Global NEST Journal, Vol 26, No X, pp 1-8

Water quality index prediction and classification using hyperparameter tuned deep learning approach

Sathya Preiya V. M.[1], Subramanian P.[2], Soniya M.[3]and Pugalenthi R.[4]

[1]Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, 600123, India

[2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India

[3]Department of Information Tecnology, Vel Tech Multi Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India

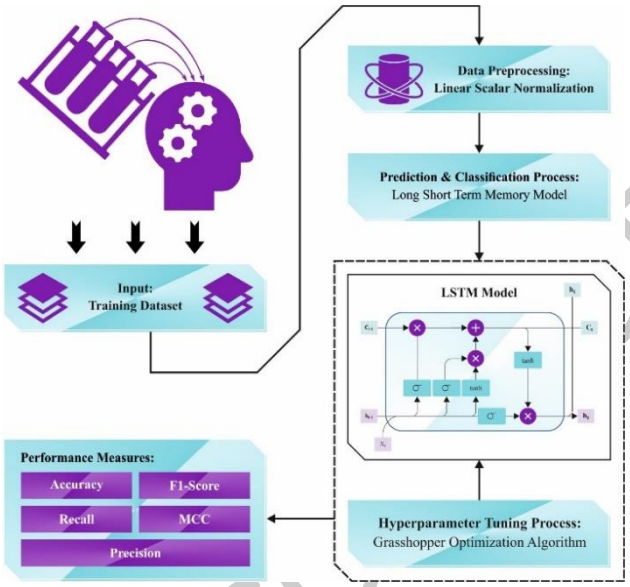
[4]Department of Artificial Intelligence and Data Science, St. Joseph's College of Engineering Chennai, 600119, India

Received: 13/02/2024, Accepted: 11/04/2024, Available online: 21/04/2024

*to whom all correspondence should be addressed: e-mail: vsathyapreiya@panimalar.ac.in

https://doi.org/10.30955/gnj.005821

Graphical abstract



Abstract

Water quality (WQ) is hugely important for animals, humans, plants, industries, and the environment. In the past few years, the WQ has been compressed by pollution and contamination. Usually, WQ is assessed utilizing costly laboratory and arithmetical processes, making real observation ineffective. Whereas, the poor WQ wants a more real and cost-effective resolution. Water pollution is a critical problem, so, it is vital to generate a method that estimates WQ in order to manage water pollution and notify users on the occasion of the recognition of poor water superiority. For effectual WQ management, it is vital to precisely estimate the WQ type. We use the advantage of machine learning (ML) models to build a model proficient in forecasting the WQ index and class. Therefore, this paper presents an automated Water Quality Index Prediction and Classification using Hyperparameter Tuned Deep Learning (WQIPC-HTDL) Approach. The purpose of the WQIPC-HTDL technique is to estimate WQI and classify the WQ into multiple levels. In the WQIPC-HTDL technique, the linear scaling normalization (LSN) approach is used. Besides, the long short-term memory (LSTM) technique is employed for the prediction and classification process. To enhance the efficacy of the LSTM model, the grasshopper optimizer algorithm (GOA) can be used. To point out the enhanced performance of the WQIPC-HTDL technique, a detailed simulation analysis was made. The obtained values inferred the rule of the WQIPC-HTDL technique when equated to other models.

Keywords: Water quality index; deep learning; grasshopper optimization algorithm; linear scaling normalization; machine learning

1. Introduction

Water is a major source of life, essential for helping the life of most present creatures and human beings. To continue their lives, living organisms require water with sufficient quality (Wang *et al.* 2024). There are specific restrictions on pollution that aquatic types are tolerated. These restrictions affect the presence of such living beings and threaten their survival. The majority of the environment's water bodies like streams, lakes, and rivers have particular quality values that show their quality (Prasad *et al.* 2022). Additionally, water conditions for other utilization retain their standards. For example, irrigation water should be neither too salt water nor comprise poisonous materials, which will be transported to soil or plants and therefore destroy the environment (Wong *et al.* 2023). Water quality (WQ) for industries also needs various aspects dependent upon the particular industrial methods. A few of the lower-cost resources of pure water namely surface and groundwater are real water resources. However, these sources could be polluted by manufacturing activities or humans and alternative natural methods (Saeed *et al.* 2024).

Accordingly, fast industrial expansion will influence the degradation of WQ at a disturbing rate (Georgescu *et al.*

2023). Additionally, surroundings with the lack of public awareness, and lesser hygienic qualities, mainly affect the quality of drinking water. Indeed, the significance of contaminated drinking water will be more risky and seriously affect the environment, infrastructure, and health (Rustam *et al.* 2022) Consequently, it is very significant to suggest a novel method for analysing and predicting the WQ. This can be suggested to analyze the temporal dimensions to predict the WQ patterns to ensure the observing of the seasonal variant of the WQ (Xu *et al.* 2024). Nevertheless, WQ could be analyzed employing conventional methods like gathering manually the water samples followed by examining them in a laboratory (Zamani *et al.* 2023). Then, it is considered expensive and time-consuming. Sensors are also categorized as alternative traditional methods. But, with the help of sensors can be deliberated expensive to test each WQ sample and frequently indicate lower accuracy (Arepalli Naik 2024). One more solution for monitoring WQ can be predictable modeling using machine learning (ML) and deep learning (DL) models. By comparison with other traditional techniques, it has numerous benefits: fewer costs, effective with respect to the time needed for travel and assortment, allows prediction on diverse stages of a method, and forecasts required values while retrieving a location will be difficult (Debow *et al.* 2023).

This research develops an automated Water Quality Index Prediction and Classification using the Hyperparameter Tuned Deep Learning (WQIPC-HTDL) Approach. The purpose of the WQIPC-HTDL technique is to estimate WQI and classify the WQ into multiple levels. In the WQIPC-HTDL technique, the linear scaling normalization (LSN) approach is used. Besides, the long short-term memory (LSTM) method is employed for the prediction and classification process. To enhance the efficacy of the LSTM approach, the grasshopper optimizer algorithm (GOA) can be used. To point out the enhanced performance of the WQIPC-HTDL technique, a detailed simulation analysis was made. The obtained values inferred the supremacy of the WQIPC-HTDL technique compared to other models.

2. Literature survey

Arepalli and Naik (2024) developed an improved Dilated Spatial-temporal CNN (DSTCNN) method. The WQ data taken by employing the IoT sensors are considered as per the WQ index values for exploration. The labeled data was efficiently categorized into 2 types by the developed DSTCNN method. Additionally, the developed technique utilizes a hybrid activation function that synchronously integrates ReLU and sigmoid function. In (Islam and Irshad 2022), an artificial ecosystem optimization with a DL-assisted WQ Prediction and Classification (AEODL-WQPC) system was introduced. In a primary processing step, the data normalization method was employed. Along with this, an optimum stacked BiGRU (OS-BiGRU) algorithm was employed for predicting, and the Adam optimizer was employed for tuning. AEO with an improved ENN (AEO-IENNs) system was implemented in the classification. Talukdar *et al.* (2023) developed a stacking ensemble method dependent upon the DL method by incorporating 3 techniques namely Gradient Neural Network (GNN), Generalized Linear Model (GLM), and Boosting Machine (GBM). The inclusion of a DNN method that could be employed for the primary time in water pollution exploration for executing the uncertainty and sensitivity analysis in forecasting WQI, included the novel dimensions for the workflow.

Shin *et al.* (2024) introduced an AI method by forecasting dissolved organic carbon (DOC) elimination and decontamination byproduct formations, and relatively examined present experimental systems and predicted outcomes for analyzing the utility of the AI method. This article improved experimental methods for forecasting DOC removal and disinfection byproduct formations. Six AI methods have been implemented and examined employing real-time data. In (Khullar and Singh 2022), a DL–based BiLSTM (DLBL-WQA) technique was presented. The developed method exhibits a new model which comprises missing values attribution in the major phase, the secondary phase produces the feature maps from the specified input data, the last phase comprises a BiLSTM model for increasing the learning method.

Moeinzadeh *et al.* (2023) projected a DL-based method for reconstructing these seven factors from four parameters for estimating the WQI. The technique also enables the analysis of sample qualities by computing the WQI through 7 synthesized factors with verified possible hydrogen and complete dissolved solids values. Correct evaluation of these parameters will be crucial to estimating the correctness of water for diverse functions. In (El-Shebli *et al.* 2023), a DNN method was developed for predicting WQI. Statistical modeling and unsupervised ML approaches have been employed. This modeling comprises the PCA or Factor Analysis (FA) that will be employed for interpreting the seasonal variations and sources of springs. Another modeling method was employed by using Hierarchical Cluster Analysis (HCA).

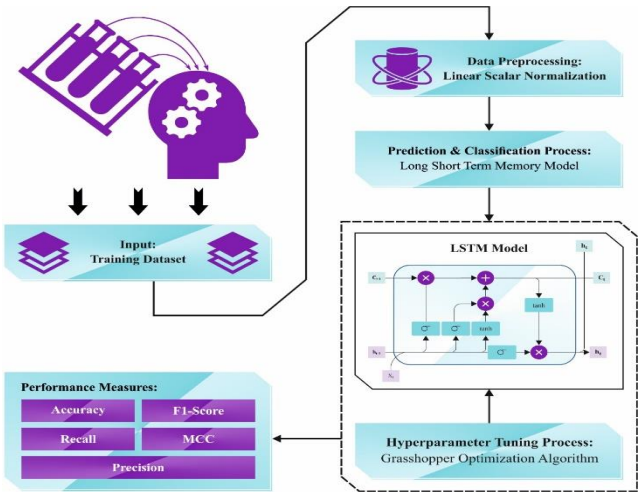


Figure 1. Workflow of WQIPC-HTDL methodology

3. The proposed method

In this study, we have presented an automated WQIPC-HTDL model. The main intention of the WQIPC-HTDL approach is to estimate WQI and classify the WQ into multiple levels. It contains three different procedures

namely LSN-based preprocessing, LSTM-based classification, and GOA-based parameter tuning process. Figure 1 demonstrates the workflow of the WQIPC-HTDL method.

3.1. Preprocessing

Initially, the WQIPC-HTDL technique undergoes the LSN approach is used. LSN is a data pre-processing model generally employed in numerous areas, containing image processing and ML (Sorguli and Rjoub 2023). The main aim of LSN is to normalize the arithmetical range of features or pixel values within a database. This technique involves linearly altering the original values so that they decrease within a definite range, normally between 0 and 1. By using a linear scaling alteration, LSN certifies that all data points are evenly adjusted, averting the dominance of definite features with higher arithmetical scales and enabling the convergence of algorithms during model training.

3.2. WQI Prediction using LSTM Model

In this paper, the LSTM technique is applied for the prediction and identification process. Though the RNN can develop sequential data, it is prone to gradient explosion or disappearance problems (Li *et al.* 2021). To solve these problems, the LSTM model is introduced. In comparison to RNN, LSTM comprises three logic gates (forget gate f_t, input gate i_t, and output gate o_t) along with memory unit C_t. It defines the output data at the present moment via the output data at the prior moment and the input data at the present moment and utilizes output data at the existing moment as input data. Using three logic gates and a memory unit, LSTM decides what amount of data and the input data at present to be kept, which can better abandon and transmit the data. However, the LSTM processes only the data in one direction. Thus, there is a need for data processing in the reverse and forward directions. Next, the Bi-LSTM is introduced so that the data in reverse and forward directions can be simultaneously processed. It implies the output of the Bi-LSTM network has context data. Where x_t and $h_t$are the input and hidden vectors at time *t*, correspondingly, *U* and *W* are the weight matrices, and b indicates the term bias.

The forget gate expresses what amount of data to be forgotten by outputting the value within [0,1] as follows:

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{1}$$

The input gate decides what data to retain by evaluating i_t and C_t and combine them based on the subsequent:

$$i_t = \sigma\left(W_i h_{t-1} + U_{i^{x_t}} + b_i\right) \tag{2}$$

$$C_t = \tanh\left(W_c h_{t-1} + U_c x_t + b_c\right) \tag{3}$$

$$A_t = f_t \odot A_{t-1} + i_t \odot C_t. \tag{4}$$

The output gate decides which part of the data to be outputted according to the equations:

$$o_t = \sigma\left(W_o h_{t-1} + U_{o^{x_t}} + b_o\right) \tag{5}$$

$$h_t = o_t \odot \tanh\left(A_t\right) \tag{6}$$

Bi-LSTM fuses the reverse and forward hidden states as a last hidden representation at *t* moment. Thus, contextual information can be better learned, and it contributes to the information flow in both directions.

3.3. Hyperparameter tuning

Finally, the GOA can be used to enhance the efficacy of the LSTM model. In the wild, grasshoppers exhibit the capability to find food sources and combine in clusters for reproduction and movement (Meraihi *et al.* 2021). A distinguishing representative of GOA is its calculation of velocities and positions for virtual grasshoppers, each targeted at enhancing the main function value of the specified issue. The formula to upgrade the grasshoppers's location can be given below:

$$X_i = S_i + G_i + A_j \tag{7}$$

S_i refers to the interaction of social which indicates the relationship among the *ith* grasshopper; X_{ii} represents the position of the *ith* grasshopper; G_i describes the gravitational attraction applied under *ith* grasshopper; A_i refers to the effect of air and wind circulation in the *ith* grasshopper. It should be noted that combines stochastic behavior, the mathematical formula could be expressed below:

$$X_i = r_1 S_i + r_2 G_i + r_3 A_i \tag{8}$$

r_1, r_2 and r_3 denotes the random amount within the interval of [0, 1]

The S element in Eq. (7) can be calculated through the following equation:

$$S_i = \sum_{\substack{j=1 \\ j \neq 1}}^{N} s\left(d_{ij}\right) d_{ij} \tag{9}$$

represents the distance among the ith and jth grasshopper, computed as below: dij=xj−xi; s signifies the power of social interface; dij=xj−$xidij$ shows the unit vector from the ith to *jth* grasshopper

The magnitude of social interactive powers could be calculated by the function s that has been calculated based on the next equation:

$$s\left(r\right) = fe^{\frac{-r}{l}} - e^{-r} \tag{10}$$

r refers to the distance value, *l* denotes the amount at which the power of social interaction decreases with distance, and *f* describes the strength of attraction of social interaction that affects the level of mutual attraction and interaction amongst grasshoppers.

The G element in Eq. (7) can be estimated employing the following equation:

$$G_i = -ge_g \tag{11}$$

Here, e_g is the unit vector directed to the Earth's centre and g denotes the gravitational constant.

The A module in Eq. (7) is estimated by employing the subsequent formulation:

$$A_i = ue_w \tag{12}$$

Whereas, u defines the constant drift; e_w stands for unit vector combined with the path of the wind.

With the help of replacing, *G*, and A into the equation given in Eq. (7), the mathematical expression can be described in Eq. (13):

$$X_i = \sum_{\substack{j=1 \\ j\neq 1}}^{N} s\left(\left|x_{j-} x_i\right|\right)\frac{x_{j-x_i}}{d_{ij}} - ge_g + ue_w \tag{13}$$

Where, eg is the unit vector directed to the Earth center, d_{ij} denotes the distance amongst the *ith* and *jth* grasshopper, g refers to the gravitational constant; s is the strength of social interaction forces; u refers the constant drift; ew represents the unit vector united with the wind's direction.

Within the context of the optimizer method, Eq. (13) has been deliberately prevented because of its trend to constrain the method's capability to systematically discover and exploit the adjacent areas within the solution space. This particular nymph grasshopper system is complexly created to overcome a grasshopper swarm function in an infinite space. The significant mathematical model could not be directly implemented for solving optimization glitches, as the grasshoppers rapidly unite to their comfort regions and the group can not be changed to a singular point. An adapted version of Eq. (13) can be utilized to successfully overcome optimization challenges:

$$X_i^d = c\left(\sum_{\substack{j=1 \\ j\neq 1}}^{N} c\frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + T_d \tag{14}$$

c indicates the coefficient value, *ubd* represents the upper limit, *Td* is the desired value and *lbd* describes the lower limit.

To calculate the following grasshopper position, data including the target's location, the existing grasshopper's places, and the location of each grasshopper will be employed. According to Eq. (14), the following place of a separate grasshopper could be resolved by an integration of its existing place, the global finest outcome, and the position information of other search agents. It denotes that the GOA needs an effective contribution of each search agent in modeling the path of all grasshoppers. Particularly, the initial measure of Eq. (14) considers the comparative positioning of the existing grasshopper with compared to complements within the field. On the other hand, the following segment bounds the level of movement near the target position. The opposition emphasizes the technique's search of both wide-ranging exploration and considered exploitation with the complete swarm cantered on the target.

To simplify, c_1 denotes the restriction level executed on grasshopper actions near the objective, attaining a stable equilibrium among exploitation and exploration in the group. Conversely, c_2 donates to the reduction of repulsion, comfort zones, and attraction amongst grasshoppers, efficiently decreasing the spatial range. Accordingly, c_2 directs the grasshoppers to navigate the search space to the optimum outcome.

Important is the adaptive nature of c_1 that gradually diminishes the impact of repulsion and attraction forces between grasshoppers in ratio to the iteration count. Simultaneously c_2 progressively decreases the width of the comfort region with improving rounds. The tactical interaction arises, where c_1 improves exploitation in the next optimization phases, and c_2 progressively contracts the zones for improving proximity to the optimum solution. Both c_1 and c_2 are combined as only one parameter that can be considered to alteration as given below:

$$c = c_{\max} - l\frac{c_{\max} - c_{\min}}{L} \tag{15}$$

L states the total number of iterations, *c*min denotes the lower boundary of the parameter *c*, *c*max describes the upper limit *c*; *l* represents the existing iteration. Figure 2 demonstrates the steps involved in GOA.

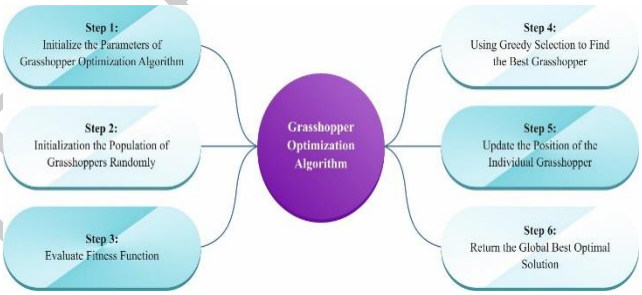


Figure 2. Steps involved in GOA

The fitness function (FF) is the significant factor manipulating the GOA performance. The hyperparameter range method contains the solution encoding technique for evaluating the efficiency of the candidate solution. In this work, the GOA reflects accuracy as the main standard to propose the FF, which can be expressed below.

$$Fitness = \max(P) \tag{16}$$

$$P = \frac{TP}{TP + FP} \tag{17}$$

From the above formulae, TP and FP denotes the true positive and false positive values.

4. Result analysis

The WQI detection outcomes of the WQIPC-HTDL system can be assessed on the WQ dataset from Kaggle [https://www.kaggle.com/datasets/mssmartypants/water-quality]. It includes 1600 samples with two classes as defined in Table 1.

Table 1. Details on database

Classes	No. of Samples
WQI-Not Safe	800
WQI-Safe	800
Total Samples	1600

Figure 3 establishes the confusion matrices formed by the WQIPC-HTDL approach below 80:20 and 70:30 of TRAPH/TESPH. The outcomes specify that the WQIPC-HTDL has effective detection of the WQI-Not Safe and WQI-Safe samples under all classes.

The WQI recognition of the WQIPC-HTDL technique is demonstrated under 80% of TRAPH and 20% of TESPH in Table 2 and Figure 4. The results demonstrate the proficient capability of the WQIPC-HTDL system in the detection of the WQI.

Table 2. WQI recognition of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$	MCC
TRAPH (80%)					
WQI-Not Safe	84.48	97.47	84.48	90.51	83.07
WQI-Safe	97.82	86.38	97.82	91.75	83.07
Average	91.15	91.93	91.15	91.13	83.07
TESPH (20%)					
WQI-Not Safe	89.51	94.77	89.51	92.06	84.52
WQI-Safe	94.94	89.82	94.94	92.31	84.52
Average	92.22	92.30	92.22	92.19	84.52

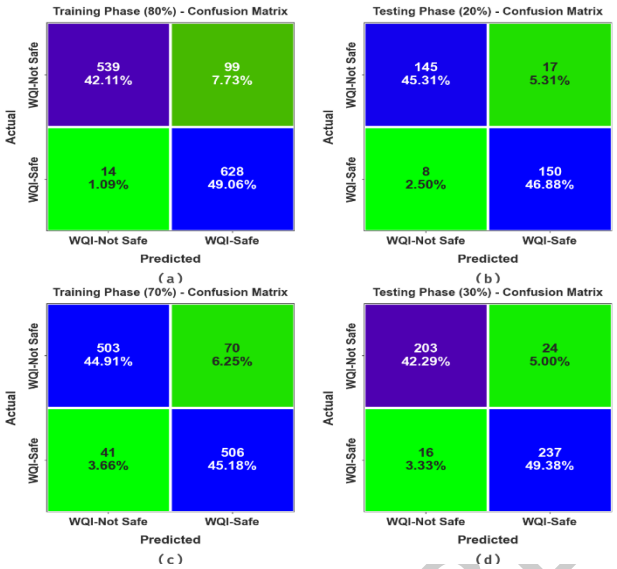


Figure 3. Confusion matrices of (a-b) 80:20 of TRAPH/TESPH and (c-d) 70:30 of TRAPH/TESPH

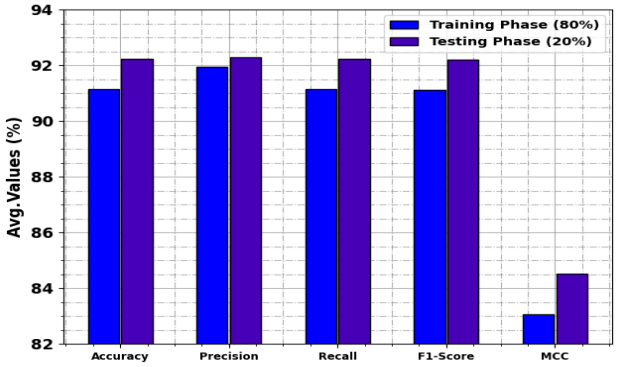


Figure 4. Average of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

With 80% of TRAPH, the WQIPC-HTDL technique gains average $\alpha accu_y$, $prec_n$, $reca_l$, and F_{score}, and MCC of 91.15%, 91.93%, 91.15%, 91.13%, and 83.07%, respectively. Additionally, with 20% of TESPH, the WQIPC-HTDL model gains average $accu_y$, $prec_n$, $reca_l$, and F_{score}, and MCC of 92.22%, 92.30%, 92.22%, 92.19%, and 84.52%, correspondingly.

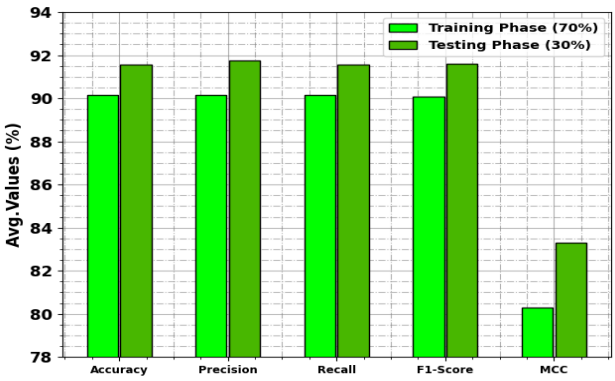


Figure 5. Average of WQIPC-HTDL approach under 70:30 of TRAPH/TESPH

The WQI detection of the WQIPC-HTDL method is verified below 70% of TRAPH and 30% of TESPH in Table 3 and Figure 5. The results determine the proficient capability of the WQIPC-HTDL model on the recognition of the WQI. With 70% of TRAPH, the WQIPC-HTDL method acquires average $accu_y$, $prec_n$, $reca_l$, and F_{score}, and MCC of 90.14%, 90.16%, 90.14%, 90.19%, and 80.30%, correspondingly. Moreover, with 30% of TESPH, the WQIPC-HTDL system gains average $accu_y$, $prec_n$, $reca_l$, and F_{score}, and MCC of 91.55%, 91.75%, 91.55%, 91.62%, and 83.30%, respectively.

The performance of the WQIPC-HTDL technique below 80:20 of TRAPH/TESPH is graphically offered in Figure. 6 in the method of training accuracy (TRAA) and validation accuracy (VALA) curves. The figure displays a beneficial analysis into the behavior of the WQIPC-HTDL technique over numerous epoch counts, representing its learning procedure and generalization abilities. Remarkably, the figure infers a stable development in the TRAA and VALA with a development in epochs. It safeguards the adaptive nature of the WQIPC-HTDL system in the pattern recognition procedure on both TRA and TES data. The rising trend in VALA sketches the ability of the WQIPC-HTDL technique to adjust to the TRA data and also excels in providing precise classification of hidden data, indicating strong generalization skills.

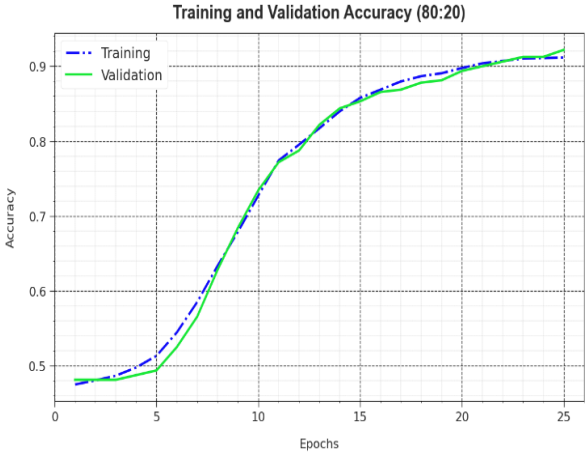


Figure 6 $Accu_y$ curve of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

Figure 7 exhibits a comprehensive representation of the training loss (TRLA) and validation loss (VALL) results of the WQIPC-HTDL approach below 80:20 of TRAPH/TESPH over different epochs. The progressive reduction in TRLA highlights the WQIPC-HTDL system enhancing the weights and minimalizing the classification error on the TRA and TES data. The figure specifies a clear understanding of the WQIPC-HTDL model's association with the TRA data, emphasizing its ability to take patterns within both datasets. Remarkably, the WQIPC-HTDL method repeatedly increases its parameters in diminishing the alterations among the forecast and actual TRA class labels.

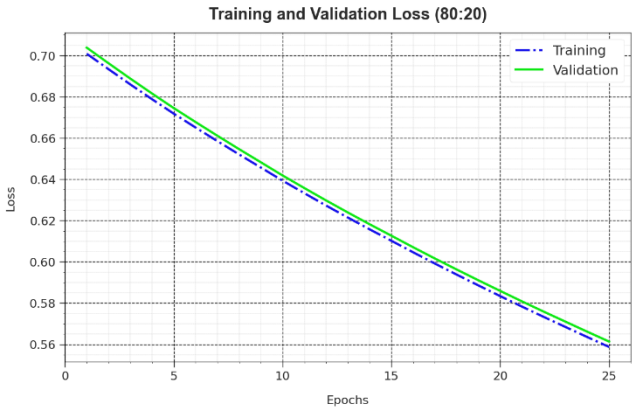


Figure 7. Loss curve of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

Inspecting the PR curve, as exposed in Figure 8, the results certified that the WQIPC-HTDL method below 80:20 of TRAPH/TESPH gradually achieves enhanced PR values under every class. It confirms the improved skills of the WQIPC-HTDL approach in the classification of dissimilar classes, demonstrating proficiency in the recognition of classes.

Table 3. WQI recognition of WQIPC-HTDL approach under 70:30 of TRAPH/TESPH

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$	MCC
TRAPH (70%)					
WQI-Not Safe	87.78	92.46	87.78	90.06	80.30
WQI-Safe	92.50	87.85	92.50	90.12	80.30
Average	90.14	90.16	90.14	90.09	80.30
TESPH (30%)					
WQI-Not Safe	89.43	92.69	89.43	91.03	83.30
WQI-Safe	93.68	90.80	93.68	92.22	83.30
Average	91.55	91.75	91.55	91.62	83.30

Table 4. Comparative analysis of WQIPC-HTDL technique with other approaches

Algorithm	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
WQIPC-HTDL	92.22	92.30	92.22	92.19
MLP Algorithm	85.09	56.60	56.42	56.51
Logistic Regression	84.03	55.21	55.96	55.50
KNN Model	72.72	47.36	47.84	47.52
Decision Tree	79.51	53.00	52.52	52.69
Random Forest	75.88	50.65	50.13	50.29
SVM Model	79.80	51.88	53.28	52.30

Besides, in Figure 9, ROC curves formed by the WQIPC-HTDL system under 80:20 of TRAPH/TESPH outperformed the classification of different labels. It delivers a comprehensive understanding of the tradeoff among TPR and FRP over dissimilar detection threshold values and epoch counts. The figure emphasized the higher classifier outcomes of the WQIPC-HTDL technique under all classes, outlining the efficacy in addressing many classification problems.

Table 4 and Figure 10 report a detailed comparison study of the WQIPC-HTDL technique (Ahmed *et al.* 2019). It is noticed that the KNN, DT, RF, and SVM models gain ineffectual results. Next to that, the MLP and LR models have reached considerable performance. But the WQIPC-

HTDL technique demonstrates its superior performance with increased $accu_y$, $prec_n$, $reca_l$, and F_{score} of 92.22%, 92.30%, 92.22%, and 92.19%, respectively. Thus, the WQIPC-HTDL technique can be applied to the automated WQI detection process.

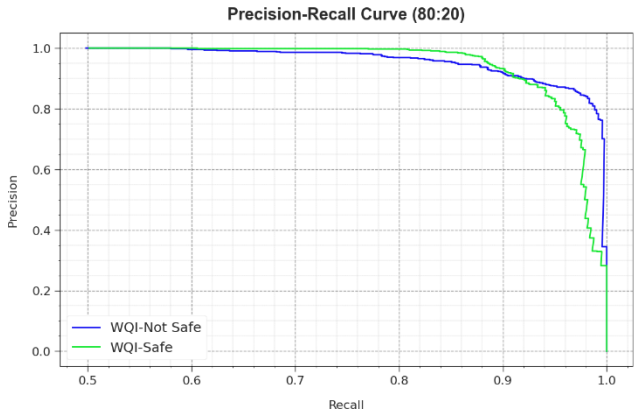


Figure 8. PR curve of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

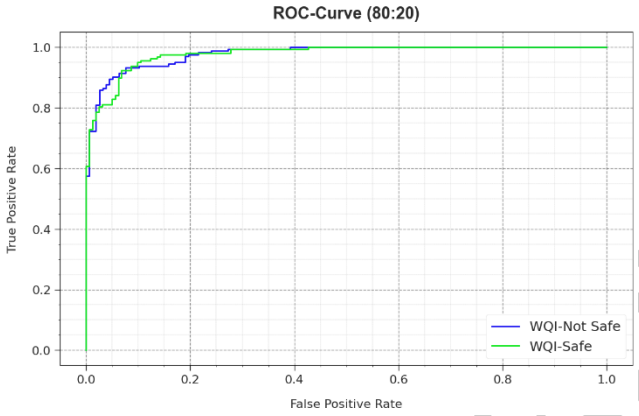


Figure 9. ROC curve of WQIPC-HTDL approach under 80:20 of TRAPH/TESPH

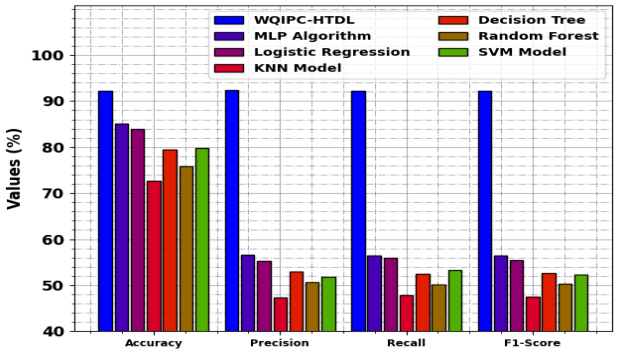


Figure 10. Comparative analysis of WQIPC-HTDL technique with other approaches

5. Conclusion

In this research, we have presented an automated WQIPC-HTDL model. The main intention of the WQIPC-HTDL method is to estimate WQI and classify the WQ into multiple levels. It contains three different procedures namely LSN-based preprocessing, LSTM-based classification, and GOA-based parameter tuning process. Initially, the WQIPC-HTDL technique undergoes the LSN approach is used. Besides, an LSTM model is applied for the prediction and classification process. To enhance the efficacy of the LSTM model, the GOA can be used. To point out the enhanced performance of the WQIPC-HTDL technique, a detailed simulation analysis was made. The obtained values inferred the supremacy of the WQIPC-HTDL technique compared to other models.

References

Ahmed U., Mumtaz R., Anwar H., Shah A.A., Irfan R. and García-Nieto J. (2019). Efficient water quality prediction using supervised machine learning. *Water*, **11**(11), 2210.

Arepalli P.G. and Naik K.J. (2024). An IoT based smart water quality assessment framework for aqua-ponds management using Dilated Spatial-temporal Convolution Neural Network (DSTCNN). *Aquacultural Engineering*, **104**, 102373.

Arepalli P.G. and Naik K.J. (2024). Water contamination analysis in IoT enabled aquaculture using deep learning based AODEGRU. *Ecological Informatics*, **79**, 102405.

Debow A., Shweikani S. and Aljoumaa K. (2023). Predicting and forecasting water quality using deep learning. *International Journal of Sustainable Agricultural Management and Informatics*, **9**(2), 114–135.

El-Shebli M., Sharrab Y. and Al-Fraihat D. (2023). Prediction and modeling of water quality using deep neural networks. Environment, *Development and Sustainability*, 1–34.

Georgescu P.L., Moldovanu S., Iticescu C., Calmuc M., Calmuc V., Topa C. and Moraru L. (2023). Assessing and forecasting water quality in the Danube River by using neural network approaches. *Science of The Total Environment*, **879**, 162998.

https://www.kaggle.com/datasets/mssmartypants/water-quality

Islam N. and Irshad K. (2022). Artificial ecosystem optimization with deep learning enabled water quality prediction and classification model. *Chemosphere*, **309**, 136615.

Khullar S. and Singh N. (2022). Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. *Environmental Science and Pollution Research*, **29**(9), 12875–12889.

Li H., Ma Y., Ma Z. and Zhu H. (2021). Weibo text sentiment analysis based on bert and deep learning. *Applied Sciences*, **11**(22), 0774.

Meraihi Y., Gabis A.B., Mirjalili S. and Ramdane-Cherif A. (2021). Grasshopper optimization algorithm: theory, variants, and applications. *Ieee Access*, **9**, 50001–50024.

Moeinzadeh H., Jegakumaran P., Yong K.T. and Withana A. (2023). Efficient water quality prediction by synthesizing seven heavy metal parameters using deep neural network. *Journal of Water Process Engineering*, **56**, 104349.

Prasad D.V.V., Venkataramana L.Y., Kumar P.S., Prasannamedha G., Harshana S., Srividya S.J., Harrinei K. and Indraganti S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of the Total Environment*, **821**, 153311.

Rustam F., Ishaq A., Kokab S.T., de la Torre Diez I., Mazón J.L.V., Rodríguez C.L. and Ashraf I. (2022). An Artificial Neural Network Model for Water Quality and Water Consumption Prediction. *Water*, **14**(21), 3359.

Saeed A., Alsini A. and Amin D. (2024). Water quality multivariate forecasting using deep learning in a West Australian estuary. *Environmental Modelling & Software*, **171**, 105884.

Shin H., Byun Y., Kang S., Shim H., Oak S., Ryu Y., Kim H. and Jung N. (2024). Development of water quality prediction model

for water treatment plant using artificial intelligence algorithms. *Environmental Engineering Research*, **29**(2).

Sorguli S. and Rjoub H. (2023). A Novel Energy Accounting Model Using Fuzzy Restricted Boltzmann Machine—Recurrent Neural Network. *Energies*, **16**(6), 2844.

Talukdar S., Ahmed S., Naikoo M.W., Rahman A., Mallik S., Ningthoujam S., Bera S. and Ramana G.V. (2023). Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms. *Journal of Cleaner Production*, **406**, 136885.

Wang J., Xue B., Wang Y., Yinglan A., Wang G. and Han D. (2024). Identification of pollution source and prediction of water quality based on deep learning techniques. *Journal of Contaminant Hydrology*, **261**, 104287.

Wong W.Y., Hasikin K., Khairuddin M., Salwa A., Razak S.A., Hizaddin H.F., Mokhtar M.I. and Azizan M.M. (2023). A Stacked Ensemble Deep Learning Approach for Imbalanced Multi-Class Water Quality Index Prediction. *Computer. Materials. Contion*, **76**, 1361–1384.

Xu R., Hu S., Wan H., Xie Y., Cai Y. and Wen J. (2024). A unified deep learning framework for water quality prediction based on time-frequency feature extraction and data feature enhancement. *Journal of Environmental Management*, **351**, 119894.

Zamani M.G., Nikoo M.R., Jahanshahi S., Barzegar R. and Meydani A. (2023). Forecasting water quality variable using deep learning and weighted averaging ensemble models. *Environmental Science and Pollution Research*, **30**(59), 124316–124340.