© March 2025 | IJIRT | Volume 11 Issue 10 | ISSN: 2349-6002

Godavari River Water Quality Prediction Using Machine Learning Algorithm

Dr.K.N.S Lakshmi[1], A. Akshaya[2], Ch.Pavan Saketh[3], T. Manikanta[4], B. Jyothi[5]
[1]*Professor, Computer Science and Engineering, Sanketika Vidya Parishad Engineering college*
[2,3,4,5] *Computer Science and Engineering, Sanketika Vidya Parishad Engineering college*

***Abstract*—Water quality is critical for public health and environmental safety, necessitating effective monitoring methods. This study presents a methodology for predicting river water quality using integrated geospatial and statistical approaches. High-resolution satellite imagery, turbidity data from USGS Earth Explorer, and hydrological features extracted from QGIS, such as NDWI, were used to train a Random Forest model. The model demonstrated high accuracy in predicting water quality across diverse conditions. Spatial distribution of predictions was visualized in QGIS for intuitive interpretation. This framework showcases the effectiveness of combining remote sensing, GIS, and machine learning for scalable and efficient water quality assessment.**

***Index Terms*—Water quality, Remote sensing, GIS, Machine learning, Random Forest, Turbidity, NDWI, QGIS, Environmental monitoring**

I. INTRODUCTION

Monitoring water quality in riverine systems is essential for managing water resources and safeguarding public health. This study aims to predict turbidity levels in the Godavari River, a major water source in India, using satellite remote sensing and machine learning techniques. The Normalized Difference Water Index (NDWI), derived from the Green and Near-Infrared (NIR) bands of satellite imagery, is used to assess variations in water surface characteristics over time.

Ground truth turbidity data (measured in NTU) serves as the target variable for model validation. A Random Forest model was trained using Delta NDWI as the primary feature to predict turbidity levels, achieving promising results. Model performance was evaluated using Mean Absolute Error (MAE) and R^2 metrics, demonstrating a strong correlation between Delta NDWI and turbidity fluctuations. This study highlights the potential of Delta NDWI as a remote sensing index for water quality monitoring and suggests further enhancements through additional swater quality indices and advanced machine learning techniques.

1.1 Study Area

Godavari River The Godavari River, one of India's longest rivers, flows through Maharashtra, Telangana, Andhra Pradesh, Chhattisgarh, and Odisha, ultimately draining into the Bay of Bengal. It serves as a vital water source for agriculture, industry, and drinking purposes. However, pollution from agricultural runoff, industrial discharge, and domestic waste threatens its water quality. Monitoring water quality is crucial for sustaining the river ecosystem and ensuring safe water resources.

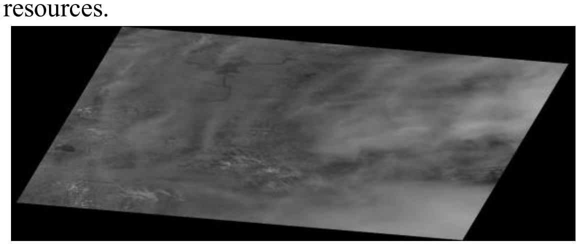


Fig1.1:Study area of Godavari River

1.2 Methodology Flowchart

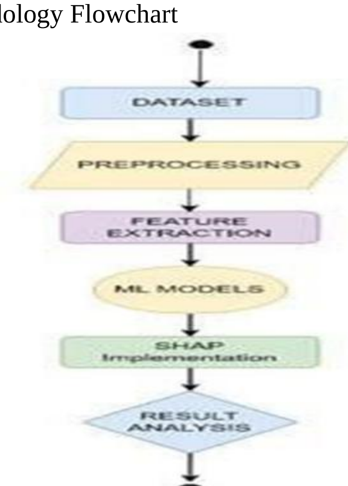


Fig1.2: Block diagram

II. LITERATURE SURVEY

Water quality prediction is critical for environmental monitoring, impacting public health, agriculture, and ecosystems. Machine learning (ML) techniques have been widely applied to enhance predictive accuracy, particularly for parameters such as pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), turbidity, and conductivity.

Random Forest in Water Quality Prediction

Random Forest (RF), an ensemble learning method, has demonstrated superior predictive performance in various studies.

Fig-2 Godavari river

III. NDWI AND WATER QUALITY ANALYSIS

3.1NDWI and Water Quality Analysis

The Normalized Difference Water Index (NDWI) is a widely used remote sensing index for detecting and monitoring water bodies. Calculated using the reflectance values from the Green **and** Near-Infrared (NIR) bands, NDWI helps distinguish water surfaces from land and vegetation.

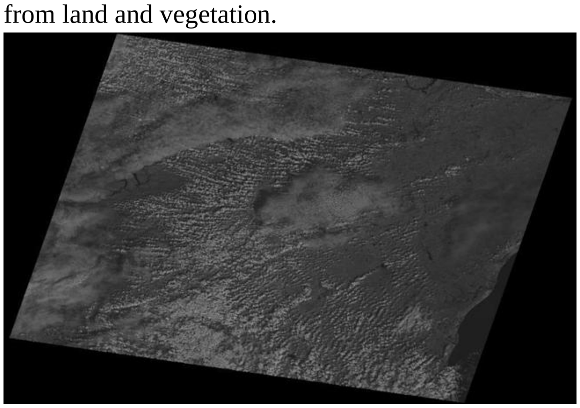


FIG:3.1 NDWI Calculation Illustration

3.2Delta NDWI for Water Quality Assessment

Delta NDWI (ΔNDWI) represents the temporal difference between NDWI values at two distinct time points, enabling the analysis of water quality changes, particularly turbidity variations. Turbidity, a key water quality parameter, influences light penetration and aquatic ecosystems.

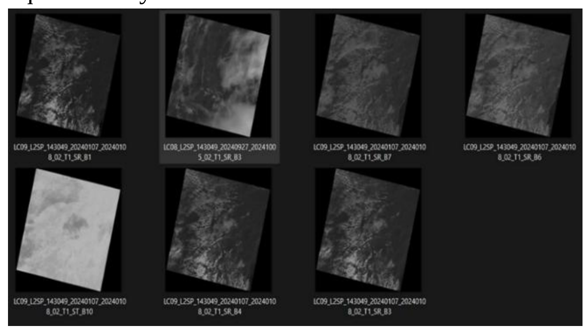


Fig 3.2 Godavari River (USGS)

3.3Methodology for Delta NDWI Analysis

1. Data Collection **–** Acquiring satellite imagery (Landsat/Sentinel) and ground truth turbidity data.
2. Preprocessing – Atmospheric corrections and NDWI computation for different time periods.
3. ΔNDWI Calculation **–** Computing changes in NDWI values to assess water clarity variations.
4. Machine Learning Modeling **–** Using Random Forest (RF) to predict turbidity based on ΔNDWI.
5. Evaluation **–** Model performance assessed using Mean Absolute Error (MAE) and R^2 metrics**.**

IV. DATA PROCESSING

4.1 Raster Data Processing (QGIS & Python)

Satellite Imagery Preprocessing – Landsat/Sentinel images were corrected for atmospheric effects, and Green & NIR bands were extracted for NDWI calculation.

ΔNDWI Computation – The temporal difference between NDWI values was calculated to analyze water quality changes.

4.2 Ground Truth Data Preprocessing

Data Collection – Turbidity values (NTU) were gathered from in-situ measurements.

Data Cleaning – Missing values were handled, and datasets were aligned with ΔNDWI values for consistency.

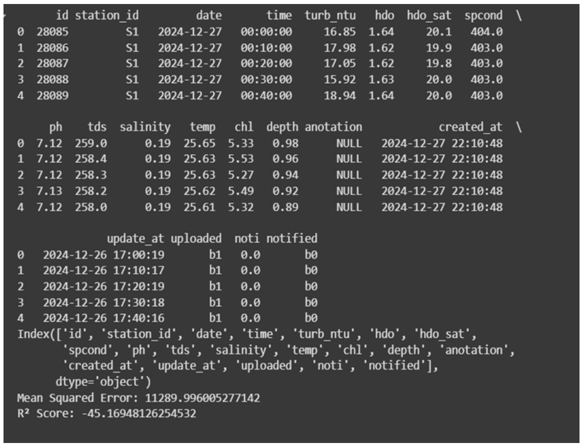
4.3 Feature Extraction

Key Feature Selection – Mean, standard deviation, and min/max ΔNDWI values were extracted as predictors for machine learning.

Normalization – Features were scaled to ensure uniformity across data inputs.

4.4 Data Splitting for Model Training

Training & Testing Sets – Data was split (80%-20%) to evaluate model performance, ensuring robust validation.



```
      id station_id        date      time  turb_ntu  hdo  hdo_sat  spcond  \
0  28085         S1  2024-12-27  00:00:00     16.85 1.64     20.1   404.0
1  28086         S1  2024-12-27  00:10:00     17.98 1.62     19.9   403.0
2  28087         S1  2024-12-27  00:20:00     17.05 1.62     19.8   403.0
3  28088         S1  2024-12-27  00:30:00     15.92 1.63     20.0   403.0
4  28089         S1  2024-12-27  00:40:00     18.94 1.64     20.0   403.0

     ph    tds  salinity   temp   chl  depth anotation           created_at  \
0  7.12  259.0      0.19  25.65  5.33   0.98      NULL  2024-12-27 22:10:48
1  7.12  258.4      0.19  25.63  5.53   0.96      NULL  2024-12-27 22:10:48
2  7.12  258.3      0.19  25.63  5.27   0.94      NULL  2024-12-27 22:10:48
3  7.13  258.2      0.19  25.62  5.49   0.92      NULL  2024-12-27 22:10:48
4  7.12  258.0      0.19  25.61  5.32   0.89      NULL  2024-12-27 22:10:48

            update_at uploaded  noti notified
0  2024-12-26 17:00:19       b1   0.0       b0
1  2024-12-26 17:10:17       b1   0.0       b0
2  2024-12-26 17:20:19       b1   0.0       b0
3  2024-12-26 17:30:18       b1   0.0       b0
4  2024-12-26 17:40:16       b1   0.0       b0
Index(['id', 'station_id', 'date', 'time', 'turb_ntu', 'hdo', 'hdo_sat',
       'spcond', 'ph', 'tds', 'salinity', 'temp', 'chl', 'depth', 'anotation',
       'created_at', 'update_at', 'uploaded', 'noti', 'notified'],
      dtype='object')
Mean Squared Error: 11289.996005277142
R² Score: -45.16948126254532
```

Fig:4.1 Raster Data Design

V. MODELLING APPROACH

Random Forest was chosen as the preferred model for turbidity prediction in the Godavari River due to its ability to handle non-linearity, robustness to noise, interpretability, and ease of implementation. Compared to Linear Regression, SVM, ANN, and GBM, RF provides a reliable and computationally efficient solution for predicting water quality using remote sensing data. Future work could involve integrating additional machine learning techniques, such as hybrid models combining RF with deep learning approaches, to further enhance predictive performance.

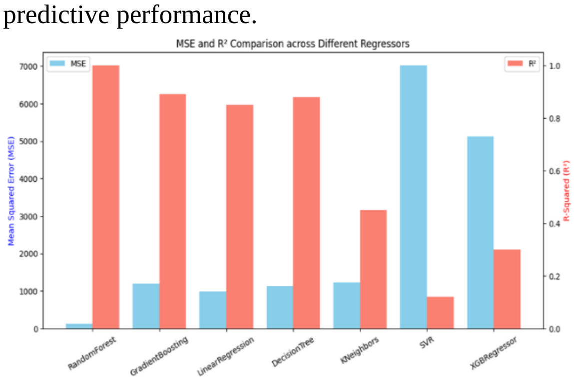


Fig 5.1 Regressor Models

5.1 Random Forest For Turbidity Prediction

This straightforward approach helps determine whether a significant linear relationship exists between Delta NDWI and turbidity, providing valuable insights into the feasibility of using remote sensing data for water quality prediction.

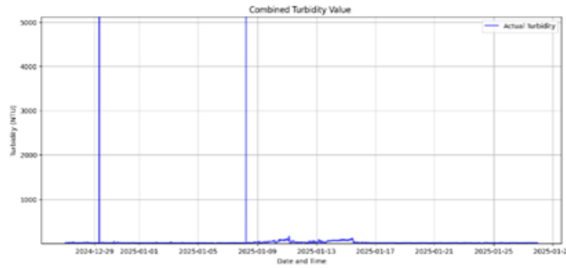


Fig:5.2 Combined Turbidity

VI. RESULTS AND DISCUSSION

6.1 Delta NDWI Visualization

The Delta NDWI (ΔNDWI) serves as an indicator of water quality changes over time. For this project, ΔNDWI values were calculated as the difference between NDWI measurements taken from two time periods, allowing for a visual assessment of changes in the Godavari River's water content and turbidity.

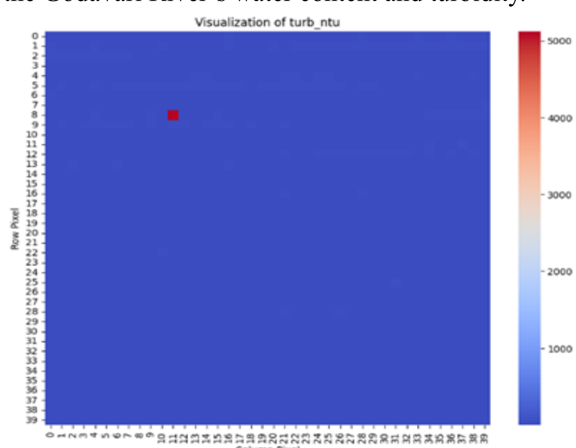


Fig:6.1 Delta NDWI Visualization

6.2 Predicted vs. Actual Turbidity

After training the Random Forest model using Delta NDWI as the predictor, the model's predictions for turbidity were compared with actual turbidity measurements (in NTU) from the ground truth data.

Scatter Plot of Predicted vs. Actual Turbidity

A scatter plot comparing predicted and actual turbidity values was generated to evaluate model performance visually.

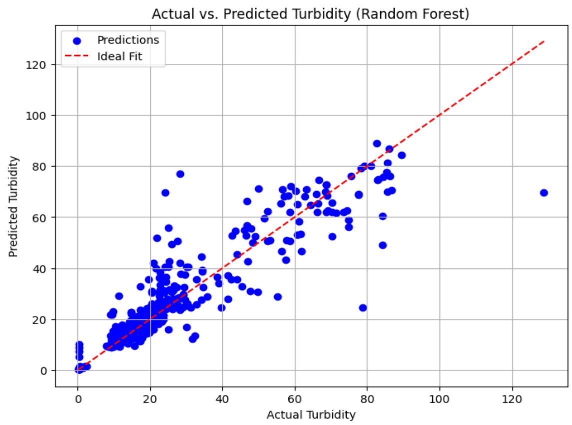


Fig:6.2 Scatter Plot of Predicted vs. Actual Turbidity

6.3 SHAP VALUE IMPACT

SHapley Additive exPlanations (SHAP) is a powerful interpretability technique used to analyze feature contributions in machine learning models. In this study, SHAP was implemented to understand the impact of Delta NDWI (ΔNDWI) and other features on turbidity prediction using the Random Forest model. The SHAP diagram provides a visual representation of feature importance, showing how each variable influences the model's predictions. Higher SHAP values indicate a greater contribution of a specific feature to the predicted turbidity levels. By leveraging SHAP, we identified that ΔNDWI plays a crucial role in water quality prediction, reinforcing its effectiveness as a remote sensing-based indicator. This implementation enhances model transparency, helping stakeholders interpret and trust the predictions for environmental monitoring.

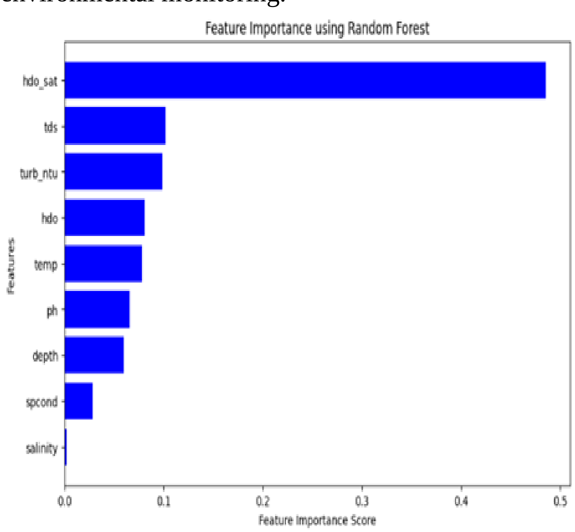


Fig 6.3: Feature importance of SHAP

VI. CONCLUSION

The research took a systematic approach to predict raw water quality by carefully choosing essential features for accurate estimations. Seven different regression models were tested, each showing diverse performances. The SHAP analysis revealed crucial features influencing accurate predictions in the Random Forest model, while Lime offered specific explanations, making the model's decisions easier to understand. Overall, this comprehensive approach not only improved prediction accuracy but also provided a deeper understanding of key features, making the models more reliable for assessing raw water quality.

REFERENCE

[1] Z. Luo, Q. Shao, Q. Zuo, and Y. Cui, ''Impact of land use and urbanization on river water quality and ecology in a dam dominated basin,'' J. Hydrol., vol. 584, May 2020, Art. no. 124656.

[2] S. Loos, C. M. Shin, J. Sumihar, K. Kim, J. Cho, and A. H. Weerts,''Ensemble data assimilation methods for improving river water quality forecasting accuracy,'' Water Res., vol. 171, Mar. 2020, Art. no. 115343.

[3] M. Okumah, A. S. Yeboah, and S. K. Bonyah, ''What matters most? Stakeholders' perceptions of river water quality,'' Land Use Policy, vol. 99, Dec. 2020, Art. no. 104824.

[4] S. B. H. S. Asadollah, A. Sharafati, D. Motta, and Z. M. Yaseen, ''River water quality index prediction and uncertainty analysis: A comparative study of machine learning models,'' J. Environ. Chem. Eng., vol. 9, no. 1, Feb. 2021, Art. no. 104599.

[5] Z. Z. Al-Janabi, A.-R. Al-Kubaisi, and A.-H. M. J. Al-Obaidy, ''Assessment of water quality of Tigris River by using water quality index (CCME WQI),'' Al-Nahrain J. Sci., vol. 15, no. 1, pp. 119–126, 2012.

[6] Y. R. Ding, Y. J. Cai, P. D. Sun, and B. Chen, ''The use of combined neural networks and genetic algorithms for prediction of river water quality,'' J. Appl. Res. Technol., vol. 12, no. 3, pp. 493–499, Jun. 2014.

[7] Tiyasha, T. M. Tung, and Z. M. Yaseen, ''A survey on river water quality modelling using

artificial intelligence models: 2000–2020,'' J. Hydrol., vol. 585, Jun. 2020, Art. no. 124670.

[8] T. R. Zolnikov, ''The maladies of water and war: Addressing poor water quality in Iraq,'' Amer. J. Public Health, vol. 103, no. 6, pp. 980–987, Jun. 2013.

[9] A. J. Kadhem, ''Assessment of water quality in tigris river-iraq by using GIS mapping,'' Natural Resour., vol. 4, no. 6, pp. 441–448, 2013.

[10] A. S. Alsaqqar, B. H. Khudair, and S. K. Ali, ''Evaluating water stability indices from water treatment plants in Baghdad City,'' J. Water Resource Protection, vol. 6, no. 14, p. 1344, 2014.

[11] K. A. Rahi and T. Halihan, ''Changes in the salinity of the euphrates river system in Iraq,'' Regional Environ. Change, vol. 10, no. 1, pp. 27–35, Mar. 2010.

[12] S. H. Abbas, B. H. Khudair, and M. S. Jaafar, ''River water salinity impact on drinking water treatment plant performance using artificial neural network,'' J. Eng., vol. 25, no. 8, pp. 149–159, Jul. 2019.

[13] Z. F. Makki, A. A. Zuhaira, S. M. Al-Jubouri, R. K. S. Al-Hamd, and L. S. Cunningham, ''GIS-based assessment of groundwater quality for drinking and irrigation purposes in central Iraq,'' Environ. Monitor. Assessment, vol. 193, no. 2, pp. 1–27, Feb. 2021.

[14] S. I. Abba, S. J. Hadi, S. S. Sammen, S. Q. Salih, R. A. Abdulkadir, Q. B. Pham, and Z. M. Yaseen, ''Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination,'' J. Hydrol., vol. 587, Aug. 2020, Art. no. 124974.

[15] A. Azad, H. Karami, S. Farzin, A. Saeedian, H. Kashi, and F. Sayyahi, ''Prediction of water quality parameters using ANFIS optimized by ntelligence algorithms (case study: Gorganrood River),'' KSCE J. Civil Eng., vol. 22, no. 7, pp. 1–8, 2017.

[16] Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, ''A review of the artificial neural network models for water quality prediction,'' Appl. Sci., vol. 10, no. 17, p. 5776, Aug. 2020.

[17] T. Rajaee, S. Khani, and M. Ravansalar, ''Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review,'' Chemometric Intell. Lab. Syst., vol. 200, May 2020, Art. no. 103978.

[18] R. Barzegar, J. Adamowski, and A. A. Moghaddam, ''Application of wavelet-artificial intelligence hybrid models for water quality prediction: A case study in Aji-Chay River, Iran,'' Stochastic Environ. Res. Risk Assessment, vol. 30, no. 7, pp. 1797–1819, Oct. 2016.

[19] M. Ehteram, S. Q. Salih, and Z. M. Yaseen, ''Efficiency evaluation of reverse osmosis desalination plant using hybridized multilayer perceptron with particle swarm optimization,'' Environ. Sci. Pollut. Res., vol. 27, no. 13, pp. 15278–15291, May 2020.

[20] Z. Yaseen, M. Ehteram, A. Sharafati, S. Shahid, N. Al-Ansari, and A. ElShafie, ''The integration of nature-inspired algorithms with least square support vector regression models: Application to modeling river dissolve doxygenconcentration,''Water, vol.10, no.9, p. 1124, Aug.2018.

[21] M. Bayatvarkeshi, M. A. Imteaz, O. Kisi, M. Zarei, and Z. M. Yaseen,''Application of M5 model tree optimized with excel solver platform for water quality parameter estimation,'' Environ. Sci. Pollut. Res., vol. 28, no. 6, pp. 1–18, 2020.

[22] S. Park, S. Jung, H. Lee, J. Kim, and J.-H. Kim, ''Large-scale water quality prediction using federated sensing and learning: A case study with realworld sensing big-data,'' Sensors, vol. 21, no. 4, p. 1462, Feb. 2021.

[23] C. S. Akratos, J. N. E. Papaspyros, and V. A. Tsihrintzis, ''An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands,'' Chem. Eng. J., vol. 143, nos. 1–3, pp. 96–110, Sep. 2008.

[24] T. Deng, K.-W. Chau, and H.-F. Duan, ''Machine learning based Marine water quality prediction for coastal hydro-environment management,'' J. Environ. Manage., vol. 284, Apr. 2021, Art. no. 112051.

[25] A. Naseri, M. Jamei, I. Ahmadianfar, and M. Behbahani, ''Nanofluids thermal conductivity prediction applying a novel hybrid data-driven model validated using Monte Carlo-based sensitivity analysis,'' Eng. Comput., pp. 1–25, Oct. 2020.

[26] I. Ahmadianfar, M. Jamei, M. Karbasi, A. Sharafati, and B. Gharabaghi, ''A novel boosting ensemble committee-based model for local scour

depth around non-uniformly spaced pile groups,'' Eng. Comput., pp. 1–23, Mar. 2021.
[27] M. Gholizadeh, M. Jamei, I. Ahmadianfar, and R. Pourrajab, ''Prediction of nanofluids viscosity using random forest (RF) approach,'' Chemometric Intell. Lab. Syst., vol. 201, Jun. 2020, Art. no. 104010.