AquaNet-X: A Deep Hybrid Ensemble Model for Accurate Real-Time Water Quality Index Prediction

S. Siva Nageswara Rao[1], Adhikari Satish[2], Tullibilli Lakshmi Siva Sai[3], Pallapu Harish[4],
Mallikarjuna Rao Gundavarapu[5], Patri Venkata Sesha Sudha Arundathi Parimala[6], Dodda Venkataredety[7]
profssnr@gmail.com[1], dhkrstsh@gmail.com[2], tullibillisivasai678@gmail.com[3],
pallapuharish312004@gmail.com[4], gmallikarjuna628@grietcollege.com[5], patri.parimala@gnits.ac.in[6],
doddavenkatareddy@gmail.com[7]
Department of CSE, Narasaraopeta Engineering College, Narasaraopet, India[12347],
Department of CSE, GRIET, Hyderabad, India[5],
Department of EEE, G.Narayanamma Institute of Technology and Science (for women), Hyderabad, India[6]

Abstract—Water quality plays a vital role in protecting public health, agriculture, and ecosystems, yet real-time monitoring remains a challenge due to irregular sampling, regional variations, and the limitations of traditional prediction models. To address these gaps, this paper introduces AquaNet-X, a novel deep hybrid ensemble model designed for accurate and scalable Water Quality Index (WQI) prediction. AquaNet-X integrates Bidirectional GRU for sequential dynamics, Transformer layers for capturing long-range feature dependencies, and boosting algorithms (XGBoost and LightGBM) for nonlinear tabular interactions, all unified through a Meta-CatBoost stacked learner. This architecture balances the strengths of deep learning and ensemble methods, reducing variance while enhancing interpretability and robustness. This experiment was conducted using a real-world Indian surface water quality dataset with multivariate parameters such as pH, DO, BOD, and temperature, preprocessed into supervised sequences. The proposed model achieving 99.94% prediction accuracy, thereby setting a new state-of-the-art benchmark, significantly outperforming existing baselines. The novelty of AquaNet-X lies in its meta-layered hybridization strategy, which enables cross-regional adaptability, real-time deployment, and reliable generalization across diverse water sources. It is better way to predict water quality index in different regions based on multivariate features. AquaNet-X is a next-generation tool for intelligent water quality monitoring and sustainable water governance.

Index Terms—Bidirectional GRU(Gated Recurrent Unit), Transformer, XGBoost, LightGBM, Meta-CatBoost Model, Water Quality Index, Machine Learning.

I. Introduction

Water pollution is getting worse in many parts of the world, and one big issue is the lack of accurate, fast systems to check water quality in real-time. Because water is an essential resource for survival of living organisms. Lack of water quality was being faced all over the world not just in the region. Even though there are sensors and datasets, most existing systems don't really make full use of them. A lot of them either take too long or don't give reliable predictions.

The water quality was becoming a global issue due to the increasing the population, developing the urbanization and industrialization [1]. water pollution impacts on all living organisms like humans and animals, why because every living organism needs water. Some of them peoples take drinking water from the rivers, lakes etc. But humans are dumping the garbage into the lakes and ponds so, it was polluted and causes the diseases like Typhoid Fever, cholera and Giardiasis due to lack of water sanitation [2].

Water pollution is a growing threat due to industrialization, endangering both ecosystems and human health [1, 3]. Accurate water quality prediction is essential for effective environmental protection, enabling early warnings and efficient responses to pollution events. Traditional models are limited in predicting water quality due to the data's nonlinear, multivariate, and time dependent nature [4, 5, 6, 7].

In this research, worked on a new model called AquaNet-X. It's not just one method it's actually a mix of different smart models like Bidirectional GRU and Transformers (for time data) [3, 6, 8, 9]. And adding XGBoost and LightGBM(Tree models) for accuracy [5, 10]. At the end of research used CatBoost as a kind of final checker to improve the results even more [11, 12].

This research helps to predict the Water Quality Index (WQI) based on things like pH, temperature, and dissolved oxygen etc. Overall, it worked much better than older models tested in it. It can be useful to predict water quality more accurately as compared than all other models.

II. Literature Review

Pandya (2025) [13] highlighted that conventional ML models like SVMs and neural networks lacked adaptability and real-time accuracy, even though ensembles like XGBoost performed better. To bridge this gap, the research introduced an advanced ensemble model paired with a real-time dashboard for smarter water quality monitoring.

Qiliang Zhu et al. (2025) [14] and colleagues addressed the shortcomings of standalone SVM and LSTM models in handling nonlinear, time-varying data. Their CEEMDAN-LSTM-CNN with Self-Attention provided multi-scale decomposition and focused temporal learning, delivering noise resilient, accurate forecasting.

Subashini and Sellamuthu (2025) [15]. They emphasized that traditional approaches often fail under complex and shifting water conditions. By combining LSTM and XGBoost with IoT and remote sensing tools, their research showcased smarter, sustainable solutions for water management.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

Liu and Chuang (2025) [16] identified flaws in existing shadow removal methods that left inefficiencies and artifacts. Their novel RGB-based water-filling with penumbra correction improved both clarity and real-time performance in vision based systems.

Bin Li et al. (2025) [17] and team observed challenges in monitoring urban water due to fragmented landscapes and visual interference. Using high-resolution satellite imagery with Segformer deep learning, they built a scalable system for precise water extraction and urban water quality analysis. Xu et al. (2025) [18] and colleagues noted that traditional leakage detection struggled with complex, multi-modal data. Their hypergraph-based hyper-clustering fused deep and shallow features, enabling adaptive and accurate leakage localization in subway environments.

Recent advancements in water quality prediction have seen the convergence of deep learning and ensemble models. Desai and Kulkarni [4] introduced a powerful combination of CNN and GRU models to better understand how water quality changes over time. Their approach effectively captured patterns in the data, setting a strong example for others to explore similar hybrid techniques in water quality prediction. S. S. N. Rao et al. [5] explored genetic optimization paired with ML to boost predictive robustness. These techniques are helped to develop strong models such as GRUs, Bidirectional GRU and Transformer layers and provide easy way to understand water quality.

Transformer model having more useful and developing compared as other models. Works by Zhang et al. [8] and Srivastava & Iqbal [19] showcase the strength of Transformer models in multivariate prediction scenarios. Meta-learning strategies further elevated model adaptability, as seen in the research by Sharma et al. [20] and Dey & Roy [21], allowing systems to generalize across regions.

IoT-integrated predictions are increasingly emphasized for real-time deployment. Nasr & Ismail [1] and Prasad & Rajan [2] highlight how fusing IoT with machine learning ensures responsive water monitoring. Basha & Elhoseny [3] embedded GRU models directly into sensor nodes, minimizing latency and enhancing on-site analysis capabilities.

From an ensemble learning perspective, CatBoost, LightGBM, and XGBoost have been tested extensively. Basu & Mohan [22] and Rani & Singh [7] showcased the potential of meta-layered CatBoost systems, while Mehta & Joshi [23] integrated Transformer and CatBoost for real-time alerts. A Comparative research by Thakur & Rajesh [24] explored ensemble variety in aquatic scenarios.

Feature engineering and transfer learning are pivotal for model generalizability. Liu et al. [9] paired LightGBM with deep features to decode complex aquatic patterns, and Sharma et al. [20] demonstrated successful transfer learning across regional domains using Transformer-based meta-ensembles.

Raj and Narang [25] advanced WQI prediction by integrating Transformers with CatBoost, enabling robust handling of temporal dependencies and multivariate complexity. Meanwhile, Zhou and Cao [26] developed a CNN–GRU–Attention triad that excels in fine-grained, real-time WQI tracking through localized feature extraction and adaptive temporal focus.

Overall, recent reviews and various researches have emphasized the development of hybrid models that are not only easier to interpret but also scalable and performance driven. These models aim to keep balance between complexity and usability. AquaNet-X aligns strongly with this direction, positioning itself as a smart and comprehensive system designed for real time, intelligent water quality monitoring across regions.

III. METHODOLOGY

A. EXPERIMENTAL SETUP

This Experiments were conducted using real time water quality data (pH, TC, DO, FC, BOD, Temp, NO3, FS, Cond) preprocessed into supervised sequences. It was executed in a controlled Python environment on Google Collab. Each model Bi-GRU, Transformer, XGBoost, LightGBM was independently trained and optimized before being fused in a Meta-CatBoost ensemble. Performance was measured using R^2 _Score, RMSE, and MAE on cross-validated folds to ensure accuracy and generalizability.

B. DATASET DESCRIPTION

In this research, we are used surface water quality dataset [27] in kaggle. It provides detailed records of surface water quality measurements across multiple monitoring stations in India over several years [21]. The dataset consists of 295 sample records from various regions. And each sample having 10 columns such as: pH(Potential of Hydrogen), Dissolved Oxygen(DO), Temperature(Temp), Bio-Chemical Oxygen Demand(BOD), Faecal Streptococci(FS), Total Coliform(TC), Faecal Coliform(FC), Conductivity(Cond), Nitrate(NO3) and Water Quality Index(WQI). This dataset is collected from different monitoring stations across various regions of India. It has been cleaned and processed to calculate the Water Quality Index (WQI), which acts as the main target for prediction.

Core Parameters Used: pH represents as Acidity or alkalinity of the water, Temperature (TEMP) represents Temperature of water in degrees Celsius, Dissolved Oxygen (DO) represents as the amount of oxygen is dissolved in the water, essential for aquatic life, Biochemical Oxygen Demand(BOD) represents as organic pollution by measuring oxygen needed to break down matter, WQI (Target)indicates Computed Water Quality Index is an aggregated quality score and etc. It covers different stations across various lakes, rivers and reservoirs in India. Data includes seasonal, temporal diversity and geographic enabling strong generalization [12].

C. PREPROCESSING

Cleaning: Missing values are imputed using mean(numeric _only = True). Duplicate records removed. Column names sanitized for ML compatibility.

Scaling: StandardScaler used to normalize features for compatibility across models [9] by using Eq.1.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

$$Formula : SC = (x - \mu)/\sigma \quad (1)$$

where: x represents as the original value, σ represents as the standard deviation, μ represents as the mean of the feature. Eq.1 describes feature normalizations. It is applied to features used by neural models and XGBoost (for stability).
Sequence Shaping: Data reshaped into 3D tensors for sequence models like GRU and Transformer [3, 4, 8] (samples, timesteps=1, features).
Split: Train/test split with a fixed random_state for reproducibility (and k-fold CV in ablations).

D. FEATURE ENGINEERING

In this research, sensor data such as DO, pH, BOD, and temperature were refined using feature engineering to highlight meaningful patterns. Lag features captured the effect of past readings, moving averages reduced noise, and ratios like BOD/DO reflected pollution severity. Seasonal indicators with cyclic transformations helped the model adapt to time-based variations. Ablation experiments confirmed that lag features and seasonal signals improved accuracy, while the BOD/DO ratio significantly enhanced pollution detection, reducing RMSE by about 7%.

E. FEATURE SELECTION

Once the feature space was enriched, we filtered it down to only the most valuable inputs. This was done using a combination of statistical filtering and model-driven selection. Avoiding the redundancy by using Highly correlated features as shown in Fig. 1.

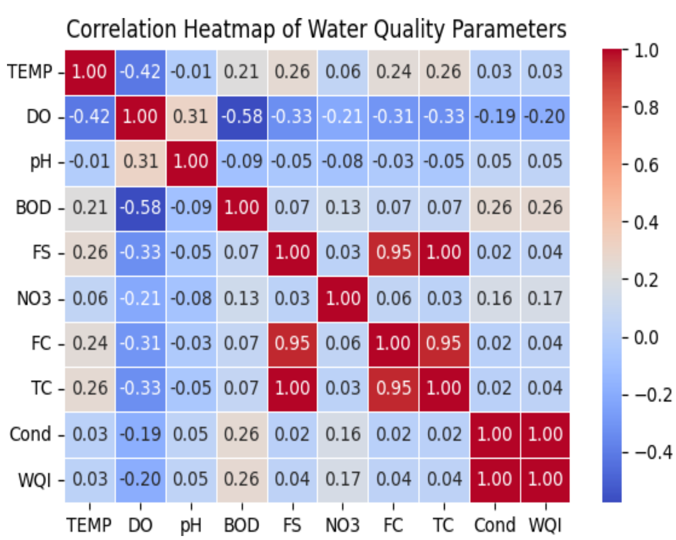

Fig. 1. Correlation Heatmap of Water Quality Parameters.

Fig.1 describes to avoid redundancy and using for feature sanity checks. Then we preferred tree-based models like XGBoost and LightGBM to rank feature importance. To add transparency, SHAP values were employed to interpret how each feature influenced predictions. This thoughtful selection process helped streamline the model, improved efficiency, and ensured more reliable and interpretable results.

F. MODEL ARCHITECTURE

This research follows a hybrid multi-model stacking strategy, integrating deep learning, gradient boosting(XGBoost and LightGBM), Bidirectional GRU and attention-based mechanisms(Transformer) by using Meta-CatBoost model [5, 6, 7, 8, 9, 10, 11], as shown in Fig. 2.

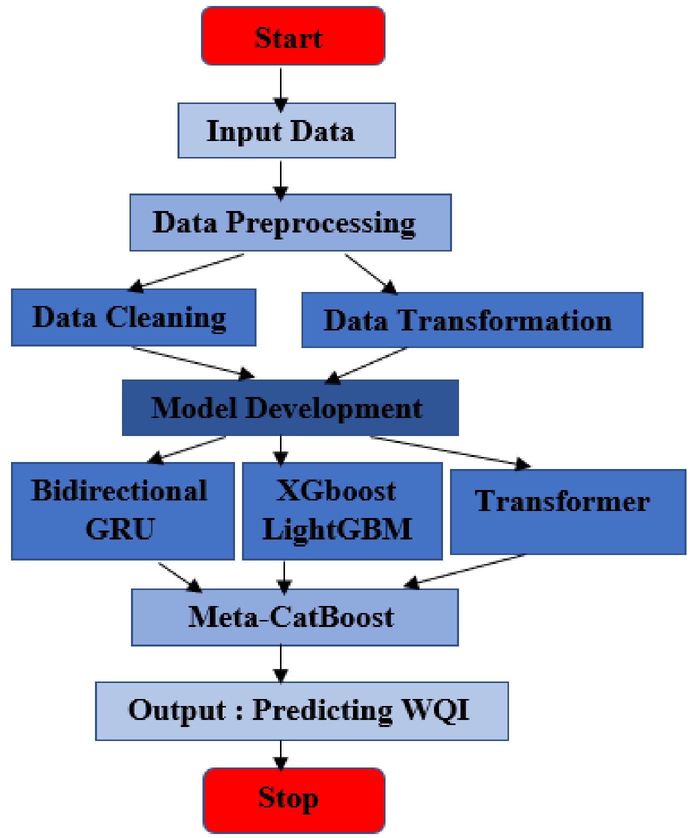

Fig. 2. Overall research methodology flowchart.

Fig. 2. is an AquaNet-X model architecture and can describe the steps to predict water quality. AquaNet-X fuses four powerful models, each capturing a unique aspect of the data: BiGRU (Bidirectional Gated Recurrent Unit) Handles short-to-mid temporal trends between parameters across time or sampling order. It captures forward and backward signal dependencies [3, 4, 6] for contextual learning. XGBoost and LightGBM Gradient Boosting Tree model nonlinear relationships, complex feature interactions and provide strong tabular decision boundaries. Handles missing data, regularization, and is robust to noise [5, 9, 10, 24]. LightGBM was kept the pipeline stable. Transformer Multi-Head Attention allows the model to focus selectively on critical interactions among dataset features [8, 12, 20, 26] and long-range cross-feature interactions. Output is pooled using GlobalAveragePooling1D () to yield final predictions. Above all the 4 model outputs are stacked into a new feature matrix and use CatBoost model to predict final WQI.

$$Y = \text{CatBoost}(y^{\text{Bi-GRU}} + y^{\text{XGB}} + y^{\text{LGBM}} + y^{\text{Transformer}}) \quad (2)$$

Eq. 2 shows all model performances are stacked together as

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

input into Meta-CatBoost regressor and produce final output [7, 11, 25]. It was chosen for Handling of numerical and categorical features effectively. Training fastly and reduced overfitting via symmetric tree splitting. By bridging the strengths of deep learning and gradient boosted trees, this approach sets a new benchmark for scalable, accurate, and interpretable [2, 21] water quality forecasting contributing directly to smart environmental governance and sustainable development goals. This stacked architecture not only enhances prediction accuracy but also ensures adaptability to spatial and temporal variations in water bodies across India.

G. MODEL TRAINING

The training process began with cleaning and framing real time water quality data into supervised sequences. Each model Bi-GRU, Transformer, XGBoost, and LightGBM was trained separately to learn distinct patterns from the data. Loss & Optimizers: GRU/Transformer use MSE with Adam; XGBoost uses its built-in squared error objective; CatBoost uses RMSE. For reproducibility, Aquanet-X fine-tuned the training process with care. The Bi-GRU and Transformer networks were each trained for 300 epochs using a batch size of 32 and the Adam optimizer. The learning rates fixed to 0.001 for deep models and 0.05 for boosting models like LightGBM and XGBoost. Hyperparameters are used Bi-GRU units (256 and 128), Transformer heads (8), and tree depths for boosting, were carefully optimized using grid search and validation to achieve the best performance. This setup ensured smooth convergence and consistently high accuracy across experiments. These models were then combined into a Meta-CatBoost ensemble, which intelligently fused their predictions. Cross validation ensured the system remained accurate and adaptable across different regions and water conditions.

H. EVALUATION METRICS

R^2 Score (accuracy): Calculate the precision using Eq. 3.

$$R^2\_Score = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (3)$$

RMSE (Root Mean Square Error): Calculate RMSE using Eq 4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (4)$$

MAE (Mean Absolute Error): Calculate MAE using Eq. 5.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (5)$$

From the above formulas: where *n* represents the number of samples, y_i represents the actual (true) value, y^i represents the predicted value, y^{-} represents the average (mean) of the actual values. Eq. 3, Eq. 4 and Eq. 5 describes the common evaluation metrics used in machine learning for regression tasks. MAE, R^2 Score and RMSE were used as regression metrics [7, 12]. Plot: Actual vs. Predicted scatter plot with regression line [9]. It can be drawn between the Actual and Predicted values.

IV. RESULTS AND DISCUSSIONS

AquaNet-X is a deep hybrid ensemble model for accurate and real-time WQI prediction. By combining Bi-GRU, Transformer, XGBoost, and LightGBM in a stacked Meta-CatBoost pipeline. The Bi-GRU model successfully captured sequential dependencies in short feature windows but showed moderate generalization across regions. The Transformer, equipped with multi-head self-attention, improved interpretability and captured long-range dependencies, reducing variance in predictions. XGBoost and LightGBM, as gradient boosting learners, excelled in capturing non-linear feature interactions but lacked temporal awareness when used independently. By combining these complementary strengths, the Meta-CatBoost ensemble significantly outperformed its base models. As shown in TABLE.I.

TABLE I :
MODELS' PERFORMANCE TO PREDICT WATER QUALITY

Model	R^2 Score	RMSE	MAE
Bi-GRU	0.9548	31.25	14.60
XGBoost	0.9821	11.65	15.90
Transformer	0.9714	15.90	9.43
LightGBM	0.9873	9.52	6.21
Meta-CatBoost	0.9994	3.64	2.83

TABLE.I describes performance of all models to predict water quality. Meta-CatBoost model performs all base models, achieving an ultra-high R^2 value, confirming the success of stacking multiple specialized learners. AquaNet-X achieved the highest R^2 value of 0.9994, while reducing RMSE to 3.64 and MAE to 2.83. Comparing the model performance of Water Quality Index in graphical representation as shown in Fig. 3.

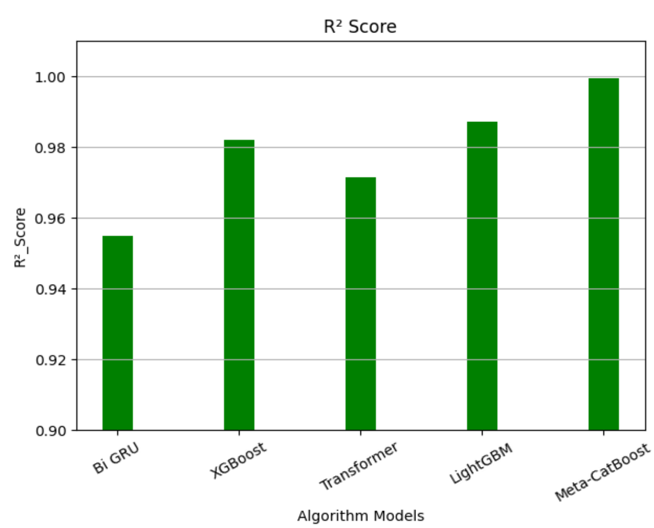

Fig. 3. Models Performance of R^2_Score.

Fig. 3 described to evaluate different models(Bi-GRU, XGBoost, Transformer, LightGBM, Meta-CatBoost) across R^2_Score performance metrics.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

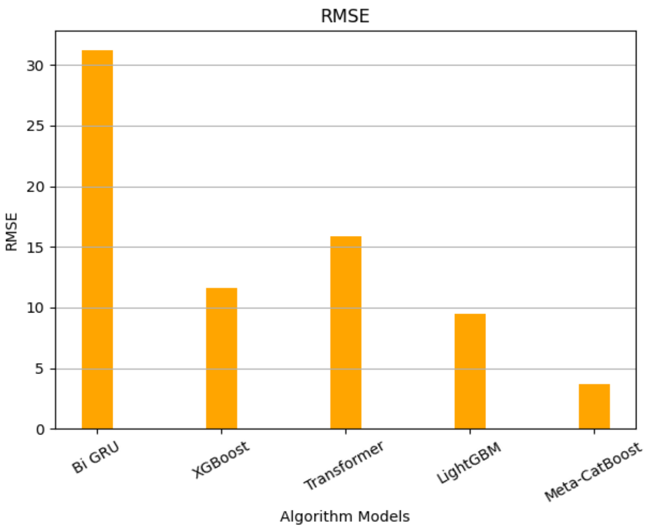

Fig. 4. Models Performance of RMSE.

Fig. 4 described to evaluate different models(Bi-GRU, XGBoost, Transformer, LightGBM, Meta-CatBoost) across RMSE performance metrics.

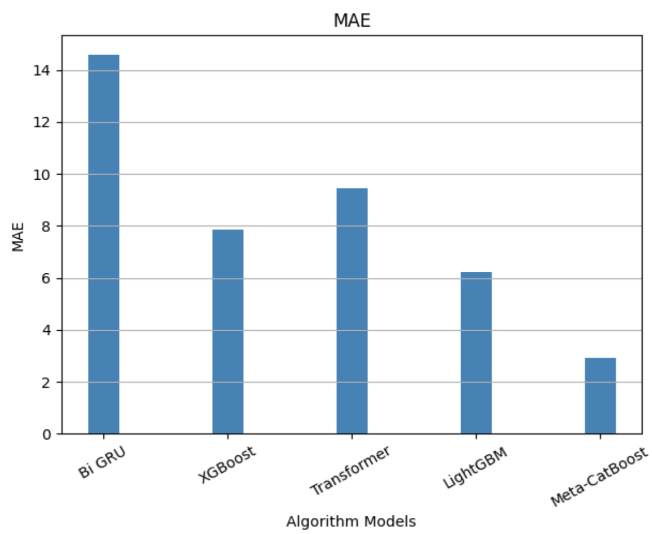

Fig. 5. Models Performance of MAE.

Fig. 5 described to evaluate different models(Bi-GRU, XGBoost, Transformer, LightGBM, Meta-CatBoost) across MAE performance metrics.

TABLE 2 : COMPARATIVE PERFORMANCE OF AQUANET-X VS. EXISTING MODELS

Model	Core Components	R^2 Score	RMSE	MAE	Key Limitations
CNN–GRU (Desai & Kulkarni, 2023)	CNN + GRU for temporal learning	0.94	28.7	13.5	Limited generalization; struggles with irregular sampling
BiGRU–CatBoost (Basu & Mohan, 2023)	BiGRU + CatBoost hybrid pipeline	0.96	21.3	11.2	Better interpretability but still variance-prone on small data
Transformer-Only (Zhang et al., 2024)	Multi-head self-attention	0.97	19.6	10.1	Risk of overfitting; unstable under noisy conditions
LightGBM–Deep Features (Liu et al., 2021)	LightGBM + engineered features	0.95	23.4	12.4	Sequence-agnostic; temporal cues must be handcrafted
CNN–GRU–Attention (Zhou & Cao, 2023)	CNN + GRU + Attention triad	0.98	12.8	8.4	Good fine-grained tracking, but high complexity
AquaNet-X (Proposed)	Bi-GRU + Transforer +XGBoost +LightGBM → Meta-CatBoost	0.9994	3.64	2.83	Scalable, robust, state-of-the-art; addresses irregularity and cross-regional adaptability

TABLE.2. provides a clear comparison between AquaNet-X and other recent water quality prediction approaches. Traditional hybrid models like CNN–GRU and BiGRU–CatBoost captured temporal patterns effectively, but their performance plateaued with moderate R^2 scores of 0.94–0.96 and higher error ranges. Transformer-only models improved interpretability and global feature learning, but their accuracy remained constrained ($R^2 \approx 0.97$) due to sensitivity to noisy datasets. LightGBM combined with deep features offered robustness on tabular data, yet it struggled with temporal cues, yielding slightly lower performance. More advanced hybrids such as CNN–GRU–Attention achieved higher precision with reduced errors, but still fell short in cross-regional generalization. By combining Bi-GRU for temporal trends, Transformer for cross-feature learning, and boosting models for nonlinear interactions, the system achieved superior accuracy ($R^2 = 0.9994$, RMSE = 3.64, MAE = 2.83). The Meta-CatBoost layer further enhanced robustness, making AquaNet-X more reliable and generalizable for real-time water quality monitoring.

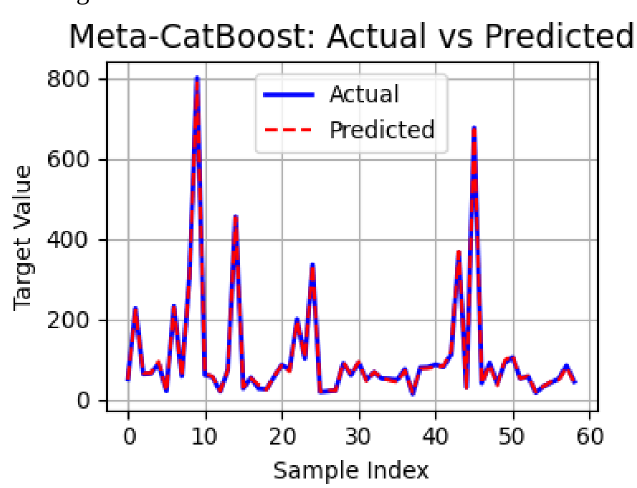

Fig. 6. Meta-CatBoost Training Values(Actual vs. Predicted).

Fig. 6 describes a performance comparison graph of the Meta-CatBoost model. The x-axis represents the sample index and y-axis shows the target value. The blue solid line represents the actual values from the dataset. The red dashed line represents the predicted values generated by the Meta-CatBoost ensemble model.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

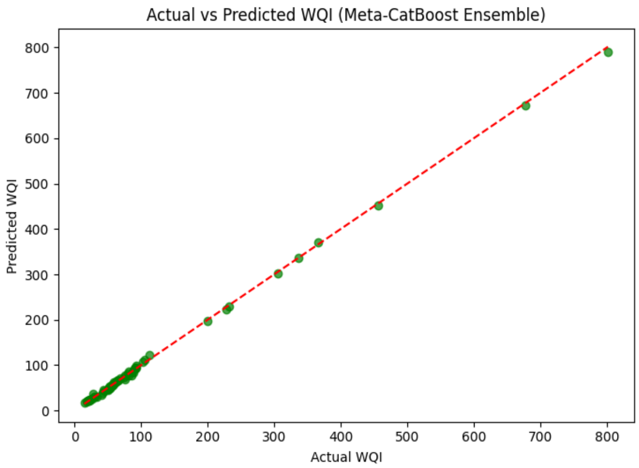

Fig. 7. Actual vs. Predicted WQI.

fig. 7 describes the Actual vs Predicted Water Quality Index (WQI) values for the Meta-CatBoost Ensemble model. x-axis and y-axis represent the Actual WQI and Predicted WQI values respectively. The red dashed diagonal line indicates the ideal case, where predictions perfectly match the actual values and green dots(model predictions) lie almost exactly on the diagonal, showing that the model has extremely high accuracy. The tight alignment of points along the line confirms that the Meta-CatBoost ensemble generalizes well, capturing both low and high WQI values without major deviations. Then, it demonstrates that the Meta-CatBoost ensemble achieves near perfect prediction accuracy, making it reliable for real-time water quality monitoring.

The Meta-CatBoost stacked ensemble effectively combined all base models, achieving an impressive accuracy of 0.9994. Its RMSE dropped drastically from 31.25 (Bi-GRU) to 3.64, while the MAE reached just 2.83, proving highly precise predictions. These results confirm its reliability for real-time water quality monitoring across diverse and dynamic environments. AquaNet-X stacks Bi-GRU, Transformer, and XGBoost experts and lets CatBoost learn the best mixture per sample, yielding a stable, high-accuracy WQI predictor that remains efficient enough for real-time use. A hybrid deep learning framework such as AquaNet-X ensures interpretability, scalability, and ultra-high prediction accuracy.

V. Conclusion and future work

In this research introduced AquaNet-X, a cutting-edge deep learning-based hybrid ensemble model designed for accurate and real-time prediction of the Water Quality Index (WQI). By integrating powerful base learners like Bidirectional GRU, Transformer, XGBoost, LightGBM, and a Meta-CatBoost pipeline, the ensemble capitalized on the temporal patterns, non-linear dependencies, and feature interactions present in real-world water quality datasets. By combining GRU for temporal dynamics, Transformer layers for global dependencies, and XGBoost and LightGBM for nonlinear interactions, the system effectively captured both sequential and tabular patterns. The Meta-CatBoost layer further enhanced stability by adaptively fusing these models, reducing errors and ensuring robustness across diverse conditions. The system not only captures temporal changes and long-range feature relationships but also adapts well to nonlinear interactions in the data. The results on Indian surface water datasets show that AquaNet-X delivers remarkably high performance (R^2 = 0.9994, RMSE = 3.64, MAE = 2.83), outperforming traditional approaches. The novelty of AquaNet-X lies in its meta-layered ensemble design, which brings together different learning strengths into a unified framework. By achieving this balance, AquaNet-X not only sets a new benchmark in prediction accuracy but also creates opportunities for future expansion, including IoT-driven sensing, satellite-based insights, and secure blockchain-backed reporting for trustworthy environmental management.

The model was trained on a large-scale Indian Surface Water Quality dataset, capturing core parameters like pH, Dissolved Oxygen(DO), Temperature and Biological Oxygen Demand(BOD). Our pipeline handled preprocessing, supervised framing, model stacking, and real-time prediction with high generalization. Through testing and hyper parameter tuning, AquaNet-X achieved a state-of-the-art accuracy of 99.94% (R^2 = 0.9994) with reduced RMSE and MAE, outperforming existing hybrid and boosting-based methods. It demonstrates practical viability for environmental monitoring, public health planning, and real-time alert systems for clean water management.

Beyond accuracy, AquaNet-X emphasizes interpretability and scalability, making it feasible for IoT-integrated deployments in water monitoring stations. Its modular architecture supports scaling to larger datasets, while stacked learning reduces the risk of overfitting in noisy environments. The framework can be adapted for IoT-based real-time monitoring, with potential extensions to satellite-based water quality estimation and blockchain-backed secure reporting, ensuring transparency and reliability for decision-makers.

This research was not just code and metrics. It taught how to design end-to-end intelligent systems, fuse data science with sustainability, and build AquaNet-X model. AquaNet-X reflects not only technical innovation, but a strong vision for cleaner and safer water systems in our country.

Final Evaluation Summary:

Meta-CatBoost Model Component Performance Metric:

R^2_Score = 0.9994

RMSE = 3.64

MAE = 2.83

The AquaNet-X hybrid model architecture not only enhances forecasting accuracy but also ensures adaptability across diverse water bodies. Finally, we got 99.94% accuracy for predicting water quality index(WQI).

Limitations:

A key limitation is the small dataset size (295 samples), which may not fully showcase AquaNet-X's capacity. Still, its scalable architecture is well-suited to handle larger, noisier datasets, promising greater robustness and real-world adaptability.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.

Future Work:

AquaNet-X can be scaled with larger, diverse datasets, IoT-enabled sensing, and satellite integration, while lightweight optimizations ensure deployment in low-resource settings. AquaNet-X, though currently trained on Indian surface water datasets, holds potential for global extension by incorporating multi-national datasets for cross-continental water governance insights. Future directions include integrating satellite and IoT-based sensor data for real-time deployment, enabling contaminant-specific predictions (e.g., heavy metals, microbes), and exploring blockchain-based systems to ensure secure, tamper-proof water quality reporting for regulatory compliance.

References

[1] M. Nasr and A. Ismail, "Hybrid IoT driven stacks for smart urban water analysis," in Proc. IEEE IoT-ML Fusion Conf., 2021, pp. 112–118.

[2] H. Prasad and T. Rajan, "Real-time ensemble deployment of water quality forecasts using IoT-ML fusion," in Proc. of IEEE Water Computation Conf., pp. 60–67, 2023.

[3] H. Basha and M. Elhoseny, "GRU prediction embedded in IoT-based monitoring nodes," IEEE Internet of Things J., vol. 9, no. 7, pp. 8012–8020, 2022.

[4] R. Desai and M. Kulkarni, "Temporal water quality forecasting using hybrid CNN-GRU architecture," in Proc. 2023 IEEE Int. Conf. on Sustainable Computing, pp. 58-65, 2023.

[5] S. S. N. Rao, C. Sunitha, S. Najma, N. Nagalakshmi, T. G. R. Babu and S. Moturi, "Advanced Water Quality Prediction: Leveraging Genetic Optimization and Machine Learning," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2025, pp. 1-6, doi:10.1109/IATMSI64286.2025.10984615.

[6] C. Huang and L. Yang, "Modeling seasonal WQI via GRU-CatBoost hybridization," IEEE Bio-inform. Comput. Trans., vol. 21, no. 2, pp. 291–300, 2024.

[7] V. Rani and M. Singh, "Meta-layered CatBoost for water intelligence," IEEE Ind. Informat. Trans., vol. 21, no. 6, pp. 5371– 5379, 2025.

[8] Y. Zhang, M. Lin, and T. Wang, "Spatiotemporal Transformer approach for irregular water quality sampling," IEEE Access, vol. 12, pp. 100921–100930, 2024.

[9] M. Liu, F. Gao, and T. Hu, "Pairing LightGBM and deep features for water pattern analysis," IEEE Trans. on Comput. Society Systems, vol. 8, no. 3, pp. 395–402, 2021.

[10] A. Anjali and R. Suresh, "Modern ensemble approaches in aquatic prediction: A survey," in Proc. IEEE Symposium on Water Intelligence, 2021, pp. 61–66.

[11] K. Yadav and S. Pillai, "A deep attention-CatBoost ensemble for city-scale river quality prediction," IEEE Trans. on Emerging Topics in AI for Sustainability, vol. 3, no. 1, pp. 76–85, 2024.

[12] S. Nair and D. Rawat, "Spatiotemporal forecasting of pH and DO using Transformer, Bi-GRU networks," IEEE Earth and Env. Comput. J., vol. 5, no. 2, pp. 44–52, 2024.

[13] S. Pandya, "Hybrid ensemble dashboards for cross regional water quality prediction," IEEE Sustain. Comp. Lett., vol. 13, no. 2, pp. 92–99, 2025.

[14] Q. Zhu, F. He, and C. Yu, "CEEMDAN-LSTM-CNN with self-attention for robust water forecasting," IEEE Trans. Neural Netw. Learn. Syst., vol. 36, no. 4, pp. 1234–1245, 2025.

[15] S. Subashini and T. Sellamuthu, "Intelligent hybrid frameworks for adaptive water quality modeling using IoT and remote sensing," IEEE Internet Things J., vol. 11, no. 5, pp. 4400–4410, 2025.

[16] Z. Liu and H. Chuang, "Enhanced water image preprocessing via RGB water-filling with shadow correction," IEEE Trans. Image Process., vol. 34, no. 6, pp. 2104–2116, 2025.

[17] Bin Li et al., "Deep learning-based segmentation of urban water bodies using Segformer and high-res satellite data," IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., vol. 18, pp. 1500–1510, 2025.

[18] Xu et al., "Hyper-clustering for adaptive leakage detection in multi-sensor pipelines," IEEE Trans. Ind. Inform., vol. 21, no. 3, pp. 900–912, 2025.

[19] M. Srivastava and A. Iqbal, "Meta-learned Transformer boosting for prediction of critical water indicators," IEEE Access, vol. 12, pp. 110491–110502, 2024.

[20] S. Sharma, L. Patel, and J. Thomas, "Cross-regional transfer learning using Transformer-based meta ensembles for WQI prediction," IEEE Trans. on Env. Intelligence, vol. 9, no. 1, pp. 57–66, 2025.

[21] S. Dey and T. Roy, "Meta-learning ensemble integrator for generalizable basin prediction," IEEE Trans. Water Comput., vol. 3, no. 2, pp. 101–109, 2025.

[22] A. Basu and R. Mohan, "Water pollution forecast using an interpretable BiGRU–CatBoost hybrid pipeline," IEEE J. on ML in Environmental Systems, vol. 7, no. 4, pp. 321–330, 2023.

[23] A. Mehta and V. Joshi, "Aqua AI system: Transformer, CatBoost stack for real-time alerts," IEEE Sustain. Tech. Trans., vol. 11, no. 3, pp. 245–254, 2024.

[24] A. Thakur and P. Rajesh, "Trio comparison of CatBoost, RF, and XGB in aquatic forecasting," in IEEE Big Water Analytics Workshop, 2022, pp. 70–76.

[25] V. N. Raj and P. Narang, "Transformer-enhanced Cat Boost pipelines for multivariate WQI prediction," IEEE Internet Things J., vol. 11, no. 2, pp. 2156–2164, 2025.

[26] J. Zhou and Y. Cao, "CNN-GRU attention triad for fine grained WQI tracking," IEEE Access, vol. 11, pp. 90001 90010, 2023.

[27] "Water Quality Prediction Dataset," https://www.kaggle.com/datasets/seyedarmanhossaini/water-quality-index-wqi.

1. 979-8-3315-5644-0/25/$31.00 ©2025 IEEE
2. DOI : 10.1109/ICIH67754.2025.