International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 03, MAR, 2025

Comparative Analysis of Deep Learning Models for Enhanced Water Quality Prediction

Mr. Nanduri Ashok Kumar[1], Assistant Professor, Dept. Of IT

Vasireddy Venkatadri Institute of Technology, Namburu, Guntur, Andhra Pradesh -522508.

Gonuguntla Sai Gopi[2], Prabhu Das Gera[3], Vinay Gottam[4], Rajesh Devarakonda[5]

UG Students, Dept. Of IT, Vasireddy Venkatadri Institute of Technology, Namburu,
Guntur, Andhra Pradesh -522508.

ABSTRACT

Environmental safety and public health depend on the evaluation of water quality. Utilizing the Indian Aqua Attributes Dataset, which include variables such as temperature, dissolved oxygen, pH, conductivity, B.O.D., nitrate levels, and coliform counts, this work applies deep learning techniques to forecast water quality. Three models—CNN, LSTM, and CNN-LSTM—were trained and assessed following preprocessing. With an AUC-ROC score of 0.95, the CNN-LSTM model outperformed the others, indicating its dependability in classification.The trained model was saved in.h5 format and deployed using Streamlit to provide real-time predictions. This allows users to enter water parameters and receive immediate responses. The results demonstrate deep learning's efficiency in monitoring water quality. Future research could improve accuracy by using bigger datasets and more features.

KEYWORDS

Water Quality Prediction, Deep Learning, CNN-LSTM, Indian Aqua Attributes Dataset, Streamlit, ROC Curve, Confusion Matrix, Environmental Monitoring.

I. INTRODUCTION

Water quality is essential for industrial processes, environmental preservation, and drinking water safety. Contaminants in water sources can lead to severe health hazards, making accurate and timely water quality prediction essential. Traditional monitoring approaches depend on personal sampling and testing in laboratories, which can be laborious, expensive, inconsistent due to environmental conditions. Automatic prediction models have become effective substitutes for evaluating water quality due to developments in deep learning and AI. Using the Indian Aqua Attributes Dataset, this work investigates the use of CNN, LSTM, and CNN-LSTM models to categorize water quality according to important factors like pH, temperature, dissolved oxygen, conductivity, B.O.D., nitrate levels, and coliform counts. The CNN-LSTM model outperformed the others, showcasing deep learning's promise for real-time water quality monitoring. The proposed system is integrated into a Streamlit-based web application, allowing users to input water parameters and obtain instant predictions. By leveraging deep learning-based classification, this approach aims to enhance decision-making in water management while reducing dependence on manual testing.

a. Background on Water Quality Prediction

Predicting the quality of water entails evaluating its physical, chemical, and biological properties to see if it is appropriate for human consumption and other uses. Various parameters, including pH levels, temperature, dissolved oxygen, and bacterial contamination, serve as key indicators of water quality. Traditionally, water quality assessments require extensive sample collection, laboratory analysis, and expert interpretation, which can be inefficient and costly. As deep learning and machine learning evolved, data-driven models can now predict water quality with high accuracy. These models analyze historical and real-time data to identify patterns and classify water samples effectively. In this study, deep learning architectures like CNN, LSTM, and CNN-LSTM were trained to classify water quality based on the Indian Aqua Attributes Dataset. The CNN-LSTM model demonstrated the highest classification accuracy, making it a reliable tool for automated water quality prediction.

b. Importance of Tracking of Water Quality in India:

India faces significant challenges in maintaining clean and safe water sources due to rapid industrialization, urbanization, and population growth. Contaminants from industrial discharge, agricultural runoff, and sewage often degrade water quality, leading to health risks such as waterborne diseases, gastrointestinal infections, and long-term toxic effects. Effective water quality monitoring is essential to prevent such risks and ensure compliance with environmental and public health regulations. The Indian government has implemented initiatives like the National Water Quality Monitoring Programme (NWQMP) to track water quality; however, traditional methods rely on manual sampling, which is inefficient for large-scale monitoring. By utilizing deep learning-based prediction models, water quality assessment can be automated, improving accuracy and efficiency. The proposed CNN-LSTM-based system offers real-time analysis through a web-based application, providing a scalable and accessible solution to check India's water quaity.

c.Challenges in Traditional Analysis of Water Quality

Traditional evaluation of water quality methods involves manual sample collection, laboratory testing, and expert analysis, posing several challenges:

1) Time-Consuming Process – Laboratory analysis requires significant time for chemical and biological tests, delaying critical decision-making.
2) High Costs – Regular sample collection and testing involve substantial financial investment, making it impractical for continuous monitoring.
3) Geographical Limitations – Remote areas lack access to well-equipped laboratories, leading to delays in water quality assessment.
4) Human Errors and Inconsistencies – Manual sampling can introduce errors due to variations in collection techniques and environmental conditions.
5) Scalability Issues – Traditional methods struggle to monitor water quality across vast regions and multiple sources simultaneously.

To overcome these limitations, deep learning models provide an efficient alternative by analyzing water quality data in real time. The CNN-LSTM model, deployed via a Streamlit web application, enhances accuracy and accessibility while reducing dependence on manual testing.

d.Scope of the Proposed Work

This research focuses on developing an automated water quality prediction system using deep learning techniques. The Indian Aqua Attributes Dataset was preprocessed by removing irrelevant features and normalizing key parameters like pH, temperature, dissolved oxygen, conductivity, and nitrate levels. After training and evaluating The CNN-LSTM model outperformed the CNN, LSTM, and CNN models in terms of categorization with an AUC-ROC score of 0.95. The trained CNN-LSTM model was deployed using Streamlit, creating an interactive web-based platform where users can input water parameters and receive instant predictions. This system aims to:

1) Improve Accuracy – Achieve high precision in water quality classification using deep learning.
2) Enable Real-Time Monitoring – Provide instant predictions through a user-friendly web interface.
3) Enhance Scalability – Allow deployment across multiple regions with minimal infrastructure requirements.
4) Reduce Manual Effort – Minimize dependence on traditional sample collection and testing.

Future enhancements may include expanding the dataset, incorporating additional water quality parameters, and integrating IoT sensors for real-time data collection, further improving prediction capabilities.

e.Deep Learning-Based Approaches in Water Quality Analysis

Deep learning has revolutionized water quality prediction by providing automated, scalable, and accurate models. Studies have explored different DL architectures to classify and predict water contamination levels. CNN models are widely used for feature extraction, while LSTM models handle time-series dependencies. Hybrid models, such as CNN-LSTM, have demonstrated superior performance in various environmental monitoring applications. The proposed work utilizes a CNN-LSTM model trained on the Indian Aqua Attributes Dataset, achieving high accuracy in water quality classification. The model's integration into a Streamlit-based web application enables real-time water quality assessment, reducing reliance on traditional methods.

II. RELATED WORK

a.Overview of Existing Methods for Water Quality Prediction

Forecasting water quality has proven an area of extensive research due to its importance in public health, environmental sustainability, and industrial applications. Conventional techniques for evaluating the quality of water depend on laboratory-based chemical and biological analysis, requiring manual sampling and testingAlthough these techniques yield precise findings, they are time-intensive and laborious and do not allow for continuous surveillance. Automating the evaluation of water quality has become more popular in recent years thanks to machine learning (ML) and deep learning (DL) models. ML models like random forests (RF), support vector machines (SVM), and decision trees have demonstrated improved efficiency in classifying water quality based on historical data patterns. Yet, such models frequently need substantial feature engineering and are limited in their ability to manage complicated time-series data. Methods for deep learning, especially CNN, LSTM, and hybrid CNN-LSTM models or Convolutional Neural Networks, have shown superior performance in water quality prediction . These algorithms improve forecast accuracy and eliminate the need for manual preprocessing by automatically extracting geographical and time-based data from water quality datasets.

To categorize water quality, **Umair Ahmed et al. (2019)** suggested a supervised machine learning method based on various chemical and physical parameters. Decision trees, support vector machines, and random forests were among the machine learning methods used in the study demonstrating high classification accuracy. However, the model required extensive feature engineering and manual parameter tuning. Compared to this study, our proposed CNN-LSTM model removes the necessity of choosing attributes by hand by leveraging deep learning's automated feature extraction capabilities. Additionally, the Streamlit-based web interface enhances real-time usability.

Heelak Choi et al. (2021) investigated the efficacy of computational algorithms like CNN and LSTM in short-term water quality forecasting. The research demonstrated how LSTM can model time-series dependencies, making it superior to conventional ML methods. In our research, we extend this approach by integrating CNN and LSTM into a hybrid CNN-LSTM model, achieving better performance for multiclass water quality classification. The experimental results demonstrate that CNN-LSTM outperforms individual deep learning algorithms for forecasting trends in the health of water.

Choi et al. (2021) the effectiveness of deep learning techniques for predicting the state of surface water in the short run,comparing models like RNN, GRU, LSTM, and hybrid CNN-GRU with the traditional ARIMA model. The study found that deep learning models outperformed ARIMA in univariate data sets, achieving lower mean absolute percentage error (MAPE) for parameters like T-P (total phosphorus) and BOD (biochemical oxygen demand). However, the inclusion of additional water quality variables in multivariate data sets did not consistently improve prediction accuracy, with performance varying across monitoring stations. Factors such as the number of input parameters, sliding window size, and inclusion of statistics on weather and outflow significantly influenced model performance. The study highlights the need for iterative optimization of input parameters and model configurations to improve the forecasting accuracy of water quality.

A Multi-Task Deep Learning (MTL-CNN-LSTM) model was presented by **Wu et al. (2022)** for predicting the water quality of the Yellow River in China. Their model leverages the correlation between multiple river sections to improve prediction accuracy by sharing and learning from water quality data across sections simultaneously. Comparing the combined CNN-LSTM architecture to single-task models, the mean absolute error (MAE) and root mean square error (RMSE) are reduced by 13.2% and 15.5%, respectively, demonstrating the successful capture of both temporal and spatial data. This method shows how multi-task learning can improve the stability and generalization of water quality prediction models.

III. DATASET AND PREPROCESSING

a. Dataset Description

The data used for the research was collected via HydroShare, an application for collaboration for sharing hydrological datasets. The dataset, referred to as the Indian Aqua Attributes Dataset, contains detailed water quality measurements collected from various monitoring stations across India. These records provide insights into the physicochemical properties of water, which are essential for assessing its appropriateness for use in industry, agriculture, & human consumption. The dataset consists of multiple attributes, including station codes, location details, temperatures, values of pH, electrical conductivity (EC), total dissolved solids (TDS), dissolved oxygen (DO), and biochemical oxygen demand (BOD). These characteristics are important markers of water quality that aid in dividing water into safe, moderate, and hazardous categories. Given the importance of these attributes, the dataset is ideal for developing a deep learning-based water quality prediction system using CNN-LSTM models.

The dataset used in this project has been sourced from **HydroShare.hydroshare.org/resource**, a widely recognized repository for hydrological and environmental data. HydroShare provides open access to water quality datasets, enabling researchers to develop models that enhance water resource management and pollution control strategies. This dataset has been curated from **multiple water monitoring stations across India**, covering diverse **geographical regions and seasonal variations**. The inclusion of multiple data points from different regions allows for a more generalized and robust prediction model, making it applicable to a variety of water sources across the country.

b.Explanation of Features

The dataset comprises several key features that are crucial for **predicting water quality**. Below is a description of the most relevant attributes:

- **Station Code:** A unique identifier assigned to each water quality monitoring station.
- **Location:** The geographical coordinates or name of the water source from which samples were collected.
- **Temperature (°C):** The water temperature, which influences chemical reactions and biological processes.
- **pH:** On a scale of 0 to 14, it indicates how acidic or alkaline water is. Acidity is indicated by values less than 7, and alkalinity is indicated by values more than 7.
- **Dissolved Oxygen (DO) (mg/L):** Reflects the quantity of oxygen that aquatic life may access in the water. Pollution is indicated by low DO levels.
- **Biochemical Oxygen Demand (BOD) (mg/L):** Shows how much oxygen is needed by microbes to break down organic materials. Elevated BOD levels indicate pollution.
- **Electrical Conductivity (EC) (µS/cm):** Evaluates the water's electrical conductivity, which is affected by the minerals and salts that are dissolved in it.
- **Total Dissolved Solids (TDS) (mg/L):** Indicates the overall amount of dissolved material in water. Poor water quality may be indicated by excessive TDS rates.

The CNN-LSTM model is trained using these features, enabling it to categorize water quality according to patterns seen in the data.

c.Data Cleaning and Preprocessing

To ensure the dataset is well-structured and suitable for deep learning, several preprocessing steps were performed. Initially, unnecessary columns that did not contribute to water quality prediction, such as administrative remarks and sampling time, were removed to reduce data complexity. Handling missing values was a crucial step, as incomplete data could lead to inaccurate predictions. Rows with more than 40% missing values were discarded, while those with minimal missing values were calculated based on the kind of attribute using statistical techniques such as mean, median, or mode. For categorical variables like station locations, missing values were replaced using the most frequent value to maintain consistency. Additionally, since the dataset contains attributes with varying numerical scales, feature scaling was applied to standardize the values. Min-Max scaling was used to normalize numerical features, bringing all values within a range of 0 to 1, ensuring balanced learning and keeping the training of the model from being dominated by any one characteristic. These preprocessing steps significantly improved the dataset's quality, enhancing the CNN-LSTM model's stability and performance in predicting water quality.

IV. METHODOLOGY

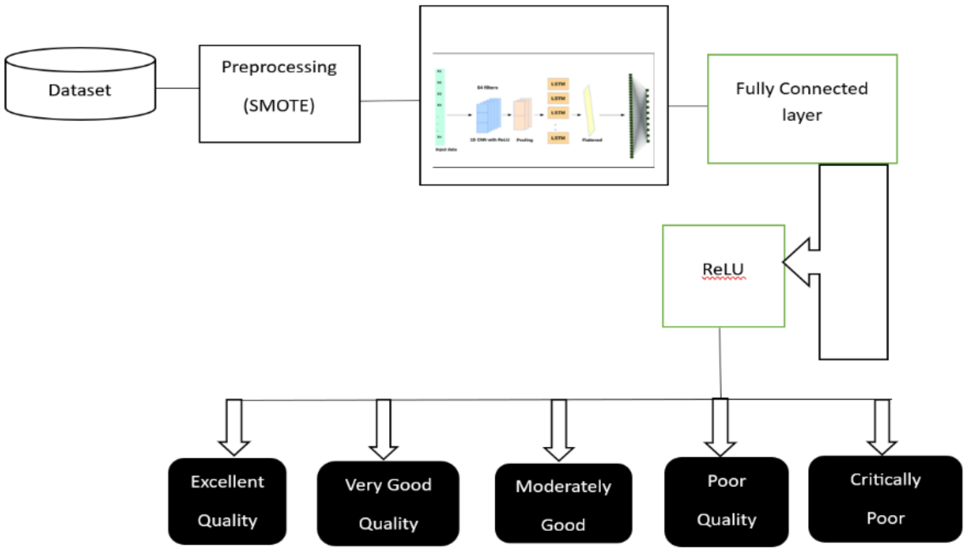

Fig 1: *System Architecture*

a. Convolutional Neural Network (CNN)

The water quality dataset's spatial features were determined through Convolutional Neural Network. Although CNNs are primarily used for image processing, their ability to identify patterns and correlations in structured data makes them useful for water quality prediction. The dataset, consisting of numerical values for parameters like temperature, pH, and conductivity, was reshaped into a format suitable for CNN input. Convolutional layers extracted relevant feature representations, followed by activation functions and pooling layers that reduced dimensionality while retaining crucial information. The collected characteristics were further processed for classification by the fully linked layers.

b.Long Short-Term Memory (LSTM)

An LSTM model was used to depict the fluctuations in changes in water quality. One kind of recurrent neural network (RNN) made to effectively handle sequential information is called an LSTM network. Water quality data fluctuates over time due to environmental and human factors, making LSTM suitable for learning these dependencies. The dataset was structured as time-series input, where each sample represented water quality parameters at a specific time step. The memory cells in LSTM retained important past information while filtering out irrelevant data, allowing the model to make predictions based on historical trends.

c.Hybrid CNN-LSTM Model

A CNN-LSTM hybrid model was created to take advantage of both CNN's and LSTM's advantages. CNN component extracted spatial relationships between water quality attributes, while the LSTM component captured temporal dependencies. The model architecture began with convolutional layers to learn feature representations, LSTM layers are then used to process the dataset's sequential nature. This combination allowed the model to make more accurate predictions by utilizing both spatial and temporal dependencies. The final output layer classified water quality into predefined categories.

d.Training Strategy and Hyperparameter Tuning

An 80-20 train-test split was used to train the models, guaranteeing that the characteristics of water quality were distributed evenly. Adam served as the optimizer for the CNN, LSTM, and CNN-LSTM models, and the loss function was categorical cross-entropy. Several hyperparameters were fine-tuned, including number of LSTM units, learning rate, and batch size. Batch normalization and dropout were incorporated to prevent overfitting, ensuring better generalization on unseen data.

e.Model Selection Based on Performance

Based on the evaluation metrics, Compared to individual CNN and LSTM models, the CNN-LSTM model fared better. While CNN efficiently extracted spatial features and LSTM captured temporal dependencies, their hybrid combination offered a more thorough comprehension of patterns in water quality. Reliable predictions were ensured by the final model's high AUC score, accuracy, and low loss.

f.Building the Streamlit Web Application

Frontend Development with Streamlit A user-friendly interface was developed using Streamlit, enabling easy interaction with the model. The web application featured an input panel where users could enter water quality parameters such as temperature, pH, and conductivity. The interface was designed to be intuitive, allowing non-technical users to access predictions effortlessly. Backend Integration with the Trained Model The trained CNN-LSTM model was loaded in the backend to process user inputs. When users entered water quality attributes, the backend preprocessed the data, applied the model for prediction, and displayed the output. The .h5 model was utilized for inference, ensuring that predictions were generated in real-time with minimal latency. User Input and Prediction Output Users provided key water quality parameters, which were fed into the model for analysis. The system then classified the water quality based on learned patterns, displaying the predicted category along with confidence scores. The prediction results helped assess water quality conditions, aiding decision-making for environmental monitoring and public health safety.

This methodology, which combined deep learning with an interactive web-based solution, guaranteed an effective, precise, and user-friendly approach to water quality prediction.

V. RESULTS AND DISCUSSION

To find the best method for predicting water quality, the CNN, LSTM, and hybrid CNN-LSTM models' performances were assessed. CNN is renowned for its capacity to identify spatial characteristics, performed well in recognizing patterns among water quality attributes. However, it struggled with temporal dependencies, leading to limitations in long-term trend predictions. The LSTM model, designed for sequential data processing, effectively recorded changes in water quality measurements over time, but lacked CNN's capacity to extract spatial features. The hybrid CNN-LSTM model, which combined both approaches, outperformed the individual models by leveraging spatial and temporal dependencies. This model achieved higher accuracy and lower loss, demonstrating its ability to generalize well across different water quality scenarios. To assess model performance, various metrics were used. Accuracy measured overall classification performance, while loss monitored how the actual values differ from the projected values. The trade-off between true positives and false positives was assessed using the ROC curve, whose area under the curve (AUC) signifies the efficacy of the model. A thorough analysis of true positive, true negative, false positive, and false negative values was given by the confusion matrix, helping analyze misclassification rates. The best outcomes were obtained using the CNN-LSTM model, with a high accuracy and AUC score, demonstrating its superior predictive ability. The models were evaluated using standard performance metrics, including accuracy, loss, the confusion matrix and the ROC curve. The percentage of accurate predictions was called accuracy, and the difference between expected and actual values was called loss. The CNN-LSTM model outperformed CNN and LSTM in terms of accuracy. The reduction in loss for the CNN-LSTM model indicated effective learning and minimal overfitting. These results confirmed that integrating convolutional layers with LSTM significantly improved the predictive capability of the model.

Each model's classification efficacy was evaluated using the ROC curve, which plots the true positive rate against the false positive rate. Better class discrimination is shown by a larger area under the curve (AUC). The model with the highest AUC value was CNN-LSTM, confirming its superior predictive power. The ROC curve showed that the hybrid model maintained a strong balance between sensitivity and specificity, making it the most reliable choice for water quality classification.

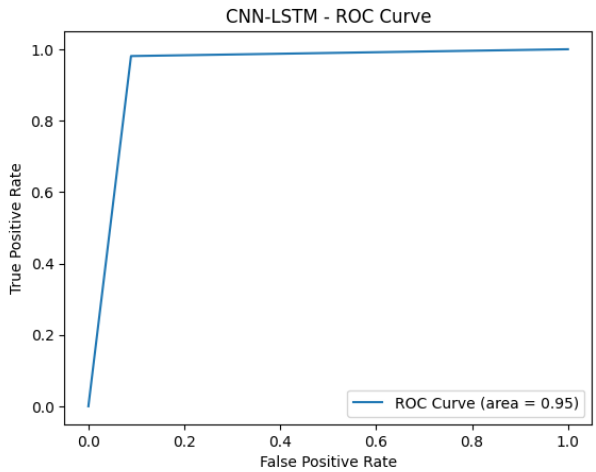

Fig 2: *CNN-LSTM ROC Curve*

The confusion matrix described true positives, false positives, true negatives, and false negatives, offering information about the model's classification performance. The CNN-LSTM model produced more accurate forecasts for water quality because it had the lowest rates of false negatives and false positives. The standalone CNN and LSTM models, while performing well, had slightly higher misclassification rates, reinforcing the effectiveness of the hybrid model. The matrix visualization illustrated the improved decision boundaries achieved by combining CNN's spatial learning with LSTM's sequential learning capabilities.

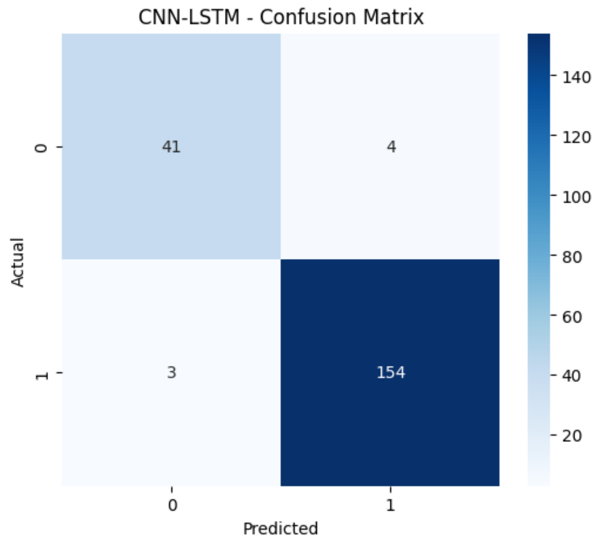

Fig 3: *CNN-LSTM Confusion Matrix*

The training and validation performance across several epochs was shown by the accuracy and loss graphs. The CNN-LSTM model showed a steady increase in accuracy with minimal fluctuations, indicating stable learning. Conversely, CNN and LSTM models had slower convergence rates, with slight overfitting observed in later epochs.

The loss graph confirmed that the hybrid model minimized error effectively, achieving a smoother decline compared to the standalone models. These results confirmed that a more reliable and broadly applicable model for predicting water quality was produced by merging CNN and LSTM. These findings demonstrate the benefits of combining CNN with LSTM, offering a trustworthy and precise method for evaluating the quality of water. With 99.92% accuracy, the CNN-LSTM model's exceptional performance makes it a good option for real-time water quality monitoring applications.

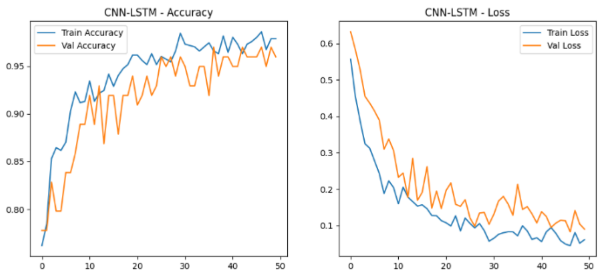

Fig 4: *CNN-LSTM Accuracy Loss Graph*

VI. CONCULSION AND FUTURE WORK

This work used the Indian Aqua Attributes dataset to provide a deep learning-based method for predicting water quality. We successfully examined temporal and geographical dependencies in water quality measures using CNN, LSTM, and a CNN-LSTM hybrid model. The CNN model demonstrated strong feature extraction capabilities, while LSTM efficiently captured sequential dependencies. The CNN-LSTM hybrid model outperformed the standalone models by integrating both spatial and temporal learning, achieving higher accuracy and lower loss. The deployment of the trained model using Streamlit provided a user-friendly interface for real-time water quality prediction, ensuring accessibility for stakeholders. This study advances the field by showcasing the benefits of hybrid modeling approaches and proving the efficacy of deep learning in water quality evaluation.

Future work can focus on addressing the limitations by incorporating a more diverse and extensive dataset covering multiple regions to enhance generalizability. Implementing real-time data collection and updating the model dynamically could improve accuracy and adaptability to environmental changes. To lower computing costs and facilitate deployment on edge devices, additional optimization strategies like model reduction and quantization can be investigated. Additionally, integrating external environmental parameters like rainfall, industrial activities, and climate conditions could provide a more holistic approach to water quality assessment.

VII. REFERENCES

[1]. Shams MY, Tarek Z, Elshewey AM et al (2023) A Machine Learning-Based Model for Predicting Temperature Under the Effects of Climate Change. In: Hassanien AE, Darwish A (eds) The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations. Springer Nature Switzerland, Cham, pp 61–81.

[2]. Elshewey AM, Shams MY, Elhady AM et al (2023) A Novel WD-SARIMAX Model for Temperature Forecasting Using Daily Delhi Climate Dataset. Sustainability 15:757. https://doi.org/10.3390/su15010757

[3]. Kumar, N.A., Vellaichamy, J. (2025). A deep learning approach for water quality assessment: Leveraging Gated Linear Networks for contamination classification. Ingénierie des Systèmes d'Information, Vol. 30, No. 2, pp. 349-357. https://doi.org/10.18280/isi.300206.

[4]. Tarek Z, Shams MY, Elshewey AM et al (2023) Wind Power Prediction Based on Machine Learning and Deep Learning Models. Comput Mater Contin 74:715–732. https://doi.org/10.32604/cmc.2023.032533

[5]. Elshewey AM, Shams MY, Tarek Z et al (2023) Weight Prediction Using the Hybrid Stacked-LSTM Food Selection Model. Comput Syst Sci Eng 46:765–781. https://doi.org/10.32604/csse.2023.034324

[6]. Waqas M, Tu S, Halim Z et al (2022) The Role of Artificial Intelligence and Machine Learning in Wireless Networks Security: Principle, Practice, and Challenges. Artificial Intelligence Review 55:5215–5261. https://doi.org/10.1007/s10462-022-10143-2

[7]. Wu J, Wang Z (2022) A Hybrid Model for Water Quality Prediction Based on an Artificial Neural Network, Wavelet Transform, and Long Short-Term Memory. Water 14:610.

[8]. Nanduri, A.K., Sravanthi, G.L., Pavan Kumar, K.V.K.V.L., Babu, S.R., Rama Krishna, K.V.S.S. (2021). Modified fuzzy approach to automatic classification of cyber hate speech from the online social networks (OSN's). Revue d'Intelligence Artificielle, Vol. 35, No. 2, pp. 139-144. https://doi.org/10.18280/ria.350205.

[9]. Malek NHA, Wan Yaacob WF, Md Nasir SA, Shaadan N (2022) Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. Water 14:1067.

[10]. Slatnia A, Ladjal M, Ouali MA, Imed M (2022) Improving Prediction and Classification of Water Quality Indices Using Hybrid Machine Learning Algorithms with Features Selection Analysis. Online International Symposium on Applied Mathematics and Engineering (ISAME22), vol 1, Istanbul-Turkey, pp 16–17.

[11]. Khullar S, Singh N (2022) Water Quality Assessment of a River Using Deep Learning Bi-LSTM Methodology: Forecasting and Validation. Environmental Science and Pollution Research 29:12875–12889.

[12]. Arepalli P.G., Narayana V.L., Venkatesh R., Kumar N.A. (2019). Certified node frequency in social network using parallel diffusion methods, Ingenierie des Systemesd'Information, Vol. 24, No. 1, pp. 113-117. https://doi.org/10.18280/isi.240117.

[13]. Elbeltagi A, Pande CB, Kouadri S, Islam ARM (2022) Applications of Various Data-Driven Models for the Prediction of Groundwater Quality Index in the Akot Basin, Maharashtra, India. Environmental Science and Pollution Research 29:17591–17605.

[14]. Nosair AM, Shams MY, AbouElmagd LM et al (2022) Predictive Model for Progressive Salinization in a Coastal Aquifer Using Artificial Intelligence and Hydrogeochemical Techniques: A Case Study of the Nile Delta Aquifer, Egypt. Environmental Science and Pollution Research 29:9318–9340.

[15]. Mounika, N. M., LSahithi, N., Lakshmi, N. K., KPraveenya, N., & Kumar, N. N. A. (2023). A CNN CLASSİFİCATİON APPROACH FOR POTATO PLANT LEAF DİSEASE DETECTİON. EPRA International Journal of Research & Development (IJRD), 277–281. https://doi.org/10.36713/epra14773.

[16]. Khoi DN, Quan NT, Linh DQ et al (2022) Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. Water 14:1552.

[17]. Wang S, Peng H, Liang S (2022) Prediction of Estuarine Water Quality Using Interpretable Machine Learning Approach. Journal of Hydrology 605:127320.

[18]. Deng T, Chau K-W, Duan H-F (2021) Machine Learning-Based Marine Water Quality Prediction for Coastal Hydro-Environment Management. Journal of Environmental Management 284:112051.

[19]. Garabaghi FH, Benzer S, Benzer R (2021) Performance Evaluation of Machine Learning Models with Ensemble Learning Approach in Classification of Water Quality Indices. Research Square 1:1–35. https://doi.org/10.21203/rs.3.rs-876980/v2

[20]. Nanduri, A. K., Kumar, S., L, G., Ratna, S., & V, K. (2020). IoT based Automatic Damaged Street Light Fault Detection Management System. International Journal of Advanced Computer Science and Applications, 11(8). https://doi.org/10.14569/ijacsa.2020.0110853.

[21]. Hassan MM, Hassan MM, Akter L et al (2021) Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. Human Centric Intelligent Systems 1:86–97.

[22]. Khan MSI, Islam N, Uddin J et al (2021) Water Quality Prediction and Classification Based on Principal Component Regression and Gradient Boosting Classifier Approach. Journal of King Saud University – Computer and Information Sciences 34(8):4773–4781. https://doi.org/10.1016/j.jksuci.2021.06.003

[23]. Jain D, Shah S, Mehta H et al (2021) A Machine Learning Approach to Analyze Marine Life Sustainability. In: Proceedings of International Conference on Intelligent Computing, Information and Control Systems. Springer, pp 619–632.

[24]. Hmoud Al-Adhaileh M, Waselallah Alsaade F (2021) Modelling and Prediction of Water Quality by Using Artificial Intelligence. Sustainability 13:4259.

[25]. Nanduri, A. K., Kondru, J. R., Rambhupal, M., & Ashok, N. (2023). Lung cancer detection using Computer-Aided Diagnosis (CAD). In Lecture notes in electrical engineering (pp. 501–510). https://doi.org/10.1007/978-981-99-0412-9_45.

[26]. Zhou J, Wang Y, Xiao F et al (2018) Water Quality Prediction Method Based on IGRA and LSTM. Water 10:1148.

[27]. Liu P, Wang J, Sangaiah AK et al (2019) Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. Sustainability 11:2058.

[28]. Hu Z, Zhang Y, Zhao Y et al (2019) A Water Quality Prediction Method Based on the Deep LSTM Network Considering Correlation in Smart Mariculture. Sensors 19:1420.

[29]. Lee S, Lee D (2018) Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. International Journal of Environmental Research and Public Health 15:1322.

[30]. Bhardwaj D, Verma N (2017) Research Paper on Analysing Impact of Various Parameters on Water Quality Index. International Journal of Advanced Research in Computer Science 8(5):2496–2498.

[31]. Prakash R, Tharun VP, Devi SR (2018) A Comparative Study of Various Classification Techniques to Determine Water Quality. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE, pp 1501–1506.

[32]. Radhakrishnan N, Pillai AS (2020) Comparison of Water Quality Classification Models Using Machine Learning. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, pp 1183–1188.

[33]. Halim Z, Waqar M, Tahir M (2020) A Machine Learning-Based Investigation Utilizing the In-Text Features for the Identification of Dominant Emotion in an Email. Knowledge-Based Systems 208:106443. https://doi.org/10.1016/j.knosys.2020.106443

[34]. Clark RM, Hakim S, Ostfeld A (2011) Handbook of Water and Wastewater Systems Protection. Springer, pp 1–29. https://doi.org/10.1007/978-1-4614-0189-6

[35]. Friedman JH (2002) Stochastic Gradient Boosting. Computational Statistics & Data Analysis 38:367–378.

[36]. Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water Quality Assessment in Terms of Water Quality Index. American Journal of Water Resources 1:34–38.