Global Journal of Engineering and Technology [GJET].
ISSN: **2583-3359** (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

Modeling and Simulation of Univariate for "Groundwater", and Multivariate for "Rainfall, Temperature, Root and Surface Soil Witness, Depth to Groundwater level analytics" applying Deep Learning and Machine Learning Analytics of Time Series Forecasting in the Neural Network Model in the Mymensingh area of Bangladesh.

By

Ashraf Shahriar

Department of Electrical and Computer Engineering (ECE), ID:2025310050
North South University, Plot # 15, Block # B, Bashundhara, Dhaka – 1229, Bangladesh.

Article History

Received: 05/11/2024
Accepted: 21/11/2024
Published: 23/11/2024

Vol – 3 Issue – 11

PP: - 05-11

DOI:10.5281/zenodo. 14210207

Abstract

This research explores the modeling and simulation of groundwater dynamics applying the univariate and multivariate time series forecasting. The univariate analysis focuses on groundwater levels, whereas the multivariate analysis integrates related variable quantity such as rainfall, temperature, root and surface soil witness, and depth to groundwater level. The incorporation of progressive computational systems such as deep learning and machine learning offers significant enhancements in analytical exactness and model robustness compared to traditional numerical approaches. Key results of this study comprise the expansion of projecting models that can be used to estimate groundwater levels based on the existing and historic data of related variable quantity. The developed models can support policymakers and stakeholders in making informed results concerning groundwater usage and maintenance.

Keywords: *Univariate, Multivariate, Temperature, Humidity, Rainfall, Surface Soil Witness, Time Series Forecasting.*

I. INTRODUCTION

This study offers a comprehensive view of how advanced analytics techniques like Deep Learning (DL) and machine Learning (ML) can be applied to the time series forecasting, particularly in groundwater management and environmental monitoring. Developing methods to interpret the decision-making processes of DL models to better understand the relationship between variables. The relationships between multiple climatic and environmental factors, modeled together to improve prediction accuracy (Adhikari and Ikeda, 2020). Rainfall modeling and simulation are decisive tools for considerate and forecasting rainfall patterns. Particular temperature modeling supports in predicting weather, reviewing climate variation, and dealing agricultural performs. More than a few machine learning algorithms can be castoff for research to find the above statements analysis, as well as Support Vector Regression (SVR), Random Forest (RF), K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). The emphasis of the research has been fascinated on current enlargements, advancements, margins and insufficiencies of Advanced Neural Networking (ANN) tactics by using the Deep Learning system (Gharbi and Bouaziz, 2023). The relevant and expected measurements' level data were employed to train and test the Neural Network. With the usage of the proficiency principles, mean square error (MSE), root mean square error (RMSE), and other metrics, each network structure's prediction accuracy was evaluated (R2). Results have been demonstrated the time series forecasting in neural network (NN) model in the area of Mymensingh that is located in the Latitude 24.88833 and Longitude 90.46861.

II. METHODOLOGY

Data Collection

Appropriate data was collected from NASA for multivariate and Bangladesh Water Development Board (BWDB) for univariate time series analysis.

A. Loss and accuracy function

R-squared (R^2) is a numerical formation where indicates, 1 = best, 0 or < 0 = worse. If the extent of loss function is high, it means algorithm is presentation a lot of alterations the consequence and desires be amended.

*Corresponding Author: Ashraf Shahriar

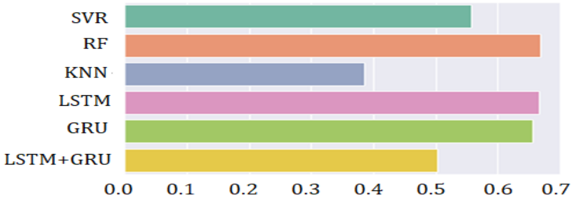


Figure 1: **Application of algorithms, loss and accuracy**

Observations: RF appears (Figure:1) the highest Test R2 Score, close to 0.7. The. LSTM and GRU perform well, scores slightly lower than RF. SVR and the LSTM+GRU have similar performance, with scores around 0.5. And the KNN seems the lowest score among the algorithms**.**

A. Split data for Training and Testing

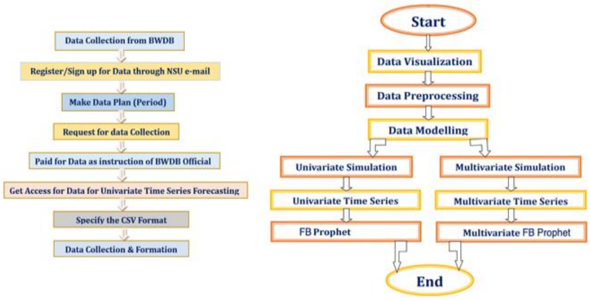
The input data is split as training and testing:

Figure 2: Split data for training and testing

Figure 3: **Split data for training and testing with periodical analysis**

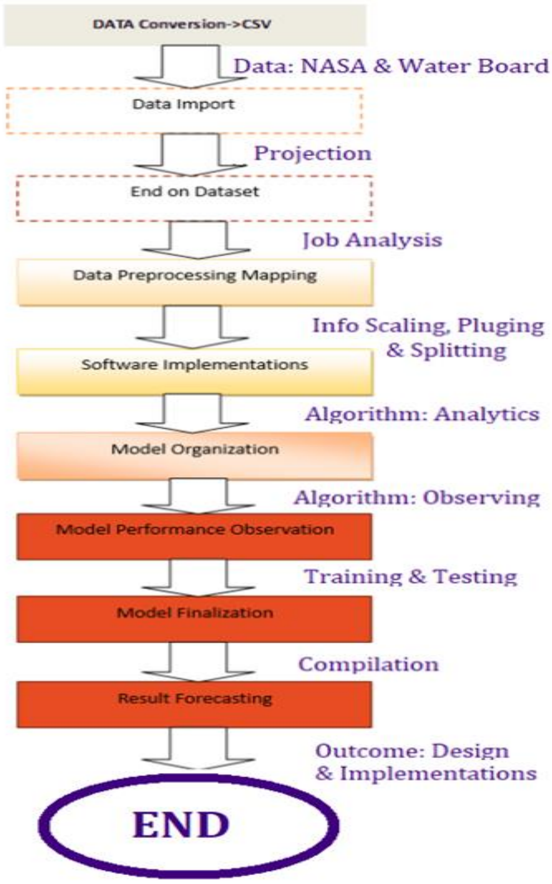
Observations: All graphs show (Fugure:3) a cyclical pattern in the data, suggesting seasonal variations in water levels. The main difference between the left and right columns is the training size approach. The training data shows a cyclical pattern with peaks and troughs, likely indicating seasonal variations in the water levels.

A. Model Setup and Data format (Flowchart:1)



Flowchart 1: **Data collection method and applicable research forming**

D. Research Design and process setup (Flowchart:2)



Flowchart 2: **Research and process flow**

E. Time series prediction to forecast the future groundwater levels, considering factors such as water table depth, parapet height, and geographical directs (Table:1).

	SL	DISTRICT	UPAZILA	WELL ID	OLD ID	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
0	1	Mymensingh	Mymensingh Sadar	GT6152020	MY024	7.80	13.63	0.69	35.98	24.88833	90.46861
1	2	Mymensingh	Mymensingh Sadar	GT6152020	MY024	8.18	13.63	0.69	35.98	24.88833	90.46861
2	3	Mymensingh	Mymensingh Sadar	GT6152020	MY024	8.46	13.63	0.69	35.98	24.88833	90.46861
3	4	Mymensingh	Mymensingh Sadar	GT6152020	MY024	8.87	13.63	0.69	35.98	24.88833	90.46861
4	5	Mymensingh	Mymensingh Sadar	GT6152020	MY024	9.42	13.63	0.69	35.98	24.88833	90.46861

Table 1**: water table depth, parapet height, and geographical directs.**

Application and Summary of SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithms. *G*raph provides a visual comparison of the performance of different machine learning algorithms in predicting GWL in Mymensingh (Table:2).

Algorithms Name	Train RMSE	Test RMSE	Train MSE	Test MSE	Train MAE	Test MAE	Train VRS	Test VRS	Train R2 Score	Test R2 Score	Train MGD	Test MGD	Train MPD	Test MPD
Support Vector Regression	1.283726	1.543685	1.647951	2.382962	1.007969	1.154555	0.723821	0.402937	0.723599	0.389824	0.019854	0.030871	0.177329	0.267364
Random Forest	0.813173	1.459691	0.661250	2.130697	0.661501	1.161981	0.889123	0.465690	0.889092	0.454418	0.008328	0.026973	0.072033	0.237325
K Nearest neigbor	2.021792	1.506218	4.087642	2.268692	1.686453	1.275634	0.316049	0.470364	0.314403	0.419084	0.050257	0.029878	0.447005	0.257825
LSTM	2.158069	1.443908	4.657261	2.084871	1.800295	1.183253	0.224129	0.466418	0.218865	0.466152	0.056764	0.026892	0.507445	0.234231
GRU	2.101055	1.400112	4.414431	1.960313	1.711538	1.103032	0.266032	0.498902	0.259593	0.498046	0.054165	0.024886	0.482384	0.218532
LSTM+GRU	2.274672	1.546368	5.174131	2.391254	1.844850	1.157177	0.223900	0.443971	0.132173	0.387700	0.063085	0.029748	0.564285	0.264186

Table 2: **visual comparison of the performance of different machine learning algorithms**

III. MODELING AND SIMULATION

Univariate Time Series Forecasting for Groundwater Level (GWL)

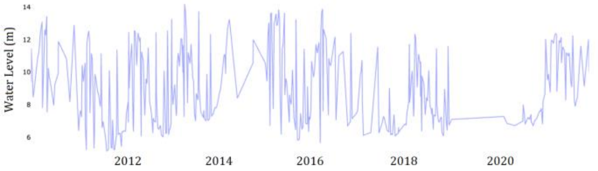


Figure 4: **GWL chart, water level Mymensingh**

Observations: Relatively consistent across the years (Figure:4), though there are some variations in the extremes from year to year. There appears to be more frequent and rapid fluctuations in some years compared to others. Recent data (around 2020) shows a sharp rise in water levels after a period of relative stability.

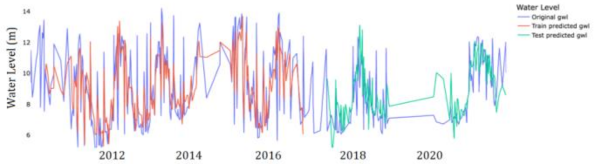


Figure 5: **Original Vs predicted GWL of Mymensingh by SVR**

Observations: The model seems (Figure:5) to capture the overall trends and seasonality, but it doesn't always match the exact peaks and troughs of the original data. There appears to be more variation in the prediction accuracy during certain periods, particularly in the test set.

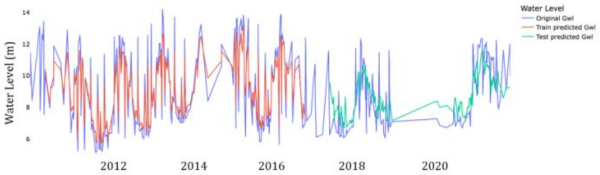


Figure 6: **Comparison between original GWL vs predicted GWL chart of Rajshahi by RF**

Observations: *The model seems (Figure:6) to capture the overall trends and seasonality, but doesn't always match the exact peaks and troughs of the original data. There appears to be more variation in the prediction accuracy during certain periods, particularly in the test set.*

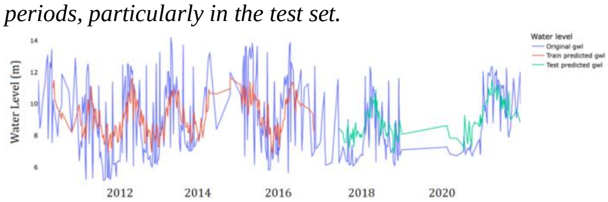


Figure 7: **Comparison between original GWL vs predicted GWL with chart by KNN**

Observations: The KNN model seems (Figure:7) to capture the overall trends and seasonality, but doesn't always match the exact peaks and troughs of the original data. There appears to be more variation in the prediction accuracy during certain periods, particularly in the test set.

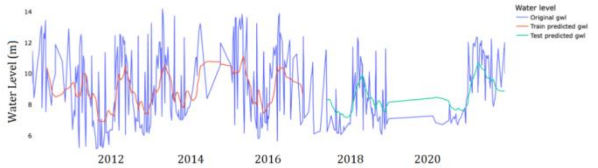


Figure 8: **Comparison of original GWL price vs predicted GWL chart of by LSTM**

Observations: Both prediction lines (red and green) seem (Figure:8) to capture the overall seasonal patterns of the original data but smooth out the short-term variations. The original data shows peaks reaching around 14 meters and troughs as low as 6 meters.

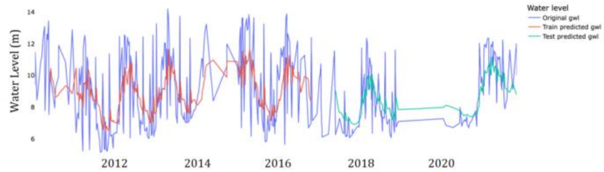


Figure 9: **Comparison between original GWL vs predicted GWL chart of GRU**

Observations: The model's predictions seem (Figure:9) to capture the overall trend but miss some of the short-term variability present in the original data. There's a noticeable gap in the original data around 2020, where the blue line is absent.

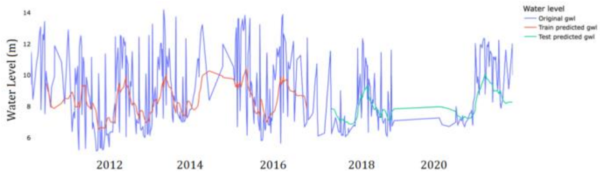


Figure 10: **Comparison between original GWL vs predicted GWL chart of Mymensingh by LSTM+GRU**

Observations: The model seems (Figure#10) to capture the overall trend and seasonality of the groundwater levels, but it doesn't always match the exact peaks and troughs of the original data. There's a noticeable gap in the original data around 2019, which the model appears to interpolate. The seasonal pattern shows higher water levels (peaks) occurring roughly annually, possibly corresponding to wet seasons or monsoons.

B. Multivariate Time Series Forecasting for Groundwater Level, Rainfall, Temperature, Root and Surface Soil Witness, Depth to Groundwater level.

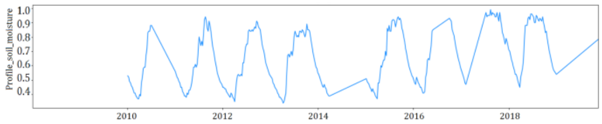


Figure 11: **Multivariate analysis, profile soil moisture, Rajshahi zone**

Observations: The amplitude of these cycles varies (Figure#11) from year to year, with some years showing more extreme highs and lows than others. There appears to be a

*Corresponding Author: Ashraf Shahriar

slight overall increasing trend in the peaks over the years, though this would need statistical verification. The cycles are not perfectly regular, with some years showing broader peaks or multiple peaks

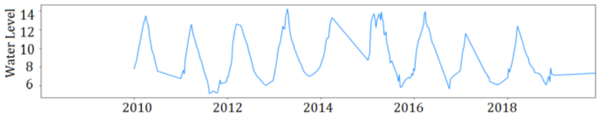


***Figure12:* Multivariate analysis, water level, Mymensingh zone**

Observations: There's a clear cyclical pattern (Figure:12), likely representing seasonal variations in water level. Peaks generally occur around same time, possibly corresponding to wet seasons. Amplitude of cycles varies from year to year, with some years showing higher peaks than others.

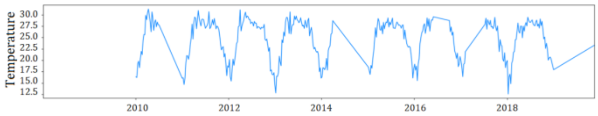


***Figure 13:* Multivariate analysis, temperature, Mymensingh zone**

Observations: The overall pattern suggests (Figure:13) seasonal temperature changes, with higher temperatures in summer months and lower in winter. Temperature range is quite wide, indicating a climate significant seasonal variation. Like the top graph, there's no clear long-term trend visible in the temperature data.

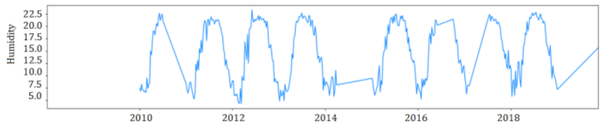


***Figure 14:* Multivariate analysis, humidity, Mymensingh zone**

Observations: There's a clear cyclical pattern (Figure:14), likely representing seasonal variations in humidity. The peaks generally occur at regular intervals, corresponding to annual wet seasons. The amplitude of the cycles is fairly consistent year to year, with peaks typically reaching around 20-22 units and troughs around 5-10 units.

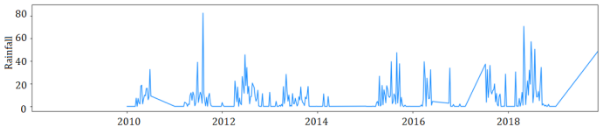


Figure 15: Multivariate analysis, Rainfall, Mymensingh zone

Observations: There are clear spikes in rainfall (Figure:15), with some years showing more intense or frequent rainfall events than others. The rainfall appears to be highly seasonal, with periods of little to no rain followed by intense rainfall events. The highest rainfall spike appears to be in 2012, reaching about 80 units.

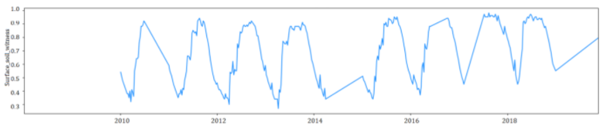


Figure 16: Multivariate Analysis (surface soil witness), Mymensingh

Observations: The pattern in graph is remarkably similar (Figure:16), indicating a strong correlation between surface and root zone soil moisture. There's no obvious long-term trend visible in either data set over this time period. Data gaps: There appears to be a gap in data around early 2014.

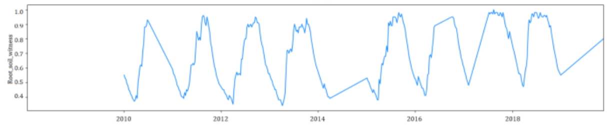


Figure 17: Multivariate Analysis (root soil), Mymensingh

Observations: The amplitude of the cycle varies from year to year (Figure:17), with some years showing more extreme highs and lows than others. The pattern in graph is remarkably similar, indicating a strong correlation between surface and root zone soil moisture. There's no obvious long-term trend visible in either data set over this time period. There appears to be a gap in data around early 2014.

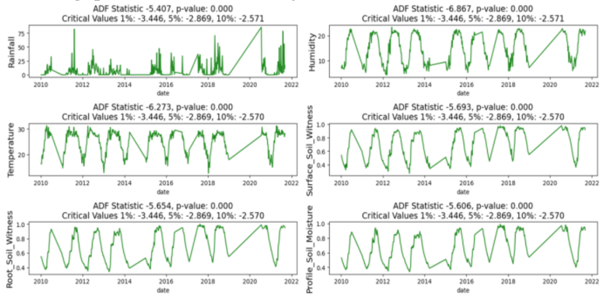


***Figure 18:* Multivariate Series (Soil, Temperature, Humidity, Rainfall, Surface Soil witness)**

Observations: Each graph includes the results of an Augmented Dickey-Fuller (ADF) test (Figure:18), which is used to test for stationarity in time series data. a p-value underneath 0.05. which Examine the ADF indicators range in next of kin to fundamental echelons. Breakdown of each graph:

All graphs have ADF test results indicating stationarity This suggests that these time series don't have unit roots and their statistical properties are constant over time. The cyclical patterns in most graphs likely represent seasonal variations. The rainfall graph stands out as the most irregular, which is typical for precipitation data. These visualizations are useful for understanding the temporal patterns and relationships between different environmental variables, which could be important for hydrological or agricultural studies.

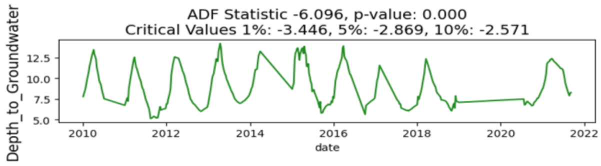


***Figure 19:* Time Series Analysis (periodical status)**

Observations: The ADF test results suggest (Figure:19) that the time series is stationary (p-value < 0.05 and test statistic

*Corresponding Author: Ashraf Shahriar

more negative than critical values), which means it doesn't have a unit root and its statistical properties like mean, variance, and autocorrelation are constant over time.

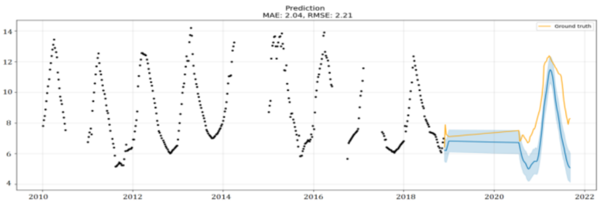


Figure 20: **Prediction – MAE & RMSE, Depth to GWL analysis, Multivariate - Mymensingh**

Observations: This visualization is useful for assessing the model's performance (Figure:20) in predicting groundwater levels and understanding long-term trends and seasonal variations in the data. The relatively higher MAE and RMSE values indicate that there's room for improvement in the model's predictions.

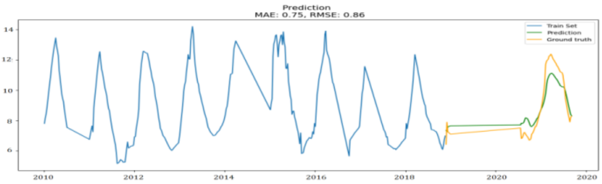


Figure 21: **Multivariate Time Series Analysis (train set, prediction, ground truth)**

Observations: The historical data shows (Figure:21) a consistent seasonal pattern with peaks & troughs occurring regularly. Prediction line follows ground truth, suggesting good model performance. There's a notable spike in groundwater levels around 2021, captured by both the prediction and ground truth lines. The model seems to perform well in predicting both the overall trend and the seasonal fluctuations.

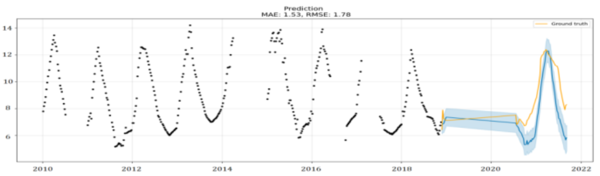


Figure 22 **Multivariate Time Series Analysis (Prediction: MAE, RMSE), ground truth**

Observations: The model's ability to capture (Figure:22) this cyclical pattern in its predictions. A reasonable match between predictions & ground truth in later years, though some discrepancies. Increased variability & potentially some anomalous behavior in the most recent data (2020-2022).

IV. FINDINGS&RECOMMENDATIONS

A. Major Findings

- **Temporal Dependencies:** The DL and ML models could categorize the lagged relationships and seasonality patterns that should be influenced the behavior of the variables over time.
- **Nonlinear Relationships:** The study could expose nonlinear relationships and interactions between the ecological variables such as neural networks, are well-suited for capturing complex, nonlinear changing aspects in multivariate time series data.
- **Variable Importance:** By investigating feature importance in the predicting models, the most significant influences driving changes in the groundwater levels and soil moisture content. This could support the prioritize managing actions and involvements aimed at protective water resources and enhancing ecology resilience
- **Prediction Accuracy:** The research and study area might validate the superior calculation accuracy of DL and ML models associated to the traditional statistical approaches for time series forecasting (Karthikeyan, Khosa and Singh, 2020).
- **Long-Term Trends:** By studying the historic data and projecting future circumstances, the research could classify the long-term trends and fluctuations in the groundwater accessibility, precipitation regimes, and soil moisture dynamics associated with climate variation and anthropogenic effects. This information could update the adaptation approaches and water resource forecasting ingenuities.
- **Uncertainty Analysis:** The study might calculate the uncertainty related with the model forecasts and assess the sensitivity of outcomes to the differences in input parameters, model structures, and figures sources. Uncertainty breakdown could afford to the decision-makers with a more nuanced considerate of the reliability and restrictions of predicting consequences.
- **Management Strategies:** Grounded on the findings, the study could recommend adaptive managing approaches and policy references for sustainable water source management, and ecological protection. This could contain enhancing irrigation plans, applying water-saving technologies, and protective groundwater recharge zones.
- **Interdisciplinary Insights:** The interdisciplinary nature of the study could foster relationships among hydrologists, climatologists, agronomists, and data scientists to address the complex water-related challenges from multiple viewpoints (Mojid, Parvez, Moinuddin and Hodgson, 2019).
- **Performance of the applied Algorithms:** The utilization of the different predicting models, including SVR, RF, KNN, LSTM, GRU, and LSTM+GRU, was assessed effectively and have found the predictable outcome. The models confirmed varying degrees of correctness in forecasting the groundwater levels, with some outperforming others in the certain circumstances.

B. Major recommendations

- **Data Quality and Availability:** Confirm the high-quality and consistent data for all variable quantity

*Corresponding Author: Ashraf Shahriar

of awareness, including groundwater levels, rainfall, temperature, and soil moisture. Apply multiple data sources, such as weather stations, soil sensors, and groundwater monitoring wells, to capture spatial and temporal changeability.

- **Feature Engineering:** Conduct systematic feature engineering to the abstract expressive predictors from the raw data. Reflect integrating lagged variable quantity, seasonal indicators, and meteorological indices to capture chronological forms and dependencies.
- **Multivariate Modeling:** Implement multivariate time series estimating models that concurrently forecast multiple variables of interest This allows for capturing the complex interactions and response loops amid ecological variable quantity (Abdollahi, Bazrafshan and Razmjooy, 2020).
- **Data Preprocessing:** Preprocess and ensures the data sensibly, handling missing principles, outliers, and seasonality. Normalize the input variables to ensure that they are on a similar scale and facilitate convergence during training (Lundberg and Lee,2017).
- **Model Training and Validation:** Divided the dataset into training, validation, and test sets to assess the evaluate of the models. Train the models by means of the training data and authenticate them using the validation set.
- **Joint Approaches:** Reflect using collective approaches to combine estimates from multiple models or model alternatives. Ensemble methods such as bagging, boosting, and stacking can expand the predicting accuracy and robustness by leveraging the strengths of diverse algorithms.
- **Interpretability and Explainability:** Develop the interpretability of the predicting models investigating the feature importance, variable contributions, and model forecasts.
- **Model Evaluation Metrics:** Select the proper valuation metrics to evaluate the exactness and reliability of the predicting models. Common metrics consist of MAE, MSE, RMSE, $R^{2.}$
- **Uncertainty Quantification:** Measure the uncertainty associated with the model forecasts and evaluate the sensitivity of outcomes to differences in input parameters and model assumptions (Aishwarya and Vasudevan, 2023).
- **Validation and Application:** Authenticate the predicting models using independent datasets and real-world observations to govern their generalizability and applicability in the effective backgrounds.

V. RESULTS

A. Loss Score

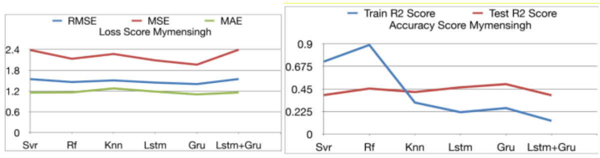


***Figure 23*: Loss and Accuracy Score formation**

Observations: The quantity (Figure:23) fluctuated amongst 1.11 and 2.40 for the LSTM+GRU and KNN consistently. By distinction, the MSE and the RMSE usual less losses. The quantity of losses which originate the MAE persisted identical which is 0.16 for SVR, RF, KNN and LSTM algorithms. At that moment the loss augmented in GRU which is 1.10 and diminished in LSTM+GRU which is 0.14. We can realize that the inclination for the MAE was fluctuated. The value for SVR, RF, KNN and LSTM algorithms are 1.11, 1.12, 1.11, 1.12 correspondingly. Then the loss augmented in GRU which is 1.11 and diminished in GRU is 1.11. Test R2 Score is the authentic accuracy score of the algorithms. As can be seen from the graph, there were different trends for Train R2 Score and Test R2 Score. The value of Train R2 Score augmented 0.68 to 0.90 for SVR and RF correspondingly. Subsequently that, the Train R2 Score intensely deteriorated to 0.20 in the GRU algorithms. This is the bottommost training accurateness over the above six algorithms. Then the value augmented to 0.220 for LSTM+GRU algorithms. On the other hand, the Test R2 Score persisted almost same for SVR, RF, KNN and LSTM which is about 0.44. It diminished to 0.44 for SVR and increased to 0.46 for GRU. The lowest accuracy is found in GRU algorithm and the highest in LSTM+GRU.

A. ***Accuracy Score Heatmap:*** The below defined Accuracy Score Heatmap correlation of algorithms of Mymensingh: This shows the heat map correlation of SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithm.

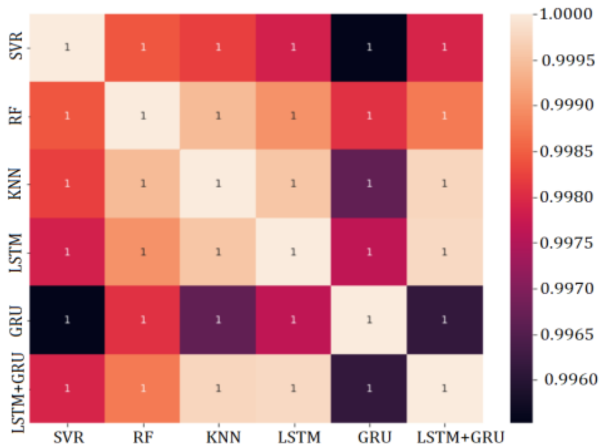


Figure 24: Accuracy Heatmap of Rajshahi

Observations: The high correlations suggest (Figure:24) that most algorithms perform similarly, but the subtle differences could be important for selecting the most appropriate model for specific applications.

*Corresponding Author: Ashraf Shahriar

VI. CONCLUSION

Predicting the future of Artificial Intelligence (AI) and Neural Networks (NN) contains inspecting current trends, technical developments, and probable applications to make informed forecasts about their future expansion. Computational models of NN help to test models of reasoning developments, such as consideration, intellectual, and decision-making. Forecasting the precise consequences can be challenging, understanding current trends and developing technologies can provide valued insights into the potential instructions of AI expansion. In this research, I have allied the performance of immeasurable machine learning algorithms, as well as SVR, RF, KNN, LSTM, GRU, and LSTM+GRU for Modeling and Simulation by Applying Deep Learning Univariate and Multivariate Time Series Forecasting in Neural Network Model for the leading divisional area of that is Rajshahi. Considerate the fundamental mechanisms and values governing the behavior of these models can be challenging, principally for deep learning architectures with millions of constraints (Cheng and Castelletti, 2020). Train the particular models by means of the training data and tune hyper restrictions using the authentication set. Employ methods such as cross-validation and grid search to improve the model performance. Abstract appropriate structures from the raw data that can capture the changing aspects and mutuality between the variables. These may include lagged variables, periodic indicators, and meteorological indices. I think this article to serve as an insightful and comprehensive resource for researchers and experts in the area.

REFERENCES

1. Abdollahi, A., Bazrafshan, J., & Razmjooy, N. (2020). *A deep learning method for temperature forecasting based on Bi-LSTM. Applied Soft Computing*, 97, 106782.
2. Acharyya, A., (2017), *Sustainable Development of Groundwater Resources in India*: A Relook at Policy Initiatives, Vol-VIII.
3. Adhikari, U., & Ikeda, M. (2020). XGBoost and LSTM model for multivariate time series forecasting of groundwater level. Water, 12(11), 3082. DOI.
4. Aishwarya, R., & Vasudevan, V. (2023). A comprehensive review of multi-source data integration for groundwater modeling. Journal of Hydrology, 617, 128812. DOI.
5. Cheng, L., Li, G., & Castelletti, A. (2020). *GWL forecasting with LSTM networks: A study in agricultural water management.* Hydrological Sciences Journal, 65(9), 1505-1516.
6. Fawaz, H. I., Forestier, G., Weber, J., et al. (2019). *Deep learning for time series classification: A review*. Data Mining and Knowledge Discovery, 33(4), 917-963.
7. Gharbi, S., & Bouaziz, M. (2023). Real-time data assimilation for GWL prediction using machine learning. Water Resources Management, 37(5), 1461-1478. DOI.
8. Karthikeyan, L., Khosa, R., & Singh, V. P. (2020). *Deep learning models for soil moisture retrieval from remote sensing data: A review. Environmental Earth Sciences*, 79(7), 193.
9. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).
10. Lim, B., & Zohren, S. (2021). Time series forecasting with deep learning: *A survey. Philosophical Transactions Royal Society: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200209.
11. Mojid, M.A., Parvez, M.F., Mainuddin, M. and Hodgson, G., (2019), Water Table Trend- *A Sustainability of GWL Development in North-West Bangladesh*, Water, Vol 11, pp-1182.

*Corresponding Author: Ashraf Shahriar