

A MULTI-PHASE HYBRID DEEP LEARNING AND OPTIMIZATION FRAMEWORK FOR WATER QUALITY INDEX PREDICTION AND CLASSIFICATION

S. Janani,

Ph.D. Research Scholar,

Department of Computer Science,

Dr. SNS Rajalakshmi College of Arts and Science

(Autonomous),

Coimbatore, Tamil Nadu, India.

Dr. S.S. Suganya,

Associate Professor,

Department of Computer Science,

Dr. SNS Rajalakshmi College of Arts and Science

(Autonomous),

Coimbatore, Tamil Nadu, India.

Abstract: This study presents a multi-stage hybrid deep learning and optimization approach to predict and classify the Water Quality Index (WQI). The goal was to identify a single process to create a model pipeline that will allow for the use of both feature selection and the use of deep learning combined with metaheuristics in the assessment of water quality. The objective of this research was to develop a 3-stage predictive model using Relief-GWO feature selection, optimized using ANN/DANN with ISLO and DNN using PSO or CCPSO or a Chaos-Driven Bat algorithm. Eighteen physical, chemical, and biological characteristics of water from 3,673 sampling sites were studied. The results obtained by the baseline models developed in Phase 1 showed that SVM was the best classifier at a rate of 91.29%. Phase 2 hybrid ANN-Maxout obtained an overall classification accuracy of 93.74%, whereas DANN-ISLO achieved a better regression R^2 value of 0.9806 than all other models. In phase 3, the DNN-CCPSO model had the best performance for regression as measured by the R^2 value of 0.9813, and DNN-ChaosBat was the best in terms of F1-Macro at 0.7269. Overall, this study demonstrated that a proposed multi-hybrid approach will provide superior results to many individual models used alone, providing significant improvements over time throughout each phase. The work presented here demonstrates how the use of artificial intelligence can be applied to support automatic water quality monitoring, detect contamination early on, and support government decision-making relative to environmental policies.

Keywords : Deep learning, Water Quality Index, Relief-GWO feature selection, Metaheuristic optimization, Hybrid neural networks.

I. INTRODUCTION

Water quality assessments are crucial to protect both public health and sustainable development across the world. Access to safe drinking water continues to be a major global challenge, with millions of people dying each year due to water-borne disease. The Water Quality Index (WQI) is used to develop an overall index value that integrates multiple physico-chemical and biological parameter values to provide a standard method of assessing the quality of water for all uses. Water Quality Index classifications consist of five categories: Excellent (0 – 50), Good (51 – 100), Fair (101 – 150), Poor (151 – 200), and Very Poor (>201). The traditional water quality index was developed using laboratory methods for assessing water quality. These methods were very time consuming and cost prohibitive in addition to having no predictive capabilities. As a result, there is a need to develop automated systems for real-time monitoring of water quality that can be used to monitor growing amounts of industrial, agricultural, and household pollutant discharges [6].

Traditional machine learning methods (e.g. K-nearest neighbor, decision trees, support vector machines) have shown to be inadequate when predicting water quality using complex high-dimensional datasets. In contrast, deep learning approaches (i.e., artificial neural networks, deep neural networks) are capable of discovering both linear and nonlinear associations within large-scale multi-variable environmental datasets. Dimension reduction is necessary to increase the

efficiency of these machine learning algorithms; feature selection can aid this process by determining which variables are the most informative relative to each other [7], [8].

Intelligent hyperparameter tuning methods, such as Improved Sea Lion Optimization (ISLO) and Constriction Coefficient Particle Swarm Optimization (CCPSO), along with the Chaos-Driven Bat Algorithm, have been used to improve the performance of deep learning by optimizing the hyperparameters. There is currently no work that has combined both phase-by-phase feature selection and phase-by-phase deep learning optimization to predict the Water Quality Index (WQI). This study develops an optimized three-phase hybrid approach incorporating feature engineering, deep learning, and bio-inspired optimization. The goal of this research is to provide an intelligent, automated method to bridge the gap between the traditional methods of assessing water quality and providing an accurate predictive model of the future state of water quality [1], [2], [3].

II. REVIEW OF LITERATURE AND RESEARCH FRAMEWORK

A. Related Works

Industrial development, city growth, and large-scale agriculture have greatly harmed the environment through pollution of both surface waters and underground water sources. The traditional methods of assessing water quality by using laboratory testing are a lengthy and costly process that has limitations in terms of being able to assess the full scope of

an area's water quality at one time. Because of these factors, AI/ML models can be used to predict and classify WQIs. This allows for timely, cost-effective assessments of water quality[6], [7].

Early research into the use of artificial intelligence (AI)-based systems to predict water quality primarily focused on using a single model type (such as Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Feed Forward Neural Networks (FFNN), Radial Basis Function (RBF) networks, or K-Nearest Neighbors (KNN)) to identify relationships among various water quality variables. As these early models were data-driven, they generally captured the complex interactions of many water quality factors well and performed better than traditional statistical methodologies [6][7]. Research has increasingly transitioned from singular models toward ensembles, which are multiple models trained together, and toward deep learning techniques. Ensemble models such as Random Forest, Gradient Boosting, XGBoost, Support Vector Machine (SVM), and multilayer perceptron (MLP) have exhibited high performance capabilities, with some achieving accuracy rates over 95% [3][4][9]. Additionally, the application of convolutional neural network (CNN) and long short-term memory (LSTM) deep learning architectures for predicting spatiotemporal patterns of water quality characteristics has also been explored [3][4][9].

The effect that the parameters that are selected (feature selection) have on the performance of a model has been recognized as one of the key factors. A number of parameter selection methods, for example, Pearson correlation, PCA, and Relief-based methods, have been applied in order to determine which input parameters have the largest influence on the Water Quality Index (WQI). Furthermore, Explainable AI (XAI) tools (such as SHAP) have been employed to provide an explanation of why certain model predictions were made. Although there is still much room for improvement regarding the development of frameworks that combine feature selection techniques with advanced optimization techniques [3][5]. Optimization of hyperparameters through metaheuristics (for instance, PSO, GA, and IPSO), while effective, has also shown that newer optimizations such as the Chaos Driven Bat algorithm and CCPSO have not yet had extensive use in predicting the Water Quality Index, therefore representing a large area of research that needs to be addressed [2][4]. VidhyaJanani et al. [10] introduced an artificial intelligence-based (AI) system that predicts/assesses water quality; FFNN outperformed KNN in all instances by reaching 100% accuracy. Thakur et al. [11] used Random Forest and XGBoost; both models reached 80% vs. 78% accuracy. Akasam et al. [12] found that when data was balanced, the random forest model reached 89% accuracy. The most accurate model to date is represented by Venkataraman et al. [17]; they reported a 98.01% accuracy via the use of an optimization-driven deep learning model. Prasad et al. [18] presented a hybrid model consisting of a quadratic discriminant analysis (QDA), along with bagging, which yielded as much as 100% accuracy. Sutanto et al. [19] showed stacking gradient boosting was able to achieve 96% accuracy. Overall, these studies support the idea of utilizing hybrid or optimized approaches; however, they indicate the lack of a single unified multi-phased framework.

B. Problem Definition

Traditional water quality monitoring uses lab-based testing (slow, expensive and reactive) and conventional machine learning methods use all parameters in their input (no effective feature extraction; noise and model overfitting). All single-ML model based systems are incapable of fully capturing the variability present in water quality measurements for a wide range of water quality categories. There have been few studies conducted where the optimal values of hyper-parameters for Deep Learning were systematically determined with bio-inspired algorithms. There has been no study reported which establishes a multi-phased framework to improve predictive capability of an ML system in stages from a baseline ML system to a highly optimized DL system. [3], [4] There is a major knowledge gap in the literature regarding the lack of inclusion of Relief-GWO feature selection into existing ANNs, DANNs, and DNN-CCPSOs; a fully integrated, computationally efficient method needs to be developed in order to provide accurate predictions of WQIs, and classify them on a large scale in order to create an automated monitoring system for contaminants in water resources, which could be used by government agencies for environmental policy development and for detecting water contamination.

C. Objectives of the Study

1. To develop a more accurate and reliable WQI prediction and classification methodology by utilizing hybrid machine learning, deep learning, and advanced optimization methods.
2. To determine important features and to optimize model performance by means of Relief-GWO feature selection and metaheuristics like ISLO, PSO, CCPSO, and the chaos-driven bat algorithm.

III. DATASET DESCRIPTION

A total of 3,673 water sample records were compiled, each representing one or more of eighteen different physico-chemical and biological parameters measured at a variety of water locations. These include pH, dissolved oxygen, biochemical demand for oxygen (BOD), total dissolved solids (TDS), conductivity, turbidity, sulfate, nitrate, water hardness, chloramines, organic carbon, trihalomethanes (THMs), total coliforms, fecal coliforms, fecal streptococci, water temperature, salinity, and Secchi depth. Using WHO guidelines, BIS standards (IS 10500), and CPCB Class C standards, the Water Quality Index (WQI) was calculated by means of the weighted arithmetic index method. As may be seen in Table 1, the frequency distribution of the WQI classes is Preprocessing of data included the treatment of missing values, IQR-based outlier capping, and min-max normalization to give all variables a common scale (i.e., range = 0 – 1). Stratified random sampling was used to divide the sample into an 80% training set (n=2938) and a 20% test set (n=735). Inter-parameter correlation plots are shown in Fig. 2. Boxplot outlier analysis can be found in Fig. 3. In addition, an illustration of how the WQI data is distributed within its respective class limits can be visualized in Fig. 4.



Fig. 1. Parameter Distributions of Water Quality Dataset

TABLE 1. WQI CLASSIFICATION DISTRIBUTION

WQI Range	Class	Count	Percentage
0–50	Excellent	15	0.41%
50–100	Good	831	22.63%
100–150	Fair	1944	52.93%
150–200	Poor	763	20.77%
>200	Very Poor	120	3.27%

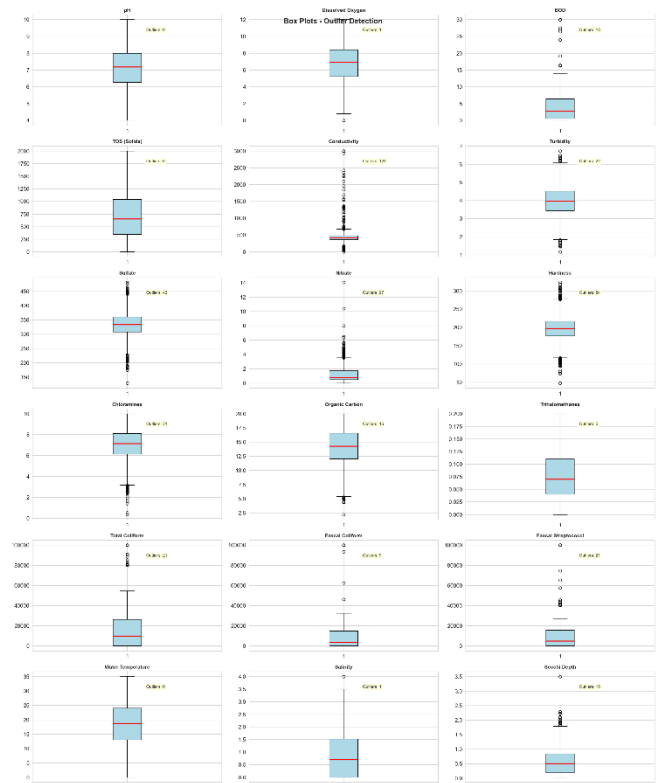


Fig. 3. Boxplots Showing Outlier Distribution

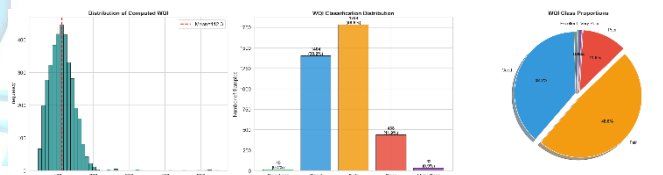


Fig. 4. WQI Distribution and Classification

IV.METHODOLOGY

The proposed multi-phase hybrid deep learning and optimization framework for Water Quality Index (WQI) prediction and classification is structured into three systematic phases, each building upon the previous to progressively enhance model performance.

A. Phase 1: Baseline Machine Learning Benchmarking

Phase 1 of this research provides a baseline machine learning benchmarking with the use of a Relief algorithm (iterations = 1000) that ranks all 18 potential input variables by degree of association with the Water Quality Index. This ranking is shown in Fig. 5. Following the application of Relief, the Grey Wolf Optimizer (number of wolves = 20; number of iterations = 50) was utilized to determine the optimal set of input features out of the original 18. In doing so, the 18 original features were reduced to 9. The convergence characteristics of GWO during this process are presented in Fig. 6. The identified input features were next utilized as inputs to train four basic machine learning models. These included K-Nearest Neighbor (KNN), Decision Trees, Random Forests, and Support Vector Machines (SVM). Both classification (a five-class water quality index) and regression (a water quality index predicted continuously) metrics were examined.

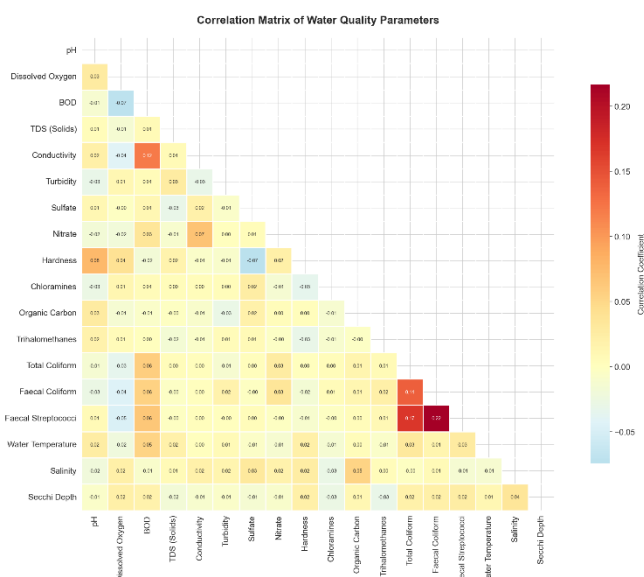


Fig. 2. Correlation Heatmap of Water Quality Parameters

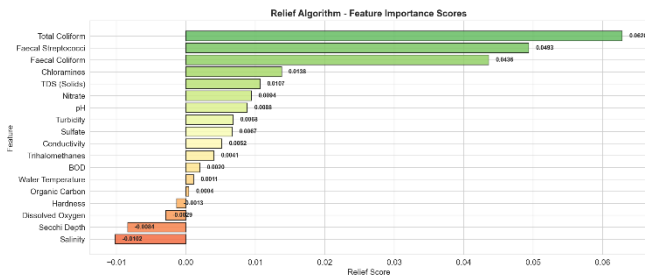


Fig. 5. Relief Feature Importance Scores

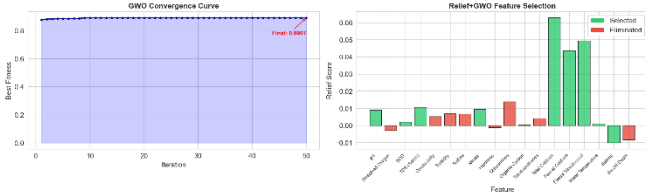


Fig. 6. GWO Convergence for Feature Selection

B. Phase 2: Hybrid Deep Learning with ISLO Optimization

Phase 2 introduces different artificial neural network (ANN) variations that utilized the Tanh, ReLU, and Maxout activation functions (see Table 3) plus a "dynamic" ANN (DANN), which was optimized by improved sea lion optimization (ISLO). ISLO dynamically adjusted hyperparameters such as learning rate, number of hidden layers, number of neurons in each hidden layer, and batch size. Furthermore, the DANN-ISLO model used dynamic weights adjustment and a population-based search algorithm for finding the best possible configuration of the neural networks. Training convergence curves are provided in Fig. 9.

C. Phase 3: Metaheuristic Optimized Deep Neural Network

Phase 3 utilized an advanced set of metaheuristics to optimize a deep neural network. Three metaheuristic algorithms were used, including Particle Swarm Optimization (PSO), Cooperative Coevolutionary PSO (CCPSO), and Chaos-Driven Bat Algorithm. CCPSO was developed to break down the large-dimensional hyperparameter space into several smaller-dimensional subpopulations that would improve the efficiency of the exploration. Chaotic sequences were also added to a chaos-driven bat algorithm to create additional diversity during the search. All of the abovementioned deep neural networks were then tested on their ability to perform well as both classifiers (Table 4) and regressors (Table 6).

V. RESULTS AND DISCUSSION

The proposed study presents a multi-phase predictive framework combining feature selection, hybrid deep learning, and metaheuristic optimization for WQI prediction and classification. Each phase targets progressive improvement in model performance. This section discusses the outcomes of all three modeling phases, followed by a comprehensive cross-phase comparative evaluation.

A. Phase 1: Baseline ML Results

The Relief-GWO method resulted in a reduction from 18 to nine optimum characteristics that improved model efficiency while maintaining accuracy, which is crucial for enhancing the performance of machine learning algorithms in classification tasks. Classification results for each of the four baseline classifier algorithms (Table 2) indicated that the Support Vector Machine (SVM), with an accuracy of 91.29%, had the

highest accuracy of all four baseline classifiers and an F1-Macro of .6650. It was the random forest algorithm, which produced the best regression results with a coefficient of determination (R^2) of .9649 and root mean squared error (RMSE) of 6.308. K-Nearest Neighbors and Decision Trees demonstrated significantly poorer results than Random Forest, indicating a greater necessity for the use of more complex predictive modeling techniques. Fig. 7 contains a graphical representation of phase one model evaluation, and Fig. 8 contains a graphical representation of the confusion matrix associated with each of the models evaluated during this phase.

TABLE 2. PHASE 1: BASELINE ML CLASSIFICATION RESULTS

Model	Accuracy	Precision	Recall	F1-Score	F1-Macro
KNN	0.8639	0.8601	0.8639	0.8598	0.5845
Decision Tree	0.8367	0.8276	0.8367	0.8320	0.6427
Random Forest	0.8993	0.8973	0.8993	0.8971	0.6968
SVM	0.9129	0.9095	0.9129	0.9112	0.6650

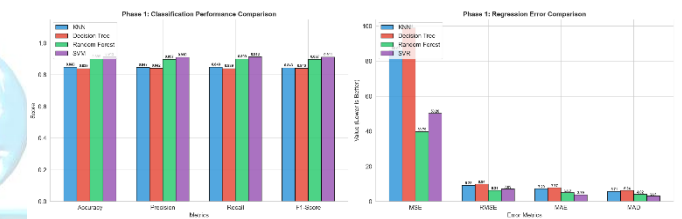


Fig. 7. Phase 1 – Baseline ML Comparison

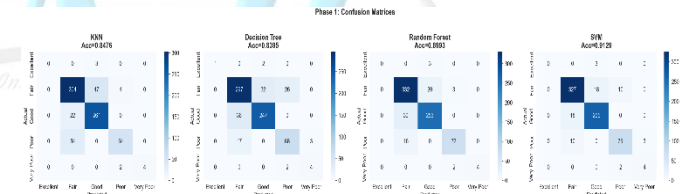


Fig. 8. Phase 1 – Confusion Matrices

B. Phase 2: Hybrid Deep Learning Results

The DANN-ISLO model and all ANN variants displayed an increased performance compared to baseline methods (as reported in Table 3). Of these ANN models, the ANN-Maxout variant had the largest increase in classification accuracy, with an increase of 2.45% from phase one SVM to 93.74%. The ANN-Maxout also yielded the best F1 macro score of .7083. The DANN-ISLO model showed excellent regression performance as indicated by $R^2=0.9806$ and $RMSE=4.692$. ISLO optimization was able to optimize the hyperparameters for DANN, which resulted in faster convergent results than those obtained using traditional gradient descent and better overall generalization ability.

TABLE 3. PHASE 2: HYBRID DEEP LEARNING CLASSIFICATION RESULTS

Model	Accuracy	Precision	Recall	F1-Score	F1-Macro
ANN-Tanh	0.9197	0.9159	0.9197	0.9178	0.6822
ANN-ReLU	0.9293	0.9275	0.9293	0.9280	0.6873
ANN-Maxout	0.9374	0.9343	0.9374	0.9354	0.7083

DANN-ISLO	0.9320	0.9283	0.9320	0.9300	0.6902
-----------	--------	--------	--------	--------	--------

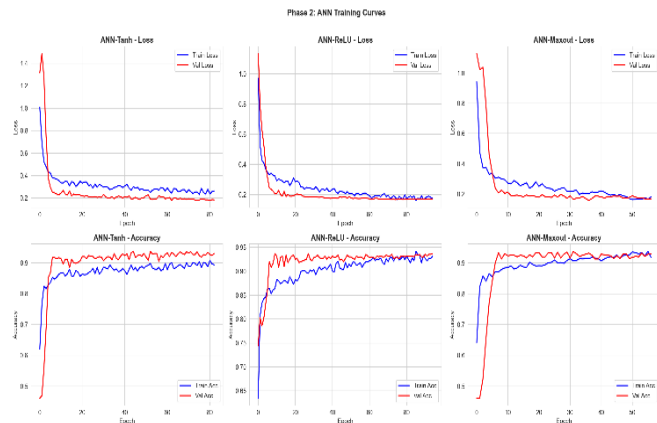


Fig. 9. ANN Training Curves

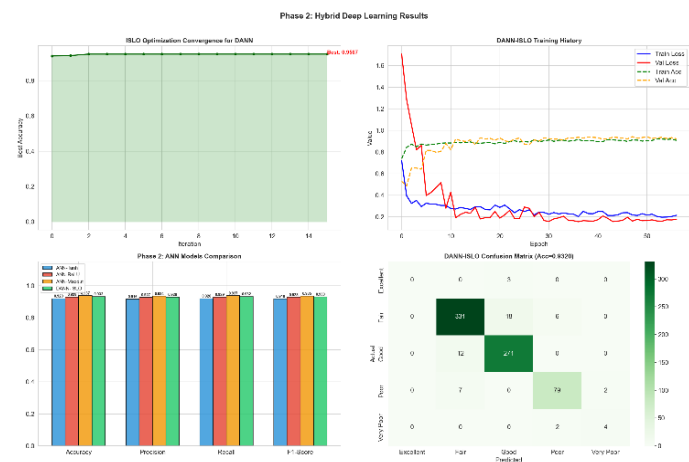


Fig. 10. Phase 2 – Hybrid Deep Learning Results

C. Phase 3: Optimized DNN Results

The results for the metaheuristics-based DNN models that were classified similarly to the decision tree and better than the decision tree in terms of regression are shown in Table 4. DNN-ChaosBat (the best F1-macrovalue, 0.7269) was better than other models at balancing accuracy for each class of WQI because it had the best macro F1-score. DNN-CCPSO (best $R^2=0.9813$; best RMSE=4.611) produced the best regression results by achieving an MSE improvement of 78% over those obtained from the decision tree model. In addition, the Phase 3 optimized DNN performance is illustrated graphically in Fig. 11.

TABLE 4. PHASE 3: OPTIMIZED DNN CLASSIFICATION RESULTS

Model	Accuracy	Precision	Recall	F1-Score	F1-Macro
DNN-PSO	0.9279	0.9246	0.9279	0.9259	0.7126
DNN-CCPSO	0.9184	0.9149	0.9184	0.9165	0.7032
DNN-ChaosBat	0.9320	0.9291	0.9320	0.9299	0.7269

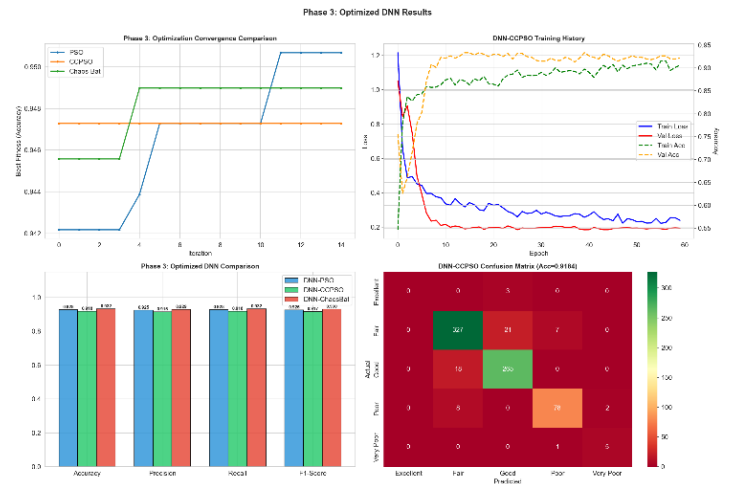


Fig. 11. Phase 3 – Optimized DNN Results

D. Cross-Phase Comparison and Analysis

The comparison of the best classification models for each of the three phases can be seen in Table 5, while a comparison of the regression results is shown in Table 6. Accuracy in classification increased from 91.29% (Phase 1 SVM) to 93.74% (Phase 2 ANN-Maxout), and macro-F1 score also increased from 0.6968 (Phase 1 Random Forest) to 0.7269 (Phase 3 DNN-ChaosBat). All of the regression measures were consistently better than their corresponding values in previous phases. Specifically, R^2 was higher in Phase 3 (0.9813) than it was in Phase 1 (0.9649) using random forest as the predictor model, and MSE was lower in Phase 3 (21.26) than it was in Phase 1 (98.84) when decision trees were used. Fig. 12 provides an overview of the proposed model, while a radar chart comparing each of the phase's models on key evaluation metrics is provided in Fig. 13.

TABLE 5. CROSS-PHASE CLASSIFICATION COMPARISON (TOP MODELS)

Model	Accuracy	Precision	Recall	F1-Score	F1-Macro
P2: ANN-Maxout	0.9374	0.9343	0.9374	0.9354	0.7083
P2: DANN-ISLO	0.9320	0.9283	0.9320	0.9300	0.6902
P3: DNN-ChaosBat	0.9320	0.9291	0.9320	0.9299	0.7269
P3: DNN-PSO	0.9279	0.9246	0.9279	0.9259	0.7126
P1: SVM	0.9129	0.9095	0.9129	0.9112	0.6650
P1: Random Forest	0.8993	0.8973	0.8993	0.8971	0.6968

TABLE 6. CROSS-PHASE REGRESSION COMPARISON

Model	MSE	RMSE	MAE	R^2	MAPE
DNN-CCPSO	21.26	4.611	3.419	0.9813	3.26
DANN-ISLO	22.01	4.692	3.580	0.9806	3.43
ANN	27.13	5.209	3.941	0.9761	3.91
Random Forest	39.79	6.308	4.997	0.9649	4.62
SVR	50.20	7.085	3.789	0.9557	3.53
KNN	86.09	9.278	7.197	0.9241	6.40
Decision Tree	98.84	9.942	7.773	0.9129	7.19

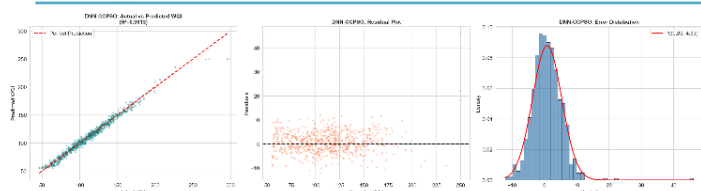


Fig. 12. Proposed Model Analysis

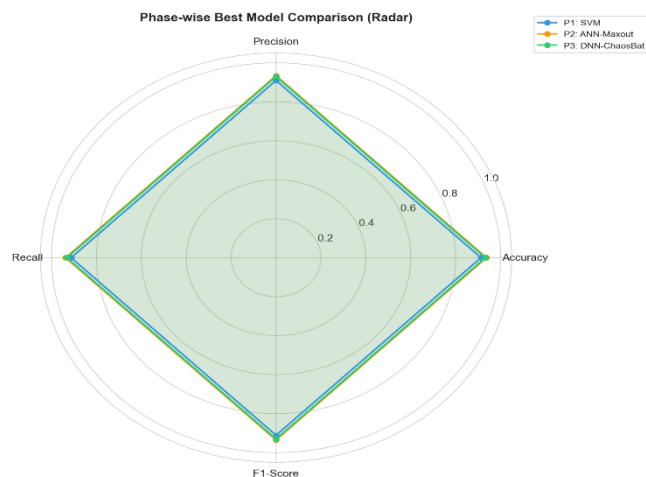


Fig. 13. Radar Comparison of All Models

VI.CONCLUSION

A multi-phase hybrid deep learning and optimization method for predicting and classifying water quality index (WQI) has been accomplished utilizing 3,673 water sampling data points, from which 18 physicochemical and biological parameters were used in the analysis. In phase 1 the Relief-GWO method utilized to perform feature selection on these 18 physico-chemical/biological parameters yielded 9 of the most important parameters (Fig. 5, Fig. 6). Thus, eliminating some of the less important parameters in the analysis resulted in improved efficiency of the model while preserving its overall accuracy. Baseline models in Phase 1 (Table 2) identified Support Vector Machine as the highest-performing traditional classifier at an accuracy rate of 91.29% and Random Forest as the highest-performing regressor at an R^2 value of 0.9649.

In phase 2, the hybrid deep learning models (Table 3) showed an impressive development in comparison to the previous phase. The Maxout Neural Network for Classification (ANN-Maxout) had the largest percentage of correct classifications at 93.74%. In addition, Deep Learning-based ANNs for Regression (DANN-ISLO) performed a high degree of correlation in regression ($R^2=0.9806$). In phase 3, the Metaheuristic Optimized DNN Models (Table 4) were able to perform the highest R^2 values for both regression and the lowest RMSE in regression by utilizing CCPSO as the optimization method on their DNN model ($R^2=0.9813$, $RMSE=4.611$). In addition, the Chaos Bat Optimization Method was utilized as the optimization method on their DNN model to achieve the highest F1 Macro score for classification ($F1\text{-Macro}=0.7269$). All three cross-phase comparisons (Tables 5 & 6) and model analyses (Fig. 12 & Fig. 13) demonstrate that there is progression from one phase to another.

A systematic evaluation of this progressive three-stage framework has shown it provides a practical, scalable method for optimizing the automated monitoring of water quality. This

framework can be used to improve the sustainability of water resources; support early detection of contaminants in water sources; and provide valuable information to those making environmental policies.

VII.FUTURE SCOPE

Future research will focus on validation of the proposed framework with the use of multiple regional and longitudinal datasets so that the results are generalizable. In addition, the incorporation of explainable AI (XAI) methods such as SHAP and LIME will improve understanding and clarity of the decision-making process. Future developments could also include the application of transformer models or other attention-based models to perform temporal analysis on time series water quality data. The framework can also be extended to develop an Internet of Things (IoT)-based real-time monitoring system for continual evaluation. Finally, class imbalance techniques such as SMOTE or cost-sensitive training can be utilized to increase accuracy of predictions related to minority classes.

VIII.REFERENCES

- [1] M. Ehteram, A. N. Ahmed, M. Sherif, and A. El-Shafie, "An advanced deep learning model for predicting water quality index," *Ecological Indicators*, vol. 160, p. 111806, 2024.
- [2] A. H. Alharbi et al., "Hybrid deep learning optimization for smart agriculture: Dipper throated optimization and polar rose search applied to water quality prediction," *PLOS One*, vol. 20, no. 7, p. e0327230, 2025.
- [3] S. Talukdar et al., "Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms," *Journal of Cleaner Production*, vol. 406, p. 136885, 2023.
- [4] M. Y. Shams et al., "Water quality prediction using machine learning models based on grid search method," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35307–35334, 2023.
- [5] A. Elmotawakkil et al., "Advanced machine learning models for robust prediction of water quality index and classification," *Journal of Hydroinformatics*, vol. 27, no. 2, pp. 299–319, 2025.
- [6] M. H. Al-Adhaileh and F. W. Alsaade, "Modelling and prediction of water quality by using artificial intelligence," *Sustainability*, vol. 13, no. 8, p. 4259, 2021.
- [7] A. N. Ahmed et al., "Machine learning methods for better water quality prediction," *Journal of Hydrology*, vol. 578, p. 124084, 2019.
- [8] D. T. Bui et al., "Improving prediction of water quality indices using novel hybrid machine-learning algorithms," *Science of The Total Environment*, vol. 721, p. 137612, 2020.
- [9] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 2, pp. 415–433, 2020.
- [10] S. Janani et al., "Artificial Intelligence in Water Quality Monitoring for Sustainable Resource Management," *International Journal of Environmental Sciences*, vol. 11, no. 14, pp. 1528–1534, 2025.

- [11] S. N. Thakur et al., "Analysis of Water Quality Using Machine Learning Techniques," Springer Science and Business Media LLC, 2025, doi: 10.21203/rs.3.rs-6452922/v1.
- [12] S. Akasam et al., "Optimising Water Quality Assessment: A Comparative Study of Machine Learning Algorithms for Imbalanced Potability Datasets," Journal of Mines, Metals and Fuels, vol. 73, no. 6, pp. 1765–1770, 2025.
- [13] S. V. Patil et al., "Water Quality Analysis and Prediction Using Machine Learning," Journal of Information Systems Engineering and Management, vol. 10, no. 39, pp. 1069–1073, 2025.
- [14] S. Veshala and K. K. Mam, "Machine Learning-Based Prediction of Drinking Water Quality Using Mineral and pH Data," International Journal of Research Publication and Reviews, vol. 6, no. 8, pp. 6077–6082, Aug. 2025.
- [15] R. M. Frincu, "Artificial intelligence in water quality monitoring: A review of water quality assessment applications," Water Quality Research Journal, vol. 60, no. 1, pp. 164–176, 2025.
- [16] V. Sangwan and R. Bhardwaj, "Machine learning framework for predicting water quality classification," Water Practice & Technology, vol. 19, no. 11, pp. 4499–4521, 2024.
- [17] M. Venkataraman et al., "Water quality prediction and classification using attention based deep differential recurflownet with logistic giant armadillo optimization," Global NEST Journal, vol. 27, no. 1, pp. 1–13, 2024.
- [18] D. V. V. Prasad et al., "Water quality prediction using statistical, ensemble and hybrid models," Global NEST Journal, vol. 26, no. 2, pp. 1–9, 2024.
- [19] T. Sutanto et al., "Comparison of Logistic Regression, Random Forest, SVM, KNN Algorithm for Water Quality Classification Based on Contaminant Parameters," Journal of Data Science, vol. 2024, no. 48, pp. 1–7, 2024.

