International Journal of Informatics, Information System and Computer Engineering

INJIISCOM

Analysis and Prediction of Seasonal Water Quality of Nepal Using Machine Learning Approach with SHAP Analysis

*Biplov Paneru *, Bishwash Paneru **, Sanjog Chhetri Sapkota ***

[*]Dept. of Applied sciences and Chemical Engineering, Institute of Engineering, Nepal
[**]Dept. of Electronics and Communication Engineering, Pokhara University
[***]Nepal Research and collaboration centre
*Corresponding Email: Biplov Paneru (biplovp019402@nec.edu.np)

ABSTRACTS

Water quality is a crucial concern worldwide, including in Nepal, where efficient monitoring is essential for safe drinking water and preventing waterborne illnesses. This study employs machine learning to analyze and forecast the seasonal water quality index (WQI) of Nepalese well water. Hybrid models with nested cross-validation were introduced, using methods like CatBoost, Decision Tree, Logistic Regression, MLP-GRU, and LSTM-GRU hybrids. Performance metrics included R^2, accuracy, and RMSE. CatBoost achieved the highest classification accuracy (99.35%), while the LSTM-GRU hybrid excelled in capturing complex temporal patterns. Nested cross-validation demonstrated 96.13% accuracy with low standard deviation. Additionally, SHAP analysis identified key predictive factors using the SVM model. This research highlights machine learning's potential in predicting and managing water quality effectively.

ARTICLE INFO

Article History:
Received 08 Jul 2024
Revised 30 Aug 2024
Accepted 19 Sept 20xx
Available online 24 Dec 2024
Publication date 01 Dec 2025

Keywords:
Water quality,
Machine Learning,
LSTM-GRU
SHAP analysis
DO
pH

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

1. INTRODUCTION

In underdeveloped countries, a variety of issues affecting the water supply process can lead to water tap contamination. While machine learning techniques have gained popularity for making accurate water quality predictions, gathering the necessary data for modeling in underdeveloped nations has proven to be difficult (Kuroki, et al., 2023)[1]. Water quality is extremely important to humans, animals, plants, industries, and the environment. Water quality is measured through Water Quality Index (WQI). In this study, the parameters were optimized and tuned to increase the accuracy of numerous machine learning models and techniques that were used to measure WQI and WQC (Shams, et al., 2024). Conversely, lakes and reservoirs serve as essential sources of water. These reservoirs play a vital role in sustaining life, offering clean water and supporting a rich diversity of aquatic ecosystems (Solanki, et al., 2015). Groundwater plays an essential role in maintaining natural water reserves, serving as a vital resource for drinking water, farming, and diverse industrial uses. However, industrial and agricultural activities greatly impact groundwater quality, often leading to contamination. This highlights the necessity of evaluating water quality to ensure safe consumption and efficient irrigation practices (Abbas, et al., 2014). To limit harmful components' access into water bodies, particularly rivers, timely monitoring and rapid decision-making are crucial. Conventional approaches to assessing water quality can occasionally be expensive and require significant time (Najafzadeh, & Basirian, 2023).

Fig. 1. Global water dryness problem

The analysis and prediction of water quality of a country or area using machine learning arpproach highlights a significant step towards utilizing artificial intelligence methods for water quality prediction and analysis making.

An important development in the use of artificial intelligence for environmental management is the analysis and forecast of water quality by machine learning techniques. Large datasets containing several water quality indicators, like dissolved oxygen (DO), pH, temperature, and electrical conductivity (EC) can be effectively examined to find trends and forecast future conditions by utilizing machine learning techniques. This makes it possible to identify possible problems with the quality of the water early on, which facilitates prompt intervention and improved resource management. Additionally, machine learning reduces the need for labor-intensive human sampling and testing by improving the speed and accuracy of water quality tests. In general, a more proactive, data-driven

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

approach to guaranteeing safe and sustainable water supplies is encouraged by the incorporation of machine learning into water quality analysis. Previously various works are carried out in field of water and few countries-based research used Machine Learning techniques for predictions. Many research like (Kuroki, et al., 2023; Shams, et al., 2024; Solanki, et al., 2015; Abbas, et al., 2024; Najafzadeh, & Basirian, 2023; Rahat, et al., 2023; Anand, et al., 2023; Perumal, et al., 2023) indicate huge research gap for the prediction and analysis for WQI for Nepal.

We used LightGBM, Random Forest (RF), and Support Vector Machine (SVM) and, which are based on both tap water quality and water source data collected by the government of Nepal. Additionally, logistic regression (LR) was employed to forecast E. coli contamination in water taps. By utilizing input data derived from the pseudo-pipeline network, SVM demonstrated solid performance, achieving an accuracy of 70% across 26 cities and 79% across 25 cities when Kathmandu was excluded. LR's accuracy was much lower for all cities (61%) than for 25 cities (79%) (Kuroki, et al., 2023) [1]. The dataset used in this investigation has 1991 cases and 7 characteristics. Furthermore, five assessment measures were used to evaluate the effectiveness of the classification approaches: precision, F1 score, recall, accuracy, and Matthews' Correlation Coefficient (MCC). The performance of the regression models was evaluated using four metrics: Mean Absolute Error (MAE), Median Absolute Error (MedAE), Coefficient of Determination (R^2), and Mean Square Error (MSE). Regarding classification, the testing results indicated that the GB model delivered the highest performance, predicting WQC values with an impressive accuracy of 99.50%. According to the experimental findings, the MLP regressor model surpassed the other regression models, reaching an R^2 value of 99.8% in forecasting WQI values (Shams, et al., 2024). Using the WEKA software, we processed secondary data provided by a third party concerning the Chaskaman River located near Nasik, Maharashtra, India. The research revealed that unsupervised deep learning methods demonstrated higher accuracy compared to supervised learning approaches. The results demonstrate that denoising autoencoders and deep belief networks can attain robustness and effectively handle data unpredictability (Solanki, et al., 2015).

Various machine learning classifiers were employed to predict the WQI, yielding results that show Gradient Boosting and Random Forest achieving the highest accuracies of 96% and 95%, respectively. SVM follows closely with a 92% accuracy, while KNN achieves 84%, and Decision Trees attain 77%. Traditional methods of water quality assessment are both time-consuming and prone to errors. However, the application of artificial intelligence and machine learning offers a disruptive solution, effectively overcoming these limitations. The study not only aimed to predict the Water Quality Index (WQI) but also performed an uncertainty analysis of the models using the R-factor, providing valuable insights into the consistency and reliability of the predictions. By combining accurate WQI predictions with uncertainty assessment, this approach offers a more thorough

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

understanding of water quality in Mirpurkash (Abbas, et al., 2014).

For the Hudson River, the WQI was determined using Landsat 8 OLI-TIRS imagery and four Artificial Intelligence (AI) models: Evolutionary Polynomial Regression (EPR), Gene Expression Programming (GEP), M5 Model Tree (MT), and Multivariate Adaptive Regression Spline (MARS). This process involved analyzing 13 water quality parameters (WQPs)—including dissolved oxygen arsenic, pH, alkalinity, magnesium, nitrate, sulfate, turbidity, potassium, sodium, fluoride, hardness, and chloride—within the New York area. Initially, Multiple Linear Regression (MLR) models were developed to establish relationships between these WQPs and the spectral indices derived from Landsat 8 OLI-TIRS images. The most strongly correlated spectral indices were then selected as inputs for the AI models. Subsequently, the WQI was computed based on the measured WQP values (Najafzadeh, & Basirian, 2023).

This study emphasizes predicting water quality through machine learning techniques. The approach evaluates water's color and overall condition to assess its usability for drinking or other purposes. Leveraging Convolutional Neural Networks (CNN), Keras, and TensorFlow, the model is trained to forecast water quality. Designed to be economical and efficient, the project offers a preliminary water quality assessment via an image processing application. The technique can analyze water samples using images from mobile devices and Google Earth. Furthermore, this paper introduces a method that utilizes a Long Short-Term Memory Network (LSTM) trained with Moderate Resolution Imaging Spectroradiometer (MODIS) satellite reflectance data, calibrated against Total Suspended Solids (TSS) measurements provided by the Ohio River Valley Water Sanitation Commission (ORSANCO). This methodology facilitates an in-depth empirical analysis and data-driven models capable of addressing spatial variability within watersheds while delivering reliable water quality predictions under uncertain conditions (Rahat, et al., 2023; Anand, et al., 2023).

According to experimental data, the proposed model demonstrated superior performance in predicting extreme values compared to both the mechanism and LSTM models. This model was tested in the Thamirabarani River basin. The maximum relative errors between the predicted and observed values were 7.58% for dissolved oxygen, 18.45% for chemical oxygen demand, and 22.25% for NH3─N. Regarding computational efficiency, the LSTM-GWO-FSO model developed surpassed the artificial neural network (ANN), recurrent neural network (RNN), and back propagation neural network (BPNN) models in performance (Perumal, et al., 2023).

Our work involves prediction and analysis of water quality of Nepal using a data published in a paper. The physical water quality of shallow groundwater in the southern Kathmandu Valley was assessed and analyzed. During the dry season, temperatures vary between 15.3 and 24.2 °C, pH levels range from 5.67 to 8.07, electrical conductivity spans from 230 to 2860 µS/cm, and dissolved oxygen levels are between 0.09 and 9.1 mg/L. In contrast, the wet season sees temperatures ranging from 19.6 to 27.3 °C, pH levels from 5.92 to 8.3, electrical conductivity from 183 to 3030 µS/cm, and dissolved oxygen between 0.19 and

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

7.9 mg/L. Water Quality Index (WQI) maps show that areas upstream of rivers generally have higher water quality compared to those downstream. Locations such as Kalanki and Satdobato exhibit lower water quality based on the Nepal Drinking Water Quality Standard criteria (Bohara, 2016). On a older dataset we apply ML algorithms which might not be so, much relative to present water quality context of country still, it makes a huge effect in fulfilling a research gap underlying in water quality prediction and analysis using AI based techniques (Pant, 2011). The results presented in the paper can be used for useful future analysis of water quality of Nepal, using ML based approach and apply AI based techniques for environmental sustainability.

2. METHOD

2.1. Data Collection

The dataset used in this study was obtained from a publication that provided detailed water quality measurements from various wells across Nepal. The dataset included parameters such as Total Hardness, Arsenic Content, Iron Content, Nitrate, Ammonia, Total Coliform Count, pH, Temperature, Turbidity, Conductivity, and Chloride Content.

2.2. Data Preprocessing

The dataset was highly imbalanced, with certain water quality categories underrepresented and a high imbalance was seen on data. To address this, SMOTE was applied to balance the data. The data was then split into training and testing sets for model evaluation.

2.3. Feature Selection

Feature selection was performed to identify the most relevant parameters for WQI prediction and classification. This step involved statistical analysis and correlation studies to ensure that the selected features contribute significantly to the model's performance.

2.4. Regression Algorithms for WQI Prediction

Various regression algorithms were explored, including:

- Random Forest Regression
- Support Vector Regression (SVR)
- Decision Tree Regression
- Linear Regression
- Gradient Boosting Regression

2.5. Linear Regression

Linear Regression is a statistical technique employed to represent the relationship between a dependent variable, y, and one or more independent variables, X.

2.6. Decision Tree Regression

Decision Tree Regression uses a tree-like model of decisions and their possible consequences to predict a continuous target variable. The tree splits the dataset into subsets based on the feature values.

Equation:

A decision tree does not have a single equation but rather a set of if-then-else decision rules.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

2.7. Gradient Boosting Regression

Gradient Boosting Regression creates an ensemble of trees in a sequential manner, with each new tree aimed at correcting the mistakes of the previous ones. By adding trees that fit the negative gradient of the loss function, it works to minimize the loss.

2.8. Support Vector Regression (SVR)

Support Vector Regression (SVR), a variation of Support Vector Machine (SVM), is designed for regression tasks (Vörösmarty, et al., 2010; Wang & Nguyen, 2017). Its goal is to identify a function that deviates no more than ϵ from the actual target values for all training data, while remaining as flat as possible (Sun, et al., 2017; Ghahramani, 2015).

2.9. Classification Algorithms for Water Quality Classification

For the classification of water quality, the following algorithms were used:

- **Logistic Regression**

Logistic Regression is a statistical technique used to solve binary classification problems. It predicts the likelihood that a given input belongs to a specific class by utilizing the logistic function.

- Decision Tree Classifier

  A Decision Tree Classifier classifies data into various categories using a tree-like model of decisions and their potential outcomes. It divides the dataset into subsets based on feature values.

- Random Forest Classifier

  Random Forest Classifier is an ensemble method that generates multiple decision trees during the training phase and classifies by selecting the most frequent class across all trees.

- Gradient Boosting Classifier

  The Gradient Boosting Classifier constructs an ensemble of trees in sequence, where each new tree corrects the errors made by the previous ones by minimizing a loss function.

- Support Vector Machine (SVM)

  Support Vector Machine (SVM) is a supervised learning model applied to classification tasks. It identifies the hyperplane that best divides the classes in the feature space.

- Neural Networks (Multi-Layer Perceptron)

  The Multi-Layer Perceptron (MLP) is a type of feedforward artificial neural network (ANN) with multiple layers of nodes. Each node, or neuron, performs a mathematical function.

- LSTM-GRU hybrid model:
  The LSTM-GRU Hybrid Model merges the advantages of both the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures for processing and classifying sequential data.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

- MLP-GRU hybrid model:
  The MLP-GRU Hybrid Model is a sequential neural network that merges the advantages of recurrent layers with fully connected dense layers to enhance feature extraction and classification. It combines the layers of Multi-Layer Perceptron (MLP) and Gated Recurrent Unit (GRU).

Performance metrics such as accuracy, precision, recall, F1-score, and support were used to evaluate the classification models.

2.10. Evaluation metrices

A confusion matrix is a table utilized to evaluate the performance of a classification model. It presents a summary of the counts for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This matrix simplifies the understanding of the various types of errors made by the model (Rezaie-Balf, et al., 2020; Zhao, et al., 2021).

Table 1. Evaluation Metrics.

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

Accuracy:

Accuracy: The accuracy is determined by taking the percentage of all forecasts that were correct. Despite being a commonly used statistic for classification systems, misleading results may arise from an imbalanced dataset.

Accuracy $= \frac{TP+TN}{(FP+FN+TP+TP)}$…………..Equation 2.

Precision:

Positive predictive value is another name for precision, which is the ratio of real positive forecasts to all positive predictions. It displays how accurately the model forecasted the positive results.

Precision$=\frac{TP}{(FP+TP)}$ ……….Equation 8.

Sometimes called sensitivity or true positive rate, recall measures the proportion of real positive cases that the model correctly identified. It demonstrates how effectively the model can represent commendable cases.

Recall = $\frac{TP}{(FN+TP)}$……………….Equation 9.

F1-Score

Sometimes called sensitivity or true positive rate, recall measures the proportion of real positive cases that the model correctly identified. It demonstrates how effectively the model can represent commendable cases.

F1 Score$=2 \times \frac{Precision*Recall}{Precision+Recall}$ …………Equation 3.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

2.11. Water Quality evaluation features

Electrical conductivity (EC):

The electrical conductivity (EC) of water, which measures its ability to conduct electricity, is closely related to the concentration of dissolved ions within it. Elevated EC values suggest elevated concentrations of dissolved salts and minerals, perhaps signaling contamination or pollution (Mohammadpour, 2015; Tung & Yaseen 2020). High EC can have an impact on the suitability of local and well water for industrial, agricultural, and drinking purposes. Because many aquatic creatures are sensitive to variations in ion concentration, it may also have an effect on aquatic ecosystems. In order to determine the overall salinity of the water and make sure it satisfies the requirements for its intended usage, EC monitoring is crucial.

pH:

The pH scale ranges from 0 to 14, with 7 being neutral, and it measures the acidity or alkalinity of water. Water with a pH below 7 is classified as acidic, while water with a pH above 7 is deemed alkaline. The pH level of local and well water is significant because it affects chemical reactions within the water and the solubility of various pollutants. For instance, acidic or low pH water can make heavy metals more soluble, increasing their availability and potential toxicity to humans and aquatic life. On the other hand, precipitation of minerals caused by high pH (alkaline water) can block pipes and decrease the effectiveness of water distribution systems.

Dissolved Oxygen (DO):

The phrase "dissolved oxygen" (DO) refers to the amount of oxygen present in water, typically measured in milligrams per liter (mg/L). DO is a crucial indicator of aquatic ecosystem health, as oxygen is essential for the survival of most aquatic organisms, such as fish and invertebrates. When DO levels drop too low, hypoxic conditions can occur, which may stress or even lead to the death of aquatic life. Adequate DO levels are required in well and local water quality to inhibit the growth of anaerobic bacteria, which can result in toxic byproducts and unpleasant odors. Monitoring DO contribute to maintaining the health of water bodies and their ability to sustain a variety of thriving ecosystems.

2.12. Temperature

One important factor affecting water quality that affects both the chemical and biological processes in the water is temperature. The temperature of both local and well water affects the solubility of gases, such as oxygen, and the metabolic rates of aquatic organisms (Sharma, et al., 2021; Elbaz, et al., 2023). Higher temperatures usually cause an increase in an organism's metabolic rate, which might raise its oxygen demand, and a decrease in the solubility of oxygen, which results in lower DO levels. Furthermore, the growth of hazardous algal blooms and the toxicity of certain contaminants can also be impacted by temperature.

2.13. SMOTE

Machine learning practitioners frequently utilize Synthetic Minority Over-sampling Technique (SMOTE), to

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

solve the issue of imbalanced datasets. By creating artificial examples for the minority class, it successfully balances the distribution of classes without just copying already-existing data (Elbaz, et al., 2023). By interpolating between current minority class samples that are near to one another in the feature space, SMOTE generates new instances. To accomplish this, choose at random a point on the line segment that connects a sample of a minority class to one of its closest neighbors. By offering a more balanced training set, SMOTE serves to enhance the performance of classification algorithms by promoting better generalization and reducing bias towards the majority class.

Nested Cross validation

Nested cross-validation is a dependable method for assessing the performance of machine learning models, particularly when hyperparameter tuning is necessary. It consists of two layers of cross-validation: the outer loop and the inner loop. The outer loop splits the data into training and testing sets, while the inner loop performs the hyperparameter tuning on the training data (Du, et al., 2023). This tiered technique ensures that the hyperparameter tuning procedure does not affect the evaluation results, which aids in delivering an objective assessment of the model's performance.

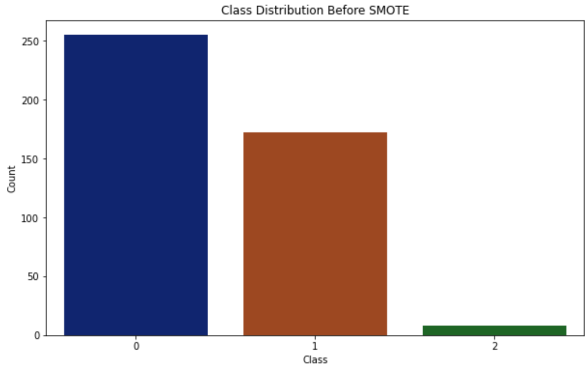

Fig. 2. Class distribution before SMOTE

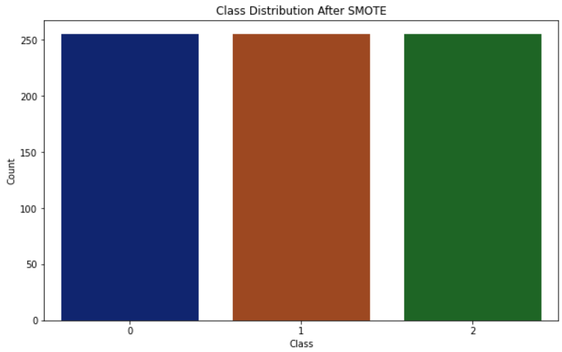

Fig. 3. Cass distribution after SMOTE

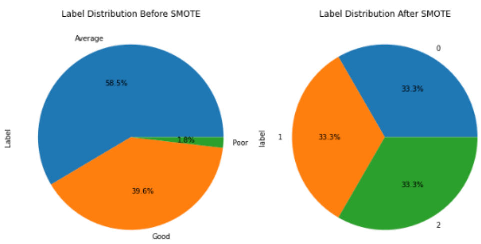

Fig. 4. Class volume before and after SMOTE

We can see the classes distribution from the above pie chart, after label encoding the corresponding class labels are represented in numerical representations. After SMOTE nearly 800 rows of data including 5 columns were formed for both regression and classification purpose. The both season data were included for training various models and on small dataset model performance were to be analyzed.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Table 2. Nested Cross Validation

Class	Label encoded representation
Average	0
Good	1
Poor	2

Water Quality Index

Water Quality Index (WQI) is a quantitative measure that reflects the overall quality of water based on various physical, chemical, and biological parameters. It is designed to simplify the representation of complex water quality data by aggregating multiple parameters into a single numerical value. The calculation of the Water Quality Index (WQI) typically involves parameters like temperature, pH, turbidity, dissolved oxygen, and the concentrations of various pollutants. Environmental agencies and water resource managers commonly use the resulting index to evaluate and compare the quality of various water sources, track pollution trends, and make informed decisions related to water treatment and environmental protection. A higher WQI indicates better water quality, while a lower WQI signals poorer quality, potentially posing risks to human health and aquatic life.

$$WQI = \frac{\sum Qi \cdot Wi}{\sum Wi} \text{........................Equation 4.}$$

SHAP analysis:

A machine learning technique called SHAP (SHapley Additive exPlanations) analysis offers comprehensible insights into model predictions. It uses ideas from game theory, particularly Shapley values, to calculate how much each feature contributes to a prediction. Each feature is given an importance value by SHAP, which indicates how much it influences the model's output for a given instance. Its ability to comprehend the behavior of the model, maintain transparency, and pinpoint which features influence predictions makes it an effective tool for debugging and model interpretation.

3. RESULTS AND DISCUSSION

3.1. Regression Results

The regression models' performance was assessed using various evaluation metrics. Among them, the Random Forest Regression model stood out with the highest accuracy, boasting an R^2 score of 0.92, which reflects a strong correlation between the predicted and actual WQI values for water quality classification.

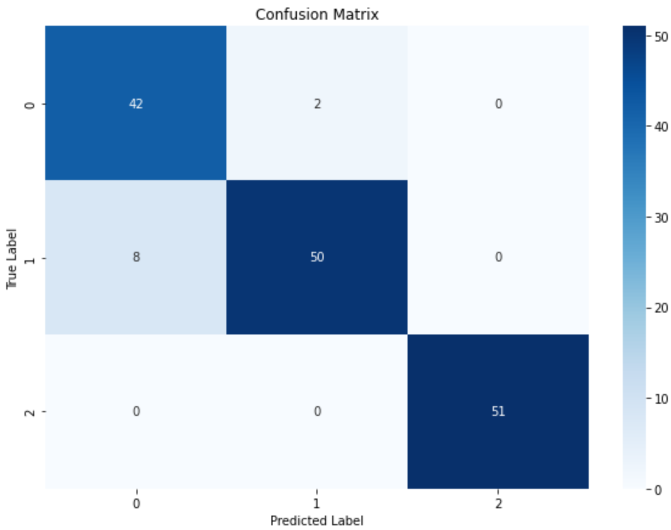

Fig 5. Confusion matrix for SVM

Overally, the confusion matrix for SVM shows a good performance on different classes, 42 samples our of 44 distributions were correctly predicted as class 0 and similarly, 50 out of 58 were

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

correctly predicted as class 1 and finally, 51 distributions were predicted correctly for class 2 with no incorrect predicted results on any of the classes.

Logistic Regression

Training Accuracy: 93.10%

Testing Accuracy: 93.46%

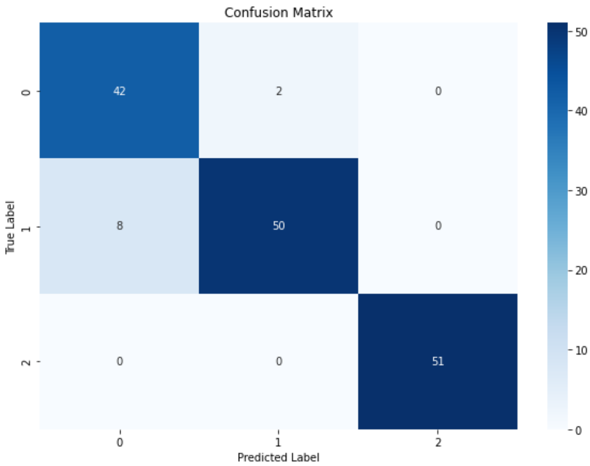

Fig. 6. LR classified confusion matrix plot

Similar results were seen for logistic regression in which, 42 out of 44 samples were predicted for class '0' and 50 out of 58 were predicted correctly for class '1' and finally, 51 were the totally correct predictions as seen in class '2' for our models development.

Decision Tree

- Training Accuracy: 100.00%
- Testing Accuracy: 98.69% plot

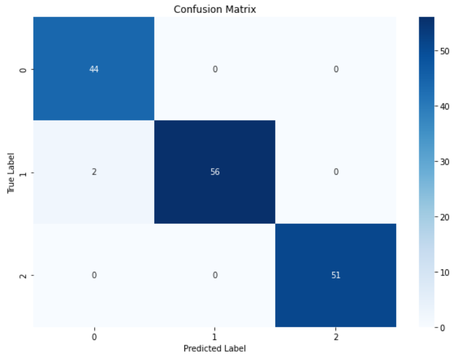

Fig. 7. Decision tree confusion matrix plot

In the decision tree algorithm the 44 predicted distributions among the whole were totally correct predicted for class '0' and similarly, 56 out of '58' samples were predicted correctly for class '1' and finally, 51 were correctly predicted on total 51 distributions on the dataset.

Catboost:

- Training Accuracy: 100.00%
- Testing Accuracy: 99.35%

Model confusion matrix results

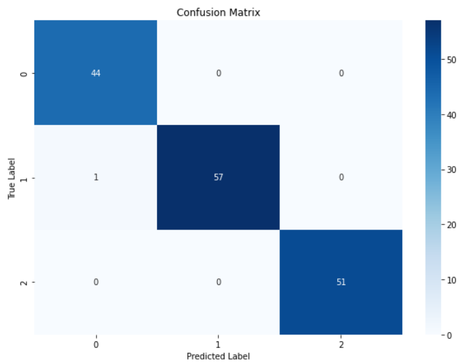

Fig. 8. Catboost confusion matrix plot

In the decision tree algorithm the 44 predicted distributions among the whole were totally correct predicted for class '0'

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

and similarly, 57 out of '58' samples were predicted correctly for class '1' and finally, 51 were correctly predicted on total 51 distributions on the dataset.

MLP (Multi-layer Perceptron)

- Training Accuracy: 98.03%
- Testing Accuracy: 96.73%

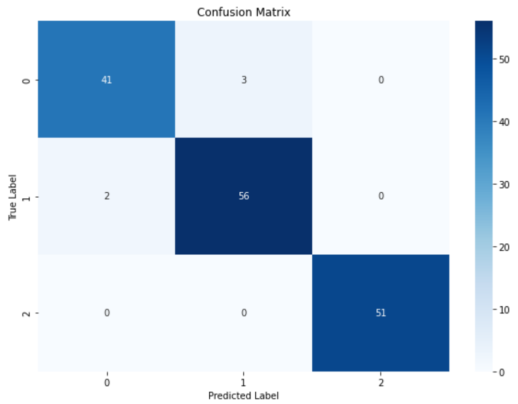

Fig. 9. MLP (ANN) confusion matrix plot

Trained for 50 epochs In the MLP deep learning results the '41' out of the 44 predicted distributions were totally correct predicted for class '0' and similarly, 56 out of '58' samples were predicted correctly for class '1' and finally, 51 were correctly predicted on total 51 distributions on the dataset.

MLP-GRU results

- Training Accuracy: 99.18%
- Testing Accuracy: 96.73%

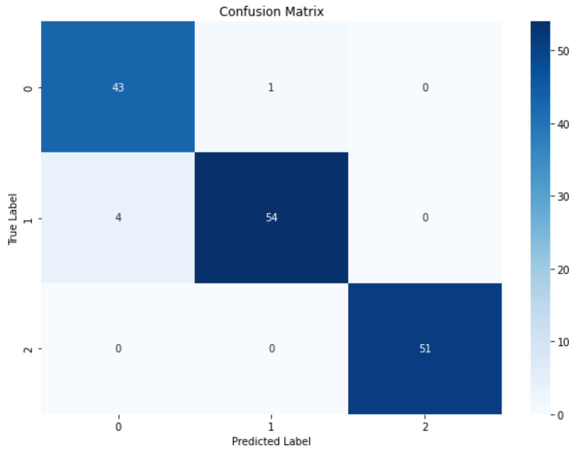

Fig. 10. MLP-GRU confusion matrix plot

A hybrid model MLP-GRU was developed and tested on the dataset indicating a very good performance as shown in above confusion matrix plot, 43 out of 44 were correctly precited similary, 54 out of 58 were correct predictions for class 1 and class 2 was predicted correctly for whole distributions in the testing set.

The models' performance in classification was assessed by calculating their accuracy scores on training and testing datasets. With training and testing accuracies of 93.10% and 93.46%, respectively, the Logistic Regression model demonstrated strong performance and a high degree of generalizability to new data. The Decision Tree model showed near-perfect accuracy (98.69%) on the testing set and perfect accuracy (100.00%) on the training set, indicating possible overfitting. With a testing accuracy of 99.35%, the CatBoost model surpassed the others, demonstrating its dependability and efficiency for this classification assignment.

High accuracy was also attained by the Multi-Layer Perceptron (MLP) model, which had testing and training accuracy of 96.73% and 98.03%, respectively. The

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

MLP-GRU hybrid model demonstrated competitive performance with training and testing accuracies of 96.06% and 96.73%, respectively, merging dense layers with GRU (Gated Recurrent Unit) layers.

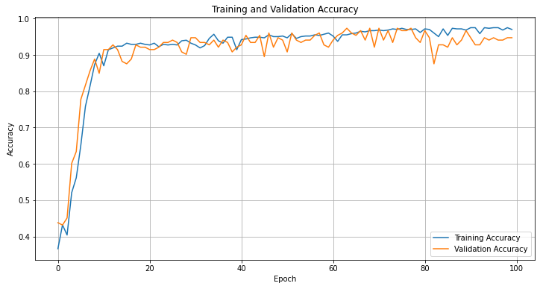

Fig. 11. MLP-GRU Model training history

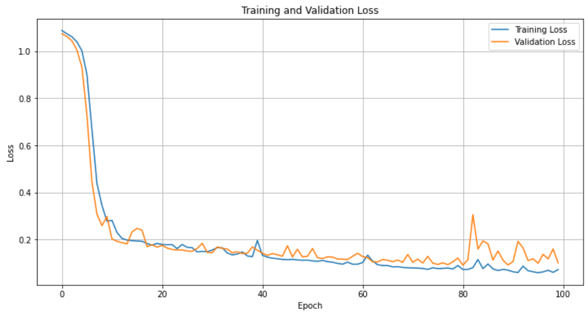

Fig. 12. MLP-GRU hybrid Model loss history

We can see that model training is increasing along with validation performance and on validation inclidated by orange curve on plot, the loss is decreasing linearly. The model was trained for 100 epochs and evaluated for performance. The result yields model great potentiality in predicting water quality and in real world applications development.

Table 3. Nested Cross validation results

Test Run	Accuracy (%)
1	96.1039
2	98.7013
3	94.7368
4	97.3684
5	97.3684
6	98.6842
7	97.3684
8	94.7368
9	96.0526
10	97.3684
Mean Accuracy	**96.8489**
Standard Deviation	**1.3413**

The table displays the outcomes of a model's evaluation through 10 separate test runs, utilizing nested cross-validation. 96.10%, 98.70%, 94.74%, 97.37%, 97.37%, 98.68%, 97.37%, 94.74%, 96.05%, and 97.37% are the accuracy values attained in each run. The model's average performance is shown by the 10 runs' mean accuracy of 96.85%. Furthermore, the accuracy scores' variability is demonstrated by the standard deviation of 1.34%, which indicates how consistently the model performs across various data splits. The model's performance appears to be fairly consistent, with minimal variance between test runs, as indicated by the low standard deviation, which supports the validity of the model's predictions.

Table 4. Accuracy Comparison for Classification Models

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Model	Training Accuracy (%)	Testing Accuracy (%)
Logistic Regression	93.10	93.46
Decision Tree	100.00	98.69
CatBoost	100.00	99.35
MLP	98.03	96.73
MLP-GRU	99.18	96.73

3.2. Regression Results

- Decision tree regressor:
- Training RMSE: 0.00
- Training R-squared: 1.00
- Testing RMSE: 1.95
- Testing R-squared: 0.99.........

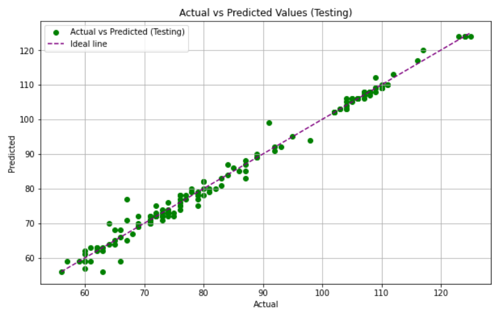

Fig. 13. Decision tree regressor actual vs predicted result

Linear regression

- Training RMSE: 3.23
- Training R-squared: 0.97
- Testing RMSE: 2.95
- Testing R-squared: 0.97

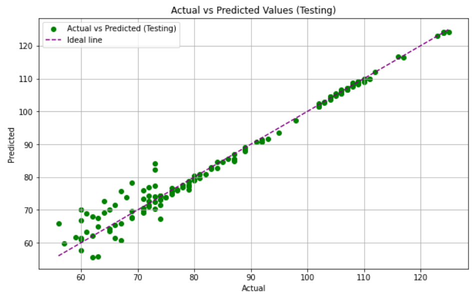

Fig. 14. Linear Regression (actual vs predicted plot)

LSTM-GRU regressor

- Training RMSE: 2.61
- Training R-squared: 0.98
- Testing RMSE: 2.43
- Testing R-squared: 0.98

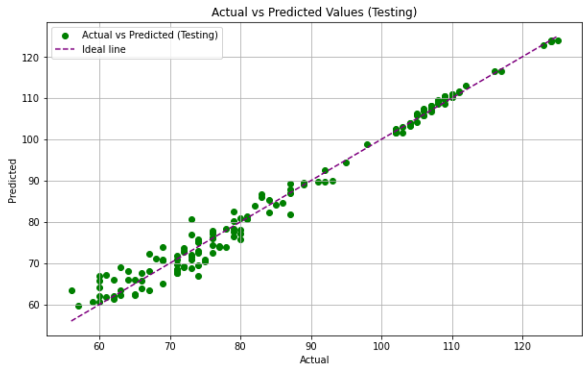

Fig. 15. LSTM-GRU (actual vs predicted result)

For this hybrid model above shown, is a scatter plot, for actual vs predicted results in LSTM-GRU model which is commonly used to evaluate a regression model's performance, is shown, showing the relationship between the actual and projected values from a testing dataset. The predicted values produced by the model are displayed on the y-axis, while the x-axis displays the actual values. A visual comparison of the model's performance is made possible by the green dots on the plot, each of which represents a pair of actual and anticipated values. The situation when the

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

anticipated values exactly match the actual values is represented by an ideal line, which is dashed purple in color (i.e., the perfect prediction line where predicted = actual).

The model performs better the closer the green dots are to this ideal line. The majority of the data points overlap along this line, demonstrating a significant connection between the actual and anticipated values, suggesting that the model predictions are fairly accurate. Deviations from this line would indicate model performance deviations or forecast errors.

Hyperparameters used:

The performance of regression models was assessed using metrics such as R2 (Coefficient of Determination) and RMSE (Root Mean Squared Error). The Decision Tree Regressor, exhibiting an RMSE of 0.00 and an R2 of 1.00 on the training set, showed a perfect fit. It also maintained strong performance on the testing set, achieving an RMSE of 1.95 and an R2 of 0.99.

This shows that generalization is still good but slightly overfitted. The Linear Regression model delivered excellent results, with training and testing R2 scores of 0.97 and RMSE values of 3.23 and 2.95, respectively, indicating a strong correlation between the predicted and actual values. Meanwhile, the LSTM-GRU hybrid regressor showed its ability to capture complex temporal dependencies in the data, achieving R2 scores of 0.98 and RMSE values of 2.61 and 2.43 for training and testing. These impressive results were achieved through both models.

Table 5. Different model hyperparameters used

Model Type	Hyperparameters
SVM	kernel='linear', C=0.001
MLP	hidden_layer_sizes=(100, 50), max_iter=500, random_state=42
Dense Neural Network	Dense(100, activation='relu'), Dense(50, activation='relu'), Dense(len(label_encoder.classes_), activation='softmax'), optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy']
Random Forest Regression	n_estimators=100, max_depth=10, min_samples_leaf=3, random_state=42
LSTM-GRU	LSTM(50, return_sequences=True), GRU(50, activation='relu'), optimizer='adam', loss='mse',

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Model Type	Hyperparameters
	EarlyStopping(monitor='val_loss', patience=5)
MLP-GRU	Dense(100, activation='relu'), Dense(50, activation='relu'), GRU(50, activation='relu'), Dense(len(label_encoder.classes_), activation='softmax'), optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy']
Other models	(No hyperparameters were found necessary)

The various hyperparameters can be observed as in table above, the hyperparameters applied on the model development was essential for building more accurate models for predictions.

Table 6. Regression Models Performance Comparison

Model	Training RMSE	Training R^2	Testing RMSE	Testing R^2
Decision Tree	0.00	1.00	1.95	0.99
Linear Regression	3.23	0.97	2.95	0.97
LSTM-GRU	2.61	0.98	2.43	0.98

3.3. Impact of SMOTE

The application of SMOTE significantly improved the performance of the classification models by addressing the class imbalance issue. The balanced dataset led to more accurate and reliable classification results.

3.4. Seasonal Variation Analysis

After a comprehensive review of paper (Bohara, 2016) various analysis can be made on water quality context of the capital city, Kathmandu, Nepal. As we analyzed the dataset created our review study looked at dissolved oxygen (DO), pH, temperature, electrical conductivity (EC), and other water quality characteristics in groundwater from wells that were excavated in the southern Kathmandu Valley. Below is a thorough analysis of the findings:

Temperature: The groundwater's temperature was discovered to be within a range that is thought to be typical for the area, suggesting that it is favorable for microbial development. The water's ambient temperature indicates that it is neither excessively hot or cold, which is ideal for preserving a balanced aquatic ecology and is normally suitable for residential usage.

pH

The groundwater's average pH was quite basic. On the other hand, the northwest of the research region and Hanumante Khola showed acidic pH levels. The corrosion of plumbing systems and pipelines caused by acidic water poses a threat to water safety and infrastructure upkeep. In general, the slightly basic pH in other locations is suitable for domestic use.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Total Quality and Adherence The majority of the measured values fell between the permitted ranges established by the World Health Organization (WHO) and Nepal's National Drinking Water Quality Standards (NDWQS). There was minimal fluctuation observed in the Water Quality Index (WQI) maps for all seasons, suggesting consistent physical water quality parameters all year round.

Recommendations

Although most of the study area's groundwater is adequate for domestic and agricultural usage, some regions need special treatment:

pH correction in Nakhu Khola EC control in some Kalanki Area wells DO enhancement in the majority of river sections upstream

3.5. SHAP Analysis results

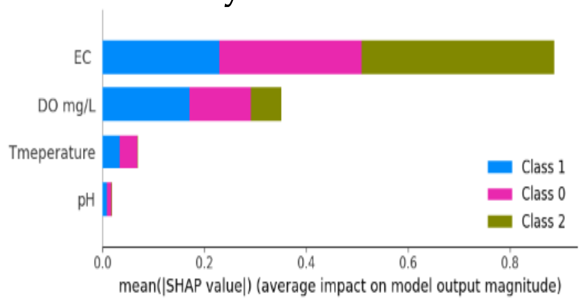

Fig. 16. SHAP analysis plot result

The SHAP (SHapley Additive exPlanations) analysis was performed on SVM model and the figure 15. plot provides insights into how different features contribute to the predictions made by a water quality model. In this case, the plot showcases the average impact of various water quality parameters on the model's output across three different classes (Class 2, Class 1, and Class 0). The bars represent the mean absolute SHAP values for each feature, highlighting their importance in the model's decision-making process.

EC: The model's most significant component, electrical conductivity (EC), is shown to be critical in establishing the quality categorization. It significantly affects the forecasts for all three classes, with Class 2 showing the most contribution (shown in green). This suggests that, especially for Class 2, EC has a significant role in differentiating the water quality among the classes. Furthermore, EC significantly affects Class 0 (pink) and Class 1 (blue), highlighting its significance in the model.

DO: Another significant component that mostly influences the forecasts for Classes 0 and 2 is the amount of dissolved oxygen (DO) mg/L. The feature makes a significant contribution, especially for Class 0, where it is essential to the categorization. Although it has less of an effect on Class 1, DO still has a big influence on the model's overall performance.

The model's predictions are comparatively less affected by temperature and pH. A little more than pH is contributed by temperature, which especially affects Class 0 and Class 1 predictions. But when compared to EC and DO, the effects of these traits are negligible. Specifically, pH has the least impact, suggesting that it is not a significant component in differentiating between the various water quality classes in our model.

This SHAP study shows that the two most important parameters for this model's prediction of water quality are Electrical Conductivity and Dissolved

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Oxygen. The least significant factors are pH and temperature, with pH having the least impact. These observations, which highlight the most significant characteristics, can direct additional research and model improvement.

3.6. Further Studies

In order to obtain a thorough comprehension of the water quality, more research concentrating on chemical and microbiological tests is suggested. This will guarantee that the water is safe for all planned uses and assist in identifying any potential contaminants not covered by the physical characteristics.

The study's overall findings show that although the groundwater quality is suitable for most uses, more research and focused treatments are required to guarantee that it satisfies all safety and quality requirements.

3.7. Future Work

The dataset should be enlarged in the future to include more comprehensive and varied water quality measurements from various locations and times of year. The predictive power of the model may be improved by adding other parameters, such as microbiological contaminants and developing pollutants. Furthermore, investigating deep learning models and sophisticated ensemble learning strategies may help to increase prediction robustness and accuracy. It may be possible to create machine learning algorithms that are connected with real-time water quality monitoring systems to deliver precise and quick water quality assessments. Lastly, incorporating stakeholders and local communities in the process of gathering data and developing models helps guarantee that the solutions are workable and customized to the unique requirements of the various areas.

4. CONCLUSION

This study concludes by showing the great potential of machine learning methods for categorizing and forecasting water quality metrics. High accuracy and efficiency were demonstrated by the models, especially CatBoost and LSTM-GRU hybrid, in completing the regression and classification tasks, respectively. Other models, such as Logistic Regression and MLP, demonstrated great generalizability by offering a fair balance between training and testing accuracy, whereas the Decision Tree model showed evidence of overfitting. The problem of data imbalance was successfully resolved by using the Synthetic Minority Over-sampling Technique (SMOTE), which enhanced the performance of the models even further. These results highlight the importance of using machine learning to regulate water quality, providing Nepal and comparable regions with a promising means of ensuring clean drinking water and reducing the spread of waterborne illnesses.

Declaration of conflict of interest:
The authors declare that thre is no conflict of interest

Author’s contribution statemtent
Bishwash Paneru: Investigation, Data collection, Data curation and validation, review and parametric analysis, final draft preparation
Biplov Paneru: Method and conceptualization, model development, data analysis, review and editing, final draft preparation

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Sanjog Chhetri Sapkota:
Validation, Review and analysis, final draft preparation

REFERENCES

Abbas, F., Cai, Z., Shoaib, M., Iqbal, J., Ismail, M., Alrefaei, A. F., & Albeshr, M. F. (2024). Machine learning models for water quality prediction: a comprehensive analysis and uncertainty assessment in Mirpurkhas, Sindh, Pakistan. *Water*, *16*(7), 941.

Anand, M. V., Sohitha, C., Saraswathi, G. N., & Lavanya, G. V. (2023, May). Water quality prediction using CNN. In *Journal of Physics: Conference Series* (Vol. 2484, No. 1, p. 012051). IOP Publishing.

Bohara, B. (2016). Water Quality Index of southern part of the Kathmandu Valley, Central Nepal; evaluation of physical water quality parameters of shallow wells. *Bulletin of the Department of Geology*, *19*, 45-56.

Du, J., Zhang, J., Castro-Lacouture, D., & Hu, Y. (2023). Lean manufacturing applications in prefabricated construction projects. *Automation in Construction*, *150*, 104790.

Elbaz, K., Shaban, W. M., Zhou, A., & Shen, S. L. (2023). Real time image-based air quality forecasts using a 3D-CNN approach with an attention mechanism. *Chemosphere*, *333*, 138867.

Elbaz, K., Zhou, A., & Shen, S. L. (2023). Deep reinforcement learning approach to optimize the driving performance of shield tunnelling machines. *Tunnelling and Underground Space Technology*, *136*, 105104.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452-459.

Kuroki, S., Ogata, R., & Sakamoto, M. (2023). Predicting the presence of E. coli in tap water using machine learning in Nepal. *Water and Environment Journal*, *37*(3), 402-411.

Mohammadpour, R., Shaharuddin, S., Chang, C. K., Zakaria, N. A., Ghani, A. A., & Chan, N. W. (2015). Prediction of water quality index in constructed wetlands using support vector machine. *Environmental Science and Pollution Research*, *22*, 6208-6219.

Najafzadeh, M., & Basirian, S. (2023). Evaluation of river water quality index using remote sensing and artificial intelligence models. *Remote Sensing*, *15*(9), 2359.

Pant, B. R. (2011). Ground water quality in the Kathmandu valley of Nepal. *Environmental monitoring and assessment*, *178*, 477-485.

Perumal, B., Rajarethinam, N., Velusamy, A. D., & Sundramurthy, V. P. (2023). Water Quality Prediction Based on Hybrid Deep Learning Algorithm. *Advances in Civil Engineering*, *2023*(1), 6644681.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584

Rahat, S. H., Steissberg, T., Chang, W., Chen, X., Mandavya, G., Tracy, J., ... & Ray, P. (2023). Remote sensing-enabled machine learning for river water quality modeling under multidimensional uncertainty. *Science of the Total Environment*, *898*, 165504.

Rezaie-Balf, M., Attar, N. F., Mohammadzadeh, A., Murti, M. A., Ahmed, A. N., Fai, C. M., ... & El-Shafie, A. (2020). Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: Comparative assessment of a noise suppression hybridization approach. *Journal of Cleaner Production*, *271*, 122576.

Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, *83*(12), 35307-35334.

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, *2*(1), 24-28.

Solanki, A., Agrawal, H., & Khare, K. (2015). Predictive analysis of water quality parameters using deep learning. *International Journal of Computer Applications*, *125*(9), 0975-8887.

Sun, Y., Chen, Z., Wu, G., Wu, Q., Zhang, F., Niu, Z., & Hu, H. Y. (2016). Characteristics of water quality of municipal wastewater treatment plants in China: implications for resources utilization and management. *Journal of Cleaner Production*, *131*, 1-9.

Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, *585*, 124670.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... & Davies, P. (2010). Global threats to human water security and river biodiversity. *nature*, *467*(7315), 555-561.

Wang, J., & Nguyen, A. V. (2017). A review on data and predictions of water dielectric spectra for calculations of van der Waals surface forces. *Advances in Colloid and Interface Science*, *250*, 54-63.

Zhao, S., Zhang, S., Liu, J., Wang, H., Zhu, J., Li, D., & Zhao, R. (2021). Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture*, *540*, 736724.

DOI: https://doi.org/10.34010/injiiscom.v6i2.13396
p-ISSN 2810-0670 e-ISSN 2775-5584