

Analyse of Weather Forecasting Models

Using LSTM with Noise Removal Methods

Nge^{a*}, Nyein Nyein Oo^b

^aDepartment of Information Technology Engineering, Technological University (Thanlyin), Yangon, Myanmar

^bDepartment of Computer Engineering and Information Technology, Yangon Technological University, Yangon, Myanmar

^aEmail: dawnge0722@gmail.com

^bEmail: nyeinnyeinoo2024@gmail.com

Abstract

Weather forecasting is one of the most important fields for all sectors such as transportation (air traffic, marine), industry, agriculture, forestry and even public health sector. The aim of this study is to analyse the weather forecasting models using filtering techniques which can remove the noise involved in the time series weather data. We utilize ten years weather data (2013 to 2022) in Hmawbi region, Yangon, Myanmar. The weather dataset includes maximum and minimum temperatures, humidity, wind speed, cloud amount and weather type. It is taken from the Department of Meteorology and Hydrology, Myanmar. The original data includes noise data. Thus, two noise removal methods, Simple Moving Average and Weighted Moving Average, are used for data cleaning. In this paper, three prediction models are developed by using Long Short-Term Memory (LSTM) with different datasets: one with the original data (without noise removal), one with data cleaned using the Simple Moving Average method, and another with data cleaned using the Weighted Moving Average method. The research goal is to compare and evaluate the performance of these three weather prediction models to determine which one gives the better results. To evaluate these models, Root Mean Square Error (RMSE) and Mean Square Error (MSE) are calculated for each model. LSTM achieves RMSE of 5.398 and MSE of 29.020, SMA achieves RMSE of 3.767 and MSE of 14.841, WMA achieves RMSE of 0.798 and MSE of 0.534. According to the experimental results, it is found that model with data cleaned using the Weighted Moving Average method is lower error rates than other two models and gives the better predicting results than other two models.

Received: 2/3/2025

Accepted: 4/10/2025

Published: 4/22/2025

* Corresponding author.

Keywords: Accuracy Analyze; Long Short-Term Memory; Data Analysis; Deep Learning; Prediction Models; Simple Moving Average; Time Series Analysis; Weighted Moving Average.

1. Introduction

Forecasting, in context of data analysis and machine learning, refers to the process of predicting future values or trends based on historical data and patterns. It's a crucial aspect of decision-making in various domains, such as economics, finance, supply chain management, weather prediction, and more. Deep learning, a subset of machine learning, has emerged as a powerful tool for forecasting due to its ability to capture complex patterns and relationships in data. It can handle intricate temporal dependencies and non-linear relationships that are challenging for traditional statistical methods. Traditional forecasting methods often struggle to capture intricate patterns in complex datasets. Deep learning has revolutionized time series forecasting, enabling more accurate predictions and informed decisions.

This paper utilizes a long short-term memory (LSTM) to enhance weather forecasting, employing simple moving average (SMA) and weighted moving average (WMA) methods for noise reduction in time series data to achieve highly accurate classification outcomes. The paper's contributions include:

- (i) The application of LSTM is taken for weather forecasting.
- (ii) The utilization of SMA and WMA are applied for noise reduction in time series data, thereby enhancing the accuracy of classification results. This experiment uses previous 10 years weather data to avoid large difference in weather changes along the years.

2. Materials and Methods

In this paper, the researchers gathered historical weather data from various sources, such as temperature, humidity, wind speed, precipitation, and atmospheric pressure, specific the Shenzhen region [1]. They developed a deep learning model, which is a type of artificial neural network, to analyse and learn from this vast dataset. The deep learning model utilizes a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal patterns in the weather data. CNNs are effective in recognizing spatial patterns, such as temperature variations across different geographical locations, while RNNs excel in capturing temporal patterns, such as daily or seasonal weather changes. Once the deep learning model is trained, it can make accurate predictions of various weather parameters for the next day in Shenzhen City. These predictions include temperature, humidity, wind speed, precipitation, and other relevant factors. The model considered both the historical data and the current weather conditions to generate forecasts. The study evaluated the performance of the deep learning model by comparing its predictions with actual observed weather data. It measured various statistical metrics, such as mean absolute error (MAE) and root mean square error (RMSE), to assess the accuracy and reliability of the forecasts.

The authors implemented, the Deep Learning-Based Weather Prediction using Long Short-Term Memory (LSTM) which is an artificial Recurrent Neural Network (RNN) architecture [2]. Like numerical weather

prediction, deep learning-based weather prediction is quite sensitive to the commencement value. The precision of forecast results is strongly influenced by the quality of the input datasets. Sensors and autonomous observing platforms, spanning ocean-based, ground-based, air-based, and space-based platforms, have accumulated Pressure

Barometric (PB) of meteorological observation data. The authors described data mining techniques for agricultural meteorological features collected from Bengaluru district's meteorological centre [4]. The feature extraction with hierarchical clustering and K-means which play an important role in the decision making for the experimental result showed that the hierarchical approach supports better than K-means clustering gives effectiveness for prediction of weather.

The author studied how CNN could classify weather from images and evaluated the recognition output of the layers of both the ImageNet-CNN and the Weather trained CNN [5]. In the role of weather classification, the method outperforms the latest technology by a large margin. They also looked at how all of the layers of CNNs behaved, and some interesting results were discovered.

The author introduced the application of the geostatistical interpolation technique known as Kriging for the short-term forecasting of weather conditions [7]. It has proven effective in accurately representing the spatial-temporal distribution of wind and temperature, thereby enabling the production of high quality, localized short-term weather predictions, complete with measurements of uncertainty. A thorough evaluation, including cross-validation using multiple methods, was carried out to demonstrate the precision of capturing the spatial-temporal distribution of these weather variables through the generated wind and temperature models.

In [8] presented that deep learning can be used to identify extreme weather patterns through climate data. Deep CNN architecture was created to identify tropical cyclones, weather fronts, and atmospheric rivers. It's the first-time deep CNN has been used to solve problems with climate pattern recognition. This promising application could pave the way for tackling a wide range of pattern detection issues for climate science. The findings of this study will be used to measure the current and future trends in climate extreme weather events, and also investigate changes in thermodynamics and dynamics of extreme events in the face of global warming.

- **Deep Learning**

Deep learning is a kind of artificial intelligence and machine learning that imitates the path humans take to achieve specific kinds of knowledge [3]. This learning is a potential thing in data science that contains modelling of predictive and statistical methods. This is more efficient for data scientists who are tasked with the analysis, translation, and collection of a huge volume of data; this learning makes this procedure easier and faster.

Learning descriptions from raw datasets is one of the primary activities of deep learning neural networks. The neural network model has the ability to automatically discover the descriptions in data needed to detect features and classify them; this is called feature learning or description. A deep learning neural network may be specified with a class of machine learning methods that apply many layers of operational elements for

continuous learning and extract features from a raw dataset [6]. As the movement from the lower end to the higher end of the layers is done, the extracted features start resulting in more and more pronounced results in the learning model for inferring desired results for the given classification or forecasting job.

- **Long Short-Term Memory (LSTM)**

Long Short-Term Memory (LSTM), is a recurrent neural network (RNN), trained using Backpropagation Through Time that overcomes the vanishing gradient problem.

There are three types of gates within a unit:

Input Gate: conditionally decides which values from the input to update the memory state

$$i_t = \sigma(W_i [X_t, h_{t-1}] + b_i) \quad (1)$$

Forget Gate: conditionally decides what information to throw away from the block

$$f_t = \sigma(W_f [X_t, h_{t-1}] + b_f) \quad (2)$$

Output Gate: conditionally decides what to output based on input and the memory of the block

$$o_t = \sigma(W_o [X_t, h_{t-1}] + b_o) \quad (3)$$

where X_t is input vector at time t , h_{t-1} is hidden state at time $t-1$, i_t, f_t, o_t are input gate, forget gate, output gate vectors at time t , and W_i, W_f, W_o are weight matrices for the input, forget and output gates, respectively. b_i, b_f, b_o are bias vectors for the input, forget and output gates, respectively and σ is sigmoid function. Each unit is like a mini-state machine where the gates of the units have weights that are learned during the training procedure.

- **Simple Moving Average**

Simple Moving Average (SMA) is a fundamental and widely used time series in weather forecasting systems. It serves as a valuable tool for smoothing noisy time series data and identifying underlying trends or patterns. In weather forecasting, time series data represents a sequence of measurements or observations taken at regular time intervals, such as temperature readings recorded daily or hourly. The Simple Moving Average is a smoothing technique that calculates the average of a fixed number of consecutive data points, often referred to as a "window" or "period." The moving average window moves through the time series, one step at a time, to create a series of smoothed values. To begin using SMA, the initial step involves deciding on the window size, which dictates the number of data points to factor into each average calculation. The choice of window size depends on the desired level of smoothing and the characteristics of the data. For each time point in the time series, the Simple Moving Average is calculated by taking the arithmetic mean of the data points within the moving window. The formula for calculating the SMA for a specific time point t is:

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n} \quad (4)$$

where, A_n is the sum of data points, and n is number of total periods.

- **Weighted Moving Average**

The weighted moving average (WMA) is a technical tool used to guide decisions regarding trend directions [9]. WMA calculates the average of input values over specified time intervals, assigning greater importance to recent data points. It's employed to smooth time series data, effectively filtering out noise. In our case, we utilize a 10-year weather dataset spanning from 2013 to 2022. The formula for calculating the weighted moving average is as follows:

$$M = \frac{\sum_{t=1}^n W_t * V_t}{\sum_{t=1}^n W_t} \quad (5)$$

where, M is the average value, V is actual value, W is weighting factor, and n is number of periods in the weighting group.

3. Proposed System Architecture

In the proposed system, three types of weather prediction models; Long Short-Term Memory (LSTM) without noise reduction method, LSTM with Simple Moving Average Method (SMA) and LSTM with Weighted Moving Average (WMA) are trained. The testing of each model is evaluated. Initially, the dataset is partitioned into training (80%) and testing (20%) subsets for the system. Noise removal in time series weather forecasting systems is a crucial preprocessing step aimed at enhancing the accuracy and reliability of weather predictions. Noise in weather data can arise from various sources, including sensor inaccuracies, measurement errors, and environmental factors. The system design is described in Figure 1.

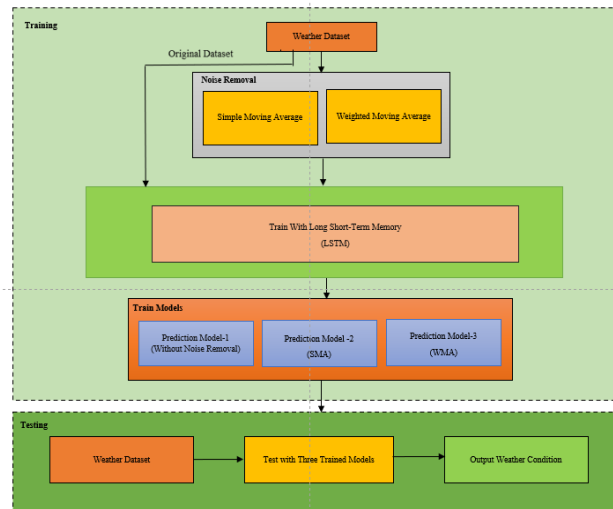


Figure 1: System Block Diagram

In the preprocessing step, the original dataset is preprocessed to eliminate noise with SMA and WMA. The proposed system trains the training dataset with three approaches: LSTM, LSTM with SMA and LSTM with WMA for the weather information of the years 2013 to 2022 in the region of Hmawbi, Yangon in Myanmar. This weather dataset consists of 6 attributes: date, maximum temperature, minimum temperature, humidity, wind speed, cloud amounts. Table 1. presents the original sample weather dataset.

Table 1: Sample Weather Dataset

Date	Maximum Temperature (°C)	Minimum Temperature (°C)	Relative Humidity (%)	Wind Speed (mph)	Cloud Status
1/1/2013	32.5	17.0	90	0.0	5
2/1/2013	32.0	17.0	90	0.0	4
3/1/2013	32.0	18.0	91	0.0	5
...	...	...	...	...	...
30/12/2022	32.0	13.0	100	1.2	6
31/12/2022	30.0	16.0	100	0.0	6

4. Performance Evaluation

The experiments will be evaluated based on the root mean square error (RMSE) and mean square error (MSE) in order to measure the accuracy of prediction.

- **Measurement Parameters**

In this system, historical data of weather condition in Hmawbi, Yangon, Myanmar is used as input data source. This system model is trained in Kera's which is based on TensorFlow. The system is implemented by the ratio of data size: (Training – 80%, Testing – 20%). The key metrics of performance measures (mean squared error and root mean squared error) are evaluated for this proposed system analysis. Mean Squared Error (MSE) is a widely used metric for evaluating the accuracy of a predictive model's forecasts, including time series predictions. It measures the average squared difference between predicted values and actual values over a set of data points. Mathematically, MSE is calculated as shown in formulas for this proposed system analysis.

$$MSE = \frac{1}{n} * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

where, n is the total number of data points, y represents the actual (observed) values at each time step, $\hat{y}$ represents the predicted values at each time step. Root Mean Squared Error (RMSE) is a variation of Mean Squared Error (MSE) and is particularly useful when you want the error metric to be in the same units as the original data. RMSE is the square root of the MSE and is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \|y_i - \hat{y}_i\|^2}{n}} \quad (7)$$

where n is the number of data points, $y(i)$ is the i^{th} measurement, and $\hat{y}(i)$ is its corresponding prediction.

5. Experimental Results

The performance comparison of three models for the next day prediction is evaluated according to the different weather attributes; maximum temperature, minimum temperature, humidity, wind speed, cloud amount based on daily weather data. As depicted in the Figure 2, the prediction result in maximum temperature is quite different from the actual weather data. Although noise removal method was used in SMA, the prediction results in all attributes are quite similar in LSTM. The prediction results of LSTM with WMA in all attributes are almost identical with the actual weather data whereas the prediction results of both methods (LSTM, LSTM with SMA) are quite different from the actual weather data. According to the results of Figure 2, the prediction of evaluation results was found that RMSE of LSTM, LSTM with SMA and LSTM with WMA are 5.398, 3.767 and 0.798, respectively while MSE of LSTM, LSTM with SMA and LSTM with WMA are 29.020, 14.841 and 0.534, respectively. According to evaluation results, WMA with LSTM model is considered acceptable for time series data analysis system.

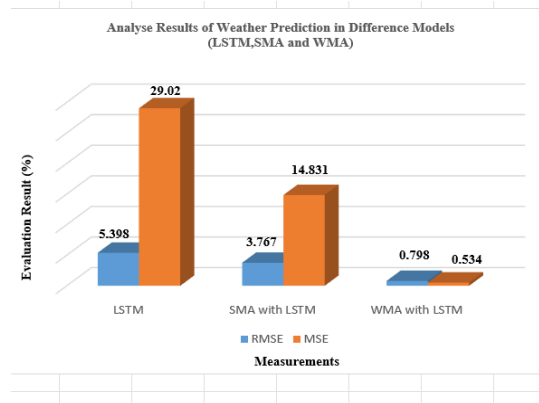


Figure 2: Analysis Results of Weather Prediction in Difference Models LSTM, SMA and WMA)

6. Conclusion

This study utilized LSTM, LSTM with SMA and LSTM with WMA models to predict daily weather attributes; maximum temperature, minimum temperature, humidity, wind speed, and cloud amount in Hmawbi, Yangon, Myanmar. The required weather data was sourced from the Department of Meteorology and Hydrology in Yangon, Myanmar, spanning from 2013 to 2022. The analysing accuracies of these models were compared

using RMSE and MSE metrics. The findings revealed that the LSTM with WMA model outperformed both LSTM and LSTM with SMA models in terms of both RMSE and MSE due to its superior forecasting accuracy and minimal error.

References

- [1] Guici Chen, Sijia Liu, and Feng Jiang, "Daily Weather Forecasting Based on Deep Learning Model: A Case Study of Shenzhen City, China", Special Issue Industrial Air Pollution Control in China, Atmosphere2022,13(8),1208.
- [2] Keerthana. P. J, Rajeswari. P, Amirtharaj. R, PandiyaRajan. G, "Weather Forecasting using Deep Learning Algorithm", INTERNATIONAL JOURNAL OF SPECIAL EDUCATION, Vol.37, No.3, 2022.
- [3] Y. Cheng, L. Thi, H. Nguyen, A. Ozaki, and T. V. Ta, "Deep learning-based method for weather forecasting: A case study in Itoshima".
- [4] N. Shobha and T. Asha, "Monitoring weather based meteorological data: Clustering approach for analysis", the International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017, pp.75-81.
- [5] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," in IEEE International Conference on Image Processing (ICIP), , 3349-3353, 2015.
- [6] V. Leijen, M. Boone and K. Horgan," Making numerical weather predictions portable compression of weather data for use in radar propagation modeling", USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), 2017, pp.12-19.
- [7] R. Dalmau, M. Perez-Batlle, and X. Prats, "Estimation and prediction of weather variables from surveillance data using spatio-temporal Kriging", IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), 2017, pp.1-8.
- [8] Y. Liu et al., "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," 2016," in Int'l Conf. on Advances in Big Data Analytics, 81-88, 2016.
- [9] <https://corporatefinanceinstitute.com/resources/equities/moving-average-wma>.
- [10] <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network>.

