Full Paper of the WasteSafe 2025 – 9th International Conference on Sustainable Solutions for Waste Management in Developing Countries
22-23 February 2025, Dhaka, Bangladesh

QUANTUM - BAYESIAN CONJUGATE NEURAL NETWORK FOR TDS ESTIMATION IN WASTE WATER TREATMENT PLANT

Shuvendu Pal Shuvo[*1], Samanta Shahrin[2], Tauhidul Islam[3], Chaitee Das[4], Shirshendu Pal Shibazee[5]

[1, 2,3] Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
[4]Department of Materials Science and Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
[5]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
[*]Corresponding Author
shuvenduce@gmail.com

ABSTRACT

This study aims to develop a Total Dissolved Solids estimation model using a conjugate solution approach that integrates Bayesian optimization and Quantum-Inspired Evolutionary algorithms (QEA) with Artificial Neural Networks and Seasonal-Trend Decomposition. The proposed method will overcome the deficiency of the available literature in which most of the research contributions are dependent on highly advanced techniques to analyses the TDS data of wastewater treatment plants, which is nonlinear and complex. This study focuses on data from a treatment facility located in Khulna, Bangladesh. For the construction of the model, Seasonal-Trend Decomposition (STD) is first applied to extract the trend and seasonal components from the TDS dataset. These components are then used for the training of the ANN. The Bayesian optimization algorithm is employed to optimize the ANN's hyperparameters, such as the configuration of the hidden layers and transfer functions, whereas QEA updates the network weights. The first 70% of the dataset was used for training, and the subsequent 15% was used for validation. The last 15% was reserved for testing. The results show that the proposed model achieves an R^2 value of 0.96, indicating a very good accuracy in simulating TDS levels. This reflects the efficiency of the combined approach in addressing the challenges of TDS prediction in wastewater treatment applications. The importance of this study involves providing a robust, efficient, and accurate method for the estimation of TDS for optimization in wastewater treatment processes. By combining the powers of Bayesian optimization, QEA, and ANN with STD, the proposed approach enables a practical solution to the nonlinear and dynamic water quality data management for improved environmental management practices.

INTRODUCTION

A TDS prediction model for a wastewater treatment plant is critical in terms of optimizing the treatment processes, achieving regulatory compliance, and environmental health protection. TDS levels are related to water quality and impact aquatic ecosystems, irrigation practices, and industrial reuse potential. Accurate predictions allow for proactive adjustments in treatment operations that reduce costs and resource consumption. It helps in the identification of potential problems at an early stage, such as scaling, corrosion, or salinity buildup, which can cause severe damage to infrastructure and further harm downstream ecosystems. Further, it helps in decision-making regarding wastewater reuse and discharge to ensure that environmental standards are met. With increased efficiency in operations and reduced environmental impact, a TDS prediction model becomes an asset for water management with sustainability and also a critical player in the global challenge of water scarcity.

This research work focuses on the Bangabandhu Wastewater Treatment Plant, one of the most important water supply facilities in Khulna. The development of the plant started in 2019 and contributes a lot to solving the regional water problems. However, developing a reliable prediction model for this plant is quite a big challenge due to its short operational time, which is not adequate to provide sufficient historical data. The limited availability of data restricts the correct analysis of trends, optimization of operations, and predictive solutions. Of the critical parameters in this context, TDS modeling is of particular importance since it is one of the key indicators of water quality and directly impacts plant performance and the safety of the water supply. Effective TDS modeling will, therefore, help maintain water quality standards, optimize treatment processes, and support long-term resource management in the region. Overcoming data scarcity is necessary to enhance operational efficiency at the plant and ensure the sustainability of water resources in Khulna. It underlines that these challenges will require innovative approaches, by combining alternative data sources, adopting advanced machine learning techniques, or resorting to simulation tools able to compensate for historical data shortcomings.

Various previous studies have utilized different methods to predict TDS and other water quality parameters. Research such as (Solaimany et al., 2013), (Lotfi et al., 2019), and (Banadkooki et al., 2020) utilized machine learning models to predict TDS by leveraging multiple parameters. ML models are especially well-suited for this task because they are able to grasp complex nonlinear relationships, which are basic to the inherently nonlinear nature of water quality data (Zaman Zad Ghavidel & Montaseri, 2014). Other works applied techniques for time series modeling, such as those found in (Ahmadpour et al., 2023; Yadav et al., 2023). However, these methods suffer from a limitation to model high nonlinear and noisy water quality datasets like that of the TDS variation. For this reason, they are barely capable of catching the intrinsic complexities in it (Grande ; Abudu & King, 2011). Amongst various machine learning approaches, ANNs (Yadav et al., 2023), SVMs (Koranga et al., 2021), LSTM (Hafezparast Mavadat & Marabi, 2021), and ANFIS (Tiwari et al., 2018) have shown promising accuracy in predicting water quality parameters. However, the noisy nature of water quality data due to measurement errors and other external factors has made the performance of these models very challenging. Noise brings variability in the data, hence affecting the performance of the predictive models by weakening their reliability and robustness. The resultant performance of this model, therefore, is not very accurate, with varying results. Thus, noise is the major problem that demands attention for improvement in machine learning models to effectively predict the water quality parameters reliably.

The study will address the deficiency of the literature by incorporating ANNs for TDS modeling. ANNs are one of the common machine learning algorithms that can handle highly complex signals. The ANN performance has been enriched in this study by integrating Bayesian optimization (BO) and Quantum-Inspired Evolutionary Algorithm (QEA) for optimizing its hyperparameters, weights, and biases. The use of these conjugate optimization techniques introduces novelty in the presented methodology by reducing runtime and complexity. Besides this, a preprocessing technique, namely STD, is used for removing the noise from the TDS data. Seasonal Trend Decomposition segregates the noisy TDS data into seasonal, trend, and residual, thus allowing the intrinsic patterns and trends to be caught which will enhance the performance of the ANN model. The main novelties of this study are:

- ✓ In this section, a conjugate optimization solution is applied by combining QEA and BO.
- ✓ Application of the STD as a preprocessing method for removing existing noise may enable better detection of a trend and pattern.

RESEARCH METHODOLOGY

Study Area and Data Collection

Khulna District in Bangladesh is one of the most salinity-affected and climate change-vulnerable areas, and the safe availability of drinking water has come under serious threat. This area faces rapidly increasing populations that have caused a large increase in demand for assured water supply. Groundwater is still the main source of drinking water supply in Khulna, and water purification is done at the Bangabandhu Water Treatment Plant situated at Pathorghata, Khulna. This plant draws its water from the Madhumati River, which gets a complete treatment in raw water filtration, waste removal, and distribution of purified water to the residents of Khulna City (Shuvo et al., 2024). In the study, weekly water quality data collected from the said treatment plant were used, focusing on TDS, which was monitored routinely to ensure proper assessment of inlet water quality. The data is shown in Figure 1.

Its importance lies in being a vital indicator of the purity of water, which reflects on the entire treatment's effectiveness. A model to predict the TDS at the outlet was necessary to optimize performance at the plant. A prediction model in this regard shall not only facilitate monitoring the quality of water but will also extend control and optimization of treatments to ensure improved reliability and operational efficiency. The study also analyzed data for the identification and understanding of individual and combined factors influencing TDS levels. In this regard, the dataset had been split into three parts: 70%

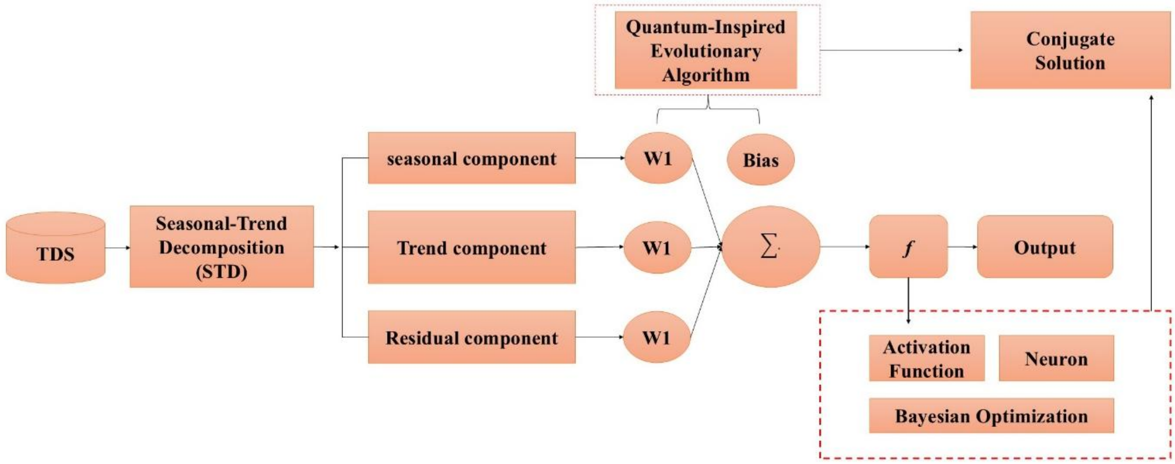

Figure 1 Collected Historical TDS Data

of the data was for training, to enable the model to learn intricate patterns; 15% was for validation, or

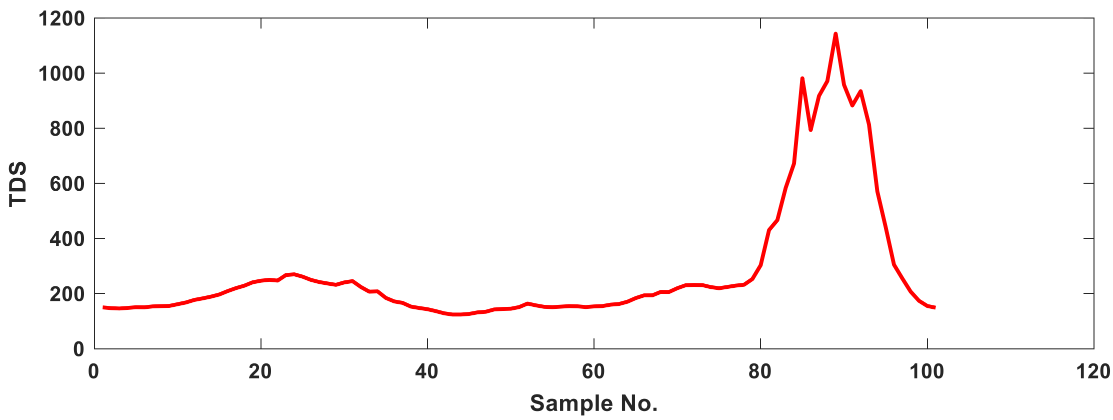

Figure 2 Overall Methodology of the proposed model

fine-tuning, to enhance the accuracy of the model; and the remaining 15% was for testing. The steps in this structured approach would make sure that this model is put to task when applied to new, unseen data to predict its strength with practical implementation for water quality management and decision-making.

Proposed Method

The research adopted the ANN methodology to model and predict TDS. The selection of the input features is one of the crucial factors in building an efficient prediction model. Traditional methods using different input parameters or the lag-time approach often fall short, especially in cases involving noisy data. Realizing the limitation of these traditional methods, this work used a preprocessing step to enhance the selection of input features. Preprocessing consisted of applying Seasonal Trend Decomposition to the TDS dataset.

The STD method decomposes the TDS data into three independent components: the seasonal component, the trend component, and the residual component. Each of these components reflects different aspects of the variability in the data. It shows that the seasonal component represents the patterns that recur over fixed time intervals, the trend component indicates the long-term behavior in which the data are evolving, and the residual component reflects irregular fluctuations or noise in the data. In the next step, the decomposed components were fed into the ANN model, making sure that the inputs were in a structured, meaningful format to enhance their predictive performance. Bayesian

Optimization was implemented to overcome the challenge of hyperparameter tuning, such as finding the optimal number of hidden layers and choosing the most appropriate transfer function.

This advanced optimization technique will search through the hyperparameter space in a structured manner seeking the best configuration for the ANN that will result in an enhanced predictive accuracy and efficiency of the model. The Quantum-Inspired Evolutionary Algorithm was used to fine-tune the weights and biases of the model. This innovative approach, inspired by principles from quantum computing, effectively optimizes the ANN parameters to ensure robust learning with improved model generalization. The integration of these advanced methodologies resulted in a comprehensive framework for TDS prediction throughout the study, which overcame most of the noisy data challenges and complex model parameterizations. The overall methodology is shown in Figure 2.

STD-BO-QEA-ANN

Seasonal-trend decomposition is a statistical technique for decomposing time series data into three key components: trend, seasonal, and residual. The trend component reflects the long-term direction or movement in the data and shows whether the data is increasing, decreasing, or remaining stable over time. The seasonal component is the repeated patterns or cycles that occur at fixed periods, which are normally of periodic influences, such as seasons, holidays, or other periodic factors. The residual component, which is sometimes referred to as the irregular component, accounts for the random noise or unexplained variations that remain after the removal of the trend and seasonal effects.

The ANN is a computational model that is based on the structure and functioning of the human brain, designed to process complex data and recognize patterns (Kouadri et al., 2022). It consists of interconnected layers of nodes or neurons, with each performing basic computations. Normally, the architecture includes an input layer to receive data, one or more hidden layers where learning occurs, and an output layer to produce results (Unnikrishnan & Jothiprakash, 2020). Each neuron within a layer is connected with other neurons in adjacent layers via weighted connections, which denote the strength of their influence on one another. These weights are modified during the learning process through optimization algorithms, such as gradient descent, directed by a loss function that measures the difference between the predicted and actual outcome. ANNs use nonlinear functions such as sigmoid, ReLU, or tanh as activation functions to introduce non-linearity, enabling them to model complex relationships. Among the common ways of training, supervised learning involves feeding the network with labeled data and iteratively refining weights by minimizing errors (Shad et al., 2022).

Bayesian optimization of the hyperparameters of a machine learning model has many applications: for example, trying to identify how many neurons it should have per layer in an ANN and a preferred choice for a transfer function (Di et al., 2022). This is very helpful for optimization in huge spaces where computations may take painfully long. The procedure creates a surrogate probabilistic model, many times with the Gaussian process model as an estimator of the objective function with a mapping in hyperparameter performance (Chaibi et al., 2021). Bayesian optimization iteratively selects hyperparameter configurations to explore, balancing the exploitation of promising areas and the exploration of less certain regions. This is an approach to finding an efficient combination of neurons and transfer functions. Second, tuning the number of neurons assures the right model complexity is sufficiently expressive with no overfitting or computational waste. Similarly, the optimization of more transfer functions like ReLU, sigmoid, or tanh tailors the network to capture the non-linear patterns within the data. The acquisition function, such as Expected Improvement or Upper Confidence Bound, guides the search for optimal configurations (Rong et al., 2020). This approach avoids the need for an exhaustive grid or random search, saving immense resources. Bayesian optimization now has state-of-the-art results for various ANN applications, thanks to the developed methodology that allows systematic refinements in the architecture for increased predictive accuracy and generalization.

Quantum-Inspired Evolutionary Algorithms are one of the advanced methods of optimization that incorporates principles of quantum computing into solving hard problems, including finding weights and biases of ANNs. QIEA differs from traditional gradient-based optimization approaches in the exploration of solutions using quantum-inspired principles of superposition and probability amplitude. These algorithms maintain a population of quantum individuals encoded as a probability distribution that describes the potential solution. Iterative operations in quantum rotation, mutation, and measurement drive the population toward optimum solutions. When applied to ANNs, QIEAs optimize weights and biases by searching for configurations that minimize a loss function, which quantifies the difference between predicted and actual outputs. The probabilistic nature of QIEAs allows them to efficiently explore vast and complex search spaces, avoiding local minima that often challenge gradient-based methods. Additionally, QIEAs are robust to irregular loss landscapes and can adapt to non-differentiable or discontinuous objective functions. By iteratively adjusting the weights and biases, QIEAs enable the ANN to reach a better level of accuracy and performance in generalization. Thus, it is of great use for problems involved with multi-objective optimizations or noisy environments, providing a novel perspective for neural network training algorithms in pursuing both efficiency and robustness.

RESULT AND DISCUSSION

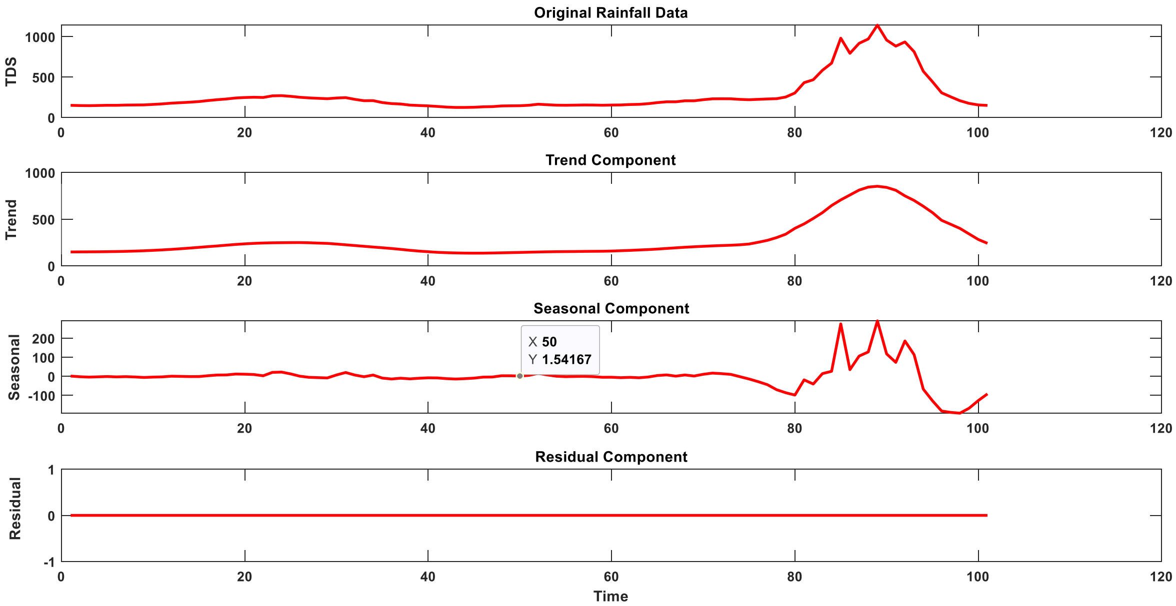

Figure 3 STD Decomposition of TDS

This work presents the application of a double-step optimized neural network in the prediction of TDS in wastewater treatment plants. The modeling here is informed by the Seasonal Trend Decomposition technique, which has been used in determining the trend and seasonal components contained in the dataset. The Figure 3 is shown the STD decomposed signal of TDS. It allows for the proper extraction of these components from the data; thus, it offers an excellent opportunity to understand the underlying behavior, hence giving a good basis for training the model. By performing the decomposition, the dataset becomes prepared for training by means of an ANN. As for the ANN, there are a series of steps undertaken to fine-tune the hyperparameters so it predicts optimally.

The Bayesian Optimization algorithm is put to work in finding that version of the neural network topology with the optimal number of hidden layers and the optimal activation function. Through this process, the model converges towards an architecture with 21 hidden neurons that uses an activation function called "leaky ReLU," very efficient in handling nonlinear data relationships. The optimization of model weights is performed by means of QEA; a state-of-the-art approach aimed at enhancing the learning efficiency of ANN for convergence towards global minima. The Figure 4 shows the BO optimized outcome of best transfer function and hidden neuron.

Some key performance metrics mark the effectiveness of the model developed. The MSE presents 4948.7671, while the Root Mean Squared Error of these presents 70.3475, which is fairly low and indicates low magnitudes of prediction error levels. Similarly, the mean of Absolute Error presents 57.3327, reflecting minimality in the average amount of deviation between predicted values and actual values. Most importantly, the R^2 is 0.9602, which reflects that the model explains 96.02% of the variance in TDS, showing a high degree of accuracy and reliability. This study has pointed out the potential of combining advanced preprocessing techniques such as STD with innovative optimization algorithms like Bayesian Optimization and QEA to enhance ANN performance. The model, by leveraging these approaches, is able to predict TDS in wastewater treatment with very high accuracy, which can be instrumental in improving the efficiency and sustainability of plant operations. Such methodologies as STD, BO, QEA, and ANN have a lot of advantages in predictive modeling. The Figure 5 illustrated the proposed model test outcome.

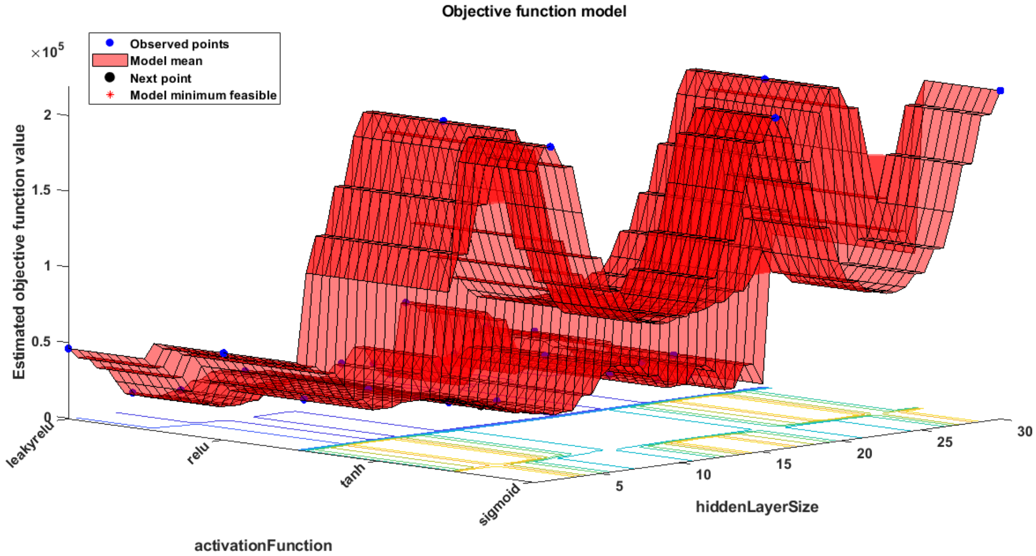

Figure 4 BO optimized best activation function and Hidden Layer Size

STD enhances data understanding by decomposing it into trend, seasonal, and residual components for clearer analysis of patterns and anomalies. This decomposition simplifies the data by reducing noise and enhancing model accuracy by ensuring that the neural network focuses on the relevant relationships. BO allows for more efficient hyperparameter tuning as the optimal configuration of search-such things like a number of hidden neurons, best activation function-comes with minimal computational costs, reducing manual trial-and-error and thus enhancing model generalization. Meanwhile, QEA performs equally well in global optimization in an attempt to avoid local minima in order for convergence towards the optimum. By leveraging quantum principles, QEA improves learning efficiency and adapts to complex problems, making it ideal for refining ANN weights. ANN, in turn, provides robust predictive power by capturing nonlinear relationships in data, scaling efficiently to handle large datasets, and delivering reliable predictions. When combined, these techniques create a powerful modeling framework. While STD ensures quality inputs, BO fine-tunes the architecture, QEA optimizes model weights, and ANN offers accurate, scalable predictions. Together, these methods enhance performance, making them highly suitable for complex applications like predicting Total Dissolved Solids in wastewater treatment plants. The practical implementation of STD, BO, QEA, and ANN can bring huge improvements in the performance of WWTPs. For the implementation of this framework, consistent data collection and preprocessing are required, where TDS data is collected from sensors and decomposed using STD to extract trend and seasonal components.

The prepared data is used to train an ANN model, whose architecture is optimized by BO and the weights further fine-tuned by QEA for better accuracy and efficiency. Once trained, the model can be deployed for real-time TDS prediction, enabling proactive adjustments in treatment processes to improve efficiency and compliance. The integration with SCADA allows automated responses, such as the adjustment of chemical dosages, based on predicted TDS values. The future needs are scaling up the model for larger WWTPs with more complicated operations, integrating IoT and cloud computing to enable real-time processing of data, and expanding the model to predict additional parameters such as BOD, COD, and pH. Hybrid models combining ANN with techniques like Support Vector Machines or Long Short-Term Memory networks improve performance for time-series data. Moreover, the integration of energy consumption data into the predictive framework can be done in order to optimize energy use and promote sustainability. The recommendations for implementation are updating the model regularly for better accuracy, user-friendly interfaces that will help the operators in visualizing predictions and trends, robust error-handling mechanisms to handle issues related to sensor data, and a collaborative effort among researchers, policymakers, and WWTP operators toward standardization and promoting the adoption of the predictive systems.

By addressing these practical needs and future directions, advanced optimization and predictive models can drive significant improvements in wastewater treatment efficiency, environmental compliance, and sustainability.

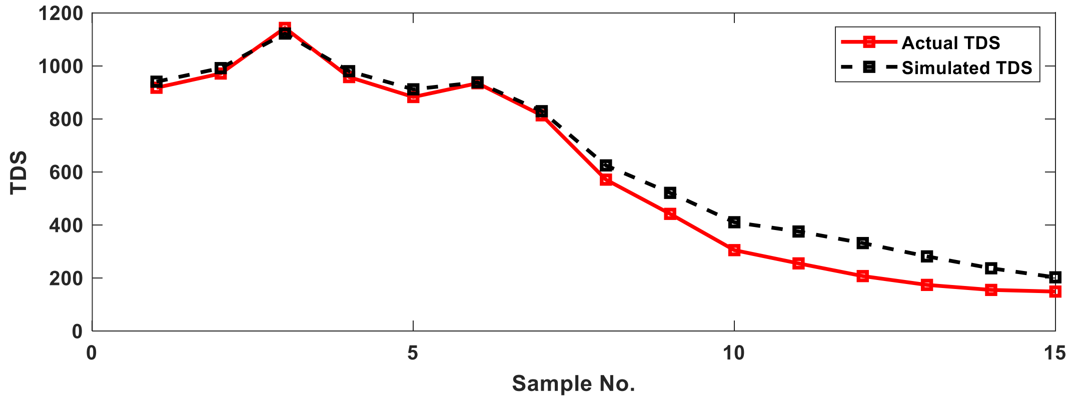

Figure 5 Actual and predicted value of Test sample.

CONCLUSION

The research study shows the efficiency of a double-step optimized ANN framework in predicting TDS occurring in the wastewater treatment plant. The model has achieved high accuracy and reliability by using Seasonal Trend Decomposition on data preprocessing, Bayesian Optimization on hyperparameter tuning, and Quantum Evolutionary Algorithm on weight optimization. An R^2 value of 0.9602 and key metrics with low error values further substantiate the strength of this model in modeling complex data relationships and reducing prediction errors. High-order methodologies represented in the study epitomize the possible incorporation of preprocessing, optimization, and machine learning into the enhancement of predictive modeling regarding the process of wastewater treatment. The practical advantages when the implementation of this framework will be performed in the field are proactive adjustment of the treatment process, integration to SCADA systems for automation, and scalability to bigger operations. Future scope for the work could thus be done by scaling this model for other water quality parameters, considering IoT and cloud computing for real-time processing, and incorporating energy consumption data for further sustainability. This approach highlights a promising pathway toward more efficient and environmentally compliant wastewater treatment systems with considerable advantages in operational efficiency and sustainability.

REFERENCES

Ahmadpour, A., Mirhashemi, S. H., & Panahi, M. (2023). Comparative evaluation of classical and SARIMA-BL time series hybrid models in predicting monthly qualitative parameters of Maroon river. *Applied Water Science*, *13*(3). https://doi.org/10.1007/s13201-023-01876-8

Banadkooki, F. B., Ehteram, M., Panahi, F., Sh. Sammen, S., Othman, F. B., & EL-Shafie, A. (2020). Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *Journal of Hydrology*, *587*. https://doi.org/10.1016/j.jhydrol.2020.124989

Chaibi, M., Benghoulam, E. M., Tarik, L., Berrada, M., & El Hmaidi, A. (2021). Machines Learning Models Based on Feature Selection and Bayesian Optimization for Predicting Daily Global Solar Radiation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3892714

Di, Y., Gao, M., Feng, F., Li, Q., & Zhang, H. (2022). A New Framework for Winter Wheat Yield Prediction Integrating Deep Learning and Bayesian Optimization. *Agronomy*, *12*(12). https://doi.org/10.3390/agronomy12123194

Grande ; Abudu, R., & King, J. P. (2011). Comparison of the Performance of Statistical Models in Forecasting Monthly Total Dissolved Solids in the Rio Grande. *Journal of the American Water Resources Association (JAWRA)*, *48*(1), 10–23. https://doi.org/10.1111

Hafezparast Mavadat, M., & Marabi, S. (2021). How to cite: M. Hafezparast Mavadat, S. Marabi, Prediction of SAR and TDS parameters using LSTM-RNN model: A case study on Aran station Prediction of SAR and TDS parameters using LSTM-RNN model: A case study on Aran station, Iran. In *Journal of Applied Research in Water and Wastewater* (Vol. 8, Issue 2).

Koranga, M., Pant, P., Pant, D., Bhatt, A. K., Pant, R. P., Ram, M., & Kumar, T. (2021). SVM Model to Predict the Water Quality Based on Physicochemical Parameters. *International Journal of Mathematical, Engineering and Management Sciences*, *6*(2), 645–659. https://doi.org/10.33889/IJMEMS.2021.6.2.040

Kouadri, S., Pande, C. B., Panneerselvam, B., Moharir, K. N., & Elbeltagi, A. (2022). Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environmental Science and Pollution Research*, *29*(14), 21067–21091. https://doi.org/10.1007/s11356-021-17084-3

Lotfi, K., Bonakdari, H., Ebtehaj, I., Mjalli, F. S., Zeynoddin, M., Delatolla, R., & Gharabaghi, B. (2019). Predicting wastewater treatment plant quality parameters using a novel hybrid linear-nonlinear methodology. *Journal of Environmental Management*, *240*, 463–474. https://doi.org/10.1016/j.jenvman.2019.03.137

Rong, G., Alu, S., Li, K., Su, Y., Zhang, J., Zhang, Y., & Li, T. (2020). Rainfall induced landslide susceptibility mapping based on bayesian optimized random forest and gradient boosting decision tree models—a case study of shuicheng county, china. *Water (Switzerland)*, *12*(11), 1–22. https://doi.org/10.3390/w12113066

Shad, M., Sharma, Y. D., & Singh, A. (2022). Forecasting of monthly relative humidity in Delhi, India, using SARIMA and ANN models. *Modeling Earth Systems and Environment*, *8*(4), 4843–4851. https://doi.org/10.1007/s40808-022-01385-8

Shuvo, S. P., Sultana, N., Dip, M. M. F., Shibazee, S. P., & Sarker, S. (2024). *Optimizing pH Prediction in Water Treatment Plant Through A Hybrid PSO-SVM Approach With Empirical Mode Decomposition* (pp. 18–31). https://doi.org/10.2991/978-94-6463-478-5_3

Solaimany, M., Aminabad, ----Aminabad, Aminabad, A., Maleki, A., & Hadi, M. (2013). Application of artificial neural network (ANN) for the prediction of water treatment plant influent characteristics Introduction 1. In *J Adv Environ Health Res* (Vol. 1, Issue 2). http://jaehr.muk.ac.ir

Tiwari, S., Babbar, R., & Kaur, G. (2018). Performance Evaluation of Two ANFIS Models for Predicting Water Quality Index of River Satluj (India). *Advances in Civil Engineering*, *2018*. https://doi.org/10.1155/2018/8971079

Unnikrishnan, P., & Jothiprakash, V. (2020). Hybrid SSA-ARIMA-ANN Model for Forecasting Daily Rainfall. *Water Resources Management*, *34*(11), 3609–3623. https://doi.org/10.1007/s11269-020-02638-w

Yadav, P., Chandra, M., Fatima, N., Sarwar, S., Chaudhary, A., Saurabh, K., & Yadav, B. S. (2023). Predicting Influent and Effluent Quality Parameters for a UASB-Based Wastewater Treatment Plant in Asia Covering Data Variations during COVID-19: A Machine Learning Approach. *Water (Switzerland)*, *15*(4). https://doi.org/10.3390/w15040710

Zaman Zad Ghavidel, S., & Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. *Stochastic Environmental Research and Risk Assessment*, *28*(8), 2101–2118. https://doi.org/10.1007/s00477-014-0899-y