017)], this system used mass moments ellipses and cluster analysis, the study examines tropical cyclone (TC) model tracks in the Western North Pacific. It selects a subset for investigation by comparing the TC tracks of models to observed ones. Two changes in track are identified by the study: an eastward shift of storm tracks around Hawaii in a warmer climate and an increase in North-South expansion caused by weak Coriolis force near the Equator. The outcomes highlight how crucial it is to take into account a variety of models and circumstances.

In [Richman, MB. B. et.al (2017)], this system proposed support vector regression (SVR) on a basic predictor pool, a study that seeks to lower seasonal prediction errors for tropical cyclones (TCs). Techniques for wrapper and attribute selection are used to optimize the strategy. The findings demonstrate that appropriate attribute selection improves connections between the predicted and observed yearly TC count by significantly lowering prediction errors. The method is adaptable and can be used in other TC basins. This system can reduce the root mean squared error (RMSE) from 3.8 to 2.7, while the mean absolute error (MAE) decreases from 2.9 to 1.8.

In [Zhang, Y. et.al (2018)], researchers developed matrix neural networks and these are ideal for image recognition and cyclone tangency prediction in the South Indian Ocean. They can handle spatial correlations, allowing for convenient input without vectorization. The results show that these networks outperform recurrent neural network architectures in preserving spatial correlation.

Using spatial locations and several meteorological parameters, the research suggests a data-driven deep learning model for tropical storm track prediction. Using real-world tropical storm datasets from 1945 to 2017, the model which consists of a CNN layer, a GRU layer, and a multidimensional feature selection layer was trained. The accuracy of the model surpasses that of certain deep learning techniques without the feature selection layer, and it performs better than conventional forecasting techniques [Lian, J.; Dong, P.; Zhnag, Y.; Pan, J. (2020)].

An ensemble machine learning model was created in this study to forecast tropical cyclone tracks in the Western North Pacific. The climatology and persistence (CLIPER) model were compared with the model, which was based on TC climatology and persistence components. During the three prediction times, the GBDT model demonstrated good performance with minor forecast errors and outstanding TC movement direction features. For storm acceleration and deceleration, it was insufficient, nevertheless. Operational TC track projections are supported by solid information from the GBDT model [Tan, J.; Chen, S.; Wang, J. (2020)].

Deep learning method is also a popular approach for large amounts of data. When dealing with massive volumes of data, traditional tropical cyclone track prediction techniques suffer. There are benefits to using deep learning techniques for capturing intricate spatial characteristics. In this paper [Dong, P. et.al (2022)], researchers implement the system by using satellite images as input data by using deep learning approach. To forecast TC tracks, a novel encoding-to-forecasting model incorporating a spatial attention network and convolutional long-term and short-term memory was put forth. With an absolute means error of 30 km less, the model performed better in real remote sensing photos than competing models.

In [Ravindra, V.; Nag, S.; (2020)] paper, an ensemble-guided cyclone track (EGCTF) technique for tracking tropical cyclones remotely was presented. Cyclone center coordinates and numerical weather prediction models are used by the method. The EGCTF demonstrated good forecast skills for more than 290 test scenarios when tested using GEFS and National Hurricane Center data, getting better with longer forecast periods.

3. Data and Proposed System

3.1. *Dataset*

The data used in this research are trajectory and metrological factors. The 146 TC track data originate from the North Indian Ocean Best Track Data provided by the Joint Typhoon Warning Center [JTWC] between 1945 and 2022. In each historical data, there are six hourly tropical cyclone landing locations of 0.1 by 0.1 degrees and several meteorological factors, including the radius of the maximum winds (MRD), the minimum sea level pressure (MSLP), the level of tropical cyclone development (TY), the pressure in millibars of the last closed isobar (RADP), the maximum sustained wind speed (VMAX), the wind intensity (kts) for the radii (RAD), storm speed (SPEED), the eye diameter (EYE), storm direction (DIR), the maximum sea level (MAXSEAS), storm name (STORMNAME), system depth (Depth), wave height for the radii given in SEAS1-SEAS4, and radius code (SEASCODE). Meteorological factors were represented as one-dimensional points on the track at each timestamp.

YYMMDDHH	Lat N/S	Lon E/W	VMAX	MSLP	STORMNAME
2008042512	105N	903E	20	1007	INVEST
2008042518	108N	895E	20	1007	INVEST
2008042600	107N	887E	20	1004	INVEST
2008042606	112N	885E	25	1002	NARGIS
2008042612	115N	879E	30	1000	NARGIS

Table 1. Data representation of cyclone Nargis

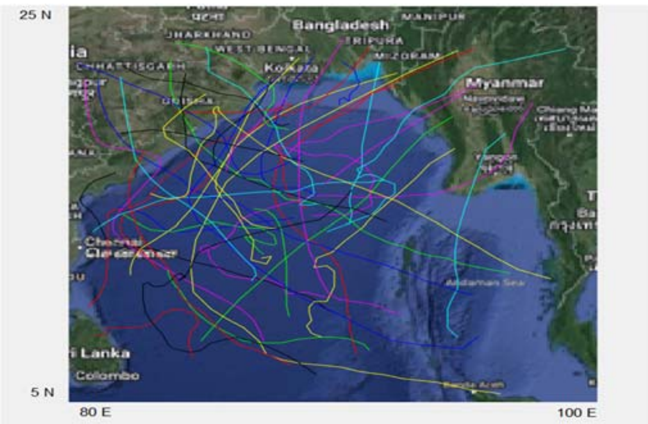


Fig. 1. Cyclones' tracks over North Indian Ocean

Predictors	Description
X_0	Current longitude
Y_0	Current latitude
P_0	Current minimum sea level pressure
WS_0	Current maximum sustained wind speed
D	Direction
M	Magnitude
TDI	Test Data Input
D_1	Direction of correlated cyclone 1
D_2	Direction of correlated cyclone 2
M_1	Magnitude of correlated cyclone 1
M_2	Magnitude of correlated cyclone 2

Table 2. Tropical cyclones' climatology elements and persistence elements in proposed system

3.2. *Proposed System*

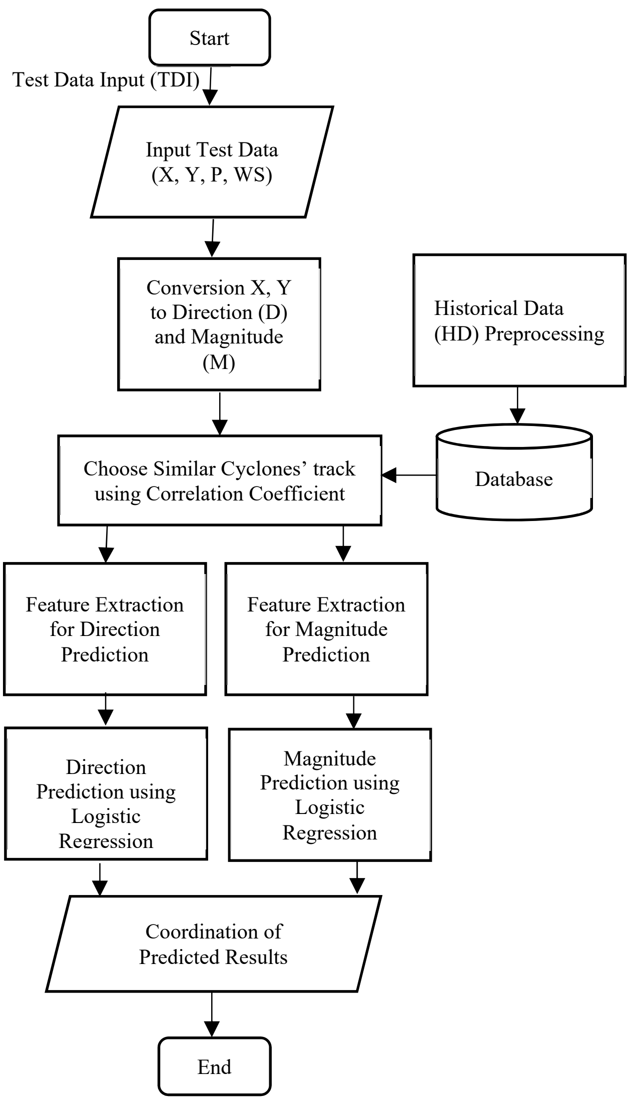


Fig. 2. Overview of the proposed system

Figure 2 depicts an overview of the proposed system, and the detailed steps are given in the following subsections. The latitude and longitude of all historical data and input test data are changed into direction and magnitude. The suggested approach mainly utilizes multiple logistic regression with pre-processed input data to generate useful features using mathematical equations. Cyclones occurred in the North Indian Ocean located from E40° to 100°, N 0° to 30°. Conversion of Latitude(Y) and Longitude(X) to direction and magnitude.

Direction (D) = θ

$$\tan\theta = \frac{Y}{X} \tag{1}$$

$$\theta = \tan^{-}\left(\frac{Y}{X}\right) \tag{2}$$

$$Magnitude(M)\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{3}$$

By using equation (2) latitude of the cyclone position is considered as Direction(D) and the longitude is transformed as Magnitude(M) by using equation (3).

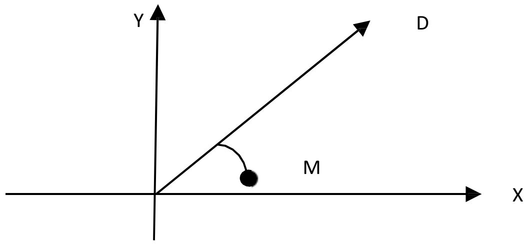


Fig. 3. Direction and magnitude

3.3. *Selection of cyclone with similar tracks*

Similar cyclone tracks are selected by using the value of the correlation coefficient. This method is very simple and effected to reduce the most similar cyclones with the tested cyclone's track. One way to define correlation is the strength of the relationship between two variables [Aguero, A. G.; Sayago, A. Gonzalez, A. G. (2006)]. For nominal and ordinal variables (as well as for time-series research), there are several measures of connection, and there are several correlation coefficients to address the unique properties of variables like dichotomies. To understand the notation, consider that the coefficient of determination is equal to the square of the correlation coefficient between x and y. That is then only true for a straight line [Kim, H-J.; Moon, I-J.; Kim, M (2019)].

$$r_{xy} = \pm\sqrt{R^2} = \sqrt{1 - \frac{S_{YY} - a_1{}^2 S_{XX}}{S_{YY}}} = a_1\sqrt{\frac{S_{XX}}{S_{YY}}} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \tag{4}$$

A popular statistical tool for describing straightforward relationships without demonstrating cause or effect, correlation is a measure of how much two variables are linearly related; correlations are also examined for statistical significance. The sample correlation coefficient, or r, indicates the stability of the connection. In this step, compute the correlation coefficient between the direction of test data input (TDI) and all historical data (HD). Select the two most correlated arrays from all historical data. Pair the Direction of the test data input DTDI with directions D1, and D2, and the Magnitude of test data input MTDI with M1 and M2. The degree of link between two variables is known as correlation.

e-ISSN : 0976-5166
p-ISSN : 2231-3850

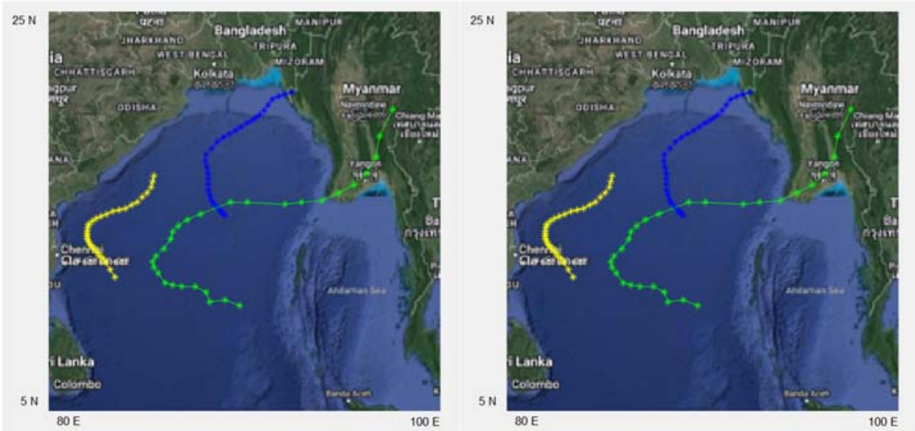


Fig. 4. Tested cyclone of Nargis and Mala with two correlated cyclones

3.4. *Feature Extraction*

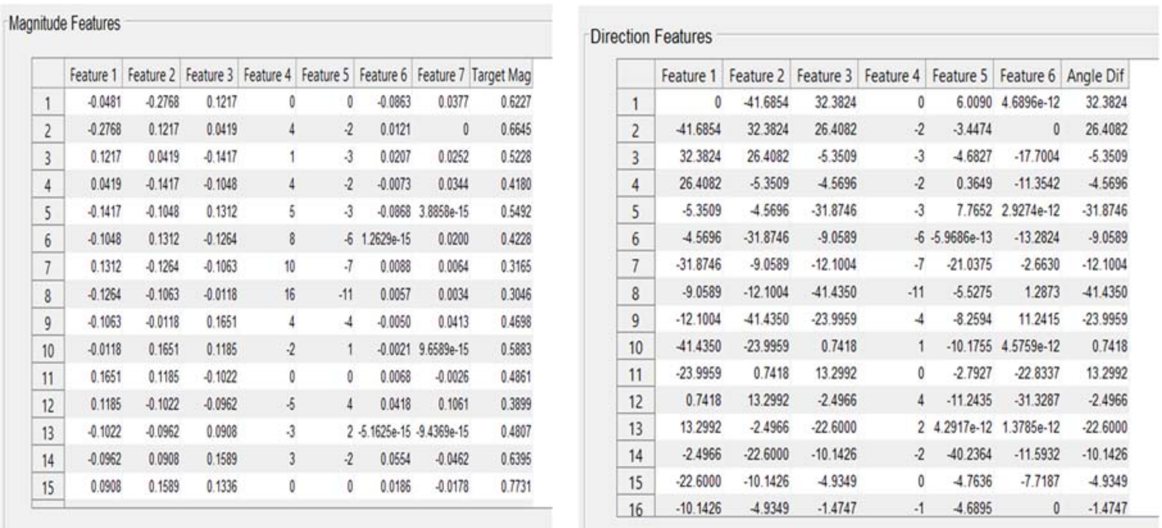
Feature extraction is an important stage for this research. Depending on the input extracted features, there will be different accuracy results. In this paper, direction is based on the sea level pressure and magnitude may be changed upon maximum sea level pressure and minimum wind speed of tested data input (TDI).

FE1	FE2	FE3	FE4	FE5	FE6	FE7(Target)
$D_{t_1-t}^{TDI}$	$D_{t_2-t_1}^{TDI}$	$D_{t_3-t_2}^{TDI}$	$P_{t_3}^{TDI}$	$D1_{t_3-t_2}$	$D2_{t_3-t_2}$	$D_{t_4}^{TDI}$

Table 3. Seven features for direction

FE1	FE2	FE3	FE4	FE5	FE6	FE7	FE8(Target)
$M_{t_1-t}^{TDI}$	$M_{t_2-t_1}^{TDI}$	$M_{t_3-t_2}^{TDI}$	$WS_{t_3}^{TDI}$	$P_{t_3}^{TDI}$	$M1_{t_3-t_2}$	$M2_{t_3-t_2}$	$M_{t_4}^{TDI}$

Table 4. Eight features for magnitude



Magnitude Features

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Target Mag
1	-0.0481	-0.2768	0.1217	0	0	-0.0863	0.0377	0.6227
2	-0.2768	0.1217	0.0419	4	-2	0.0121	0	0.6645
3	0.1217	0.0419	-0.1417	1	-3	0.0207	0.0252	0.5228
4	0.0419	-0.1417	-0.1048	4	-2	-0.0073	0.0344	0.4180
5	-0.1417	-0.1048	0.1312	5	-3	-0.0868	3.8858e-15	0.5492
6	-0.1048	0.1312	-0.1264	8	-6	1.2629e-15	0.0200	0.4228
7	0.1312	-0.1264	-0.1063	10	-7	0.0088	0.0064	0.3165
8	-0.1264	-0.1063	-0.0118	16	-11	0.0057	0.0034	0.3046
9	-0.1063	-0.0118	0.1651	4	-4	-0.0050	0.0413	0.4698
10	-0.0118	0.1651	0.1185	-2	1	-0.0021	9.6589e-15	0.5883
11	0.1651	0.1185	-0.1022	0	0	0.0068	-0.0026	0.4861
12	0.1185	-0.1022	-0.0962	-5	4	0.0418	0.1061	0.3899
13	-0.1022	-0.0962	0.0908	-3	2	-5.1625e-15	-9.4369e-15	0.4807
14	-0.0962	0.0908	0.1589	3	-2	0.0554	-0.0462	0.6395
15	0.0908	0.1589	0.1336	0	0	0.0186	-0.0178	0.7731

Direction Features

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Angle Dif
1	0	-41.6854	32.3824	0	6.0090	4.6896e-12	32.3824
2	-41.6854	32.3824	26.4082	-2	-3.4474	0	26.4082
3	32.3824	26.4082	-5.3509	-3	-4.6827	-17.7004	-5.3509
4	26.4082	-5.3509	-4.5696	-2	0.3649	-11.3542	-4.5696
5	-5.3509	-4.5696	-31.8746	-3	7.7652	2.9274e-12	-31.8746
6	-4.5696	-31.8746	-9.0589	-6	-5.9686e-13	-13.2824	-9.0589
7	-31.8746	-9.0589	-12.1004	-7	-21.0375	-2.6630	-12.1004
8	-9.0589	-12.1004	-41.4350	-11	-5.5275	1.2873	-41.4350
9	-12.1004	-41.4350	-23.9959	-4	-8.2594	11.2415	-23.9959
10	-41.4350	-23.9959	0.7418	1	-10.1755	4.5759e-12	0.7418
11	-23.9959	0.7418	13.2992	0	-2.7927	-22.8337	13.2992
12	0.7418	13.2992	-2.4966	4	-11.2435	-31.3287	-2.4966
13	13.2992	-2.4966	-22.6000	2	4.2917e-12	1.3785e-12	-22.6000
14	-2.4966	-22.6000	-10.1426	-2	-40.2364	-11.5932	-10.1426
15	-22.6000	-10.1426	-4.9349	0	-4.7636	-7.7187	-4.9349
16	-10.1426	-4.9349	-1.4747	-1	-4.6895	0	-1.4747

Fig. 5. Seven Features of direction and eight features of magnitude

3.5. *Multiple logistic regression*

A supervised machine learning technique called logistic regression is utilized for classification problems in which estimating the likelihood that an instance will belong to a particular class or not is the objective. A statistical method for analyzing the correlation between two variable components is called logistic regression. Logistic regression predicts the outcome of a categorical dependent variable. As a result, the outcome needs to be discrete or categorical. That can indicate Yes or No, 0 or 1, true or false, etc., but rather than providing a precise value between 0 and 1, it provides probabilistic values that fall in that range [Strzelecka, A; Kurdys-Kujawska, A: Zawadzka, D (2020)]. In logistic regression, the sigmoid function fits a logistic function with an "S" shape, which predicts two maximum values instead of a regression line (0 or 1). When there are two result categories and numerous independent feature variables, multiple logistic regression is employed. For multiple logistic regression, the following equations are applied.

$$P(y^{(i)}=1 \mid x^{(i)};\theta)=\sigma(\theta_0 x_0^{(i)}+\theta_1 x_1^{(i)}+\ldots+\theta_n x_n^{(i)})$$

$$=\sigma(\theta^T x^{(i)})=\frac{1}{1+e^{-\theta^T x^{(i)}}} \quad (5)$$

$$\theta^T=\left[\theta_0,\ldots.,\theta_n\right]$$

$$x^{(i)}=\begin{bmatrix} x_0^{(i)} \\ \ldots. \\ x_n^{(i)} \end{bmatrix} \quad (6)$$

i = i^{th} training example
n = number of features

$$J(\theta)=\frac{-1}{m}\sum_{i=1}^{m} y^{(i)}\log\left(\sigma\left(\theta^T x^{(i)}\right)\right)+\left(1-y^{(i)}\right)\log\left(1-\sigma\left(\theta^T x^{(i)}\right)\right)$$

$$\frac{\partial\left(J(\theta)\right)}{\partial\left(\theta_j\right)}=\frac{-1}{m}\sum_{i=1}^{m}\left(y^{(i)}-\theta^T x^{(i)}\right)x_j^{(i)}$$

(7)

$$\theta_j := \theta_j-\alpha\frac{\partial\left(J(\theta)\right)}{\partial\theta_j} \quad (8)$$

m = number of example
i = an example
j = feature j
α = learning rate

The logistic regression model's output is converted into a probability using the non-linear sigmoid function. In this research, three types of thresholds are used in the experiential result.

$$sigmoid(x)=1/(1+e^{(-x)}) \quad (9)$$

$$sigmoid(x)=1/(1+e^{(-x*0.5)}) \quad (10)$$

$$sigmoid(x)=1/(1+e^{(-x*1.5)}) \quad (11)$$

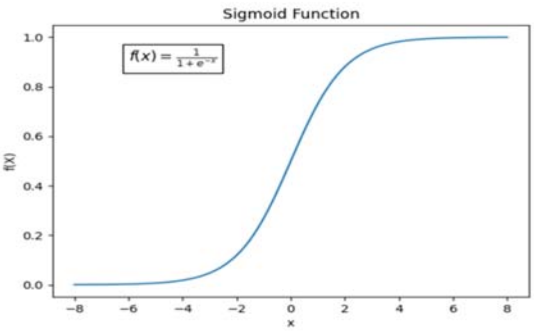


Fig. 6. Sigmoid Function

4. Experiment

4.1. *Performance Measurement*

To evaluate each method's performance between the forecast and actual tracks, the following metrics were selected: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). The MAE, which is determined by Equation (12), is the average of the absolute values of the deviations between all predicted and actual values [Kim,K.(2020)]. Poor model performance is shown by larger errors.

$$MAE = \frac{\sum_{i=1}^{h} |F_i - A_i|}{h} \tag{12}$$

where h is the number of historical data, Fi is the forecast value and Ai is the actual vale.

The average error rate for actual values is denoted by MAPE. It considers the proportion of the error to the actual tangency as well as the error between the forecast and actual tracks. Equation (13) provides the computation. A perfect model is indicated by a MAPE value of 0, and a poor model is indicated by a number greater than 1.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{F_i - A_i}{A_i} \right| \tag{13}$$

The square root of the difference between actual and forecast values as well as the number of data is known as the root mean square error, or RMSE. The following is the mathematical formulation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (F_i - A_i)^2}{n}} \tag{14}$$

The experimental results were forecasted by using the past twelve-hour locations of the tropical storm that occurs in North Indian Ocean. In this result, threshold value 0.5/ in using the sigmoid function is the best prediction result. MAE, MAPE, and RMSE are used to measure the performance of the proposed system.

Sigmoid Function	Longitude			Latitude		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Threshold 1	0.002757	0.003368	0.00397	0.003752	0.007548	0.001375
Threshold 0.5	0.002585	0.003365	0.003945	0.00358	0.007452	0.001348
Threshold 1.5	0.002627	0.003370	0.004949	0.003696	0.0076	0.001359

Table 5. Prediction results of Cyclone Nargis using MAE, MAPE and RMSE

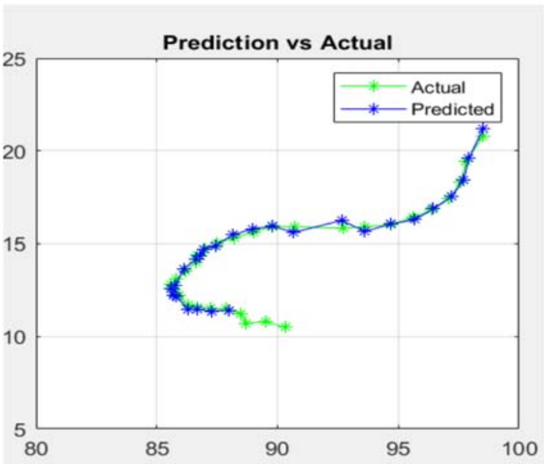
The accuracy of the result also depends on the number of similar tropical cyclones' tracks. Two, three and five similar tropical cyclones are tested with three threshold functions by using three measurement formulations. According to the result, three similar tropical cyclones are the best accuracy for this system and more similar cyclones can get less accuracy of the result.

Similar Cyclones	Sigmoid Function	Longitude			Latitude		
		MAE	MAPE	RMSE	MAE	MAPE	RMSE
Two Similar Cyclones' Tracks	Threshold 1	0.002728	0.003624	0.004357	0.003948	0.007828	0.001994
	Threshold 0.5	0.002635	0.003442	0.00420	0.003729	0.007739	0.001639
	Threshold 1.5	0.002836	0.003555	0.005382	0.003886	0.008362	0.001561
Three Similar Cyclones' Tracks	Threshold 1	0.002757	0.003368	0.00397	0.003752	0.007548	0.001375
	Threshold 0.5	0.002585	0.003365	0.003945	0.00358	0.007452	0.001348
	Threshold 1.5	0.002627	0.003370	0.004949	0.003696	0.0076	0.001359
Five Similar Cyclones' Tracks	Threshold 1	0.003286	0.004027	0.005836	0.004481	0.008824	0.001949
	Threshold 0.5	0.003629	0.003820	0.005183	0.003259	0.008272	0.002107
	Threshold 1.5	0.003925	0.004464	0.005982	0.004028	0.008661	0.001849

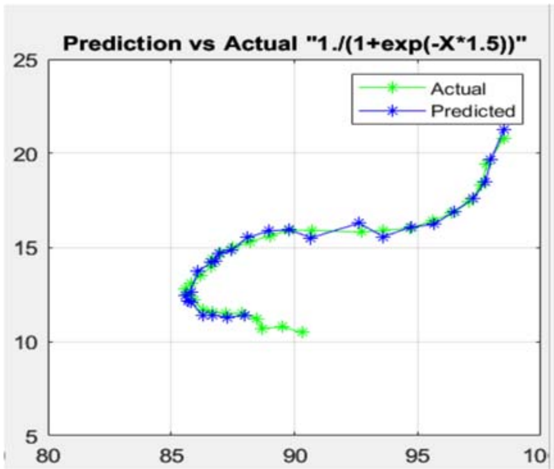
Table 6. Prediction results of Cyclone Nargis using Two, Three and Five Similar Cyclones Tracks

4.2. *Experimental Result*

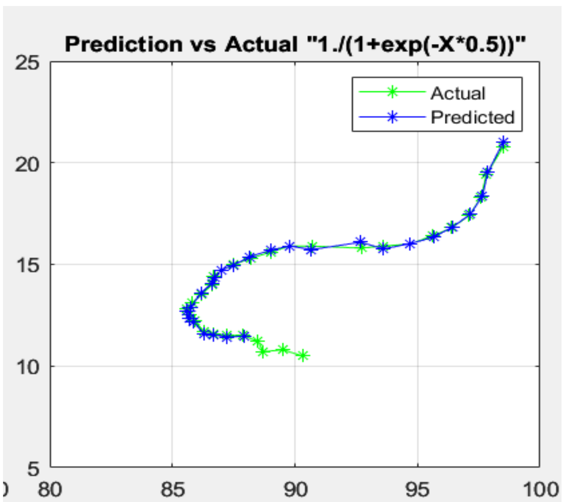
To verify which meteorological values were affected by the cyclone track, combined with the direction and magnitude (changing from latitude and longitude) of the tested cyclone and the two most correlated cyclones, maximum wind speed, and minimum sea level pressure were also used as the model input.



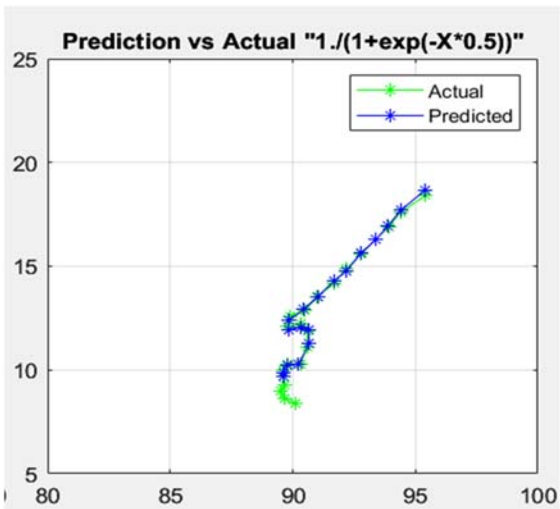
(a)



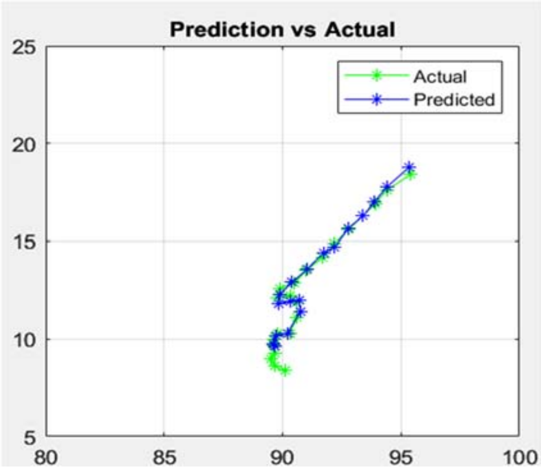
(b)



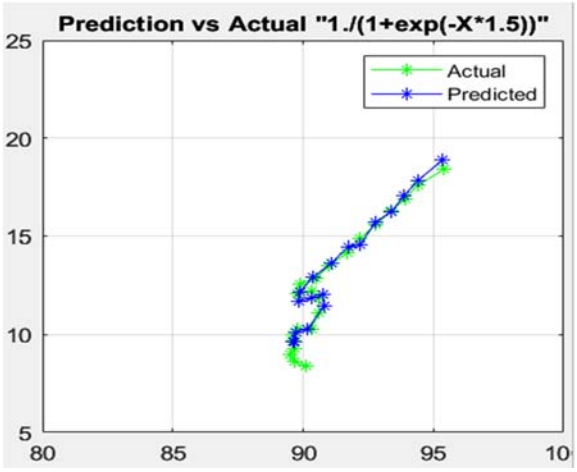
(c)



(d)



(e)



(f)

Fig. 7. Twenty-four-hour forecast result (a) Threshold value 1 of sigmoid function for cyclone Nargis (b) threshold value 0.5 (c) threshold value 1.5 (d) threshold value 0.5 for cyclone Mala (e) Threshold value 1 (f) Threshold value 1.5

5. Conclusion

In this paper, predicting tropical cyclone tracks in the North Indian Ocean are a big challenge for the cyclone track. Historical data are obtained from the Joint Typhoon Warning Center (JTWC) and the previous 24-hour tropical cyclone track is predicted using simple multiple logistic regression. To reduce data redundancy and complexity of the model, the main contribution is considering the tested cyclone and correlated cyclones. The second idea is changing the location of the cyclone to consider as direction and magnitude. Changing three threshold values of sigmoid function were used to compare the results. Another benefit of this proposed system provides a quick response with a few seconds when a new tropical cyclone track occurs in the ocean for short term prediction. In the future, integrating new data and extracting more useful features will be added to this system and a deep learning model will be applied to raise the accuracy of the predicted cyclone route.

Acknowledgments

I want to express my sincere gratitude to my best supervisor for all of their help, inspiration, and support during this study. Their vast expertise and creative ideas have been a huge help to me throughout my career.

Conflicts of Interest

The authors have no conflict of interest to declare.

References

[1] Ali, MM.; Kishtawal. C.Predicting cyclone tracks in the north Indian Ocean: An artificial neural network approach, Geophysical Research Letters, Vol, 34, L04603, (2007).

[2] Alemany, S.; Beltran, J.; Perez, A.; Ganzfried, S (2018). Predicting hurricane trajectories using a recurrent neural network. arXiv 2018, arXiv:1802.02548.

[3] Asuero, A. G.; Sayago, A. Gonzalez, A. G (2006), The Correlation Coefficient: An Overview, Critical Reviews in Analytical Chemistry, 36:41–59, 2006.

[4] Dong, P. et.al (2022), Tropical Cyclone Track Prediction with an Encoding-to-Forecasting Deep Learning Model, Donget Al, American Meteorological Society, June 2022.

[5] Kim, K. (2020), Improvement of Tropical Cyclone Track Forecast over the Western North Pacific Using a Machine Learning Method, master thesis, Department of Urban and Environmental Engineering, Graduate School of UNIST.

[6] Kim, H-J.; Moon, I-J.; Kim, M (2019), Statistical prediction of typhoon induced accumulated rainfall over the Korean Peninsula based on storm and rainfall data, research article, Metrological Application, 16, October, 2019.

[7] Lian, J.; Dong, P.; Zhnag, Y.; Pan, J. (2020). A Novel Deep Learning Approach for Tropical Cyclone Track Prediction Based on Auto-Encoder and Gated Recurrent Unit Networks, Article, Applied Sciences.

[8] Moradi Kordmahalleh, M.; Gorji Sefidmazgi, M.; Homaifar, A (2016). A sparse recurrent neural network for trajectory prediction of Atlantic hurricanes. In Proceedings of the ACM Genetic and Evolutionary Computation Conference 2016, Denver, CO, USA, 20–24 July 2016; pp. 957–969.

[9] Nakamuar, J. et.al. (2017), Western North Pacific tropical cyclone model tracks in present and future climates, J Geophys Res Atmos 122(18): 9721-9744.

[10] Richman, MB. B.; Lesile, L.M; Ramsay, H.A; Klotzbach, P.J (2017): Reducing Tropical Cyclone Prediction Errors Using Machine Learning Approaches, Prodedia Comput, Sci. 114, 314-323.

[11] Ravindra, V.; Nag, S.; Li, A (2020), Ensemble-Guided Tropical Cyclone Track Forecasting for Optimal Satellite Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing.

[12] Strzelecka, A; Kurdys-Kujawska, A: Zawadzka, D (2020): Application of logistic regression models to assess household finicical decisions regarding debt, Proceida computer science, 176, ppp.3418-3427.

[13] Tan, J.; Chen, S.; Wang, J. (2020), Western North Pacific tropical cyclone track forecasts by a machine learning model, Springer-Verlag GmbH Germany, part of Springer Nature 2020.

[14] Wang, L., Wan, B., Zhou, S., Sun, H., and Gao, Z(2023). Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model, Copernicus Publications, 16, 2167-2179.

[15] Zhang, Y. et.al (2018), Cyclone Track Prediction with Matrix Neural Networks, 2018 International Joint Conference on Neural Networks (IJCNN).

Authors Profile

Thu Zar Hsan received the M.C.Tech, degree from the University of Computer Studies, Mandalay in 2009. She is currently working as an associate professor in the University of Computer Studies, Myeik, Myanmar. Her other research interests are Image processing, Deep Learning and Embedded system. She can be contacted at email: thuzarhsan@ucsy.edu.mm and thuzarhsan@gmail.com.

Dr. Thin Lai Lai Thein is from the University of Computer Studies, Yangon (UCSY). She is working a professor in the Faculty of Information Science and has obtained degrees such as B.Sc(Maths), D.C.Sc, M.I.Sc, and Ph.D(IT). She is started working at University of Computer Studies, Yangon (25-8-1997) as a teacher. She has over 26 years of experience. She is teaching responsibility of the Faulty of Information Science and also being in charge of Data Analytic Lab (DAT Lab), She is also teaching doctoral students, supervision, and the international universities conducting research in collaboration with research organization. Now, the students of the Master's degree have been closely supervised and she is conducting research as the supervisor of 12 doctoral students. She can be contacted at email tllthein@ucsy.edu.mm.