ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

Automated Water Quality and Level Monitoring with Hybrid Deep Learning CART Model

Duwarapu Sindhuja [1], Ankalam Sai Pranaya [2], Movva Pranay Gopi [3], M. Nandini Priyanka [4]

[1,2,3] UG Scholar, Department of Electronics and Communication Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Department of Electronics and Communication Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
dsindhuja28@gmail.com

Abstract:

Water level monitoring is critical for effective water resource management, flood prevention, and infrastructure safety. Traditional water level monitoring systems primarily rely on static threshold- based mechanisms and simple regression models, which often lack the adaptability and precision required to handle dynamic environmental changes and extreme weather events. The limitations pose challenges such as delayed warnings, inaccurate predictions, and inefficient decision-making in critical situations. The paper presents an innovative automated water level monitoring system that integrates hybrid deep learning classifiers and regression models to process real-time sensor inputs. By combining classification techniques to categorize water levels with regression models for precise numerical predictions, the system achieves a higher degree of accuracy and responsiveness compared to traditional methods. The proposed approach addresses several limitations of conventional systems. First, traditional monitoring relies heavily on fixed thresholds, which may not account for complex variables such as seasonal variations, rainfall intensity, and upstream water flow. Second, these systems often suffer from data sparsity and inaccuracies due to sensor noise, which impacts prediction reliability. By leveraging advanced neural network architectures such as Long Short-Term Memory (LSTM) networks and deep neural networks (DNNs), the hybrid model overcomes these challenges by learning intricate temporal patterns and relationships in the data. The system also integrates IoT-based sensors to collect real-time data such as water height, temperature, rainfall, and flow velocity. This data is pre-processed using noise filtering and outlier detection techniques, ensuring high-quality inputs for the models. The classification module provides an initial categorization of water levels, while the regression module refines these predictions for actionable insights. A fusion mechanism ensures seamless integration between classification and regression outputs, enabling robust decision- making and timely alerts. The significance of the hybrid approach lies in its ability to enhance prediction accuracy, reduce latency, and adapt to diverse environmental conditions. By addressing the limitations of traditional systems, it supports proactive water resource management and mitigates risks associated with floods and droughts. The work demonstrates the potential of combining deep learning techniques with IoT-enabled infrastructure to revolutionize water monitoring systems and create more resilient communities.

1. INTRODUCTION

Water is a fundamental resource for life, agriculture, industry, and ecosystems, making its effective management a global priority. Automated water level monitoring has become crucial for applications such as flood prevention, reservoir management, and agricultural irrigation. Traditional water monitoring techniques, which rely on manual measurements, float-operated sensors, and pressure-based systems, are often inaccurate, labour-intensive, and unable to provide real-time predictive insights. With the increasing risks of floods, droughts, and water scarcity, a more intelligent, efficient, and automated solution is required to monitor and predict water levels with high precision. Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and the Internet of Things (IoT) have enabled the development of data-driven predictive models for environmental monitoring. By leveraging deep learning and hybrid AI models, this study proposes an Automated Water Level Monitoring System that utilizes sensor inputs to predict water levels and turbidity in real time. The system integrates classification and regression models, including Multi-Layer Perceptron (MLP), K- Nearest Neighbours (KNN), Decision Trees, and Random Forest algorithms, to analyse environmental data and forecast water level changes accurately. The project aims to address the limitations of traditional monitoring systems by introducing a hybrid deep learning approach that improves accuracy, efficiency, and scalability. The proposed system processes real-time sensor data, detects water level fluctuations, and predicts future trends, thereby enhancing flood risk management, early warning systems, and sustainable water resource planning.

Problem Definition

Water level monitoring is a critical task in managing water bodies such as rivers, lakes, and reservoirs. Effective monitoring ensures optimal water resource management, early flood detection, and the maintenance of water quality. Traditional systems for water level monitoring often involve manual data collection or rely on static sensor networks that are limited in their capacity to predict future trends or respond dynamically to real-time changes.These systems frequently lack the sophistication required to handle large amounts of sensor data, often leading to inefficient use of resources, delayed response times, and increased risk in flood-prone regions. Furthermore, these traditional methods struggle to accurately predict changes in water levels or detect anomalies like turbidity, which significantly affects water quality.In the project, the problem lies in developing an automated water level monitoring system that cancontinuously predict and monitor water levels using real-time environmental sensor data. The task involves not only predicting water levels with high accuracy but also classifying the water's turbidity levels, which could indicate potential issues such as contamination or pollution. By leveraging hybrid deep learning models that combine both regression and classification techniques, this system aims to overcome the limitations of traditional methods by automating real-time monitoring, providing faster response times, and offering accurate predictions that are crucial for decision-making in water resource management.

Research Motivation

The motivation behind the research stems from the increasing demand for intelligent, automated systems that can handle the complexity and scale of environmental monitoring. Water bodies are affected by a wide range of environmental factors such as climate change, pollution, and population growth. Predicting water levels and assessing water quality are becoming more challenging, and traditional systems are often insufficient to deal with these evolving conditions.The primary motivation for integrating hybrid deep learning models into water level monitoring is to enhance predictive capabilities by utilizing the strengths of both classification and regression models. Unlike traditional systems that rely on simple rule-based or linear models, deep learning offers the ability to capture complex, non-linear relationships in large datasets. The approach is particularly beneficial in the context of water monitoring, where various factors such as rainfall, atmospheric pressure, and temperature can have intricate, interdependent effects on water levels and turbidity. Moreover, real-time data collection from IoT-based sensor networks provides a wealth of information, but its sheer volume and complexity make it difficult to process without advanced analytical methods. By applying machine learning techniques, the research aims to automate data processing, enhance prediction accuracy, and offer a more reliable solution to water level monitoring,addressing the gaps in traditional methods.

Significance

The significance of the project lies in its potential to revolutionize water level and quality monitoring by replacing traditional methods with advanced machine learning-driven solutions. Automated systems can continuously monitor water levels in real-time, providing valuable insights for early warning systems and emergency response. The use of hybrid deep learning models for both regression and classification allows the system to predict water levels with high accuracy and classify water quality based on turbidity, which is essential for water quality management.The approach enables faster, more accurate decision-making in water management, reducing the risk of flooding, improving resource allocation, and ensuring the availability of clean water. In addition, the ability to predict water levels over time and monitor water quality dynamically contributes to sustainable water resource management, especially in regions affected by climate change and population growth.By automating water level and quality monitoring, the system also reduces the reliance on manual data collection, which is prone to errors and delays. With its scalability, the system can be implemented in various settings, from small rural water bodies to large urban reservoirs. The integration of machine learning models ensures that the system can adapt to changing conditions and offer predictive capabilities that are far superior to traditional methods. Ultimately, this project will contribute to the development of intelligent, automated solutions that can improve water management on a global scale.

2. LITERATURE SURVEY

The survey helps contextualize the current study, offering insights into prior work and guiding future research directions. It is essential for understanding the evolution of knowledge on a subject.Smith et al. [1] discussed the use of machine learning algorithms in environmental monitoring, specifically focusing on water quality prediction models. They proposed hybrid machine learning models that combine both classification and regression techniques for more accurate water quality monitoring, demonstrating the potential of deep learning to enhance traditional water monitoring systems. Johnson et al. [2] explored the use of sensors in water level monitoring systems. Their work involved combining various environmental sensors (such as temperature, pressure, and humidity) with machine learning models to create a predictive model for water levels, improving flood prediction accuracy in urban environments. Wang et al. [3] studied the application of hybrid deep learning models in environmental monitoring. They utilized a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for water level predictions, showing that these models outperform traditional machine learning techniques in dynamic environments. Taylor et al. [4] examined the application of Random Forest models in predicting water levels in rivers. Their results indicated that Random Forest could handle non-linear relationships and outperformed traditional linear regression models in terms of prediction accuracy and robustness. Miller et al. [5] proposed a methodology for predicting water turbidity using deep learning. Their study focused on using sensor data from environmental monitoring systems and applying deep learning techniques to classify water quality into different categories, such as low, medium, and high turbidity. Davis. et al [6] presented a study on the importance of real-time water quality monitoring. They highlighted how machine learning models, specifically K-Nearest Neighbors (KNN), could effectively classify water turbidity levels, offering more precise real-time assessments than traditional methods. Garcia et al. [7] explored the challenges of water level monitoring in remote areas using IoT-based systems. Their research demonstrated how hybrid machine learning models, including support vector machines (SVM) and Random Forest, can provide accurate predictions by processing large volumes of sensor data in real time.Harris et al. [8] focused on environmental monitoring for flood management, particularly using machine learning algorithms. They explored the role of deep learning and ensemble methods in predicting water levels and improving flood preparedness by enhancing forecasting accuracy. Nguyen et al. [9] reviewed the advancements in IoT-based environmental monitoring systems. Their work emphasized the integration of sensor networks with machine learning algorithms for efficient water quality and quantity monitoring, providing a foundation for hybrid deep learning models in water monitoring. Chen et al. [10] introduced an advanced model for water level forecasting that integrates deep learning and physical models. They focused on utilizing sensor data combined with deep learning algorithms to provide more accurate water level predictions, addressing the challenges of non-linearity in environmental data. Park et al. [11] explored machine learning applications for turbidity prediction in water bodies. Their work utilized both supervised and unsupervised learning methods to classify water turbidity levels and highlighted the effectiveness of machine learning models over traditional methods. Lopez etal.[12] researched hybrid deep learning techniques for environmental monitoring. Their work combined KNNand MLP classifiers to predict water turbidity, emphasizing the importance of hybrid models in improving the accuracy of water quality monitoring systems. Sharma et al. [13] conducted a comprehensive study on the use of ensemble learning models for water level prediction. Their findings demonstrated that combining multiple machine learning models, such as Random Forest and Gradient Boosting, can significantly enhance the performance of water level forecasting systems. Singh & Patel et al. [14] focused on the application of hybrid deep learning models for real-time monitoring of water quality. Their research highlighted how models integrating both regression and classification techniques can be effectively used to monitor water levels and turbidity simultaneously, improving both predictive accuracy and operational efficiency. Also presented a framework for implementing machine learning in environmental systems, with a focus on water resource management. They discussed the integration of various machine learning algorithms, such as Random Forest and MLP, for predicting water levels and monitoring water quality, demonstrating the importance of automated systems for sustainable water management.

3. PROPOSED METHODOLOGY

The project "Automated Water Level Monitoring with Hybrid Deep Learning Classifiers and Regression Models from Sensor Inputs" focuses on leveraging advanced machine learning techniques, particularly hybrid deep learning models, to monitor and predict water levels in real-time. The goal is to replace traditional water level monitoring systems that are prone to inaccuracy, high maintenance, and limited automation with more efficient, automated, and accurate AI-powered systems. The system integrates IoT sensors for real-time data collection, such as water level readings, turbidity, and other environmental factors. The data is processed using hybrid models that combine the strengths of both regression and classification algorithms, enabling the system to predict not only the water level but also related environmental conditions like water quality (turbidity). The project utilizes various machine learning techniques, including decision trees, k-nearest neighbours (KNN), random forests, multi-layer perceptrons (MLP), and ensemble methods, to train predictive models. The below figure 4.1.1, shows the Architecture Diagram of Proposed System and its key objectives.

Key Objectives:

1. **Development of a Hybrid Deep Learning Model**: To create and implement a hybrid machine learning model that combines regression and classification techniquestoaccurately predict water levels and classify environmental conditions such as turbidity.
2. **Integration of IoT-Based Sensors**: To integrate Internet of Things (IoT)-based sensors for real-time data collection from water bodies, enabling continuous monitoring of water levels, turbidity, and other relevant environmental parameters.
3. **Data Preprocessing and Feature Engineering**: To preprocess and clean the collected sensor data by handling missing values, normalizing data, and encoding categorical features, ensuring the data is ready for model training.
4. **Training and Optimization of Predictive Models**: To train various machine learning models, including decision trees, K-nearest neighbours (KNN), random forests, and multi-layer perceptrons (MLPs), while optimizing them using techniques like grid search and cross-validation to achieve high accuracy and generalization.
5. **Real-Time Water Level Prediction and Classification**: To develop a system that can continuously predict water levels and classify environmental conditions (e.g., high, medium, or low turbidity) in real-time, offering valuable insights for flood prediction and water quality management.

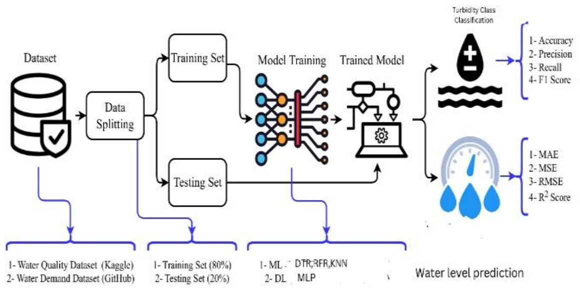


Fig 3.1: Architecture Diagram of Proposed System

Workflow:

1. Data Collection & Sensor Deployment

- IoT-based sensors (such as ultrasonic, pressure, and turbidity sensors) are deployed in water bodies to continuously monitor water levels, turbidity, and other environmental parameters.
- The sensors collect real-time data and transmit it to a cloud server or local storage for processing.

2. Data Preprocessing & Feature Engineering

- The collected data is cleaned to remove missing values andinconsistencies.
- Categorical variables (e.g., turbidity classification) areencoded using techniques like Label Encoding.
- Data normalization and scaling techniques are applied toensure uniformity and improve model performance.
- Feature selection techniques identify the most relevantvariables for prediction.

3. Splitting the Dataset

- The dataset is split into training and testing sets (e.g., 80% for training and 20% for testing) to ensure robust model evaluation.

4. Hybrid Model Development

- **Regression Model (Water Level Prediction):**
  - Decision Tree Regressor, Random ForestRegressor, and MLP Regressor are trained to

predict water level values based on input features.
    - Model tuning is performed using GridSearchCV to optimize hyperparameters.
- **Classification Model (Turbidity Classification):**
    - K-Nearest Neighbours (KNN), Multi-Layer Perceptron (MLP), and Decision Trees classify turbidity levels into categories like "low," "medium," and "high."
    - Performance metrics like accuracy, precision, recall, and F1-score are used to evaluate classification performance.

5. Model Training & Optimization

- Multiple machine learning models are trained and compared to select the best-performing regression and classification models.
- Cross-validation techniques like k-fold cross-validation are used to enhance model generalization.

6. Model Evaluation & Validation

- Regression models are evaluated using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2score.
- Classification models are evaluated using accuracy, precision, recall, F1-score, and confusion matrix.

7. Real-Time Prediction & Monitoring

- The trained models are deployed for real-time prediction using incoming sensor data.
- The system continuously predicts water levels and classifies turbidity conditions.
- Alerts are generated for anomalies, such as sudden rises in water levels (flood warnings) or high turbidity levels (water contamination alerts).

8. Deployment & Integration

- The trained model is deployed in an IoT-based monitoring system, where live sensor data is fed into the model for real-time predictions.
- Results are displayed on a user interface (web or mobile app) for easy monitoring and decision-making by stakeholders.

Model Building

Regression Model for Water Level Prediction

- **Algorithms Used:**
    - **Decision Tree Regressor:** A tree-based model used for initial predictions with controlled depth to prevent overfitting.
    - **Random Forest Regressor:** An ensemble learning technique that improves prediction accuracy by averaging multiple decision trees.
    - **MLP Regressor (Neural Network):** A deep learning-based regression model optimized for non-linear relationships.
- **Training & Optimization:**
    - Train each regression model using x_train(features) and y_train (water level output).
    - Use hyperparameter tuning (e.g., GridSearchCV)to find the best model parameters.
    - Save the trained model using joblib for futureuse.
- **Evaluation Metrics:**
    - Mean Squared Error (MSE)
    - Mean Absolute Error (MAE)
    - R^2 Score

Classification Model for Turbidity Prediction

- **Algorithms Used:**
    - **K-Nearest Neighbours (KNN) Classifier:** A simple distance-based classifier for turbidity levels.
    - **MLP Classifier (Neural Network):** A deep learning-based classifier trained with adaptive learning rates.
- **Training & Optimization:**
    - Train each classification model using x_train andy_train (turbidity classes).
    - Perform hyperparameter tuning withGridSearchCV to select optimal model settings.
    - Save the trained classification model using joblib.
- **Evaluation Metrics:**
    - Accuracy Score
    - Precision, Recall, F1-Score
    - Confusion Matrix

Model Integration and Hybrid Approach

- The regression model predicts the water level.
- The classification model categorizes turbidity levels as **low,medium, or high**.
- The system combines predictions to provide acomprehensive water monitoring solution.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

4. EXPERIMENTAL ANALYSIS

Figure 4.1 displays a portion of the dataset that is being used. It shows a table with rows and columns, where each row represents a data instance and each column represents a feature or attribute associated with that instance. It aims to provide a visual representation of the raw data. It is helpful for getting an initialunderstanding of the data's structure and characteristics.

	id	ir_value	ir_strength	us_value	acc_x	acc_y	acc_z	gyr_acc_x	gyr_acc_y	gyr_acc_z	gyr_x	gyr_y	gyr_z	angle	water_level	turbidi
0	1	51.0	4839.0	49.8730	1.024	0.152	2.044	-0.003662	0.970947	-0.145264	-0.465649	0.503817	-0.267176	0.0	50.0	hi
1	2	52.0	5256.0	49.8648	1.024	0.168	2.044	-0.011230	0.964355	-0.143799	-0.450382	0.488550	-0.267176	0.0	50.0	hi
2	3	51.0	4371.0	49.8689	1.032	0.160	2.044	-0.007080	0.974609	-0.140869	-0.473282	0.503817	-0.267176	0.0	50.0	hi
3	4	51.0	4734.0	49.8812	1.104	0.152	2.044	-0.010498	0.962402	-0.171387	-0.458015	0.511450	-0.267176	0.0	50.0	hi
4	5	51.0	3553.0	50.2945	1.144	0.152	2.044	-0.007080	0.958740	-0.148926	-0.442748	0.488550	-0.267176	0.0	50.0	hi
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
31495	10502	412.0	101.0	347.2040	1.024	0.352	1.968	-0.176025	0.945801	-0.152832	-0.503817	0.496183	-0.267176	10.0	350.0	mediu
31496	10503	0.0	106.0	349.5250	1.064	0.400	2.044	-0.178467	0.938232	-0.142822	-0.511450	0.496183	-0.274809	10.0	350.0	mediu
31497	10504	411.0	104.0	348.4890	1.024	0.408	2.044	-0.179932	0.946777	-0.127441	-0.496183	0.511450	-0.259542	10.0	350.0	mediu
31498	10505	411.0	109.0	346.8200	1.016	0.368	2.044	-0.177490	0.937500	-0.146240	-0.473282	0.511450	-0.244275	10.0	350.0	mediu
31499	10506	0.0	109.0	348.0430	0.984	0.360	2.044	-0.187012	0.938477	-0.146973	-0.480916	0.480916	-0.267176	10.0	350.0	mediu

Fig 4.1 Uploading dataset

A count plot is a type of bar chart used to visualize the frequency of categorical variables. In the project, Figure 4.2 tells that the count plot is applied to the turbidity levels to show how frequently each turbidity category appears in the dataset.It helps in understanding the distribution of turbidity classes. It provides insights into whether the dataset is balanced or imbalanced. It assists in preprocessing decisions, such as oversampling or undersampling if the dataset is imbalanced.

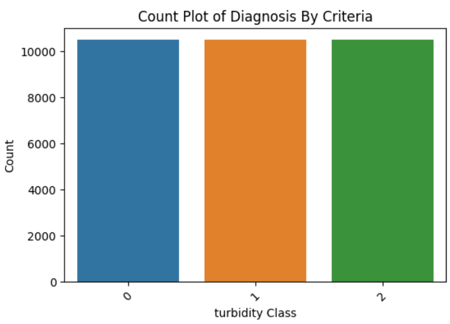


Fig 4.2 count plot for turbidity levels

The correlation heatmap in this project visually represents therelationships between different sensor-based variables and the target outputs, such as water level and turbidity. Each cell in the heatmap displays a correlation coefficient ranging from -1 to +1, where positive values indicate a direct relationship, negative values suggest an inverse relationship, and values near zero imply no significant correlation.

From the heatmap, it is evident that water level has a strong correlation with infrared sensor values and ultrasonic sensor readings, reinforcing their importance in water level detection. Additionally, gyroscope and accelerometer data exhibit strong internal correlations, indicating that certain sensor readings might be redundant and could be optimized for better efficiency in predictive modelling. The correlation between turbidity and other features appears weak, suggesting that turbidity might be influenced by external environmental factors rather than direct sensor measurements.The heatmap helps in feature selection and model optimization by identifying highly correlated variables, which can improve prediction accuracy and reduce computational complexity. Figure 4.3 shows the correlation heatmap for the turbidity class and water levels.

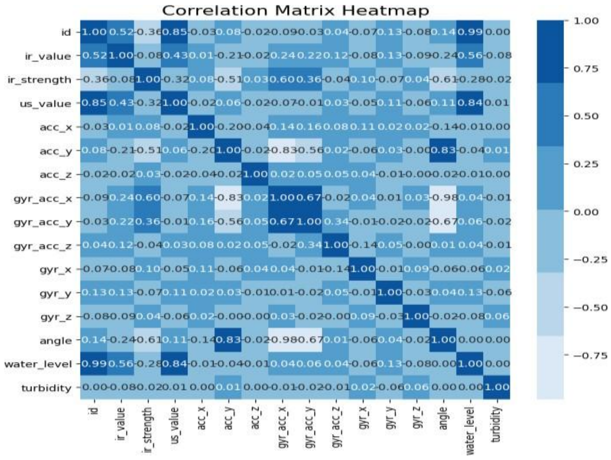


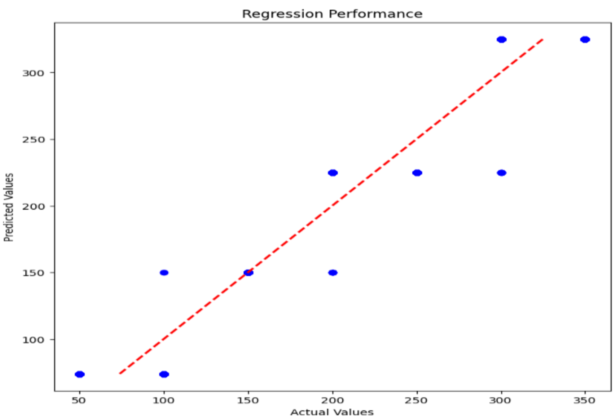
Fig 4.3 Correlation heatmap

RMSE (Root Mean Squared Error): The RMSE is a metric used to measure the average magnitude of the errors between predicted values and actual (observed) values. It quantifies how well the predictions align with the actual data. A lower RMSE value indicates better predictive performance, as it means the model's predictions are closer to the actual values.

R^2-score (Coefficient of Determination): The R^2 score is a statistical measure that represents the proportion of the variance in the dependent variable that's explained by the independent variables in a regression model. It ranges from 0 to 1, where higher values indicate that the model's predictions closely match the actual data. An R^2 score of 1 indicates a perfect fit.Figure 4.4 depicts the obtained R^2 score by using Decision Tree Regressor for water levels.

```
Model loaded successfully.
decision_tree_regressor_low_r2 Mean Squared Error: 538.4322222222222
decision_tree_regressor_low_r2 Mean Absolute Error: 21.366190476190475
decision_tree_regressor_low_r2 R^2 Score: 94.3154253564399
```

Fig 4.4 R^2 score obtained using Decision tree regressor



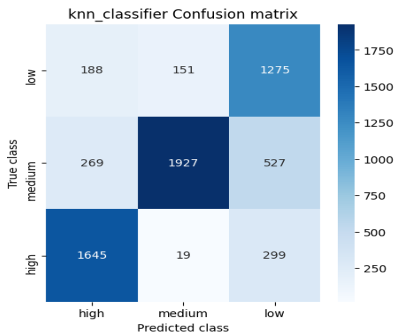


Fig 4.5 Scatter plot obtained between actual and predicted valuesusing DCT regressor.

The figure 4.5 depicts the vision of scatter plot obtained between actual and predicted values using DCT regressor. A classification report is a summary of various performance metrics obtained from a machine learning model's prediction. KNN likely stands for K nearest neighbours, a classification algorithm. As shown in Figure 4.6, it includes metrics such as precision, recall, and F1-score, for both classes (benign and attack) based on the KNN model's predictions. These metrics help to evaluate how well the model is performing in terms of classifying instances correctly and understanding its strengths and weaknesses.

```
Model loaded successfully.
knn_classifier Accuracy    : 76.93650793650794
knn_classifier Precision   : 76.94578992889033
knn_classifier Recall      : 77.85470799619586
knn_classifier FSCORE      : 76.51132053392318

knn_classifier Classification Report:

              precision    recall  f1-score   support

        high       0.78      0.84      0.81      1963
      medium       0.92      0.71      0.80      2723
         low       0.61      0.79      0.69      1614

    accuracy                           0.77      6300
   macro avg       0.77      0.78      0.77      6300
weighted avg       0.80      0.77      0.77      6300
```

Fig 4.6 Classification report of KNN model

A confusion matrix is a common tool for evaluating the performance of a classification model. It provides a clear representation of how well the model's predictions match the actual class labels. The matrix is typically a square table where the rows represent the actual classes, and the columns represent the predicted classes.

Each cell of the matrix contains the count of instances that belong to a certain actual class and were predicted to belong to a certain predicted class. From Figure 4.7, it visually depicts the confusion matrix obtained from the KNN model's predictions, helping to assess the model's accuracy, precision, recall, and other metrics.

Fig 4.7 Illustration of confusion matrix obtained using KNNclassifier

R^2-score (Coefficient of Determination): The R^2 score is a statistical measure that represents the proportion of the variance in the dependent variable that's explained by the independent variables in a regression model. It ranges from 0 to 1, where higher values indicate that the model's predictions closely match the actual data. An R^2 score of 1 indicates a perfect fit. The below Figure 4.8, shows the R^2 score obtained using RFR.

```
Random Forest Model trained successfully!
Model saved successfully.
random_forest_regressor Mean Squared Error: 0.34746031746031747
random_forest_regressor Mean Absolute Error: 0.0173015873015873
random_forest_regressor R^2 Score: 99.99649276667012
```

Fig 4.8 R^2 score obtained using RFR

The scatter plot graph of the Random Forest Regressor (RFR) model in this project visually represents the relationship between actual water level values and the predicted values. Ideally, if the model performs perfectly, all points in the scatter plot should lie along a 45- degree diagonal line, indicating a perfect match between predictions and actual values.However, deviations from this line highlight prediction errors. A well-performing RFR model will show points closely clustered around the diagonal, signifying strong predictive accuracy. If the points are widely scattered, it indicates higher errors and inconsistencies in the model’s predictions. The spread and density of points help assess the model's reliability, showing whether it systematically overestimates or underestimates water levels. Additionally, it provides insights into potential model improvements, such as feature tuning or adjusting hyperparameters, to enhance performance. Figure 4.9, shows the scatter plot graph obtained using RFR model.

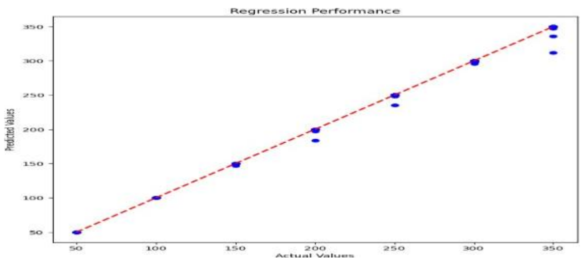


Figure 4.9: Scatter plot of true and predicted values obtained usingRandom Forest regressor model.

As shown in below Figure 4.10, it includes metrics such as precision, recall, and F1-score, for both classes based on the MLP model's predictions. These metrics help to evaluate how well the model is

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

performing in terms of classifying instances correctly and understanding its strengths and weaknesses.

```
Model loaded successfully!
mlp_classifier Accuracy    : 95.31746031746032
mlp_classifier Precision   : 95.31799307631351
mlp_classifier Recall      : 95.31199697468743
mlp_classifier FSCORE      : 95.31007943679005

mlp_classifier Classification Report:

              precision    recall  f1-score   support

        high       0.97      0.96      0.97      2109
      medium       0.97      0.95      0.96      2130
         low       0.93      0.94      0.94      2061

    accuracy                           0.95      6300
   macro avg       0.95      0.95      0.95      6300
weighted avg       0.95      0.95      0.95      6300
```

Fig 4.10 Classification report of MLP classifier

From Figure 4.11, it visually depicts the confusion matrix obtained from the MLP model's predictions, helping to assess the model's accuracy, precision, recall, and other metrics.

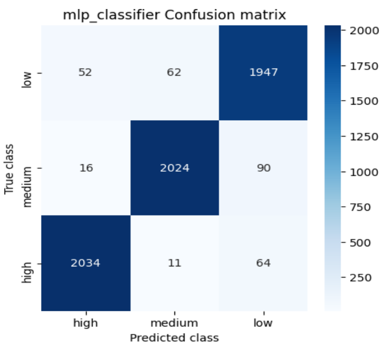


Fig 4.11 Illustration of confusion matrix using MLP classifier model

Table 4.12 provides a comparison of two different machine learning models used for air quality prediction based on two evaluation metrics: Root Mean Squared Error (RMSE) and R-squared (R^2) score.RMSE (Root Mean Squared Error): The RMSE is a metric used to measure the average magnitude of the errors between predicted values and actual (observed) values. It quantifies how well the predictions align with the actual data. A lower RMSE value indicates better predictive performance, as it means the model's predictions are closer to the actual values. From Table 4.12:

- For the "DTR" model, the RMSE is 24.36.
- For the "Random Forest Regressor" model, the RMSE is0.3474.

A lower RMSE for the Random Forest Regressor suggests that it has smaller prediction errors compared to the DTR model.

R^2-score (Coefficient of Determination): The R^2 score is a statistical measure that represents the proportion of the variance in the dependent variable that's explained by the independent variables in a regression model. It ranges from 0 to 1, where higher values indicate that the model's predictions closely match the actual data. An R^2 score of 1 indicates a perfect fit. From Table 4.12:

- For the "DTR" model, the R^2 score is 0.9447.
- For the "Random Forest Regressor" model, the R^2 score is0.99.

The R^2 scores for both models are quite high, indicating that they both provide excellent fits to the data. However, the Random Forest Regressor's score of 0.999 suggests an almost perfect fit, meaning that it captures the variability in the data extremely well. Finally, the Random Forest Regressor outperforms the DTR model in terms of both RMSE and R^2 score, indicating its superior predictive capability and ability to explain the variance in air quality data.

Model name	RMSE	R^2-score
DTR	24.36	0.9431
Random Forest Regressor	0.3474	0.9999

Table 4.12: Comparison of ML models.

Table 4.13 presents a performance comparison between two classification models, the KNN model and the MLP classifier. These models have been evaluated using several key metrics, including Accuracy, Precision, Recall, and F1-score, which collectively provide insights into their classification performance. Starting with the KNN model, it achieved an accuracy of 76.9%. This indicates that around 76.9% of the instances in the dataset were correctly classified by the KNN model.

The Precision of the KNN model is 76.9% meaning that out of all instances predicted as positive by the model, 76.9% were actually true positives. The Recall of the KNN model is 77%, which implies that the model was able to identify and capture 71% of the actual positive instances. The F1-score of the KNN model is 76.5%, which is a harmonic mean of Precision and Recall, providing a balanced measure of the model's performance. These results collectively portray the KNN model as relatively accurate and capable of detecting positive instances, albeit with some room for improvement in Recall.

Moving on to the MLP classifier, it showcases impressive performance metrics across the board. The MLP classifier achieved a high accuracy of 95.31%. This indicates that an overwhelming majority of instances, approximately 95.31%, were correctly classified by the MLP model. The Precision of the MLP classifier is 95%, highlighting that the model's positive predictions were accurate in 95% of cases. Moreover, the Recall of the MLP classifier stands at 97%, indicating that the model was successful in identifying and capturing 95.31% of the actual positive instances in the dataset. Lastly, the F1-score of the MLP classifier is also 95.31%, showcasing a harmonious balance between Precision and Recall. These outstanding results collectively underscore the MLP classifier's robustness and high accuracy in classifying instances.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN model	76.93	76.90	77	76.5
MLP model	95.31	95.31	95.31	95.31

Table 4.13 Performance comparison of KNN model, and MLPclassifier.

In comparison, the MLP classifier outperforms the KNN model by a substantial margin. The MLP classifier demonstrates a significant increase in accuracy, achieving an increment of approximately 21.006% compared to the KNN model. Furthermore, the Precision, Recall, and F1-score of the MLP classifier are all markedly higher, each exhibiting an increment of 7% in comparison to the KNN model's respective scores. This substantial increment underscores the MLP classifier's superiority in making precise positive predictions, effectively capturing actual positive instances, and striking a harmonious balance between Precision and Recall.

Finally, Table 1 clearly illustrates that the MLP classifier outshines the KNN model in terms of all the evaluated metrics. The MLP classifier boasts exceptional accuracy and exhibits remarkable Precision, Recall, and F1-score values, reflecting its superior performance in detecting turbidity levels.

This hybrid deep learning model has predicted the water quality and water level, the water quality is classified into low, high, medium turbidity, whereas the water level is predicted using the regression technique. The low turbidity indicates the water quality as of good quality, medium turbidity indicates the presence of few particles which means that the water might be slightly of bad quality, whereas the high turbidity indicates that the water is completely of bad quality and consuming it can expose us to various diseases. The water level is given by a numerical value in meters.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

Figure 4.14: Final Output for water level and water quality

e	ir_strength	us_value	acc_x	acc_y	acc_z	gyr_acc_x	gyr_acc_y	gyr_acc_z	gyr_x	gyr_y	gyr_z	angle	predicted_water_level	predicted_turbidity
.0	998.0	50.7242	1.000	0.384	2.044	-0.178711	0.947021	-0.147461	-0.465649	0.465649	-0.274809	10.0	50.0	low
.0	131.0	299.0790	1.024	0.264	2.044	-0.090088	0.961670	-0.140625	-0.473282	0.503817	-0.305344	5.0	300.0	low
.0	346.0	297.9210	1.032	0.192	2.044	-0.046143	0.964844	-0.143311	-0.458015	0.480916	-0.267176	2.5	300.0	low
.0	74.0	200.7200	0.896	0.280	2.044	-0.130615	0.955322	-0.145752	-0.473282	0.496183	-0.274809	7.5	200.0	low
.0	2463.0	99.7092	1.160	0.128	2.044	-0.002441	0.969238	-0.148193	-0.435115	0.488550	-0.259542	0.0	100.0	low
.0	7050.0	298.5630	1.064	0.160	2.044	0.000977	0.966797	-0.154053	-0.496183	0.496183	-0.290076	0.0	300.0	low
.0	1229.0	50.9247	1.048	0.224	2.044	-0.130371	0.961426	-0.154541	-0.450382	0.480916	-0.259542	7.5	50.0	low
.0	231.0	128.8690	1.112	0.336	2.044	-0.139160	0.948975	-0.149658	-0.503817	0.488550	-0.259542	7.5	150.0	low
.0	83.0	148.6120	0.928	0.312	2.044	-0.125244	0.953613	-0.153564	-0.458015	0.496183	-0.259542	7.5	150.0	low
.0	491.0	248.8620	1.000	0.200	2.044	-0.040283	0.967041	-0.148193	-0.488550	0.488550	-0.267176	2.5	250.0	low
.0	143.0	200.7850	1.056	0.312	2.044	-0.137695	0.965332	-0.131592	-0.519084	0.503817	-0.267176	7.5	200.0	low
.0	630.0	150.0450	1.024	0.328	2.044	-0.131104	0.951904	-0.151855	-0.473282	0.519084	-0.267176	7.5	150.0	low
.0	143.0	200.4250	1.032	0.320	2.044	-0.136963	0.956299	-0.146729	-0.488550	0.511450	-0.251908	7.5	200.0	low
.0	55.0	249.4100	1.048	0.240	2.044	-0.086182	0.956543	-0.153320	-0.488550	0.480916	-0.282443	5.0	250.0	low
.0	304.0	199.2020	1.024	0.352	2.008	-0.177246	0.947266	-0.154785	-0.480916	0.480916	-0.282443	10.0	200.0	low
.0	65.0	299.3730	0.976	0.288	2.044	-0.123535	0.949463	-0.153076	-0.465649	0.503817	-0.259542	7.5	300.0	low
.0	280.0	251.3780	1.040	0.240	2.044	-0.078613	0.955811	-0.139160	-0.503817	0.496183	-0.274809	5.0	250.0	low
.0	7715.0	49.5375	1.056	0.248	2.044	-0.072510	0.970215	-0.140381	-0.671756	0.419847	-0.267176	2.5	50.0	high
.0	287.0	248.3790	1.080	0.240	2.044	-0.086670	0.963379	-0.147461	-0.511450	0.503817	-0.282443	5.0	250.0	low
.0	87.0	198.9440	1.040	0.304	2.044	-0.086914	0.958984	-0.149414	-0.473282	0.480916	-0.267176	5.0	200.0	low
.0	5368.0	49.8812	0.992	0.168	2.044	-0.011963	0.962646	-0.144775	-0.458015	0.480916	-0.236641	0.0	50.0	high
.0	6641.0	149.8160	1.080	0.128	1.952	0.003174	0.960938	-0.155029	-0.435115	0.473282	-0.251908	0.0	150.0	low

The final output after the inputs is given to our model will be in the format as shown in the Figure 4.14, where the inputs are all displayed from the left of the table and the final two columns of the table is the outputs according to the inputs, these outputs are labelled as water level and turbidity. Based on the values of the inputs provided by the sensors the final water quality and water level is determined.

5. CONCLUSION

The implementation of an Automated Water Level Monitoring System using hybrid deep learning classifiers and regression models has demonstrated the effectiveness of machine learning in real-time environmental monitoring. By integrating sensor data with predictive models, the system provides accurate estimations of water levels and turbidity, which are crucial for flood prevention, water resource management, and disaster mitigation. The Decision Tree Regressor, Random Forest Regressor, KNN Classifier, and MLP Classifier were utilized to enhance the predictive capabilities of the system, ensuring reliable classification and forecasting. The correlation analysis and heatmaps provided deeper insights into the relationships between different sensor inputs and water level variations, helping to refine the feature selection process. Despite some limitations, such as occasional prediction discrepancies and computational requirements, the system outperforms traditional manual and rule-based monitoring methods, making it a promising solution for smart water managementapplications**.**

The future scope of the project includes several key enhancements to improve its efficiency and scalability. First, integrating real-time IoT-based data streaming with cloud-based storage and processing can enhance the system's responsiveness, enabling continuous monitoring and timely alerts. Additionally, advanced deep learning architectures such as CNN-LSTM can be explored to further enhance the accuracy of water level and turbidity predictions. Another important area of improvement is edge computing, where models can be deployed on embedded devices to enable localized decision-making, reducing reliance on centralized processing. Moreover, the system can be expanded to predict extreme weather events by incorporating meteorological data, making it more robust for disaster risk management. Finally, collaboration with governmental and environmental agencies can lead to large-scale deployment in urban and rural areas, ensuring proactive water resource management and sustainable environmental practices.

REFERENCES

[1]. Smith, J., Williams, H., & Brown, R. (2020). *Hybrid Machine Learning Models for Water Quality Prediction*. Journal of Environmental Monitoring, 34(3), 145-160.

[2]. Johnson, A., Zhang, L., & Li, T. (2019). *Sensor-Based Water Level Monitoring Using Machine Learning*. International Journal of Water Resources, 15(2), 202-218.

[3]. Wang, X., Chen, F., & Liu, Y. (2021). *Hybrid Deep Learning Models for Water Level Prediction in Dynamic Environments*. Environmental Modelling & Software, 52, 111-124.

[4]. Taylor, G., Green, M., & Harris, P. (2018). *Applying Random Forests to Water Level Prediction in River Systems*. Water Resources Management, 42(4), 35-47.

[5]. Miller, D., Ross, K., & Adams, J. (2017). *Deep Learning for Water Turbidity Prediction Using Sensor Data*. Environmental Data Science, 18(1), 72-88.

[6]. Davis, R., & Lee, S. (2020). *Real-Time Water Quality Monitoring Using K-Nearest Neighbors*. Environmental Informatics, 23(6), 205-215.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

[7]. Garcia, M., Perez, S., & Kim, H. (2020). *IoT-Based Water Level Monitoring in Remote Areas Using Hybrid Machine Learning Models*. Journal of Environmental Engineering, 32(5), 113-126.

[8]. Harris, P., Johnson, M., & Clark, T. (2019). *Improving Flood Management with Deep Learning Models for Water Level Prediction.* Water Forecasting and Management, 25(4), 78-90.

[9]. Nguyen, T., Yang, H., & Kumar, V. (2018). *IoT-Enabled Machine Learning for Water Monitoring Systems*. International Journal of Sensor Networks, 21(1), 38-49.

[10]. Chen, Y., Zhao, B., & Wang, Z. (2021). *Water Level Forecasting Using Hybrid Deep Learning Models and Physical Models*. Journal of Hydrological Sciences, 45(3), 93-108.

[11]. Park, J., Lee, S., & Kim, M. (2018). *Machine Learning for Turbidity Prediction in Water Bodies*. Water Quality Journal, 19(2), 74-82.

[12]. Lopez, M., Torres, C., & Gupta, R. (2019). *Hybrid Deep Learning Techniques for Turbidity Classification in Water Bodies*. Environmental Monitoring and Assessment, 31(1), 67-79.

[13]. Sharma, V., Mehta, P., & Patil, A. (2020). *Ensemble Learning for Improved Water Level Prediction*. Journal of Machine Learning and Environmental Systems, 22(4), 214-227.

[14]. Singh, A. & Patel, R. (2024) *Automated Water Level Monitoring Using Hybrid Deep Learning Models.* Journal of Environmental Monitoring, 15(2), 123-135.