International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. IV, No. 07, pp. 37-47, July

AquaSpikeNet: Bio-Inspired Spiking Neural Network Framework for High-Resolution Water Quality Assessment

Ramya R R[1], N. Rathika[2], R. Sahila Devi[3]

[1]Assistant Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi. ramyarr@veltech.edu.in
[2]Associate Professor, Department of Electrical and Electronics Engineering, SKP Engineering College, Tiruvannamalai. rathika.nsd@gmail.com
[3]Assistant Professor, Department of Computer Science and Engineering, Rohini College of Engineering and Technology, Kanyakumari, India. r.sahiladevi@gmail.com

***Corresponding Author E-mail:** ramyarr@veltech.edu.in

ABSTRACT: Water quality (WQ) is to serious environmental and health concern as a result of worldwide population growth. A predictive model for WQ is essential for the timely prevention and management of water pollution. This study proposes a novel Bio-Inspired Spiking Neural Network (SNN) for identifying water pollution in the environment. The data pre-processing stage finds the missing data and shows the duplicate WQ values through data imputation and data cleaning methods. Also, the data normalization where WQ data are normalised as mounting data into nominal data and transformed the better WQ data using data transformation. Then the WQ data are split as 80% train and 20% test using a data splitting technique. The Bio-Inspired SNN model is trained and tested on processes WQ data, enabling to accurately predict pollution. Implemented using Python software, the WQ data is chosen from WQ dataset and effectively evaluated pollution sources with an improved prediction accuracy of 93%.

Keywords: Spiking Neural Network, Data cleaning, Data imputation, Data Normalization, Data Transformation, Data Splitting, and Water Quality.

1. Introduction

Water is one of the most valuable resources in daily life that is measured by a health of aquatic life. Consequently, it is impacted by water contamination, which lowers the quality of the water. Since monitoring WQ and ensuring the survival of marine species are crucial [1]. Whereas, WQ concerns and challenges require mitigating and controlling water pollution. Many governments throughout the world have started to develop ecological water management programs in an effort to understand the state of the nautical ecosystem [2]. Approximately one billion people lack access to safe drinking water, and two million people die each year as a result of contaminated water, inadequate sanitation, and unclean conditions. Therefore, maintaining the quality of the freshwater is essential for daily use [3]. To ensure safe drinking water at affordable prices, there is an increasing demand for improved water management and WQ control. Systematic evaluations of freshwater, disposal systems and organizational monitoring practices are required to address these problems [4].

However, WQ forecasting involves predicting the features of variations in the health of a water system. Planning and regulating WQ depend heavily on the results of WQ assessments [5]. Additionally, predicting WQ data at various pollution levels is crucial for developing practical measures to prevent further contamination. Thus, the effective pollution control strategies are built on three pillars: monitoring, modelling, and management [6-7]. The conventional techniques for assessing WQ count on comparing parameter values that have been found experimentally with the current recommendations [8]. Consequently, Water Quality Indices (WQI) are the methods that greatly reduce the amount of data and make it easier to communicate the state of the WQ [9]. Physico-chemical and bacteriological characteristics are used to calculate the WQ by the following techniques [10].

Using various existing classification techniques, the WQ data is classified and analysed to assess the different levels of WQ and support effective decision-making. Using Decision Tree (DT) technique, the WQ assessment in diverse water body type is calculated with higher accuracy. Low requirements for data preparation, interpretability and the capacity to deal with missing values and non-linear connections are happened in this classifier. Whereas, over fitting, is susceptible to fluctuations in data, and not effective with huge datasets or high-dimensional data [11]. A Support Vector Machine (SVM) classifier is used as a predictor for the WQ data. SVM have a higher value of relative index prediction error and classify the data with higher accuracy. However, computationally costly, prone to outliers and noise, have difficulty using large datasets [12]. To improve the classification and prediction of WQ, CAT-Boost model is used. It is a flexible and effective gradient boosting algorithm with excellent accuracy and precision, reduces over fitting, and offers feature importance analysis. Nonetheless, includes the possibility of data quality loss and the requirement for meticulous parameter adjustment [13].

Consequently, a LightGBM classifier is used to predict the WQ data with higher accuracy. LightGBM uses the large dataset for prediction and produces higher accuracy and performance. However, it takes more processing time to access a single data and over fitting for noise [14]. Whereas, using Random Forest (RF) classifier in WQ data is calculated to more accurate allocation of pollution sources and identifying possible causes of variances. It's provides higher accuracy and adaptability in managing problems involving regression and classification. Nevertheless, its drawbacks include its high processing cost and lack of interpretability [15]. To overcome these issues the proposed Bio-Inspired SNN classification technique used to predict higher WQ accuracy and good performance.

2. Related Work

Hmoud Al-Adhaileh et al [16] (2021) have proposed a Feed-Forward Neural Network (FFNN) and K-Nearest Neighbours (KNN) applied to classify WQ. FFNN and K-NN that estimate water pollution concentrations in various urban locations by utilizing WQ data. The model's dependence on data, sensitivity to WQ, and extensive network. However, the effectiveness this method it fails to achieve the necessary level on hyper parameters.

Wu et al [17] (2022) have developed an Artificial Neural Network (ANN), Discrete Wavelet Transform (DWT), and Long Short-Term Memory (LSTM) to predict the WQ. ANN-DWT-LSTM techniques classify and predict the WQ with better performance. The Prediction of WQ data is analysed by using the lake water with high accuracy. However, it have over fitting problem when the noise reduction process going on and computationally expensive.

Alfwzan *et al* [18] (2023) have presented a Deep Learning based Bi-directional Long Short Term Memory (Bi-LSTM) to classify the ground WQ. To find the ground WQ Bi-LSTM process in both forward and backward direction and it is an effective as well as precious method. Its benefit is finding the quality of ground water upward and down ward direction with high accuracy. Nevertheless, higher memory needs and computational complexity in comparison to conventional LSTMs.

Pérez-Beltrán et al [19] (2024) have proposed an ANN to calculate the WQ with high accuracy and performance. ANN is an effective machine learning tool for WQ control and problem identification. ANN finds the problem depth wise and control the WQ for lakes, ponds, sea etc. Whereas, it have good sensitivity and specificity value when controls the WQ. However, increased computational burden, then observe the character of model building, and susceptibility to over fitting.

Arepalli et al [20] (2024) have developed an enhanced Dilated Spatial-Temporal Convolution Neural Network (DSTCNN) for WQ monitoring in aquaculture. DSTCNN captures the real-time data from the aqua ponds and aqua river. Capacity to identify temporal and regional patterns in data, as well as the application of mixed activation functions to increase accuracy. However, possibility of high blunt expenditures, integration difficulty, and a lack of appropriate data.

The Contributions of the Work are as follows:

- Data pre-processing involves Data cleaning and imputation to handle missing WQ values and remove outliers.
- Data normalization converts the WQ data's from mounting data into nominal data and data transforms convert the raw WQ data into processed WQ data.
- Employs a Bio-Inspired SNN to calculate the WQ, provides precise and robust predictions for WQ related metrics.

3. Proposed Work

Figure 1 shows the block diagram of the proposed Bio-Inspired SNN model for environmental analysis that uses advanced learning method. Bio-Inspired SNN efficiently predicts WQ using the WQ dataset. Firstly, the data pre-processing technique uses data cleaning, data imputation to find the missing WQ values and remove outliers. By data normalization the WQ data's are normalised as mounting data into nominal data. Then transforms the raw WQ data in to processed WQ data using data transformation. Following to data transformation the processed data is test and trained using data splitting technique. Finally, the trained WQ data is given as an input to proposed Bio-Inspired SNN classifier. The Bio-Inspired SNN classifier classifies the WQ spikes, which are more realistic in terms of biology and energy efficiency.

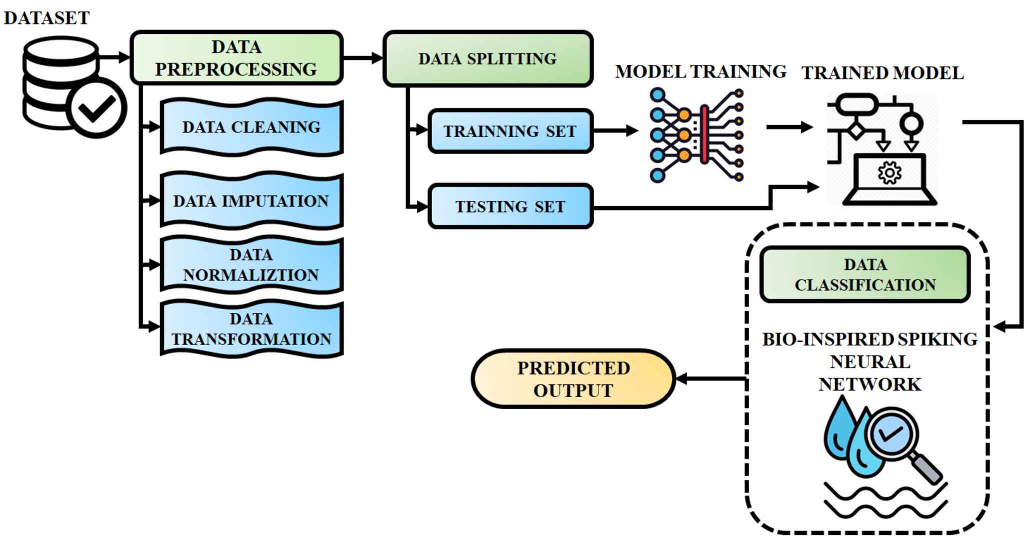


Figure 1: *Block diagram for proposed Bio-Inspired SNN*

3.1 Data Pre-Processing for WQ

The first step in the data pre-processing includes WQ is data cleaning, data imputation, data normalization and data transformation.

3.1.1 Data Cleaning for WQ

The process of finding and fixing errors, variations, and inaccuracies in a WQ dataset, in order to enhance its data quality and usability, is called data cleaning, referred to as data cleansing. Some of the data cleaning techniques are finding and fixing problems including inaccurate data entry, missing numbers, duplicates, and poor formatting.

3.1.2 Data Imputation

Data imputation models assessed, and their loss functions or missing WQ values computed. It decreases the size of the WQ dataset and result in skewed findings and the loss of important information, particularly if the dataset have a significant proportion of missing values.

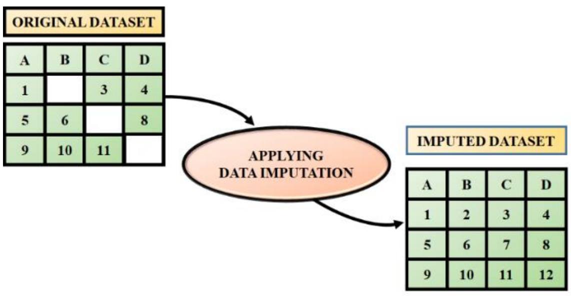


Figure 2: *Data imputation for WQ*

Figure 2 shows the data imputation for WQ. The simplest imputation methods include weighted averaging, arithmetic, and mean imputation and linear interpolation, which only use the time-series data that is available to carry out the imputation.

3.1.3 Data Normalization

The process of data normalization, is converting the rising data into nominal data. Each value in the instances is scaled to an equal contribution as part of the normalizing procedure in figure 3.

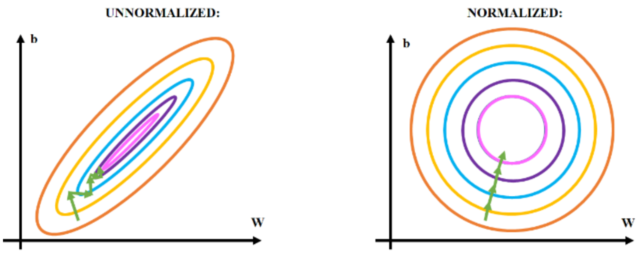


Figure 3: *Data normalization for WQ*

It addresses outliers, which delay the training process of classifier design, and decreases the bias of characteristics, allowing numerical differentiating to effectively distinguish between classes. The data normalization classes are mini-max scalar, Z-score normalization that normalizes the WQ data.

3.1.4 Data Transformation

The process of transforming raw data into structure for analysis in order to enhance its WQ dataset is known as data transformation. Log and square root transformations are common strategies for mitigating the consequences of outliers and skewed distributions. Log transformation significantly lowers the impact of extreme values, whereas square root transformation provides a less forceful adjustment and is especially good for positively skewed data. Following to this, the pre-processed data is given to the data splitting.

3.2 Data Splitting for WQ

In data splitting two parts of the dataset are separated as a test set and a training set. 20% of the total data in the test set, whereas 80% in the training dataset. To predict or select an alternative, the machine learning algorithm establishes a relationship with the independent and dependent parameters. To get around this issue, every model is constructed 20 periods using various random subsets of training and testing data. Followed by the data splitting WQ data is given to the proposed Bio-Inspired SNN.

3.3 Data classification by Bio-Inspired SNN for WQ

The proposed Bio-Inspired SNN is utilized for effective classification of WQ assessment.

3.3.1 Introduction to Bio-Inspired SNNs

In order to create artificial intelligence that is more realistic and energy-efficient, SNNs are made to behave similarly to biological neurons in the water. The general architecture moves from simple to sophisticated representations by progressively extracting and refining features from the incoming data. Convolutional and pooling layers work together to improve the network's generalization and pattern recognition, which makes it useful for image recognition applications. Figure 4 shows the basic bio-inspired SNN. A spike is produced when a membrane potential hits a specific threshold. In this work, numerous attempts have been made to model and replicate the water's computational operations. The timing of spikes is the basis for SNN computation; spikes that fire simultaneously are linked together and become more strongly connected.

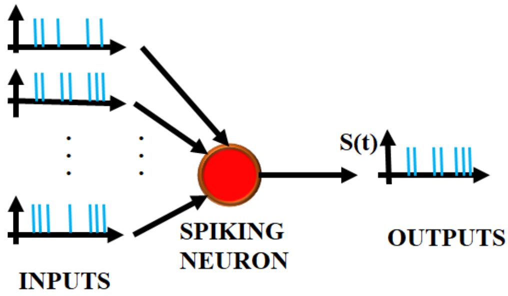


Figure 4: *Basic Bio-inspired SNN*

3.3.2 Bio-Inspired SNN Architecture:

In Figure 5, the Bio-Inspired SNN to WQ, this includes data preparation and classification. The first layer processes raw data, encodes it into spike trains, and applies a WQ data. After that, it is linked to a convolutional layer, which separates each input into many identically sized features and a stride window that moves over the

input.

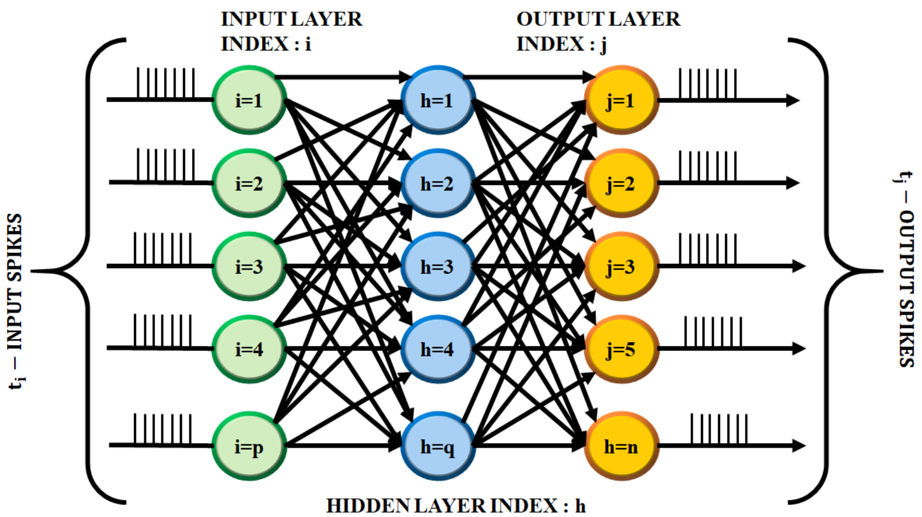


Figure 5: *Bio-Inspired SNN architecture*

A convolutional feature, which serves as an input to the excitatory layer, is formed by each convolutional window. The following formula is used to determine the number of neurons O in the convolutional layer:

$$O = \frac{(in_{size} - c_{size}) + 2P}{c_{stride}} + 1 \qquad (1)$$

Where c_{stride} the size of the stride in the convolutional layer, P indicates padding, c_{size} is the size of each feature in the convolutional layer, and in_{size} is the size of the input image in the input layer. The convolution output size O, is the square root of the convolutional layer's neuron count in figure 6.

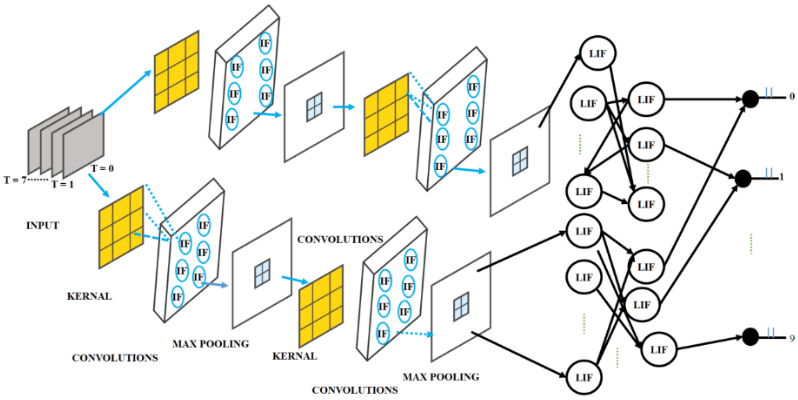


Figure 6: *Bio-Inspired SNN with nine output strides*

All feature patches other than the one to which a neuron is linked are suppressed in the third layer, which is an inhibitory layer. The number of excitatory layer patches is directly correlated with the number of neurons in the inhibitory layer. By using the Bio-Inspired SNN the WQ is calculated depth wise with good performance.

4. Result and Discussion

In this section the proposed Bio-Inspired SNN is to predict the WQ data. For WQ assessment all of the data divided into two sets: 20% were used for testing, and the remaining 80% used for training. WQ datasets used in the models' execution using Python software. Moreover, the model's computation time for future prediction algorithm implementation in performance metrics is another crucial factor.

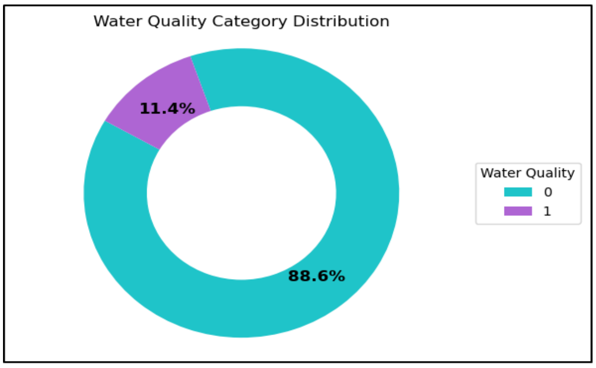


Figure 7: *WQ category distribution*

Figure 7 shows the WQ category distribution as 0 and 1, where 0 shows the class value of 88.6% and 1 have the class value of 11.4%. The dataset is taken from the water quality dataset.

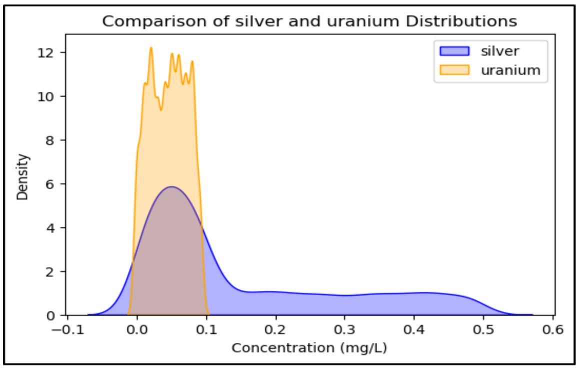


Figure 8: *Comparison of silver and uranium distribution for WQ*

Figure 8 shows the comparison of silver and uranium distribution for WQ. The X-axis shows the concentration value and Y-axis shows the density value. Silver is indicated in blue color and uranium is indicated in orange color. Uranium value is 12 which are increased in-between 0 to 0.1. Then the silver value is 6 which are increased in-between 0 to 0.1 and decreased still 0.55.

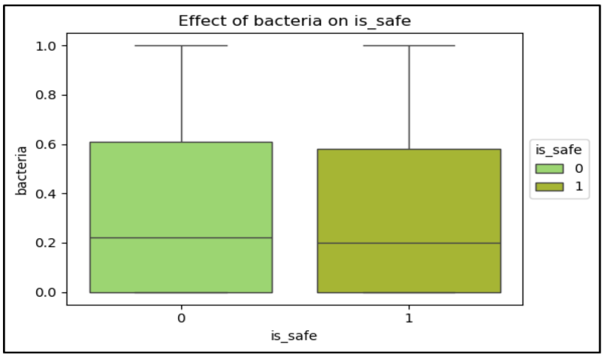


Figure 9: *Effect of bacteria on is safe*

Figure 9 shows the effect of bacteria on is_safe for WQ. The X-axis shows the is_safe 0 and 1 and Y-axis shows the bacteria. 0 is indicated in light green color and 1 is indicated in dark green color. 0 have the bacteria of 0.6 and 1 have the bacteria of 0.56.

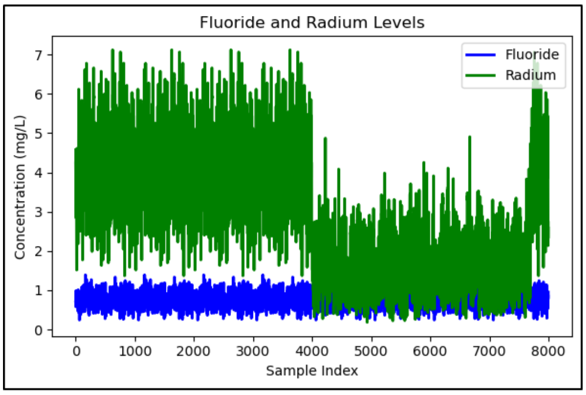
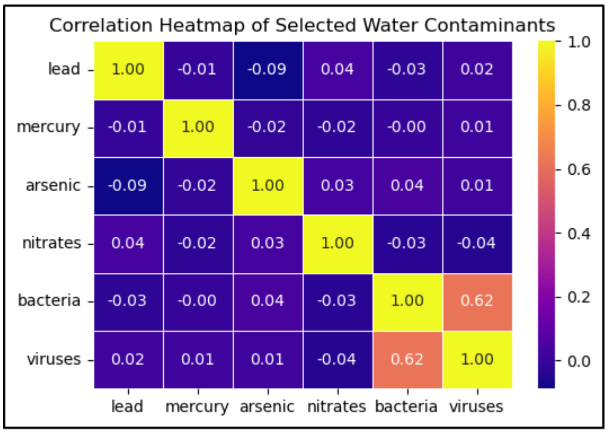


Figure 10: *Fluoride and radium level in water*

Figure 10 shows the fluoride and radium level in water. The X-axis indicates the sample index and Y-axis indicates the concentration. Fluoride is indicated in blue color and radium is indicated in green color. The sample indexes have the values up to 8000 and concentration have the value up to 7. Radium have the vibrations in the concentration between 2 to 7, sample index is 0 to 4000 then concentration 0.5 to 7 and sample index is 4000 to 8000. Fluoride concentration is in between 0.5 to 1.5 and sample index is up to 8000.



	lead	mercury	arsenic	nitrates	bacteria	viruses
lead	1.00	-0.01	-0.09	0.04	-0.03	0.02
mercury	-0.01	1.00	-0.02	-0.02	-0.00	0.01
arsenic	-0.09	-0.02	1.00	0.03	0.04	0.01
nitrates	0.04	-0.02	0.03	1.00	-0.03	-0.04
bacteria	-0.03	-0.00	0.04	-0.03	1.00	0.62
viruses	0.02	0.01	0.01	-0.04	0.62	1.00

Figure 11: *Correlation heatmap of selected water contaminants*

Figure 11 shows the correlation heatmap of selected water contaminants for WQ. The WQ is calculated as lead, mercury, arsenic, nitrates, bacteria and viruses. The value for lead, mercury, arsenic, nitrates, bacteria and viruses is 1.

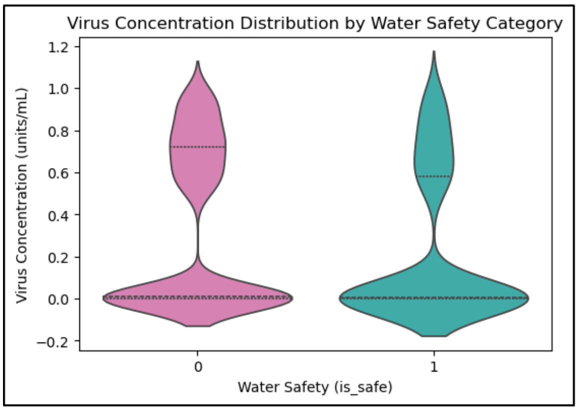


Figure 12: *Virus concentration distribution by water safety category*

Figure 12 shows the virus concentration distribution by water safety category for WQ. The water safety class is distributed as 0 and 1. 0 indicated in purple color and 1 indicated in blue color, virus concentration in between 0 to 1.2.

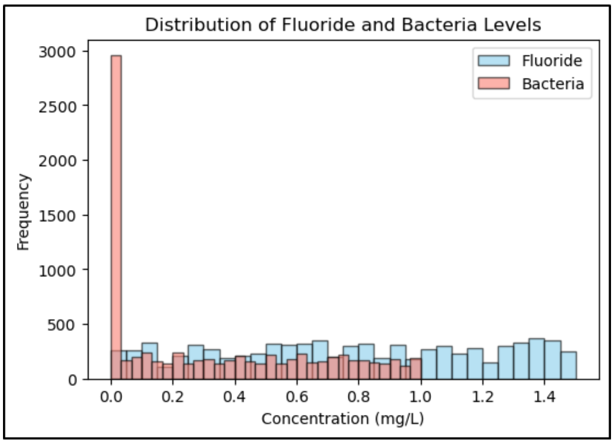


Figure 13: *Distribution of fluoride and bacteria levels*

Figure 13 shows the distribution of fluoride and bacteria levels for water. The X-axis indicates the concentration and Y-axis indicates the frequency. The bacteria are concentrated in between 0 to 1 and it is very high up to 3000 in the concentration 0.0. Fluoride is increased and decreased in between the concentration 0 to 1.45 and frequency up to 500.

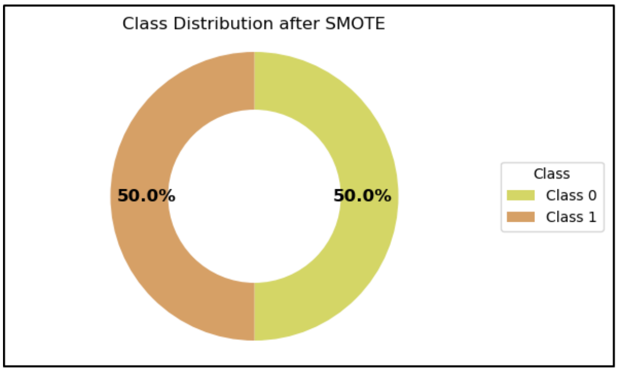


Figure 14: *Class distribution after SMOTE*

Figure 14 shows the class distribution after SMOTE for WQ. The class 0 have the percentage of 50 and class 1 have the percentage of 50 for WQ.

4.1 Performance Metrics

Performance metrics like accuracy, sensitivity and specificity which indicate WQ be correctly detected and appropriately analysed as the probabilities calculated by the Bio-Inspired SNN. The true positive, true negative, false positive and false negative results are denoted by TP, TN, FP, and FN, respectively. Sensitivity$[TP/(TP + FN)]$, specificity $[TN/(TN + FP)]$, F-value $[2 \times (recall \times precision)/\,(recall + precision)]$, and accuracy $[(TP + TN)/(TP + FP + FN + TN)]$.

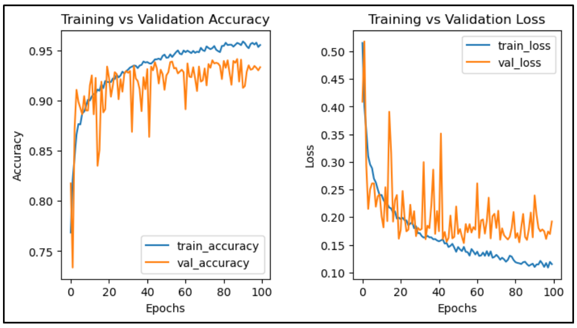


Figure 15: *Training vs validation accuracy, training vs validation loss*

Figure 15 shows the training vs validation accuracy, training vs validation loss. The X-axis denotes the epochs and Y-axis is accuracy for training and validation accuracy simultaneously for training and validation loss X-axis for epochs and Y-axis for loss. The Model Accuracy graph illustrates increasing accuracy of 93%, with the model achieving high performance and minimal over fitting. Training and validation loss is reliably declining in training and validation loss, representing improved learning and reduced error.

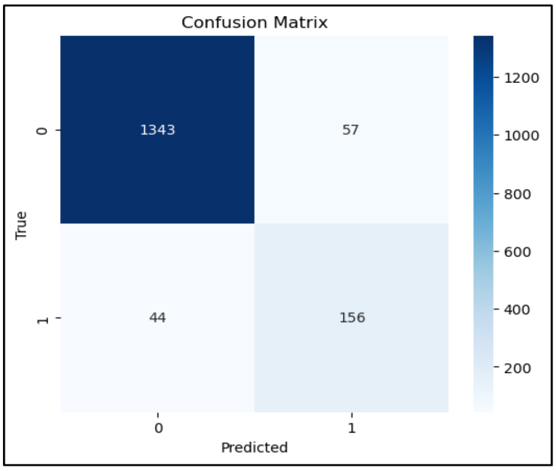


Figure 16: *Confusion matrix for proposed work*

The confusion matrix for WQ is shown in figure 16. Proposed method evaluates the model's performance in predicting class 0 and class 1 of WQ. It shows high accuracy for all classes, with 1343 for class 0 and 156 for class 1.

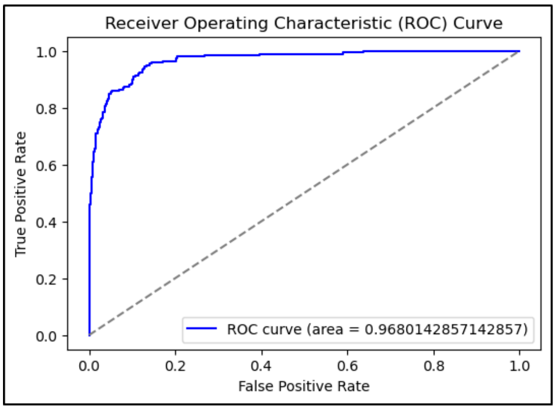


Figure 17: *ROC curve*

Figure 17 shows the ROC curve for proposed work. The suggested Bio-Inspired SNN continues to have the highest AUC value of 0.99 for WQ demonstrating the approach.

4.2 Comparison for the Proposed Method

The comparison of proposed method with existing method is presented in table 1. The CATBoost have the classification accuracy of

90.5% [13] and LightGBM have the accuracy of 91.08% [14]. The proposed Bio-Inspired SNN have the higher accuracy of 93% compared to the existing method.

Table 1: *Comparison for the Proposed Method*

Study	Method	Accuracy
Nasir *et al* [13]	CATBoost	90.5%
Zegaar *et al.*[14]	LightGBM	91.08%
Proposed approach	Bio-Inspired SNN	93%

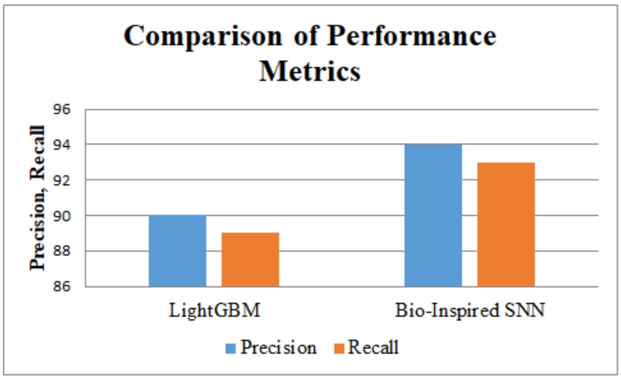


Figure 18: *Comparison of performance metrics*

Figure 18 shows the comparison for Precision and recall value for proposed work. The LightGBM have the precision value of 89% and recall value of 90% [14] and proposed Bio-Inspired SNN have the precision value of 94% and recall of 93% for accessing the WQ.

5. Conclusion

In this paper, WQ data is classified and identified using the proposed Bio-Inspired SNN method. The data preprocessing techniques using data cleaning and data imputation, it effectively finds the missing values and outliers by data cleaning, finds the duplicate WQ values by data imputation. By data normalization and data transformation, the WQ data is effectively normalized and transferred. Finally, the proposed Bio-Inspired SNN to improve the network's generalization and produce higher accuracy. The performance evaluation, conducted on the WQ dataset, demonstrates the superior capabilities of the Bio-Inspired SNN model compared to existing state-of-the-art methods. By using Python software the proposed Bio-Inspired SNN model, it improved the high accuracy of 93% compared to the existing method.

References

1. Zhaoshi Wu; Xijun Lai; Kuanyi Li, Year: 2021, "Water quality assessment of rivers in Lake Chaohu Basin (China) using water quality index", Ecological Indicators, Vol: 121, pp. 107021.

2. Ali El Bilali; Abdeslam Taleb; Youssef Brouziyne, Year: 2021, "Groundwater quality forecasting using machine learning algorithms for irrigation purposes", Agricultural Water Management, Vol: 245, pp. 106625.

3. Saber Kouadri; Ahmed Elbeltagi; Abu Reza Md Towfiqul Islam; Samir Kateb, Year: 2021, "Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast)", Applied Water Science, Vol: 11, No: 12, pp. 190.

4. Segun O. Olatinwo; Trudi-H. Joubert, Year: 2024, "Water quality assessment tool for on-site water quality monitoring", IEEE Sensors Journal, Vol: 24, No: 10, pp. 16450– 16466.

5. Qingqing Zhang; Xue-yi You, Year: 2024, "Recent advances in surface water quality prediction using artificial intelligence models", Water Resources Management, Vol: 38, No: 1, pp. 235– 250.

6. Javed Mallick; Saeed Alqadhi; Hoang Thi Hang; Majed Alsubih, Year: 2024, "Interpreting optimised data-driven solution with explainable artificial intelligence (XAI) for water quality assessment for better decision-making in pollution management", Environmental Science and Pollution Research, Vol: 31, No: 30, pp. 42948– 42969.

7. Angela Gorgoglione; Alberto Castro; Christian Chreties; Lorena Etcheverry, Year: 2020,

"Overcoming data scarcity in earth science" Data, Vol: 5, No. 1, pp. 5.

8. Jaegyu Shim; Seokmin Hong; Jiye Lee; Seungyong Lee; Young Mo Kim; Kangmin Chon, Sanghun Park; Kyung Hwa Cho, Year: 2023, "Deep learning with data preprocessing methods for water quality prediction in ultrafiltration", Journal of Cleaner Production, Vol: 428, pp. 139217.

9. Chenguang Song; Leihua Yao, Year: 2022, "Application of artificial intelligence based on synchrosqueezed wavelet transform and improved deep extreme learning machine in water quality prediction", Environmental Science and Pollution Research, Vol: 29, No: 25, pp. 38066– 38082.

10. Md Saikat Islam Khan; Nazrul Islam; Jia Uddin; Sifatul Islam; Mostofa Kamal Nasir, "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach", Journal of King Saud University-Computer and Information Sciences, Vol: 34, No: 8, pp. 4773– 4781.

11. L. Díaz-González; R. A. Aguilar-Rodríguez; Julio Cesar Pérez-Sansalvador; N. Lakouari, "AQuA-P: A machine learning-based tool for water quality assessment" Journal of Contaminant Hydrology, Vol: 269, pp. 104498.

12. Md Galal Uddin; Stephen Nash; Mir Talas Mahammad Diganta; Azizur Rahman; Agnieszka I. Olbert, Year: 2022, "Robust machine learning algorithms for predicting coastal water quality index", Journal of Environmental Management, Vol: 321, pp. 115923.

13. Nida Nasir; Afreen Kansal; Omar Alshaltone; Feras Barneih; Mustafa Sameer; Abdallah Shanableh; Ahmed Al-Shamma'a, Year: 2022, "Water quality classification using machine learning algorithms", Journal of Water Process Engineering, Vol: 48, pp. 102920.

14. Aymen Zegaar; Samira Ounoki; Abdelmoutia Telli, Year: 2024, "Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process", Water Resources Management, Vol: 38, No: 2, pp. 621– 637.

15. Han Zhang; Xingnian Ren; Sikai Chen; Guoqiang Xie; Yuansi Hu; Dongdong Gao; Xiaogang Tian; Jie Xiao; Haoyu Wang, Year: 2024, "Deep optimization of water quality index and positive matrix factorization models for water quality evaluation and pollution source apportionment using a random forest model", Environmental Pollution, Vol: 347, pp. 123771.

16. Mosleh Hmoud Al-Adhaileh; Fawaz Waselallah Alsaade, Year: 2021, "Modelling and prediction of water quality by using artificial intelligence", Sustainability, Vol: 13, No: 8, pp. 4259.

17. Junhao Wu; Zhaocai Wang, Year: 2022, "A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory", Water, Vol: 14, No: 4, pp. 610.

18. Wafa F. Alfwzan; Mahmoud M. Selim; Saad Althobaiti; Amira M. Hussin, Year: 2023, "Application of Bi-LSTM method for groundwater quality assessment through water quality indices", Journal of Water Process Engineering, Vol: 53, pp. 103889.

19. C. H. Pérez-Beltrán; A. D. Robles; Nicolás Artemio Rodríguez; F. Ortega-Gavilán; A. M. Jiménez-Carvelo, Year: 2024, "Artificial intelligence and water quality: From drinking water to wastewater", TrAC Trends in Analytical Chemistry, Vol: 172, pp. 117597.

20. Peda Gopi Arepalli; K. Jairam Naik, Year: 2024, "An IoT based smart water quality assessment framework for aqua-ponds management using Dilated Spatial-temporal Convolution Neural Network (DSTCNN)", Aquacultural Engineering, Vol: 104, pp. 102373.