



# OPTIMIZATION OF NEURAL NETWORK CLASSIFIERS BY LEVERAGING THE SEQUENTIAL FEATURE ENGINEERING FOR ROBUST WATER QUALITY PREDICTION SYSTEM

Maheswara Rao V V R<sup>1</sup>  
Silpa N  
Kranthi Addanki  
Shiva Shankar Reddy  
Ramachandra Rao Kurada  
Pachipala Yellamma

Received 09.10.2023.

Revised 04.12.2023.

Accepted 10.01.2024.

Keywords:

*Water Quality Prediction, Neural Networks, Feature Engineering, Data analytics, Predictive Modelling.*

ABSTRACT

*Rapid population growth increases water demand, intensifying extraction from wells and rivers. The Water Quality Index (WQI) assesses water suitability for drinking based on multiple parameters. Accurate assessment of pollution in water is imperative for effective management of water quality. The present research on the Neural Network-based Robust Water Quality Prediction System (NN-RWQPS) exploits the capabilities of neural networks and advances in feature engineering, positioning it at the forefront of WQI. Venturing into the new world of predictive modelling armed with four different neural network classifiers: Wide, Bilayer, Trilayer, and an Optimized Neural Network. Further the study harness the power of feature selection, deploying four distinct methods. A champion feature selection method is scientifically validated for each neural network, and then the neural networks are fine-tuned by training them across a range of feature dimensions, unveiling an empirically supported set of optimal features. Study advances water quality prediction using neural networks and feature engineering.*



© 2025 Published by Faculty of Engineerin

## 1. INTRODUCTION

In the ever-evolving landscape of environmental science and data analytics, the research is poised at the forefront, where the pursuit of optimizing water quality prediction and unlocking the potential of neural networks and sequential feature engineering takes center

stage. Water, the elixir of life, presents an intricate puzzle encapsulated within a vast dataset brimming with its subtleties, a puzzle that continues to challenge even the most adept scientists and researchers.

A neural-network classifier emulates the human brain's decision-making process. The system as depicted in

<sup>1</sup> Corresponding author: Maheswara Rao V V R  
Email: [mahesh\\_vvr@yahoo.com](mailto:mahesh_vvr@yahoo.com)

Figure 1 consists of three essential layers: the input layer receives data, hidden layers process information, and the output layer produces results. The input layer receives and preprocesses data, transmitting it to the hidden layers. Hidden layers, consisting of interconnected artificial neurons, perform complex computations through weighted connections. Finally, the output layer produces predictions, often in the form of class probabilities, which represent the category or label. These networks are trained on labeled datasets, adjusting their internal parameters through backpropagation to minimize prediction errors. They are widely used for diverse classification tasks, benefiting from their capacity to handle intricate and non-linear patterns.

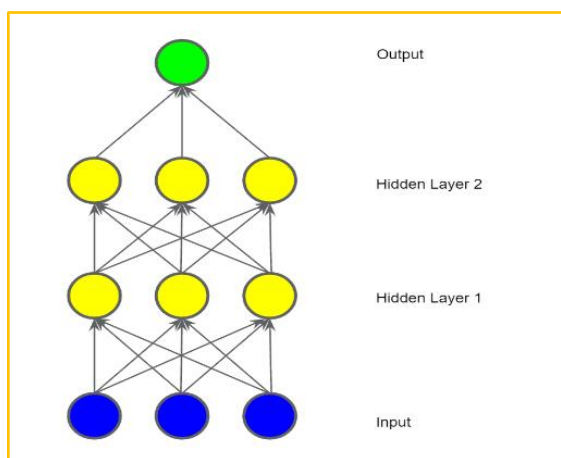


Figure 1. Basic Neural Network Architecture

In the exploration of this cutting-edge domain of predictive modeling, our research has armed us with a formidable array of four unique neural network classifiers: the Wide, Bilayer, Trilayer, and the Optimized Neural Network. These distinct neural networks serve as our companions, adeptly navigating the intricate maze of water quality prediction. Each neural network brings its own strengths and characteristics to the table, allowing us to unlock a more comprehensive understanding of this dynamic field. Through their guidance and proficiency, we embark on a journey towards enhancing our predictive capabilities and shedding light on the complexities inherent in the realm of water quality forecasting.

However, this voyage doesn't end with the selection of neural networks. To navigate this enigmatic terrain effectively, the research taps into the potent art of feature selection, utilizing four meticulous and rigorous methods: MRMR, Chi-square, Anova, and Kruskal. Each of these methods serves as a finely honed instrument, akin to precision tools in the hands of a water quality analyst.

A feature selection method, rigorously and scientifically validated for each neural network, stands as a testament to our commitment to precision and accuracy. With this

champion method in hand, we embark on the fine-tuning of our neural networks. Through meticulous training across a spectrum of feature dimensions, a symphony of performance optimization unfolds, revealing a set of optimal features empirically supported by our research.

The findings of this study represent a significant leap forward, a cornerstone in our understanding and practical application of neural networks and sequential feature engineering within the realm of water quality prediction. In the spirit of scientific collaboration, this paper serves as an invaluable guide, offering the broader scientific community a roadmap to harness these advanced techniques for the enhancement of environmental monitoring and management, ushering in a brighter future for our planet.

This scholarly work follows a structured framework, consisting of five distinct sections. Section 2 initiates with an in-depth review of existing research in the domain of water quality estimation through neural networks. Section 3 delves into the proposed system, providing comprehensive insights into the intricacies of data preprocessing, feature engineering techniques, and the specific neural network classifier models utilized. Section 4 shifts the focus towards the experimental outcomes. In the concluding Section 5, the authors consolidate their findings, providing conclusive insights and observations drawn from the entire research endeavor, offering a holistic view of their contributions to the field.

## 2. RELATED WORK

The authors advanced every stage of the present study by thoroughly reviewing the literature from 2008 to 2023, improving the field of water quality evaluation with an emphasis on neural-network classifiers.

The authors of this literature (Mengyuan et al., 2022; Zhu et al., 2022; Paptsov and Popkova, 2022) stress the significance of thorough data analysis as the basis for precise water quality estimates that are related to health.

To improve the accuracy of WQI, a number of researchers have recently investigated novel machine learning-based techniques (Aldhyani et al., 2020; Khoi, et al., 2022; Bonthu et al., 2023; Flores et al., 2023; Goodarzi et al., 2023).

With the object of significantly enlightening the high-accuracy of WQI, the authors (Patel et al., 2022; Kalaivanan et al., 2022; Kurra, et al., 2022; Maheswara Rao, et al., 2022; Uddin et al., 2022) have presented a number of methods that make use of ML Based Classification techniques. According to research publications (Rao et al., 2011; Silpa & Rao, 2021), the data investigation stage in ML based for WQI is a crucial phase.

To improve prediction accuracy, the integration of sophisticated ML techniques has also been investigated (Silpa & Rao, 2022; Rao et al., 2023; Reddy et al., 2022; Rao et al., 2023). Recent years have seen a notable increase in interest in the field of ML for water quality investigation (Maheswara Rao et al., 2022; Rao et al., 2022; Silpa et al., 2023) in the feature engineering process (Fei et al., 2023; Gong et al., 2022; Rao et al., 2023).

### 3. PROPOSED METHODOLOGY

The current study on the Neural Network-based Robust Water Quality Prediction System (NN-RWQPS) leverages the potential of neural networks and sequential feature engineering, establishing it as a pioneering approach in water quality analysis. Embarking on predictive modelling, it employs four

diverse neural network classifiers: Wide Neural Network Classifier (WNNC), Bilayered Neural Network Classifier (BNNC), Trilayered Neural Network Classifier (TNNC), and an Optimized Neural Network Classifier (ONNC). The study also harnesses the power of feature selection, utilizing four robust methods: MRMR, Chi-Square, ANOVA, and Kruskal, akin to finely tuned research instruments. Each neural network is rigorously assessed and optimized, validated through a scientific algorithm, revealing an empirically supported set of optimal features. As illustrated in Figure 2, it covers the entire process from gathering and refining data to establishing a robust machine-learning framework for evaluating water quality. Notably, this approach emphasizes the importance of combination of Network classifier and feature selection to optimize achieves performance.

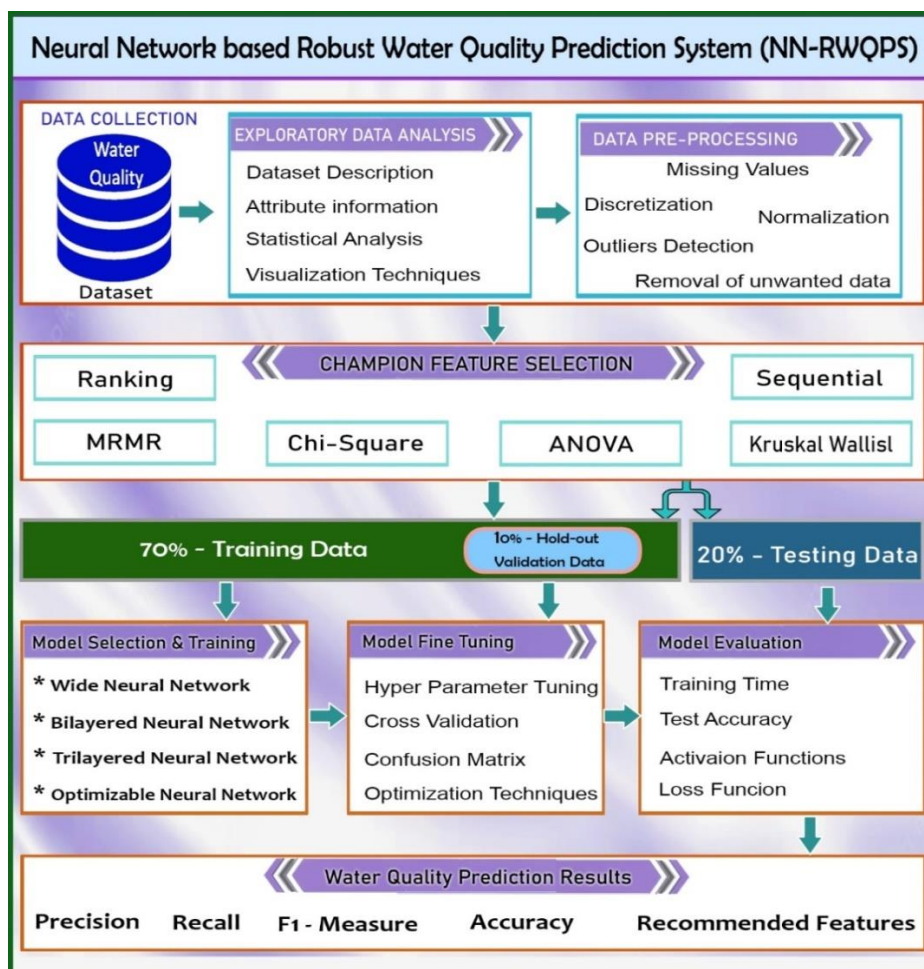


Figure 2. Architecture of proposed NN-RWQPS

The process of NN-RWQPS begins with collecting essential water quality data, the data is rigorously pre-processed as presented in Algorithm 1. During the model selection phase, various neural network classifiers are explored to determine the most suitable approach for the task. The models are subsequently trained and evaluated, allowing for the assessment of their performance using a

variety of evaluation metrics. As the pursuit of optimization continues, the models are refined, aiming to identify the best combination of features and network classifier. By comparing the performance of different neural network models using a range of evaluation metrics, valuable insights are gained into the most effective approach for water quality prediction.

---

**Algorithm 1: Proposed NN-RWQPS**

**Inputs:** Water Quality Dataset; Neural Network Classifiers: WNNC, BNNC, TNNC and ONNC; Feature Selection Methods; Evaluation Metrics: F1-Score, Precision, Recall, Testing Accuracy

**Output:** Optimal Feature Selection Method and Feature Set and for each Neural Network Classifier

---

**01. NN-RWQPS: Initialization**

- a. Initialize an empty set to store optimal feature sets for each NN Classifier
- b. Initialize a dictionary to store performance metrics for each model.

**02. NN-RWQPS: Data-Understanding**

- a. Load & preprocess the water quality dataset.
- b. Perform data cleaning, normalization, and splitting into training and testing sets.

**03. NN-RWQPS: Model Selection**

- a. Select the neural network Classifiers: WNNC, BNNC, TNNC and ONNC.
- b. Initialize an empty list to store performance metrics for each model.

**04. NN-RWQPS: Champion Feature Selection**

- a. For each feature-selection method
- b. Select features using the current feature selection method.

**05. NN-RWQPS: Neural Network Training**

- a. For each neural network model using Algorithm 2, 3, 4 and 5
  - ✓ Initialize the model with the selected architecture and hyperparameters.
  - ✓ Train and Evaluate the Neural Network using the training dataset with all features.

**06. NN-RWQPS: Fine-Tuning Loop**

- For each neural network model:
  - For each feature selection method:
    - For a range of feature counts:
      - ✓ Fine-tune the model by selecting a specific number of features.
      - ✓ Train the model with the reduced feature set.

**07. NN-RWQPS: Optimal Feature Set**

- a. For each neural network model:
  - ✓ Identify the optimal feature set with the best performance.
  - ✓ Add this feature set to the set of optimal feature sets for that model.

**08. NN-RWQPS: Model Evaluation**

- a. Investigate the NN-model's efficiency using the chosen evaluation metrics.
  - b. Track the best-performing feature count and corresponding performance metrics.
- 

**09. NN-RWQPS: Performance Comparison**

- a. Compare the performance of different neural network models with their respective optimal feature sets.
  - b. Utilize the chosen evaluation metrics to assess model performance.
  - c. Store the performance metrics in the dictionary.
- 

In the current research article, the focus was on training different neural network classifiers, namely the WNNC, BNNC, TNNC, and ONNC. The researchers meticulously detailed the training process for each classifier, providing a step-by-step procedure encapsulated in Algorithms 2, 3, 4, and 5. This comprehensive approach facilitates a better understanding of the training methodologies and paves the way for insightful comparisons and application of water quality prediction.

**Wide Neural Network Classifier (WNNC):**

Training a "WNN Classifier" is a research-intensive process aimed at harnessing the potential of expansive network architectures. As presented in Algorithm 2, the model employs a greater number of neurons in its hidden layers, allowing it to capture intricate data patterns and relationships. Optimization techniques are crucial to prevent overfitting in these wider networks. WNN classifiers exhibit limitations, notably susceptibility to overfitting owing to an abundance of parameters, computational intensity, difficulties with high-dimensional data, and sensitivity to noise.

---

**Algorithm 2: WNNC Training**

**Input:** Water Dataset, WNNC with all hyperparameters, Optimization Algorithm, Batch Size, Number of training epochs

**Output:** Trained WNNC for Classification

---

**1. WNNC: Initialization**

- a. Create a WNN Classifier with the configuration of an input, wide hidden, and an output layers.

**2. WNNC: Loss Function Selection**

- a. Choose the most suitable loss function for water quality classification as binary cross-entropy.

**3. WNNC: Training**

Initialize the parameters NN-Classifier.

For each training epoch:

- a. Shuffle the training data to introduce randomness.
  - b. Divide the training data into batches of a predefined batch size for mini-batch training.
  - c. For each batch:
    - ✓ Conduct a forward pass to compute predictions for the batch.
-

- ✓ Calculate the loss using the selected loss function.
- ✓ Perform backpropagation to update the network's parameters, including weights and biases.
- d. Calculate the average loss for the epoch as a performance metric.

**4. WNNC: Validation and Early Stopping**

- a. Throughout the training process, evaluate the model's performance on the validation dataset.
- b. Employ early stopping based on the validation loss to mitigate overfitting.

**5. WNNC: Evaluation**

- a. Following the training process, assess the model's performance using the test dataset to measure its classification accuracy and generalization capabilities.

**6. WNNC: Deployment**

- a. After the model meeting the predefined performance criteria, consider deploying it for the present study.

**Bilayer Neural Network Classifier (BNNC):**

The training of the "BNN Classifier" is a technically intricate process involving the optimization of weights and biases within a neural network architecture comprising two layers: an input layer and a hidden layer. During training, data is propagated forward through the input layer and then through the hidden layer, employing nonlinear activation functions and backpropagation for error minimization as mentioned in Algorithm 3.

Parameters such as learning rates and batch sizes are iteratively adjusted to achieve convergence and optimal performance. This advancement of bi-layer neural network methodologies in the field of water quality analysis is shown high accuracy in classification. BNN limitations include limited complexity, struggle with non-linearity, potential under fitting, and challenges in capturing intricate hierarchical features.

**Algorithm 3: BNNC Training**

**Input:** Water Quality Dataset, BNNC with number of input and output neurons, Activation Function, Loss Function, Optimization Algorithm, Batch Size, Number of training epochs

**Output:** Trained BNNC for Classification

**1. BNNC: Initialization**

- a. Create a BNNC with an input an output layer.
- b. Configure the input layer with the appropriate number of neurons and activation function.
- c. Configure the output layer with the parameters for the number of classes and an appropriate activation function.

**2. BNNC: Loss Function Selection**

- a. Choose the loss function for water dataset

**3. BNNC: Training**

Initialize the neural network's weights and biases.

For each training epoch:

- a. Shuffle the training Water Quality data.
- b. Divide the training data into batches of the specified batch size.
- c. For each batch:
  - ✓ Perform a forward pass to compute predictions for the batch.
  - ✓ Calculate the loss using the selected loss function.
  - ✓ Backpropagate the gradients to update the network's parameters.
- d. Calculate the average loss for the epoch.

**4. BNNC: Validation and Early Stopping**

- a. Assess the model's effectiveness on the validation dataset while training.
- b. Employ early stopping by monitoring validation loss to safeguard against potential overfitting issues.

**5. BNNC: Evaluation**

- a. After training is complete, assess the model's performance on the test dataset to measure its classification accuracy and generalization.

**6. BNNC: Deployment**

- a. If the model meets the performance criteria, deploy it for real-world applications.

**Trilayer Neural Network Classifier (TNNC):**

Training the "TNN Classifier" involves a meticulous process of optimizing weights and biases across three layers to achieve superior classification performance. Data is propagated through the input layer, followed by the two hidden layers, utilizing advanced activation functions and backpropagation algorithms as presented Algorithm 4. The parameters are fine-tuned through iterative epochs. Tri-layer neural network limitations encompass potential overfitting, and challenges in learning complex hierarchical representations in data.

**Algorithm 4: TNNC Training**

**Input:** Water Dataset, BNNC with number of input neurons, hidden neurons, output neurons, Activation functions, Loss Function, Optimization Algorithm, Batch Size, Number of training epochs

**Output:** Trained TNNC for Classification

**1. TNNC: Initialization**

- a. Create a tri-layer neural network with an input layer, one or more hidden layers, and an output layer.
- b. Configure the input layer with the appropriate number of neurons and activation function.
- c. Configure the hidden layers with the number

of neurons and activation functions.

- d. Configure the output layer with the parameters for the number of classes and an appropriate activation function.

## 2. TNNC: Loss Function Selection

- a. Choose the loss function for water Quality Prediction

## 3. TNNC: Training

**Initialize the neural network's weights and biases.**

For each training epoch

- e. Shuffle the training Water Quality Data.
- f. Divide the training data into batches of the specified batch size.
- g. For each batch:
  - ✓ Perform a forward pass to compute predictions for the batch.
  - ✓ Calculate the loss using the selected loss function.
  - ✓ Back propagate the gradients to update the network's parameters.
- h. Calculate the average loss for the epoch.

## 4. TNNC: Validation and Early Stopping

- c. During training, evaluate the model's performance on the validation dataset.
- d. Apply early stopping based on validation loss to prevent overfitting.

## 5. TNNC: Evaluation

- b. After training is complete, assess the model's performance on the test dataset to measure its classification accuracy and generalization.

## 6. TNNC: Deployment

- a. If the model meets the performance criteria, deploy it for present study.

on independent water quality datasets, enabling accurate and tailored water quality assessment.

---

### Algorithm 5: ONNC Training

**Input:** Water Dataset, ONNC with number of layers, neurons per layer, Activation Function, Loss Function, Optimization Algorithm, Mini Batch Size, Number of training epochs

**Output:** Trained ONNC for Classification

---

#### 1. ONNC: Initialization

- a. Construct a neural network architecture with the specified hyperparameters, which may include the number of layers, neurons per layer, and activation functions.
- b. Initialize the network's weights and biases using a suitable initialization method

#### 2. ONNC: Loss Function Selection

- a. Choose a loss function tailored to the present classification problem

#### 3. ONNC: Training

Initialize epoch counter.

While the maximum number of training epochs is not reached:

- a. Shuffle the training data to introduce randomness.
- b. Divide the training data into mini-batches of a predefined size.
- c. For each mini-batch:
  - ✓ Perform a forward pass to compute predictions.
  - ✓ Calculate the loss using the selected loss function.
  - ✓ Conduct a backward pass to compute gradients.
  - ✓ Update the network's weights and biases using the chosen optimization algorithm.
- d. Increment the epoch counter.
- e. Monitor the training process and display or record the loss and performance metrics.

#### 4. ONNC: Validation and Early Stopping

- a. Continuously evaluate the model's performance on the validation dataset during training.
- b. Implement early stopping based on the validation loss to prevent overfitting.

#### 5. ONNC: Evaluation

- a. After training is completed, assess the model's performance on the test dataset to measure classification accuracy and generalization.

#### 6. ONNC: Deployment

- a. Save the trained model parameters and performance criteria, consider to deploy for the present study.
- 

### Optimizable Neural Network Classifier (ONNC):

The "ONN Classifier" demands specialized attention in the present research due to its capacity for dynamic optimization. In the context of water quality analysis, training an "ONN Classifier" involves the customization of the neural network's architecture and hyper parameters to optimize the precision and efficiency of water quality parameter assessment as shown in Algorithm 5.

The present Researchers establish a comprehensive search space, encompassing various network configurations. A relevant performance metrics are chosen to gauge the classifier's effectiveness in water quality analysis. Optimization techniques are applied iteratively to adjust the network's architecture and hyper parameters, aligning it with the chosen metrics. The process continues until the best-performing configuration is identified through convergence. Subsequently, it undergoes rigorous real world testing

In this research study, a selection of Neural Network Classifiers, including WNNC, BNNC, TNNC, and ONNC, undergo rigorous execution alongside their respective best feature selection methods. Each classifier undergoes training with feature sets ranging from null feature set to incorporating all available features set.

The primary goal is to determine the optimal feature set for each classifier based on its performance. This meticulous evaluation process involves a comprehensive analysis to identify the most accurate and efficient feature combinations for the given task, thereby providing empirical evidence of feature relevance and classifier effectiveness. Subsequently, the performance of all neural networks, each fine-tuned with its optimal feature set, is compared. This comparative analysis helps identify the top-performing classifier when utilizing the most suitable feature set. The research findings contribute to the field of machine learning and shed light on the intricate relationship between feature selection and classifier performance.

#### 4. RESULTS AND DISCUSSIONS

This section unveils the results of the NN-RWQPS model and its thorough validation. The empirical study delves into the synergy of forward feature engineering and Neural Network classifiers for assessing water quality. The proposed system entails training and evaluating an NN-powered Robust Water Quality Prediction System using real-world data. Upcoming sections will detail the insights and conclusions derived from this investigation of each neural network classifier.

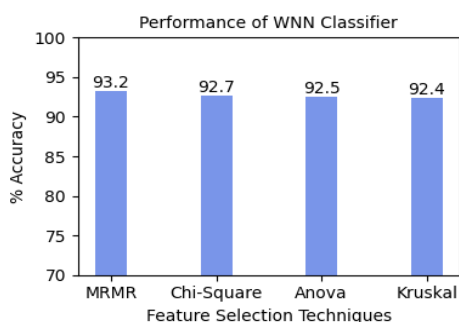


Figure 3. Performance of WNNC with FS Techniques

The research results offer insights into the performance of feature selection techniques with a focus on accuracy. MRMR stands out with the highest accuracy of 93.20%, making it the top choice for researchers prioritizing classification accuracy as showcased in Figure 3. Chi-Square, ANOVA, and Kruskal are competitive with accuracies. MRMR's slight advantage in accuracy makes it the preferred selection if high accuracy is the objective.

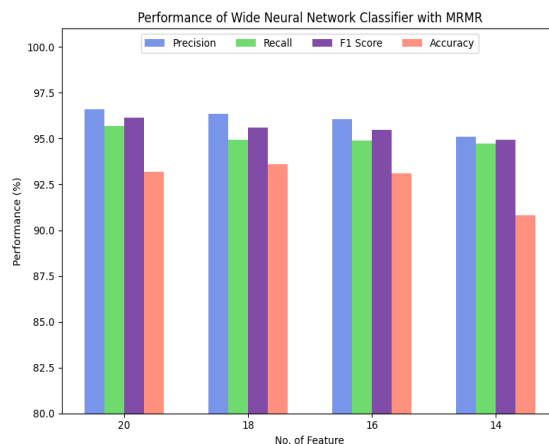


Figure 4. Performance of WNNC with MRMR

The selected dataset for this study encompasses 21 numerical attributes, systematically collected to analyse diverse aspects of water quality analysis systems. These attributes cover a broad range of parameters crucial for comprehending and assessing water quality. The presented results demonstrate the impact of varying the number of features on classification performance metrics. As the number of features decreases from 20 to 14, accuracy exhibit a corresponding decline as shown in Figure 4.

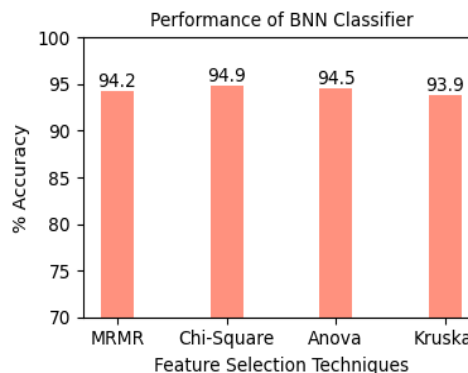


Figure 5. Performance of BNNC with FS Techniques

The research findings reveal that among the feature selection techniques, Chi-Square attains the highest accuracy at 94.90%, making it the top choice for maximizing classification accuracy. MRMR and Anova closely follow with commendable accuracies of 94.60% and 94.80%, respectively. Kruskal, while still achieving a respectable 93.90% accuracy, slightly lags behind the other methods. The decision is clear for researchers aiming to optimize accuracy - Chi-Square is the preferred feature selection technique as presented in Figure 5.

The research outcomes provide valuable numerical insights into the effect of feature reduction on classification performance metrics. As the number of features decreases from 20 to 14, precision remains relatively stable, with a minimal variation of approximately 1%. Similarly, recall shows only a slight

deviation of around 1%. F1 Score, a measure of the balance between precision and recall, maintains a high value across different feature counts, with a minor change of about 1%. This emphasizes the trade-off between model simplicity and overall classification performance, while considering precision, recall, F1 Score, and accuracy in the research analysis presented in Figure 6.

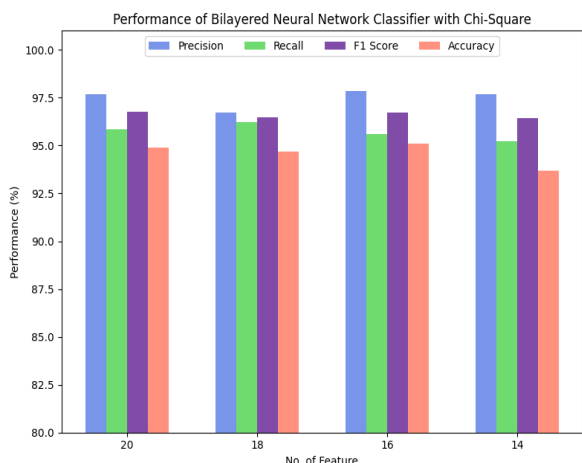


Figure 6. Performance of BNNC with Chi-Square

The research results demonstrate varying accuracy levels among different feature selection techniques. MRMR stands out with the highest accuracy at 95.20%, making it a robust choice for researchers seeking optimal classification accuracy. Chi-Square closely follows with a commendable accuracy of 94.60%. Anova exhibits a slightly lower accuracy at 93.90%, while Kruskal lags behind at 92.90%. The decision for accuracy optimization is clear - MRMR is the preferred feature selection method as depicted in Figure 7.

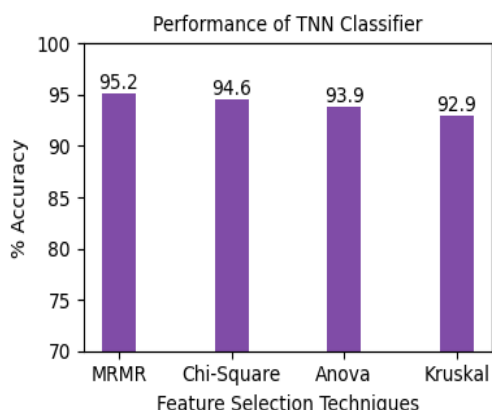


Figure 7. Performance of TNNC with FS Techniques

The research results provide a meticulous exploration of the interplay between varying feature counts and classification performance metrics. Precision consistently maintains high values, ranging from 96.34% to 97.56%, indicating the model's proficiency in correctly identifying positive cases. Likewise, recall remains stable, with values between 95.47% and

96.75%, reflecting the model's capacity to capture true positive instances effectively. The F1 Score consistently records values around 96.37%, underscoring the model's ability to balance precision and recall harmoniously. However, accuracy demonstrates a declining trend as the feature count decreases, highlighting the intricate trade-off between model simplicity and overall performance. These findings emphasize the importance of selecting the optimal feature count at 16 as shown in Figure 8.

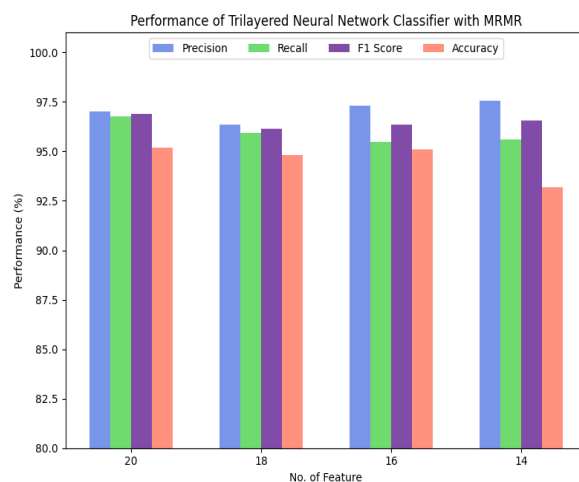


Figure 8. Performance of TNNC with MRMR

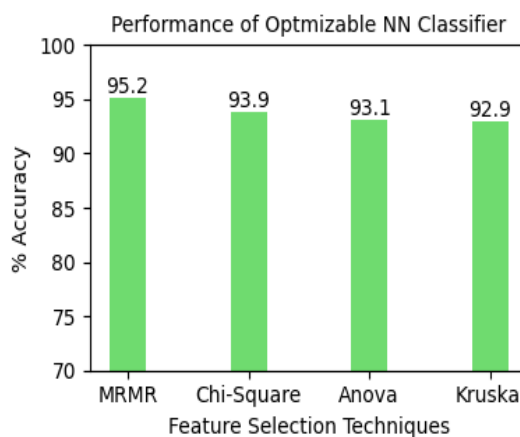


Figure 9. Performance of ONNC with FS Techniques

The research findings depict the accuracy performance of various feature selection techniques. MRMR stands out with the highest accuracy of 95.20%, showcasing its efficacy in enhancing classification accuracy as presented in Figure 9. Chi-Square, while strong, records a slightly lower accuracy of 93.90%. Anova follows with an accuracy of 93.10%, and Kruskal trails closely at 92.90%. The decision for this study focusing on accuracy enhancement is clear – so, MRMR is the optimal choice.

The research findings offer a detailed exploration of the impact of varying feature counts on classification performance metrics. Precision consistently achieves

high values, ranging from 97.56% to 99.32%, indicating the model's proficiency in correctly identifying positive cases. Recall maintains stability, with values between 94.76% and 95.81%, reflecting the model's effective capture of true positive instances. The F1 Score consistently records values around 96.40% to 97.53%, showcasing the model's ability to balance precision and recall harmoniously as showcased in Figure 10.

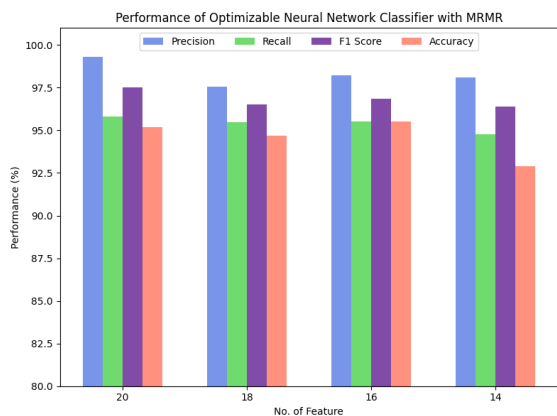


Figure 10. Performance of ONNC with MRMR

However, accuracy exhibits a gradual decrease as the feature count diminishes, highlighting the trade-off between model simplicity and overall classification performance. These insights underscore the importance of selecting the optimal feature count, considering research goals.

## References:

- Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). [Retracted] Water Quality Prediction Using Artificial Intelligence Algorithms. *Applied Bionics and Biomechanics*, 2020(1), 6659314. doi:10.1155/2020/6659314
- Bonthu, S., Kiran, K. B., Deenakonda, M., Rao, V. M., & Jagadeesh, S. V. V. D. (2023, May). Multi-label and Multi-class Classification on a Custom Dataset using Convolution Neural Networks. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 576-579). IEEE. doi: 10.1109/ICICCS56967.2023.10142828.
- Fei Han, Fanyu Li, Qinghua Ling, et al. (2023). A Feature Selection Method Based on Feature-Label Correlation Information and Self-Adaptive MOPSO. *Research Square*, September 14, 2023. doi: 10.21203/rs.3.rs-3348206/v1.
- Flores, V., Bravo, I., & Saavedra, M. (2023). Water Quality Classification and Machine Learning Model for Predicting Water Quality Status—A Study on Loa River Located in an Extremely Arid Environment: Atacama Desert. *Water*, 15(16), 2868. doi: 10.3390/w15162868.
- Gong, L., Xie, S., Zhang, Y., Wang, M., & Wang, X. (2022). Hybrid feature selection method based on feature subset and factor analysis. *IEEE Access*, 10, 120792-120803. doi: 10.1109/ACCESS.2022.3222812.
- Goodarzi, M. R., Niknam, A. R. R., Barzkar, A., Niazkar, M., Zare Mehrjerdi, Y., Abedi, M. J., & Heydari Pour, M. (2023). Water quality index estimations using machine learning algorithms: a case study of Yazd-Ardakan Plain, Iran. *Water*, 15(10), 1876. doi:10.3390/w15101876.
- Kalaivanan, K., & Vellingiri, J. (2022). Survival Study on Different Water Quality Prediction Methods Using Machine Learning. *Nature Environment & Pollution Technology*, 21(3). doi:10.46488/NEPT.2022.v21i03.032.
- Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water*, 14(10), 1552. doi:10.3390/w14101552
- Kurra, S. S., et al. (2022). Water quality prediction using machine learning. *International Research Journal of Modernization in Engineering Technology and Science*. 53(1): 3–13. doi:10.2166/wqrj.2018.025

## 5. CONCLUSIONS

The present study has demonstrated the significant potential of integrating feature engineering with neural network classifiers for water quality prediction. The research findings reveal that the proposed approach enhances the robustness and accuracy of water quality prediction.

Through rigorous experimentation, have ascertained that the feature engineering process, when combined with neural networks, effectively identifies and selects relevant features, leading to improved model performance. This is particularly crucial in real-world scenarios where water quality can be influenced by diverse factors. The empirical results highlight the ability of the optimized neural network classifiers to provide reliable and accurate predictions, enabling better-informed decision-making in water quality management. Overall, this research contributes to the development of a more robust and efficient Water Quality Prediction System, offering valuable insights for environmental monitoring and resource management.

## Acknowledgements:

First author, Dr. Maheswara Rao V V R, would like to express their sincere thanks to the management, Shri Vishnu Educational Society (SVES), of Shri Vishnu Engineering College for Women (A), Bhimavaram-534202, India for their logistic support that helped us to carry out this work.

- Maheswara Rao, V. V. R., et al. (2022). An Enhanced Machine Learning Classification System to Investigate the Status of Micronutrients in Rural Women. In *International Conference on Recent Trends in Computing, Lecture Notes in Networks and Systems, 341* (Springer). doi: 10.1007/978-981-16-7118-0\_4
- Maheswara Rao, V. V. R., N. Silpa, Mahesh Gadiraju, Reddy Shiva Shankar, Dr. Vijaya Kumar, & Dr. K. B. V. Brahma Rao (2023). An Optimal Machine Learning Model Based On Selective Reinforced Markov Decision to Predict Web Browsing Patterns. *Journal of Theoretical and Applied Information Technology, 101*(2), 859-873.
- Patel, J., Amipara, C., Ahanger, T. A., Ladhva, K., Gupta, R. K., Alsaab, H. O., ... & Ratna, R. (2022). A Machine Learning-Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable AI. *Computational Intelligence and Neuroscience, 2022*(1), 9283293. doi: 10.1155/2022/9283293
- Paptsov, A. G., & Popova, K. Yu. (2022). Management of the quality of water resources for sustainable development based on industrial and manufacturing engineering. *International Journal for Quality Research, 16*(2), 311–328. <https://doi.org/10.24874/IJQR16.02-01>
- Rao M. V. V. R., et al. (2023). An Innovative Machine Learning based Heart Disease Assessment System by Sequential Feature Selection Approach. *International Conference on Intelligent Technologies (CONIT)*. doi: 10.1109/CONIT59222.2023.10205817
- Rao M. V. V. R., S. N. M. Gadiraju, S. S. Reddy, S. Bonthu, & R. R. Kurada (2023). A Plausible RNN-LSTM based Profession Recommendation System by Predicting Human Personality Types on Social Media Forums. In *2023 7<sup>th</sup> International Conference on Computing Methodologies and Communication (ICCMC)*. doi: 10.1109/ICCMC56507.2023.10083557
- Rao, V. V. R. M., V. V. Kumari & K. V. S. V. N. Raju (2011). An Intelligent System for Web Usage Data Preprocessing. In: Meghanathan, N., Kaushik, B. K., Nagamalai, D. (eds) *Advances in Computer Science and Information Technology. CCSIT 2011*. doi:10.1007/978-3-642-17857-3\_47
- Reddy, S. S., Gadiraju, M., & Maheswara Rao, V. V. R. (2022). Analyzing student reviews on teacher performance using long short-term memory. In *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021* (pp. 539-553). Singapore: Springer Nature Singapore. doi:10.1007/978-981-16-7167-8\_39
- Silpa, N., & Rao, V. M. (2021). Enriched big data pre-processing model with machine learning approach to investigate web user usage behaviour. *Indian J Comput Sci Eng [Internet], 12*(5), 1248-56., doi: 10.21817/indjce/2021/v12i5/211205050
- Silpa, N., & Rao, V. M. (2022). Machine learning-based optimal segmentation system for web data using Genetic approach. *Journal of Theoretical and Applied Information Technology, 100*(11), 3552-3561.
- Silpa, N., VVR, M. R., Subbarao, M. V., Pradeep, M., Grandhi, C. R., & Karunasri, A. (2023, June). A Robust Team Building Recommendation System by Leveraging Personality Traits Through MBTI and Deep Learning Frameworks. In *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)* (pp. 1-6). IEEE. doi: 10.1109/ICICAT57735.2023.10263718
- Uddin, M. G., Nash, S., Diganta, M. T. M., Rahman, A., & Olbert, A. I. (2022). Robust machine learning algorithms for predicting coastal water quality index. *Journal of Environmental Management, 321*, 115923. doi:10.1016/j.jenvman.2022.115923.
- Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., ... & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health, 1*(2), 107-116. doi:10.1016/j.eehl.2022.06.001.
- Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., ... & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health, 1*(2), 107-116. doi: 10.1016/j.eehl.2022.06.001

---

**Maheswara Rao V V R**

Shri Vishnu Engineering College for Women,  
Bhimavaram,  
India  
[mahesh\\_vvr@yahoo.com](mailto:mahesh_vvr@yahoo.com)  
ORCID 0000-0002-0503-7211

**Shiva Shankar Reddy**

S.R.K.R. Engineering College,  
Bhimavaram,  
India  
[shiva.shankar591@gmail.com](mailto:shiva.shankar591@gmail.com)  
ORCID 0000-0001-5439-0348

**Silpa N**

Shri Vishnu Engineering College for Women,  
Bhimavaram,  
India  
[nrusimhadri.silpa@gmail.com](mailto:nrusimhadri.silpa@gmail.com)  
ORCID 0000-0003-3411-0358

**Ramachandra Rao Kurada**

Shri Vishnu Engineering College for Women,  
Bhimavaram,  
India  
[ramachandrarao.kurada@gmail.com](mailto:ramachandrarao.kurada@gmail.com)  
ORCID 0000-0002-7014-8313

**Kranthi Addanki**

James Cook University,  
James Cook Drive,  
Australia  
[kranthi.addanki@my.jcu.edu.au](mailto:kranthi.addanki@my.jcu.edu.au)  
ORCID 0000-0003-4921-8783

**Pachipala Yellamma**

Koneru Lakshmaiah Education Foundation,  
Vaddeswaram, Guntur,  
India  
[pachipala.yamuna@gmail.com](mailto:pachipala.yamuna@gmail.com)  
ORCID 0000-0002-6291-7418

---