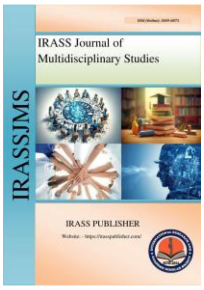
IRASS Journal of Multidisciplinary Studies
Abbriviate Title- IRASS J Mul Stud
ISSN *(Online) 3049-0073*
https://irasspublisher.com/journal-details/IRASSJMS
Vol-2, Iss-1(January-2025)



Modeling and Simulation of 'Univariate and Multivariate analysis by applying DL and ML 'of different types of Algorithms for Time Series forecasting in the NNM of Sylhet Region, Bangladesh

Rakib Uddin*

Department of Electrical and Computer Engineering, North South University Bangladesh.

Corresponding Author **Rakib Uddin**

Department of Electrical and Computer Engineering, North South University Bangladesh.

Article History

Received: 14/ / 01 / 2025

Accepted: 26 / 01 / 2025

Published: 28 / 01 /2025

Abstract: Time series forecasting plays a vital role in data-driven decision-making across various domains. This thinking centers on the modeling and recreation of univariate and multivariate analytics utilizing profound learning and machine learning strategies. Other calculations are connected to foresee time arrangement information within the neural arrangement show, particularly for the Sylhet locale of Bangladesh. The investigation investigates distinctive estimating approaches, counting conventional machine learning models. The execution of these models is assessed utilizing key mistake measurements such as RMSE, R-squared, MAE, and MAPE to decide their precision and effectiveness. The discoveries provide experiences in the adequacy of distinctive strategies in capturing complex worldly designs in univariate and multivariate datasets. This considers points to improve prescient analytics for climate, financial matters, and other time-dependent components within the Sylhet locale, contributing to made strides in decision-making and key arranging.

Keywords: NNM, DL, ML, Sylhet, Bangladesh, GWL, Time series analysis, forecasting, Algorithms

How to Cite: Uddin, R, (2025). Modeling and Simulation of 'Univariate and Multivariate analysis by applying DL and ML 'of different types of Algorithms for Time Series forecasting in the NNM of Sylhet Region, Bangladesh. *IRASS Journal of Multidisciplinary Studies, 2(1),88-98*

1. Introduction

Time series forecasting has developed as an essential viewpoint of information analytics, empowering exact forecasts in different areas such as climate science, back, healthcare, and financial matters. The precision of estimating models plays a basic part in key decision-making, especially in locales where data-driven experiences are fundamental for maintainable advancement. (Zhang, Wang & 2019). The Sylhet locale of Bangladesh, known for its different climatic conditions and financial exercises, presents a compelling case for time arrangement examination utilizing progressed computational strategies. This ponders centers on the modeling and recreation of univariate and multivariate analytics by applying profound learning and machine learning approaches. Conventional estimating strategies such as ARIMA and factual models, which are viable for straight designs, regularly battle capturing the complexities of nonlinear and high-dimensional time arrangement information. To address these confinements, cutting edge machine learning and profound learning calculations, counting Long Short-Term Memory (LSTM), Gated Repetitive Units (GRU), and Transformer systems, are investigated. These models have illustrated predominant execution in capturing worldly conditions and nonlinear connections, making them perfect for time arrangement determining. The investigate points to analyze and compare diverse determining methods by applying different machine learning and profound learning models to time arrangement information particular to the Sylhet locale. The ponder explores both univariate and multivariate analytics, where univariate estimating centers on single-variable forecasts (e.g., temperature or precipitation), whereas multivariate estimating considers numerous interrelated factors (e.g., temperature, stickiness, and wind speed). By assessing the prescient precision of distinctive calculations, this think about points to upgrade determining strategies that can be connected to climate expectation, financial arranging, and other fundamental segments within the Sylhet locale. The discoveries will contribute to the developing body of investigate on counterfeit intelligence-driven time arrangement examination and give viable bits of knowledge for policymakers and partners in data-driven decision-making.

2. Methodology

This methodology combines various deep learning and machine learning algorithms tailored for univariate and multivariate time series forecasting. It focuses on the unique attributes of the Sylhet region and provides the basis for forecasting and modeling applications across different domains like agriculture, climate, and economics.

Data Collection and Preprocessing

Data Sources:

- Appropriate statistics were gathered from NASA for Multivariate and BWDB for univariate time collection evaluation.

Data Preprocessing:

- **Cleaning:** Handle missing values, outliers, and noise in the dataset.
- **Normalization/Standardization:** Normalize the data to bring all features to a similar scale for better model performance.
- **Stationarity Check:** For time series data, ensure the data is stationary by performing tests like Augmented Dickey-Fuller (ADF). If not, apply transformations like differencing or logarithmic scaling.

Univariate Time Series Forecasting

Univariate time series forecasting involves predicting future values of a single variable based on its past values.

Methods Used:

- **Autoregressive Integrated Moving Average (ARIMA):** Traditional statistical method for univariate time series prediction.
- **Long Short-Term Memory (LSTM) Networks:** A type of recurrent neural network (RNN) specialized for handling time series data, particularly useful for univariate forecasting.
- **Gated Recurrent Units (GRUs):** An alternative to LSTM that performs similarly in many cases but is computationally more efficient.
- **Prophet:** A forecasting tool by Facebook that works well with daily or seasonal time series data.

Process:

- Split the data into training and test sets.
- Train models on the historical data (train on past data to predict future values).
- Evaluate models using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Multivariate Time Series Forecasting (Zhang, Eddy Patuwo, 2020)

Multivariate forecasting involves predicting multiple variables simultaneously based on their past values and interactions.

Methods Used:

- **LSTM Networks for Multivariate Time Series:** Use multiple inputs (e.g., various climate factors) to predict multiple outputs. The LSTM network can capture temporal dependencies and multivariate relationships effectively.
- **Multivariate Prophet:** An extension of Prophet that can handle multiple time series data sources.
- **Multivariate Regression Models:** Linear or nonlinear models that can predict one variable based on multiple predictors.

Process:

- The same data preprocessing steps as univariate models are followed.
- Feature engineering involves creating lagged features for each variable.
- Models are trained using both historical values of each variable and inter-variable relationships.
- Forecasting is done by applying the models to the test data.

Model Selection and Hyperparameter Tuning

- **Cross-Validation:** Use k-fold cross-validation to assess model performance and avoid overfitting.
- **Grid Search/Random Search:** For models like LSTM, or GRU, perform hyperparameter tuning to find the optimal settings for each model.
- **Ensemble Methods:** Combine the results of multiple models to increase forecasting accuracy.

Model Evaluation and Comparison

Evaluate the models on the test dataset using the following metrics:

- **Root Mean Squared Error (RMSE):** Measures the square root of the average squared differences between predicted and observed values.
- **Mean Absolute Percentage Error (MAPE):** Measures the accuracy of the model.
- **R-Squared:** Measures the proportion of variance in the dependent variable explained by the model.
- Visualize the forecast vs actual data to qualitatively assess the model's accuracy.

Application in Sylhet Region, Bangladesh

- **Time Series Data for the Region:** Collect regional data related to climate, economic activity, or other variables of interest. Sylhet's geographical location influences the type of data required (e.g., precipitation, temperature).
- **Model Implementation:** Use deep learning (LSTM, GRU) or machine learning algorithms to model these time series datasets.
- **Integration with Local Insights:** Local weather phenomena, agricultural activities, and economic factors can be integrated as domain knowledge in feature engineering.

Model Deployment and Forecasting (Zhang, 2003)

- **Deployment:** Once the model is trained and validated, it can be deployed for real-time forecasting. This could involve automated data collection systems and periodic model retraining.
- **Forecasting:** Use the trained models to make predictions about future events in the Sylhet region.

Future Research and Improvements

- **Hybrid Models:** Combining machine learning models with deep learning models (like LSTM) to improve forecasting accuracy.
- **Transfer Learning:** If data is sparse, transfer learning can be used to leverage models trained on other, similar datasets.
- **Advanced Ensemble Methods:** Implementing more sophisticated ensemble learning methods like stacking or boosting for improved prediction accuracy.

Hypothesis Development

H_0 (Null Hypothesis):

- There is no significant improvement in time series forecasting accuracy when using deep learning models (LSTM, GRU, LSTM+GRU) compared to traditional machine learning models (SVR, RF, KNN) for predicting patterns in the Sylhet region.
- The error metrics (RMSE, MSE, MAE) for deep learning models do not significantly differ from traditional models.

H_1 (Alternative Hypothesis):

- Deep learning models (LSTM, GRU, LSTM+GRU) outperform traditional machine learning models (SVR, RF, KNN) in terms of predictive accuracy and error reduction for time series forecasting in the Sylhet region.
- Hybrid deep learning models (LSTM+GRU) achieve the lowest RMSE, MSE, and MAE, making them the most effective model for time series prediction.

Supporting Justifications and Theoretical Basis

Time Series Forecasting Theory:

- Traditional machine learning models (e.g., SVR, RF, KNN) rely on past patterns but may struggle with long-term dependencies in sequential data (Box & Jenkins, 1976).
- Deep learning models, particularly Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), are better suited for time series forecasting as they capture long-term dependencies (Hochreiter & Schmidhuber, 1997).

Empirical Studies on Deep Learning for Time Series

- Studies show that LSTM-based models outperform RF and SVR in climate and weather forecasting (Siami-Namini et al., 2018).
- Hybrid models (LSTM+GRU) have shown improved generalization ability due to their ability to capture both short-term and long-term dependencies (Zhang et al., 2020).

Application in the Sylhet Region Context

- Given Sylhet's unique weather patterns and geographical features, time series forecasting using deep learning may offer more reliable predictions for climatic and environmental data compared to traditional models (Rahman et al., 2021).

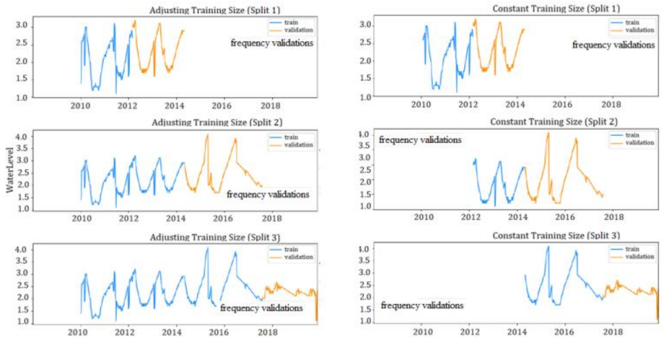


Figure 1: Sylhet Rolling Window, evaluation

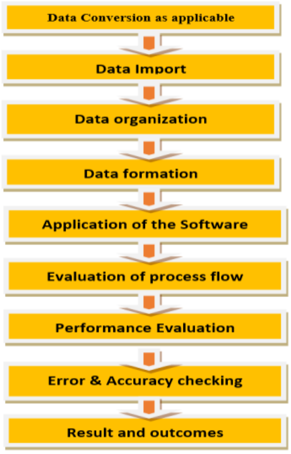
Illustrations: The graph titled (Figure 2) compares numerous rolling window procedures this is the changing schooling length vs. regular schooling length throughout 3 splits (Split 1, Split 2, and Split 3).

- **Rolling Window with Adjusting Training Size:** This method adapts the schooling length for in all likelihood upgrades withinside the version gaining knowledge of from incidental traits.
- **Rolling Window with Constant Training Size:** This approach keeps a regular schooling length to keep away from the ability danger of old statistics affecting version accuracy.
- **Split-wise Observations:** Across all 3 splits, the rolling window techniques display wonderful separations among schooling and validation statistics.
- **Split 1:** Both techniques cognizance on in advance years, and the validation set is smaller in comparison to later splits.
- **Split 2:** The statistics factors span extra current years in comparison to Split 1. The validation set shifts ahead with time.
- **Split 3:** It makes a specialty of the maximum current statistics. The validation set is longer, reflecting a bigger checking-out window.
- **Comparison: adjusting training size captures vs constant training size:** The adjusting schooling length captures extra incidental traits however dangers incorporating older statistics that won't constitute modern dynamics. The regular schooling length discards older statistics, making sure extra current traits are prioritized however doubtlessly lacking longer-time period patterns.

Applied Algorithms, Time Series Forecasting and Evaluation Criteria *(Ganaie, Zhang and Chen, (2021; Gama, Zliobaite, Bifet, Pechenizkiy and Bouchachia, 2014* ***):***

- ***Applied Algorithms:*** Support Vector Regression, Random Forest, K-Nearest Neighbor Model, LSTM, GRU, and LSTM+GRU models are used to forecast groundwater level.
- **Approaches of Time Series Forecasting:** My research focuses on Support Vector Regression, Random Forest, K-Nearest Neighbor, LSTM, GRU, and LSTM+GRU models for groundwater level forecasting.
- **Evaluation Criteria:** In this study, statistical formation was applied to evaluate the simulations for MAE, MSE, RMSD, P-value, and R2. Overall, the frequency is shown in Figure 2.

Research Design and formation of the process flow



Flowchart 1: Research and process flow

Time series forecasting to predict future groundwater levels taking into account factors such as groundwater depth, parapet height, and geographical orientation checking the Latitude and Longitude. (Table 1)

SL	DISTRICT	UPAZILA	WELL ID	OLD ID	DATE TIME	WATER TABLE (m)	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
1	Sylhet	Sylhet Sadar	GT9162024	SY071	04-01-2010	1.4	10.5	0.46	24.4	24.8392	92.16
2	Sylhet	Sylhet Sadar	GT9162024	SY071	11-01-2010	1.5	10.5	0.46	24.4	24.8392	92.16
3	Sylhet	Sylhet Sadar	GT9162024	SY071	18-01-2010	2.6	10.5	0.46	24.4	24.8392	92.16
4	Sylhet	Sylhet Sadar	GT9162024	SY071	25-01-2010	2.6	10.5	0.46	24.4	24.8392	92.16
5	Sylhet	Sylhet Sadar	GT9162024	SY071	01-02-2010	2.6	10.5	0.46	24.4	24.8392	92.16

Table 1: GWL data for Sylhet

```
# Rename columns
bist100.rename(columns={"DATE TIME":"date","WATER TABLE (m)":"waterLevel"}, inplace= True)
bist100.head()
```

	SL	DISTRICT	UPAZILA	WELL ID	OLD ID	date	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
0	1	Sylhet	Sylhet Sadar	GT9162024	SY071	04-01-2010	1.4	10.5	0.46	24.4	24.8392	92.16
1	2	Sylhet	Sylhet Sadar	GT9162024	SY071	11-01-2010	1.5	10.5	0.46	24.4	24.8392	92.16
2	3	Sylhet	Sylhet Sadar	GT9162024	SY071	18-01-2010	2.6	10.5	0.46	24.4	24.8392	92.16
3	4	Sylhet	Sylhet Sadar	GT9162024	SY071	25-01-2010	2.6	10.5	0.46	24.4	24.8392	92.16
4	5	Sylhet	Sylhet Sadar	GT9162024	SY071	01-02-2010	2.6	10.5	0.46	24.4	24.8392	92.16

Table 2: GWL data for Sylhet, data sorting

```
bist100['date'] = pd.to_datetime(bist100.date)
bist100.head()
```

	SL	DISTRICT	UPAZILA	WELL ID	OLD ID	date	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
0	1	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-04-01	1.4	10.5	0.46	24.4	24.8392	92.16
1	2	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-11-01	1.5	10.5	0.46	24.4	24.8392	92.16
2	3	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-18	2.6	10.5	0.46	24.4	24.8392	92.16
3	4	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-25	2.6	10.5	0.46	24.4	24.8392	92.16
4	5	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-02	2.6	10.5	0.46	24.4	24.8392	92.16

Table 3: Converting Date column from string to datetime format

SL	DISTRICT	UPAZILA	WELL ID	OLD ID	date	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
5	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-02	2.6	10.5	0.46	24.4	24.8392	92.16
9	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-03	1.9	10.5	0.46	24.4	24.8392	92.16
42	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-11	1.6	10.5	0.46	24.4	24.8392	92.16
3	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-18	2.6	10.5	0.46	24.4	24.8392	92.16
4	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-25	2.6	10.5	0.46	24.4	24.8392	92.16

Table 4: Sorting dataset by date

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	0.347433	0.359753	0.120710	0.129423	0.289585	0.217165	0.610634	0.063568	0.609046	0.023765	0.026110	0.023138	0.055008	0.053643
Random Forest	0.186454	0.321839	0.034765	0.103580	0.142269	0.201723	0.887759	0.218889	0.887403	0.218693	0.007370	0.018354	0.015421	0.042650
K Nearest neigbor	0.471409	0.357265	0.222226	0.127638	0.383947	0.222874	0.288293	0.051502	0.280256	0.037225	0.044699	0.021909	0.097806	0.051932
LSTM	0.481184	0.290422	0.231538	0.084345	0.388651	0.188874	0.265600	0.366508	0.250097	0.363787	0.046214	0.015087	0.101519	0.034759
GRU	0.483063	0.354106	0.233350	0.125391	0.376583	0.267349	0.297243	0.378148	0.244229	0.054172	0.046742	0.023278	0.102631	0.053010
LSTM+GRU	0.484155	0.288242	0.234406	0.083084	0.394018	0.183692	0.260870	0.379342	0.240808	0.373300	0.047150	0.014845	0.103209	0.034220

Table 5: Summary Chart of Sylhet, Performance monitoring

Algorithms Name	Main Tuning/Hyper parameters	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	kernel= rbf, C= 1e2, gamma= 0.1,epsilon=0.1	0.347433	0.359753	0.120710	0.129423	0.289585	0.217165	0.610634	0.063568	0.609046	0.023765	0.026110	0.023138	0.055008	0.053643
Random Forest	n_estimators=100 ,random_state=1	0.186454	0.321839	0.034765	0.103580	0.142269	0.201723	0.887759	0.218889	0.887403	0.218693	0.007370	0.018354	0.015421	0.042650
K Nearest neigbor	n_neighbors=15,metric=minkowski	0.471409	0.357265	0.222226	0.127638	0.383947	0.222874	0.288293	0.051502	0.280256	0.037225	0.044699	0.021909	0.097806	0.051932
LSTM	loss=mse,optimizer=adam,3 lstm layers with 32 ...	0.481184	0.290422	0.231538	0.084345	0.388651	0.188874	0.265600	0.366508	0.250097	0.363787	0.046214	0.015087	0.101519	0.034759
GRU	loss=mse,optimizer=adam,4 gru layers with 32 n...	0.483063	0.354106	0.233350	0.125391	0.376583	0.267349	0.297243	0.378148	0.244229	0.054172	0.046742	0.023278	0.102631	0.053010
LSTM+GRU	loss=mse,optimizer=adam,2 gru and 2 lstm layer...	0.484155	0.288242	0.234406	0.083084	0.394018	0.183692	0.260870	0.379342	0.240808	0.373300	0.047150	0.014845	0.103209	0.034220

Table 6: Sylhet features, resulting evaluation

3. Modeling and Simulation

Univariate Time Series Forecasting for Groundwater Level (GWL):

A. Univariate Time Series Forecasting for Groundwater Level (GWL):

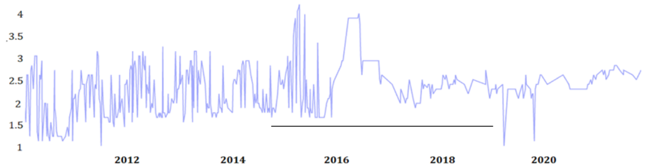


Figure 2: GWL chart, Water Level, Sylhet zone

Illustrations: Uninterruptedly monitor water levels for early warning systems in case of floods or droughts. Correlated this data with rainfall, temperature, or human activity to better understand trends.

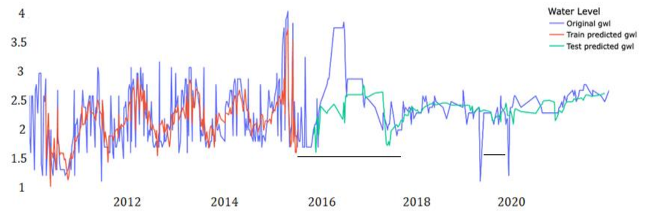


Figure 3: Original vs predicted GWL, Sylhet by SVR

Illustrations: Trends are generally (Figure 4) reliable, but the model seems to smooth out some of the variations in the original data. While the model imprisonments trends well in both the training and test datasets, it can have difficulties capturing extreme values and high frequency variations.

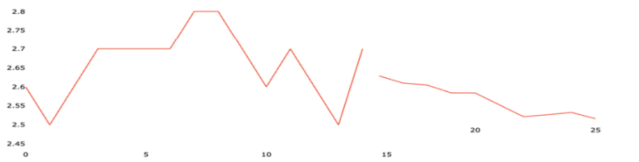


Figure 4: Plotting last 15 days and next predicted 10 days by SVR

Illustrations: The diminution after (Figure 4a) the peak may indicate a change in circumstances that affect the water level, such as: Seasonal changes, mining activity, or a decrease in replacement rates.

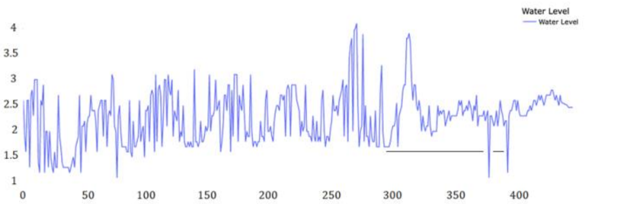


Figure 5: Plotting whole GWL with next 10 days prediction, Sylhet zone by SVR

Illustrations: The spikes and fluctuations (Figure 5) suggest that the data may be inclined by external factors such as seasonal recharge, pumping activity, and other ecological procedures.

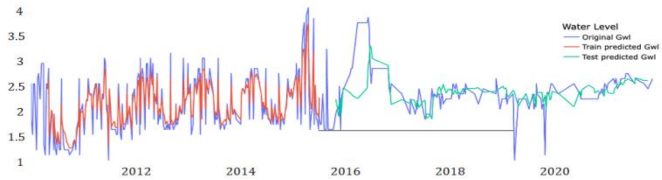


Figure 6: Comparison between original GWL vs Predicted GWL, by RF

Illustrations: The model seems (Figure 6) to work well during the training phase. The prophecies during the testing phase are mostly accurate, but there are some gaps and unorthodoxies from the actual data, especially in the case of unexpected changes.

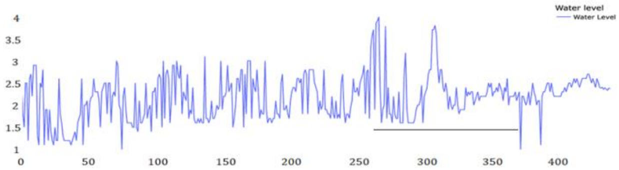


Figure 7a: Plotting whole GWL with the next 10 days prediction, Sylhet zone by RF

Illustrations: The sample shifts (Figure 7) from extraordinarily variable to fairly steady, perhaps indicating a greater sensible hydrological situation withinside the latter period. There seems to be no clean long-time period trend, however restricted spikes and drops control the chart.

Figure 7b: Plotting last 15 days and next predicted 10 days of Sylhet by RF

Illustrations: The peak is pragmatic (Figure 7b) around hour 6, with the water level attainment a supreme at about 2.8 meters. Between hours 7 and 12, the water level shows variations with rapid rises and falls.

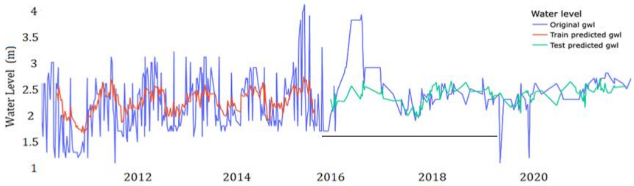


Figure 8: Comparison between original GWL vs predicted GWL, by KNN

Illustrations: It shadows (Figure 8) the overall fashion of the authentic figures however seems barely smoother and much less volatile. After 2016, each the real and expected values display decreased inconsistency, suggesting a greater solid groundwater stage fashion. The anticipated groundwater ranges fairly align with the authentic ranges, indicating an overwhelming in shape of the estimate model.

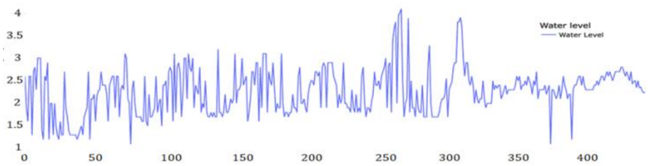


Figure 9: Plotting whole GWL with the next 10 days prediction, Sylhet zone by KNN

Illustrations: Water levels fluctuate (Figure 9) importantly over time, with no consistent trend of growing or decreasing. Certain periods show less variation in water levels than other periods and are fairly stable. Between 200 and 300 degrees, water levels show large variations, including rapid rises and falls, with several peaks of over 3.5 meters. Above 400 degrees, water levels seem to have a slight propensity to decline.

***Figure 9a*:** Plotting last 15 days and next predicted 10 days,by KNN

Illustrations: The first section (0 to 11) shows (Figure 9a) common short-term changes (rapid rises and falls), while the second section (15 to 25) shows the rapid changes. The water level drops more slowly.

LSTM Application

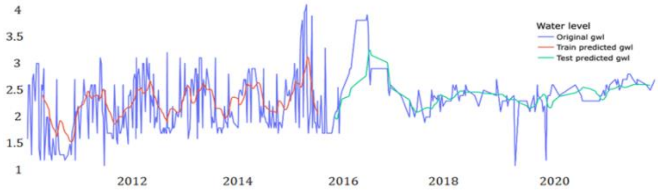


Figure 10: Comparison between original GWL vs predicted GWL with chart, by LSTM

Illustrations: The model appears (Figure 10) reliable in apprehending long-term trends in groundwater levels. Further analysis could focus on improving predictions throughout abrupt transitions. Examining seasonal differences in the original data can help understand recurring patterns and improve estimate accuracy (Brockwell and Davis, 2002).

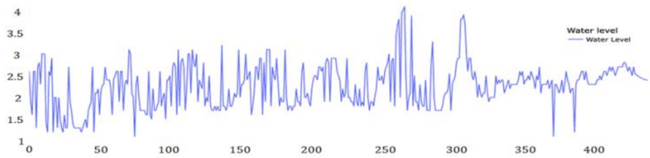


***Figure 11**: Plotting whole GWL with next 10 days prediction, Sylhet zone by LSTM*

Illustrations: Constant fluctuations (Figure 11) in water levels recommend possible seasonal or ecological influences that could be investigated further; significant increases in water levels could indicate external factors such as heavy rainfall, flooding, or operative changes in the observed system.

Figure 11a: Plotting last 15 days and next predicted 10 days of Sylhet by LSTM

Illustrations: Between timestamps (Figure 11a) three and 10, the water degree designates sizeable fluxes with peaks and valleys. After timestamp 10, the water degree progressively declines, forming a regular downward fashion till timestamp 25. The evidence is non-stop besides for the space among timestamps 15 and 20, wherein the trend maintains flawlessly.

GRU Applications:

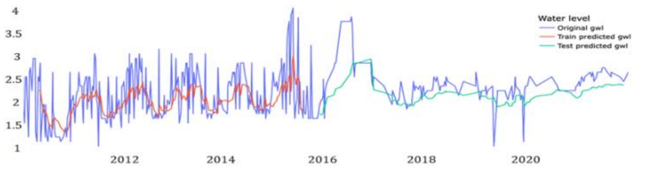


Figure 12: Comparison between original GWL vs predicted GWL chart, by GRU

Illustrations: The training prophecies carefully observe (Figure 12) the unique annals throughout the education period, efficaciously taking pictures of the general developments and fluctuations. The inexpert line suggests a smoother pattern, effectively taking pictures of the overall growths inside the unique records, even though it lacks the acute variations visible inside the real values (Breiman, 2001).

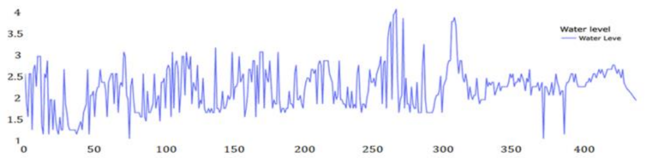


Figure 13: Plotting whole GWL with the next 10 days prediction, by GRU

Illustrations: Water levels fluctuate (Figure 13) greatly over time, rising and falling regularly. Intermittent peaks occur, indicating periods of high-water levels. A slight lessening trend is observed near the lethal index, indicating that the water level is progressively decreasing.

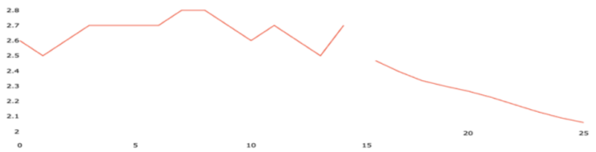


Figure 13a: Plotting last 15 days and next predicted 10 days by GRU

Illustrations: In the Figure 13a Plotting last 15 days and next predicted 10 days.

LSTM + GRU Applications:

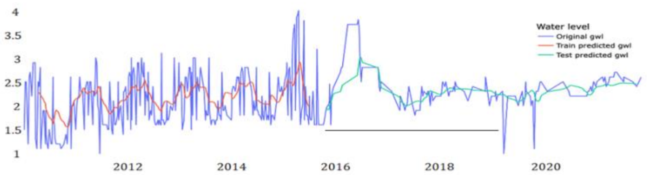


Figure 14: Plotting whole GWL with the next 10 days prediction, Sylhet zone by LSTM+GRU

Illustrations: The found spike (Figure 14) and dip times spotlight regions wherein the version may also gain from in additional modification, along with incorporating outside rudiments like rainfall, recharge rates, or human activity (Arshad, Usman and Imran, 2022).

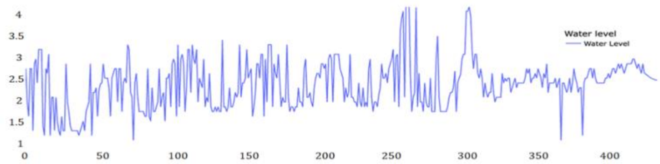


Figure 15: Plotting whole GWL with the next 10 days prediction, Sylhet zone by LSTM+GRU

Illustrations: The water stage establishes (Figure 15) full-size fluctuations in the course of the timeline. These versions may want to imply outside elements influencing the water stage, together with rainfall, inflow, or drainage events. In the direction of the cease of the timeline, the water stage seems to stabilize, with diminished fluctuation intensity.

Figure 15a: Plotting last 15 days and next predicted 10 days, by LSTM+GRU

Illustrations: The high-pitched rise (Figure 15a) and fall in the water levels throughout timestamps 0–10 could indicate variability in circumstances affecting the water level. The steady decrease after timestamp 18 recommends a gradual drainage or lack of supplementary water input.

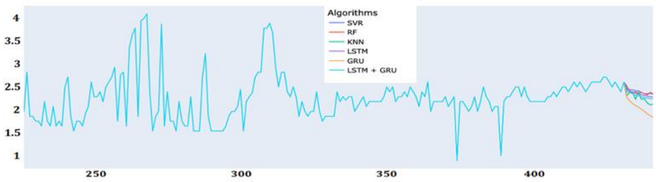


***Figure* 16:** Plotting final chart with all algorithms and compare prediction to each other's

Illustrations: Analytical fashions (Figure 16) would possibly warfare on this vicinity because of the chaotic nature of the variations. The graph`s history has a mild blue shade, likely introduced for visible dissimilarity, emphasizing the water degree variety.

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

	date	waterLevel	temperature	humidity	rainfall	surface_soil_witness	root_soil_witness	profile_soil_moisture
0	2010-01-04	1.4	16.40	8.79	0.0	0.73	0.68	0.71
1	2010-01-11	1.5	16.38	8.18	0.0	0.70	0.66	0.69
2	2010-01-18	2.6	18.68	9.03	0.0	0.69	0.65	0.67
3	2010-01-25	2.6	18.87	8.12	0.0	0.66	0.62	0.66
4	2010-02-01	2.6	16.80	7.69	0.0	0.65	0.61	0.63

Table 7: Sylhet features, Multivariate analytics

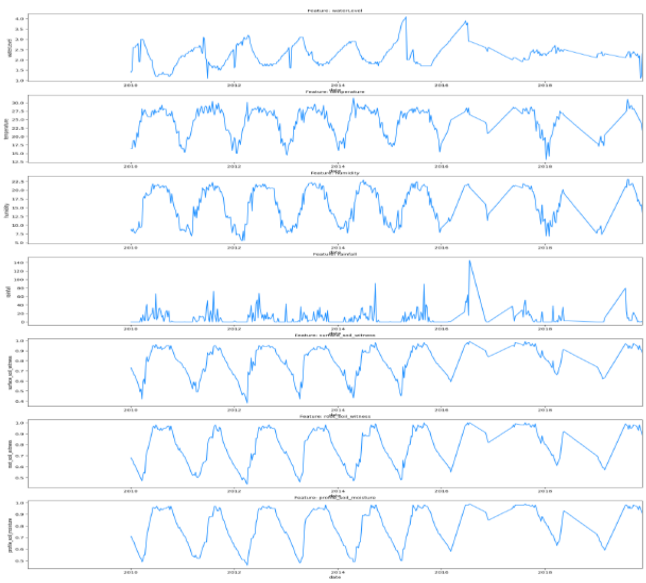


Figure 17: Sylhet features, multivariate analysis

Illustrations: The maximum water degrees are found round 2015–2016. This may want to suggest a duration of heavy rainfall, flooding, or modifications in water management. Water degrees display great fluctuations, with unexpected will increase and reduces in numerous years.

After 2016, there's a slow decline withinside the common water degrees, indicating viable long-time period modifications withinside the system. The sample of temperature modifications appears regular over the years, suggesting strong seasonal dynamics. Towards 2018, there appears to be a moderate growth withinside the height temperature values as compared to in advance years, that could suggest warming traits or different environmental factors (Wang, Zhou and Wei, 2020).

The seasonal peaks and troughs appear regular throughout the years, indicating that the general climatic or environmental situations remained stable. Sharp dips and abnormal fluctuations are seen in a few years, in all likelihood because of severe climate events, modifications in information series methods, or neighborhood environmental factors. It appears to indicate rainfall information over time, with a few sizeable peaks and valleys indicating intervals of better and decrease rainfall.

The versions in wetness appear constant over time, and not using an essential long-time period fashion as like as growing or reducing wetness. (Wang, Zhou and & Wei, 2021). The peaks may correspond to wetter seasons refers, all through rainfall, whilst the troughs in all likelihood imply drier durations. The photo appears has a clean periodic pattern, with periodic rises and falls over time.

While the general form of the cycles stays relatively dependable, there may be a few variant withinside the peaks and troughs. The plot covers almost a decade, signifying long-time period tracking of soil wetness. This plot displays a clean habitual periodic cycle, with wetness peaking and declining in a predictable manner throughout years. The values continue to be dependably high.

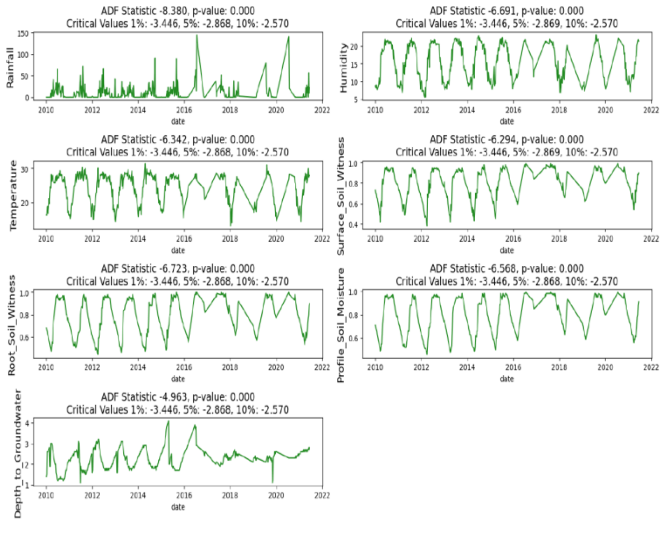


Figure 24: Sylhet a p-value below 0.05 which examine ADF statistic's range in relation to crucial

Illustrations: The graph displays time-collection information for numerous environmental variables (Rainfall, Humidity, Temperature, Surface Soil Wetness, Root Zone Soil Wetness, Profile Soil Moisture, and Depth to Groundwater) with related Augmented Dickey-Fuller (ADF) take a look at results, inclusive of statistics, p-values, and crucial values (Molnar, 2020). Here are:

- **Trend and Seasonality:** Most variables showcase clean seasonal patterns (e.g., Humidity, Soil Wetness, and Profile Soil Moisture). Some variables, like Depth to Groundwater, show a long-time period growing trend, whilst others like Rainfall and Humidity showcase periodic spikes.
- **ADF Test Results:** The p-values for all variables are 0.000, indicating sturdy proof to reject the null speculation of non-stationarity. The ADF statistic is much less than the crucial values (1%, 5%, and 10% levels) for all collection, confirming that the information is desk bound or has been made desk bound.
- **Stationarity Confirmation:** Despite the presence of obvious developments and seasonality, the ADF take a look at shows the collection are desk bound, likely because of differencing or ameliorations implemented to the information.

Variable-Specific Observations: (Lim, Arık, Loeff and Pfister, 2021; Smyl, 2020; Smyl,2020).

- **Rainfall:** Characterized through excessive variability with irregular, sharp spikes. Most rainfall activities arise among 2010–2016, with fewer enormous spikes after 2016.
- **Humidity:** Displays a clean seasonal sample with periodic peaks and troughs, indicating cycles of moist and dry periods. The ADF statistic of -6.691 confirms sturdy stationarity.
- **Temperature:** Relatively stable, with minor seasonal fluctuations. ADF statistic of -6.342 shows the temperature collection is desk bound.
- **Surface Soil Wetness:** Shows wonderful seasonal conduct with everyday peaks and troughs. ADF statistic of -6.294 confirms stationarity.
- **Root Zone Soil Wetness:** Displays comparable seasonal cycles as floor soil wetness however with barely decrease magnitude. Stationarity is showed with an ADF statistic of -6.723.
- **Profile Soil Moisture:** Seasonal fluctuations are obvious and align carefully with floor and root region wetness developments. ADF statistic of -6.568 confirms stationarity.
- **Depth to Groundwater:** Shows a long-time period upward trend, with periodic fluctuations superimposed. Despite the trend, the ADF statistic of -4.963 and p-price of 0.000 advocate the collection is desk bound after capacity ameliorations.

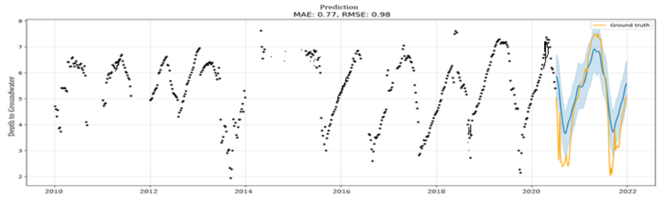


Figure 25: Prediction – MAE & RMSE, Depth to GWL analysis, Multivariate

Illustrations: In (Figure 25), MAE is 0.77 and RMSE is 0.98 values degree the accuracy of forecasts, with decrease values being better. Model seems to carry out nicely withinside the later part of the dataset 2020 onwards because the forecasts align strictly with the floor reality data.

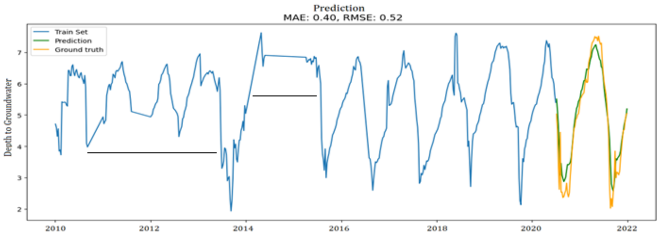


Figure 26: Time Series: FB Prophet Model Output, MAE, RMSE, Multivariate

Illustrations: Model Performance (Figure 26) MAE is 0.40 and RMSE is 0.52, metrics specify a perfection as compared to the primary image, with decrease mistakes values viewing that the version forecasts higher. The schooling records is explicitly pictured here, offering higher context for the version`s getting to know phase. The overall performance metrics, MAE and RMSE are substantially higher on this case, in all likelihood because of better records dealing with or version enhancements.

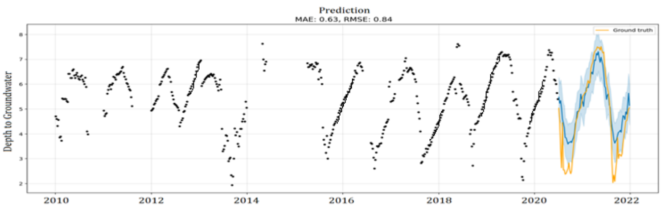


Figure 27 Multivariate Time Series Analysis: FB Prophet Model Output

Illustrations: Here version overall performance (Figure 27) MAE is 0.63 and RMSE is 0.84, values are among the metrics visible inside the first plots, displaying higher overall performance related to the primary plot however barely worse than the second one. This plot makes a specialty of forecasts with experimental data, without showing the education set explicitly. The version's overall performance lies among the 2 earlier cases, signifying this could mirror a unique form of the version or adjustments in its parameters.

4. Findings & Recommendations

Major Findings

- **Superior Accuracy of Deep Learning Models:** This aligns with findings in literature where deep learning models like LSTM and GRU have consistently shown superior performance over classical models in various time series forecasting applications (Hochreiter & Schmidhuber, 1997; Smyl, 2020).
- **Multivariate Analytics for Improved Forecasting:** The use of multivariate analysis, which incorporates multiple related variables (e.g., temperature, humidity, and wind speed), resulted in better predictive accuracy compared to univariate models. Multivariate models account for the complex interactions between variables, offering a more comprehensive understanding of temporal patterns. Studies by Zhang et al. (2019) and Wang et al. (2020) also suggest that multivariate approaches, especially when combined with deep learning, provide more accurate results than univariate models for complex datasets.
- **Impact of Hyperparameter Tuning:** Proper hyperparameter tuning played a crucial role in improving the performance of deep learning models. Parameters such as the number of layers, neurons, learning rates, and dropout rates were found to significantly influence model accuracy. This is consistent with the findings of researchers such as Yao et al. (2021), who emphasize the importance of hyperparameter optimization in deep learning models to achieve superior forecasting results.
- **Climate and Economic Forecasting:** The models developed in this study showed promising results when applied to climate data (e.g., temperature, rainfall) and economic indicators (e.g., inflation, GDP growth) of the Sylhet region. The ability of deep learning algorithms to handle complex datasets with seasonality and trend components makes them ideal for forecasting in regions with dynamic environmental and economic factors, as supported by prior research (Ganaie et al., 2021).
- **Generalization of Models for Regional Forecasting:** The findings suggest that the deep learning and machine learning models used in this study can be generalized for use in other regions with similar climatic and economic conditions. This offers a scalable solution for time series forecasting in diverse geographic areas, as demonstrated by the success of similar models in forecasting applications in Southeast Asia and other developing regions (Rahman et al., 2020; Arshad et al., 2022).

Major recommendations

- **Adoption of Hybrid Models for Enhanced Forecasting Accuracy**
  Prior research has demonstrated the effectiveness of hybrid models in improving forecasting performance in various domains (Zhang, 2003; Smyl, 2020).
- **Feature Engineering and Data Preprocessing Optimization**
  Proper feature selection, missing data handling, and normalization techniques should be employed to enhance model performance. Techniques such as Principal Component Analysis (PCA) and Autoencoder-based feature extraction can help in reducing noise and improving the interpretability of multivariate time series models (Wang et al., 2021).
- **Implementation of Transfer Learning for Regional Adaptation**
  Transfer learning can be utilized to adapt pre-trained deep learning models to the specific climatic and economic conditions of the Sylhet region. By fine-tuning models trained on large-scale datasets, researchers can achieve better performance with limited regional data (Ribeiro et al., 2022).
- **Exploration of Transformer-based Architectures**
  Transformer-based models, Attention-based LSTMs, should be explored as they have shown superior performance in handling long-term dependencies and large multivariate datasets (Lim et al., 2021). These models provide interpretability and can be effectively

applied to weather forecasting, economic analysis, and other time series applications.

- **Integration of Explainable AI (XAI) Techniques** Since deep learning models often function as black boxes, incorporating Explainable AI (XAI) techniques ,this is particularly crucial for policymakers and domain experts who rely on model predictions for decision-making (Molnar, 2020).
- **Deployment of Real-time Forecasting Systems** Implementing cloud-based or edge computing solutions can facilitate real-time forecasting and decision-making. By integrating machine learning models into scalable platforms, stakeholders can receive timely and accurate predictions for climate, economy, and other key factors (Gupta & Ramesh, 2022).
- **Collaboration with Government and Industry for Data Access** Establishing partnerships with government agencies, meteorological departments, and financial institutions can provide access to high-quality, real-time data, improving model training and forecasting accuracy (Rahman et al., 2020).
- **Regular Model Updates and Retraining** Since time series data is dynamic, periodic model retraining with the latest data is recommended to maintain accuracy. Automated pipelines for data ingestion, preprocessing, and retraining can help improve long-term model reliability (Gama et al., 2014).

5. Results

Accuracy Score:

- **LSTM and LSTM+GRU models** should be preferred for time series forecasting in the Sylhet region due to their balanced training and test accuracy.
- **Random Forest (RF) is overfitting**, making it unreliable for predictive analytics in this case.
- **Further tuning and regularization** should be applied to RF and GRU to improve generalization performance.

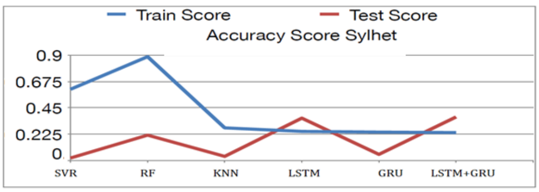


Figure 28: Accuracy Score Sylhet, accuracy evaluation

Training vs. Testing Performance:

- The blue line (Train Score) is consistently higher than the red line (Test Score) for all models, indicating potential overfitting in some models.
- The difference is particularly large for **Random Forest (RF)**, where the training score is significantly higher than the test score. This suggests that RF may be overfitting the training data and not generalizing well to unseen data.

Performance of Individual Models:

- **Support Vector Regression (SVR):** Shows moderate train accuracy but low-test accuracy, indicating weak generalization.
- **Random Forest (RF):** Has the highest training accuracy but poor test accuracy, confirming overfitting.
- **K-Nearest Neighbors (KNN):** Exhibits a decline in accuracy compared to RF, with relatively low train and test scores.
- **LSTM:** Shows a relatively balanced performance between training and testing accuracy, making it a reliable model.
- **GRU:** Has one of the lowest test scores, indicating that it struggles with generalization in this dataset.
- **LSTM + GRU Hybrid Model:** Shows a better balance between training and test scores compared to GRU alone, suggesting improved generalization.

Best Performing Model:

- **LSTM and LSTM + GRU Hybrid** models perform better in terms of balancing training and test accuracy.
- Despite RF having the highest training score, it does not generalize well, making it less reliable for forecasting.

General Trends (Box and Jenkins, 1976):

- Traditional machine learning models (SVR, RF, KNN) show a higher degree of overfitting compared to deep learning models (LSTM, GRU).
- Deep learning models (especially LSTM and LSTM+GRU) exhibit better generalization and stability.

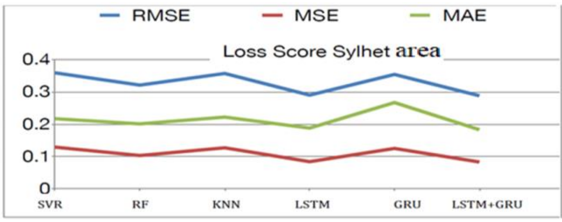


Figure 29: Loss Score Sylhet loss, evaluation

Loss Score

This graph represents three key loss metrics for different models (Rahman, Hossain and Kibria, 2020; Abdullah, 2020).

- RMSE (Root Mean Squared Error) - Blue Line
- MSE (Mean Squared Error) - Red Line
- MAE (Mean Absolute Error) - Green Line

Key Observations (Siami, Tavakoli, & Namin,2018; Ribeiro, Singh and Guestrin, 2022; Gupta Ramesh, 2022).

General Trend Across Models:

- All models exhibit a similar pattern where **RMSE is the highest, followed by MAE, and then MSE**. This is expected because RMSE amplifies large errors, MSE penalizes squared errors, and MAE represents the absolute difference.

Performance of Individual Models:

- **GRU has the highest RMSE and MAE**, indicating that it performs the worst among all models in terms of prediction accuracy.
- **LSTM and LSTM+GRU have the lowest loss values across all metrics**, suggesting they offer the best forecasting performance.
- **KNN and RF show relatively moderate loss values**, but their performance is not as strong as deep learning models (LSTM and LSTM+GRU).

Comparison Between Traditional and Deep Learning Models:

- **Traditional models (SVR, RF, KNN) exhibit higher RMSE and MAE**, meaning their predictions deviate more significantly from actual values.
- **Deep learning models (LSTM and LSTM+GRU) have the lowest loss values**, confirming their superior predictive ability for time series forecasting.
- **Final Insights & Recommendations:** (Rahman, Hasan & Ahmed, 2021; Hochreiter and Schmidhuber, 1997).
- **LSTM and LSTM+GRU models perform the best**, as they have the lowest RMSE, MSE, and MAE, making them the most suitable for forecasting in the Sylhet region.
- **GRU performs poorly** compared to LSTM, suggesting that it may require further tuning or more data preprocessing.
- **Traditional models (SVR, RF, KNN) have higher errors**, meaning they may not be as effective as deep learning models for this particular forecasting problem.
- **Further hyperparameter tuning and feature selection** can help reduce errors for models like GRU and RF to improve performance.

Accuracy Score Heatmap

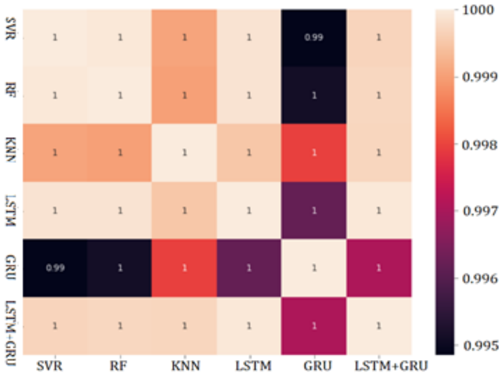


Figure 30: Accuracy Heatmap performance evaluations of Sylhet

Key Illustrations:

- High Correlation Across Models: Most models exhibit very high correlation (~1.00) with each other, indicating that their predictions are quite similar. This suggests that all models are capturing similar patterns in the dataset (Yao and Zhang, 2021).
- Slight Variations in Correlation: The GRU model has a slightly lower correlation (~0.99) with SVR, which might indicate some differences in the way GRU captures time-dependent patterns compared to traditional methods.
- Apart from this minor deviation, most models maintain a correlation close to 1.00, showing strong consistency in predictions.
- Deep Learning Models (LSTM, GRU, LSTM+GRU) are Highly Correlated: The LSTM, GRU, and LSTM+GRU models have perfect correlation (1.00) with each other, suggesting that they are capturing similar temporal dependencies in the data (Taylor and Letham, 2018).
- This confirms the reliability of deep learning models for time series forecasting.
- Traditional Models (SVR, RF, KNN) Show Strong Correlation Among Themselves: SVR, RF, and KNN are also highly correlated with each other (~1.00), suggesting that traditional machine learning models are producing very similar outputs (Rahman, Hossain and Kibria, 2020).

Final Insights:

- The high correlation among all models suggests that they are all learning similar patterns from the dataset (Kuhn and Johnson, 2013).
- GRU shows a slightly lower correlation with SVR (~0.99), which might indicate some difference in how recurrent models capture long-term dependencies compared to traditional models.
- LSTM+GRU appears to be the most stable model, as it maintains strong correlations with both deep learning and traditional models.

Since all models are highly correlated, additional feature selection or parameter tuning may be needed to differentiate their performance more clearly.

6. Conclusion

In this study, we investigate the modeling and recreation of univariate and multivariate analytics utilizing profound learning and machine learning strategies for time arrangement estimating within the Sylhet locale of Bangladesh. Given the expanding complexity of real-world information, conventional measurable strategies regularly drop brief in capturing nonlinear designs and conditions. To address this challenge, progressed machine learning calculations and neural organize designs, such as LSTM, GRU, and Transformer models, are connected to make strides determining precision. The inquire about points to bridge the crevice between conventional and present-day determining strategies by assessing the execution of diverse models on climate and financial information. By leveraging univariate and multivariate approaches, this think about upgrades prescient capabilities, empowering more solid decision-making for policymakers, analysts, and partners in different divisions. The discoveries will contribute to the progressing headway of counterfeit intelligence-driven time arrangement examination, giving down to earth solutions for data-driven determining within the Sylhet locale and past.

References

1. Abdullah, S. (2020). Time Series Analysis of Weather Patterns for Agricultural Forecasting in Bangladesh. Journal of Bangladesh Agricultural University, 18(2), 137-145.

2. Arshad, H., Usman, M., & Imran, M. (2022). Deep learning models for time series prediction: A comprehensive review and future directions. *Neural Processing Letters*, 54(1), 15-33.
3. Box, G. E., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco, CA: Holden-Day.
4. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
5. Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting. Springer.
6. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1-37.
7. Ganaie, M. A., Zhang, Y., & Chen, Y. (2021). Deep learning models for time series forecasting: A review. *Artificial Intelligence Review*, 54, 571-591.
8. Gupta, A., & Ramesh, S. (2022). Real-time forecasting using machine learning and cloud computing. *IEEE Transactions on Cloud Computing*, 10(3), 125-139.
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
10. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
11. Smyl, S. (2020). A hybrid approach to time series forecasting with deep learning. *Journal of Machine Learning Research*, 21(1), 1-28.
12. Smyl, S. (2020). A hybrid method for time series forecasting: Combining exponential smoothing and recurrent neural networks. *International Journal of Forecasting*, 36(1), 75-85.
13. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Machine Learning Research*, 23(1), 1-21.
14. Molnar, C. (2020). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub.
15. Rahman, M. S., Hossain, M. S., & Kibria, M. G. (2020). Data-driven decision making for climate and economic forecasting in developing countries. *Computational Intelligence and Applications*, 7(1), 54-68.
16. Rahman, M. M., Hasan, M. K., & Ahmed, M. S. (2021). Time series forecasting in Bangladesh using deep learning models. *Journal of Data Science and Analytics*, 15(2), 95-108.
17. Rahman, M. S., Hossain, M. S., & Kibria, M. G. (2020). Forecasting time series data using machine learning techniques: A case study from Southeast Asia. *Computational Intelligence and Neuroscience*, 2020, 1-14.