Paper ID: CE 0409

Machine Learning Approach for Accurate Forecasting of Monthly Air Temperature: A Case Study in Chapainawabganj, Bangladesh

R. Sarker[1], H. M. Rasel[2], S. A. Saki[3]

[1]Department of Civil Engineering, RUET, Bangladesh (mohammedranasarkar297@gmail.com)
[2]Department of Civil Engineering, RUET, Bangladesh (hmrruetce02@gmail.com)
[3]Department of Civil Engineering, RUET, Bangladesh (ahmad.saki007@gmail.com)

Abstract

Air temperature is a critical environmental factor that has an impact on different sectors, such as energy, agriculture and public health. With regards to climate change and global warming, precise monthly temperature prediction is gaining more significance for making informed decisions in these domains. Nevertheless, accurately predicting air temperature remains a persistent challenge due to its vulnerability to various climate-related factors. The proposed method in this thesis for forecasting monthly air temperature in Chapainawabganj, Bangladesh involves the use of a machine learning approach. The objective of this research is to use two models, Autoregressive integrated Moving Average (ARIMA) and Recurrent Neural Network (RNN), to predict the accuracy of monthly air temperature. This research considers various factors related to climate change and global warming, such as precipitation, wind speed, palmer drought severity index, and relative humidity, to predict air temperature. For model training, we use data ranging from 1981 to 2011, while data from 2012 to 2021 is used for testing purposes. The study evaluates the performance of the RNN model and ARIMA model in predicting monthly air temperature levels in Chapainawabganj, Bangladesh. Several metrics are utilized to compare the performance of two models, including coefficient of determination (R^2), Willmott's index (d), RMSE, Pearson's correlation (R), and MAE. The findings suggest that the RNN model, which is a non-linear approach, is more effective in predicting air temperature compared to the statistical ARIMA model. Consequently, the RNN model is selected as the preferred method for air temperature prediction. Our study contributes to the growing body of literature on temperature prediction using machine learning methods, particularly in the context of investigations on climate change and global warming. Our proposed model could be a valuable tool for temperature forecasting in various regions, potentially helping to mitigate the impacts of climate change on different sectors.

Keywords: *Temperature, Climate Change, Global Warming, ARIMA, RNN, Forecasting.*

1 Introduction

Precisely predicting the atmospheric temperature at a specific place and time is a significant area of study that holds great importance across various fields, including hydrology, agriculture, irrigation, and the environment (Zhou et al. 2023). Furthermore, it serves as a crucial meteorological factor that can effectively depict phenomena such as global warming and climate change (Alomar et al., 2022). The escalation of temperatures associated with climate warming results in various adverse consequences, including the melting of glaciers, elevation of sea levels, heightened likelihood of severe weather, extinction of species, disruption of food chain, and the occurrence of natural disasters like landslides, mudslides, tsunamis, and typhoons (Bonjakovi 2012). The measurement of air temperature plays a significant role in assessing the greenhouse effect, hydrological balance, and energy balance, estimation of radiation of solar and air pollution levels (Immerzeel et al. 2010; Li et al. 2013). Multiple factors influence the fluctuations in air temperatures, including geographical dispersion, air movement, ocean currents, light, and wind speed, presence of water bodies, vegetation cover and geomorphological features (Byeongseong et al. 2021). As a result, temperature change exhibits nonlinear, unpredictable, and dynamic characteristics. Air temperature plays a crucial role in governing various processes on the Earth's surface, including photosynthesis, respiration, and evaporation (Zhou et al. 2023). Precise forecasting of air temperature is essential for effective planning of agricultural operations, recreational activities, tourism, transportation, energy generation, and implementing measures to cope with temperature fluctuations (Yakut & Süzülmüş, 2020). Experiencing unhealthy air temperatures can lead to severe health issues (Lan et al., 2010; Schulte et al., 2016).

In this research, a dynamic model is constructed by utilizing both ARIMA (Autoregressive Integrated Moving Average) and RNN (Recurrent Neural Network) models. To enhance the accuracy of air temperature forecasting, the model incorporates additional covariates such as relative humidity, precipitation, wind speed, and the palmer drought severity index. The effectiveness of the models is evaluated by comparing their performance. To our knowledge, no previous studies have been conducted to compare the performance of ARIMA and RNN models that incorporate the covariates of relative humidity, precipitation, wind speed, and the Palmer Drought Severity Index. Hence, the objective of this study is to compare the ARIMA and RNN models to determine which one provides more accurate predictions of air temperature in Chapainawabganj, Bangladesh.

2 Materials and Methods

2.1 Location and Sample collection

Chapainawabganj is a district located in the northwest region of Bangladesh, specifically in the Rajshahi Division. The Rajshahi Division is responsible for its administrative oversight. The Indian state of West Bengal shares a boundary with the district. The two nations in this area are naturally separated by the Mahananda River. The cultivation of crops including rice, mango,wheat, jute, and sugarcane is just one of Chapainawabganj's many agricultural pursuits. With numerous archeological sites and notable landmarks, the neighborhood has a rich historical and cultural legacy that draws tourists. Similar to other regions in Bangladesh, the air temperature in Chapainawabganj can vary based on the season. Data regarding air temperature, relative humidity, precipitation, wind speed, and the palmer drought severity index have been collected for a period of 41 years, spanning from 1981 to 2021. The data is obtained from satellite-based information accessed through websites (https://power.Larc.Nasa.Gov/ and https://app.climateengine.org/climateEngine).

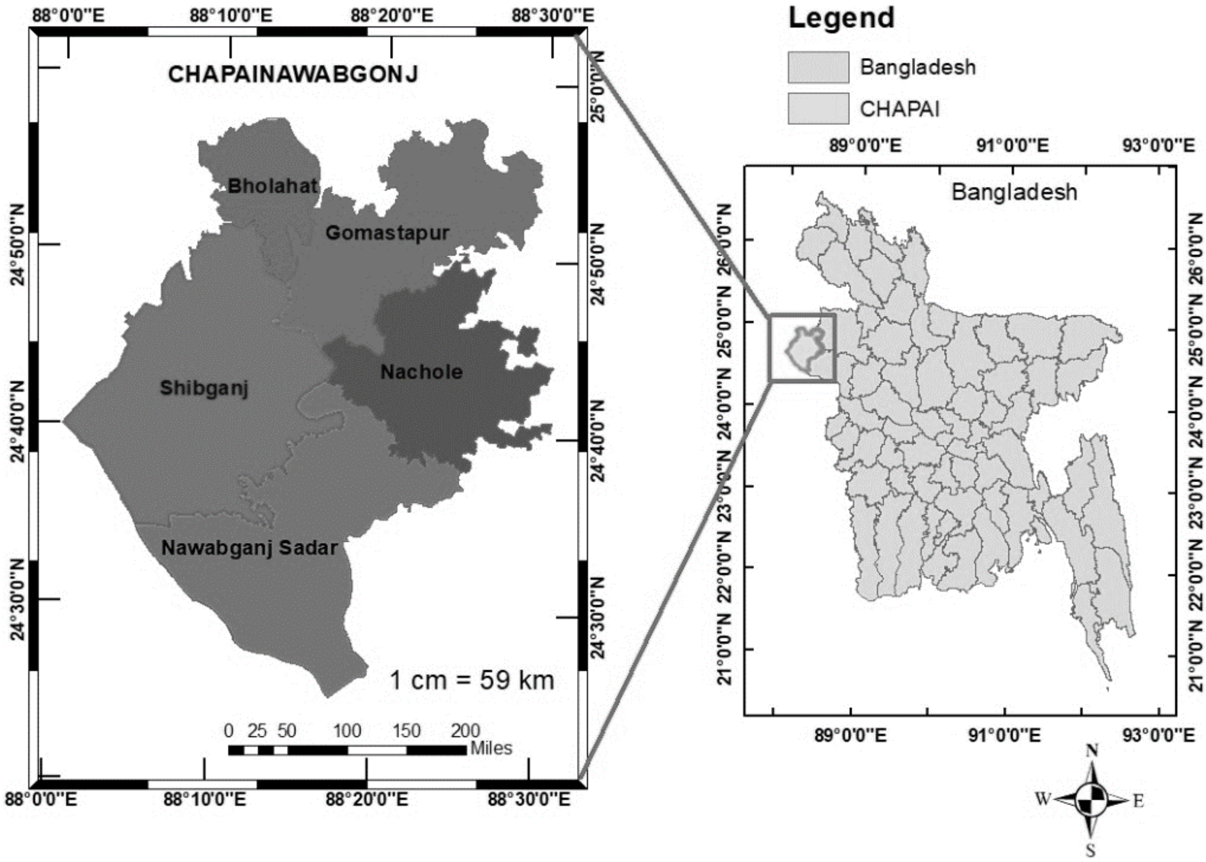

Figure 1. GIS based representation of Chapainawabganj, Bangladesh

2.2 Flowchart of Working Procedure

Figure 2 illustrates an overview of the research methodology utilized in the study.

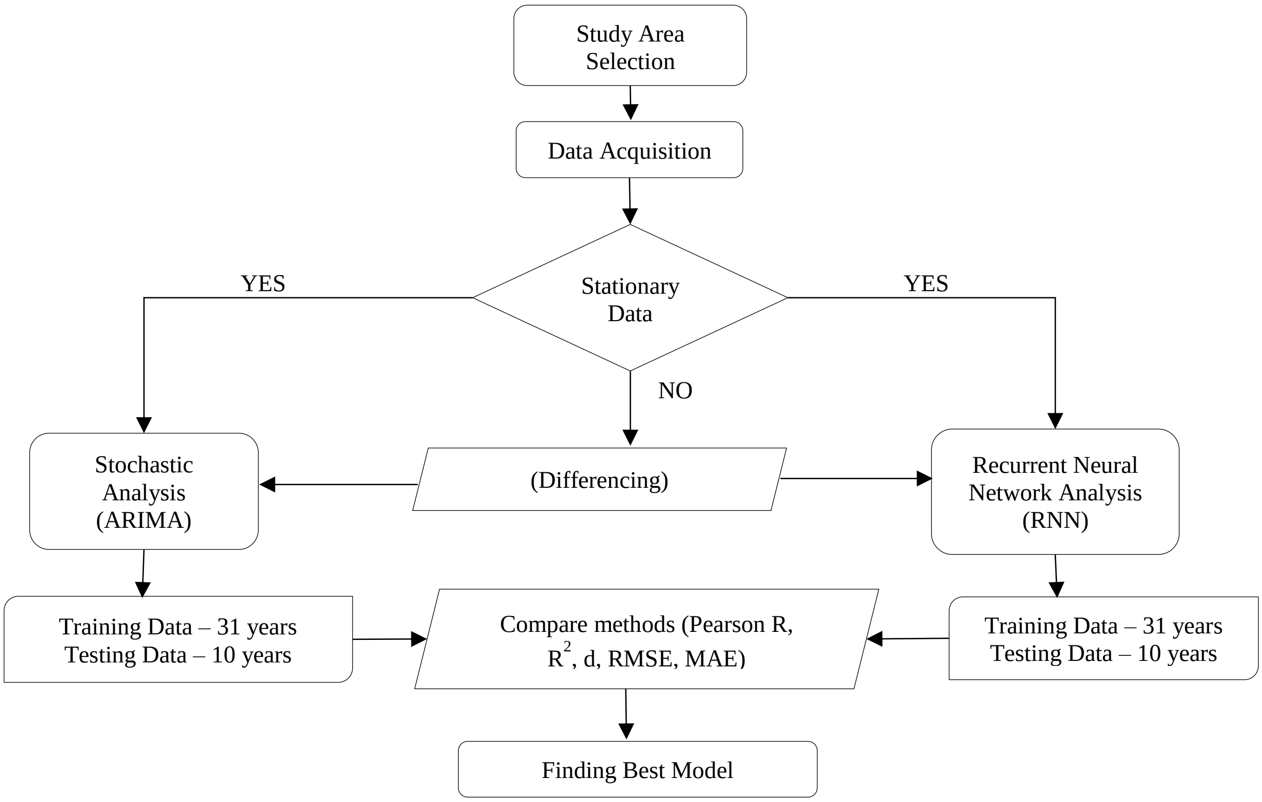

Figure 2. Flowchart of Working procedure

2.3 Methods

2.3.1 ARIMA Model

The study employed the Box-Jenkins method of ARIMA modeling, which involves an iterative process initially proposed by Box and Jenkins. This iterative process helped determine the best models through trial and error, with the aid of statistical software packages. The ARIMA model comprises three components: autoregressive (AR), integrated (I), and moving-average (MA). The AR component captures the autocorrelation between current and past observations, while the MA component describes the autocorrelation structure of the residuals. The integrated part of the ARIMA model indicates the level of differencing necessary to convert a non-stationary series into a stationary one. The non-seasonal ARIMA model is commonly represented as (p, d, q), where p represents the AR component, d signifies the required level of differencing, and q represents the MA component. In the case of the non-seasonal components of a seasonal ARIMA model, the MA operator can be expressed as follows:

$$\theta\,(B) = 1 - \theta 1B - \theta 2B^2 - \cdots \ldots . - \theta qB^q \tag{1}$$

In the equation, q represents the order of the non-seasonal Moving Average (MA) operator. The MA parameters are denoted as θ_1, θ_2, ..., θ_q. The backward shift operator, represented by B, is defined in a way that it shifts the time index of the series.

$$BZ_t = Z_{t-1} \tag{2}$$

The AR operator in the ARIMA model can be expressed as follows:

$$\emptyset(B) = 1 - \emptyset 1B - \emptyset 2B^2 - \cdots \ldots . - \emptyset pB^p \tag{3}$$

In this particular situation, p signifies the non-seasonal Autoregressive (AR) operator's order. The non-seasonal AR parameters are represented by $\emptyset_1, \emptyset_2, ..., \emptyset_p$. The non-seasonal ARIMA (Autoregressive Integrated Moving Average) model for a collection of equidistant measurements Z = [$Z_1, Z_2, ..., Z_n$] can be formulated in the following manner:

$$\emptyset(B)(1-B)^d Z_t = \emptyset(B)a_t \tag{4}$$

In the generalized form of the non-seasonal (p, q, d) ARIMA model, d represents the number of differences applied to the time series data to achieve stationarity. The discrete time index is denoted by t, and the white noise or error term in the model is represented by at. The non-seasonal ARIMA model can be expressed as follows:

$$U_t = \emptyset_1 U_{t-1} + \emptyset_2 U_{t-2} + \cdots .. + \emptyset_p U_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots ... - \theta_q\varepsilon_{t-q} \tag{5}$$

$$U_t = X_t - X_{t-d} \tag{6}$$

In the equation, $\emptyset_p$ represents the autoregressive parameter, εt represents the residual or white noise, θ represents the moving average parameter, X represents the dependent variable, and U_t denotes the d^{th} difference of the dependent variable.

2.3.2 Recurrent Neural Network (RNN) model

Deep (Machine) learning has recently gained significant traction, particularly in tasks involving sequential prediction, such as statistical language modeling, chaotic time series analysis, Finance, marketing, and ecological modeling for dynamic systems control. This success has inspired researchers to explore the application of deep learning techniques in hydrology events for time series forecasting. Recurrent Neural Networks (RNNs) lie at the core of these efforts, aiming to process input data in long sequences by iteratively performing the same task for each element while incorporating information from previous computations. RNNs possess a memory cell that retains information as the training data sequence progresses. The architecture of RNNs can vary depending on the specific application, such as the many-to-one model (used for predicting the current time step based on previous inputs) or the many-to-many model (used for predicting multiple future time steps based on previous inputs), among other variations. The choice of architecture depends on the particular problem and the underlying phenomena being addressed. In our study, we employ a many-to-one model for one-step ahead forecasting, specifically predicting the low-flow value of the current month based on the input of previous month's low-flow values. RNNs are connectionist models designed to capture temporal dependencies between inputs and outputs using internal memory. Unlike other neural network architectures, RNNs incorporate loops that enable them to retain information from past time steps and integrate it into the current computation. The memory of an RNN is encoded in its recurrent connections, enabling the persistence of information and its influence on the network's behavior throughout the sequential time series. Each neuron in an RNN generates its output by considering the accumulated information from previous time steps.

3 Results and Discussion

3.1 Performance Analysis of Multivariate ARIMA and RNN model

Based on the performance indices presented in Table 1, the accuracy of the Multivariate ARIMA and Multivariate RNN models was evaluated. The evaluation included metrics such as R, R^2, d, RMSE, and MAE. The results indicate that the Multivariate RNN model demonstrated superior performance compared to the Multivariate ARIMA model specifically in predicting air temperature.

Table 1: Performance metrics of Multivariate ARIMA Model and Multivariate RNN Model

Models	Period	R	R^2	d	RMSE	MAE
Multivariate ARIMA	Training	0.01	0.07	0.52	5.72	4.97
	Testing	0.61	0.37	0.75	4.15	3.46
Multivariate RNN	Training	0.99	0.98	0.99	0.77	0.60
	Testing	0.97	0.94	0.99	1.20	0.88

3.2 Monthly Air Temperature Model

3.2.1 Monthly Air Temperature for Multivariate ARIMA Model

Figure 3 shows the Multivariate ARIMA model illustrating the monthly output of air temperature throughout the Training phase (1981-2011) and Testing phase (2012-2021) along with the observed data. The results suggest that the Multivariate ARIMA model achieved poor accuracy and exhibited an inadequate fit during both phases.

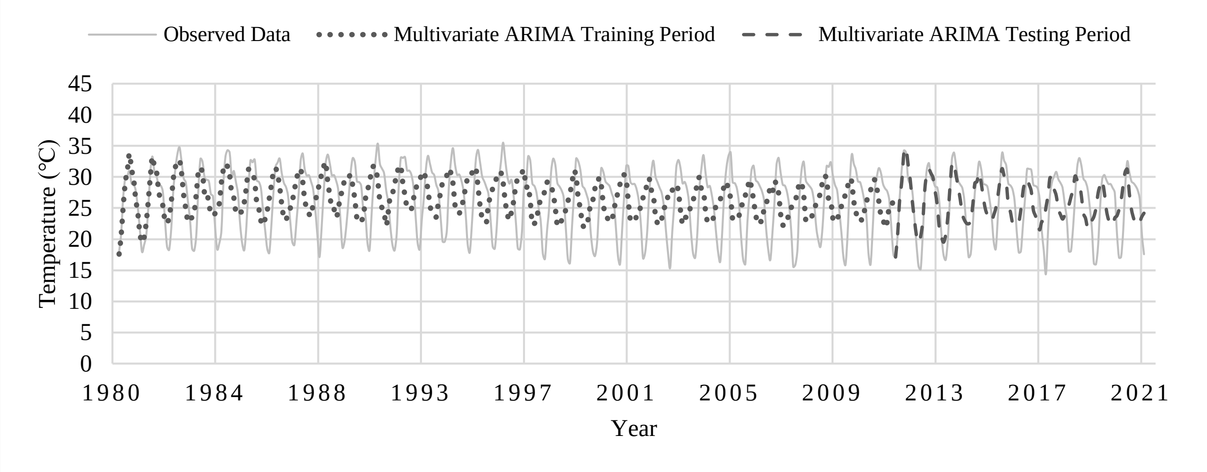

Figure 3. Variation of Multivariate ARIMA Model Output in Training and Testing Phase with the Observed data

3.2.1 Monthly Air Temperature for Multivariate RNN Model

In Figure 4, the Multivariate RNN model is presented visualizing the monthly output of air temperature during the Training phase (1981-2011) and Testing phase (2012-2021) accompanied by the observed data. The results indicate that the Multivariate RNN model achieved the highest accuracy and exhibited the best fit throughout both phases. Consequently, the Multivariate RNN model proves to be more effective in forecasting the monthly air temperature with minimal forecasting error.

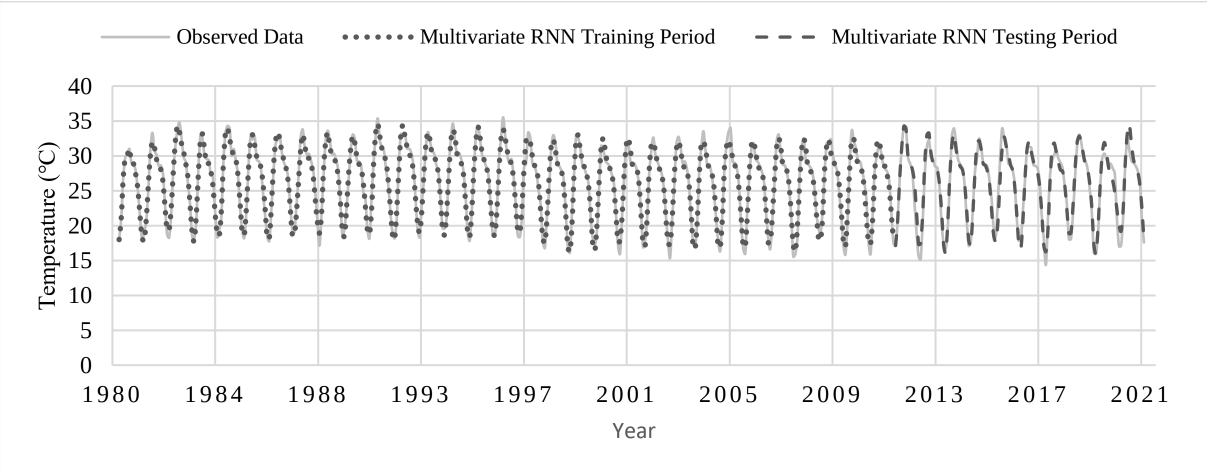

Figure 4. Variation of Multivariate RNN Model Output in Training and Testing Phase with the Observed data

4 Conclusion

The study utilizes both the Multivariate ARIMA and Multivariate RNN models to make predictions for the monthly air temperature in Chapainawabganj, Dhaka. The accuracy of the forecast provided by these models is compared by evaluating performance metrics such as R, R2, d, RMSE, and MAE. The results demonstrate that the RNN model performs the best, with values for R, R2, d, RMSE, and MAE during the training period of 0.99, 0.98, 0.99, 0.77, and 0.60, respectively, and 0.97, 0.95, 0.99, 1.20, and 0.88, respectively, during the testing period. On the other hand, the ARIMA model performs with values for R, R2, d, RMSE, and MAE of 0.07, 0.01, 0.52, 5.72, 4.97 during the training period and 0.61, 0.37, 0.75,4.15 and 3.46 during the testing period, respectively. Undoubtedly, the findings of this study indicate that the Multivariate RNN model exhibited superior performance compared to the Multivariate ARIMA model based on the assessed performance metrics. The results clearly indicated that the Multivariate RNN model achieved higher accuracy and demonstrated lower forecasting errors in comparison to the Multivariate ARIMA model. Therefore, based on the findings, it can be concluded that the Multivariate RNN model is more effective for predicting the monthly air temperature with enhanced accuracy.

References

Anisimov, O. A., Lobanov, V. A., & Reneva, S. A. (2007). Analysis of changes in air temperature in Russia and empirical forecast for the first quarter of the 21st century. Russian Meteorology and Hydrology, 32(10), 620-626.

Araujo, G., & Andrade, F. A. (2022). Post-Processing Air Temperature Weather Forecast Using Artificial Neural Networks with Measurements from Meteorological Stations. Applied Sciences, 12(14), 7131.

Ebtehaj, I., Bonakdari, H., Gharabaghi, B., & Khelifi, M. (2023). Time-Series-Based Air Temperature Forecasting Based on the Outlier Robust Extreme Learning Machine. Environmental Sciences Proceedings, 25(1), 51.

Fister, D., Pérez-Aracil, J., Peláez-Rodríguez, C., Del Ser, J., & Salcedo-Sanz, S. (2023). Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques. Applied Soft Computing, 136, 110118.

Hossain, I., Rasel, H. M., Imteaz, M. A., & Mekanik, F. (2020). Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics, 132, 131-141.

Hossain, I., Rasel, H. M., Mekanik, F., & Imteaz, M. A. (2020). Artificial neural network modelling technique in predicting Western Australian seasonal rainfall. International Journal of Water, 14(1), 14-28.

Hou, J., Wang, Y., Zhou, J., & Tian, Q. (2022). Prediction of hourly air temperature based on CNN–LSTM. Geomatics, Natural Hazards and Risk, 13(1), 1962-1986.

Kalteh, A. M. (2017). Enhanced monthly precipitation forecasting using artificial neural network and singular spectrum analysis conjunction models. INAE Letters, 2, 73-81.

Lee, S., Lee, Y. S., & Son, Y. (2020). Forecasting daily temperatures with different time interval data using deep neural networks. Applied Sciences, 10(5), 1609.

Ramadan, L., Shahrour, I., Mroueh, H., & Chehade, F. H. (2021). Use of machine learning methods for indoor temperature forecasting. Future Internet, 13(10), 242.

Rasel, H. M., & Imteaz, M. A. (2016). Application of artificial neural network for seasonal rainfall forecasting: a case study for South Australia. In Proceedings of the World Congress on Engineering (Vol. 1).

Rasel, H. M., Al Mamun, M. A., Hasnat, A., Alam, S., Hossain, I., Mondal, R. K., ... & Awual, M. R. (2023). Sustainable futures in agricultural heritage: Geospatial exploration and predicting groundwater-level variations in Barind Tract of Bangladesh. Science of The Total Environment, 865, 161297.

Ray, S. N., & Chattopadhyay, S. (2021). Analyzing surface air temperature and rainfall in univariate framework, quantifying uncertainty through Shannon entropy and prediction through artificial neural network. Earth Science Informatics, 14(1), 485-503.

Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophysica, 67(5), 1471-1481.

Tepanosyan, G., Asmaryan, S., Muradyan, V., Avetisyan, R., Hovsepyan, A., Khlghatyan, A., ... & Dell'Acqua, F. (2023). Machine Learning-Based Modeling of Air Temperature in the Complex Environment of Yerevan City, Armenia. Remote Sensing, 15(11), 2795.

Yeh, H. F., & Hsu, H. L. (2019). Stochastic model for drought forecasting in the Southern Taiwan Basin. Water, 11(10), 2041.

Zhou, J., Wang, D., Band, S. S., Mirzania, E., & Roshni, T. (2023). Atmosphere air temperature forecasting using the honey badger optimization algorithm: on the warmest and coldest areas of the world. Engineering Applications of Computational Fluid Mechanics, 17(1), 21741