IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 June 2023, accepted 29 June 2023, date of publication 6 July 2023, date of current version 13 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3292907

RESEARCH ARTICLE

Stacked Ensemble Model for Tropical Cyclone Path Prediction

KALIM SATTAR[1,2], SYEDA ZOUPASH ZAHRA[1,2], MUHAMMAD FAHEEM[3], (Member, IEEE), MALIK MUHAMMAD SAAD MISSEN[1], RAB NAWAZ BASHIR[2], (Senior Member, IEEE), AND MUHAMMAD ZAHID ABBAS[2], (Member, IEEE)

[1]Department of Information Technology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan
[3]School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland

Corresponding author: Muhammad Faheem (muhammad.faheem@uwasa.fi)

The work of Muhammad Faheem was supported in part by the University of Vaasa, and in part by the Academy of Finland.

ABSTRACT Tropical cyclones (TC) are intense circular storms that cause significant economic and human losses in the coastal areas of the equatorial region. Various statistical models have been proposed to forecast the potential path of TC. This study proposes a stacked ensemble-based method to enhance the effectiveness of predicting TC paths using temporal data. The proposed method can be divided into two phases. In the first phase, the Long Short-Term Memory Networks (LSTM) and Gated Recurrent Unit (GRU) models are optimized with stacked layers to determine the most effective configuration for Stacked LSTM and Stacked GRU. In the second phase, k-fold cross-validation is employed to construct multiple Stacked LSTM and Stacked GRU models, and a Meta learner is used to ensemble the predictions from these models. We evaluate the performance of our proposed model using the temporal China Meteorological Administration (CMA) dataset and compare its results with those obtained from other ensemble and non-ensemble techniques. The results demonstrate a significant reduction in mean square error and variance achieved by the proposed model. The code is available on GitHub: TC path prediction

INDEX TERMS Tropical cyclone, path prediction, stacked RNN, stacked ensemble.

I. INTRODUCTION

Tropical cyclones (TC) are intense circular storms that arise in the ocean and coastal areas of the equatorial region. The equatorial region is the region of the earth that exists between 23 degrees in the southern hemisphere and the northern hemisphere from the earth's equator. TC is named differently in different areas, such as cyclones, typhoons, and hurricanes. TC occurs in tropical and subtropical zone because of the vertical angle of the sunlight in the summer and autumn seasons.

International shipping in the area is impacted by tropical storms at sea because of the enormous waves, strong winds, and precipitation they produce. In contrast, the worst effects of the TC are their landfalling which causes substantial economic and life losses in the coastal areas [1]. Hence the accurate and timely prediction of the TC path is critical to saving the economy and life losses.

TC contains complex physical phenomena that make predicting their path difficult. Generally, there are six different

TABLE 1. TC path prediction techniques.

Technique	Advantages	Disadvantages
Weather Maps	Provides a visual representation of atmospheric conditions	Subjective interpretation, limited spatial and temporal resolution
Diagnostic Analysis	Identifies patterns and relationships in weather data	Requires expert knowledge and manual analysis
Mathematical Statistics	Quantifies statistical relationships in weather data	Assumes linear relationships, may not capture complex patterns Dynamic Models
(Numerical Forecasting Methods)	Incorporates physical laws and equations for accurate modeling	Computationally intensive, requires extensive data inputs
Statistical Dynamic Models	Combines statistical and dynamic modeling approaches	Balances accuracy and computational efficiency, requires calibration
Extrapolation	Simple and quick method, useful for short-term predictions	Assumes linear trend continuation, and limited accuracy for longer-term predictions

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca.

techniques used for TC path prediction. These techniques could be divided into two major categories, which are subjective and objective methods. Some of the important techniques for TC path prediction are described in 1. Much of the work has been done on statistical approaches in the last few years because they require fewer resources. The capability to provide the prediction is much faster than the dynamic models. Statistical techniques used the data collected from different sources such as previous storms, numerical simulation, synoptic analysis, and current storms. Statistical techniques are capable enough to capture the multiple combinations of parameters or variables in the obtained Dataset. Researchers developed several statistical techniques for Cyclone forecasting. Among those techniques, five popular techniques are climatology aware, statistical synoptic, steering airflow, climatology and persistence, and statistical dynamical procedures [1].

In some basins, the TC form in a particular season, and their motion is quite similar to the previous storm; climatologically-aware forecasting techniques use the observations and previous seasonal variations of the storms to forecast the movement of the current storm [2]. Using the mean velocity of previous hurricanes to forecast cyclones [3], analog cyclone forecasting [4], and cyclone forecasting through Markov chains are the methods used in climatologically aware forecasting.

Forecasting using climatology and persistence techniques is based on observing that cyclones continue to move in the same direction under stable atmospheric conditions. Climatology and persistence techniques combine data from both previous and current storms and can make short- and long-term predictions. In 1972, Neumann proposed a regression system known as (Climatology and persistence) CLIPER [5]. Many researchers used this method as a baseline in their research. In the statistical synoptic technique, pressure levels at different geopotential heights are considered one of the key features. Veigas-Miller [6], Chen-Elsberry [7], and National Hurrican Center Synoptic [8] techniques are essential statistical synoptic techniques. Steering airflow determination and statistical synoptic are quite similar for Cyclone forecasting; however, steering airflow considers the cyclone a vertex point. The surrounding winds can determine the direction and speed of the hurricane [9]. TC contains complex physical phenomena and non-linear relations that statistical models cannot deal with effectively.

In the past few years, Statistical Dynamic methods such as machine learning-based methods have provided compelling predictions in different aspects. Hence many researchers utilized the advantages of the machine learning-based model for TC path predictions. Researchers use machine learning algorithms such as linear regression [10], random forest [11], Decision tree [12], and support vector regression [13] for short-term and long-term forecasting for TC tasks. Since Neural networks can deal with the data's non-linear behaviors much more effectively, some researchers use neural network-based models for TC path prediction. Hybrid radial bases function (HRBF) with the time difference and structure learning (TDSL) is used for TC identification and trajectory mining [14]. Ali et al. [15] used the backpropagation with Multilayer perceptron (MLP) for TC track forecasting. Since RNN-based models such as LSTM and GRU have proved their effectiveness for the time series data, Alemany et al. [16] used the fully connected RNN for the TC path predictions. CNN [17], GAN [18], and Convo-LSTM [19] were also used for TC path prediction on imagery and Spatio-temporal data.

Machine learning proves an effective method for TC prediction; however, there is still room for developing new models to perform prediction more accurately and precisely. Ensemble learning has been proven effective for hydrological tasks [20]. Despite making one complex model, we can use simple models as weak learners and ensemble their predictions to achieve a better final prediction. This study presents a novel approach for TC path prediction using the stacked ensemble-based method. The contribution of this study is as follows:

- We optimized the RNN-based methods (LSTM, GRU) by using the stacking technique and proved the effectiveness of Stacked LSTM and Stacked GRU for the TC path prediction task.
- Propose a stacked ensemble-based method to combine and optimize the predictions of multiple stacked LSTM and stacked GRU.
- Evaluate the impact of different Meta learners in the stacked ensemble method and compare the prediction of the proposed method with other ensemble and non-ensemble techniques.

The content of the article is arranged as follows.Section II described the literature study relevant to TC path prediction. Section III discusses the preliminaries. Section IV explains the Stacked Ensemble based model. Section Vpresents the proposed approach's evaluation procedure, formulates the research questions to compare the proposed approach's performance outcomes and analyzes the threats. Section VI presents the conclusion of the paper.

II. RELATED WORK

Predicting the path of TC can be thought of as a trajectory prediction problem under the category of fluid objects. In this section, the techniques for path prediction are presented that are further categorized between classical and deep learning approaches.

A. TRADITIONAL TECHNIQUES FOR PATH PREDICTION

Path prediction can be considered a time series forecasting problem. The goal is to predict possible longitude and latitude after a particular time. Some of the researchers use linear systems for the path prediction task. Early studies use multivariate linear regression (MVLR) for TC path prediction [5]. However, MVLR is limited to path prediction tasks that are non-consecutive in position. Generally, the prediction of the latitude and longitude is considered a regression problem, so multiple autoregression models such as Autoregression (A.R.), Autoregression Moving Average (ARMA), Autoregression, and Integration Moving Average (ARIMA) for predicting the continuous path of TC [21]. Several machine learning-based methods have been applied for path prediction tasks under non-linear systems. Chau and Wu [22] used a support vector machine for the TC path prediction task. The author argues that a support vector machine can provide results much faster than Artificial Neural Network techniques. This method also used dimensionality reduction techniques such as principal component analysis (PCA) and wavelet decomposition to reduce the dimensionality of the data. Song et al. utilized the [23] to propose a hybrid model based on an Artificial Neural network and support vector regression for rainfall prediction. In this presented method, rainfall data were decomposed using singular spectrum analysis. The training set is distributed into low, medium, and high-intensity rainfall groups. An artificial Neural network is used for low and medium intensity while support vector regression is used for high-intensity rainfall. The proposed method is used for 1,2 and 3 days rainfall forecasting.

B. DEEP LEARNING BASED TECHNIQUES FOR PATH PREDICTION

Deep learning-based techniques have enormously gained in different domains [24], [25]. Researchers adopt several deep learning-based models to predict paths for both rigid and fluid object categories. Lee et al. [14] proposed a hybrid model for TC identification and track mining. This model is based on an oscillatory neural graph and a hybrid radial basis function. Ali et al. [15] used a multi-layer perceptron with backpropagation for the path prediction task. Recurrent Neural Network-based models such as Long Short-Term Memory (LSTM) [26] and Gated Recurrent Unit (GRU) [27] are the most popular models. Both LSTM and GRU were designed with a gated mechanism capable of remembering helpful information. Hense both models performed better on Time-series tasks such as Path prediction. Zhang et al. [28] used a matrix neural network to extract spatial information for TC track forecasting in the south Indian oceans. However, a matrix neural network is more suitable for image recognition tasks. Rüttgers et al. [29] used a generative adversarial network for cyclone path prediction on image data. CNN is another type of Neural network known for its spatial feature extraction ability [30]. Nishant et al. [31]used CNN for robot navigation and trajectory prediction; however, this model is based on 1D CNN which is too simple to capture essential features in complex phenomena. Giffard et al. [17] also used CNN and a fused network for TC path prediction on reanalysis data. In this study, three different CNN was used named Wind CNN, Pressure CNN, and Past Track and Meta CNN, the output of all three CNN, is combined with a fused Neural network. Zyner et al. [32] use the Recurrent Neural network and mixture density network to predict the driver's intention and possible path. The clustering algorithm ranks the possible paths based on their likelihood.

C. ENSEMBLE-BASED TECHNIQUES FOR PATH PREDICTION

Huang et al. [33] proposed a model for vehicle trajectory prediction. In this work, the author used GPS data as a Dataset and used two ensemble-based models, random forest and Adabooster, to predict the possible trajectory of the vehicle. Xiao et al. [34] also utilized ensemble-based classifiers random forest, gradient boosting, Decision tree, and XGBoost for human trajectory prediction tasks based on the transport used. The experiments show XGboost Shows better performance. However, XGBoost relies heavily on expert domain knowledge, which can cause unstable generalization. Wan et al. [35] designed a hybrid ensemble learning method for tourist route recommendation. In this method, location history, weather conditions, and temperature are used as features for the KNN classifier, and the Bayes classifier is used to estimate the prior probability.

III. PRELIMINARIES

This section introduces the basic concepts of the techniques that are the foundation of our proposed method for temporal data's TC path prediction task that includes stacked RNN and Stacked Ensemble. This section also discusses the Dataset and adopted pre-processing techniques.

A. TIME SERIES FORECASTING

Forecasting time series is the analytic process of predicting future values or trends in a succession of data points collected over time. This method of predicting is frequently employed in a variety of fields, including finance, economics, meteorology, and supply chain management. By studying historical data, recognizing trends, and comprehending seasonal patterns, time series forecasting tries to reveal significant insights and create accurate predictions of future occurrences. The capacity of RNN-based algorithms to capture temporal dependencies in sequential data makes them excellent for time series analysis. By retaining an internal memory, RNNs excel in processing and predicting based on historical context, making them ideally suited for forecasting, anomaly detection, and sequence synthesis in time series data [36].

B. STACKED RNN

RNN is robust neural network architecture often used to solve sequential and time series problems. Generally, the single-layer RNN is defined as follows.

$$h_t = f(w_x x_t + W_h h_{(t-1)} + b_h) \tag{1}$$

$$y_t = g(w_o h_t + b_o) \tag{2}$$

where f and g are the activation functions, h_t is the hidden state for the time step t, y_t is the output, W_x, W_h are input weights, W_o,are the output weights, b_h, b_o These are biased terms.

The key idea in stacked RNN is to increase the depth of the network by using multiple layers of RNN and the hidden state of one layer is used as an input state for another [37]. For k number of layers, Stacked RNN can define as follows.

$$h_t^{(1)} = f(W_x^{(1)} x_t + W_h^{(1)} h_{(t-1)}^{(1)} + b_h^{(1)}) \tag{3}$$

$$h_t^{(i)} = f(W_x^{(i)} h_t^{(i-1)} + W_h^{(i)} h_{(t-1)}^{(i)} + b_h^{(i)}) \tag{4}$$

$$y_t = g(w_o h_t^k + b_o) \tag{5}$$

For the initial layer, the input state will be x_t while the $h_t^{(1)}$ is the hidden state of the first layer, here i Denote the number of layers, and each layer will use the hidden state of the previous layer as an input state. h_t^k The hidden state for the k^{th} layer will be used to produce the network's output [38].

1) LSTM

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) architecture designed to detect long-term dependencies in sequential data. It accomplishes this by employing specialized memory cells that store and propagate data over extended time periods. The LSTM network consists of input, forget, and output gates that regulate the flow of data within the network [36].

2) GRU

The Gated Recurrent Unit (GRU) is an RNN architecture optimized for sequential data processing applications. It provides a simpler structure with fewer gates as an alternative to the more sophisticated Long Short-Term Memory (LSTM) paradigm. Utilizing update and reset gates to govern the flow of information inside the network, the GRU model successfully captures dependencies in sequential data. These gates allow the GRU to selectively store and update knowledge from prior time steps, making it a potent instrument for applications such as natural language processing and time series analysis [39].

C. STACKED ENSEMBLE

Stacking is one of the ensemble learning techniques in Machine learning, used to perform prediction in two different phases as shown in Fig 1. In phase 1, several classifiers are used as a base learner using training data. In phase 2, a Meta learner is used; a Meta learner uses the predictions of the base learners as input and makes final predictions. The significant advantage of the stacked ensemble, it involves the predictions of multiple models rather than only choosing the best model [40].

D. DATASET AND PRE-PROCESSING METHODS

1) DATASET

This study focuses on 2D feature extraction, so we only focus on the temporal dataset. The dataset we used for our experiments is the CMA TC Best Track dataset [41], [42]. This dataset is published by CMA (China Meteorological Administration) TC Data Center. This dataset is collected

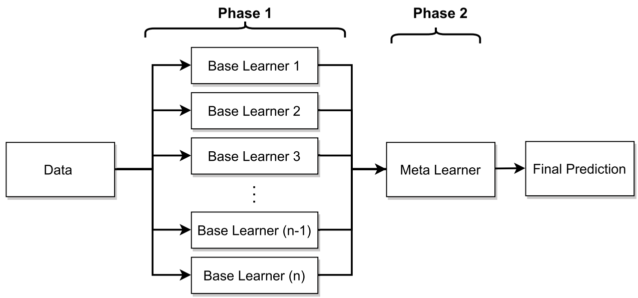


FIGURE 1. Layout of stacked ensemble.

TABLE 2. CLIPER method features [5].

Features ID	Features Description
1-5	Latitude for every 6-hour difference in the last 24 hours.
6-10	Longitude for every 6-hour difference in the last 24 hours.
11-15	Wind speed for every 6-hour difference in the last 24 hours.
16	Current month.
17-20	Historical latitude's 1^{st} order difference.
21-24	Historical longitude's 1^{st} order difference.
25-28	Historical wind speed 1^{st} order difference.
29-30	The sum of squares of 1st-order latitude and longitude difference.
31-32	The square root of features 29 and 30.
33-34	The square root of the present location.
35-38	Physical acceleration of historical location for every 6 h difference in the last 24 hours.
39-42	Zonal angle for every 6h difference in the last 18 hours.
43-46	Meridional angle for every 6h difference in the last 18 hours.
47-50	The angle between every 6h Historical location in last 24 hours.
51-53	The angle between every 6h Historical path in the last 24 hours.

from the Northwestern Pacific, including the South China Sea. The dataset contains six different attributes for every 6-hour difference for every typhoon occurring in this region since 1949. Attributes included in this dataset are DateTime, Intensity, Latitude, Longitude, Pressure, and Wind.

2) PRE-PROCESSING

Firstly, data cleaning is performed on the dataset to remove missing and duplicated values. After performing data cleaning, CMA Best Track dataset features are transformed into CLIPER method features [5], resulting in 53 features. Table 2 shows the details of the CLIPER method features.

IV. PROPOSED METHODOLOGY

The proposed methodology can be divided into 2 phases. In phase 1, stacking is applied on RNN-based neural networks (LSTM and GRU), and multiple stacked LSTM and GRU are trained as base learners on pre-processed data using the k-fold cross-validation technique. In phase 2, the meta-learner is trained by using the outputs of the multiple stacked neural networks. The step-by-step process of the proposed methodology is described in Fig 2.

A. STACKING OF LSTM AND GRU

We used a stacked-RNN-based technique defined above in the Preliminaries section to build stacked LSTM and stacked GRU. Two layers based on stacked LSTM and stacked GRU are shown in Fig 3.

Algorithm 1 Algorithm for Proposed Stacked Ensemble Model

Input: Let D be the CMA TC Best track dataset.
Output: Cyclone path
Initialisation:
1: Split D into train and test data such that $D = D_{trian}, D_{test}$.
2: Let N be the Total Number of Folds
3: F_N are folds created by K-fold cross-validation for D_{train}. Where $F_N \subseteq D_{Train}$ for each value of N.
4: Training Stacked LSTM and Stacked GRU as base learners.
5: Set $i = 1$
6: **while** $i <= len(F_N)$ **do**
7: $Stacked_{LSTM_{(i)}}(F_{[i]})$
8: $Stacked_{GRU_{(i)}}(F_{[i]})$
9: $i = i + 1$
10: **end while**
11: $Trained_{Stacked_{LSTM}} = \{Stacked_{LSTM_1}, Stacked_{LSTM_2}, \ldots, Stacked_{LSTM_N}\} = N$
12: $Trained_{Stacked_{GRU}} = \{Stacked_{GRU_1}, Stacked_{GRU_2}, \ldots, Stacked_{GRU_N}\} = N$
13: $Trained_{Models} = \{Trained_{Stacked_{LSTM}}, Trained_{Stacked_{GRU}}\} = 2N$
14: Train the Meta learner based on the output of the $Trained_{Models}$
15: $output[]$
16: Set $i = 1$
17: **while** $i <= len(Trained_{Models})$ **do**
18: $output[i] = Trained_{Models}[i](D_{train})$
19: $i = i + 1$
20: **end while**
21: $lat, long = Meta_{learner}(output[1], output[2], \ldots, output[2N])$
22: Repeat from step 15 to step 21 for D_{test} to get predicted latitude and longitude.
23: **return** $lat, long$

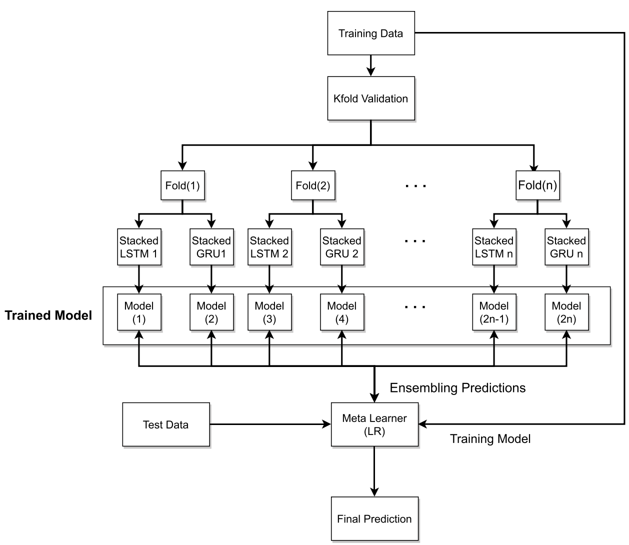


FIGURE 2. Proposed stacked ensemble based technique.

B. META LEARNER

In phase 2 of the proposed technique, a meta-learner is used to ensemble the output of the base learners. Linear Regression is used as the meta-learner in the proposed approach. In the training phase of the Linear Regression as a meta-learner, training data are passed to the base learners. In our approach, trained models work as base learners. Base learners predict the latitude and longitude of the given data. Each trained model can perform predictions independently. For each value, trained models predict 2N different latitudes and longitudes. These 2N predictions are used as inputs for the meta-learner that predicts the final latitude and longitude for the next 24 hours. Further, Algorithm 1 explains the proposed technique in detail.

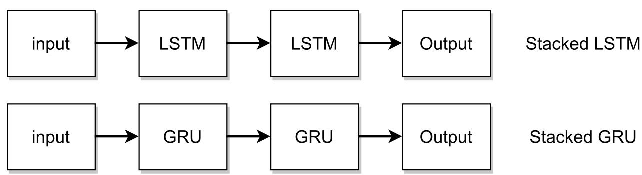


FIGURE 3. Two layer stacked LSTM and GRU.

V. EVALUATIONS

A. RESEARCH QUESTIONS

The evaluation of the proposed approach looks into the following questions:

Research Question 1: How does the proposed approach perform by comparing with existing techniques using temporal data?

Research Question 2: What effect does RNN-based stacking have on base learners in the proposed model?

Research Question 3: What effect does the stacked ensemble layer have on the proposed model?

Research Question 4: What effect does linear regression have as a meta-learner on the proposed model?

To evaluate the **Research Question 1**, modern machine learning/deep learning methodologies are compared to the proposed method. To evaluate the **Research Question 2**, stacked networks with a different number of layers are constructed for both LSTM and GRU, and the impact of stacking on base learners is evaluated. To evaluate the **Research Question 3**, the stacked layer is replaced with the BMA layer, and their performances are compared. To evaluate **Research Question 4**, the performance of different classifiers as meta-learners is evaluated.

B. PROCESS

In our experiments, we targeted the 24-hour prediction of TC tracks. We used the CMA best track dataset from 1949 to 2018. We employed 46332 samples from 2233 unique TC between (1949 and 2014) as training data and 2234 samples from 121 unique TC between (2015 and 2018) as test data. In k-fold cross-validation, we used 10 folds that produced 20 independent trained models. For stacked LSTM and stacked GRU, we used 4 and 3 stacked layers, respectively, further explained in Research Question 2 in the results section. We used linear regression as a meta-learner that predicts the final latitude and longitude location for the next 24 hours. The rest of the hyperparameters used for training the base learners in our experiments are as follows: batch size was 128, the optimizer was Adam, the learning rate was 0.001, input size was 53, hidden size was 128, and epoch size was 128.

C. METRICS

The proposed model predicts the latitude and longitude values that correspond to the typhoon's location for the next 24 hours for the given feature set. To evaluate the performance of the proposed model, we used Mean Squared Error and Variance as evaluation metrics.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2 \tag{6}$$

where MSE denotes mean squared error, n denotes the number of data points, Y_i denotes observed values, and $\tilde{Y}_i$ denotes the predicted values.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{(n-1)} \tag{7}$$

where S^2 denotes sample variance, x_i denotes a single observation value, $\bar{x}$ denotes the mean of all observation values, and n denotes the size of observations.

D. RESULTS

1) RESEARCH QUESTION 1: PERFORMANCE OF THE PROPOSED MODEL

To evaluate Research Question 1, we compare the results of our proposed model with other recent methods; existing methods contain both ensemble and non-ensemble-based approaches. We used conventional machine learning-based approaches in ensemble approaches: Random Forest, Decision Tree, Gradient Boosting, and voted regression-based ensemble methods. While in the non-ensemble-based techniques, we compare our technique with Linear Regression, LSTM, and GRU.

TABLE 3. 24 Hours prediction error (km).

Method	2015	2016	2017	2018
Random Forest	174.25	221.74	222.17	240.62
Decision Tree	265.47	315.06	301.57	324.45
Gradient Boosting	176.67	207.59	216.23	235.86
Linear Regression	170.75	215.52	216.44	212.81
LSTM	170.89	209.10	217.78	218.51
GRU	165.63	202.92	209.16	206.11
XGBoost	201.16	226.23	247.65	246.28
LightGBM	186.85	220.54	217.37	236.31
VR Ensemble	170.99	214.76	226.58	228.00
Stacked Ensemble	**161.88**	**195.62**	**200.57**	**196.29**

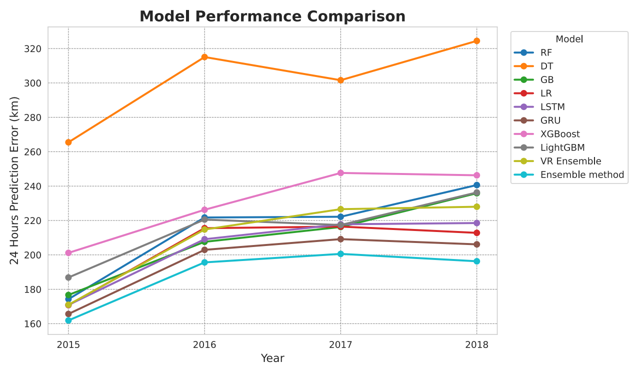


FIGURE 4. Model comparison with baseline models.

Table 3 displays the predicted errors in kilometers (km) spanning four years for several models. The proposed ensemble model consistently exhibits the lowest prediction error throughout all four years, suggesting its improved accuracy in estimating the TC path. In addition to the individual years, the ensemble model has the lowest average prediction error across all four years compared to the other models. This indicates that the ensemble model offers consistently more precise predictions.The visual representation of the results is depicted in 4 and 5

To validate the significant difference between the proposed approach and baseline approaches, we employ single-factor ANOVA analysis. ANOVA (Analysis of Variance) is a statistical technique used to assess whether or not there are statistically significant differences between the means of two or more groups. It evaluates the variability within and across groups in order to establish whether the observed differences are likely due to random chance or to true group differences. The prediction we used for ANOVA analysis

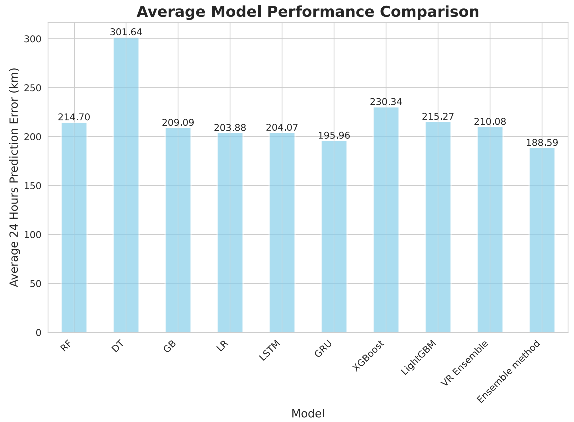


FIGURE 5. Model comparison Based on average error.

is obtained on the same dataset and objective to prove the significant difference among different approaches. We used Microsoft Excel to compute ANOVA analysis with its default setting without doing any manual adjustments. The results of ANOVA analysis are described in Fig 6. That shows $F > F_{cric}$ and $p - value < (alpha = 0.05)$, where $F = 7.40$, $F_{cric} = 2.41$, $p - value = 0.000013$. It proves that there is a significant difference between multiple approaches. Furthermore, the results in Fig 6 also show the average mean squared error and variance for Table 3. That shows the stacked ensemble method achieved an average mean squared error of **188.59 km** and variance of **321.96**, the lowest in all models. It is also notable that the proposed approach significantly reduced the average mean squared error by**3.9%** and variance up to **25.3%** against the baseline GRU model that performs best among baseline models.

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Random Forest	4	858.7972	214.6993	804.5164
Decision Tree	4	1206.581	301.6453	669.676
Gradient Boosting	4	836.3694	209.0924	606.9966
Linear Regression	4	815.5384	203.8846	490.2073
LSTM	4	816.2986	204.0746	507.629
GRU	4	783.8441	195.961	415.1834
XGBoost	4	921.3435	230.3359	473.9922
LightGBM	4	861.07	215.2675	427.5203
VR Ensemble	4	840.3317	210.0829	714.3913
Ensemble method	4	754.3798	188.5949	321.9682

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	36194.36589	9	4021.596	7.403417	1.32E-05	2.210697
Within Groups	16296.24243	30	543.2081			
Total	52490.60831	39				

FIGURE 6. ANOVA analysis for proposed model and baseline models.

To evaluate the Real-time performance of the proposed model we choose TC "KONG-REY" with unique TID 2350 originated in the western North Pacific (13.6 °N and 138.7 °E) on September 28, 2018. The visualization of the historical and predicted path for TC "KONG-REY" is shown in Fig 7. The proposed model achieved a prediction error of **148.0 km** which is **9.2%** less compared to the baseline GRU model.

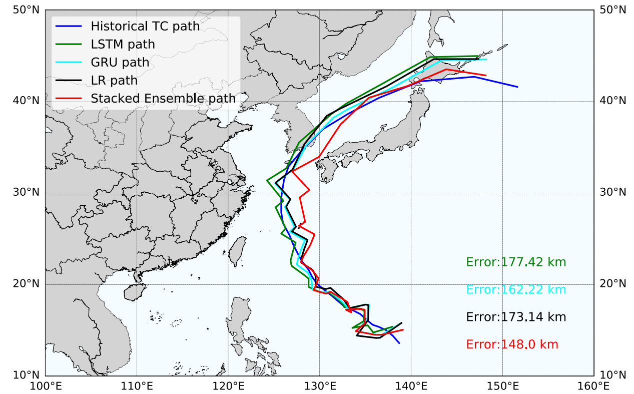


FIGURE 7. Visualization of prediction error for TC KONG-REY.

2) RESEARCH QUESTION 2: INFLUENCE OF STACKING ON BASE LEARNERS

To answer research question 2, we evaluate the performance of stacked LSTM and Stacked GRU models with different numbers of stacked layers. We used layer combinations between 2 and 8 and chose the best possible layer combination for stacked LSTM and GRU. Visualization of the stacking impact on base learners is shown in Fig 8.

Stacked LSTM: Stacked LSTM with four layers produced an average mean squared error of **191.26** km, which was the lowest among other layer combinations.

Stacked GRU: A stacked GRU with three layers outperforms other layer combinations and produced an average mean squared error of **191.35** km,

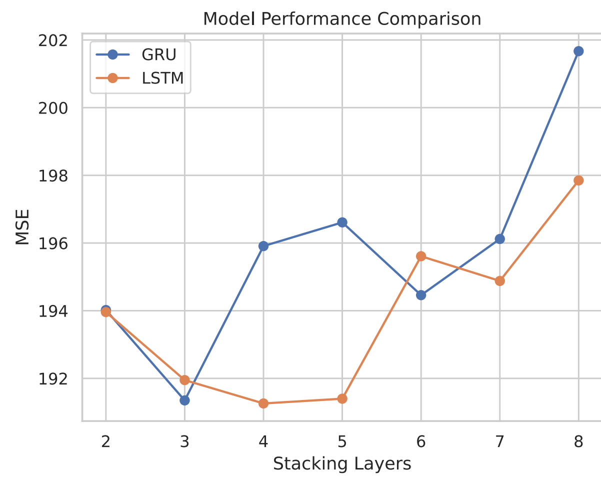


FIGURE 8. Impact of stacking on base-learners.

3) RESEARCH QUESTION 3: IMPACT OF STACKED ENSEMBLE LAYER

To evaluate the impact of the stacked ensemble layer, we compare our ensemble model with the Bayesian Model Averaging technique by replacing the stacked ensemble layer with the BMA layer. A comparison between Bayesian Model Averaging and the Stacked Ensemble Model is shown in Fig 9. The stacked ensemble model significantly outperforms the BMA model in average mean squared error with a difference of **23.86 km**, which is **11.9 %** lower than the BMA ensemble method.

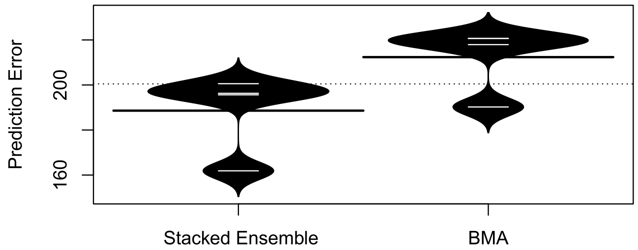


FIGURE 9. Plot for ensemble methods.

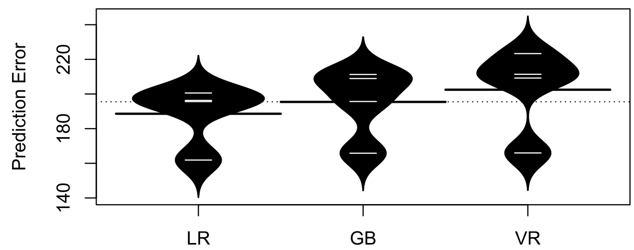


FIGURE 10. Plot for candidate meta learners.

4) RESEARCH QUESTION4: INFLUENCE OF LINEAR REGRESSION AS META LEARNER IN THE PROPOSED APPROACH

To evaluate the impact of the meta-learner on the stacked ensemble method, we used different machine-learning classifiers as meta-learners. We chose linear regression, and gradient boosting, and voted regression as candidates for the meta-learner to evaluate research question 4.

The impact of candidate meta-learners is shown in Fig 10. It is notable that linear regression performed best among other candidates concerning the average mean squared error.

5) THREATS

A threat to external validity is that TC is a complex physical phenomenon. The dataset we used in this research was collected in the western North Pacific. However, the performance of the proposed approach is significant. However, results may not hold for other basins. A threat to the construct validity is that the features we used in this research were based on the CLIPER method, which may not correctly define the physical phenomena of TC among different basins. Numerous elements, such as air conditions, sea surface temperatures, geography, and interactions with other weather systems, influence the behavior of TC. The CMA best track dataset may not capture all of the complexities and uncertainties associated with these interactions, making it challenging to accurately predict cyclone tracks using this data alone.

VI. CONCLUSION

TC track prediction is one of the complex physical phenomena. Accurate track prediction can save human life and financial loss. This paper has presented a novel technique to improve the TC track prediction based on temporal data. Our proposed stacked ensemble-based method uses stacked LSTM and stacked GRU as base learners and linear regression as a meta-learner. China Meteorological Administration (CMA) dataset is used to evaluate the proposed model. The proposed technique was compared with the state-of-the-art ensemble and non-ensemble-based techniques. Compared to the other state-of-the-art models, the stacked ensemble-based method significantly reduced both average mean squared error and variance. For future work, we plan to use an ensemble-based method for other TC path prediction tasks, such as intensity prediction and storm surge prediction. Furthermore, we plan to use the ensemble-based method for spatial-temporal data.

ACKNOWLEDGMENT

The authors are highly grateful to their affiliated universities and institutes for providing research facilities.

REFERENCES

[1] C. Roy and R. Kovordányi, "Tropical cyclone track forecasting techniques—A review," *Atmos. Res.*, vols. 104–105, pp. 40–69, Feb. 2012.
[2] B. Harper, J. Kepert, and J. Ginger, "Guidelines for converting between various wind averaging periods in tropical cyclone conditions," Citeseer, Princeton, NJ, USA, World Meteorolog. Org. (WMO), Geneva, Switzerland, Tech. Rep. P364772, 2010.
[3] R. A. Jeffries and R. J. Miller, "Tropical cyclone forecasters reference guide. 3. Tropical cyclone formation," Naval Res. Lab. Monterey, CA, USA, Tech. Rep., AD-A265216, 1993.
[4] J. R. Hope and C. J. Neumann, "An operational technique for relating the movement of existing tropical cyclones to past tracks," *Monthly Weather Rev.*, vol. 98, no. 12, pp. 925–933, Dec. 1970.
[5] C. J. Neumann, "An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system," Nat. Weather Service., Southern Region, Sci. Services Division, USA, Tech. Rep. 62, 1972.
[6] K. W. Veigas, R. G. Miller, and G. M. Howe, "Probabilistic prediction of hurricane movements by synoptic climatology," Travelers Insurance Company, New York, NY, USA, Tech. Rep. 169, 1959.
[7] J.-M. Chen, R. L. Elsberry, M. A. Boothe, and L. E. Carr, "A simple statistical-synoptic track prediction technique for western north Pacific tropical cyclones," *Monthly Weather Rev.*, vol. 127, no. 1, pp. 89–102, Jan. 1999.
[8] C. J. Neumann, "A guide to Atlantic and eastern Pacific models for the prediction of tropical cyclone motion," World Meteorolog. Org. (WMO), Geneva, Switzerland, Tech. Rep. 541, 1979.
[9] G. Holland, "Global guide to tropical cyclone forecasting," Bureau Meteorol. Res. Center, Melbourne, VIC, Australia, Tech. Rep. 6372, 2011.
[10] J. S. Wijnands, G. Qian, and Y. Kuleshov, "Variable selection for tropical cyclogenesis predictive modeling," *Monthly Weather Rev.*, vol. 144, no. 12, pp. 4605–4619, Dec. 2016.
[11] D. Ahijevych, J. O. Pinto, J. K. Williams, and M. Steiner, "Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique," *Weather Forecasting*, vol. 31, no. 2, pp. 581–599, Apr. 2016.

[12] W. Zhang, B. Fu, M. S. Peng, and T. Li, "Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis," *Weather Forecasting*, vol. 30, no. 2, pp. 446–454, Apr. 2015.

[13] M. B. Richman and L. M. Leslie, "Adaptive machine learning approaches to seasonal prediction of tropical cyclones," *Proc. Comput. Sci.*, vol. 12, pp. 276–281, Jan. 2012.

[14] R. S. T. Lee and J. N. K. Liu, "Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid RBF network track mining techniques," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 680–689, May 2000.

[15] M. M. Ali, C. M. Kishtawal, and S. Jain, "Predicting cyclone tracks in the north Indian ocean: An artificial neural network approach," *Geophys. Res. Lett.*, vol. 34, no. 4, 2007, Art. no. L04603.

[16] S. Alemany, J. Beltran, A. Perez, and S. Ganzfried, "Predicting hurricane trajectories using a recurrent neural network," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 468–475.

[17] S. Giffard-Roisin, M. Yang, G. Charpiat, C. K. Bonfanti, B. Kégl, and C. Monteleoni, "Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data," *Frontiers Big Data*, vol. 3, p. 1, Jan. 2020.

[18] M. Rüttgers, S. Lee, and D. You, "Prediction of typhoon tracks using a generative adversarial network with observational and meteorological data," 2018, *arXiv:1812.01943*.

[19] S. Kim, H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat, "Deep-Hurricane-tracker: Tracking and forecasting extreme climate events," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1761–1769.

[20] M. K. Tiwari and C. Chatterjee, "Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach," *J. Hydrol.*, vol. 394, nos. 3–4, pp. 458–470, Nov. 2010.

[21] W. W. Wei, "Time series analysis," in *The Oxford Handbook of Quantitative Methods in Psychology*, vol. 2. Sweden: Adlibris, 2006.

[22] K. W. Chau and C. L. Wu, "A hybrid model coupled with singular spectrum analysis for daily rainfall prediction," *J. Hydroinformatics*, vol. 12, no. 4, pp. 458–473, Oct. 2010.

[23] H.-J. Song, S.-H. Huh, J.-H. Kim, C.-H. Ho, and S.-K. Park, "Typhoon track prediction by a support vector machine using data reduction methods," in *Proc. Int. Conf. Comput. Inf. Sci.* Cham, Switzerland: Springer, 2005, pp. 503–511.

[24] M. Woźniak, M. Wieczorek, J. Siłka, and D. Połap, "Body pose prediction based on motion sensor data and recurrent neural network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2101–2111, Mar. 2021.

[25] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder–decoder architecture," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1672–1678.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[28] Y. Zhang, R. Chandra, and J. Gao, "Cyclone track prediction with matrix neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[29] M. Rüttgers, S. Lee, S. Jeon, and D. You, "Prediction of a typhoon track using a generative adversarial network and satellite images," *Sci. Rep.*, vol. 9, no. 1, Apr. 2019.

[30] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[31] N. Nikhil and B. Tran Morris, "Convolutional neural network for trajectory prediction," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 1–11.

[32] A. Zyner, S. Worrall, and E. Nebot, "Naturalistic driver intention and path prediction using recurrent neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1584–1594, Apr. 2020, doi: 10.1109/TITS.2019.2913166.

[33] K. Huang and M. Sun, "Machine learning based path prediction system—Adapting one model for all intersections," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Nov. 2018, pp. 64–69.

[34] Z. Xiao, Y. Wang, K. Fu, and F. Wu, "Identifying different transportation modes from trajectory data using tree-based ensemble classifiers," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 2, p. 57, Feb. 2017.

[35] L. Wan, Y. Hong, Z. Huang, X. Peng, and R. Li, "A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks," *Int. J. Geographical Inf. Sci.*, vol. 32, no. 11, pp. 2225–2246, Nov. 2018.

[36] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1394–1401.

[37] J. Lambert, "Stacked RNNs for encoder–decoder networks: Accurate machine understanding of images," FL, USA, Tech. Rep., Dec. 2018.

[38] J. Turek, S. Jain, V. Vo, M. Capotă, A. Huth, and T. Willke, "Approximating stacked and bidirectional recurrent architectures with the delayed recurrent neural network," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 9648–9658.

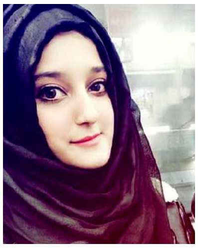
[39] P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between ARIMA, LSTM, and GRU for time series forecasting," in *Proc. 2nd Int. Conf. Algorithms, Comput. Artif. Intell.*, Dec. 2019, pp. 49–55.

[40] L. Rokach, *Ensemble Learning: Pattern Classification Using Ensemble Methods*. Singapore: World Scientific, 2019.

[41] M. Ying, W. Zhang, H. Yu, X. Lu, J. Feng, Y. Fan, Y. Zhu, and D. Chen, "An overview of the China meteorological administration tropical cyclone database," *J. Atmos. Ocean. Technol.*, vol. 31, no. 2, pp. 287–301, Feb. 2014.

[42] X. Lu, H. Yu, M. Ying, B. Zhao, S. Zhang, L. Lin, L. Bai, and R. Wan, "Western north Pacific tropical cyclone database created by the China meteorological administration," *Adv. Atmos. Sci.*, vol. 38, no. 4, pp. 690–699, Apr. 2021.

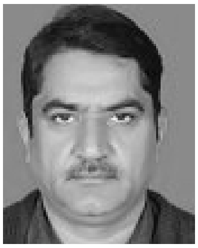
KALIM SATTAR received the B.S. degree in computer science from the COMSATS Institute of Information Technology, Pakistan, in 2014, and the M.S. degree in computer science from the Institute of Southern Punjab, Pakistan, in 2017. He is currently a Ph.D. Scholar with the Department of Information Technology, The Islamia University of Bahawalpur. His research interests include natural language processing, machine learning, and data mining.



SYEDA ZOUPASH ZAHRA received the B.S. degree in electrical engineering from UET Lahore, and the M.Phil. degree from NCBAE Lahore. She is currently a Ph.D. Scholar with the Department of Information Technology, The Islamia University of Bahawalpur. Her research interests include fuzzy integral, ensemble learning, and the applications of machine learning.

MUHAMMAD FAHEEM (Member, IEEE) received the B.Sc. degree in computer engineering from the University College of Engineering and Technology, Bahauddin Zakariya University (BZU), Multan, Pakistan, in 2010, the M.S. degree in computer science from Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2012, and the Ph.D. degree in computer science from the Faculty of Engineering, UTM, in 2021. In the past, he was a Lecturer with the COMSATS Institute of Information and Technology, Pakistan, from 2012 to 2014. In addition, he was an Assistant Professor with the Department of Computer Engineering, Abdullah Gul University (AGU), Kayseri, Turkey, from 2014 to 2022. He is currently a Researcher with the School of Computing Science (Innovations and Technology), The University of Vaasa, Vaasa, Finland. He has numerous publications in journals and international conferences. His research interests include cybersecurity, industry 4.0, smart cities, smart grids, and underwater sensor networks.

MALIK MUHAMMAD SAAD MISSEN is currently an Associate Professor with the Department of Information Technology, The Islamia University of Bahawalpur. His research interests include information retrieval/processing, web usability engineering, and software quality assurance.



RAB NAWAZ BASHIR (Senior Member, IEEE) was born in Pakistan. He received the Ph.D. degree in computer science from The Islamia University of Bahawalpur, Pakistan, in 2021.

From 2015 to 2023, he was an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Vehari campus. He is the author of more than 20 articles. His research interests include the IoT and machine learning applications in agriculture to support sustainable developments in agriculture.

MUHAMMAD ZAHID ABBAS (Member, IEEE) received the B.Sc. degree from BZU Multan, Pakistan, the master's degree in computer science from PUCIT, University of the Punjab, Lahore, Pakistan, in 2004, the M.S. degree in computer science from Uppsala University, Uppsala, Sweden, in 2008, and the Ph.D. degree in computer science from UTM, Malaysia, in 2017. He is currently an Assistant Professor with the Faculty of Computer Science, COMSATS University Islamabad, Pakistan. His research interests include routing and monitoring in wireless sensor networks, UWSN, LSN, the IoT, and security.