IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 March 2025, accepted 9 April 2025, date of publication 16 April 2025, date of current version 28 April 2025.
Digital Object Identifier 10.1109/ACCESS.2025.3561593

RESEARCH ARTICLE

Gated-LNN: Gated Liquid Neural Networks for Accurate Water Quality Index Prediction and Classification

SREENI CHADALAVADA[1], SULEYMAN YAMAN[2,3], ABDULKADIR SENGUR[4], ABDUL HAFEEZ-BAIG[5], RU-SAN TAN[6,7], PRABAL DATTA BARUA[5,8], RAVINESH C. DEO[3], (Senior Member, IEEE), MAKIKO KOBAYASHI[8,9], (Senior Member, IEEE), AND U. RAJENDRA ACHARYA[3]

[1]School of Engineering, University of Southern Queensland, Springfield, QLD 4300, Australia
[2]Biomedical Department, Vocational School of Technical Sciences, Fırat University, 23119 Elâzığ, Türkiye
[3]School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia
[4]Electrical-Electronics Engineering Department, Technology Faculty, Fırat University, 23119 Elâzığ, Türkiye
[5]School of Business, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[6]National Heart Centre Singapore, Singapore 169609
[7]Duke-NUS Medical School, Singapore 169857
[8]International Research Organization for Advanced Science and Technology (IROAST), Kumamoto University, Kumamoto 860-8555, Japan
[9]Graduate School of Science and Technology, Kumamoto University, Kumamoto 860-8555, Japan

Corresponding author: Abdulkadir Sengur (ksengur@firat.edu.tr)

ABSTRACT Surface water quality is of utmost significance to ensure public health and facilitate sustainable economic development. Traditional water quality assessment methods are typically time-consuming and labor-intensive and require numerous field measurements and laboratory analyses, which are costly and impractical to implement in large-scale water quality monitoring. Recent advances in machine learning (ML) have brought new approaches to predicting water quality index (WQI) and classifying water quality in real time to enhance decision-making in environmental management. In this study, we propose a novel gated liquid neural network (gated-LNN) that can predict WQI and classify water quality with high accuracy. As opposed to typical ML models, the proposed gated-LNN includes a gating mechanism that enhances temporal learning and noise robustness, making it well-suited for dynamic environmental data. For ascertaining the effectiveness of the proposed approach, we conducted rigorous experiments on a publicly available water quality dataset with 1897 examples collected from varied water bodies of India between the years 2005 and 2014. The dataset comprises seven most significant parameters of water quality, i.e., dissolved oxygen, pH, conductivity, biological oxygen demand, nitrate, fecal coliform, and total coliform. The proposed gated-LNN model achieved a high R^2 of 0.9995 for WQI prediction and 99.74% accuracy for three-class water quality classification into "Good," "Poor," and "Unsuitable" classes, outperforming state-of-the-art models in both regression and classification tasks. While these results highlight the model's potential as a highly accurate and efficient tool for real-time water quality assessment, its generalizability to different regions remains an important consideration. Future work will focus on enhancing computational efficiency and conducting generalization tests on datasets from diverse geographic regions and time periods to evaluate adaptability.

INDEX TERMS Gate mechanism, liquid neural networks, machine learning, water quality prediction.

I. INTRODUCTION

Surface waters in river systems constitute the primary source of freshwater on earth and are vital to human existence on the planet [1]. Water quality is determined by

The associate editor coordinating the review of this manuscript and approving it for publication was Yougan Chen.

natural processes, such as precipitation and erosion, and anthropogenic activities that pollute the aquatic ecosystem, including agriculture, urbanization and industrialization [2], [3]. Degradation of surface water quality compromises access to potable water, posing serious challenges to public health and economic development [4], [5]. The water quality index (WQI), which conglomerates several physical and chemical variables, is widely used to grade the quality of ground and surface waters in water resource management and environmental protection [6], [7], [8].

Artificial intelligence (AI) can be harnessed to model water quality in environmental engineering [9], [10], [11], [12]. Hameed et al. [13] applied neural networks to a dataset that contained monthly information on water quality in Malaysia. The radial basis function neural network yielded the best performance, with the highest coefficient of determination (R^2) of 0.9872 for WQI prediction. Asadollah et al. [14] used a novel ensemble of machine learning (ML) methods—the extra-tree regression ensemble model—to predict WQI on a dataset that collected monthly data in Hong Kong. Among various experiments, the highest R2 value of 0.98 was reported. Majnooni et al. [15] applied four deep learning (DL) models—gated residual variable selection model (GRVS), cross-deep model, wide-deep model, and deep base model—on a WQI dataset that contained ten parameters. The best model, GRVS, attained the highest 0.98 R2 value. Talukdar et al. [16] optimized gradient boosting (GB), random forest (RF), and deep neural network (DNN) with an entropy-based weighting approach and applied them to a water quality dataset sourced from Loktak Lake. After extensive experiments, the best model, RF, yielded the highest R2 value of 0.90 for others. Wang et al. [17] proposed a long short-term memory (LSTM) network-based ML approach for WQI monitoring of a semi-arid river. After applying Apriori algorithm-based optimization for tuning training parameters, the LSTM-based model attained 89.07% accuracy for WQI prediction. Shams et al. [18] used various ML approaches for WQI prediction, including RF and several boosting models. A grid search mechanism was used to tune the parameters of the boosting models. The GB attained the best accuracy of 99.5% for WQI prediction.

Khan et al. [19] used principal component analysis-based regression and gradient boosting for WQI prediction and water quality classification. On a water quality dataset sourced from the Gulshan Lake, the model attained an excellent 95% accuracy for WQI prediction and 100% accuracy for water quality classification. Bi et al. [20] proposed a hybrid model that contained a LSTM-based encoder-decoder network. Using the Savitzky-Golay filter to eliminate noise in the dataset, the model attained 0.679 root mean square error (RMSE) for WQI prediction. Saeed et al. [21] proposed an LSTM-based deep-learning approach for WQI forecasting. They applied data pre-processing and aggregation on a West Australian estuary dataset, followed by data transformation, which converted the pre-processed data for time-series forecasting. An excellent 0.957 R2 value for WQI prediction was reported. Tian et al. [22] monitored the water quality of reservoirs supplying city and agricultural water using a Fuzzy C-means clustering algorithm and the band combination model to analyze remotely sensed water quality variables. Using ML models like mixed density network (MDN), XGBoost, DNN, and support vector regression (SVR) to predict optically active and non-optically active components, the best R^2 score of 0.98 was attained by the MDN model.

While the reviewed studies have made notable progress in ML-based water quality prediction, they come with several limitations [23], [24]. Many approaches, such as neural networks and ensemble methods, require significant computational power and fine-tuning of hyperparameters, making them less practical for real-time applications. Some studies rely on datasets with low temporal resolution, like monthly measurements, which may fail to capture short-term fluctuations in water quality. DL models, despite their strong predictive capabilities, often require large amounts of training data and are prone to overfitting, limiting their ability to generalize to new environments. Similarly, tree-based models like random forests and gradient boosting, while effective, can be difficult to interpret and struggle with noisy or incomplete data. Additionally, while LSTM and hybrid models enhance temporal learning, they are computationally intensive and do not always outperform simpler models in real-world applications.

In this work, we have built a ML model for WQI prediction using a more advanced liquid neural network (LNN) [25], which is a type of artificial neural network used to model time-varying dynamic systems and uncertainties. Compared with traditional neural networks, LNN possesses the advantages of adaptability over time, flexibility, and computational efficiency. Moreover, we have incorporated a gating mechanism to create a novel gated LNN designed for the efficient processing of information over time. Gated-LNN can learn long-term dependencies more effectively by controlling the forgetting of past information and is more resilient to noisy data [26]. We trained and tested the gated-LNN model on a public water quality dataset. In our experiments, gated-LNN demonstrated excellent performance for predicting WQI and classifying water quality labels.

The main contributions of this paper are as follows:

- To the best of our knowledge, this work is the first to propose and implement gated-LNN for water quality prediction and classification tasks.
- The incorporated gating mechanism dynamically adjusts the contributions of new inputs, maintains relevant past states, and controls the outputs, thereby enhancing the model's ability to learn long-term dependencies and its resilience against noisy data.
- On a public water quality dataset with seven key water quality attributes, which had been sourced from various water bodies in India from 2005 to 2014, the model attained

excellent performance, lending support for its application in environmental engineering and water resource management.

II. MATERIALS AND METHODS

A. PROPOSED GATED LIQUID NEURAL NETWORK

The proposed gated-LNN model (Fig. 1) systematically processes water quality attributes and dynamically captures their relationships. First, water quality attributes in the study dataset (see Section D. DATASET) are processed through a two-staged sequential gating mechanism comprising a gate sigmoid layer followed by a gate hyperbolic tangent layer. The former normalizes feature importance to a range between 0 and 1 and dynamically selects and weights the most relevant information for further processing; the latter applies the non-linear transformation to the former's output and scales the relative importance of features to a range between -1 and 1, capturing both positive and negative feature contributions. The information the two gate layers have processed is then passed to the liquid layer, which (1) captures complex, non-linear, and time-varying relationships between features through recurrent connections and (2) processes and preserves temporal dependencies. These operations ensure effective modeling of dynamic and interdependent water quality parameters, allowing the gated-LNN to understand how changes in one parameter affect others. Finally, the processed features are fed to a fully connected layer for WQI prediction and water quality classification.

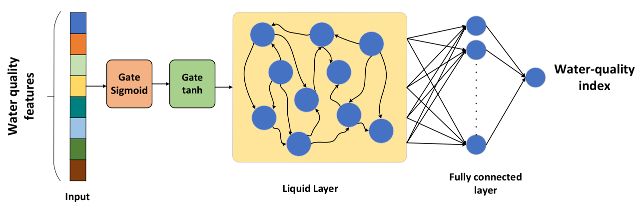


FIGURE 1. Proposed gated-LNN model for water quality index prediction.

B. LIQUID NEURAL NETWORKS

An LNN applies a differential equation to describe how a system changes over time; its primary components are time-varying activation functions and a combination of such functions defined according to the given differential equation [25], [27]. The state of the network is thus a continuous time-dependent function, where the weighted input function at a given time, (t), can be mathematically expressed as:

$$u(t) = W.x(t) \tag{1}$$

where $u(t)$ represents the input vector of the network, W, the weight matrix; and $x(t)$, the current state of the network, i.e., the present state of the dynamic system. A differential equation governs the network state at any time changes and this change,

$$\frac{\partial x(t)}{\partial t} = f\left(x(t), u(t), \theta\right) \tag{2}$$

where $x(t)$ represents the instantaneous state of the network; $f(.)$, the function that determines the nature of the dynamics of the system (and is usually considered to be nonlinear); and θ, parameters of the model. The output function is obtained by a composition of the state of the network and the weighted input function, as given below:

$$y(t) = \sigma(W_y.x(t)) \tag{3}$$

where $y(t)$ represents the network output; W_y, the weight matrix used for the output; and σ, the activation function. The LNN learns by backpropagation, which minimizes the loss between outputs of the network and target values, as given below:

$$L = \sum_{i=1}^{N} \left(y_i(t) - \hat{y}_i(t)\right)^2 \tag{4}$$

where L represents the loss function of the network; $y_i(t)$, actual output, and $\hat{y}_i(t)$, predicted outputs. The loss function is used to optimize the parameters of the gated-LNN model.

C. GATED-LNN

LNNs represent a powerful approach to learning the evolution of dynamic systems as structures modeled by differential equations [25], [27]. These structures can be enhanced with gate mechanisms [28], which control and optimize the dynamics in a much more flexible way. Over time, gate mechanisms can either activate parts of the network or limit their influence, rendering the model more controlled and sensitive to data. In a typical LNN, network dynamics are governed by a differential equation, where the state of the network, $x(t)$, is modeled as a function of time. With the addition of gate mechanisms, which control how active the network is over certain time intervals, the evolution of the network becomes more controlled. The state of the LNN is given as:

$$\frac{\partial x(t)}{\partial t} = f\left(x(t), u(t), \theta\right).g(t) \tag{5}$$

where $g(t)$ represents the gate function.

The time-varying gate function is defined mathematically as:

$$g(t) = \sigma\left(W_s.x(t) + b_s\right).tanh\left(W_t.x(t) + b_t\right) \tag{6}$$

where σ is the sigmoid activation function, which regulates information flow, while $tanh(.)$ is the hyperbolic tangent activation function, which scales and transforms input features non-linearly. Here, W_s and W_t represent the weight matrices for the sigmoid gate and the tanh activation function, respectively, while b_s and b_t are their corresponding bias terms. The sigmoid function limits the value of $g(t)$ to between 0 and 1: when $g(t) \approx 0$, the state of the network remains unchanged; when $g(t) \approx 1$, the system makes a full evolution. This gating mechanism introduces flexibility to the learning as network activity can vary dynamically over a certain time interval. As the value of the function $g(t)$ changes over time, the dynamics of the network adapt to these changes, i.e., the gate function controls how the network evolves. The output of a LNN is thus directly related to the state of the network and the gate function and can be expressed as:

$$y(t) = \sigma(W_y.x(t).g(t)) \tag{7}$$

where $y(t)$ represents the output of the network, W_y is the output weight matrix, and $x(t)$ is the transformed and gated network state. The gate function decides the amount of activity of the output. This flexibility allows the network to undergo a more efficient learning process by processing only important information at certain time intervals. In other words, the gate function allows the network to focus on more important data when needed. The effect of the gate mechanism on the learning process is observed by tuning the loss function. As learning of the network is usually done through a backpropagation algorithm, the loss function is deployed here to minimize differences between network predictions and actual values.

The loss function can be represented as:

$$L = \sum_{i=1}^{N} \left(y_i(t) - \hat{y}_i(t)\right)^2 \tag{8}$$

where L represents the loss function; $y_i(t)$, actual output; $\hat{y}_i(t)$, model's predicted output. Thus, the network can operate more efficiently by learning only at important moments.

The gating mechanism used in this design is similar to a GRU-type approach [29], utilizing a sigmoid function to control information flow and a *tanh* activation to control input transformation. However, unlike GRU, which employs separate update and reset gates to manage information flow across time steps, the proposed model uses a single gating function that modulates the entire information transfer process. This single gate architecture reduces computational complexity while still capturing temporal dependencies effectively. The fully connected layers are used as intermediate processing steps so that the network can learn appropriate gating values dynamically. This process is achieved in the liquid neural network by modulating the input before it reaches the liquid dynamics layer, with information transfer controlled and increasing the model's temporal dependency ability.

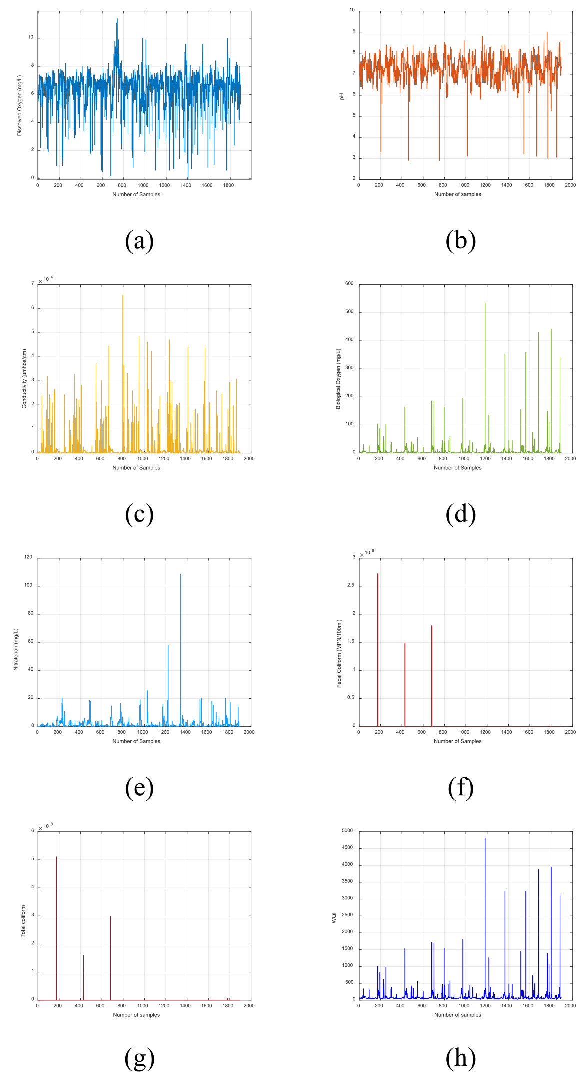


FIGURE 2. Water quality parameters and water quality indices (WQI) of water samples in the study dataset (n=1897): (a) dissolved oxygen, (b) pH, (c) conductivity, (d) biological oxygen, (e) nitrate, (f) fecal coliform, (g) total coliform, and (h) WQI.

D. DATASET

We used the public water quality dataset [30], which is sourced from various lakes and rivers in India from 2005 to 2014. It contains a total of 1897 cases, which have been characterized using seven quantitative features of water quality: (1) "dissolved oxygen", the amount of oxygen dissolved in the water, which is necessary to sustain aquatic life; (2) "pH", the level of acidity or alkalinity of the water; (3) "conductivity", the ability of water to conduct electric current, which informs on the presence and level of dissolved particles; (4) "biological oxygen", the amount of dissolved oxygen that microorganisms in the water absorb, which reflects the degree of organic pollution; (5) "nitrate", the number of nitrate ions in the water, which is an indicator of sewage or fertilizer contamination; (6) "fecal coliform", the number of coliform bacteria of fecal origin in the water, which is an indicator of fecal contamination; and (7) "total coliform", the total amount of fecal and non-fecal coliform bacteria from both in the water (Fig. 2).

As seen in Fig. 2, the dissolved oxygen ranges from 0.0 to 11.4, indicating instances of complete oxygen depletion, which may signify severe pollution or stagnation alongside well-oxygenated conditions. The pH values exhibit an extremely wide range, from 0.0 to 9,01, suggesting the presence of potential data anomalies or measurement errors. Conductivity spans from 0.4 to 65,700, reflecting diverse water salinity or ion concentration levels potentially influenced by pollution or natural mineral content. Biological oxygen demand varies between 0.1 and 534.5, highlighting significant differences in organic matter content and microbial activity. Nitrate levels, ranging from 0.0 to 108.7, suggest areas with negligible to extremely high nutrient concentrations, possibly due to agricultural runoff or industrial waste. Fecal coliform and total coliform exhibit wide ranges, from 0.0 to 27,252 and 0.0 to 51,109, respectively, indicating variations in microbial contamination, with potential health risks at higher levels. Finally, the WQI ranges from 19.3 to 99.8, reflecting water quality from poor to excellent conditions.

Data preprocessing was carried out, including mean imputation for missing values and z-score normalization using standardization to scale the data. This was done to prepare the data for subsequent processing and ensure it is suitable for further analysis.

E. CALCULATION OF WQI

WQI is a well-established metric used to assess water quality. It is calculated as [18]:

$$WQI = \frac{\sum_{i=1}^{N} q_i . w_i}{\sum_{i=1}^{N} w_i} \tag{9}$$

where N denotes the total number of parameters; and q_i and w_i represents the quality rating and the corresponding unit weight of the parameter i, respectively. q_i is calculated as:

$$q_i = 100\left(\frac{v_i - v_{id}}{s_i - v_{id}}\right) \tag{10}$$

where v_i corresponds to the observed value for the parameter i, v_{id} represents the ideal value under pure water conditions, and s_i is the standard threshold for the same parameter. w_i is calculated as:

$$w_i = \frac{k}{s_i} \tag{11}$$

where k is a proportionality constant given by:

$$k = \frac{1}{\sum_{i=1}^{N} s_i} \tag{12}$$

Categorical labels of water quality are determined based on values of WQI: a value between 0 to 50 indicates "Good"; 51 to 100, "Poor"; and 101 and above, "Unsuitable" [18].

III. EXPERIMENTS AND RESULTS

The gated-LNN model was implemented in MATLAB on a computer with an NVIDIA GeForce RTX 3090 GPU and 24 GB RAM. Model performance was evaluated using a hold-out validation approach, where 75% of the dataset was allocated for training and 25% for testing. Performance metrics calculated on the test set include RMSE, mean absolute percentage error (MAPE), mean absolute error (MAE), and R^2. RMSE measures the average magnitude of the errors between the predicted and actual values and is an indicator of model accuracy [31]: low RMSE implies that model predictions closely approximate actual data. Mathematically, the RMSE is computed as [31]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} \tag{13}$$

where N represents the total number of data points, y_i, actual data values, and $\hat{y}_i$, predicted data values. The MAE measures the average absolute difference between actual and predicted values, while the MAPE expresses this error as a percentage of the actual values [32], [33]:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left|y_i - \hat{y}_i\right| \tag{14}$$

$$MAPE = \frac{100}{N} \sum_{i=1}^{N} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{15}$$

where y_i represents the actual values, and $\hat{y}_i$ denotes the predicted values.

R^2, which ranges from 0 to 1, expresses what proportion of the original data variance is covered by the predicted data [34]. An R^2 value near 1 indicates a good fit between predicted and actual data. It is calculated using the following formula [35]:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N} \left(y_i - \bar{y}_i\right)^2} \tag{16}$$

where y_i represents the actual data points, $\hat{y}_i$ represents the predicted values, and $\bar{y}_i$ is the average of the actual data. A higher R^2 value indicates that the underlying structure and trend of the data have been better preserved after the modeling or simplification process.

To overcome the challenges posed by the additional parameters introduced by the gating mechanisms, we employed several strategies. We employed regularization techniques such as L2 weight decay and dropout to avoid overfitting via reduction of reliance on specific neurons and enhanced generalization. In addition, we controlled the capacity of the network via careful selections for the number of hidden units in the gating layers in a way that attempted to balance expressiveness and complexity. To facilitate effective training despite the increased parameter space, we employed the Adam optimizer. Furthermore, the liquid neural network's dynamic temporal dependencies with fewer trainable parameters than traditional recurrent architectures alleviated the added complexity introduced by the gating mechanism. Parameters for the gated-LNN model were determined heuristically in experiments. The model has a feature input layer configured to accept an input dimension without normalization. The gating mechanism consists of two fully connected layers, having 1024 neurons, followed by sigmoid activation and hyperbolic tangent activation functions, respectively, for nonlinearity and gating control. A custom liquid layer block contains another fully connected layer with 1024 neurons, followed by a fully connected hidden layer with 1024 neurons. The final output layer is a single fully connected layer for regression tasks, followed by a regression layer. The training was performed using the Adam optimizer with a maximum of 600 epochs, an initial learning rate of 0.0001, and a mini-batch size 256. Data shuffling at every epoch was applied, and training progress was monitored via live plots with validation data included for performance assessment.

The constant learning rate of 0.0001, combined with the Adam optimizer, ensured stable and gradual updates for effective optimization. The training progress for WQI prediction converged well (Fig. 3). RMSE and loss curves for both training and validation decreased steeply in the initial iterations and stabilized to near-minimal values as training progressed. The final achieved validation RMSE is 0.035737, indicating high accuracy. The smoothed training and validation curves suggest consistency with neither major overfitting nor divergence, which implies the generalizability of well-to-unseen data. The model attained RMSE, MAE, and MAPE of 0.0357, 0.0242, and 0.5376, respectively, which indicate minimal deviations between predicted and actual values, and R^2 of 0.9995, underscoring the model's ability to explain nearly all the variability in the data. These results lend credence to its robustness and efficiency in predicting WQI. Not only is the model able to predict the general trend of WQI values, but it is also adept at capturing local fluctuations in WQI effectively. In an analysis of the test set, the predicted WQI values closely followed the actual values across all samples, demonstrating the model's generalization ability. Even in regions with sharp peaks and variations, the deviations remained minimal, indicating the robustness of Gated-LNN (Fig. 4).

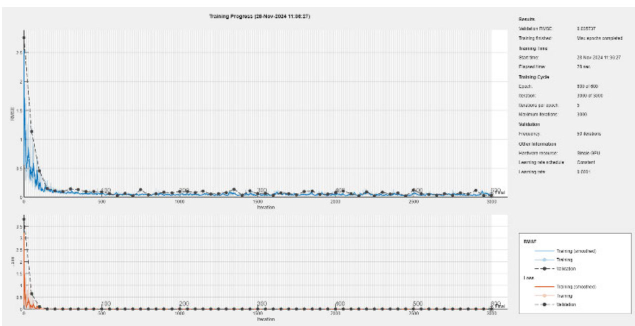


FIGURE 3. Plots of training progress for WQI prediction.

TABLE 1. Summary of evaluation metrics obtained for WQI prediction.

RMSE	MAE	MAPE	R^2
0.0357	0.0242	0.5376	0.9995

The quantitative evaluation metrics obtained for WQI prediction, presented in Table 1, highlight the high accuracy and reliability of the proposed model. The RMSE of 0.0357 and MAE of 0.0242 indicate minimal deviations between predicted and actual values. The MAPE of 53.76% demonstrates the model's exceptional prediction accuracy relative to the true values. Besides, R^2 of 0.9995 further underpins the model for explaining nearly all the variability in data, giving credence to its robustness and efficiency in predicting WQI.

Fig. 4 also compares the actual and predicted WQI values, which reflect the good predictive capability of the proposed

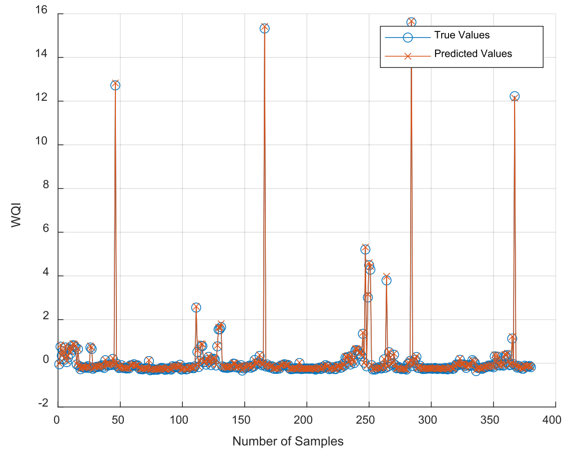
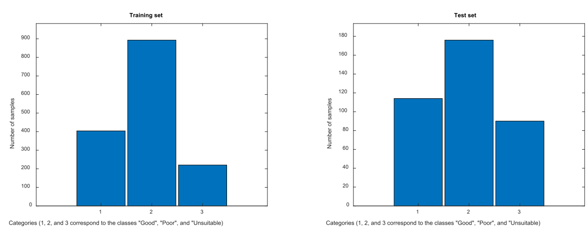


FIGURE 4. Comparison of predicted versus actual WQI on the test set.

model. The predicted values follow the actual values closely over the entire range of samples, with minimal deviations even in regions characterized by sharp peaks and variations. This is indicative of the model's ability to capture both the general trend and local fluctuations in WQI effectively.

The proposed gated-LNN model was also used in the classification of water quality. The performance evaluation of this classification task was carried out with accuracy, recall, precision, specificity, and F1-score metrics [36] to evaluate the model performance of classification of water quality into "Good", "Poor", and "Unsuitable" classes. In the training phase of the classification task, we employed an Adam optimizer with a maximum of 600 epochs, an initial learning rate of 0.0001, and a mini-batch size of 256. Data shuffling was applied at every epoch, and training progress was monitored via live plots with validation data included for performance assessment.



(a) (b)

FIGURE 5. The distribution of the class categories in training (a) and test (b) sets.

The class distributions of the training and test sets are shown in Fig. 5. As observed, the "Poor" class has the highest number of samples in both the training and test sets, while the "Good" and "Unsuitable" classes have relatively fewer samples. However, this situation did not lead to an overfitting toward the majority class. This situation can be seen in Fig. 6 and Table 2. If there were a class imbalance, the confusion matrix would show more misclassifications for underrepresented classes, with most predictions biased

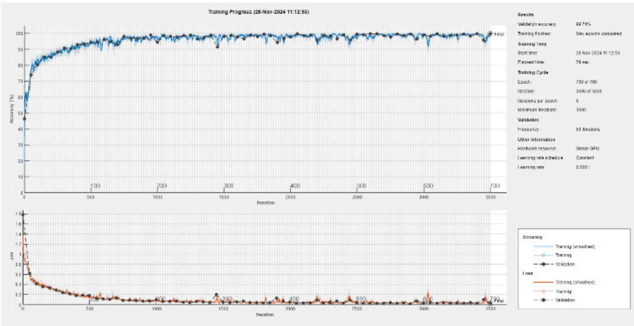


FIGURE 6. Plots of training progress for water quality classification.

TABLE 2. Summary of evaluation metrics (%) obtained for automated water quality classification using the proposed gated-LNN model.

Accuracy	Precision	Recall	Specificity	F1-score
99.74	99.81	99.71	99.84	99.76

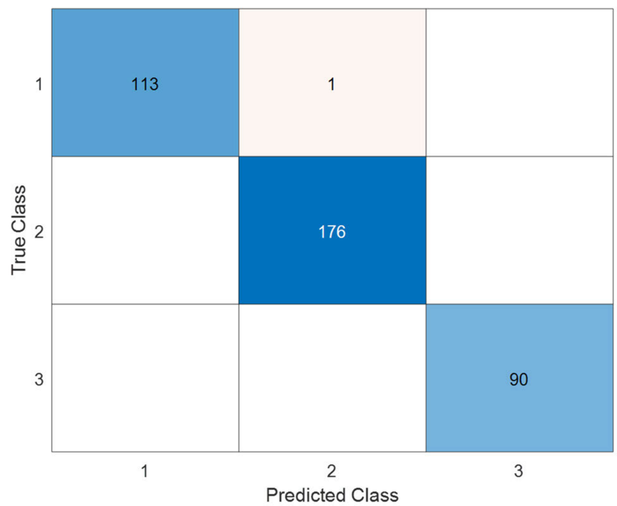


FIGURE 7. Confusion matrix obtained for water quality classification. ∗1, 2, and 3 correspond to the classes "Good", "Poor", and "Unsuitable, respectively.

toward the majority class. This would lead to lower recall for underrepresented classes and lower precision for the majority class. The F1-score would also decrease, indicating poor precision-recall balance. However, the confusion matrix shows high precision, recall, and F1-scores for all classes, indicating no significant class imbalance and that the model handles all categories well.

Training accuracy increased rapidly from about 50% to over 96% in the first 200 iterations, stabilizing at 96.7% at the end of training (Fig. 5). Validation accuracy demonstrated a similar trend, converging to about 96%. In contrast, training loss declined sharply from an initial value of about 1.2 to below 0.1 within the first 300 iterations, stabilizing at about 0.05 toward the last iterations. Validation loss demonstrated a similar trend, reaching a convergence value of approximately 0.07. The confusion matrix shows only one instance of misclassification (Fig. 6). Moreover, the model attained excellent accuracy, precision, recall, specificity, and F1-score of 99.74%, 99.81%, 99.71%, 99.84%, and 99.76%, respectively, indicating that the proposed gated-LNN model is accurate, discriminative, and sensitive for water quality high classification.

It can be observed from Fig. 6 that the proposed Gated-LNN model classified Class 1 instances as 113, Class 2 instances as 176, and Class 3 instances as 90, with only one misclassification for Class 1. These results show that the proposed model effectively discriminates between classes with high overall accuracy and maintains a minimum misclassification rate across all classes.

Table 2 lists the performance measures of water quality classification using the proposed Gated-LNN. The model yielded 99.74% in terms of accuracy, indicating the capability of the model to classify almost all instances in the dataset correctly. Besides, its precision of 99.81% indicates the high capability of the model to reduce false positives, while the recall of 99.71% reflects its efficiency in capturing the truest positives. A specificity of 99.84% indicates excellent performance in the identification of negative cases, which is critical for balanced classification. The F1-score is 99.76%, indicating a strong balance between precision and recall.

To explore the contribution of each feature to the classification of the water quality, we incorporated Shapley additive explanations [37]. The feature importance plot, as given in Fig. 7, ranks the input variables according to their mean absolute SHAP values, which represent their overall contribution to the model's predictions. Biochemical oxygen demand appears to have the largest effect, followed by dissolved oxygen and conductivity. The larger the SHAP value, the larger the influence of the feature on classification decisions, and the more it contributes towards identifying the drivers of the model's outputs.

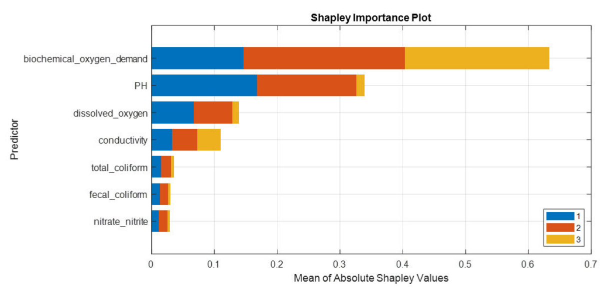


FIGURE 8. The feature importance plot.

The SHAP summary plot, as given in Fig. 8, for 'Good' provides a zoomed-in view of how feature values for specific instances add up to make predictions. Each instance in the data is depicted by a point, with colors indicating whether the feature value is low (blue) or high (pink). Horizontal spread of

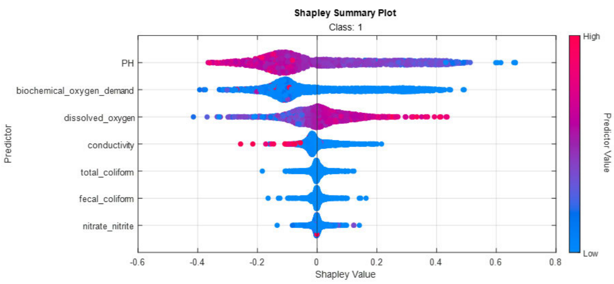


FIGURE 9. The SHAP summary plot for "Good" class.

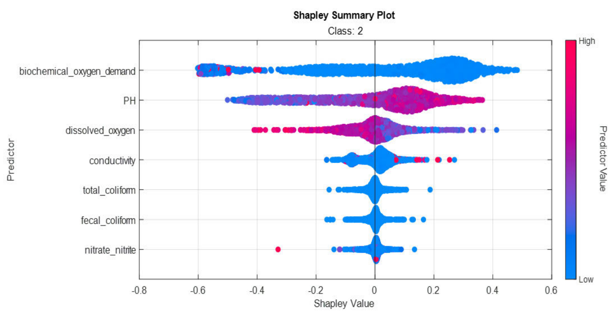


FIGURE 10. The SHAP summary plot for "Good" class.

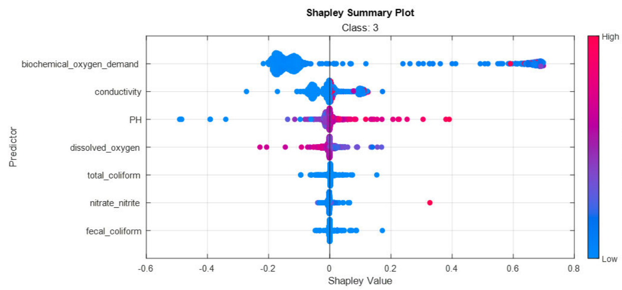


FIGURE 11. The SHAP summary plot for "Unsuitable" class.

TABLE 3. The performance of the proposed gated-LNN model on various levels of noise.

SNR	Accuracy	Precision	Recall	Specificity	F1-score
60 dB	0.9816	0.9853	0.9806	0.9890	0.9826
50 dB	0.9579	0.9722	0.9509	0.9739	0.9597
40 dB	0.9500	0.9612	0.9475	0.9703	0.9524
30 dB	0.8816	0.9026	0.8774	0.9313	0.8842

the points shows the range of the SHAP values, illustrating the degree to which variations in biochemical oxygen demand, dissolved oxygen, and other features are accountable for the classification as 'Good'.

The SHAP summary plot (Fig. 9) for 'Poor' shows the same analysis, how feature values contribute to predictions for this class. The SHAP value distribution is handy for viewing the model's sensitivity to different variables, and it identifies which features are pushing classification decisions. The features with a bigger range of SHAP values show a larger degree of influence on model predictions for 'Poor'.

The SHAP summary plot for 'Unsuitable' given in Fig. 10 illustrates the impact of each feature on predictions for this class. The color-coded points are the original feature values, and the SHAP values indicate their contribution to the classification outcome. A noticeable trend in the distribution of points suggests that certain features tend to always contribute heavily towards the prediction of this class, reinforcing the necessity to examine feature interactions in the model.

We also tested the proposed gated-LNN's performance on noisy data. To this end, various Signal-to-Noise Ratio (SNR) level artificial noise was produced and added to the input data. The previous training parameter setting was saved for this experiment too. Table 3 shows the obtained classification evaluation metrics.

Table 3 illustrates the performance of the proposed Gated-LNN model under different levels of noise. As SNR decreases, the model's performance declines steadily across all metrics. At 60 dB, the model performs optimally with an accuracy of 0.9816, in addition to high precision, recall, and specificity, indicating good performance under low noise conditions. However, at 30 dB, accuracy drops to 0.8816, and precision, recall, and F1-score also decrease, demonstrating that raising the noise level negatively impacts the model's ability to predict correctly.

Finally, we compared the performance and execution time of various ML methods with the proposed gated-LNN model. To this end, we have selected random forest (RF), support vector machines (SVM), Bayesian learning (BL), and k-nearest neighbors (kNN) algorithms, as they are popularly used in the literature [38], [39]. Table 4 shows the accuracy values and running times of each method. As observed from Table 4, the proposed method is the slowest to run, while the other ML algorithms are much faster to produce results. However, these ML models' accuracy scores are far lower relative to the performance of the proposed gated-LNN model. The superior performance of Gated-LNN stems from its gating mechanism, which adaptively controls information flow through recurrent units. Unlike standard LNNs, it selectively retains relevant information while filtering out noise, enhancing long-term dependency learning and preventing state saturation. This improves robustness to noise and allows the model to dynamically adjust to new inputs, leading to better regression and classification performance.

IV. DISCUSSION

Water quality determination using ML offers a rapid and inexpensive alternative to traditional testing, which can be used for real-time monitoring and early pollution detection. These methods can process complex datasets to predict WQI and water quality categories accurately to support informed decisions on sustainable water management while ensuring public health safety. In this work, a gated-LNN has

TABLE 4. Performance and running time comparisons of the proposed method with some of the ML methods.

Method	Accuracy (%)	Running Time (sec)
RF	95.25	0.18
SVM	91.03	0.02
kNN	65.17	0.004
BL	62.80	0.002
Gated-LNN	99.74	31.10

been constructed by adding general gating mechanisms to an otherwise LNN model to enhance both time dynamics and feature representations. These gates are configured to dynamically adjust the contributions of new inputs, maintain relevant past states, and regulate the outputs at each time step for improved handling of temporal context. The proposed gated-LNN yielded excellent performance for WQI prediction and water quality label classification, attaining salutary 0.0357 RMSE, 0.0242 MAE, 0.5376 MAPE and 0.9995 R^2 for the former task; and excellent 99.74% accuracy, 99.81% precision, 99.71% recall, 99.84% specificity and 99.76% F1-score for the latter task.

The gated-LNN model compares favorably against and outperforms recent published methods for WQI prediction and water quality classification (Table 5). Shams et al. [18] studied several ML approaches to the tasks, employing RF, XGBoost, GB, and AdaBoost for classification, and k-nearest neighbors (KNN), decision tree (DT), SVR, and multilevel perceptron (MLP) for regression. They performed preprocessing that involved mean imputation and normalization for the dataset, which contained 1897 instances and seven water quality features. GB attained 99.50% accuracy for water quality classification and MLP, R^2 of 0.9980 for WQI prediction. Nasir et al. [40] used standard DT for water quality classification. Guo et al. [41] studied various ML approaches for water quality classification and concluded that standard support vector machines outperformed other ML approaches. Ahmed et al. [42] used a standard MLP approach for water quality classification.

In Table 5, a detailed comparison of the proposed gated-LNN method with several state-of-the-art models for WQI prediction and water quality classification is given. For the classification task, the proposed gated-LNN achieved an accuracy as high as 99.75%, while the results from other methods are 99.50% by GB, 99.30% by XGBoost, and 99.00% by RF from Shams et al. [18]. Other classification models, such as DT from Nasir et al. [40] SVM from Guo et al. [41] yielded considerably less 81.62% and 88.75%, respectively. This comparison illustrates the high performance of the proposed gated-LNN in the classification task, significantly improving the state-of-the-art approach.

The regression performance of the proposed gated-LNN was also outstanding, achieving R^2 of 0.9997, outperforming all other models in comparison. Specifically, the MLP regressor from Talukdar et al. [16] is 0.9980, and the SVR is 0.9910, far lower than the gated-LNN result. The outstanding R^2 reached by the proposed method shows its strong capability in mining underlying patterns in the regression task and further confirms its superiority in performance.

TABLE 5. Summary of comparison with the state-of-the-art methods developed on the WQI prediction and water quality classification.

Ref.	Method	Task	Acc (%)	R^2
[18]	Grid search-based RF	C	99.00	-
[18]	Grid search-based XGBoost	C	99.30	-
[18]	Grid search-based AdaBoost	C	99.10	-
[18]	Grid search-based GB	C	99.50	-
[18]	Grid search-based KNN regressor	R	-	98.20
[18]	Grid search-based DT regressor	R	-	99.00
[18]	Grid search-based SVR	R	-	99.10
[18]	Grid search-based MLP regressor	R	-	99.80
[40]	DT	C	81.62	-
[41]	SVM	C	88.75	-
[42]	MLP	C	85.07	-
TS	Gated-LNN	R	-	99.95
TS	Gated-LNN	C	99.75	-

[a] C=classification, R=regression, TS=this study.

TABLE 6. Summary of comparison with the standard LNN method on the WQI prediction and water quality classification.

METHOD	RMSE	MAE	MAPE	R2
Gated-LNN	0.0357	0.0242	0.5376	0.9995
LNN	0.0867	0.0721	0.6112	0.9969

METHOD	Accuracy	Precision	Recall	Specificity	F1-score
Gated-LNN	99.74	99.81	99.71	99.84	99.76
LNN	98.68	99.08	99.18	98.67	98.67

We carried out another comparison where the proposed gated-LNN was compared with the traditional LNN. For a fair comparison, all gated LNN and LNN parameters were set to identical values. Table 6 shows the evaluation metrics for this comparison. As seen in Table 6, the gated-LNN significantly outperforms LNN across multiple evaluation metrics. Specifically, gated-LNN achieves lower RMSE, MAE, and MAPE values, demonstrating improved predictive accuracy, while also achieving a higher R^2 value of 0.9995, indicating a near-perfect fit. Additionally, in classification performance, Gated-LNN exhibits superior accuracy, precision, recall, specificity, and F1-score, with an accuracy of 99.74% compared to 98.68% for LNN.

We also conducted a further performance analysis using a Taylor diagram [43] to visually assess the predictive capability of the proposed Gated-LNN in comparison to other traditional methods for WQI prediction. The Taylor diagram, presented in Fig. 12, provides a comprehensive evaluation

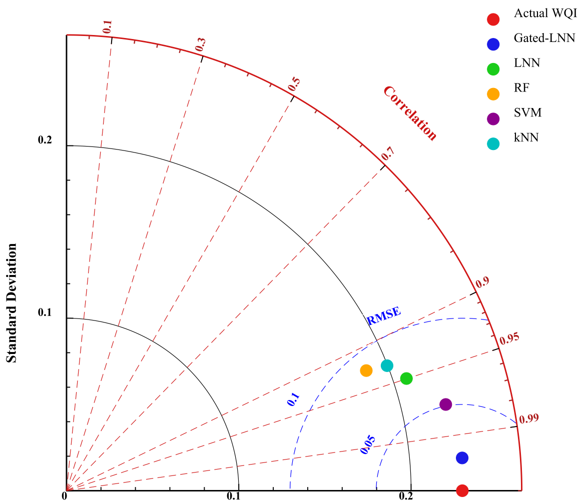


FIGURE 12. Comparison of the proposed Gated-LNN in the Taylor diagram [43] with other methods for WQI prediction.

by illustrating the relationship between model standard deviation, RMSE, and correlation coefficient in a single plot. As shown in Fig. 12, the gated-LNN exhibits the best predictive performance, achieving the highest correlation and the lowest RMSE, indicating a near-perfect fit to the actual WQI values. SVM also demonstrates strong performance, with a similarly high correlation and low RMSE, making it a competitive alternative to Gated-LNN. RF and kNN models perform moderately well, with slightly lower correlation values and higher RMSE, suggesting that while they can still provide reliable predictions, they are not as accurate as gated-LNN and SVM. Furthermore, the Taylor diagram highlights the substantial improvement achieved by the proposed Gated-LNN over the traditional LNN model for WQI prediction. The diagram clearly shows that Gated-LNN exhibits a significantly higher correlation coefficient and a lower RMSE compared to LNN, indicating that the proposed modifications have successfully enhanced the model's predictive accuracy. This improvement suggests that the gated mechanism effectively optimizes the learning process, allowing the model to better capture variations in WQI values.

The advantages and disadvantages of our proposed gated-LNN are listed below.

Advantages:

•The proposed gated-LNN model yielded excellent results in WQI prediction and water quality classification, with R^2 of 0.9995 and 99.74%, outperforming state-of-the-art methods.

•This model incorporates different gating mechanisms, which improve feature representation via dynamic adjustment of input contributions and maintaining relevant past states. These operations enhance the model's ability to learn long-range dependencies and (because of that) resilience to data noise.

•Trained and tested on a large dataset comprising multi-site data collected over many years, excellent model performance supports its application for water quality assessment in different contexts and scenarios.

•The model allows real-time assessment of water quality and can inform decision-making on the protection of public health and sustainable management of water resources.

Disadvantages:

•Gated-LNNs are computationally expensive in terms of their architecture. Their training requires powerful GPUs and, very often, extended times.

•Most DL models, like gated-LNN, often lack transparency. Users may find it hard to understand the decision-making process.

•The performance heavily relies on the quality and quantity of the input data; errors in data pre-processing or anomalies may adversely impact outcomes.

•While the model handles temporal dynamics well, it may struggle with datasets containing extreme heterogeneity or unstructured attributes.

•As the proposed model is customized for certain water quality datasets, further modifications will be needed for it to be applied in other environmental contexts.

V. CONCLUSION

In this study, a new gated liquid neural network (Gated-LNN) model was proposed for the prediction of water quality index (WQI) and classification of water quality. The proposed model demonstrated high accuracy and robustness, yielding R^2 value of 0.9995 for the prediction of WQI and a 99.74% three-class classification accuracy, outperforming state-of-the-art modern methods. The gating mechanism in gated-LNN enhances the model's ability to handle temporal dynamics, adapt to noisy datasets, and learn long-term dependencies, rendering it a powerful tool for water quality monitoring. The excellent results attest to the potential of gated-LNN to contribute to real-time water quality assessment for sustainable water resource management.

Future work should prioritize optimizing the computational efficiency of the gated-LNN model to ensure its scalability for large-scale datasets and real-time deployment. Additionally, incorporating explainable AI techniques will enhance model transparency, fostering trust and adoption among policymakers and environmental managers. To broaden its applicability, future research should explore integrating multi-source data, including remote sensing and IoT-based sensor networks, into the model framework. Such advancements would enable gated-LNN to be applied beyond water quality assessment to real-time monitoring and prediction in other environmental domains, such as air quality and soil contamination. These improvements will strengthen the model's role in environmental conservation and sustainable resource management. Moreover, evaluating the model's performance on datasets from different geographic regions and time periods remains an important direction for future research. Conducting such generalization tests will provide deeper insights into the adaptability and robustness of the proposed approach. Lastly, we will explore alternative gating mechanisms, such as fully learnable or attention-based gates, to evaluate their impact on the performance of the Gated-LNN when applied to more complex datasets.

REFERENCES

[1] R. C. Grabowski and A. M. Gurnell, "Hydrogeomorphology–ecology interactions in river systems," *River Res. Appl.*, vol. 32, no. 2, pp. 139–141, Feb. 2016, doi: 10.1002/rra.2974.

[2] N. B. Bhatti, A. A. Siyal, A. L. Qureshi, and I. A. Bhatti, "Socio-economic impact assessment of small dams based on T-paired sample test using SPSS software," *Civil Eng. J.*, vol. 5, no. 1, p. 153, Jan. 2019, doi: 10.28991/cej-2019-03091233.

[3] C. J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies, "Global threats to human water security and river biodiversity," *Nature*, vol. 467, no. 7315, pp. 555–561, Sep. 2010, doi: 10.1038/nature09440.

[4] H. Cheng, Y. Hu, and J. Zhao, "Meeting China's water shortage crisis: Current practices and challenges," *Environ. Sci. Technol.*, vol. 43, no. 2, pp. 240–244, Jan. 2009, doi: 10.1021/es801934a.

[5] M. G. Uddin, S. Nash, and A. I. Olbert, "A review of water quality index models and their use for assessing surface water quality," *Ecol. Indicators*, vol. 122, Mar. 2021, Art. no. 107218, doi: 10.1016/j.ecolind.2020.107218.

[6] A. A. Bordalo, R. Teixeira, and W. J. Wiebe, "A water quality index applied to an international shared river basin: The case of the douro river," *Environ. Manage.*, vol. 38, no. 6, pp. 910–920, Dec. 2006, doi: 10.1007/s00267-004-0037-6.

[7] E. Sánchez, M. F. Colmenarejo, J. Vicente, A. Rubio, M. G. García, L. Travieso, and R. Borja, "Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution," *Ecol. Indicators*, vol. 7, no. 2, pp. 315–328, Apr. 2007, doi: 10.1016/j.ecolind.2006.02.005.

[8] P. S. González, A. Stehr, and R. O. Barra, "Assessment of water quality trends through the application of an aggregated water quality index with historical monitored data in a Mediterranean Andean basin," *Ecol. Indicators*, vol. 166, Sep. 2024, Art. no. 112373, doi: 10.1016/j.ecolind.2024.112373.

[9] M. Khadr and M. Elshemy, "Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt," *Ain Shams Eng. J.*, vol. 8, no. 4, pp. 549–557, Dec. 2017, doi: 10.1016/j.asej.2016.08.004.

[10] Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," *J. Hydrol.*, vol. 585, Jun. 2020, Art. no. 124670, doi: 10.1016/j.jhydrol.2020.124670.

[11] H. Yoshioka and Y. Yoshioka, "Assessing fluctuations of long-memory environmental variables based on the robustified dynamic orlicz risk," *Chaos, Solitons Fractals*, vol. 180, Mar. 2024, Art. no. 114336, doi: 10.1016/j.chaos.2023.114336.

[12] H. Li, B. Zhou, X. Xu, R. Huo, T. Zhou, X. Dong, C. Ye, T. Li, L. Xie, and W. Pang, "The insightful water quality analysis and predictive model establishment via machine learning in dual-source drinking water distribution system," *Environ. Res.*, vol. 250, Jun. 2024, Art. no. 118474, doi: 10.1016/j.envres.2024.118474.

[13] M. Hameed, S. S. Sharqi, Z. M. Yaseen, H. A. Afan, A. Hussain, and A. Elshafie, "Application of artificial intelligence (AI) techniques in water quality index prediction: A case study in tropical region, Malaysia," *Neural Comput. Appl.*, vol. 28, no. S1, pp. 893–905, Dec. 2017, doi: 10.1007/s00521-016-2404-7.

[14] S. B. H. S. Asadollah, A. Sharafati, D. Motta, and Z. M. Yaseen, "River water quality index prediction and uncertainty analysis: A comparative study of machine learning models," *J. Environ. Chem. Eng.*, vol. 9, no. 1, Feb. 2021, Art. no. 104599, doi: 10.1016/j.jece.2020.104599.

[15] S. Majnooni, M. Fooladi, M. R. Nikoo, G. Al-Rawas, A. T. Haghighi, R. Nazari, M. Al-Wardy, and A. H. Gandomi, "Smarter water quality monitoring in reservoirs using interpretable deep learning models and feature importance analysis," *J. Water Process Eng.*, vol. 60, Apr. 2024, Art. no. 105187, doi: 10.1016/j.jwpe.2024.105187.

[16] S. Talukdar, Shahfahad, S. Bera, M. W. Naikoo, G. V. Ramana, S. Mallik, P. A. Kumar, and A. Rahman, "Optimisation and interpretation of machine and deep learning models for improved water quality management in lake loktak," *J. Environ. Manage.*, vol. 351, Feb. 2024, Art. no. 119866, doi: 10.1016/j.jenvman.2023.119866.

[17] J. Wang, B. Xue, Y. Wang, Y. A. G. Wang, and D. Han, "Identification of pollution source and prediction of water quality based on deep learning techniques," *J. Contaminant Hydrol.*, vol. 261, Feb. 2024, Art. no. 104287, doi: 10.1016/j.jconhyd.2023.104287.

[18] M. Y. Shams, A. M. Elshewey, E.-S.-M. El-kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, "Water quality prediction using machine learning models based on grid search method," *Multimedia Tools Appl.*, vol. 83, no. 12, pp. 35307–35334, Sep. 2023, doi: 10.1007/s11042-023-16737-4.

[19] M. S. Islam Khan, N. Islam, J. Uddin, S. Islam, and M. K. Nasir, "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 4773–4781, Sep. 2022, doi: 10.1016/j.jksuci.2021.06.003.

[20] J. Bi, Y. Lin, Q. Dong, H. Yuan, and M. Zhou, "Large-scale water quality prediction with integrated deep neural network," *Inf. Sci.*, vol. 571, pp. 191–205, Sep. 2021, doi: 10.1016/j.ins.2021.04.057.

[21] A. Saeed, A. Alsini, and D. Amin, "Water quality multivariate forecasting using deep learning in a west Australian estuary," *Environ. Model. Softw.*, vol. 171, Jan. 2024, Art. no. 105884, doi: 10.1016/j.envsoft.2023.105884.

[22] D. Tian, X. Zhao, L. Gao, Z. Liang, Z. Yang, P. Zhang, Q. Wu, K. Ren, R. Li, C. Yang, S. Li, M. Wang, Z. He, Z. Zhang, and J. Chen, "Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, south China," *Environ. Pollut.*, vol. 342, Feb. 2024, Art. no. 123104, doi: 10.1016/j.envpol.2023.123104.

[23] R. B. Singh, A. I. Olbert, A. Samantra, and M. G. Uddin, "AI-driven modelling approaches for predicting oxygen levels in aquatic environments," *J. Water Process Eng.*, vol. 66, Sep. 2024, Art. no. 105940, doi: 10.1016/j.jwpe.2024.105940.

[24] H. I. Kim, D. Kim, M. Mahdian, M. M. Salamattalab, S. M. Bateni, and R. Noori, "Incorporation of water quality index models with machine learning-based techniques for real-time assessment of aquatic ecosystems," *Environ. Pollut.*, vol. 355, Aug. 2024, Art. no. 124242, doi: 10.1016/j.envpol.2024.124242.

[25] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Liquid time-constant networks," in *Proc. 35th AAAI Conf. Artif. Intell.*, vol. 35, May 2021, pp. 7657–7666, doi: 10.1609/aaai.v35i9.16936.

[26] L. Jing, C. Gulcehre, J. Peurifoy, Y. Shen, M. Tegmark, M. Soljacic, and Y. Bengio, "Gated orthogonal recurrent units: On learning to forget," *Neural Comput.*, vol. 31, no. 4, pp. 765–783, Apr. 2019, doi: 10.1162/neco_a_01174.

[27] M. Chahine, R. Hasani, P. Kao, A. Ray, R. Shubert, M. Lechner, A. Amini, and D. Rus, "Robust flight navigation out of distribution with liquid neural networks," *Sci. Robot.*, vol. 8, no. 77, Apr. 2023, pp. 1–15, doi: 10.1126/scirobotics.adc8892.

[28] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600, doi: 10.1109/MWSCAS.2017.8053243.

[29] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[30] *Indian Water Quality Data*. Accessed: Mar. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/anbarivan/indian-water-quality-data

[31] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Feb. 2014.

[32] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[33] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.

[34] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016, doi: 10.1016/j.ijforecast.2015.12.003.

[35] R. E. Carlson and T. A. Foley, "The parameter R2 in multiquadric interpolation," *Comput. Math. Appl.*, vol. 21, no. 9, pp. 29–42, 1991, doi: 10.1016/0898-1221(91)90123-l.

[36] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Dec. 2020, pp. 1–4, doi: 10.1186/s12864-019-6413-7.

[37] S. Mangalathu, S.-H. Hwang, and J.-S. Jeon, "Failure mode and effects analysis of RC members based on machine-learning-based Shapley additive exPlanations (SHAP) approach," *Eng. Struct.*, vol. 219, Sep. 2020, Art. no. 110927, doi: 10.1016/j.engstruct.2020.110927.

[38] R. B. Singh, K. C. Patra, and A. Samantra, "GHPSO-ATLSTM: A novel attention-based genetic LSTM to predict water quality indicators," *Stochastic Environ. Res. Risk Assessment*, vol. 2024, pp. 1–16, Mar. 2024, doi: 10.1007/s00477-024-02694-3.

[39] R. B. Singh, K. C. Patra, B. Pradhan, and A. Samantra, "HDTO-DeepAR: A novel hybrid approach to forecast surface water quality indicators," *J. Environ. Manage.*, vol. 352, Feb. 2024, Art. no. 120091, doi: 10.1016/j.jenvman.2024.120091.

[40] N. Nasir, A. Kansal, O. Alshaltone, F. Barneih, M. Sameer, A. Shanableh, and A. Al-Shamma'a, "Water quality classification using machine learning algorithms," *J. Water Process Eng.*, vol. 48, Aug. 2022, Art. no. 102920, doi: 10.1016/j.jwpe.2022.102920.

[41] H. Guo, Z. Chen, and F. Y. Teo, "Intelligent water quality prediction system with a hybrid CNN–LSTM model," *Water Pract. Technol.*, vol. 19, no. 11, pp. 4538–4555, Nov. 2024, doi: 10.2166/wpt.2024.282.

[42] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, "Efficient water quality prediction using supervised machine learning," *Water*, vol. 11, no. 11, p. 2210, Oct. 2019, doi: 10.3390/w11112210.

[43] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res., Atmos.*, vol. 106, no. D7, pp. 7183–7192, Apr. 2001, doi: 10.1029/2000jd900719.

SREENI CHADALAVADA received the Ph.D. degree from the University of South Australia, in 2009, specializing in contaminant hydrogeology for optimal water resources management. He is currently a Senior Lecturer (Water Engineering) with the University of Southern Queensland. Prior to this, he was a Principal Environmental Scientist and the Research Manager of the Cooperative Research Centre for Contamination Assessment and Remediation of Environment (CRC CARE), a national center of excellence for contaminant assessment and environmental remediation headquartered with The University of Newcastle, Australia. In his current role, he is actively involved in teaching and research in the fields of hydrology, water resources, and environmental sciences. Over the past 12 years of his academic and research career with the University of South Australia and The University of Newcastle, Australia, he has played a significant role in conducting cutting-edge research in the field of water resources and environmental engineering.

SULEYMAN YAMAN received the B.Sc. and M.Sc. degrees in biomedical engineering from Afyon Kocatepe University, Türkiye, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronics Engineering, Fırat University, Türkiye. He is a Lecturer with the Department of Biomedical Engineering, Fırat University, and a Visiting Ph.D. Researcher with the University of Southern Queensland, Australia. His research interests include signal processing, image processing, machine learning, deep learning, and embedded systems.

ABDULKADIR SENGUR received the B.Sc. degree in electronics and computers education, the M.Sc. degree in electronics education, and the Ph.D. degree in electrical and electronics engineering from Fırat University, Türkiye, in 1999, 2003, and 2006, respectively. He became a Research Assistant with the Technical Education Faculty, Fırat University, in February 2001. He is currently a Professor with the Technology Faculty, Fırat University. His research interests include signal processing, image segmentation, pattern recognition, medical image processing, and computer vision.

ABDUL HAFEEZ-BAIG received the master's degree in MIS and MBA from Griffith University, Brisbane, Australia, and the Ph.D. degree in the domain of healthcare and information systems from USQ. Since joining USQ, in January 2004, he has published several refereed publications and has numerous research grants in the domains of health and education regarding technology management. He is very conversant with wireless technology as well as emerging technologies and learning management systems. He teaches information systems concepts to both undergraduate and postgraduate students, including MBA students. He also has numerous publications in academic and scholarly journals and has a vast array of scholarly conference papers, all of which have been focused on the domain of information systems. He has extensive experience in the area of information systems, especially related to the healthcare sector. He is quite interested in wireless healthcare applications, systems analysis and design, adoption, the infusion and diffusion of information technology, m-learning and e-commerce, outsourcing, networking, healthcare and information technology, and the re-engineering of business processes.

RU-SAN TAN received the degree in medicine from the National University of Singapore, in 1991. He is currently a Consultant with the Department of Cardiology, National Heart Centre Singapore, and an Adjunct Associate Professor with Duke-NUS Medical School. He specializes in non-invasive cardiac imaging: cardiovascular magnetic resonance imaging, echocardiography, and nuclear cardiology. He has been site a principal investigator and a steering committee member of several multicenter clinical trials of cardiovascular drugs. His research interests include advanced cardiovascular imaging, bioengineering, signal processing, and AI applications.

PRABAL DATTA BARUA received the Ph.D. degree in information systems from the University of Southern Queensland. He is currently an Academic and an Accredited Research Supervisor with the University of Southern Queensland, with 12 years of teaching experience, where he is also an Adjunct Professor. He is an Honourary Industry Fellow with the University of Technology Sydney. He received support from Queensland Government Innovation Connections under the Entrepreneurs Program to research "Cancer recurrence using innovative machine learning approaches." He has published several articles in Q1 journals. He is interested in AI technology development in health, education, agriculture, and environmental science. He is the Industry Leader in ICT entrepreneurship in Australia and an ICT Advisory Panel Member of many organizations.

RAVINESH C. DEO (Senior Member, IEEE) received the B.Sc. degree from The University of the South Pacific, Fiji, the M.Sc. degree from the University of Canterbury, New Zealand, and the Ph.D. degree in signal processing from The University of Adelaide, Australia. He has secured significant research grants and has published more than 300 articles that have cumulative citation of more than 15,400 and an H-index of over 71. He leads the UniSQ's Advanced Data Analytics Research Laboratory, University of Southern Queensland, Australia. His research interests include artificial intelligence and applications to the health, energy, wireless communication, national intelligence and defence, and engineering domain. His field of research is machine learning, artificial intelligence, and applied computing. He is a Highly Cited Researcher ranked in Top 1% by Clarivate Analytics and Web of Science, in 2021, and Top 2% of World Scientists by Stanford University, USA, from 2021 to 2023. He was educated and received the Senior School National Dux Award at Labasa College, Fiji. He mentored more than 40 higher degree students, many of them receiving a best thesis award. He received the Excellence in Research Award and the Ronel Erwee Memorial Award for Excellence in Postgraduate Supervision, Faculty Research Excellence, University and School Publication, Faculty Teaching Excellence, and Elsevier Highly Cited Papers for influential contributions. He received the Gold Medal, during the B.S. degree.

MAKIKO KOBAYASHI (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from the Department of Electrical and Electronic Engineering, Chiba University, Japan, in 1997 and 1999, respectively, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 2004. From 2004 to 2011, she contributed her expertise to the Industrial Materials Institute, Division of the National Research Council of Canada (NRCC). Since 2012, she has been holding an associate professor position with Kumamoto University, where she was promoted to a Professor, in 2022. She has been the Co-Founder and a Technical Advisor of CAST Company Ltd., since 2019. Her research interests include the development of porous piezoelectric materials for high-temperature ultrasonic transducers and various applications. She received the Young Scientist Award from the 26th Symposium on Ultrasonic Electronics, in 2005, and was honored with NRC's 2022 IP Achievement Award.

U. RAJENDRA ACHARYA received the D.Eng. and D.Sc. degrees. He is currently a Professor of artificial intelligence in health with the University of Southern Queensland, Australia, a Distinguished Professor with the International Research Organization for Advanced Science and Technology, Kumamoto University, Japan, an Adjunct Professor with the University of Malaya, Malaysia, and an Adjunct Professor with Asia University, Taiwan. His funded research has accrued cumulative grants exceeding six million Singapore dollars. He has authored over 600 publications, including 550 in refereed international journals, 42 in international conference proceedings, and 17 books. His research interests include biomedical imaging and signal processing, data mining, and visualization, as well as applications of biophysics for better healthcare design and delivery. He has received over 67,500 citations on Google Scholar (with an H-index of 130). He has been ranked in the top 1% of the highly cited researchers for the last seven consecutive years in computer science, according to the Essential Science Indicators of Thomson, from 2016 to 2022. He is on the editorial boards of many journals and has served as a guest editor for several AI-related issues.