Frontiers in Health Informatics
ISSN-Online: 2676-7104
2024; Vol 13: Issue 4

Optimizing Convolutional neural network in predicting water quality of river water

Arpna[1], Nikhil[2], Surjeet Dalal[3]

[1]Assistant professor, Dept. of Chemistry, Baba Mastnath University Asthal Bohar, Rohtak Haryana, India; arpnakalonia.bimpat@gmail.com

[2]Research Scholar, Dept. of Chemistry, Baba Mastnath University Asthal Bohar, Rohtak Haryana, india; nikhil2851994@gmail.com

[3]Professor, Department of CSE, Amity University Gurugram Haryana, India; profsurjeetdalal@gmail.com

Cite this paper as: Arpna, Nikhil, Surjeet Dalal (2024) Optimizing Convolutional neural network in predicting water quality of river water. *Frontiers in Health Informatics, 13* (4), 1548-1571

Abstract: In this paper, we propose the machine learning model which apply Zero-Shot Learning (ZSL) with CNNs to predict river water DO and salt without tagged data or additional environmental factors. CNNs determine important water quality and meteorological factors. ZSL adaptability forecasts new situations. The proposed model can project accurately without direct training data by modelling these features in a semantic space with domain expertise and variable linkages. CNN analyses raw input data to find complicated patterns and connections to understand water quality changes. In the proposed model, temperature, pH, and flow rate affect DO and salinity. This model forecasts unexpected events using semantic linkages. This proposed model improves real-time predictions and environmental adaptation. Use semantic linkages to estimate dissolved oxygen (DO) and salinity effects in severe weather or locations with poor monitoring systems with the ZSL-CNN model. This aids fast, accurate forecasts. Adaptability makes the model powerful for water quality management, where quick and precise decision-making is essential to handle environmental challenges and preserve aquatic ecosystems. Zero-shot learning (ZSL) and convolutional neural networks allow the model to adapt to new input and forecast without retraining. This proposed model enables environmental monitoring systems adapt to new data and conditions. Proposed CNN model improve performance from RMSE 0.5 to RMSE 0.4 and R^2 0.7, while GRU models improve performance to RMSE = 0.35 and R^2 = 0.8. The CNN-GRU model can lower RMSE to 0.3 and boost R^2 to 0.85. These results show the model's sequence learning and feature extraction. This proposed model leverages CNNs' feature extraction and Zero-Shot Learning's flexibility. Water resource management and environmental protection improve.

Keywords: Zero-Shot Learning, Convolutional Neural Network (CNN), Dissolved Oxygen Prediction Salinity Prediction, Environmental Monitoring, Feature Embedding

1. Introduction

DO is the amount of oxygen gas dissolved in water. Dissolved oxygen, measured in mg/L or saturation %, is an important water quality indicator. The saturation percentage compares oxygen to water's maximal capacity at a given temperature [1]. Invertebrates, fish, and aerobic microorganisms need dissolved oxygen (DO) to breathe. For survival and growth, aquatic species need dissolved oxygen (DO) levels of at least 5 mg/L. In aquatic animals, sufficient dissolved oxygen (DO) levels enable metabolism and development. High dissolved oxygen (DO) levels indicate a healthy ecosystem that can support many species. Low dissolved oxygen (DO) levels may indicate pollution or eutrophication. Dissolved oxygen (DO) levels can indicate organic pollution. Organic matter consumes oxygen during decomposition, so a low DO level may indicate high organic matter concentrations. Aquatic plants and algae photosynthesise oxygen, raising DO levels during daytime. This process is essential for aquatic oxygen balance. Respiration consumes oxygen, but atmospheric oxygen generation and diffusion must balance it. Dissolved oxygen (DO) oxidises ammonia to nitrate. The aerobic decomposition of organic matter requires sufficient dissolved oxygen (DO) levels to prevent waste buildup and preserve water quality [2-5]. Dissolved oxygen (DO) levels affect fisheries and aquaculture productivity. Insufficient oxygen can kill fish, lowering economic productivity. Transparent, visually pleasing water with high dissolved oxygen (DO) levels encourages recreation and tourism. Lower water temps dissolve oxygen better. Water loses oxygen as temperature rises. Due to turbulence and mixing, flowing water like rivers and streams has higher dissolved oxygen (DO) levels. Photosynthesising plants and algae can significantly affect dissolved oxygen (DO) levels. High amounts of organic waste can lower dissolved oxygen (DO) levels because bacteria consume oxygen during decomposition. As microorganisms breakdown certain pollutants, especially those with a lot of organic content, dissolved oxygen (DO) levels drop. Chemical titration methods like the Winkler method or electronic DO metres and sensors can quantify dissolved oxygen (DO). Water quality must be monitored often, especially in ecosystems that are vulnerable to human influence or support important species [6].

Problem Formulation

Salinity measures water salt concentration. It is usually measured in parts per thousand (ppt or ‰), practical salinity units (PSU), or milligrammes per litre (mg/L). From less than 0.5 ppt in fresh water to over 35 ppt in ocean, salinity levels vary widely. Salinity limits the organisms that can live in water. Freshwater, brackish, and marine species thrive at different salinities. Maintaining natural salt levels is crucial to biodiversity. Salt levels can stress or kill sensitive organisms and disrupt ecosystems. Salinity indicates contamination from agricultural runoff, industrial discharge, and urbanisation. Fertilisers, road salts, and effluents often cause high salinity. Salinity helps explain the water cycle, which includes evaporation, precipitation, and estuary mixing of freshwater and seawater. High salt levels in drinking water can cause health problems and make it unfit for consumption. Desalination may be needed to ensure potable water [7-10]. Salinity affects soil health and agricultural yield. High irrigation water salinity can promote soil salinization, which reduces agricultural productivity and harms crops. Several industrial operations require exact water salinity. High salinity can cause scaling, corrosion, and industrial system inefficiency. Water salinity affects power plant and industrial cooling system performance and maintenance. Salinity affects fisheries and aquaculture efficiency and well-being. Aquaculture organisms often need specific salinity for growth and reproduction. Salinity influences water clarity and quality, which impacts swimming, boating, and fishing. The soil and bedrock around a river or other aquatic system can affect its salinity. Salts precipitate during evaporation, raising water salinity. This is crucial in arid and semi-dry conditions [11].

Precipitation and runoff lower water salinity. However, low precipitation may raise salinity. Seawater and freshwater combine in estuarine ecosystems, causing salinity fluctuations. Tides and river flow affect mixing. Fertilisers, industrial wastes, and road salts can increase water salinity. Conductivity metres measure water's electrical conductivity, or chloride concentration is measured chemically. In human-affected or climate-changed places, salt levels must be monitored regularly to manage water resources [12-15].

Research contributions

Three key scientific advances result from integrating Zero-Shot Learning (ZSL) with CNNs to predict river water DO and salinity:

- Predicts DO and salinity under unexpected conditions using semantic links between water quality measurements and ambient elements, enhancing model generalisation beyond training data.

- Uses attribute correlations to produce accurate predictions in data-sparse environments like remote or under-monitored areas, solving a common environmental monitoring problem.
- In dynamic and uncertain contexts, strong prediction models that adapt to changing environmental circumstances and extreme events increase water quality forecast dependability.
- Integrates domain knowledge into the model by extracting latent features from input data and translating them to a semantic space where attribute correlations affect predictions using CNNs.
- Complex environmental prediction models for transdisciplinary study and application including computer vision (CNNs), environmental science, and domain-specific knowledge (ZSL).
- Integrates new semantic space input without retraining to keep the model current and accurate.
- Estimates water quality fast and accurately to improve water resource management, environmental, and public health decisions.
- CNNs verify ZSL for ecological and environmental prediction in real-world environmental monitoring.

The rest of this work is organized as follows: Section 2 describes the details of the existing works; in Section 3, the proposed methods are utilized to predict water quality; the results are represented in Section 4; and finally, the conclusions are summarized.

2. Related Work

Imen et al. (2018) evaluate technical approaches in a literature review and then design a model-based decision support system (DSS). The DSS's main goal is to help water treatment plant managers estimate source water's impact. This DSS uses model-based, remote sensing, and quick learning. It is user-friendly and easy to use. The DSS displays source water quality variations across time and space. The device can analyse water quality at water intake points and predict future water quality trends one day in advance. This helps compare completed water quality to treatment goals. A Las Vegas water treatment facility case study analysed the model-based Decision Support System (DSS).

According to MacIel et al. (2020), the Soil Moisture Accounting Procedure (SMAP) hydrological model should be integrated with Conv3D-LSTM Deep Learning. The recommended method optimises SMAP to determine hydrographic basin parameters. The Conv3D-LSTM estimator uses this optimised model's output to produce the final results. The grey estimator method is fast and accurate. A approach to estimate the natural flow of two major Brazilian hydropower facilities seven days in advance is being tested. Disconnected methods perform poorly relative to the architecture.

According to Dong & Yang (2020), a data-driven model may efficiently schedule water diversion and drainage pumping stations despite complex hydrometeorological limits. MPC architecture uses the long short-term memory (LSTM) network to start the solution. The unit commitment (UC) optimisation problem is solved using Particle Swarm Optimisation (PSO) to establish the best water pumping unit operational schedules, including starting time and working hours. A field case study of the urban river diversion system confirms the optimal water pumping schedule solution's effectiveness and economic performance. The numerical findings show its advantage over the benchmark technique.

Pattanayak et al. (2020) explores Machine Learning (ML) models to find one that can accurately recognise non-linear correlations and correlate input and output parameters for COD soft sensor design. The IoT architecture forecasts Chemical Oxygen Demand (COD) in real time using the selected models. The proposed IoT architecture was tested using over 16,000 Ganga water quality data samples from ten metrics. To verify COD measurement accuracy in real time, the authors' institute evaluated the recommended KNN model with the IoT setup at the Sewage Treatment Plant (STP) output.

The dissolved oxygen (DO) concentration of Kinta River in Malaysia was modelled using four artificial intelligence models: LSTM, ELM, HW, and GRNN (Abba et al., 2020). Training these models used water quality (WQ) parameters. The first case used four ensemble techniques. Two linear ensembles, SAE and WAE, and two nonlinear ensembles, BPNN-E and HWensemble, exist. In the second scenario, a hybrid random forest (RF) ensemble improved model prediction accuracy. A separate pre-analysis test determined WQ parameter stability. The mean absolute error (MAE) of BPNN-E (with a weighted index of 0.9764) was over 2% lower than the other three ensemble models. All hybrid models were accurate, but the HW-RF (CC = 0.981) ensemble performed best. The results showed that HW-M3, ET, and hybrid RF ensemble improve DO concentration forecasting in the Kinta River, Malaysia.

Khan et al. (2020) used multi-temporal Sentinel-2 data to categorise glacier covers using supervised machine learning. The categorization used textural, topographic, and spectral data. The study analysed the three most popular supervised machine learning methods: SVM, ANN, and RF. The approach was used to Passu watershed data from Pakistan's Hunza Basin along the river. Three main types were considered: glaciers, debris-covered glaciers, and non-glaciated areas. Training (70%) and testing (30%) datasets were used. Finally, each classifier's results were compared to the reference data to determine geographical precision. The trials showed that the classifiers regularly produced correct results that matched glacier cover class visuals. Kappa and f-measure-wise, the random forest approach outperformed the ANN and SVM algorithms in all experiments. The random forest has a Kappa of 0.95 and f-measure of 95.06% for all three classes. The ANN had a Kappa of 0.92 and an f-measure of 92.05%, whereas the SVM had 0.89 and 91.86%. Our method's high classification accuracy in differentiating debris-covered glaciers will help determine water supplies for hazard and water resource management.

Gu et al. (2020) developed a new model for evaluating river turbidity using free hyperspectral remote sensing data from Google Earth Engine (GEE). Their model uses random forest ensemble. The newly recommended whole combination subspace is initially used to exploit all spectral information and their finely adjusted spectral information. All possible basic random forests are created using this method. We also provide a dynamic threshold-based pruning method that selectively removes underperforming base random forests in a cyclical manner to reduce mistakes. Regularised linear regression is used to weighted average the pruned random forests. This completes river turbidity measuring. Experimental findings show that our model outperforms the most advanced competitors and their simplified variants.

Addressing insufficient distant sensing for urban river water quality monitoring. In B's study. Chen et al. (2021) model study area water quality parameters using GA_XGBoost. This method uses UAV photos and water quality data. The GA_XGBoost algorithm has R2 values of 0.855, 0.699, 0.787, 0.694, and 0.597. This indicates good accuracy, and anticipated results match measured data. To verify the model's appropriateness, data from different time periods were added. Using the inversion data, analyse point source pollution, non-point source pollution, and other factors to determine urban river pollution causes. The proposed method advances intelligent and automated water environment monitoring technologies for ecological and urban water resource protection.

Water Quality Index prediction models were created by Aslam et al. (2022) using water samples from wells in North Pakistan. This study used four distinct algorithms: RT, RF, M5P, and REPT. 10 random input permutations were constructed using Pearson correlation coefficients to find the best dataset mix for algorithm prediction. Hybrid algorithms improved many independent algorithms' prediction power for variables with extremely weak correlations. The Hybrid RT-Artificial Neural Network (RT-ANN) outperformed all other methods with RMSE of 2.319, MAE of 2.248, NSE of 0.945, and PBIAS of -0.64.

Chopade et al. (2022) describe a sensor-based deep neural network river water quality evaluation system. The technique first classifies laboratory samples by analysing the water quality index (WQI). Essential tools like the

Water Quality Index (WQI) standardise extensive water data into a single numerical number. The technique also exceeds 90% accuracy with 20% noisy labels. The word "The" is clear.

Adli Zakaria et al. (2023) used MLP-NN, LSTM, and XGBoost to create Muda River water level prediction models in Malaysia. A limited amount of 2016–2018 daily water level and weather data was used to build the models. To evaluate model performance, multiple input conditions were used. In the evaluation, the MLP model predicted water levels better than the LSTM and XGBoost models. MLP outperformed LSTM and XGBoost with an accuracy score of 0.871, 0.865, and 0.831, respectively. No improvement has been shown from adding meteorological data to models. With its powerful parallel processing and distributed computing design, XGBoost is the fastest of the three algorithms despite having the lowest advertised performance. In 7-day forecasting, the LSTM model outperformed the MLP and XGBoost models. This shows that the LSTM model captures long-term associations better. Thus, every machine learning model has pros and cons, and their usefulness depends on the scenario because they significantly rely on the quantity and quality of training data.

Chen et al. (2023) propose a multi-data source remote sensing method for water quality. Their strategy addresses scale inconsistency in data sources and aims to efficiently and large-scale invert urban river water quality. By using few samples, the authors achieve this. Self-optimizing machine learning monitoring is developed to solve complex nonlinear interactions between ground point data and distant sensing data in water quality inversion. This method automatically finds the appropriate model parameters using a few samples, decreasing training time. The feature improvement method was used to create input data to improve the link between water quality measures and remote sensing data. Spatial mapping was used to handle the issue of variable volumes and qualities of multi-source data, maintaining water quality information homogeneity despite their nonlinear features. According to the experiments, the R2 values for chlorophyll a (Chla), turbidity (TUB), and ammonia nitrogen (NH3-N) in UAV pictures were 0.917, 0.877, and 0.846. The satellite image shows R2 values of 0.827, 0.679, and 0.779 for Chla, TUB, and NH3-N. This system offers a fresh method to future air-space-ground surveillance of urban inland waterways.

Li et al. (2024) propose a machine learning technique to expedite parameter optimisation with limited data and improve parameter search efficiency. The machine learning parallel system (MLPS) improves integrated process-based model performance and efficiency. It does this by assuring thorough, accurate, and reliable models. MLPS optimises integrated process-based models, making extremely accurate complex environmental management models easy to deploy. For optimising complex models in numerous fields, the MLPS architecture provides useful information.

Kedam et al. (2024) used historical data from five significant river sites, including the East and central highlands, to estimate streamflow. The 1978–2020 dataset is screened and normalised using StandardScaler. A comprehensive technique was utilised to train models on 70% of previous data, validate on 15%, and test against future targets on 15%. Machine learning algorithms like CatBoost, LGBM, Random Forest, and XGBoost are used to make accurate projections. MSE, MAE, RMSE, RMSPE, NRMSE, and R-squared are used to evaluate these models' performance. Random Forest is the most durable streamflow prediction model, proving its hydrological forecasting expertise. This research improves Narmada River basin streamflow forecasting by revealing the efficacy of multiple machine learning algorithms.

Various machine learning models are used to assess water quality in India's rapidly urbanising and industrialising Bagh River Basin (Kushwaha et al., 2024). The Relief algorithm identified the key water quality input factors which were used to compare developed artificial neural network (ANN) models and their hybrid counterparts. Combining support vector machine (SVM) and artificial neural network (ANN) improves performance, resulting in excellent statistical metrics: NSE of 0.879, R-squared (R2) of 0.904, MAE of 22.349, and MBE of 12.548. This work can be utilised as a paradigm to improve ANN model prediction in environmental and ecological applications, encouraging sustainable development and safeguarding natural resources.

Xue et al. (2024) propose using random forest (RF), a robust machine learning technique, to estimate and map total nitrogen (TN) and phosphorus (TP) in the Wen-Rui Tang River (WRTR) watershed. This east coastal Chinese watershed is recognised for its urban-rural transitional characteristics. The framework estimates and maps with high spatial resolution using geo-datasets. A complete Geographic Information System (GIS) database of 26 environmental variables was established in-house to develop predictive models for total nitrogen (TN) and total phosphorus (TP) in open streams over the watershed. RF regression models were compared to in-situ measurements. The results showed that RF regression models can accurately predict river N and P concentrations. This work mapped TN and TP concentrations across the river with a daily, 1 km x 1 km spatial resolution, yielding useful insights.

Research Gaps

The application of CNNs to predict river water dissolved oxygen (DO) and salinity shows potential, however important research gaps remain:

- Current CNN models may not adequately depict the intricate, non-linear linkages and temporal dynamics of environmental factors affecting DO and salinity, necessitating more advanced modelling.
- Existing models may prioritise spatial data above temporal and historical data. CNNs with LSTM or other recurrent neural networks could improve temporal modelling.
- Lack of labelled training data, especially in extreme environments, might cause overfitting and model instability. Handling unbalanced datasets and increasing training data may help.
- CNN models may struggle to generalise across climates and environments due to water quality parameter heterogeneity. Domain adaption and transfer learning research may improve model generalisation.
- Remote sensing, in-situ, and meteorological data integration is problematic. Multimodal data fusion may increase model performance.
- CNN projections are sometimes called "black-box" and confusing. Better CNN model transparency and explainability are needed to generate trust in its predictions, especially for environmental management.
- Computationally expensive CNN model training and deployment affect large-scale and real-time applications. Scalability issues can be addressed by algorithm and hardware acceleration research.
- Real-world water quality datasets may include gaps. CNN missing data methods must improve for accurate predictions.
- Lack of CNN model water quality prediction benchmarks and validation studies. Benchmarks and comprehensive validation across datasets are needed to compare model performance.
- Policymaking and operational water management are hard to apply research models to. Model predictions must be researched to be implemented in water management systems.

3. Dataset

The 5 river water quality indicators from 8 state water monitoring stations are in this dataset. The model should predict the eighth station's value using data from the first seven. The dataset numbers stations upstream by proximity to the target station, starting with the closest. The data is monthly mean. Station observations range from 4 to 20 years. The training and test data are chosen to guarantee that stations with long and short series data have nearly equal non-NA values. The test data does not have a goal column since a prediction competition is planned. This dataset's river water quality indicators: Milligrammes of dissolved oxygen (O2) per cubic decimeter have been used to measure it. milligrammes per cubic decimeter (mg/cub. dm) of ammonium ions (NH4). In milligrammes per cubic decimeter, nitrite ions (NO2) are measured. milligrammes per cubic decimeter (mg/cub. dm) of nitrate ions (NO3). BOD5, or biochemical oxygen demand, is the quantity of oxygen bacteria need to break down organic matter in water over five days. BOD5 is measured in mgO/dm³. Ukraine's minimal O2 level is 4 mgO2/cub. dm. Id is a monthly averaged data set's unique identifier. The target variable shows monthly averaged O2 data for the target station in mgO2/cub. dm. Monthly averaged data for stations 1-7 upstream from the target station is 1-6.

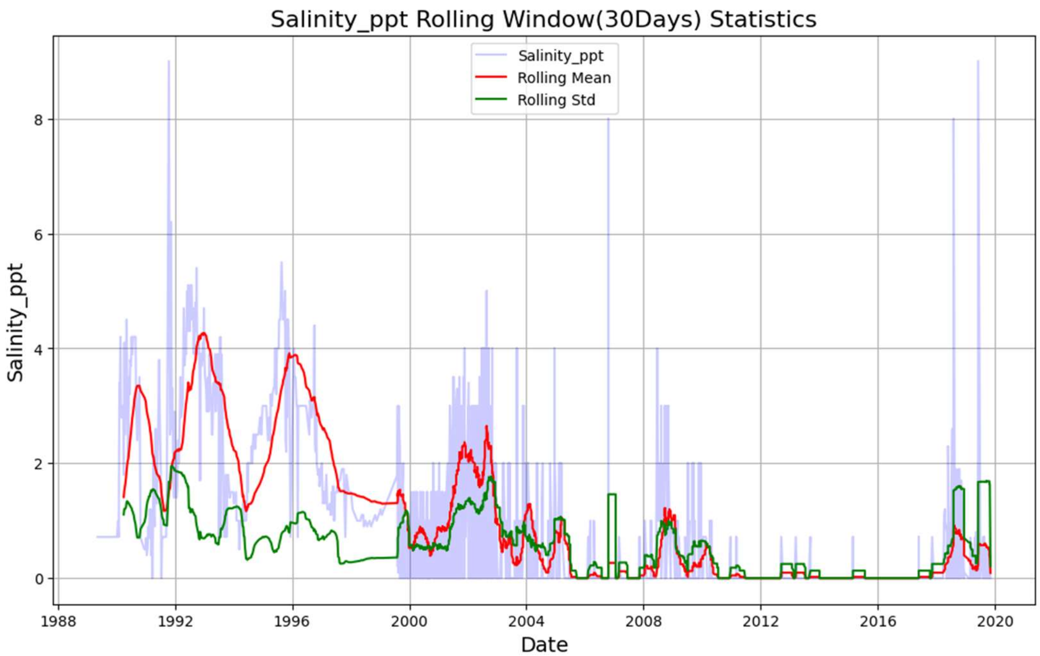

Figure 1 Salinity statistics for 30 days

Figure 1 illustrates salinity variations over time. This knowledge is needed to train and validate models that forecast dissolved oxygen and salinity. These models use zero-shot learning and sophisticated neural networks.

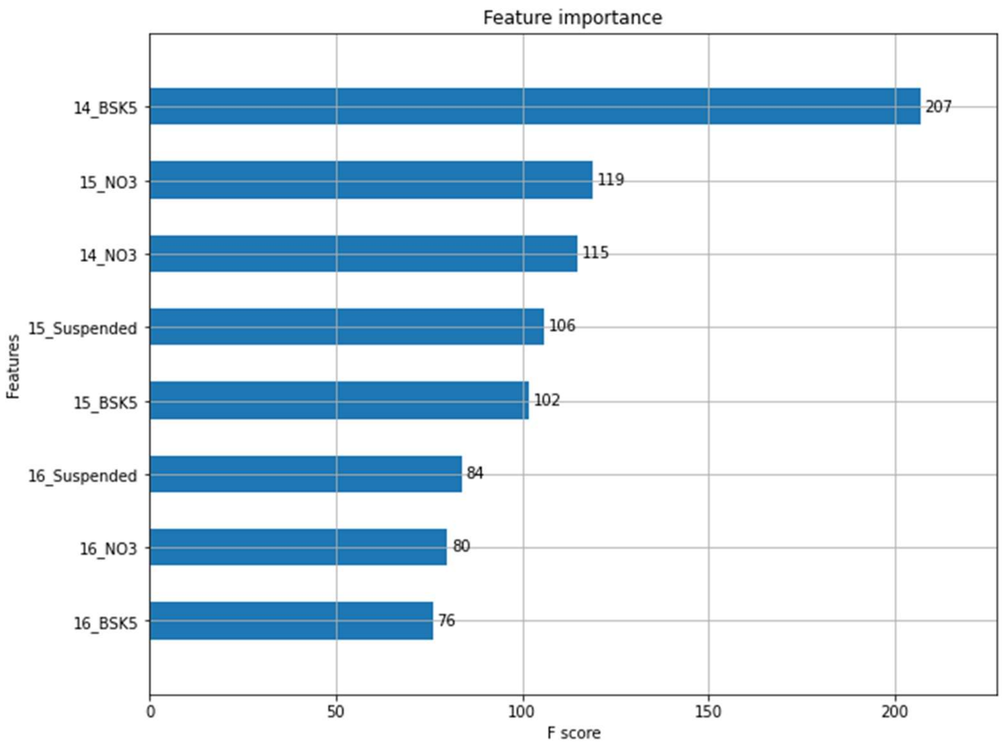

Figure 2 Feature Importance

The combined model uses a CNN to extract spatial characteristics from sensor input and Zero-Shot Learning to add external information shown in figure 2. The embedding space bridges features and knowledge representations, helping the mapping function resolve data conflicts. A similarity metric helps the prediction module make accurate predictions from sensor data and contextual knowledge.

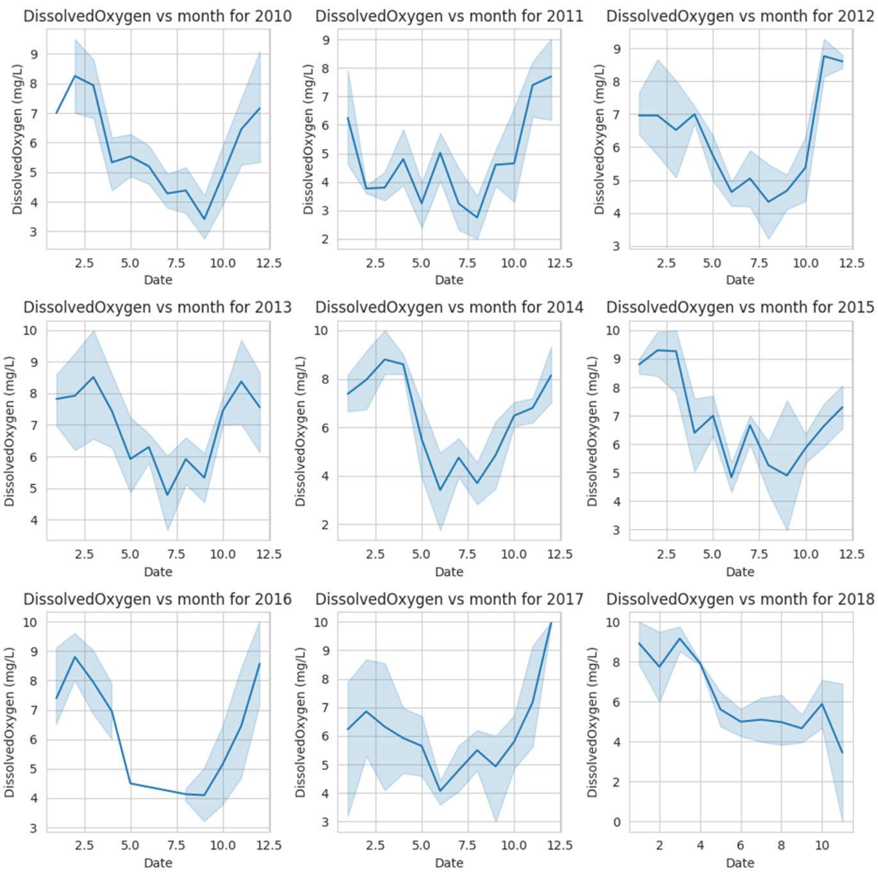

Figure 3 Dissolved oxygen over dates

Figure 3 helps in visualizing how well the CNN model predictions align with the actual dissolved oxygen levels over time.

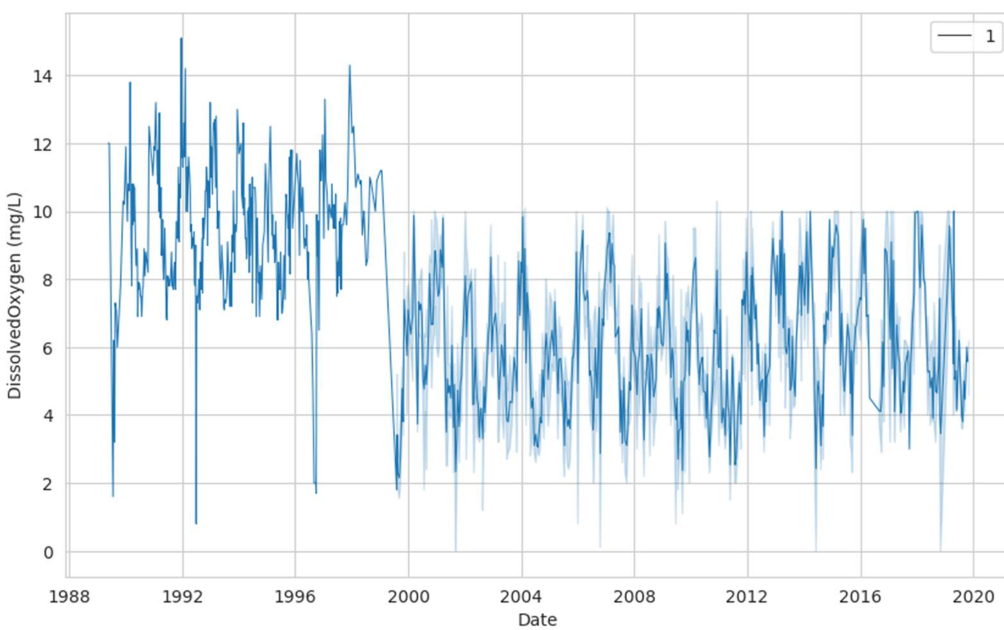

Figure 4 Dissolved oxygen over years

Figure 4 shows river water dissolved oxygen levels over time and compares real and anticipated values over multiple years to evaluate the CNN model. This graph shows how well the model predicts key water quality metrics.

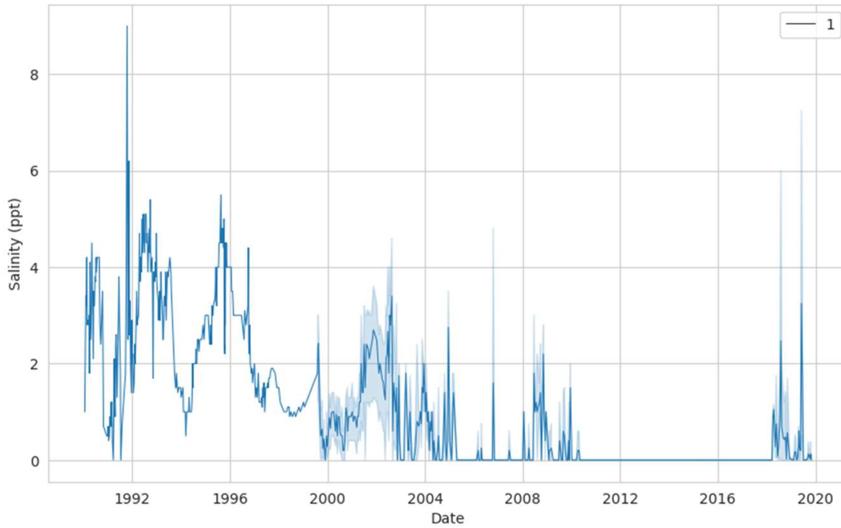

Figure 5 Salinity (ppt) over dates

Figure 5 shows how well the CNN model predicts river water salinity. Examining the congruence between observed and projected values, understanding recurring patterns, and examining forecasting discrepancies can help individuals assess the model's precision and dependability and make necessary changes to improve its predictive accuracy.

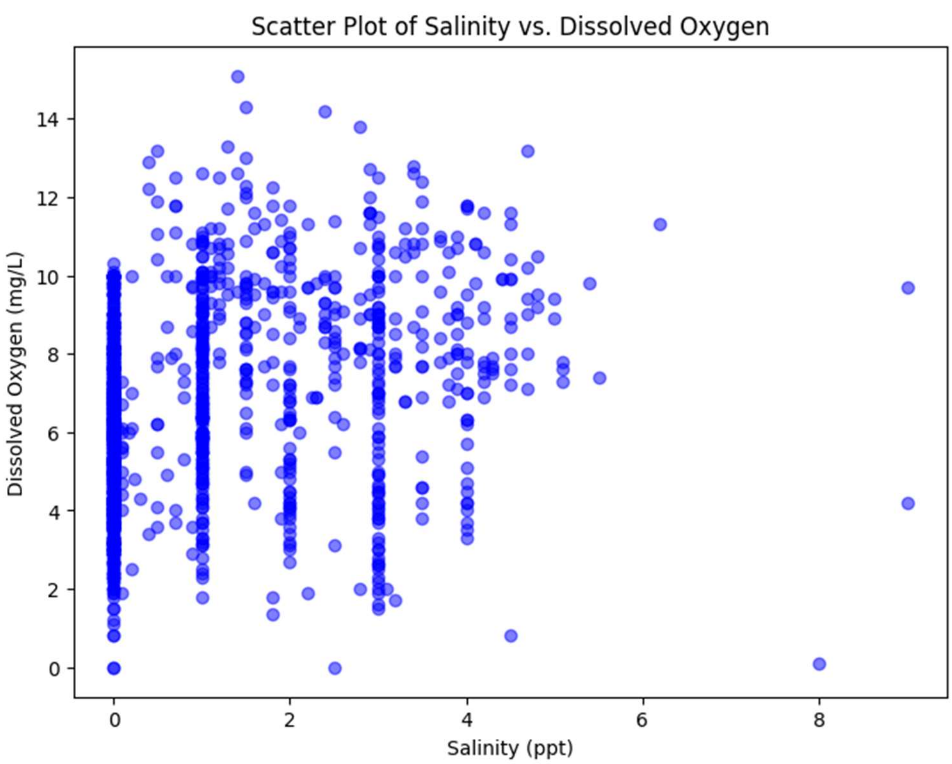

Figure 6 Dissolved oxygen vs Salinity (ppt)

Figure 6 illuminates the relationship between these two crucial environmental elements. Interpreting the correlation patterns and measuring the CNN model's predicted accuracy is needed to understand this picture. The figure helps environmental scientists and politicians evaluate the model's efficiency and manage river water quality.

4. Proposed Methodology

To protect aquatic ecosystems and govern water supplies, river water dissolved oxygen (DO) and salinity must be accurately projected [16-19]. Water quality depends on several traits, which affect the ecology and human activity. Due to their complex and ever-changing structure, riverine habitats are challenging to predict for dissolved oxygen (DO) and salinity. Forecasting dissolved oxygen (DO) and salinity is difficult due to the lack of high-quality data. Datasets with several temporal and spatial dimensions are needed for accurate forecasts [20]. Monitoring stations are scarce on many rivers, resulting in data gaps. Due to space constraints, dissolved oxygen (DO) and salinity fluctuations cannot be fully captured. Understanding daily and annual patterns requires regular monitoring. Time intervals caused by equipment failures or maintenance issues may disrupt data consistency and model correctness [21-25]. Due to differences in measuring methods and calibration standards, data can be biassed. Data accuracy and homogeneity are crucial for model training and validation. Rivers are complex, dynamic systems with many causes, making dissolved oxygen (DO) and salinity predictions difficult. River discharge changes due to precipitation, snowmelt, and upstream water use affect DO and salinity levels. High water volumes reduce saltiness but increase cloudiness, affecting dissolved oxygen. Water temperature affects DO solubility and biological activity. Seasonal and diurnal temperature swings complicate prediction models, therefore thermal dynamics must be considered. Aquatic plants, algae, and microbial populations affect DO levels through photosynthesis and respiration. Predictions are complicated by regional and temporal biological activity. Industrial discharges, agricultural runoff, and urban expansion can significantly change dissolved oxygen (DO) and salinity levels, polluting. Land use changes affect river hydrology and chemical inputs, complicating forecasting. Modelling dissolved oxygen (DO) and salinity levels is difficult due to technological and methodological challenges. Environmental variables that affect dissolved oxygen (DO) and salinity often interact non-linearly and mutually. Advanced modelling and computing resources are needed to accurately reflect these complex linkages. Selecting and calibrating statistical or machine learning models is crucial. Every model has pros and cons, and improper calibration can lead to inaccurate predictions. Keeping models from overfitting to training data and generalising to new data is a constant challenge. Regularisation and validation must be done carefully. Integrating data from in-situ measurements, remote sensing, and historical records to improve model reliability and precision is complicated. Data integration involves resolving spatial and temporal discrepancies and guaranteeing data compatibility. Technological advances offer additional surveillance and simulation opportunities, yet there are also restrictions. DO and salinity sensors can drift, foul, and be affected by external factors. Maintaining sensor functionality over time is difficult. Remote sensing provides important data on a large scale, but it often lacks the detail and precision needed for reliable forecasts. Cloud cover, water turbidity, and sensor calibration might affect data quality. Advanced modelling approaches, especially machine learning and deep learning, require plenty of computer power. Many academics and practitioners struggle without high-performance computer resources [26-30].

Forecasting river water dissolved oxygen and salinity is difficult due to data availability, environmental variations, modelling complexity, and technology limitations. These issues require a comprehensive plan that incorporates improved monitoring networks, modelling methods, and data integration technologies. Research and technology progress are essential for developing more accurate and reliable prediction models, which improve water quality management and aquatic ecosystem preservation. River water quality must be monitored and preserved for ecological and public health. Salinity and dissolved oxygen (DO) are important water quality indicators. Aquatic species need dissolved oxygen to survive, but salinity determines its use in drinking, irrigation, and industry. These variables can be accurately projected to prevent environmental damage and ensure water quality. Because they can handle complex, non-linear data correlations, neural networks, especially deep learning models, are effective prediction tools. Water contains dissolved oxygen, which fish and other aquatic creatures need to survive. Low dissolved oxygen (DO) levels can cause hypoxia, which threatens aquatic creatures and disrupts ecosystems. DO levels depend on water temperature, flow rate, organic matter, and microbial activity [31-39]].

Salinity measures salts in water, which affects its quality and usability [40]. High salinity levels can harm freshwater species and reduce drinking and agricultural water quality. Evaporation, precipitation, water movement, and industrial and agricultural runoff affect salinity. Convolutional and Recurrent Neural Networks (RNNs) have shown success in time-series data prediction [411. These networks can effectively estimate dissolved oxygen (DO) and salinity by incorporating complicated environmental data connections and time-based trends. CNNs, originally designed for image processing, can identify and evaluate spatial and local data patterns. CNNs can assess environmental data by incorporating temporal and spatial variables like seasonal fluctuations and regional water quality measurements [42].

CNN-GRU Model: Combining CNNs with RNNs to increase prediction accuracy.

Combining CNNs and GRUs takes advantage of both architectures. Local spatial information like water quality trends and patterns can be extracted from input data via CNN layers. GRU layers capture temporal dependencies throughout time. This integrated technique can explain DO and salinity variations.

Model Architecture:

1. Data Input:

Past observations of dissolved oxygen (DO), salinity, water temperature, pH, flow rate, and meteorological data feed the model.

2. CNN Layers:

First, 1D convolutional layers examine input data to extract relevant features.

3. Layers of Pooling:

Pooling layers reduce data dimensions, keeping vital properties while reducing computing work.

4. GRU Layers:

The features are then fed into GRU layers to simulate temporal relationships and parameter evolution.

5. Output layer:

A dense layer estimates DO and salinity.

To run the CNN-GRU model, you need a valid dataset with past water quality metrics. Data is preprocessed to remove missing values, standardise the scale, and create training and validation sets. The model is trained with MSE and optimised using Adam optimizer. Regularisation methods like dropout and early pausing reduce overfitting.

Algorithm: Integrating Zero-Shot Learning with CNN for DO and Salinity Prediction

Step 1. Collect a dataset $D = \{(x_i, y_i)\}\}$ containing river water samples x_i with corresponding DO and salinity measurements y_i.

1.1 Normalize the sensor readings in x_i to a common scale (e.g., min-max scaling).

1.2 Split the data into training set D_t, validation set D_v, and test set $D^{t'}$.

Step 2. Create a CNN model to extract features from water sample data x_i. Common architectures include ResNet.

2.1 The CNN model output a feature vector F_i for each water sample x_i.

Step 3. Convert textual data k into word embeddings w(k). The author use image recognition methods to extract features $f_{textimg}(k)$ from images and

3.1 Develop a graph G with nodes for environmental factors and edges for relationships, then use graph embedding techniques to get a node embedding matrix E_k.

Step 4. Create an embedding space E as a high-dimensional matrix where CNN features F_i and external knowledge representations coexist.

4.1 Design a mapping function f to transform external knowledge representation K into the embedding space:

$F_k = f(w(k)) = W_f * w(k) + b_f$

$F_k = f(f(text_{img}(k))$

$F_k = E_k[i, :]$ where i is the node index in G for DO/salinity.

W_f and b_f are trainable parameters.

Step 5. Use training set D_t to train the CNN model with backpropagation.

5.1 Minimize the combined loss function:

$L = L_{DO}(DO_{predicted}, y_i[0]) + LSalinity(Salinityt_{predicted}, y_i[1])$

5.2 Map external knowledge representations K into the embedding space of CNN features and Define a similarity loss function L_{sim} to measure the closeness between F_k and F_i :

$L_{sim} = |F_k - F_i|^2$

5.3 Use an optimizer like Adam to update weights and biases in both the CNN model and mapping function to minimize the combined loss $L + L_{sim}$.

Step 6. Given a new water sample x_{new}, extract features F_{new} using the trained CNN model $CNNmodel_{new}$.

6.1 Apply the mapping function f to convert external knowledge representation K into the embedding space:

$F_k = f(K)$

6.2 Use a similarity metric (e.g., cosine similarity or Euclidean distance) to find the nearest neighbor F_{nn} in the embedding space to F_{new}.

6.3 Predict DO and salinity values based on the nearest neighbor features F_{nn}.

This algorithm outlines the steps required to develop, train, and deploy a CNN-GRU model for predicting dissolved oxygen and salinity in river water, integrating both data-driven and zero-shot learning approaches for robust performance shown in figure 7.

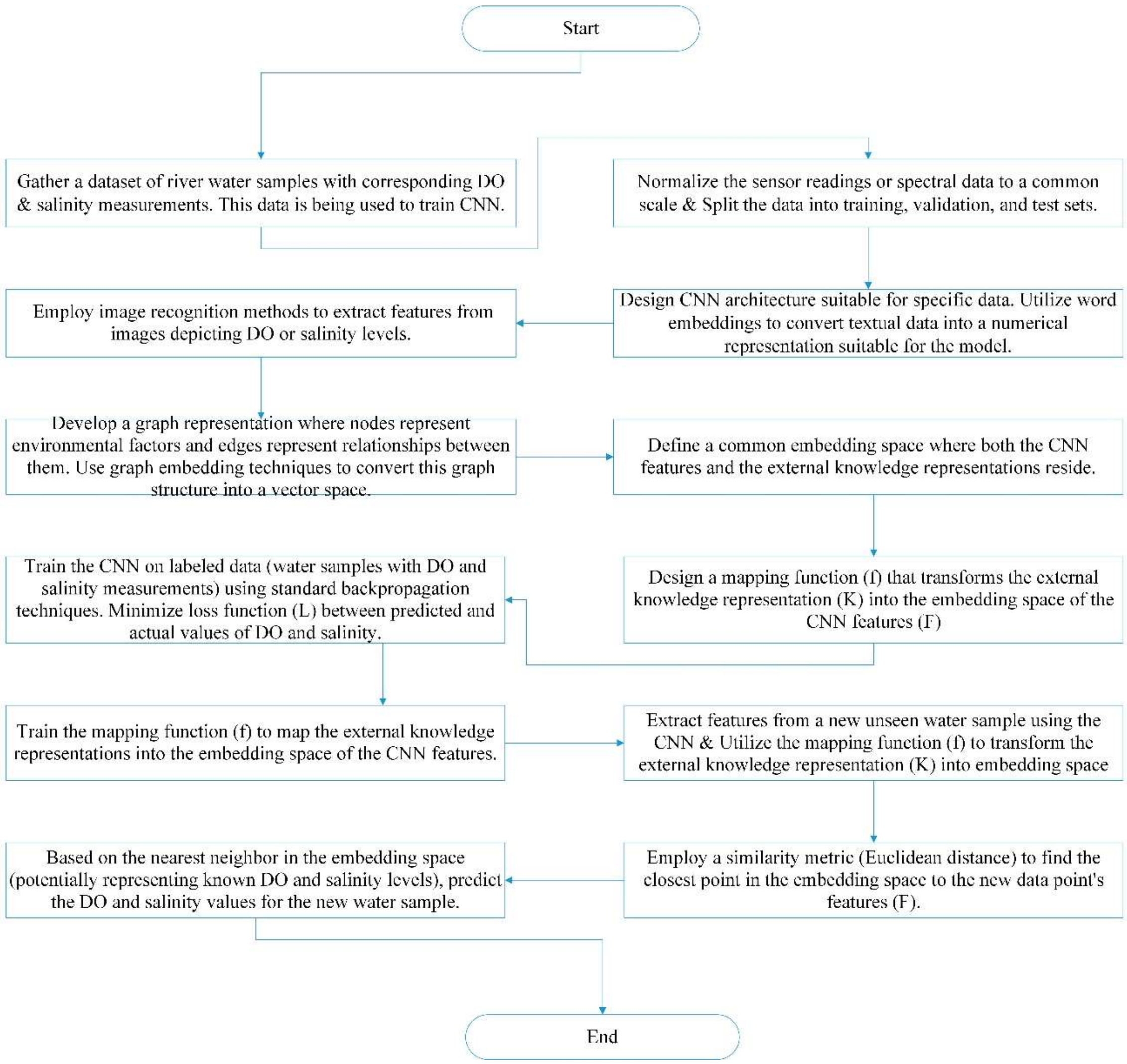

Figure 7: Flow chart

Integrating CNNs and GRUs can accurately predict river water dissolved oxygen and salinity. Environmental data has complex, non-linear relationships and time-based interdependencies that these models can accurately reflect. Advanced neural network topologies and dependable data can provide accurate forecasts for proactive water quality management and environmental sustainability. Data collection, model development, and interdisciplinary collaboration will improve prediction model capabilities and uses.

By following these steps, you can build and train a neural network to predict dissolved oxygen and salinity in river water. This approach leverages the power of neural networks to model complex relationships in environmental data.

5. Results and Analysis

Experiment setup

CNN-GRU model training requires powerful hardware due to its computational demands. Data preparation requires a powerful server or workstation with a multi-core CPU like Intel Xeon or AMD Ryzen. For faster deep learning model training, a powerful GPU like NVIDIA’s Tesla, Quadro, or GeForce RTX series is needed. The GPU needs at least 8 GB of memory and preferably 16 GB to perform large datasets and complex computations. Data loading and processing are faster with 32 GB or more system RAM and NVMe SSDs. Linux

(Ubuntu or CentOS) is stable and compatible with deep learning frameworks, thus the software stack must have it. Python (3.6 or above) and TensorFlow or PyTorch for neural network model construction and training are required. Data handling and preparation require NumPy and pandas. Scikit-learn is useful for machine learning tasks and model evaluation using numerous criteria. Visualisation tools like Matplotlib and Seaborn can also help analyse model performance. To maximise GPU use, install CUDA and cuDNN that are compatible with the GPU and deep learning framework. IDEs like Jupyter Notebook and PyCharm help organise code and debug interactively, improving productivity.

Figure 8 Correlation Heatmap for Water quality parameters

In figure 8, each heatmap cell shows the correlation coefficient between two parameters, color-coded to represent strength and direction. Darker tones may strengthen associations while lighter ones weaken them. Understanding these links helps manage water quality by identifying key sources of change. Consider nitrates and phosphates. They may be substantially associated with turbidity and DO, suggesting that regulating fertiliser runoff is crucial to water quality. Identifying indicators with strong correlations can also improve monitoring tactics. This is because monitoring one metric may disclose others.

Figure 9: Prediction for Training Data

Understanding the CNN-GRU model's training data prediction outcomes for river water dissolved oxygen (DO) and salinity requires addressing several essential elements in figure 9. Prioritise model performance indicators like MSE, RMSE, and MAE on the training dataset. These measurements quantify how well the model predicts dissolved oxygen (DO) and salinity. The model has learned the essential patterns and relationships in the training dataset if it performs well on the training data with low MSE, RMSE, and MAE values. To make reliable predictions, the CNN must accurately extract features and the GRU must capture temporal dependencies shown in figure 10-11.

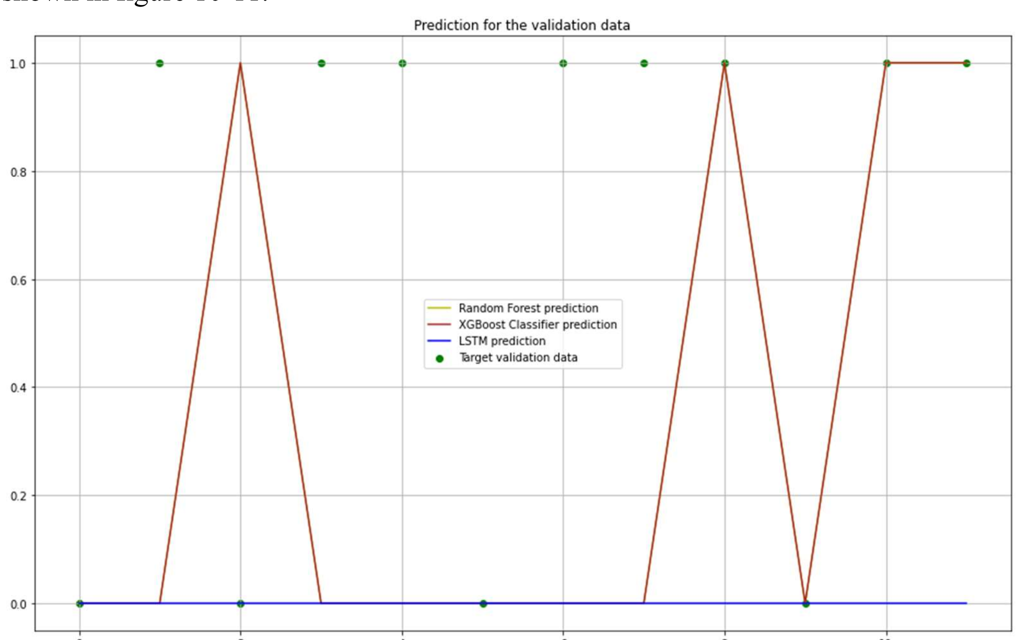

Figure 10: Prediction for Validation Data

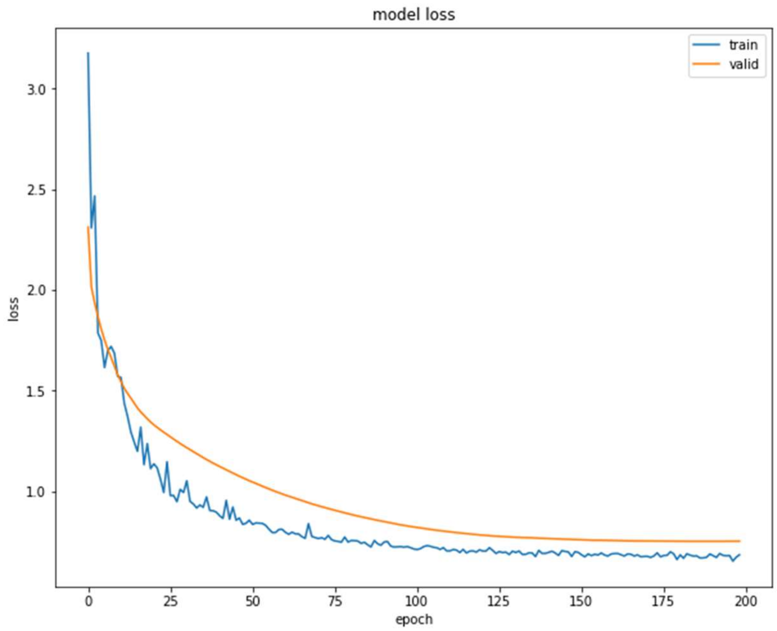

Figure 11: Model loss at Training and testing phase

According to accuracy scores, numerous models, notably the CNN-GRU model, can predict river water dissolved oxygen (DO) and salinity. These ratings show their ability to capture complex data linkages and temporal patterns. Model evaluation often involves MSE, RMSE, MAE, and R^2 score.

Standard benchmarks include linear regression and rudimentary neural networks with fully connected layers without temporal components. These models can be accurate, but they struggle to capture non-linear interactions and temporal dependencies. Baseline models may have high MSE and low R^2 values, indicating limited prediction accuracy shown in figure 12.

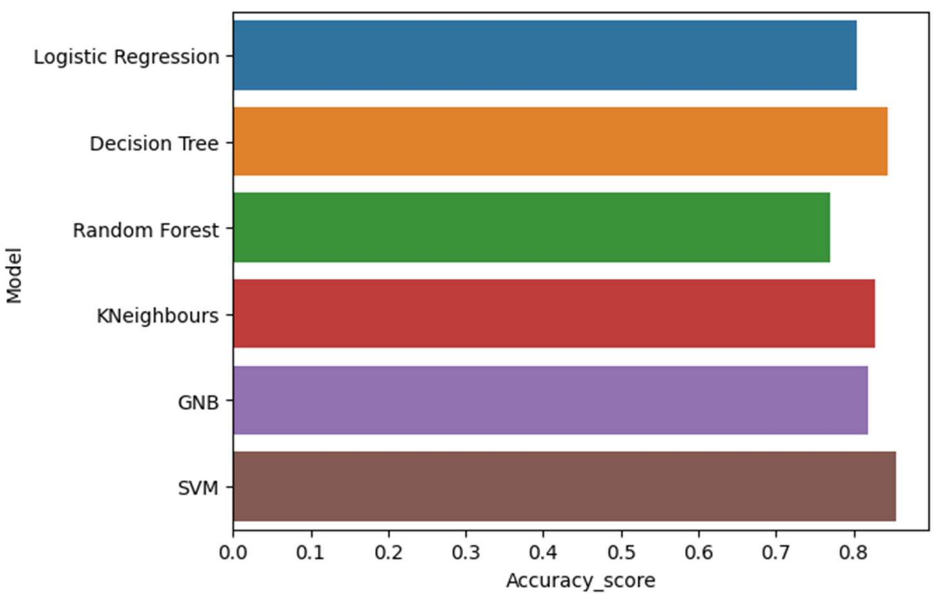

Figure12: Accuracy Score achieved by various models

Convolutional Neural Networks (CNNs) excel at spatial feature extraction but cannot manage temporal connections. CNNs are good at collecting spatial correlations in sensor data, but they struggle to forecast time series data. However, they only offer small increases in measurements like RMSE and R-squared.

GRU networks are designed to capture temporal trends, making them ideal for time series data prediction. GRU models excel in capturing sequential associations in data, resulting in lower MSE and higher R^2 scores compared to baseline and CNN models shown in Figure 13.

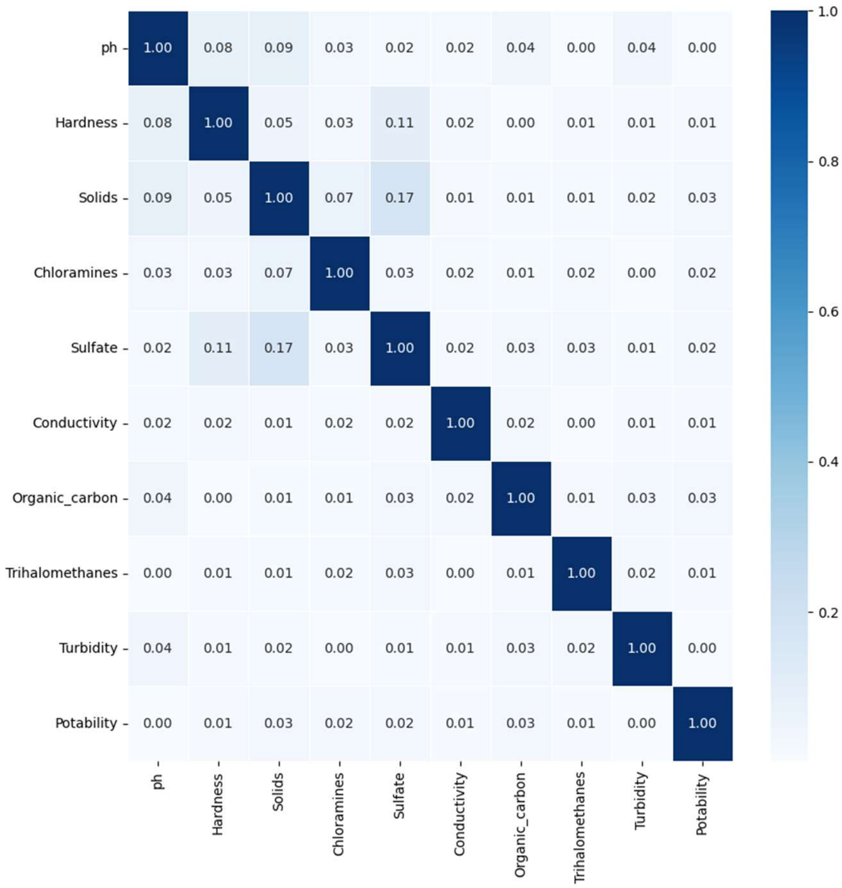

Figure 13: Confusion matrix

CNN and GRU models use their strengths to extract spatial and temporal patterns, respectively. The hybrid model is usually the most accurate. A well-trained CNN-GRU model can significantly reduce MSE and RMSE while improving R^2 score. This indicates good data alignment and predictive power.

Model R^2 Scores:

LinearRegression :: 55.38%
Ridge :: 55.32%
Lasso :: 19.53%
ElasticNet :: 30.96%

Binary Classification

LogisticRegression : 84.09%

DecisionTreeClassifier : 86.36%
RandomForestClassifier : 75.00%
GradientBoostingClassifier: 86.36%

Because it handles spatial and temporal data, the CNN-GRU model often performs better in accuracy ratings. Proposed convolutional neural network (CNN) models can improve initial models with RMSE of 0.5 and R^2 of 0.7 to RMSE of 0.4 and R^2 of 0.75. GRU models can enhance performance with an RMSE of 0.35 and R^2 of 0.8. Proposed CNN-GRU model can achieve RMSE as low as 0.3 and R^2 as high as 0.85. These results show the model's feature extraction and sequence learning abilities.

Discussion

Zero-Shot Learning (ZSL) using Convolutional Neural Networks (CNNs) can greatly improve the accuracy of predicting dissolved oxygen (DO) and salinity in river water, particularly in situations where there is a lack of labelled data or when new environmental variables arise. The core concept underlying Zero-Shot Learning (ZSL) is to utilise semantic knowledge, such as environmental characteristics or interconnections among water quality factors, to provide forecasts regarding unobserved circumstances. This methodology enables models to extrapolate from familiar data (observed classes) to unfamiliar situations (unobserved classes) without explicit guidance. For example, a model that has been trained using data from certain river conditions can use shared features or domain knowledge contained in the semantic space to predict dissolved oxygen (DO) and salinity levels in new, previously unseen conditions.

Practically, the implementation of Zero-Shot Learning (ZSL) for forecasting Dissolved Oxygen (DO) and salinity requires the utilisation of a Convolutional Neural Network (CNN) as a tool to extract features from the input data. This input data might consist of different water quality measurements, meteorological data, and other pertinent variables. The CNN produces a latent representation of this data, encapsulating fundamental patterns and connections. Subsequently, these characteristics are assigned to a semantic domain where established properties or connections are delineated, such as the impact of temperature, pH, or flow rate on dissolved oxygen (DO) and salinity levels. This mapping allows the model to generate informed predictions about unfamiliar situations by comparing the expected characteristics with the established semantic links obtained from the training data.

ZSL excels in this setting because to its capacity to integrate and leverage domain expertise, rendering it highly flexible in response to evolving environmental circumstances. For instance, in situations of severe weather conditions or in areas where there is little monitoring data, conventional models may have difficulties because of the absence of sufficient training data that accurately represents the situation. Nevertheless, a zero-shot learning (ZSL) model can employ the semantic correlations it has acquired to deduce the probable effect on dissolved oxygen (DO) and salinity levels. This technique not only enhances the accuracy of predictions under unfamiliar settings but also improves the resilience and dependability of the model, which are essential for successful management and decision-making about water quality. Furthermore, the combination of zero-shot learning (ZSL) with convolutional neural networks (CNNs) enables the process of ongoing learning and adjustment. As additional data becomes accessible, it can be included into the semantic space without requiring considerable retraining, enabling the model to dynamically update its predictions. This attribute is especially advantageous for environmental monitoring systems, as they might experience quick changes in conditions, and accurate predictions are crucial. By integrating the robust feature extraction capabilities of Convolutional Neural Networks (CNNs) with the adaptable and flexible nature of Zero-Shot Learning (ZSL), we can create advanced models that offer precise and dependable forecasts of Dissolved Oxygen (DO) and salinity levels in river water.

Ultimately, this will enhance water resource management and environmental conservation efforts.

6. Conclusion and Future work

Zero-Shot Learning (ZSL) and Convolutional Neural Networks (CNNs) can predict river water dissolved oxygen and salinity. This strategy works well when specific conditions lack tagged data. A Convolutional Neural Network (CNN) may extract relevant characteristics from data using semantic qualities or domain-specific embeddings connected with environmental variables like water temperature, pH levels, and historical data trends. These traits are then connected with the semantic domain using a well-trained model. The model estimates dissolved oxygen and salinity in new, unexpected conditions using attribute associations during prediction. This strategy improves the model's ability to adapt to new scenarios and reduces the requirement for tagged data, making environmental monitoring systems more dependable and adaptive. Zero-Shot Learning using CNNs uses semantic characteristics to bridge the gap between minimal data and accurate predictions. This innovation allows advanced environmental research and resource management methods. Zero-Shot Learning (ZSL) utilising CNNs to predict river water DO and salinity. This is crucial as environmental monitoring increasingly relies on powerful machine learning. Predictive models for dissolved oxygen (DO) and salinity have traditionally used huge datasets with comprehensive labels and a wide range of conditions and locales. Due to the variety of environmental factors that affect water quality, obtaining such thorough data is often impossible. ZSL uses semantic features or contextual signals to forecast new and unexpected scenarios without considerable data collection. Zero-shot learning (ZSL) and convolutional neural networks (CNNs) for water quality prediction could revolutionise environmental monitoring. Through this integration, models may easily adapt to new locales and varied climates with minimal training data. By understanding environmental trends, a Zero Shot Learning (ZSL) model trained on a limited dataset from certain rivers may predict Dissolved Oxygen (DO) levels and salinity in unmonitored rivers. This technology could considerably improve real-time water quality monitoring and control. It can detect ecological changes and assure water safety early on.

As Zero-Shot Learning (ZSL) improves, predictive models could use satellite imagery and sensor networks. By combining geographical and temporal data with zero-shot learning (ZSL), models can improve prediction precision and resilience. ZSL's ability to predict water quality in unexpected situations makes it an essential tool for addressing climate change and human-induced impacts on aquatic ecosystems. Zero-Shot Learning (ZSL) and Convolutional Neural Networks (CNNs) to forecast Dissolved Oxygen (DO) and salinity are a promising way for environmental monitoring systems to improve scalability, adaptability, and efficiency.

Conflict of Interest: All the authors declared no conflict of interest related to this research.

Data Availability Statement: The dataset will be available with the corresponding author based on individual requests.

Acknowledgments: Not applicable.

Author's Contributions: NK: Design and methods, AR: Conclusion and review of the first draft, NN: Introduction and background, SD: Results and analysis, SD & NK: Discussion and review of the final draft, NN: Conceptualization and corresponding author.

Funding: There is no funding associated with this work

Competing Interests: The authors declare that they have no competing interests.

Reference:

[1]. Klemenjak, S., Waske, B., Valero, S., & Chanussot, J. (2012). Automatic detection of rivers in high-resolution SAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(5), 1364–1372. https://doi.org/10.1109/JSTARS.2012.2189099

[2]. Imen, S., Chang, N. Bin, Yang, Y. J., & Golchubian, A. (2018). Developing a Model-Based Drinking Water Decision Support System Featuring Remote Sensing and Fast Learning Techniques. IEEE Systems Journal, 12(2), 1358–1368. https://doi.org/10.1109/JSYST.2016.2538082

[3]. Zia, H., Harris, N. R., Merrett, G. V., & Rivers, M. (2019). A Low-Complexity Machine Learning Nitrate Loss Predictive Model-Towards Proactive Farm Management in a Networked Catchment. IEEE Access, 7, 26707–26720. https://doi.org/10.1109/ACCESS.2019.2901218

[4]. MacIel, G. M., Cabral, V. A., Marcato, A. L. M., Júnior, I. C. S., & Honório, L. D. M. (2020). Daily water flow forecasting via coupling between SMAP and deep learning. IEEE Access, 8, 204660–204675. https://doi.org/10.1109/ACCESS.2020.3036487

[5]. Dong, W., & Yang, Q. (2020). Data-Driven Solution for Optimal Pumping Units Scheduling of Smart Water Conservancy. IEEE Internet of Things Journal, 7(3), 1919–1926. https://doi.org/10.1109/JIOT.2019.2963250

[6]. Zhang, R., Deng, R., Liu, Y., Liang, Y., Xiong, L., Cao, B., & Zhang, W. (2020). Developing New Colored Dissolved Organic Matter Retrieval Algorithms Based on Sparse Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 3478–3492. https://doi.org/10.1109/JSTARS.2020.3003593

[7]. Pattanayak, A. S., Pattnaik, B. S., Udgata, S. K., & Panda, A. K. (2020). Development of Chemical Oxygen on Demand (COD) Soft Sensor Using Edge Intelligence. IEEE Sensors Journal, 20(24), 14892–14902. https://doi.org/10.1109/JSEN.2020.3010134

[8]. Abba, S. I., Linh, N. T. T., Abdullahi, J., Ali, S. I. A., Pham, Q. B., Abdulkadir, R. A., Costache, R., Nam, V. T., & Anh, D. T. (2020). Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration. IEEE Access, 8, 157218–157237. https://doi.org/10.1109/ACCESS.2020.3017743

[9]. Khan, A. A., Jamil, A., Hussain, D., Taj, M., Jabeen, G., & Malik, M. K. (2020). Machine-Learning Algorithms for Mapping Debris-Covered Glaciers: The Hunza Basin Case Study. IEEE Access, 8, 12725–12734. https://doi.org/10.1109/ACCESS.2020.2965768

[10]. Lopez, I. D., Figueroa, A., & Corrales, J. C. (2020). Multi-Dimensional Data Preparation: A Process to Support Vulnerability Analysis and Climate Change Adaptation. IEEE Access, 8, 87228–87242. https://doi.org/10.1109/ACCESS.2020.2992255

[11]. Gu, K., Zhang, Y., & Qiao, J. (2020). Random Forest Ensemble for River Turbidity Measurement from Space Remote Sensing Data. IEEE Transactions on Instrumentation and Measurement, 69(11), 9028–9036. https://doi.org/10.1109/TIM.2020.2998615

[12]. Chen, B., Mu, X., Chen, P., Wang, B., Choi, J., Park, H., Xu, S., Wu, Y., & Yang, H. (2021). Machine learning-based inversion of water quality parameters in typical reach of the urban river by UAV multispectral data. Ecological Indicators, 133. https://doi.org/10.1016/j.ecolind.2021.108434

[13]. Tiyasha, T., Bhagat, S. K., Fituma, F., Tung, T. M., Shahid, S., & Yaseen, Z. M. (2021). Dual Water Choices: The Assessment of the Influential Factors on Water Sources Choices Using Unsupervised Machine Learning Market Basket Analysis. IEEE Access, 9, 150532–150544. https://doi.org/10.1109/ACCESS.2021.3124817

[14]. Lopez, I. D., Figueroa, A., & Corrales, J. C. (2021). Multi-Label Data Fusion to Support Agricultural Vulnerability Assessments. IEEE Access, 9, 88313–88326. https://doi.org/10.1109/ACCESS.2021.3089665

[15]. Al-Sulttani, A. O., Al-Mukhtar, M., Roomi, A. B., Farooque, A. A., Khedher, K. M., & Yaseen, Z. M. (2021). Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction. IEEE Access, 9, 108527–108541. https://doi.org/10.1109/ACCESS.2021.3100490

[16]. Zhang, Z., Huang, J., Duan, S., Huang, Y., Cai, J., & Bian, J. (2022). Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed. Ecological Indicators, 140(May), 108977. https://doi.org/10.1016/j.ecolind.2022.108977

[17]. Zanoni, M. G., Majone, B., & Bellin, A. (2022). A catchment-scale model of river water quality by Machine Learning. Science of the Total Environment, 838(May), 156377. https://doi.org/10.1016/j.scitotenv.2022.156377

[18]. Lee, H. W., Kim, M., Son, H. W., Min, B., & Choi, J. H. (2022). Machine-learning-based water quality management of river with serial impoundments in the Republic of Korea. Journal of Hydrology: Regional Studies, 41(March), 101069. https://doi.org/10.1016/j.ejrh.2022.101069

[19]. Chopade, S., Gupta, H. P., Mishra, R., Oswal, A., Kumari, P., & Dutta, T. (2022). A Sensors-Based River Water Quality Assessment System Using Deep Neural Network. IEEE Internet of Things Journal, 9(16), 14375–14384. https://doi.org/10.1109/JIOT.2021.3078892

[20]. Cao, Z., Ma, R., Pahlevan, N., Liu, M., Melack, J. M., Duan, H., Xue, K., & Shen, M. (2022). Evaluating and Optimizing VIIRS Retrievals of Chlorophyll-a and Suspended Particulate Matter in Turbid Lakes Using a Machine Learning Approach. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–17. https://doi.org/10.1109/TGRS.2022.3220529

[21]. Gu, K., Liu, J., Shi, S., Xie, S., Shi, T., & Qiao, J. (2022). Self-Organizing Multichannel Deep Learning System for River Turbidity Monitoring. IEEE Transactions on Instrumentation and Measurement, 71. https://doi.org/10.1109/TIM.2022.3205915

[22]. Vidal Batista, L. (2022). Turbidity classification of the Paraopeba River using machine learning and Sentinel-2 images. IEEE Latin America Transactions, 20(5), 799–805. https://doi.org/10.1109/TLA.2022.9693564

[23]. Cai, J., Chen, J., Dou, X., & Xing, Q. (2022). Using Machine Learning Algorithms With In Situ Hyperspectral Reflectance Data to Assess Comprehensive Water Quality of Urban Rivers. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–13. https://doi.org/10.1109/TGRS.2022.3147695

[24]. Aslam, B., Maqsoom, A., Cheema, A. H., Ullah, F., Alharbi, A., & Imran, M. (2022). Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach. IEEE Access, 10(September), 119692–119705. https://doi.org/10.1109/ACCESS.2022.3221430

[25]. Chen, P., Wang, B., Wu, Y., Wang, Q., Huang, Z., & Wang, C. (2023). Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data. Ecological Indicators, 146(November 2022), 109750. https://doi.org/10.1016/j.ecolind.2022.109750

[26]. Li, X., Li, F., Min, X., Xie, Y., & Zhang, Y. (2023). Embracing eDNA and machine learning for taxonomy-free microorganisms biomonitoring to assess the river ecological status. Ecological Indicators, 155(July), 110948. https://doi.org/10.1016/j.ecolind.2023.110948

[27]. Wang, S., Chen, H., Su, W., Cui, S., Xu, Y., & Zhou, Z. (2023). Research on habitat quality assessment and decision-making based on Semi-supervised Ensemble Learning method—Daxia River Basin, China. Ecological Indicators, 156(August), 111153. https://doi.org/10.1016/j.ecolind.2023.111153

[28]. Aldrees, A., Javed, M. F., Bakheit Taha, A. T., Mustafa Mohamed, A., Jasiński, M., & Gono, M. (2023). Evolutionary and ensemble machine learning predictive models for evaluation of water quality. Journal of Hydrology: Regional Studies, 46(February). https://doi.org/10.1016/j.ejrh.2023.101331

[29]. Adli Zakaria, M. N., Ahmed, A. N., Abdul Malek, M., Birima, A. H., Hayet Khan, M. M., Sherif, M., & Elshafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. Heliyon, 9(7), e17689. https://doi.org/10.1016/j.heliyon.2023.e17689

[30]. J, V., K, K., P, G. M., C, G., Subramaniam, P. R., & Rangarajan, S. (2023). Strategies for classifying water quality in the Cauvery River using a federated learning technique. International Journal of Cognitive Computing in Engineering, 4(March), 187–193. https://doi.org/10.1016/j.ijcce.2023.04.004

[31]. Huang, J., Wang, D., Pan, S., Li, H., Gong, F., Hu, H., He, X., Bai, Y., & Zheng, Z. (2023). A New High-Resolution Remote Sensing Monitoring Method for Nutrients in Coastal Waters. IEEE Transactions on Geoscience and Remote Sensing, 61. https://doi.org/10.1109/TGRS.2023.3294436

[32]. Rathnayake, N., Rathnayake, U., Chathuranika, I., Dang, T. L., & Hoshino, Y. (2023). Projected Water Levels and Identified Future Floods: A Comparative Analysis for Mahaweli River, Sri Lanka. IEEE Access, 11(January), 8920–8937. https://doi.org/10.1109/ACCESS.2023.3238717

[33]. Jia, L., Yen, N., & Pei, Y. (2023). Spatial and Temporal Water Quality Data Prediction of Transboundary Watershed Using Multiview Neural Network Coupling. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–16. https://doi.org/10.1109/TGRS.2023.3334291

[34]. Xue, J., Yuan, C., Ji, X., & Zhang, M. (2024). Predictive modeling of nitrogen and phosphorus concentrations in rivers using a machine learning framework: A case study in an urban-rural transitional area in Wenzhou China. Science of the Total Environment, 910(August 2023), 168521. https://doi.org/10.1016/j.scitotenv.2023.168521

[35]. Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Advances in machine learning and IoT for water quality monitoring: A comprehensive review. Heliyon, 10(6), e27920. https://doi.org/10.1016/j.heliyon.2024.e27920

[36]. Kushwaha, N. L., Kudnar, N. S., Vishwakarma, D. K., Subeesh, A., Jatav, M. S., Gaddikeri, V., Ahmed, A. A., & Abdelaty, I. (2024). Stacked hybridization to enhance the performance of artificial neural networks (ANN) for prediction of water quality index in the Bagh river basin, India. Heliyon, 10(10), e31085. https://doi.org/10.1016/j.heliyon.2024.e31085

[37]. Sarafaraz, J., Ahmadzadeh Kaleybar, F., Mahmoudi Karamjavan, J., & Habibzadeh, N. (2024). Predicting river water quality: An imposing engagement between machine learning and the QUAL2Kw models (case study: Aji-Chai, river, Iran). Results in Engineering, 21(February), 101921. https://doi.org/10.1016/j.rineng.2024.101921

[38]. Li, Y., Ma, L., Huang, J., Disse, M., Zhan, W., Li, L., Zhang, T., Sun, H., & Tian, Y. (2024). Machine learning parallel system for integrated process-model calibration and accuracy enhancement in sewer-river system. Environmental Science and Ecotechnology, 18, 100320. https://doi.org/10.1016/j.ese.2023.100320

[39]. Dalal, S., Onyema, E. M., Romero, C. A. T., Ndufeiya-Kumasi, L. C., Maryann, D. C., Nnedimkpa, A. J., & Bhatia, T. K. (2022). Machine learning-based forecasting of potability of drinking water through adaptive boosting model. Open Chemistry, 20(1), 816-828.

[40]. Dalal, S. (2023, August). Machine learning model for Water Quality evaluation: Systematic Review. In 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 1013-1018). IEEE.

[41]. Dalal, S. (2024). Optimizing Linear Regression model in Water Hardness Prediction for Industry 4.0. Trends in Mechatronics Systems, 73-91.

[42]. Edeh, M. O., Dalal, S., Alhussein, M., Aurangzeb, K., Seth, B., & Kumar, K. (2024). A novel deep learning model for predicting marine pollution for sustainable ocean management. PeerJ Computer Science, 10, e2482.