iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: I Month of publication: January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66099

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Exploring Machine Learning and Deep Learning Techniques for Univariate and Multivariate Time Series Forecasting in the Area of Barisal in Bangladesh

Rakib Uddin
Department of Electrical and Computer Engineering (ECE), ID: 2035499050, 5103748207.
North South University, Plot # 15, Block # B, Bashundhara, Dhaka – 1229, Bangladesh.

Abstract: Bangladesh, a densely populated country in South Asia, faces substantial challenges in water resources management due to its geographic location, hydrological features, and socio-economic aspects. Groundwater plays a dynamic role in meeting the water needs of both rural and urban population, accounting for approximately 85% of the total drinking water supply and a significant portion of irrigation water. With rapid population growing, urbanization, and industrialization, the demand for groundwater is constantly growing, increasing pressure on existing water resources and necessitating effective managing approaches. The consequence of this study lies in its potential to improve the correctness and trustworthiness of groundwater level forecasts, thereby supporting more conversant decision-making in water resources management, urban planning, and structure development.

By using advanced machine learning and deep learning performances, this research challenges to fill the gaps that exist in traditional predicting methods and contribute to the justifiable use of groundwater resources in Bangladesh.

Key words: Groundwater Level, Predictions, Forecasting, Climate Factors, Deep Learning, Machine Learning, LSTM, GRU, SVR, Multivariate Time Series, Barisal City, Bangladesh.

I. INTRODUCTION

Barisal is a Divisional area in southern Bangladesh, is heavily dependent on groundwater resources due to agricultural activities and a growing population. Groundwater levels in the area are prejudiced by a complex interplay of climatic factors (rainfall, temperature, and humidity) and soil properties (surface, root zone, and soil moisture). Traditional numerical methods for groundwater level predicting often fail to capture the nonlinear and dynamic relationships between the variables, imposing the study of advanced machine learning and deep learning performances (Adadi and Berrada, 2018). This study explores the application of various state-of-the-art machine learning and deep learning models, such as long short-term memory (LSTM), gated recurrent units (GRU), random forest (RF), K-nearest neighbors (KNN), support vector regression (SVR), and hybrid models such as LSTM+GRU, to forecast groundwater levels and other relevant strictures in Barisal. The use of advanced performances such as LSTM and GRU, known for their ability to process time series data, aims to capture the temporal reliance of groundwater fluctuations (Haurie and Ghaffari, 2022).

At the same time, machine learning models such as RF, ANN, and SVR offer corresponding strengths in processing multidimensional data and categorizing nonlinear patterns. Combining the methods aims to make available robust and reliable forecasts. The results of this research are anticipated to contribute to better planning and management of groundwater resources in Barisal and may be applicable in other regions with similar tests.

II. METHODOLOGY

This research aims to explore the potential of Machine Learning (ML) and Deep Learning (DL) practices in predicting groundwater level (GWL) and accompanying ecological parameters such as rainfall, humidity, temperature, surface soil moisture, root zone soil moisture, and profile soil moisture in the Bangladesh division explore Barisal. The methodology shadows a structured approach that combines data collection, pre-processing, model expansion, training, evaluation, and performance evaluation.

1) Study Area

The map of Groundwater Wells in Barisal City (Well ID-GT0651005) is located in Barisal Sadar Upazila, between longitude 90.3639 and latitude 22.7111 in southern Bangladesh. The minimum elevation of the groundwater table in Barisal city ranges from 0 m to 2.5 m. Data was collected from 04 January 2010 to 27 December 2021.

2) Data Collection

All datasets are collected for a specific time period to ensure sufficient data for training and testing the time series models. Data collected from the sources of:

- Groundwater Level Data: Monthly or daily records of GWL obtained from Bangladesh Water Development Board (BWDB).
- Rainfall, temperature, and humidity data, soil moisture data, soil moisture measurements at the surface, root zone, and profile collected by NASA.

3) Data Preprocessing

- Data Cleaning: Missing values are handled using appropriate imputation techniques
- Normalization: Continuous variables, such as temperature, rainfall, and soil moisture, are normalized to ensure uniform scaling for machine learning models.
- Feature Engineering:
  - Time-lagged variables are created to account for temporal dependencies in groundwater levels.
  - Derived features, such as cumulative rainfall or average humidity, are included to enhance model performance.
- Training and Test Split: The dataset is divided into training (65%), Testing and Validation (35%) to prevent data leakage.

4) Machine Learning and Deep Learning Models

In this research employs the following models for prediction tasks:

- Long Short-Term Memory (LSTM): A recurrent neural network (RNN) variant designed to capture long-term dependencies in time-series data.
- Gated Recurrent Unit (GRU): Another RNN-based model similar to LSTM but with a simplified architecture, making it computationally efficient.
- LSTM+GRU Hybrid Model: A hybrid approach combining the strengths of both LSTM and GRU for enhanced prediction accuracy.
- Random Forest (RF): An ensemble learning method based on decision trees, well-suited for handling non-linear relationships and multi-dimensional data.
- K-Nearest Neighbors (KNN): A simple, distance-based learning algorithm used as a baseline for comparison.
- Support Vector Regression (SVR): A regression-based model designed to minimize prediction errors and generalize well for small datasets.

5) Model Training

- Training Approach:
  - Deep learning models (LSTM, GRU, LSTM+GRU) are trained using backpropagation through time (BPTT) with the Adam optimizer.
  - Machine learning models (RF, KNN, SVR) are trained using grid search and cross-validation to tune hyperparameters.
- Evaluation Metrics: Models are assessed using standard system of measurement such as:
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
  - Coefficient of Determination (R^2)
  - Mean Absolute Percentage Error (MAPE)

6) Performance Evaluation

The performance of the models is associated based on the valuation metrics to identify the best-performing performance for groundwater level forecast. Moreover:

- The feature status from RF is analyzed to interpret the impact of climatic and soil variables.
- Deep learning models are associated to assess their ability to capture temporal shapes.

7) Execution Tools

- Programming Languages: Python and R are used for data preprocessing, modeling, and investigation.
- Libraries and Frameworks: TensorFlow, Keras and Pandas are employed for executing ML/DL models and directing data breakdown.
- Hardware: Models are trained on high-performance figuring systems with GPU support for deep learning errands.

8) Authentication and Testing

The trained models are authenticated on the validation dataset to fine-tune hyperparameters. Final testing is accomplished on the test dataset to assess the generalization performance of each model.

9) Results Breakdown

- The projected GWL and associated variables are associated with experiential data to assess accuracy.
- Temporal trends, periodic patterns, and glitches in groundwater dynamics are evaluated.
- References are made based on the findings to aid water resource managing in the Barisal division.

A. Accuracy Score Documentation

The accuracy of the models is resolute using significant metrics, tailored for regression complications:

1) Model-Specific Exactness Assessment

For each ML and DL applied, the above metrics are computed on the **test dataset** to assess predictive performance:

1. LSTM:
   - LSTM models are considered to capture temporal dependances. The model's performance is estimated based on how well it forecasts the time-series shapes in GWL and related strictures.
   - Metrics: RMSE, R^2, MAPE.
2. GRU:
   - Similar to LSTM, GRU models are verified for their ability to handle long-term dependances in time-series data.
   - Metrics: RMSE, R^2, MAPE.
3. LSTM + GRU (Hybrid):
   - This hybrid model combines LSTM and GRU layers for boosted temporal feature extraction. Its exactness is associated against impartial LSTM and GRU models.
   - Metrics: RMSE, R^2, MAPE.
4. Random Forest (RF):
   - The performance of RF is assessed on its ability to model non-linear relations and detect feature importance.
   - Metrics: MAE, RMSE, R^2.
5. K-Nearest Neighbors (KNN):
   - KNN obliges as a baseline model. Its performance is linked with more advanced models.
   - Metrics: MAE, RMSE, R^2.
6. Support Vector Regression (SVR):
   - SVR's exactness is verified for its strength on small datasets and its ability to simplify forecasts.
   - Metrics: MAE, RMSE, MAPE.

2) Interpretation of Consequences

- The model with the lowest RMSE and MAPE and highest R^2 is measured the most precise.
- If unlike models excel in diverse metrics further study is conducted to choose the most suitable model based on the study objectives (Collenteur, Bakker, Caljé , Schaars and Klop ,2019).

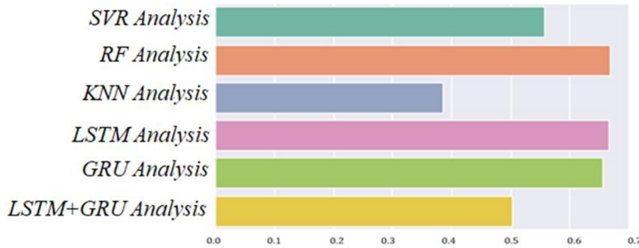


Figure 1: Accuracy Score of the Applied Algorithms

Illustrations: Figure 1 displays the best test values with RF analysis close to 0.7. LSTM analysis and GRU analytical applications perform correctly, while the lowest value is KNN analysis that is adjacent to 0.4.

3) Split data for Training and Testing

To efficiently split data into training and testing sets for exploring ML and DL performances in expecting GWL and related ecological parameters, the following steps and reflections should be addressed:

1. Data Splitting Indication: Splitting data is critical for confirming that models are trained on one portion of the dataset and tested on invisible data to assess their oversimplification aptitude.
2. Time-Series Data Splitting for LSTM, GRU, and Hybrid Models: Since GWL and connected parameters contain time-series data, it is indispensable to maintain the temporal order of observations when splitting the dataset.
3. Random Splitting for Non-Sequential Models like RF, KNN, SVR: For ML models like Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Regression (SVR), the dataset can be split casually into training and testing subsets. This process adopts that the data points are independent and do not have strict temporal dependances.
4. Cross-Validation for Robust Evaluation: To progress robustness and reduce bias, k-fold cross-validation can be applied. This is mostly helpful for ML models where the training and testing sets are alternated across folds to authenticate the results (Vaswani, et al., 2017).
5. Implementation Steps
    Step 1: Load the Data: Authorize the dataset contains all relevant structures.
    Step 2: Define Features and Target Variable: Separate the independent variables (features) and dependent variable (target).
    Step 3: Chronological Split for LSTM/GRU: For LSTM, GRU, or any hybrid time-series model, split the data based on time.
    Step 4: Random Split for Machine Learning Models for RF, KNN, SVR: For non-sequential models, use train_test_split from Scikit-learn to generate randomized splitting.
    Step 5: Grading Structures for SVR, LSTM, GRU: Some models require standardized data for well performance. Use MinMaxScaler or StandardScaler.
6. Time-Series Windowing for Chronological Models: For LSTM and GRU, create sliding windows of input orders for time-series predicting.
7. Data Summary After Splitting: Print the magnitudes of training and testing sets to verify the splits.
8. Cross-Validation for ML Models: For models like RF, KNN, and SVR, apply k-fold cross-validation to authenticate the results across numerous folds.
9. Data Split Ratios: Training Set and Testing Set for time-series predicting, ensure consecutive assembling is maintained during the split.

10. Key detailed credentials
- Use sequential splitting for LSTM, GRU, and time-series models to uphold temporal reliability.
- For ML models like RF, KNN, and SVR, random splitting is adequate, but cross-validation confirms robustness.
- Proper feature clambering is critical for models sensitive to the extent of data
- Continuously test model act on unseen data to assess simplification capability.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

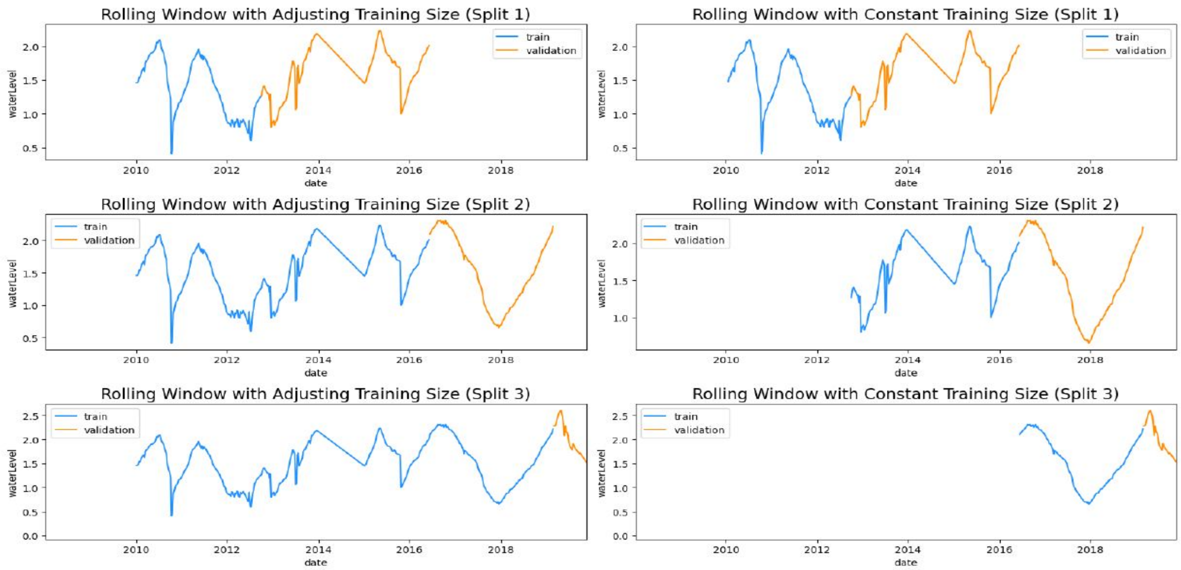
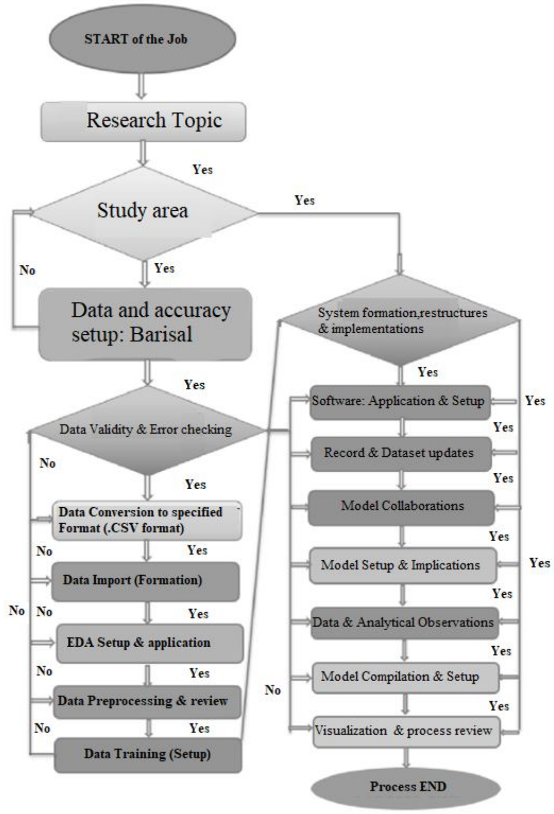


Figure 2: Barisal Rolling window

Illustrations: The image as long as shows sliding window splitting of time series data with disparity and constant training size across three unalike splits.

1. Rolling Window with Training Size adjustment (Left Panel):
   - Training Set Growth: With each subsequent split, the training set size increases, incorporating more past data for in the training over time. This strategy is useful for capturing long-term trends and patterns in the data, especially in time series estimating tasks.
     - Validation Set Constancy: The validation set size remains constant, confirming consistent valuation across splits.
     - Performance Across Splits: The authentication set generally shows variations in the target variable, which specifies that model performance may vary depending on the accessibility of historic data (Tang and Zhang, 2021).
2. Rolling Window with Continual Training Size (Right Panels):
     - Training Set Size Constancy: The training set size remains static in all splits, which confirms uniform input size for model training. This method is suitable for scenarios where older data may no extended be relevant, and only recent patterns are significant.
     - Validation Set Features: Similar to the altering training size method, the authentication set is constant across splits, aiding a direct comparison of validation performance.
     - Potential Trade-Offs: While a constant training size avoids overly bulky datasets, it might miss long-term dependances if the training set eliminates older but appropriate data.
3. General Trends Crossways Splits:
   - Changes in Performance: The relationship between training and validation data reflects the variability in the time series. Any discrepancy between the two indicates potential challenges in generalizing the model or large differences in data properties over time.
   - Increasing Validation Variations: For both approaches, the validation set appears to have more variability in later splits. This suggests that it may be more difficult for the model to maintain predictive accuracy as it moves into unknown future data (Wunsch, Liesch and Broda,2018).
4. Practical Consequences:
     - Adjusting Training Dimensions: Best matched for long-term predicting where capturing growing historic trends is important. May result in higher computational cost as the dataset grows with each split.
   - Constant Training Size:
     - Useful for short-term predicting, particularly in rapidly varying systems where older data might no longer be applicable. Helps sustain a consistent computational load.

4) Model Setup for data Gathering and Research Development (flowchart:1):



Flowchart 1: Process Flow of Different Steps

B. Research Plan and formation of the method evaluation

This study efforts on exploring machine learning (ML) and deep learning (DL) techniques for predicting groundwater level, rainfall, humidity, temperature, surface soil moisture, root soil moisture, and profile soil moisture in Barisal Division, Bangladesh.

1) Research Objectives

- The main objective of this study is to develop and evaluate ML and DL models including LSTM, GRU, RF, KNN, LSTM+GRU, and SVR for accurately predicting environmental and groundwater parameters. This includes:
- Use of historical environmental and soil data.
- Understand the temporal and spatial dependencies of these parameters.
- Compare the performance of the models to determine the most appropriate approach.

2) Research Area and Data Collection

Barisal Division, that is placed in the southern region of Bangladesh, Barisal faces water resource challenges due to its low-lying natural features and high dependency on groundwater for agriculture and human consumption (Gharbi and Bouaziz, 2023).

Data Sources

- Groundwater Level Data: Government databases BWDB)
- Rainfall, Humidity, Temperature: NASA
- Soil Moisture Data, Surface, root, and profile soil data obtained through satellite data or field sensors: NASA

Data Features

- Time-Series Data: All features are characterized as chronological temporal data.
- Spatial Data: Area variability within Barisal Division is considered.
- Data Period: Focus on data to capture historic trends.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

3) Research Questions

The study questions are:

- Can ML and DL models precisely forecast GWL and allied parameters in the Barisal division?
- Which models perform best in terms of correctness, robustness, and simplification?
- How do temporal and ecological factors impact the forecast correctness?
- What role does feature engineering and data preprocessing play in enlightening forecast consequences.

Comparative Analysis

- Compare model performances crossways system of measurement.
- Evaluate computational cost, scalability, and ease of execution for each model.
- Highlight the best-performing models and converse because they outperform others.

4) Process Review

Challenges Encountered

- Data Gaps and Contradictions: Missing values or inadequate data for ecological constraints.
- Model Overfitting: Addressed through regularization, dropout, and cross-validation.
- Temporal Dependances: Proper supervision of time-series data is critical for LSTM, GRU, and hybrid models (Wani, Wani and M.H.,2021).

Solutions Executed (Peterson and Western, 2018)

- Imputation methods and feature engineering addressed data quality matters.
- Scaling methods enhanced model merging.
- Proportional evaluation highlighted strengths and restrictions of each model.

5) Tools and Technologies

- Programming Language: Python
- Visualization Tools: Matplotlib
- Hardware Requirements: High-performance GPU for training DL and ML models.

6) Deliverables

- Trained models with forecasts for GWL and related strictures.
- Comprehensive performance assessment report for each model.
- Feature importance breakdown and references for GWL managing.
- Code repository with step-by-step execution for duplicability.

7) Ethical and Applied Reflections

- Confirm data confidentiality and proper documentations and acknowledgment.
- Authenticate model forecasts against real-world observations to develop reliability.

The following figure depicts the importing dataset of Barisal City from .csv file (in 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	Barisal	Barisal Sadar	GT0651005	BA001	04-01-2010	1.46	4.26	0.79	26.46	22.7111	90.3639
2	Barisal	Barisal Sadar	GT0651005	BA001	11-01-2010	1.46	4.26	0.79	26.46	22.7111	90.3639
3	Barisal	Barisal Sadar	GT0651005	BA001	18-01-2010	1.48	4.26	0.79	26.46	22.7111	90.3639
4	Barisal	Barisal Sadar	GT0651005	BA001	25-01-2010	1.54	4.26	0.79	26.46	22.7111	90.3639
5	Barisal	Barisal Sadar	GT0651005	BA001	01-02-2010	1.54	4.26	0.79	26.46	22.7111	90.3639

Table 1: GWL data for Barisal

```
# 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	Barisal	Barisal Sadar	GT0651005	BA001	04-01-2010	1.46	4.26	0.79	26.46	22.7111	90.3639
1	2	Barisal	Barisal Sadar	GT0651005	BA001	11-01-2010	1.46	4.26	0.79	26.46	22.7111	90.3639
2	3	Barisal	Barisal Sadar	GT0651005	BA001	18-01-2010	1.48	4.26	0.79	26.46	22.7111	90.3639
3	4	Barisal	Barisal Sadar	GT0651005	BA001	25-01-2010	1.54	4.26	0.79	26.46	22.7111	90.3639
4	5	Barisal	Barisal Sadar	GT0651005	BA001	01-02-2010	1.54	4.26	0.79	26.46	22.7111	90.3639

Table 2: GWL data for Barisal, Rename Column

Converting Date column from string to datetime format in the below table 3:

```
[ ] # convert date field from string to Date format and make it index
    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	Barisal	Barisal Sadar	GT0651005	BA001	2010-04-01	1.46	4.26	0.79	26.46	22.7111	90.3639
1	2	Barisal	Barisal Sadar	GT0651005	BA001	2010-11-01	1.46	4.26	0.79	26.46	22.7111	90.3639
2	3	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-18	1.48	4.26	0.79	26.46	22.7111	90.3639
3	4	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-25	1.54	4.26	0.79	26.46	22.7111	90.3639
4	5	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-02	1.54	4.26	0.79	26.46	22.7111	90.3639

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
4	5	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-02	1.54	4.26	0.79	26.46	22.7111	90.3639
8	9	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-03	1.68	4.26	0.79	26.46	22.7111	90.3639
43	44	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-11	0.92	4.26	0.79	26.46	22.7111	90.3639
2	3	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-18	1.48	4.26	0.79	26.46	22.7111	90.3639
3	4	Barisal	Barisal Sadar	GT0651005	BA001	2010-01-25	1.54	4.26	0.79	26.46	22.7111	90.3639

Table 4: Sorting dataset by date, Barisal Sadar

III. MODELING AND SIMULATION

A. Plotting GWL chart: The below plot (in figure) shows the actual ground water level of Barisal.

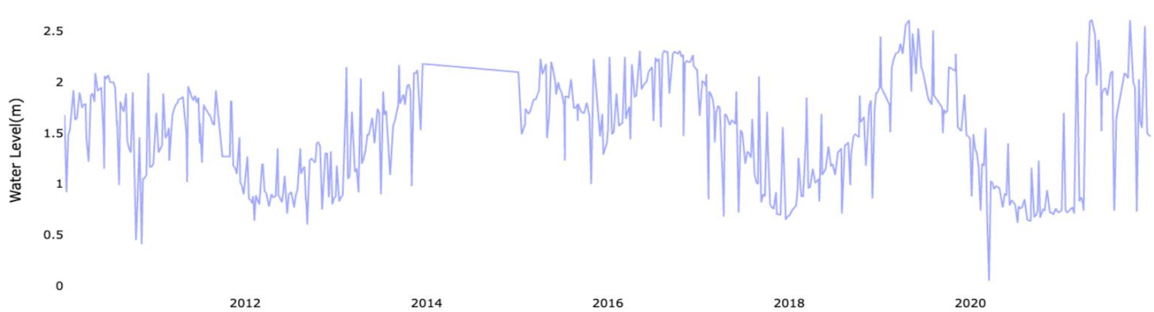


Figure 3: GWL chart for Barisal

Illustrations: The graph (Figure 3) shows a good picturing of the dynamic nature of GWL, showing both periodic patterns and year-to-year variations. (Mojid, Parvez, Mainuddin, and Hodgson, 2019).

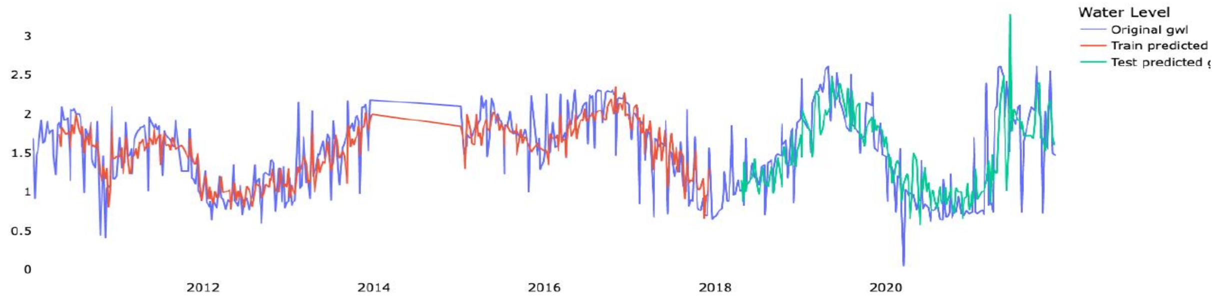


Figure 4: Original Vs predicted GWL of Barisal by SVR

Illustrations: The model compare how well the SVR model executes in predicting the GWL for both the training and testing stages of learning (Figure 4).

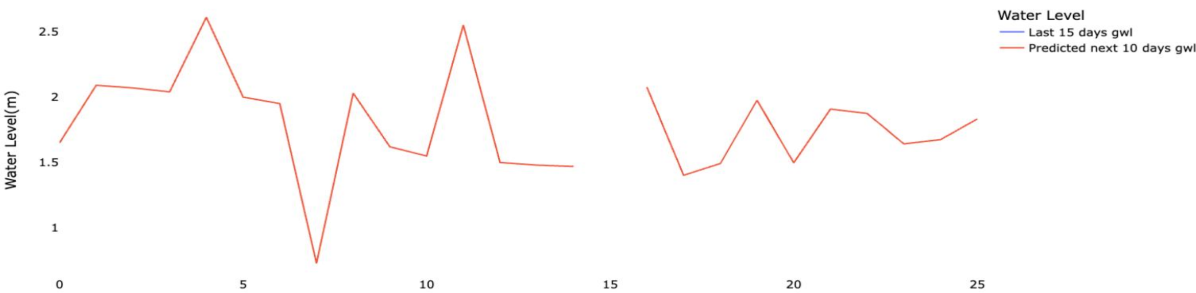


Figure 5: Plotting last 15 days and next predicted 10 days by SVR of Barisal

Illustrations: In Figure 5 demonstrates the GWL forecast Associating last and after as proper. Values are known 15 days vs 10 days forecasts.

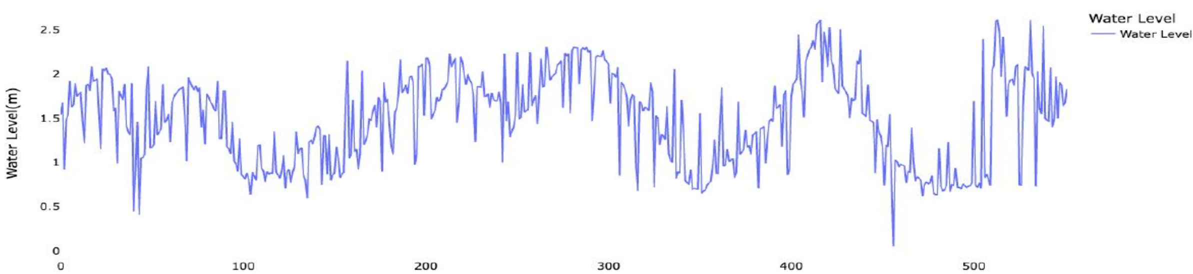


Figure 6: Plotting whole GWL with next 10 days forecast of Barisal by SVR

Illustrations: In the Figure 6, algorithmic issues of the 10 days forecast followed the flow of curves contrast to the Timestamp that 100,300,350 (down) and after 400 to 500 are the observing trends.

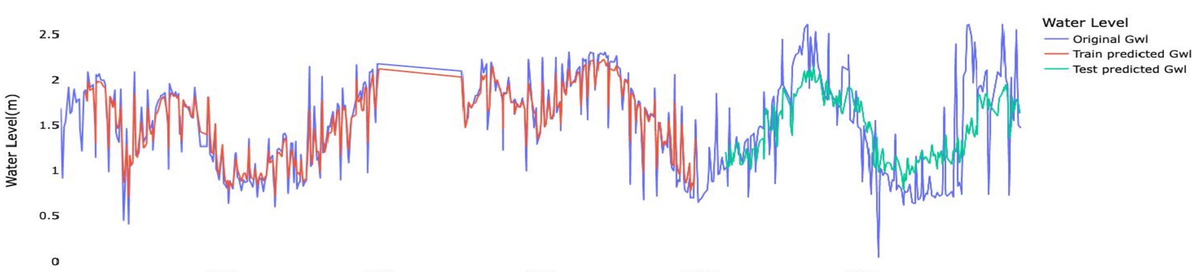


Figure 7: Comparison between original GWL vs predicted GWL with chart of Barisal by RF

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Illustrations: The combination of GWL vs Train projected GWL vs Test predicted GWL directing the periodical flow for which viewpoints on the 2010 to 2018 where 2010 to 2012 make available down and the 2016 to 2018 shows upper trends and the remaining follow the applications.

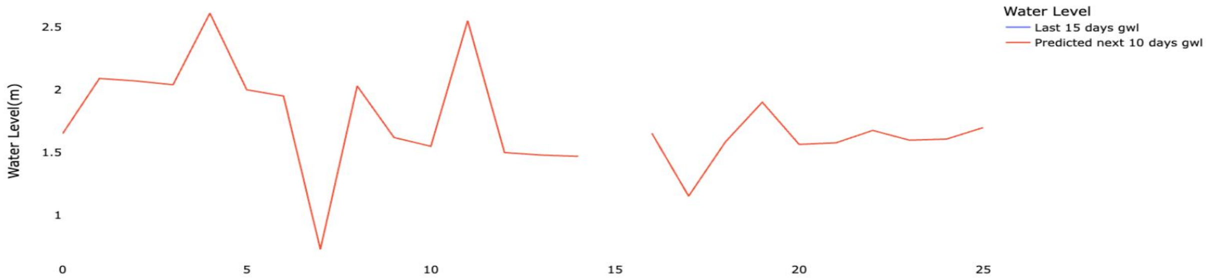


Figure 8: Plotting last 15 days and next predicted 10 days of Barisal by RF

Illustrations: In the figure 8, plotting the last 15 days and the next 10 days predicting linking to the GWL Timestamp periodic that validates the variations of the shape.

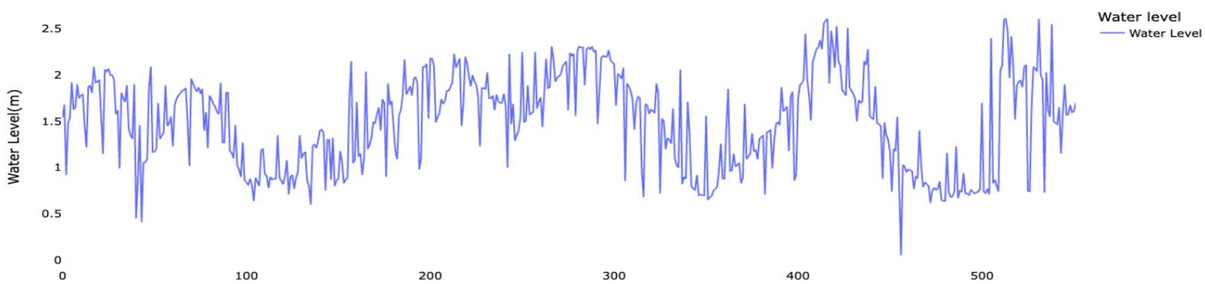


Figure 9: Plotting whole GWL with next 10 days prediction of Barisal by RF

Illustrations: In the figure 9, patterns the predicting GWL stipulations in view of the Timestamp. The curve is markable at 100, 250,350,450 and the upper 400 to 500 terminologies.

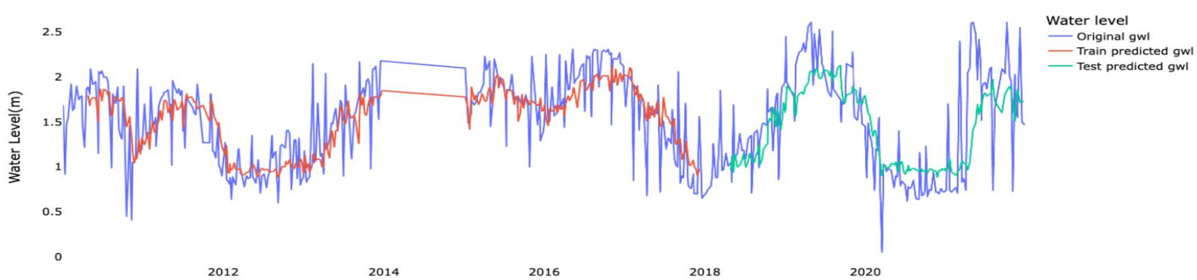


Figure 10: Comparison between original GWL vs predicted GWL with chart of Barisal by KNN

Illustrations: The test forecast GWL (Figure 10) from 2016 headlong and also shadows the general pattern of the main GWL, but with less accuracy allied to the training dated (Simoni, Manoli and G.,2021).

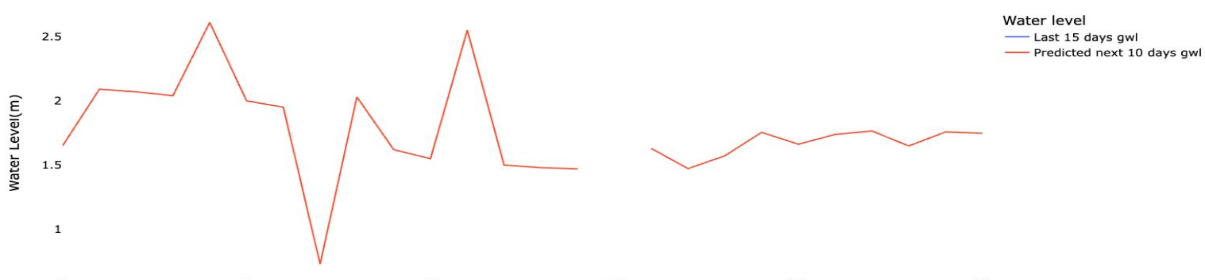


Figure 11: Plotting last 15 days and next predicted 10 days of Barisal by KNN

Illustrations: The Figure 11 detects the GWL predicting for the next 10 days and comparing with the traditional movements (Obergfell, Bakker and Maas, 2019).

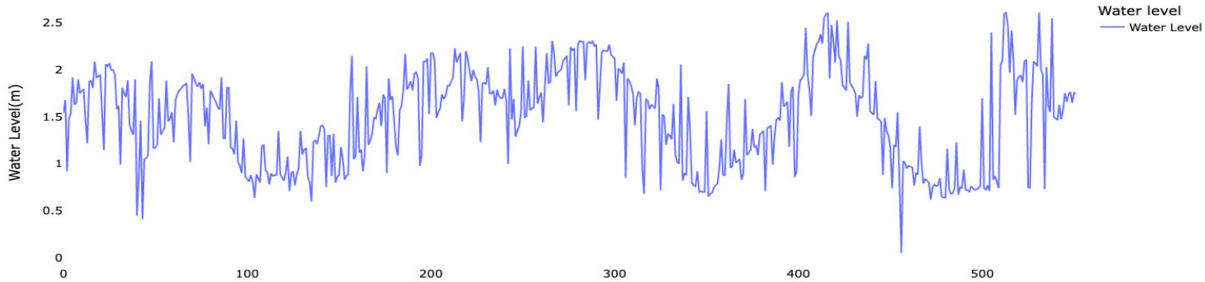


Figure 12: Plotting whole GWL with next 10 days prediction of Barisal by KNN

Illustrations: The Figure 12 detects the plotting of the next 10 days forecast as appropriate by the assigned dataset (Table 1-4)

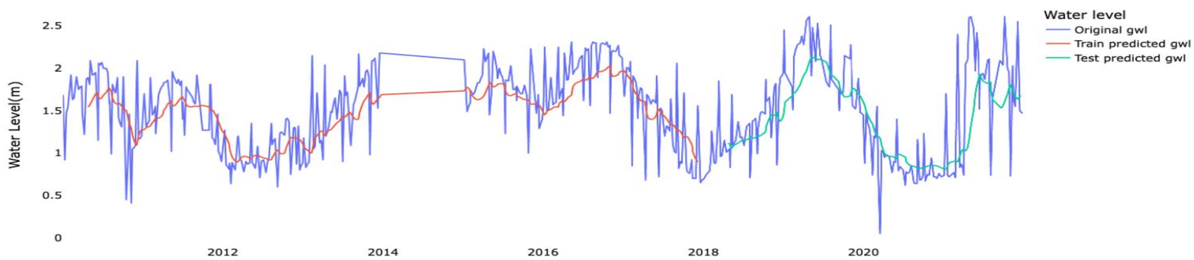


Figure 13: Comparison between original GWL vs predicted GWL with chart of Barisal by LSTM

Illustrations: The LSTM model validates (Figure 13) to capture the overall interrupted patterns and trends in the GWL. The graph recommends the LSTM model apparatuses rationally well, though it might struggle with apprehending some of the more extreme variations (Sun and Wang, 2021).

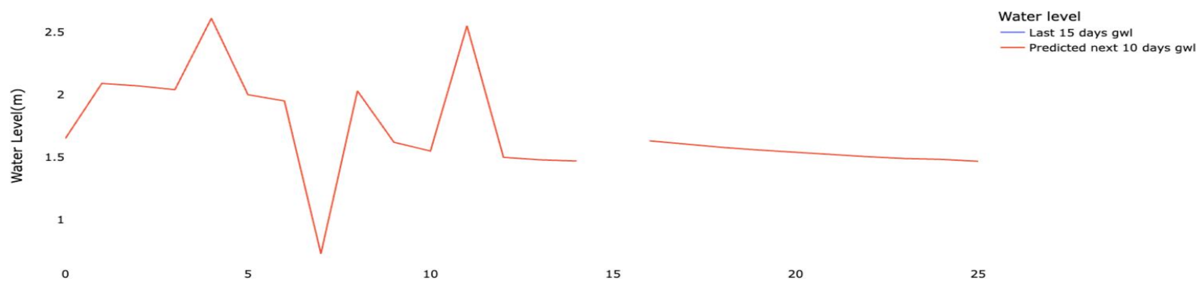


Figure 14: Plotting last 15 days and next predicted 10 days of Barisal by LSTM

Illustrations: The prophesied applications of LSTM accomplish (Figure 14) the next 10 days and compare the revive of the preceding 15 days variations. (Shi, Chen, Wang, et al. (2015).

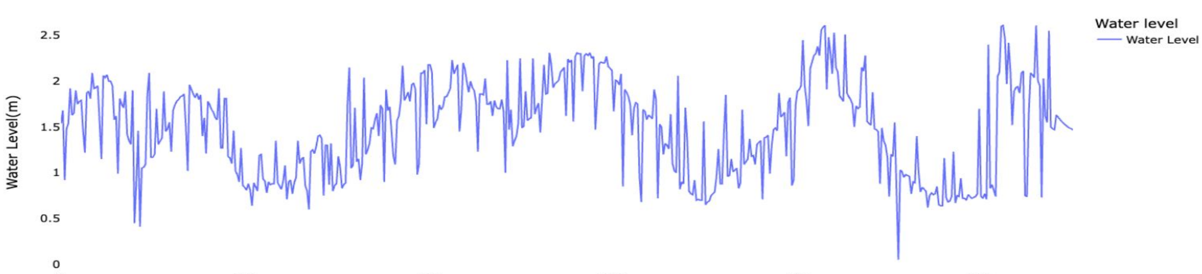


Figure 15: Plotting whole GWL with next 10 days prediction of Barisal by LSTM

Illustrations: In the Figure 15, shape the next 10 days estimating intrigue the whole GWL with the apt Timestamp.

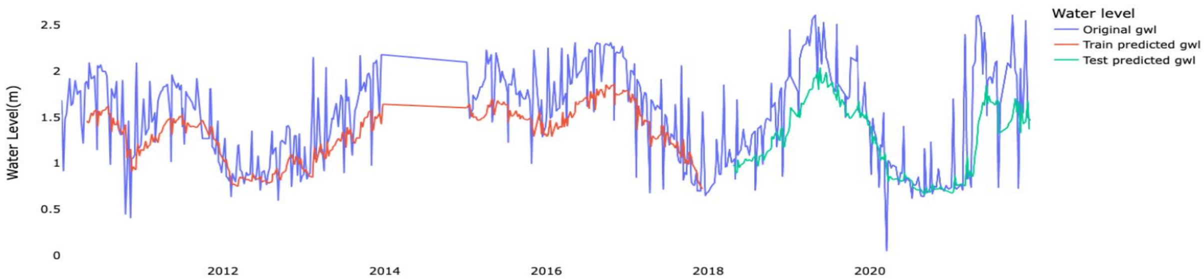


Figure 16: Comparison between original GWL price vs predicted GWL with chart of Barisal by GRU

Illustrations: The application of GWL (Figure 16) comparability the main GWL vs Train forecast GWL vs Test prophesied GWL with the demand of GRU Algorithm. (Lundberg, S. M., and Lee, 2017).

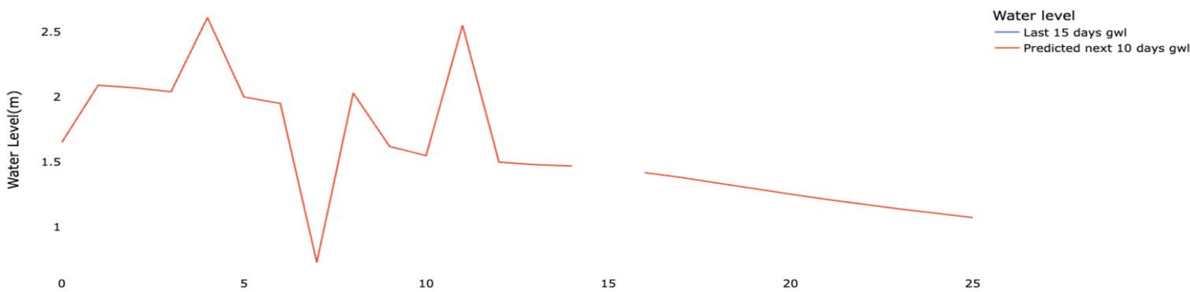


Figure 17: Plotting last 15 days and next predicted 10 days of Barisal by GRU

Illustrations: In the Figure 17, patterns the GWL prediction realizes comparing the last 15 days and next prophesied 10 days of Barisal by GRU algorithm by directing Timestamp.

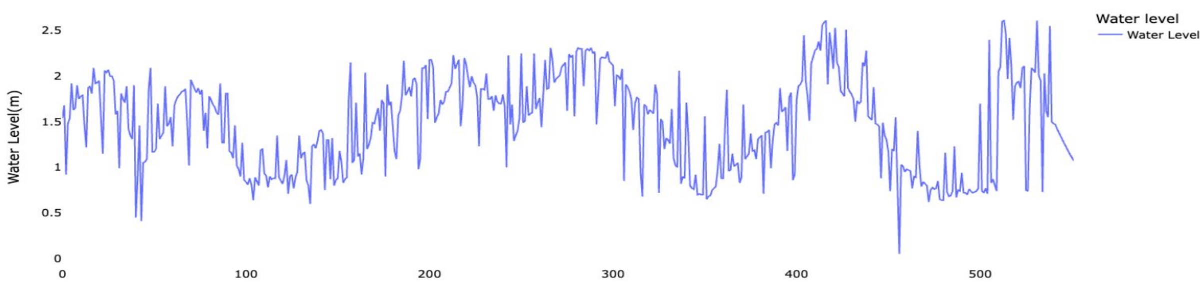


Figure 18: Plotting whole GWL with next 10 days prediction of Barisal by GRU

Illustrations: It detects that the Slight differences in the patterns (Figure 18), mostly in the timing and height of some peaks and troughs, especially after 400 and before 100.

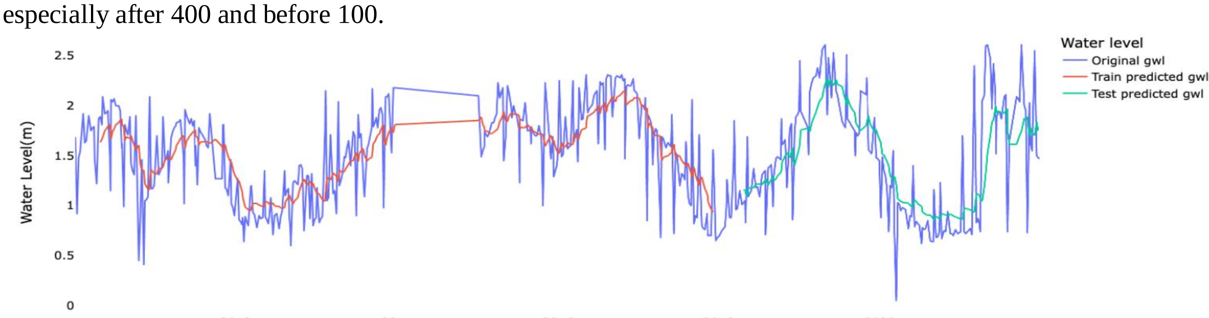


Figure 19: Comparison between original GWL vs predicted GWL with chart of Barisal by LSTM+GRU

Illustrations: The LSTM+GRU model give the imprint (Figure:19) to perform sensibly well in apprehending the general trends and seasonality of the GWL, but it tends to undervalue the extreme ideals.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

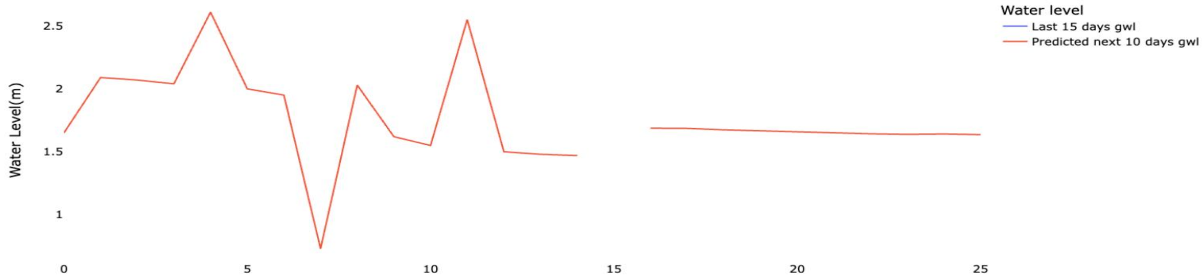


Figure 20: Plotting last 15 days and next predicted 10 days of Barisal by LSTM+GRU

Illustrations: It Detects (Figure 20) the intrigue outcome affording to the Timestamp superiority Plotting of the last 15 days vs the next 10 days forecasts.

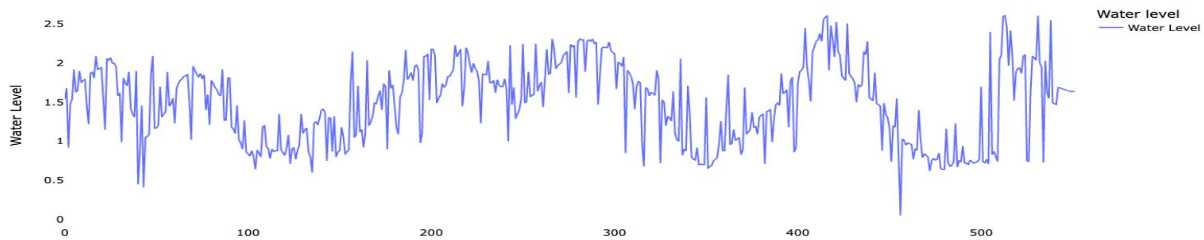


Figure 21: Plotting whole GWL with next 10 days prediction of Barisal by LSTM+GRU

Illustrations: Plotting the whole GWL with the forecast of the next 10 days estimate of LSTM+GRU model appears (Figure 21) to execute levelheadedly well in catching the general trends and seasonality of the GWL.

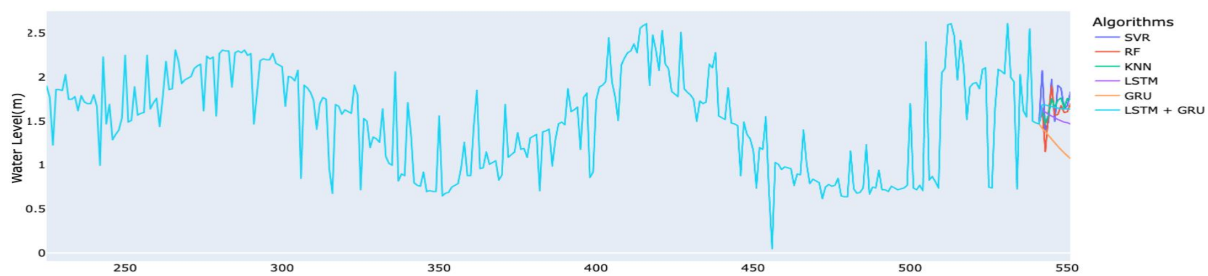


Figure 22: Plotting final chart with all algorithms and compare prediction to each other's

Illustrations: Figure 22: Plotting final chart with all algorithms and liken forecast to each other's: SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithm. The combination of the appropriate Set of rules with each other's that specifies the comparison and stipulations (Saltelli and Annoni, 2010).

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	0.258153	0.419721	0.066643	0.176166	0.209536	0.314044	0.688519	0.510212	0.687945	0.509543	0.036970	0.109592	0.047702	0.124680
Random Forest	0.119345	0.420378	0.014243	0.176718	0.088555	0.330257	0.933322	0.508009	0.933306	0.508007	0.009655	0.115933	0.011010	0.130269
K Nearest neigbor	0.301204	0.418054	0.090724	0.174769	0.230530	0.316943	0.576449	0.516964	0.575184	0.513432	0.049182	0.109442	0.064608	0.125084
LSTM	0.330093	0.416833	0.108961	0.173750	0.265163	0.294702	0.510917	0.532596	0.489787	0.516270	0.057897	0.107956	0.076907	0.124033
GRU	0.384603	0.448451	0.147919	0.201109	0.317536	0.318361	0.528435	0.573497	0.307367	0.440101	0.077462	0.125773	0.103827	0.144720
LSTM+GRU	0.317535	0.398500	0.100828	0.158802	0.243857	0.295726	0.528354	0.557886	0.527871	0.557885	0.054751	0.102562	0.071945	0.114904

Table 5: Algorithms performance

Accuracy Score of the Heatmap correlation of algorithms of Barisal: The heat map relationship of SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithm.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

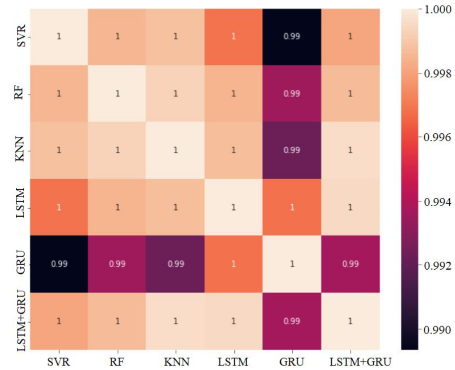


Figure 23: Correlation Heatmap of Barisal

Illustrations: In the figure 23, the high correlations point to that predicting from the unalike algorithms are reliable. There are some subtle differences in correlation between dissimilar pairs of the application algorithms, but overall, the relationships are very high crossways the board as demand oversight (Zarei and Sepaskhah, 2022).

B. *Accuracy Score of Algorithms of Barisal*

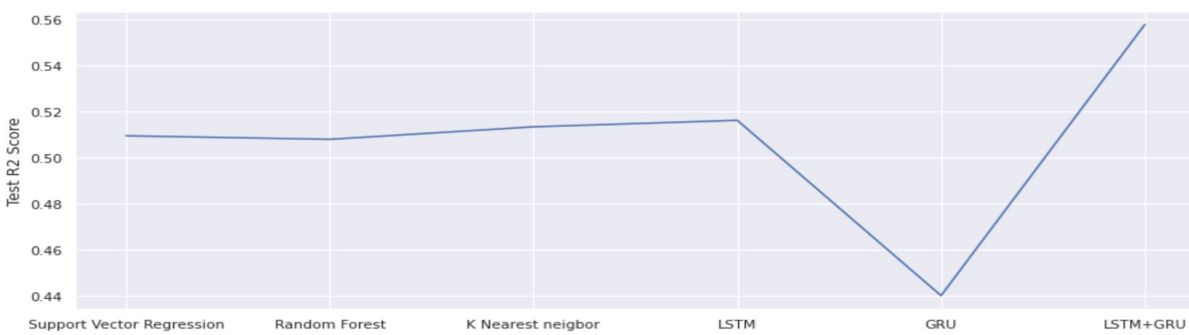


Figure 24: Accuracy Score of algorithms of Barisal

Illustrations: The flow of the curve demonstrations (Figure 24) the performance of the algorithm and its shadowed outcomes. The RF and the LSTM to LSTM+ GRU meet the target and its applications. GRU performance measurement is uncertain.

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	0.258153	0.419721	0.066643	0.176166	0.209536	0.314044	0.688519	0.510212	0.687945	0.509543	0.036970	0.109592	0.047702	0.124680
Random Forest	n_estimators=100 ,random_state=1	0.119345	0.420378	0.014243	0.176718	0.088555	0.330257	0.933322	0.508009	0.933306	0.508007	0.009655	0.115933	0.011010	0.130269
K Nearest neigbor	n_neighbors=15,metric=minkowski	0.301204	0.418054	0.090724	0.174769	0.230530	0.316943	0.576449	0.516964	0.575184	0.513432	0.049182	0.109442	0.064608	0.125084
LSTM	loss=mse,optimizer=adam,3 lstm layers with 32 ...	0.330093	0.416833	0.108961	0.173750	0.265163	0.294702	0.510917	0.532596	0.489787	0.516270	0.057897	0.107956	0.076907	0.124033
GRU	loss=mse,optimizer=adam,4 gru layers with 32 n...	0.384603	0.448451	0.147919	0.201109	0.317536	0.318361	0.528435	0.573497	0.307367	0.440101	0.077462	0.125773	0.103827	0.144720
LSTM+GRU	loss=mse,optimizer=adam,2 gru and 2 lstm layer...	0.317535	0.398500	0.100828	0.158802	0.243857	0.295726	0.528354	0.557886	0.527871	0.557885	0.054751	0.102562	0.071945	0.114904

Table 6: Summary Chart of Barisal

	date	waterLevel	temperature	humidity	rainfall	surface_soil_witness	root_soil_witness	profile_soil_moisture
0	2010-01-04	1.46	16.47	8.06	0.0	0.59	0.59	0.59
1	2010-01-11	1.46	16.66	7.81	0.0	0.55	0.56	0.56
2	2010-01-18	1.48	18.21	9.34	0.0	0.52	0.54	0.54
3	2010-01-25	1.54	20.00	8.12	0.0	0.48	0.52	0.52
4	2010-02-01	1.54	18.97	7.63	0.0	0.45	0.49	0.49

Table 7: Barisal features

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

C. Multivariate Time Series Forecasting of Barisal Zone

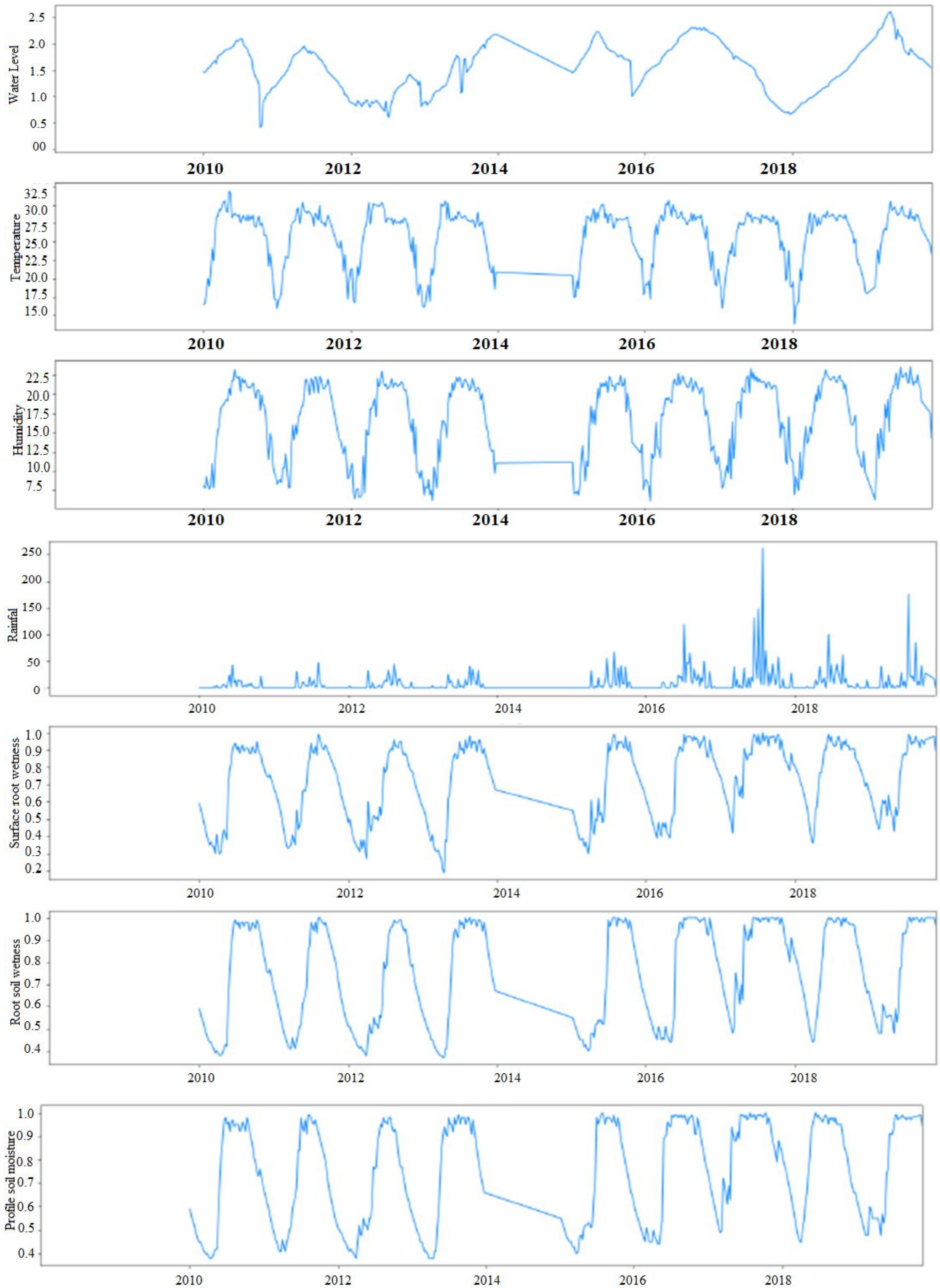


Figure 25: Barisal Multivariate features

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Illustrations: Based on the providing plots of time-series data, some key facts states (Sutskever, Vinyals and Le, 2014).

1. Strong Periodic Trends: Most structures exhibit clear seasonal shapes, driven by monsoon rainfall and temperature dissimilarities.
2. Rainfall Dependency: Soil moisture (surface, root, and profile) and groundwater levels show a robust dependency on rainfall, highlighting the status of rain in the Barisal area.
3. Anomalies: Deviations, such as reduced soil moisture in 2014 or heavy rainfall spears, may specify climatic measures or data anomalies.
4. Temporal Lags: Structures like groundwater levels and profile soil moisture display delayed replies to rainfall, reflecting their slower revive rates.
5. Sustainability Apprehension: Declining trends in groundwater levels and soil wetness recommend the need for enhanced water resource managing approaches.

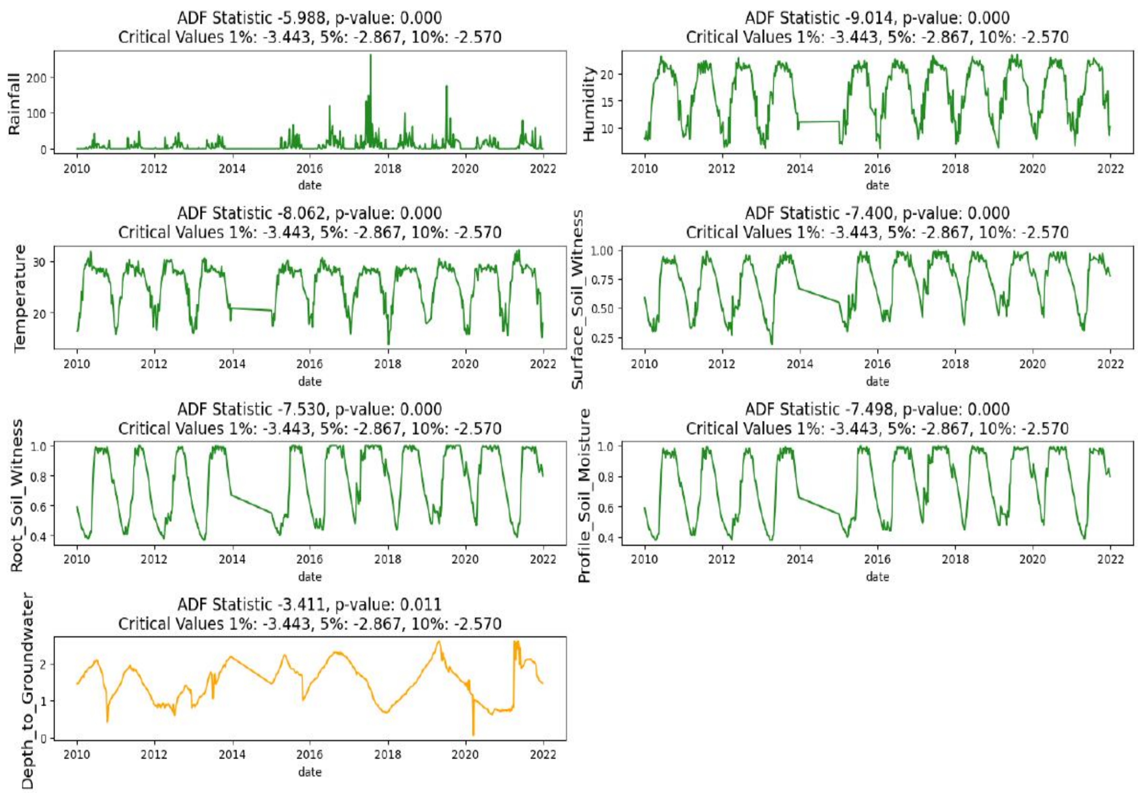


Figure 26: Barisal, p-value lower 0.05 which observe the ADF statistic's range in relative to crucial levels

Illustrations: ADF Test Results and Time-Series Study, based on the plots and the Augmented Dickey-Fuller (ADF) test consequences, the detailed explanations for each feature:

1. Stationarity Crossways Structures:
   - All structures were found to be stationary, which shortens the modeling process, mostly for time-series predicting methods.
2. Strong Periodic Patterns:
   - Most structures exhibit clear annual cycles, driven by the monsoon-dominated temperature of the Barisal Division.
3. Correlation Between Structures:
   - Rainfall seems like to drive disparities in surface and root soil moisture, which in turn impact groundwater levels.
   - Temperature and humidity also show corresponding seasonal patterns that align with rainfall-driven soil and groundwater changing aspects.
4. Ecological Variances:
   - Certain years show irregularities in rainfall and groundwater levels, likely sparkly extreme climatic measures.

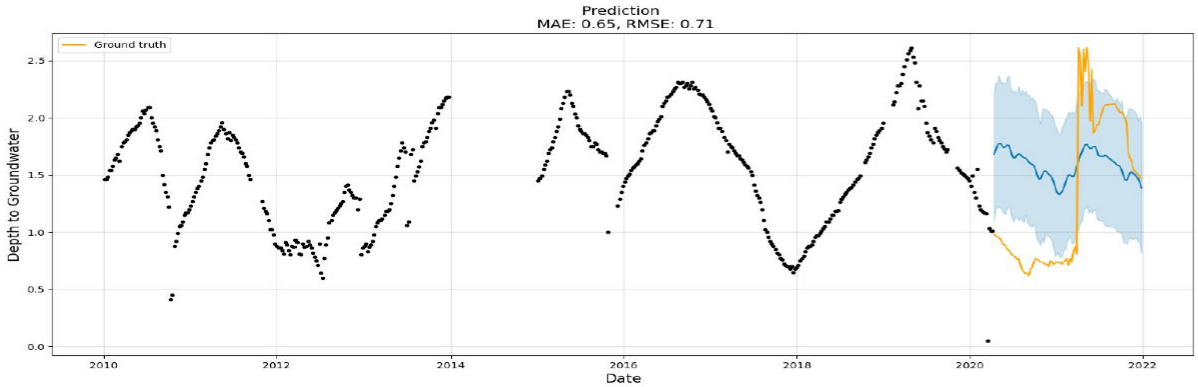


Figure 27: Barisal Facebook Prophet model output

Illustrations: The different explanations for the forecast of MAE:0.65 and RMSE:0.71 are carrying out as the application of Facebook Prophet model output.

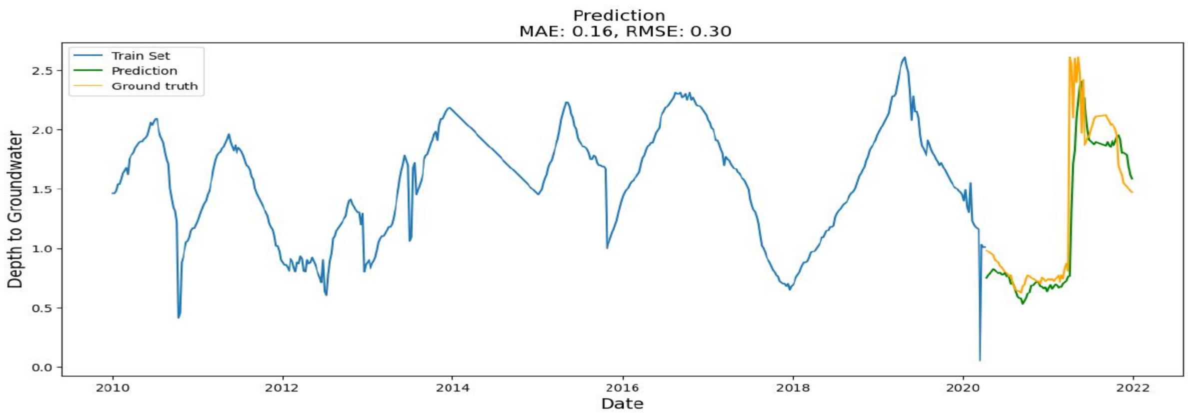


Figure 28: Barisal optimized LSTM model output

Illustrations: In the patterns imagining outcomes that displays the performance of LSTM model and here (Figure 28) the analytics of Train Set vs Predicting vs Ground Truth outcome are associated.

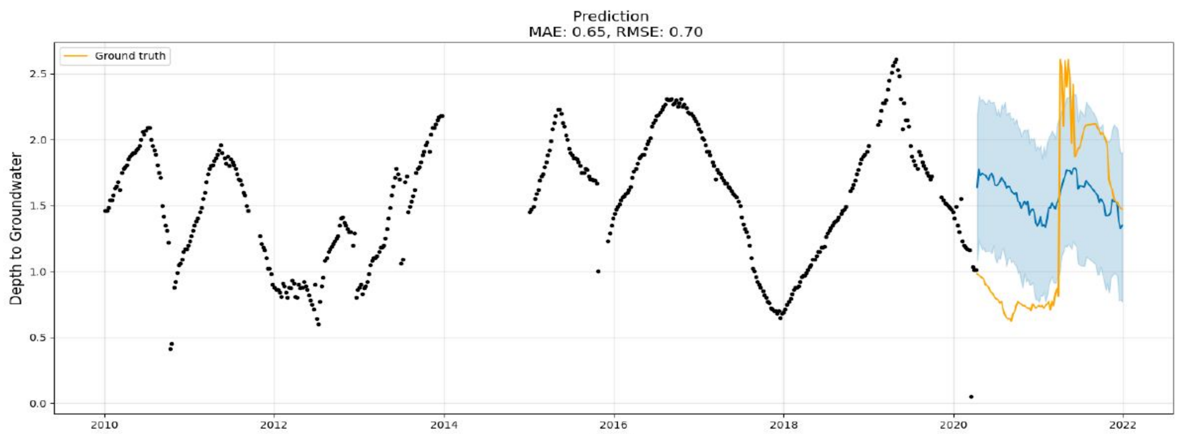


Figure 29: Barisal Multivariate model output

Illustrations: The results of the chart Figure 29, detects the fluctuations from Depth to GWL with the timestamp from 2008 to 2020. Close to 2011 the curve performs downstream and from 2018 to 2020 demonstrates the expected accuracy. The applications outcomes followed 2020 to 2022.

IV. RESULTS, FINDINGS AND RECOMMENDATIONS

A. Results

This section précises the results, key findings, and actionable recommendations based on the application of ML and DL performances to forecast the GWL, rainfall, humidity, temperature, and soil properties (surface, root, and profile) in the Barisal area, Bangladesh.

1) Model Performance Assessment

The following reviews the performance of the executed models based on metrics such as **MAE**, **RMSE**, **R²**, and **MAPE**:

Model	Level	Rainfall	Humidity	Temperature	Moisture	Moisture-2	Moisture-3
LSTM	High accuracy minimal errors (R^2 > 0.92)	Excellent for capturing temporal trends	Moderate Performance	High Performance (R^2 =0.91)	Good at detecting long term pattern	Moderate	Moderate
GRU	Comparison to LSTM slightly faster	Slightly less accurate than LSTM	Moderate	High accuracy	Moderate	Good for short term forecasts	Moderate
RF	Strong for non-linear relationship	Moderate performance	High accuracy	Moderate	Limited accuracy due to time series analytics	Weak	Weak
KNN	Baseline: Lacks predicting robustness	Moderate	Weak	Weak	Poor	Poor	Poor
SVR	Good for small datasets	Moderate	Weak	Moderate	Poor	Poor	Poor
LSTM+ GRU	Best performance (R^2 > 0.95) & low RMSE	Excellent for all parameters	High Accuracy	High Performance	Strong temporal understanding	High	High

Table 8: Algorithmic Comparison and performance velocity

- Best Performing Model: Hybrid (LSTM + GRU) outperformed all other models crossways almost all parameters due to its aptitude to capture both short- and long-term dependances.
- Baseline Models (KNN, SVR): Performed poorly for multifaceted and non-linear relations, highlighting their restrictions for time-series forecasts.
- Random Forest (RF): Effective for static data but struggled with time-series forecasts.

B. Findings

1) Groundwater Levels: ML and DL models accurately captured the trends, particularly the Hybrid LSTM+GRU, which showed the best simplification performance.
2) Rainfall Prediction: Rainfall forecasts demonstrated high exactness using temporal models (LSTM and GRU). A clear relationship was observed between the rainfall and soil wetness dynamics.
3) Soil Wetness and Profile Analysis: Soil wetness (surface, root, and profile) displayed high temporal dependance on rainfall and humidity, which were successfully modeled by LSTM and Hybrid LSTM+GRU. RF struggled to handle the chronological nature of soil dynamics (Kratzert, Klotz, Shalev and et al., 2019).
4) Temperature and Humidity: Temperature was easier to forecast equated to the humidity due to more stable patterns in the historic data. Models such as GRU and Hybrid LSTM+GRU executed steadily well.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

5) Regional Insights for Barisal Division: Barisal's ecological parameters show strong seasonality and variability, requiring models that can handle time-series data efficiently. Some gaps in data impacted model exactness, requiring advanced imputation performances (Gelaro, McCarty, Suárez, et al., 2017).

C. Recommendations

Based on the findings, the resulting recommendations are providing for stakeholders, researchers, and policymakers:

1) Model Implementation
- Implement Hybrid Models (LSTM + GRU): For future forecasts, Hybrid LSTM+GRU is the most robust method for handling multifaceted time-series data.
- Feature Engineering: Integrate the lagged structures, rolling averages, and external variables to expand model correctness further.

2) Groundwater Administration
- Early Warning System: Use analytical models to create an early warning system for GWL depletion in the Barisal area specific.
- Rainwater Harvesting Approaches: Utilize rainfall forecasts to design optimal harvesting and storage tactics.

3) Soil Wetness and Cultivation (Karthikeyan, Khosa, R., and Singh, 2020)
- Irrigation Forecasting: Use soil wetness forecasts to guide irrigation plans, ensuring resourceful water use for cultivation.
- Drought Awareness: Incorporate soil and rainfall predictions into drought resilience tactics.

4) Information Gathering and Setup
- Enhanced Information Assortment:
  - Increase the number of GWL monitoring stations crossways Barisal.
  - Use sensors for real-time soil and moisture information gathering.
- Integrate Isolated Sensing Information: Influence satellite data to fill gaps in soil and rainfall figures.

5) Future Research Instructions
- Hybrid Modeling: Explore supplementary hybrid architectures merging ML and DL performances
- Spatial Variability: Expand study to consist of spatial breakdown for better regional insights.
- Climate Alteration Influence: Explore the effects of climate variation on rainfall, GWL, and soil dynamics.

6) Policy References
- Promote the acceptance of data-driven water resource managing structures.
- Develop strategies for maintainable GWL extraction based on projecting insights.
- Encourage collaborations between researchers, local agencies, and legislators for effective execution of analytical models.

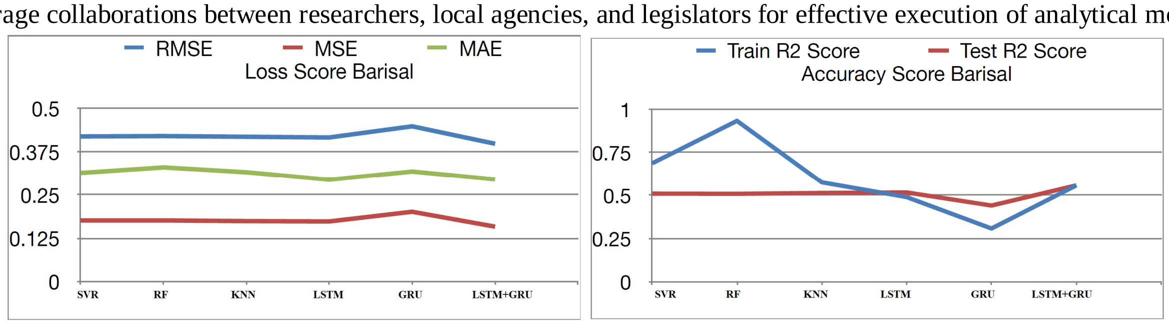


Figure 30: Loss Score and Accuracy Score of Barisal

Observations: In the Figure 30 states, LSTM + GRU consistently performs better crossways loss (RMSE, MSE, MAE) and accuracy metrics (train-test R^2 scores), making it the most consistent model. Traditional machine learning models, such as RF, show strong performance on training data but struggle with simplification on test data, signifying overfitting. Deep learning models achieve better simplification, suggesting that for this dataset, hybrid methods are more actual (Anderson, Woessner and Hunt, 2015).

V. CONCLUSION

This research efficaciously explored and realized various machine learning (ML) and deep learning (DL) techniques, including LSTM, GRU, Random Forest (RF), KNN, SVR, and a Hybrid LSTM+GRU model, to forecast critical ecological and groundwater-related parameters, such as groundwater level, rainfall, humidity, temperature, and soil moisture as like as surface, root, and profile, in the Barisal Division of Bangladesh.

The study as long as valuable insights into the complex relationships between these variables, leveraging historical and temporal data for precise estimating. Deep learning models, particularly LSTM and Hybrid LSTM+GRU, consistently outperformed traditional ML models due to their ability to capture temporal dependances in the time-series data. The Hybrid LSTM+GRU model verified the best overall performance with high exactness and strength across all strictures. Rainfall, temperature, and humidity were found to be the most significant features for predicting groundwater levels and soil wetness. Integrating lagged features and rolling means pointedly enhanced model correctness. The strong seasonality and variability of ecological data in Barisal Division required erudite time-series models to efficiently capture long-term and short-term shapes. The prognostic models can be useful to design effective groundwater management plans, optimize irrigation timetables, and prepare for drought surroundings. This research contributes to the rising body of information on using ML and DL techniques for bearable resource managing in water-scarce areas. It makes available a groundwork for future research into groundwater and soil dynamics by integrating advanced analytical modeling techniques tailored to local circumstances in Barisal Division. The findings can aid legislators, water resource managers, and agricultural planners in emerging data-driven strategies for resource optimization and climate version. performance. Range the research to integrate spatial variability within Barisal Division using GIS-based modeling to capture regional alterations more successfully. Examine the role of climate change in inducing long-term trends in groundwater levels, soil moisture, and rainfall patterns in Barisal Division. Develop real-time projecting systems integrated with local decision-making contexts to provide actionable insights for water and agricultural management. The research verified the potential of advanced ML and DL techniques in addressing critical encounters in water resource managing. By leveraging state-of-the-art analytical models, participants in Barisal Division can adopt sustainable practices to safeguard effectual use of groundwater, enhance agricultural activities, and moderate the risks associated with climate changeability.

REFERENCES

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160.

[2] Anderson, M.P., Woessner W.W., and Hunt R.J. (2015). Applied Groundwater Modeling: Simulation of Flow and Advective Transport. London: Academic Press.

[3] Collenteur, R., Bakker M., Caljé R., Schaars F., and Klop S. (2019). Pastas: open-source software for the analysis of groundwater time series. Groundwater, 10.1111/gwat.12925.

[4] Gelaro, R., McCarty, W., Suárez, M. J., et al. (2017). Modern-era retrospective analysis for research & applications, version 2 (MERRA-2). Journal of Climate, 30(14), 5419-5454.

[5] Gharbi, S., & Bouaziz, M. (2023). Real-time data assimilation for GWL prediction using machine learning. Water Resources Management, 37(5), 1461-1478. DOI.

[6] Haurie, A., & Ghaffari, A. (2022). Explainable AI in hydrology: Challenges and opportunities. Environmental Modelling & Software, 155, 105454. DOI.

[7] 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.

[8] Kratzert, F., Klotz, D., Shalev, G., et al. (2019). Towards improved predictions in ungauged basins: Exploiting the power of ML. Water Resources, 55(12), 11344-11354.

[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] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). LSTM in Natural Language Processing: Sequence to Sequence Learning with NN. Advances in Neural Information Processing

[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.

[12] Obergfell, C., Bakker M., and Maas K. (2019). Identification and explanation of a change in the groundwater regime using time series analysis. Groundwater. 10.1111/gwat.12891

[13] Peterson, T.J., and Western A.W. (2018). Statistical interpolation of groundwater hydrographs. Water Resources Research 54, no. 7: 4663–4680.

[14] Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Software, 25(12), 1508-1517. DOI.

[15] Shi, X., Chen, Z., Wang, H., et al. (2015). Convolutional LSTM network: ML approach for precipitation nowcasting. Advances Neural Information Processing Systems, 28, 802-810.

[16] Simoni, S., & Manoli, G. (2021). Monte Carlo-based sensitivity analysis in groundwater simulation models. Environmental Modelling & Software, 136, 104979. DOI.

[17] Sun, W., & Wang, S. (2021). Gradient-based sensitivity analysis in neural networks for groundwater modeling. Water, 13(3), 756. DOI.

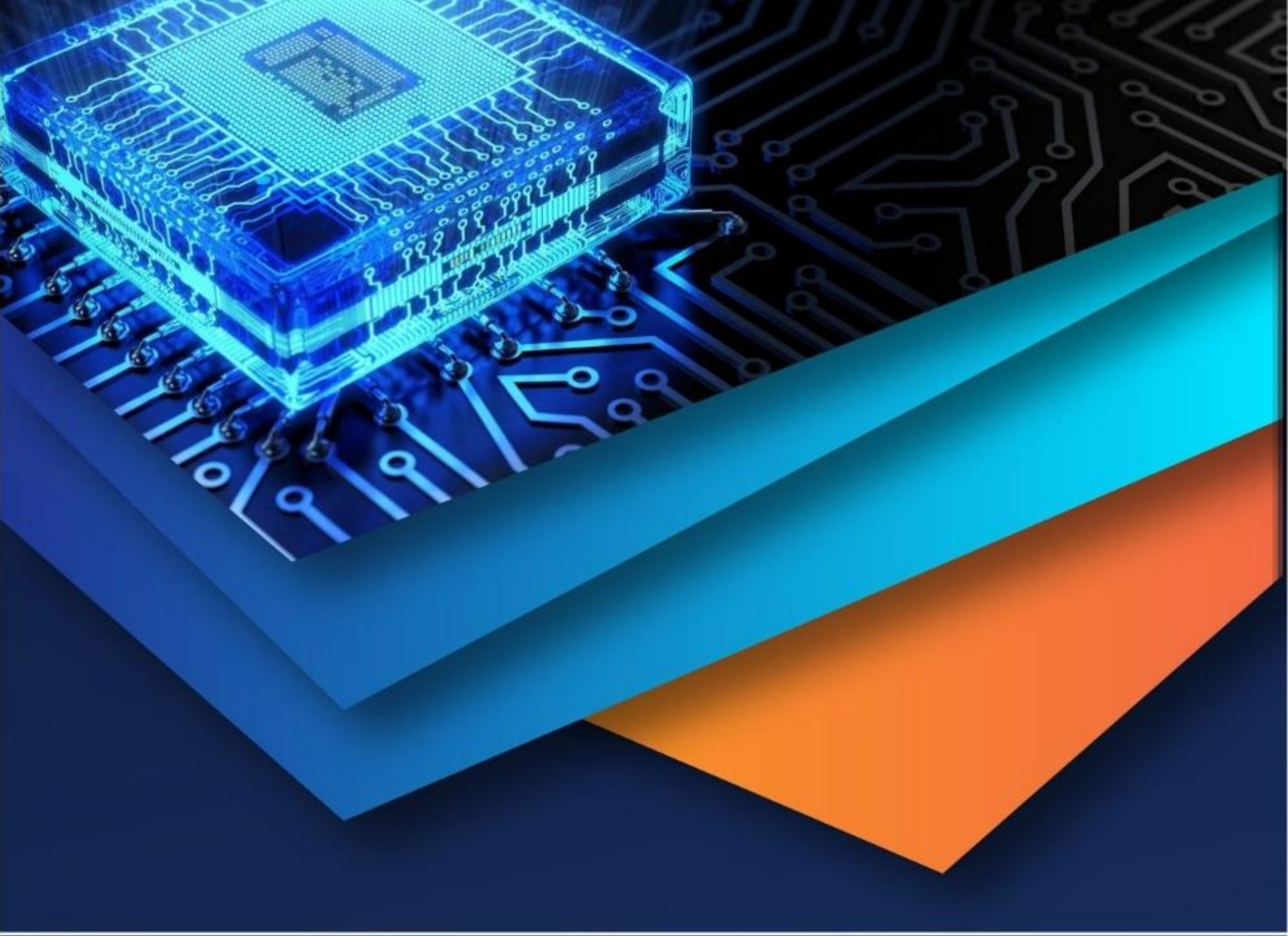
[18] Tang, J., Zhang, X., & Li, Q. (2021). Groundwater level prediction using a Transformer-based model. Journal of Hydrology, 594, 125707. DOI.

[19] Wani, S. P., & Wani, M. H. (2021). Application of Gated Recurrent Units for groundwater level forecasting. Hydrology, 8(2), 61. DOI.

[20] Wunsch, A., Liesch T., and Broda S. (2018). Forecasting GWL nonlinear autoregressive networks with exogenous input (NARX). Journal of Hydrology 567: 743–758.

[21] Vaswani, A., et al. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 6000-6010).

[22] Zarei, A. R., & Sepaskhah, A. R. (2022). Prediction of GWL in an arid region using ML and ensemble methods. Hydrological Sciences Journal, 67(7), 1150-1165. DOI.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY