Dhaka Univ. J. Sci. 73(2): 93-99, 2025 (July)
ISSN 1022 - 2502, E-ISSN 2408-8528
DOI: https://doi.org/10.3329.dujs.v73i2.82766

Weather Forecasting Using Recurrent Neural Networks and Vector Autoregressive Model: A Comprehensive Analysis of Time Series Data from Rohingya Camps and Control Areas

Arman Mahmud and Md. Israt Rayhan

Institute of Statistical Research and Training (ISRT), Dhaka University, Dhaka-1000, Bangladesh

(Received: 31 October 2024 ; Accepted: 13 March 2025)

Abstract

The study aimed to generate weather forecasts for the Rohingya camps in Ukhia and Teknaf by employing a Recurrent Neural Network (RNN) with a Long Short-Term Memory (LSTM) layer alongside a Vector Autoregressive (VAR) model. It utilized six meteorological variables-temperature, humidity, precipitation, surface pressure, wind speed, and wind direction-sourced from NASA's POWER project. The analysis focused on two sets of areas: control regions in Ukhia and Teknaf that are not impacted by the Rohingya presence, and target regions in the same locations where the Rohingya reside. The findings revealed notable differences in prediction accuracy between the models. The RNN with LSTM demonstrated superior accuracy in control areas, likely due to its ability to capture complex, nonlinear patterns. In contrast, the VAR model outperformed the RNN with LSTM in target areas, suggesting that weather patterns in these regions exhibit more linear relationships. These findings underscore the importance of selecting forecasting models based on the underlying structure of meteorological data to improve prediction accuracy.

Keywords: VAR, RNN, LSTM, Rohingya, Ukhia, Teknaf, Forecasting.

I. Introduction

Since the late 1970s, Bangladesh and Myanmar have had strained relations due to the Rohingya refugee crisis, a highly contentious issue. This crisis stems from severe human rights violations perpetrated by the military junta in Myanmar against the Rohingya Muslims in Rakhine, rendering them stateless[1]. The situation worsened on August 25, 2017, when hundreds of thousands of Rohingya fled to Bangladesh, coinciding with the release of the Advisory Commission report led by former UN Secretary-General Kofi Annan. This ongoing crisis poses a significant challenge for the international community, with Bangladesh experiencing immense pressure to accommodate such a large influx of refugees. Given the current political circumstances, long-term solutions such as voluntary repatriation appear elusive. Despite ongoing diplomatic efforts by Bangladesh, bolstered by support from the international community through the UN and various human rights evaluations and accountability initiatives, progress toward establishing conditions for a safe, dignified, and sustainable return remains minimal[2].

The Rohingya are widely recognized as the largest stateless population in the world. In August 2017, over 700,000 Rohingya are compelled to flee from Myanmar to Bangladesh, marking one of the fastest-growing refugee crises globally. Despite being neighboring countries in Southeast Asia, Bangladesh and Myanmar maintain a strained relationship, sharing a border of approximately 271 kilometers.

Cox's Bazar in Bangladesh, now the largest refugee camp in the world, has become a temporary haven for these displaced individuals, who sought safety in the camps. The sudden influx of people into Cox's Bazar, a region known for its rich biodiversity, has posed significant challenges to maintaining the local ecosystem, particularly in areas like Teknaf, which hosts a wildlife sanctuary covering over 11,615 hectares[3,4].

This article aims to outline the major environmental impacts resulting from the arrival of Rohingya refugees. Key issues include deforestation, severe water scarcity and pollution, habitat loss, fragmentation and destruction of wildlife habitats, inadequate waste management, poor drainage systems, air pollution, surface water contamination, and other environmental challenges[5]. These problems are diverse and complex, highlighting the necessity of multinational efforts to provide adequate protection and support. Contributing factors to the complexity of the Rohingya refugee crisis include the rise of Islamist insurgency, the illegal drug trade, particularly methamphetamine, and the prevalence of HIV/AIDS and other sexually transmitted infections[6].

A comparative study is conducted between VAR and LSTM models concerning weather components, alongside an examination of four different stochastic weather time series generators: first- and second-order Markov Chains (MC), vector autoregressive (VAR) models, and long short-term memory (LSTM) neural networks. These models are trained on a dataset spanning 40 years with an hourly resolution. After training, 25 years of simulated time series data are generated and evaluated using various time series metrics. The findings demonstrated that the second-order MC and VAR models excelled in replicating the original time series patterns[7].

Additionally, a comparative analysis of statistical learning models-including VAR, ARIMAX, and the deep learning model LSTM-is conducted within the context of multivariate short-term (24 hours) time series forecasting. This analysis utilized traffic volume, speed, and average waiting time while

*Author for correspondence. e-mail: israt@isrt.ac.bd

incorporating weather variables in Austin, Texas. Models are evaluated using the rolling forecast origin technique across three primary feature sets derived from feature selection, with the VAR model showing superior performance[8].

Furthermore, a comparison is made between the outcomes of two models designed for weather forecasting: one employing a VAR model and the other utilizing a Recurrent Neural Network with a LSTM layer. Six meteorological variables-temperature, humidity, precipitation, surface pressure, wind speed, and wind direction-served as inputs for the network.

II. Methodology

Data and variables

The data is obtained from NASA Prediction of Worldwide Energy Resources (POWER) project[9]. For the purpose of this investigation, meteorological data are utilized from two distinct geographical locations: the Rohingya camps in Ukhia and Teknaf, as well as two independent unaffected sites in Ukhia and Teknaf that remained undisturbed by their presence. Daily samples are taken from May 1st, 2014, through May 1st, 2024, to compile the gathered data. The data are obtained for four specific locations: the Ukhia Rohingya camp (21°12'18.0"N 92°09'06.8"E), the Teknaf Rohingya camp (20°57'22.3"N 92°15'06.2"E), and two unaffected locations independent of the Rohingya camps in Ukhia (21°03'03.5"N 92°11'22.8"E) and Teknaf (21°06'23.7"N 92°09'00.9"E). Key meteorological parameters considered included temperature, humidity, precipitation, surface pressure, wind speed, and wind direction.

The six meteorological variables: temperature, humidity, precipitation, surface pressure, wind speed, and wind direction, were chosen because they play a key role in weather. These variables are commonly used in meteorological research since they directly influence weather patterns and help capture both short-term changes and long-term climate trends. Previous studies[13] have shown that these factors are effective for weather prediction.

The selection of RNN with LSTM layer and Vector Autoregressive (VAR) models was motivated by their demonstrated effectiveness in time series forecasting. RNN with LSTM models are particularly advantageous for capturing long-term dependencies and nonlinear patterns in sequential data, making them highly suitable for complex weather prediction tasks. In contrast, VAR models are effective in modeling linear relationships among multiple time series variables, which is critical for understanding the interdependencies between various meteorological factors. These models have been shown in previous research to outperform traditional statistical methods in weather forecasting.

Selection of Target and Control Areas

The target areas refer to locations where the Rohingya camps are situated, specifically the Ukhia and Teknaf regions. The control areas in these regions were chosen based on their close geographical proximity to the Rohingya camps and similar climatic conditions. These criteria were applied to ensure that the comparison between the target and control areas would not be influenced by substantial environmental differences. Additionally, the control areas were selected to be unaffected by the presence of the Rohingya population, facilitating a clear comparison of weather patterns between the impacted and non-impacted regions.

Model Training Approach:

Separate models were trained for each region (target and control areas) to account for potential differences in weather patterns. This approach ensures that the models are tailored to the specific characteristics of each region, improving their predictive accuracy.

Recurrent Neural Networks

The Recurrent Neural Network (RNN) is a category of artificial neural network intended for time series forecasting. One notable subclass of RNNs is the Elman network, which features one or more hidden layers. The first hidden layer's weights are derived from the input layer, and each subsequent layer's weights are obtained from the previous layer. Typically, the hidden layers utilize a sigmoid bipolar activation function, while the output layer employs a linear activation function. Elman networks can handle both continuous and discontinuous activation functions. A distinctive feature of this network is the presence of a delay in the connection between the input layer and the first hidden layer at the previous time step (t-1). This delay introduces a feedback mechanism that can capture and accommodate noise from previous inputs into the next input. For a more mathematical understanding, let x(t) and y(t) represent the input and output time series, respectively. The three connection weight matrices are W_{II}, W_{HH}, and W_{IH}. The activation functions for the hidden and output units are f_I and f_H, respectively. The behavior of the recurrent network can be described by the following non-linear matrix equations:

$$h(t+1) = f_I\,(W_H\,x(t) + W_{HH}\,h(t))$$

$$y(t+1) = f_H(W_{IH}h(t+1))$$

Here, h(t) represents the state of a dynamical system, summarizing all necessary past information to predict the future behavior of the system[10].

Recurrent Neural Networks with Long Short Term Memory

The fundamental distinction between the designs of RNNs and LSTM networks lies in the fact that the hidden layer of an LSTM is a gated unit or gated cell. The cell is composed of four layers that interact with each other to generate both the output and the cell state.

$$i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i)$$
$$f_t = \sigma(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f)$$
$$o_t = \sigma(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o)$$
$$\tilde{c_t} = tanh(W_{\tilde{c},x}x_t + W_{\tilde{c},h}h_{t-1} + b_{\tilde{c}})$$

The weight matrices are $W_{i,x}, W_{i,h}, W_{f,x}, W_{f,h}, W_{o,x}$, $W_{o,h}, W_{\tilde{c},x}, W_{\tilde{c},h}$. The bias vectors are b_i, b_f, b_o *and* $b_{\tilde{c}}$.

the current input is x_t; the output of the LSTM at time t −1 is h_{t-1}, and the sigmoid activation function is σ. How much of the previous memory value to be eliminated from the cell state is decided by the forget gate. In the same way, new input to the cell state is specified by the input gate. The cell state c_t is then computed as follows:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

where ⊙ denotes the Hadamard product. The output h_t of the LSTM at time t is derived as:

$$h_t = o_t \odot \tanh(c_t)$$

Finally, The output h_t is projected to the predicted output $\tilde{y}_t$ as follows.:

$$\tilde{y_t} = W_y h_t$$

Where W_y is a projection matrix to reduce the dimension of h_t[11].

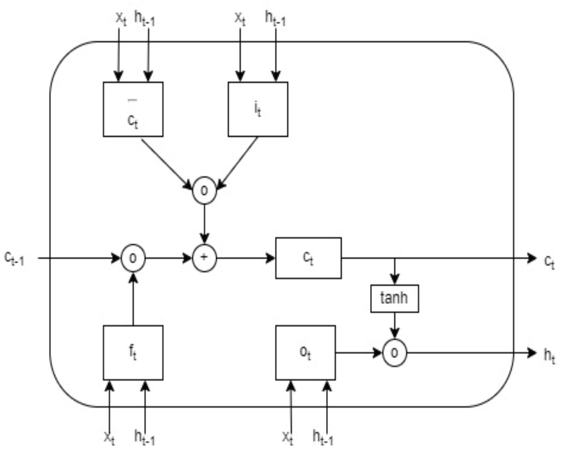

Fig. 1. RNN with LSTM layer.

Vector Autoregression Model

The VAR model is a multivariate time series approach used to forecast multiple variables simultaneously. It is applicable in situations where the variables are interdependent. In VAR modeling, each variable is represented as a linear combination of its own past values and the past values of the other variables. This creates a system of equations, with each variable assigned its own equation, which can be expressed in vector form. Let Y_t represent a vector of time series data, then a VAR model with k variables and p lags can be formulated as shown in Equation. Here, Y_t, β_0 and ϵ_t are column vectors of size k×1, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are k×k coefficient matrices.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \epsilon_t$$

If the time series is non-stationary, differencing is required before training the model, and the forecasted

values should be transformed back by reversing the differencing to obtain the actual predictions[12].

III. Analysis and Results

In this study, two models are employed: RNN with LSTM layers and the VAR model. Ten years of weather data, collected from the Rohingya camps in Ukhia and Teknaf as well as from a control location independent of these camps, are utilized. The primary objective is to compare the weather conditions between the Rohingya camps and the control area by forecasting six key weather components: temperature, humidity, precipitation, surface pressure, wind speed, and wind direction. The RNN with LSTM layers is used to capture both short-term and long-term temporal dependencies, while the VAR model is applied to understand the linear relationships among these weather components over time. The Augmented Dickey-Fuller (ADF) test is applied to each dataset, and all datasets are found to be stationary. Additionally, the process of data normalization is performed to ensure consistency in the values of the data. The tools and technologies used include Python, TensorFlow, NumPy, Pandas, Matplotlib, and Scikit-learn. The performance of the model is evaluated by computing the Root Mean Squared Error (RMSE), which is derived from the MSE by taking the square root to provide a more interpretable error metric.

Analysis and Results in Ukhia

Table 1. Comparison of RMSE and MAE between LSTM and VAR models for Target and Control Areas in Ukhia.

	RNN with LSTM		VAR	
	RMSE	MAE	RMSE	MAE
Target Area	103.19	65.08	30.91	13.81
Control Area	22.02	8.39	26.87	11.91

The table demonstrates that the VAR model exhibits higher performance in the target area, whereas the LSTM model outperforms in the control area, as evidenced by reduced RMSE and MAE values. The disparity in performance may be ascribed to the distinct properties of the time series data in each region. The weather patterns in the target area display more linear correlations, rendering VAR more appropriate, whereas the control area may demonstrate more intricate, nonlinear patterns, which are more effectively captured by LSTM's deep learning framework. Overall, VAR shows better results compared to LSTM.

Target vs Control Area Data and Future Forecast using VAR in Ukhia

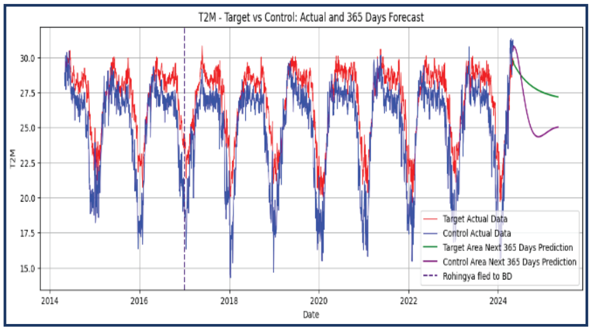

Fig. 2. *Target vs Control Area Temperature and Future Forecast*

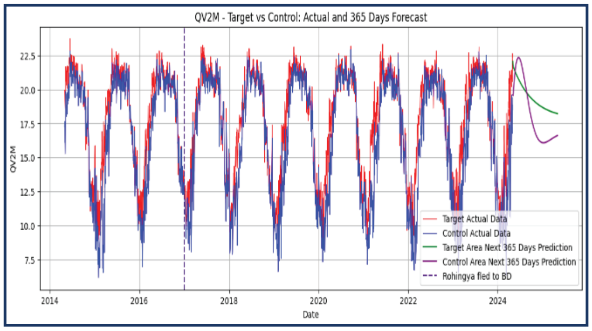

Fig. 3. *Target vs Control Area Humidity and Future Forecast*

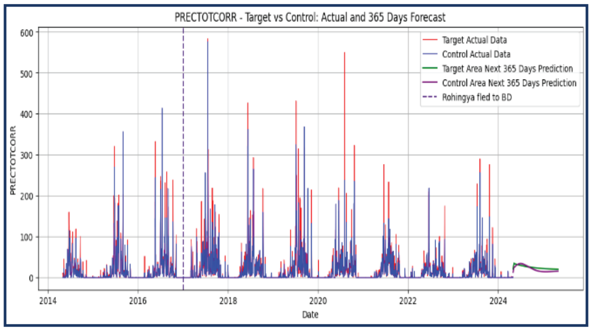

Fig. 4. *Target vs Control Area Precipitation and Future Forecast*

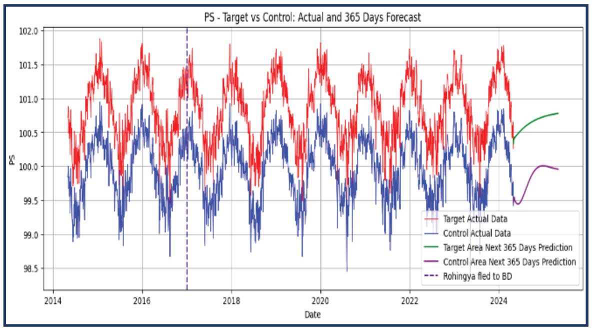

Fig. 5. *Target vs Control Area Surface Pressure and Future Forecast*

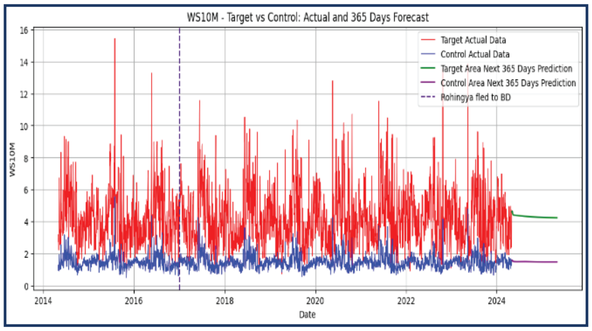

Fig. 6. *Target vs control area wind speed and future forecast.*

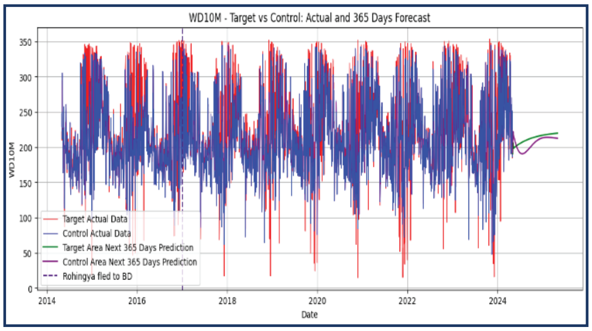

Fig. 7. *Target vs control area wind direction and future forecast.*

Analysis and Results in Teknaf

Table 2. Comparison of RMSE and MAE between LSTM and VAR models for Target and Control Areas inTeknaf.

	RNN with LSTM		VAR	
	RMSE	MAE	RMSE	MAE
Target Area	96.85	61.99	26.88	11.95
Control Area	25.17	9.57	30.91	13.76

In Teknaf, as in Ukhia, the VAR model shows better performance in the target area, while the LSTM model excels in the control area, reflected by lower RMSE and MAE values. The target area's weather patterns display more linear correlations, making the VAR model a more suitable choice. Conversely, the control area features more complex, nonlinear patterns, which are better captured by LSTM's deep learning framework, similar to the case in Ukhia. Overall, the VAR model delivers superior results compared to the LSTM in Teknaf as well

Target vs Control Area Data and Future Forecast using VAR in Teknaf

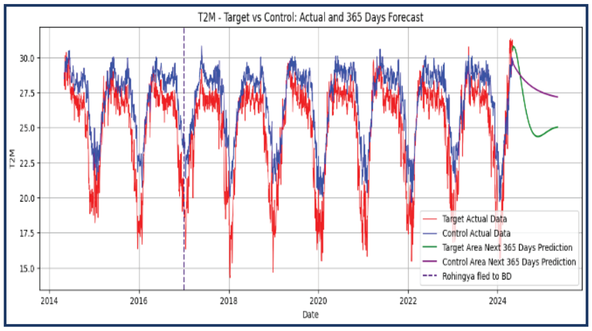

Fig. 8. *Target vs Control Area Temperature and Future Forecast*

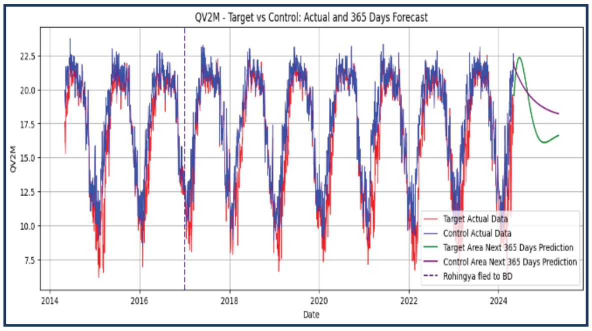

Fig. 9. *Target vs Control Area Humidity and Future Forecast*

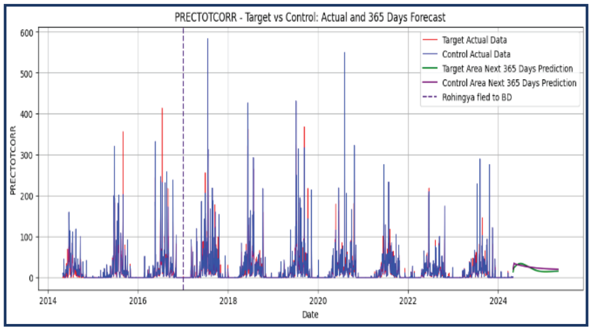

Fig. 10. *Target vs Control Area Precipitation and Future Forecast*

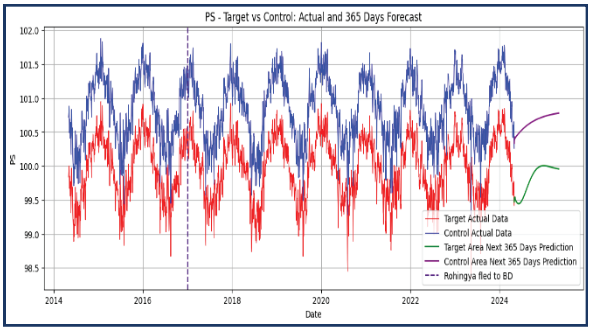

Fig. 11. *Target vs Control Area Surface Pressure and Future Forecast*

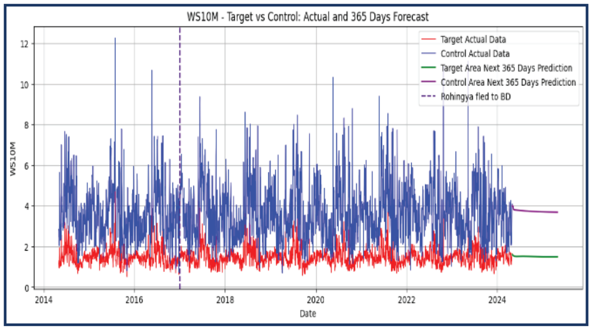

Fig. 12. *Target vs control area wind speed and future forecast.*

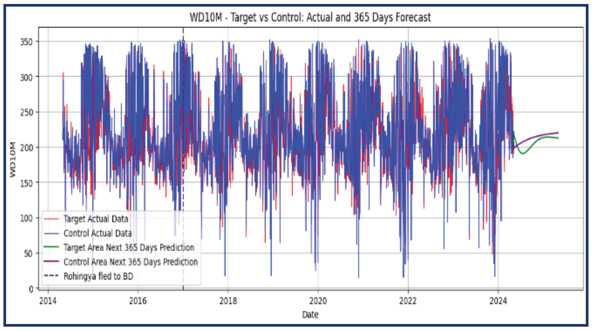

Fig. 13. *Target vs control area wind eirection and future forecast.*

IV. Discussion and Implications

This study aims to generate weather predictions using a RNN with a LSTM layer and a VAR model. Six meteorological variables temperature, humidity, precipitation, surface pressure, wind speed, and wind direction serve as inputs for these models. The weather data analyzed comes from two distinct locations: one in Ukhia and one in Teknaf, where the Rohingya population resides, along with control areas in both municipalities that are not affected by their presence. The data is obtained from NASA's POWER project, with separate datasets evaluated for Ukhia and Teknaf.

Predictions are made for all six meteorological variables in both regions, and the results are visualized. The RNN with LSTM and VAR models are applied to both target and control areas within Ukhia and Teknaf. Following the training and testing phases, the results are presented in the accompanying graph. The RNN with LSTM model is trained using normalized data through MinMaxScaler, which scales input features between 0 and 1 to enhance training efficiency. The LSTM architecture consists of two LSTM layers followed by Dense layers to handle the six weather outputs. A regression-standard Adam optimizer and mean squared error loss function are utilized for model training. Predictions are tailored to the characteristics of each area, with the model trained separately on target and control datasets. Model performance in each region is evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), providing a 365-day weather forecast. Trends are iteratively predicted using the last 60 days of available data, and line charts are created to compare actual, expected, and future data for each meteorological variable. It is possible that the influx of Rohingya refugees into Bangladesh in 2017 has influenced the weather patterns in the target area.

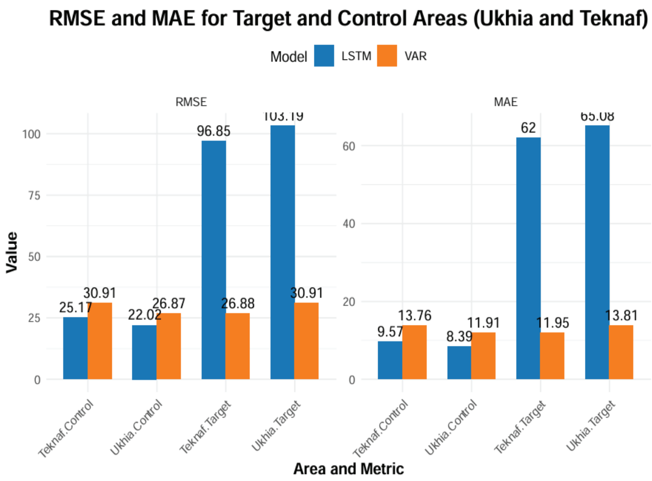

Fig. 14. Comparison *of LSTM and VAR.*

The VAR model is applied separately to both the target and control datasets. The Akaike Information Criterion (AIC) is utilized to determine the optimal lag order for each model, informing the inclusion of appropriate time delays. Utilizing historical data, weather component predictions for the next year and a half are generated for both the target and control areas. To assess the models' accuracy, RMSE and MAE are used as evaluation metrics, providing insights into forecasting performance for both datasets. Actual data, in-sample forecasts, and 365-day forecasts are visualized for each weather component.

The VAR and RNN with LSTM models are chosen due to their enhanced predictive capabilities compared to other models. The VAR model effectively captures linear relationships among various weather variables, making it suitable for short-term forecasting. The RNN with an LSTM layer is adept at modeling sequential data by retaining historical information through recurrent connections, which is crucial for time series forecasting. The LSTM component addresses long-term dependencies and patterns that differ from linear trends, while the RNN structure improves the model's ability to learn from sequential data. Additionally, the LSTM layer mitigates issues related to vanishing gradients, allowing it to effectively identify long-term weather trends. This combination of models offers a well-rounded approach that integrates both linear and nonlinear relationships for comprehensive short- and long-term weather predictions.

The findings suggest that the VAR model outperformed the Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM) in the target areas, indicating that linear dependencies among weather variables are more dominant in regions influenced by the presence of the Rohingya population. These results align with previous research by Dissanayake et al. (2021)[8], which highlights the effectiveness of VAR models in capturing linear relationships in meteorological data. Conversely, the RNN-LSTM model demonstrated superior performance in the control areas, likely due to its ability to model complex, nonlinear dependencies.

These findings have significant implications for weather forecasting in humanitarian contexts, where precise predictions are essential for disaster preparedness and resource allocation. The stronger performance of the VAR model in target areas suggests that weather patterns in these regions exhibit predominantly linear characteristics. Given that linear models like VAR are particularly effective when meteorological variables such as temperature and pressure demonstrate strong linear correlations, this reinforces the suitability of VAR for forecasting in the Rohingya-affected regions. In contrast, the LSTM model, which is adept at capturing nonlinear structures, may be less effective in environments where linear relationships prevail. These insights highlight the importance of tailoring model selection to the underlying dynamics of the local environment. Future studies could explore hybrid approaches that integrate the advantages of both VAR and RNN with LSTM models to enhance forecasting accuracy across diverse environmental settings, ensuring adaptability to both linear and nonlinear patterns. Such advancements would further empower humanitarian agencies to mitigate risks and optimize resource deployment in climate-vulnerable regions.

V. Conclusion

This study demonstrates that the VAR model is more effective for weather forecasting in Rohingya camps, where linear weather patterns dominate, while LSTM performs better in control areas with more complex, nonlinear trends. These findings have important implications for disaster preparedness and resource allocation in refugee camps, where accurate weather forecasts can help mitigate the impact of extreme weather events. Future research could explore the integration of additional data sources, such as satellite imagery or land-use data, to further improve forecasting accuracy.

VI. Limitations and Future Research

This study has certain limitations. First, it relies solely on NASA's POWER dataset, which may not include all meteorological variables that influence weather patterns. Expanding the dataset to incorporate additional factors, such as solar radiation and cloud cover, could improve forecasting accuracy. Additionally, the analysis is geographically confined to the Rohingya camps in Ukhia and Teknaf, making it uncertain whether the findings are generalizable to regions with different climatic conditions. Future research could apply these models to diverse geographical areas to evaluate their broader applicability.

Another key limitation is the assumption of stationarity in the time series data, which may not hold true for all meteorological variables. The study also operates within the constraints of the dataset's spatial resolution, potentially limiting the generalizability of the results. Furthermore, while the study compares the performance of VAR and RNN-

LSTM models, exploring hybrid approaches that integrate both could yield more accurate forecasts by capturing both linear and nonlinear relationships in weather data. Such advancements would be particularly valuable in complex humanitarian settings. Finally, model training presented challenges, notably overfitting in the LSTM model, which was mitigated using regularization techniques and early stopping.

References

1. Parnini, S. N., M. R. Othman, and A. S. Ghazali, 2013. The rohingya refugee crisis and Bangladesh Myanmar relations. *Asian and Pacific Migration Journal* **22(1)**, 133–146.
2. Gorlick, B., (2019). The rohingya refugee crisis: Rethinking solutions and accountability. *Refugee Studies Centre, Working Paper Series* (131).
3. Hossain, S. and S. Hosain, 2019. Rohingya identity crisis: A case study. *Saudi Journal of Humanities and Social Sciences.*
4. Tay, A. K., A. Riley, R. Islam, C. Welton-Mitchell, B. Duchesne, V. Waters, A. Varner, B. Moussa, A. M. Alam, M. Elshazly, et al., 2019. The culture, mental health and psychosocial wellbeing of rohingya refugees: A systematic review. *Epidemiology and Psychiatric Sciences* **28(5)**, 489–494.
5. Bashar, S., 2021. The rohingya refugee crisis in Bangladesh: Environmental impacts, policies, and practices.
6. Taufiq, H. A., 2021. Rohingya refugee crisis and the state of insecurity in Bangladesh. arXiv preprint arXiv:2107.12080.
7. Eberle, S., D. Cevasco, M.-A. Schwarzkopf, M. Hollm, and R. Seifried, 2022. Multivariate simulation of offshore weather time series: a comparison between markov chain, autoregressive, and long short-term memory models. *Wind* **2(2)**, 394–414.
8. Dissanayake, B., O. Hemachandra, N. Lakshitha, D. Haputhanthri, and A. Wijayasiri, 2021. A comparison of arimax, var and lstm on multivariate short-term traffic volume forecasting. I*n Conference of Open Innovations Association*, FRUCT, **28**, 564–570. FRUCT Oy.
9. https://power.larc.nasa.gov/data-access-viewer/
10. Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986. Learning internal representations by error prop agation, parallel distributed processing, explorations in the microstructure of cognition, ed. de rumelhart and j. mcclelland. vol. 1. 1986. *Biometrika* 71(599-607), 6.
11. Qing, X. and Y. Niu, 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy **148**, 461–468.
12. Dissanayake, B., O. Hemachandra, N. Lakshitha, D. Haputhanthri, and A. Wijayasiri, 2021. A comparison of arimax, var and LSTM on multivariate short-term traffic volume forecasting. In Conference of Open Innovations Association, FRUCT, **28**, 564–570. FRUCT Oy.
13. Ravuri, S., Lam, R., Mirowski, P., Athanasiadis, T., Kashem, S., Prudden, R., Mandhane, A., Ayzel, G., Leon, C., & Ortner, R., 2021. A deep convolutional neural network model for improving WRF forecasts. arXiv preprint arXiv:2008.06489.