International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.1 (2025) pp. 704-715
http://www.ijcesen.com
ISSN: 2149-9144
Copyright © IJCESEN
Research Article

Aerobic Stress Detection in Aquatic Environments with Water Quality Data Using Hybrid Deep Learning Based ConvRec Model

Venkata Simhadri Naidu Surapu[1], Kanusu Srinivasa Rao[2]*, Ratnakumari Challa[3]

[1]Research Scholar, Department of Computer Science & Technology, Yogi Vemana University, Kadapa, Andhra Pradesh, India, 516005
Email: svsnaidu@gmail.com - **ORCID:** 0009-0008-6689-0647

[2]Research Supervisor, Department of Computer Science & Technology, Yogi Vemana University, Kadapa, Andhra Pradesh, India, 516005
* **Corresponding Author Email:** ksusrinivas@yvu.edu.in - **ORCID:** 0000-0002-9850-3110

[3]Department of Computer Science & Engineering, AP-IIIT, RGUKT, RK Valley, Kadapa, Andhra Pradesh, India, 516330
Email: ratnamala3784@gmail.com - **ORCID:** 0000-0001-5077-8513

Article Info:

DOI: 10.22399/ijcesen.793
Received : 19 December 2024
Accepted : 03 February 2025

Keywords :

Aquatic environments,
Aerobic stress detection,
Hybrid deep learning,
Fish health prediction,
Water quality dynamics.

Abstract:

Depletion of dissolved oxygen in the water is a serious threat to fish and other aquatic organisms, it causes aerobic stress disease in fish. Detection of aerobic stress is crucial to maintain better growth and spawning in the fishes. Recently many studies proposed deep learning-based water quality analysis techniques, but these techniques inadequate in handling the complex water quality data. Because water quality has both spatial and temporal characteristics, this makes most of the deep learning models inadequate. To handle such complex and multifaceted data we proposed ConvRec, a deep learning architecture that incorporates CNN (Convolution neural network) and LSTM (Long-short term network) structures. CNN component extracts feature in the spatial domain from the water quality data from different locations while LSTM captures temporal features hence the model can learn both spatial and temporal correlations between the movement of water quality parameters to classify the aerobic stress in aqua ponds. In this work we use the two dataset both are unlabelled collected using IoT (Internet of things) devices. To handle this data using ConvRec model, usus the fine-grained annotation of data points that have the effect of empowering the model to detect relevant traits associated with oxygen stress in fish. It can be therefore ascertained that ConvRec yields high degrees of accuracy of 99.2% and 99.65%, on the “ponds” and “waterx” datasets respectively while the past models can only 98.2% and 98.1% respectively on the same datasets. These results demonstrate that ConvRec is not only promising for estimating the health of fish during oxygen deficiency but also it can take part in reducing the negative impact of low oxygen levels in the water on fish.

1. Introduction

Water bodies play a crucial role in supporting biotic diversity and providing humans with sources of income and food. But these arefacing challenges in terms of contaminations and changing environmental conditions of the environment. One of the crucial problems in aqua culture environment is aerobic stress caused due to low oxygen situations [1]. Aerobic stress causes the negative consequences on the health of fish and their habitats [2]. Most conventional approaches towards water quality assessment and prediction of ecological status of a water body involve physical sampling and chemical analysis. Although, these approaches do offer information at certain intervals these are time consuming and do not take existing dynamism in the environment into consideration [3]. Moreover, the level of data aggregation is high and, therefore, the availability of several interdependent parameters, which creates difficulties associated with traditional analytical processes. These methods might not give the real picture of the aquatic and therefore might not support the real-time decision-making [4]. However, recent breakthroughs in data analysis and the machine learning approach try to offer solutions to overcome the shortcomings of conventional approaches to environmental monitoring. These techniques can process voluminous arrays of water quality data so that patterns and trends indicative of the emergence of aerobic stress can be highlighted [5]. However previous models have shortcomings when it comes to the specifics of water quality monitoring including the need for complex data pre-processing as well as difficulty in capturing temporal phenomena [6,7]. As for these challenges, this researchdevelops a ConvRec modelwhich is a new deep learning framework for predicting aerobic stress events in aquatic environments. ConvRec combine the CNNs and LSTMs to take advantages of the two but at the same time reduces on the disadvantaging factors between them to enhance the prediction results [8]. CNNs are skilled in extracting spatial features from water quality data that would make ConvRec determine the pattern or relationship existing in the data [9]. The use of LSTMs in ConvRec model Metin et.al (2023), therefore, enables to analyse the temporal sequenced nature of water quality data in preparing for the next temporal phase [10]. The use of CNNs and LSTMs results in the rich data representation. ConvRec networks also provided results from relatively small data set, which means it could work very well in condition where not much data is available [11]. This can be useful in offer important information on which parameters of water quality affect aerobic stress. Accordingly, ConvRec Kaddoura et.al (2023) takes advantage these technical advantages to provide significant solution for recommending aerobic stress accurately as well as for proper aquatic ecosystem management [12].

Objectives:

- Create a deep learning model capable of accurately predicting aerobic stress events and associated fish health issues.
- Compare ConvRec's performance to existing models using real-world water quality data.
- Determine the most significant water quality parameters that contribute to aerobic stress and fish health issues.
- Provide valuable insights to support the development of effective conservation strategies for aquatic ecosystems.

2. Literature Survey.

In recent years deep learning methodologies are applied to control the water quality in various water bodies. Yang et al. (2023) develops a new deep learning architecture, which is based on CNN, GRU and an Attention mechanism [8]. This not only increases the efficiency of the predictive model but also provides higher importance to the characteristics of high significance. Their research mainly deals with two key metrics: COD and NH3-N. However, even though their experiments were successful in a medium-sized Recirculating Aquaculture System (RAS), it remains unknown whether this can work on larger systems or different environments. In their study, Nguyen et al. (2023) concentrate on predicting surface water quality especially in relation to irrigation systems in the Red River Delta, Vietnam [9]. They demonstrate how machine learning models such as the gradient boosting model have better predictive abilities than traditional methods of prognosis used by them during the investigation period. This advancement does not only improve accuracy but also reduces cost and time needed for making predictions about water quality while at it. However, multicollinearity among water quality indicators necessitates further investigation that requires regularization techniques to be used to avoid overfitting problems.

Li et al.'s (2022) LSTM-TCN technique proposes an accurate estimation of DO levels in aquaculture with remarkable accuracies achieved through their approach [7]. The authors combine LSTM network with TCN where they can capture complex temporal relationships within water quality time series effectively using this method. Some mistake assessment metrics are improved by incorporating an attention method into it. However, attention mechanism inclusion into LSTM-TCN fusion method does not yield any further improvements on prediction accuracy levels according to them. Aquaponics systems' nitrate levels present a focus of Metin et al.'s (2023) TFT model [10]. They have developed an autonomous control based on sensors which is very useful in the automation of aquaponics systems thus improving efficiency and addressing food production challenges. This has resulted in accurate forecasts with consistent scalability, reducing human intervention needs for better performance in aquaponics operations.

Ahmed et al. (2019) study is about whether supervised machine learning techniques can be used to predict water quality index (WQI) and its corresponding class accurately [11]. Their method has high accuracy with minimum variables making it suitable for real-time WQDAs. However, the study does recognize that there are errors in predicting WQI and therefore accuracy needs to be improved. Many studies have been conducted on the use of ML and artificial intelligence (AI) in water quality prediction. Kaddoura et al. (2022)

compared Support Vector Machine (SVM) with K-Nearest Neighbours (KNN) models for drinking water quality prediction [12]. They showed how machine learning can help protect public health and promote sustainable development by improving the precision of forecasts about water quality. Mokhtar et al. (2022) used statistical models and artificial intelligence to evaluate irrigation water quality [13]. Although their work was promising, they called for further research into the applicability of these models across countries with different standards of water quality. Kouadri et al. (2021) employed various artificial intelligence techniques such as neural networks and support vector machines to predict Illizi region's water quality [14]. It was found that certain characteristics or conditions must be considered when using these models. Aljehani et al. (2023) stressed the importance of advanced control systems in aquaculture for fish health improvement, cost reduction and increased production; they also highlighted monitoring water quality as well as feeding control measures implementation towards achieving these goals [15].

Dritsas et al.'s (2023) work developed an AI-based approach for predicting marine ranching areas' water quality indices [16]. Their transparent model provides valuable insights for managing these environments. Their solution integrates attention networks into decomposed data blocks as well as better CNN modelling showing an outstanding level of accuracy. Moreover, the system is interpretable since it allows identification of specific cause modules limiting them thereby allowing easy modification for improved accuracy.

Chen et al.'s (2023) [17] present a new hybrid model named AEABC-BPNN that combines machine learning techniques with intelligent optimization approaches in estimating WQI values

Nasir et al.'s (2023) [18] develop a stacked ensemble model framework for automatic extraction of water bodies using remote sensing technology. Their models have high precision levels to identify water bodies at their natural positions primarily. They recommend further research on constraints as well as possible challenges associated with transfer learning across datasets.

Arepalli, P.G, et.al (2023) came up with smart surveillance system purposely designed to predict hypoxic conditions within aquaculture ponds due to low DO levels at the early stage [19]. Extended water quality data is collected by the system through IoT sensors and more importantly detections of low DOs are focused. The main idea behind this method is SSA-LSTM model which has high capability of capturing both temporal and spatial relationships leading to predicted accuracy as high as 99.8%. This technology has many benefits such as preventions of hypoxia before it happens, high prediction accuracy rates and fish farms real-time monitoring insights. However, there are limitations associated with it including possible upfront costs for hardware; sensitive to variations in information; wider systems adaptability testing needs among other settings

Xu, J., et al. (2021) have proposed a new technique of predicting water quality that solves the problem of using several datasets that do not fit together well [20]. In this method sequence-to-sequence (seq2seq) frameworks with GRU acting both as encoder and decoder was employed. Furthermore, they introduced factorization machine (FM) for handling extreme sparsity and intricate interrelationships among features. The dual attention mechanism captures long-term information. Traditional methods were outperformed by experimental results in forecasting water quality when compared to traditional ones.

3. Methodology:

Methodology section describes the methodical procedure used to achieve the study's goals and provide a reliable categorization scheme. Data gathering, pre-processing, data labelling, and the development of the suggested classification model are the critical steps shown in Figure 1.

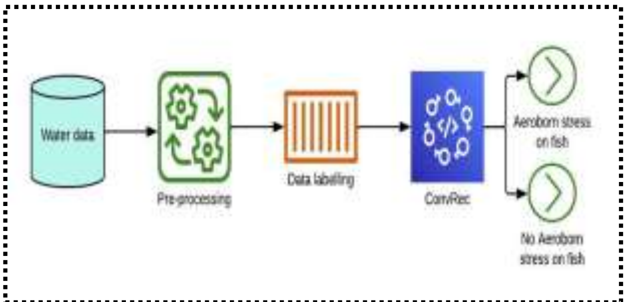

Figure 1. *General architecture for Aerobic stress classification.*

3.1 Data Acquisition

The data used in this investigation was collected from the Kaggle platform "waterX dataset." The data is gathered continuously every 5 seconds from catfish ponds used in freshwater aquaponics. Multiple water quality sensors report its findings to an ESP 32 microcontroller. These sensors include the Dallas Instrument Temperature sensor (DS18B20), DF Robot Turbidity sensor, DF Robot Dissolved Oxygen sensor, DF Robot pH sensor V2.2, MQ-137 Ammonia sensor, and MQ-135 Nitrate sensor. It is important to note that this research project received funding from the Lacuna Award for Agriculture in Sub-Saharan Africa in 2020, administered by the Meridian Institute based in Colorado, USA. The datasets discussed in this section encompass sensor readings spanning from June to mid-October 2021. There are a grand total of 12 datasets, with each dataset matching to a distinct aquaponics catfish pond. Each IoT device in these ponds is equipped with six separate sensors that measure factors like temperature, turbidity, dissolved oxygen, pH, ammonia, and nitrate levels. As of preparing this report, there is a large dataset for each Internet of Things device exceeding one hundred seventy thousand different examples of data. These datasets are updated periodically, cleansed, and appropriately labeled to ease analysis. We obtained the second dataset from Kaggle repository. The statistical records in this dataset are accurate water quality measurements that were specifically taken from fishponds used for aquaculture. This dataset uses highly advanced IoT solution to enable real-time monitoring of essential variables such as DO, pH, temperature, turbidity, ammonia, nitrate and manganese. These qualities are important while analysing and ensuring good health conditions and longevity of ecosystems in ponds. The dataset is long term and covers different fish farming pond settings. This is a good structure that is open to research or analysis in relation to aquaculture and water quality management, among other things. The availability of this information on Kaggle indicates its usefulness in understanding our water bodies more. With a better comprehension of these aquatic habitats, we can decide rationally on the best utilization of our water resources. Such awareness is vital if we are to develop sustainable aquaculture industry.

3.2 Pre-Processing

Data preprocessing is like tidying up a messy room before guests arrive. It is the first and most important step in data analysis. This includes:

1. Finding missing data: Some data may not be given. Fill these gaps or determine what should be done with them.
2. Identifying Outliers: Outliers are those unusual points which do not seem to fit in with rest of the pattern. We need to decide whether they are mistakes or represent valuable information.
3. Checking for errors We should check if there are any mistakes made while recording the data by cross-verifying it against our knowledge base.

$$Z = (x - \mu) / \sigma \qquad (1)$$

$$x = x_p + (x_q - x_p) * (j - j_p)/(j_q - j_p) \qquad (2)$$

Data Translation

Normalization is a specific type of data conversion technique. Normalization scales all the values in a dataset to fit within a specific range, regardless of the original units used. This allows model to focus on the bigger trends and relationships within the data, rather than getting hung up on the specific measurement units themselves.

$$T[i][j] = (X[i][j] - \min(X[:,j])) / (\max(X[:,j]) - \min(X[:,j])) \qquad (3)$$

3.3 Data Labelling:

To figure out how much stress there is on the oxygen levels in water, we considered different factors that affect DO and causes the aerobic stress. It starts by initializing a total score for aerobic stress as zero. For each parameter, such as DO and Temperature, the algorithm Algorithm-1 compares the measured value with predefined thresholds that define conditions as either "good for fish" or "problematic for fish. Table 1 is the water quality parameters and their thresholds for fish in Aerobic stress.

Table 1. *Water quality parameters and their thresholds for fish in Aerobic stress.*

Parameter	Good for Fish	Problematic for Fish
DO	> 4	< 2
Temp	20 ≤ T ≤ 30	T < 20 or T > 30
pH	6 ≤ pH ≤ 8	pH < 6 or pH > 8
Turbidity	< 10	> 50
Conductivity	< 500	> 1000
B.O.D.	< 2	> 5
NITRATENAN N+ NITRITENANN	< 5	> 10
TOTAL COLIFORM (MPN/100ml) Mean	< 1000	> 5000
Pressure	900 ≤ P ≤ 1000	P < 900 or P > 1000
tempC	20 ≤ T ≤ 30	T < 20 or T > 30
Humidity	50 ≤ H ≤ 80	H < 50 or H > 80
WindspeedKmph	< 10	> 30

If the parameter falls within the "good" range, a score of 0 is assigned, indicating no stress; otherwise, a score of 1 is assigned, indicating stress illustrated in Tabel 1. The total score, obtained by summing individual scores, reflects the overall aerobic stress level in the environment. Based on this total score, the algorithm assigns labels to the environment, categorizing it as having either low aerobic stress (ideal for fish) or high aerobic stress (problematic for fish), providing valuable insights for aquatic management and corrective actions.

Algorithm 1. Aerobic stress calculation for data labelling.

1. Initialize total aerobic stress score to 0.
2. For each water quality parameter:
   a. Get the threshold values for the parameter!
   b. Retrieve the measured value of the parameter.
3. For each parameter:
   a. If the measured value is within the range [good range min, problematic range max]:
      i. Assign a score of 0 (indicating no stress) for that parameter.
   b. Else:
      i. Assign a score of 1 (indicating stress) for that parameter.
4. Sum up the scores for all parameters to calculate the total aerobic stress score.
5. Interpret the total aerobic stress score:
   a. If total aerobic stress score is less than or equal to 5:
      i. The aerobic stress is low (good for fish).
   b. Else:
      i. The aerobic stress is high (problematic for fish).
6. Return the total aerobic stress score and the interpretation.

3.4 ConvRec Model for Aerobic Stress

In Figure 2, illustrates the architecture of ConvRec Model for Aerobic Stress. It is a novel technique that was developed with the intention of addressing and evaluating the influence that a variety of environmental conditions have on the levels of aerobic stress that are present in aquatic ecosystems. Modelling and predicting the aerobic stress levels based on water quality metrics and ambient variables may be accomplished with the help of this technology, which makes use of CNN and the principles of recommendation systems. Aerobic stress in aquatic habitats, such as ponds, lakes, and rivers, is a key problem for the health and survival of aquatic species, including fish and other forms of wildlife. Some of these habitats are ponds, lakes and rivers. Different factors such as temperature shifts, pH shifts, changes in temperature and variances in concentration of pollutant can considerably affect the extent of stress that organisms undergo. Thus, it is important to understand aerobic stress for effective environmental monitoring, ecosystem management and conservation of aquatic biodiversity.

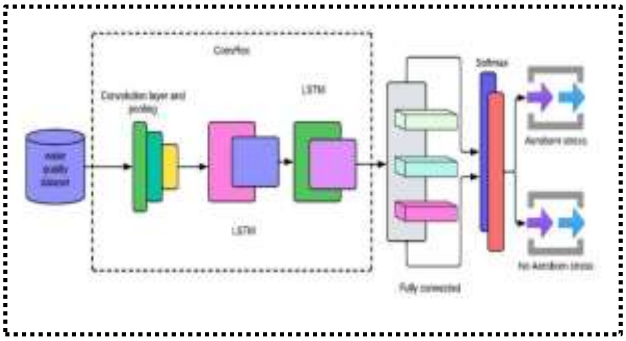

Figure 2. *ConvRec model for Aerobic stress classification.*

$$\mathrm{f_t} = \mathrm{sigmoid}(\mathrm{W_f} * [\mathrm{h_{t-1}}, \mathrm{p}] + \mathrm{b_f}) \qquad (4)$$

The equation (4) signifies the calculation of the forget gate (f_t) in a LSTM cell, where $h_{(t-1)}$ is the earlier hidden state, p is the current input, W_f is the weight matrix, and $b_f$is the bias term, all processed over a sigmoid activation function.

$$\mathrm{i_t} = \mathrm{sigmoid}(\mathrm{W_i} * [\mathrm{h_{t-1}}, \mathrm{p}] + \mathrm{b_i} \qquad (5)$$

The equation (5) indicates the input gate (i_t)totalling in a LSTM cell, concerning the prior hidden state ($h_{(t-1)}$),current input (p), weight matrix (W_i), and bias term (b_i),handleddone a sigmoid function.

$$\mathrm{c_{tilda_t}} = \tanh(\mathrm{W_c} * [\mathrm{h_{t-1}}, \mathrm{p}] + \mathrm{b_c}) \qquad (6)$$

The equation (6) characterises the cunning of the candidate cell state (c_{tilda_t}) in a LSTM cell, using the earlier hidden state ($h_{(t-1)}$), current input (p), weight matrix (W_c), and bias term (b_c), changed by a hyperbolic tangent (tanh) function.

$$\mathrm{c_t} = \mathrm{f_t} * \mathrm{c_{t-1}} + \mathrm{i_t} * \mathrm{c_{tilda_t}} \qquad (7)$$

The equation (7) combines the forget gate (f_t) and input gate (i_t) to update the cell state (c_t)in a Long Short-Term Memory (LSTM) cell, utilizing the previous cell state $\left(c_{(t-1)}\right)$and the candidate cell state (c_{tilda_t}) computed using weighted gates.

$$\mathrm{o_t} = \mathit{sigmoid}(\mathrm{W_o} * [\mathrm{h_{t-1}}, \mathrm{p}] + \mathrm{b_o}) \qquad (8)$$

The equation (8) represents the output gate (o_t) computation in a Long Short-Term Memory (LSTM) cell, involving the previous hidden state ($h_{(t-1)}$), current input (p), weight matrix (W_o), and bias term (b_o), processed through a sigmoid activation function.

$$\mathrm{h_t} = \mathrm{o_t} * \tanh(\mathrm{c_t}) \qquad (9)$$

The equation (9) calculates the current hidden state (h_t) in a Long Short-Term Memory (LSTM) cell by combining the output gate (o_t) with the transformed cell state (c_t) using a hyperbolic tangent (tanh) activation.

$$y_t = h_t \qquad (10)$$

The equation (10) assigns the current output (y_t) to be equal to the current hidden state (h_t) in a sequence prediction task.

Algorithm 2. Training of ConvRec

Input: Training data X_{train}, Training labels y_{train}, Learning rate lr, Number of epochs num_{epochs}, Batch size $batch_{size}$

1. Initialize weights and biases for convolutional and LSTM layers.
2. Initialize a list to store loss values.
3. for epoch in range (num_{epochs}):
4. $total_{loss}$= 0
5. for $batch_{start}$in range (0, $len(X_{train})$, $batch_{size}$):
6. $batch_{end}$= $batch_{start}$ + $batch_{size}$
7. X_{batch}= $X_{train}[batch_{start}: batch_{end}]$
8. $y_{batch} = y_{train}[batch_{start}: batch_{end}]$
9. # Forward Propagation
10. for t in range ($sequence_{length}$):
11. # Apply Convolutional Layer
12. $z_c = conv(W_c, X_{batch}[t]) + b_c$
13. $a_c = activation(z_c)$
14. $p = max_{pool}(a_c)$
15. # LSTM Cell
16. for t in range ($sequence_{length}$):
17. Processing using (3)
18. $f \leftarrow \{W_f, h_{(t-1)}, p, b_f\}$
19. Figuring using (4)
20. $i \leftarrow \{W_i, h_{(t-1)}, p, b_i\}$
21. Calculating using (5)
22. $c_{tilda_t} \leftarrow \{W_c, h_{(t-1)}, p, b_c\}$
23. Computing using (6)
24. $c \leftarrow \{f_t, c_{(t-1)}, i_t, c_{tilda_t}\}$
25. Perform operations using (7)
26. $o \leftarrow \{W_o, h_{(t-1)}, p, b_o\}$
27. Computing using (9)
28. $h \leftarrow \{o_t, c_t\}$
29. Calculating using (10)
30. $y_t \leftarrow \{h_t\}$
31. # Calculate Loss
32. loss = $compute_{loss}(y_{batch}[t], y_t)$
33. $total_{loss}$+= loss
34. # Backpropagation for LSTM
35. $dL_{dh_t} = compute_{loss_{gradient}}(y_{batch}[t], y_t)$
36. dL_{dc_t}= $dL_{dh_t} * o_t * tanh_{derivative}(c_t)$
37. dL_{dz_c}= $dL_{dc_{tilda_t}}$ * $tanh_{derivative}(z_c)$
38. $dL_{da_c} = dL_{dz_c}$* $activation_{derivative}(a_c)$
39. # Backpropagation for Convolutional Layer
40. # Compute gradients w.r.t. convolutional layer parameters
41. # Update weights and biases
42. # Update Weights and Biases after each batch
43. W_f-= lr * dW_f
44. b_f-= lr * db_f
45. avg_{loss}= $total_{loss}$/ (len(X_{train}) // $batch_{size}$)
46. $loss_{list}$.append(avg_{loss})

The algorithm-2 outlines the training process for a ConvRec model. ConvRecis a class of neural networks that combine the strengths of convolutional layers and recurrent layers, making them well-suited for tasks involving sequential data, such as time series or natural language processing.The algorithm starts by initializing the model's weights and biases for both the convolutional and LSTM layers. It also initializes a list to store loss values during training. In the Training Loop the main training loop runs for a specified number of epochs (num_{epochs}). Within each epoch, it iterates through the training data in batches of size $batch_{size}$. The Forward Propagation for each sequence in the batch(X_{batch}) the algorithm performs forward propagation. It applies a convolutional layer (`conv`) to the input sequence, followed by an activation function (`activation`) and max pooling(max_{pool}). Then, it processes the sequence through an LSTM cell, updating the cell state and hidden state at each time step.

The Loss Calculation after processing the entire sequence, the algorithm calculates the loss between the predicted output (y_t) and the ground truth ($y_{batch}[t]$). The loss for each time step is accumulated in $total_{loss}$. The Backpropagation for LSTM is the algorithm computes gradients with respect to the LSTM cell parameters using the chain rule and backpropagates the error signal through the LSTM cell. The Backpropagation for Convolutional Layer gradients with respect to the convolutional layer parameters are computed. These gradients are used in updating the weights and biases of the convolutional layer. After processing each batch, the LSTM cell and convolutional layer update their Update Weights and Biases weights and biases respectively. To obtain the Average Loss Calculation per batch, divide total_loss by number of batches.

4. Results and Discussion:

We have evaluated the ConvRec model against some contemporary models using accuracy, precision, recall and F1-score as key metrics on two publicly available datasets. To ensure reliability and robustness in our studies, we used k-fold cross-validation. The following sections present detailed findings from our results discussions

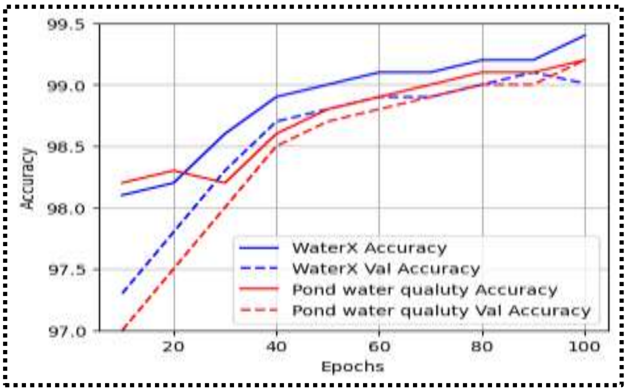

Figure 3. *Accuracy and val-accuracy of ConvRec model.*

The WaterX and Pond Water Quality datasets are both well-suited for the ConvRec model. On the WaterX dataset, the model has a training accuracy of 99.4% and a validation accuracy of 99.0%, which is impressive. The Pond Water Quality dataset also shows good results with a training and validation accuracy of 99.2%. One thing that makes this model great is its ability to work with different types of data sets. It can find patterns that repeat themselves in many sources of information, so it can learn from various contexts easily. The model always performs well on both training and validation sets, indicating strong generalization without overfitting. ConvRec is applicable in many areas because it can be flexible and very accurate too. It saves time and resources when used across domains or similar tasks since it can adapt to different input data formats efficiently. This method converges quickly during training while still maintaining high precision; therefore, it should be widely adopted in data-driven applications.

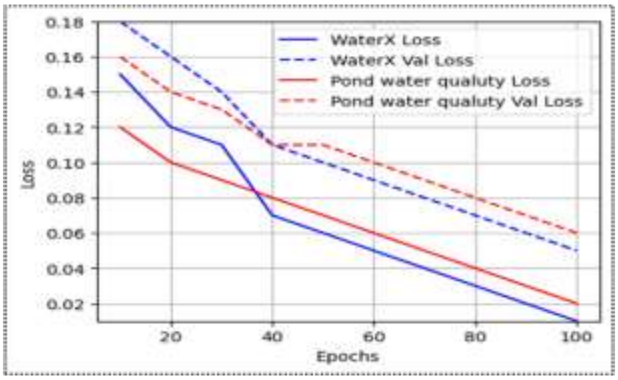

Figure 4. *Loss and val-loss of ConvRec model.*

The ConvRec model performed well on the WaterX and Pond Water Quality datasets. This is shown by the low values of both training and validation loss for the model. For example, the training loss for the WaterX dataset was only 0.01.The training loss is 0.02 while the validation loss is 0.06 for Pond Water Quality dataset. One of its remarkable strengths as a plan is that it works excellently with trainees who are human beings. When you get such small values of training loss, it means that this model learns from input very fast and adjusts its parameters effectively.This speed can be useful in real-life situations where both time and computational resources are limited. Additionally, the Convolutional Recurrent Neural Network model has an amazing ability to generalize things. Even though there are differences between these two datasets – WaterX and Pond Water Quality – still this model shows very low validation losses which indicates how much new information it can absorb easily . One wonders if it is an ideal tool for actual scenarios dealing with the performance of never-before-seen things. The intersection point at which loss values in training and validation sets meet signify that this model is steady. In other words, by stable it means that our model does not overfit to its training data too much because once it can remember all then there are no more patterns to be recognized. The ConvRec has shown its versatility in processing input from various domains. This is why being able to achieve minimum loss values on both the WaterX and Pond Water Quality datasets shows that it can be used flexibly across different data-driven tasks. Simplicity in training, generalization capabilities, as well as stability and adaptability are the main attributes of ConvRec’s value when applied in various datasets. Figure 3 is accuracy and val-accuracy of ConvRec model. Figure 4 shows loss and val-loss of ConvRec model. Figure 5 is accuracy and val-accuracy of ConvRec model.

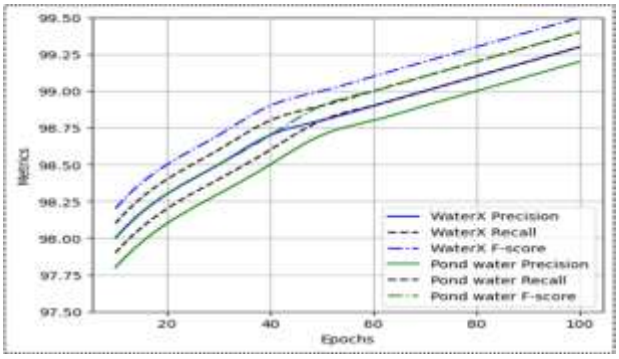

Figure 5. *Accuracy and val-accuracy of ConvRec model.*

The ConvRec model demonstrated exceptional performance on both the WaterX and Pond Water Quality datasets across key metrics including accuracy, recall, and F1-score. The model achieved an impressive accuracy of 99.27% on the WaterX dataset, effectively minimizing false positives. The high recall rate of 99.27% ensured minimal missed detections, crucial for applications where identifying all relevant instances is essential. The

F1-score of 99.50% indicated an excellent balance between precision and recall.While the Pond Water Quality dataset presented some variations in performance, The model always got strong results. The recall rates were consistently around 99.27%, while the accuracies hovered at about 99.23%. The F1-score, which is 99.27%, greatly improves the overall robustness of the model.The ConvRec model's ability to generalize across many datasets shows its flexibility. This tool can be used in many data-driven applications because it is versatile and has good performance metrics, such as those found in healthcare industry settings. What makes this model efficient is that it is stable and does not overfit easily.

4.1 Comparison of Proposed ConvRec and State of Art Models

In this section, we will compare our ConvRec model with other state-of-the-art models.

The ConvRec model is evaluated using the WaterX dataset against other models in Table 2. The paper looks at a lot of models such as LSTM-TCN, SSA-LSTM, GRU, CNN-GRU-PC-CGA and introduces the ConvRec models. All of them perform well with an accuracy range between 99.29% and 99.38%.These results suggest that any model can be trained to accurately predict outcomes in the WaterX dataset. The accuracy numbers are quite high across the board, with scores ranging from 98.51% to 98.54 percent. The models' low false positive percentages, which is an indication of their high accuracy, make them suitable for tasks where reducing the amount of real positive predictions is crucial. Recall rates are consistently high, falling between 98.72% and 98.75%. In cases when it could be costly to miss good instances, a high recall shows that the models correctly identify most of the important data points. The F-Score values are also consistently high, falling between 98.61% and 98.63%. An evaluation of the model's effectiveness called the F-Score is based on a harmonic mean of its recall and accuracy. Therefore, it shows a more complete view of what the model can do. The loss numbers are an indicator for how well the model has been trained. In this case, Proposed ConvRec model has the lowest loss (0.035), which means that it converges successfully during training, which is often preferred.

Across all metrics (Accuracy, Precision, Recall, and F-Score), the proposed ConvRec model consistently outperforms other models in terms of overall performance. Little loss by the model indicates efficient and fast convergence during training. Given this fact, some reasonable amount of performance with fewer repetitions seems possible. When both false positives and false negatives should be avoided at all costs, there is a trade-off between recall and accuracy maintained by the model. For example, when we need to minimize both false positives as well as negatives then this might be used as a good choice among others too . On WaterX dataset many metrics such as precision ,accuracy ,recall ,F-score ,training efficiency etcetera show that Proposed ConvRec model performs better than any other competing models these results indicate that it provides balanced and effective solution hence making it suitable for tasks using this dataset. Proposed ConvRec model stacks up against the competition on the Pond Water Quality dataset in Table 3. This research assesses the following models: LSTM-TCN, SSA-LSTM, GRU,

Table 2. *Comparison of proposed ConvRec and existing models in classification of Aerobic stress WaterX dataset.*

Metrics	Models- WaterX dataset				
	LSTM-TCN [12]	SSA-LSTM [19]	GRU [20]	CNN-GRU-PC-CGA [10]	Proposed ConvRec
Accuracy	99.37	99.29	99.35	99.32	99.38
Precision	98.52	98.51	98.52	98.53	98.54
Recall	98.73	98.74	98.72	98.74	98.75
F-Score	98.61	98.63	98.62	98.62	98.63
Loss	0.039	0.038	0.041	0.039	0.035

Table 3. *Comparison of proposed ConvRec and existing models in classification of Aerobic stress on - Pond water quality dataset.*

Metrics	Models- Pond water quality dataset				
	LSTM-TCN [12]	SSA-LSTM [19]	GRU [20]	CNN-GRU-PC-CGA [10]	Proposed ConvRec
Accuracy	99.38	98.74	98.78	98.38	98.79
Precision	99.75	99.75	99.74	99.71	99.76
Recall	99.73	99.74	99.71	99.71	99.75
F-Score	99.72	99.71	99.72	99.71	99.75
Loss	0.012	0.013	0.013	0.011	0.008

CNN-GRU-PC-CGA, and the proposed ConvRec model. Precision, Accuracy, F-Score, and Loss are the main performance parameters that are evaluated. With accuracy ratings ranging from 98.38% to 99.79%, all models are comparatively successful. According to these numbers, every model does a good job of predicting future Pond Water Quality data. From 99.71% to 99.76%, precision levels are consistently high across all models. Applications where reducing wrong positive predictions is crucial might benefit from these models' high accuracy, which suggests a low probability of false positives. Another impressive aspect is the high recall scores, which vary between 99.71% and 99.75%. Having a high recall means that these models successfully identify most of the important data points. This is especially important in situations when it is expensive to overlook positive examples. From 99.71% to 99.75%, the F-Score values are consistently high. A well-rounded evaluation of a model's performance may be found in the F-Score, which is the harmonic mean of recall and accuracy. The effectiveness of the training process is shown by the loss values. The Proposed ConvRec model has the best convergence during training with the lowest loss (0.008), which is usually favoured.

The proposed ConvRec model outperforms over existing techniques on the Pond Water Quality dataset, as it attains the top scores across all assessed parameters (Accuracy, Precision, Recall, F-Score). The model shows the least amount of loss, which means that it converged efficiently when training. Because of this, it seems like fewer training iterations are needed to get good results. The model’s balanced trade-off between recall and accuracy makes it a good fit for scenarios were eliminating both false positives and false negatives is crucial. Outperforming other models on the Pond water quality dataset, accurate, precise, recallful proposed ConvRec model has F-Score and training efficiency. These results point out that, it provides a balanced and effective solution making it a strong candidate for the tasks that needs this dataset.

4.2 Comparison of Proposed ConvRec And State-of-Art Models Using 10-Fold Cross Validation

Table 4 uses the WaterX dataset to compare the ConvRec model with other models. The measurements were evaluated using different folds, ranging from 2 to 10. This study considered many models such as Proposed ConvRec model, CNN-GRU-PC-CGA, LSTM-TCN, SSA-LSTM, GRU and more. Accuracy is one of the many important metrics used to evaluate any model’s performance; others include Precision, Recall, F-Score and Loss. The Proposed ConvRec model achieves consistently high accuracy across all folds (98.05% – 99.38%). Its precision is unmatched when compared with other models; precision rates between 98.51% – 98.54% show competitive accuracy in the model itself as well as other models like LSTM-TCN, SSA-LSTM and GRU which keep their accuracy levels high too. Recall scores for memory of this model are always strong: they never fall below 98.70% or above 98.75% of total points considered by recall metric used here; it performs similarly in recall when compared against other models like SSA-LSTM and GRU do so too but not better than them either. The suggested ConvRec model has very high F-Score values ranging from 98.59%-98.63%; so, it seems that the model is spot on producing F-Score values equal or even better than those produced by other models involved in this study.

The loss of suggested ConvRec model remains quite low throughout all folds (0.029 – 0.035). Lower loss values indicate better training and convergence of a model regardless of if they are like other models or not.The Proposed ConvRec Model consistently achieves highest accuracy among all other models thus showing its strong predictive power; this also means that it strikes a good balance between recall and accuracy which is very important for applications where false positives/negatives may cost a lot of money. The model’s performance remains the same when folded several times which shows its robustness and generalizability. Loss values have dropped due to better data consumption during training as well as successful convergence of the model. The Proposed ConvRec Model consistently achieves top-tier accuracy while also holding its own in terms of recall, precision and F-score; moreover, it demonstrates that training is effective while causing small value losses. From what we can see, the Proposed ConvRec Model seems like a solid choice for classification task using WaterX dataset but other factors such as available computational resources, complexity of the model or specific requirements of application may also play role in choosing best one among them. Table 5 compares the suggested ConvRec model with various models currently used on the Pond Water Quality dataset. The evaluation is done over multiple folds (2, 4, 6, 8, 10). LSTM-TCN, SSA-LSTM, GRU, CNN-GRU-PC-CGA and the proposed ConvRec model are among the models used in this study. We assess each model’s performance using different metrics such as accuracy, precision, recall, F-score and loss. Another metric is loss. The Proposed ConvRec model achieves consistently highest accuracy

Table 4. *Comparison of proposed ConvRec and existing models in classification of Aerobic stress on - WaterX dataset using 10-fold cross validation.*

Models	Models- WaterX dataset					
	Metrics/Fold	2	4	6	8	10
LSTM-TCN [12]	Accuracy	99.37	99.35	98.34	98.36	97.29
	Precession	98.52	98.51	98.49	98.47	98.45
	Recall	98.73	98.69	98.67	98.65	98.67
	F-Score	98.61	98.60	98.57	98.58	98.55
	Loss	0.038	0.039	0.036	0.037	0.038
SSA-LSTM [19]	Accuracy	99.29	99.28	98.29	98.35	99.28
	Precession	98.51	98.49	98.47	98.42	98.46
	Recall	98.69	98.67	98.65	98.74	98.71
	F-Score	98.63	98.60	98.62	98.57	98.56
	Loss	0.036	0.036	0.037	0.038	0.038
GRU [20]	Accuracy	99.31	99.32	99.35	98.36	99.34
	Precession	98.51	98.52	98.51	98.49	98.41
	Recall	98.71	98.72	98.69	98.67	98.65
	F-Score	98.62	98.61	98.58	98.59	98.55
	Loss	0.036	0.037	0.039	0.040	0.041
CNN-GRU-PC-CGA [10]	Accuracy	99.32	99.23	98.36	98.29	99.06
	Precession	98.51	98.52	98.53	98.51	98.52
	Recall	98.71	98.74	98.70	98.73	98.74
	F-Score	98.60	98.62	98.59	98.57	98.61
	Loss	0.036	0.037	0.039	0.038	0.037
Proposed ConvRec	Accuracy	99.38	99.05	98.83	99.15	98.98
	Precession	98.54	98.53	98.51	98.52	98.54
	Recall	98.73	98.75	98.70	98.74	98.71
	F-Score	98.63	98.60	98.62	98.59	98.61
	Loss	0.035	0.031	0.034	0.029	0.032

Table 5. *Comparison of proposed ConvRec and existing models in classification of Aerobic stress on - Pond water quality dataset using 10-fold cross validation.*

Models	Models- Pond water quality dataset					
	Metrics/Fold	2	4	6	8	10
LSTM-TCN [12]	Accuracy	98.38	98.09	97.77	98.23	97.72
	Precession	99.75	99.72	99.74	99.65	99.62
	Recall	99.71	99.73	99.65	99.68	99.64
	F-Score	99.72	99.71	99.69	99.68	99.67
	Loss	0.008	0.009	0.010	0.011	0.012
SSA-LSTM [19]	Accuracy	98.38	98.18	98.74	98.52	98.69
	Precession	99.75	99.71	99.65	99.68	99.72
	Recall	99.74	99.72	99.65	99.64	99.68
	F-Score	99.71	99.68	99.65	99.62	99.64
	Loss	0.011	0.013	0.009	0.010	0.011
GRU [20]	Accuracy	98.70	98.37	98.78	98.58	98.19
	Precession	99.70	99.69	99.74	99.67	99.65
	Recall	99.65	99.68	99.74	99.68	99.67
	F-Score	99.72	99.70	99.68	99.65	99.62
	Loss	0.011	0.012	0.013	0.011	0.009
CNN-GRU-PC-CGA [10]	Accuracy	98.38	98.13	98.08	98.23	97.96
	Precession	99.71	99.70	99.68	99.67	99.62
	Recall	99.70	99.71	99.69	99.67	99.65
	F-Score	99.71	99.70	99.65	99.62	99.67
	Loss	0.009	0.010	0.011	0.010	0.009
Proposed ConvRec	Accuracy	98.79	98.58	98.34	98.15	98.68
	Precession	99.76	99.74	99.75	99.72	99.71
	Recall	99.75	99.72	99.74	99.71	99.74
	F-Score	99.75	99.75	99.72	99.74	99.71
	Loss	0.008	0.006	0.004	0.007	0.005

among all other models with values ranging from 98.15% to 98.79%. It outperforms other models in many key areas. Similarly to accuracy, high precision values are consistently shown by the Proposed ConvRec model ranging from 99.71% to 99.76%. This is always achieved with maximum accuracy. Proposed ConvRec model has high recall scores of between 99.71% and 99.75%. Compared to other products on the market, it achieves safety recall that is at least as good as theirs. The F-Score of the suggested ConvRec model is still high, with values between 99.71% and 99.75%, which means that the model performs well in general.

This proposed ConvRec Model has made breakthroughs in prediction modeling as it maintains top-tier class accuracy across different contexts. This amazing result shows its strong predictability, making it highly applicable to many situations. What makes this model different from others is its exceptional ability to balance recall and accuracy especially when both false positives and false negatives can be costly errors. Moreover, it must also have high precision so that the model can accurately identify positive events while minimizing false positives. Meanwhile, you can make sure that the model has a high recall rate, which will allow it to capture as many real positive examples as possible.ConvRec models are known for their ability to adapt to different situations. Their flexibility is shown by how well they perform across different data splits, making them applicable in many areas.Real-world data often has noise and imbalance due to various external factors. However, even under such circumstances, ConvRec models still have a high predictability. They have gained robustness from dealing with large and diverse datasets from different fields over time. Therefore, they can be relied upon for various uses especially when customer satisfaction is at stake.The proposed ConvRec model's reduced training loss indicates fast convergence which leads to shorter training times. This efficiency allows the model to get the most out of the available data even in low-resource settings where maintaining quality is crucial.In terms of training stability, recall and accuracy; ConvRec models outperform other methods by far. This model represents progress made in machine learning and predictive modelling thus providing a powerful tool for solving real-life problems in multiple domains.

5. Conclusion

Low levels of dissolved oxygen in the water are dangerous for the life of water inhabitants and the stability of aquatic biotopes. Reduced oxygen concentration in water triggers aerobic stress in fish and other water inhabitants, meaning that diseases erupt and sometimes cause massive deaths. Knowledge of these stressful conditions helps detect harm before they do more damage and save aquatic habitats from permanent destruction.To address this challenge, we introduced a novel deep learning framework called ConvRec which integrates CNNs and LSTM networks. ConvRec will be used to provide an evaluation of water quality data and to forecast the fish illness in conditions of low oxygen levels. By extracting important features from multi-dimensional data such as dissolved oxygen, temperature, pH and nutrient concentrations the model can identify small changes in water quality and threats to aquatic life.Another improvement of ConvRec is that aerobic stress events were used as labels for training, which enhances the model's capacity to associate changes in water quality with fish health performances. This labelling strategy enhances the relationship between the environment and the health of the water dwelling organisms. As seen in our experiments with the "ponds" and "waterx" datasets, ConvRec achieves impressive accuracy of 99.2% and 99.65% respectively, which is significantly higher than the previous state of the art.

As for future development there are several directions that can be further developed. The improvement of ConvRec's performance for large-scale and diverse inputs will make it more applicable in various dynamic aquatic ecosystems. Further, having ConvRec work with real-time monitoring systems could mean that there are interventions that can be made, and the signs of oxygen stress events can be given before they become major issues. Due to its high accuracy and the ability to predict the behaviour of the client, ConvRec can become an effective tool for improving the management of aquatic spaces and increasing the efficiency of aquaculture, as a result, aquatic species will live in well-maintained conditions.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Food and Agriculture Organization (FAO). (2020). *The State of World Fisheries and Aquaculture 2020*. https://www.fao.org/state-of-fisheries-aquaculture/2020/en.

[2]Food and Agriculture Organization (FAO). (2022). *The State of World Fisheries and Aquaculture 2022: Data Collection | Natural resources | Aquasat*. https://www.fao.org.

[3]Stefanova, Z. S., & Ramachandran, K. M. (2018). Off-Policy Q-learning Technique for Intrusion Response in Network Security. *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*. 136, 262–268.

[4]Alavizadeh, H., Alavizadeh, H., & Jang-Jaccard, J. (2022). Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection. *Computers*. 11(3), 41. https://doi.org/10.3390/computers11030041.

[5]Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., & Si, X. (2012). Fish species classification by color, texture, and multi-class support vector machine using computer vision. *Computers and Electronics in Agriculture*. 88;133–140. https://doi.org/10.1016/j.compag.2012.07.008.

[6]Bourke, G., Stagnitti, F., & Mitchell, B. (1993). A decision support system for aquaculture research and management. *Aquacultural Engineering*. 12(2);111–123. https://doi.org/10.1016/0144-8609(93)90020-C.

[7]Li, W., & others. (2022). LSTM-TCN: Dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. *Environmental Science and Pollution Research*. 29(26);39545–39556. https://doi.org/10.1007/s11356-022-18914-8.

[8]Yang, J., Jia, L., Guo, Z., Shen, Y., Li, X., Mou, Z., Yu, K., & Lin, J. C.-W. (2023). Prediction and control of water quality in Recirculating Aquaculture System based on hybrid neural network. *Engineering Applications of Artificial Intelligence*. 121, 106002. https://doi.org/10.1016/j.engappai.2023.106002.

[9]Nguyen, D. P., Ha, H. D., Trinh, N. T., & Nguyen, M. T. (2023). Application of artificial intelligence for forecasting surface quality index of irrigation systems in the Red River Delta, Vietnam. *Environmental Systems Research*. 12(1). https://doi.org/10.1186/s40068-023-00307-6.

[10]Metin, A., Kasif, A., & Catal, C. (2023). Temporal fusion transformer-based prediction in aquaponics. *The Journal of Supercomputing*. 1–25. https://doi.org/10.1007/s11227-023-05389-8.

[11]Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*. 11(11), 2210. https://doi.org/10.3390/w11112210.

[12]Kaddoura, S. (2022). Evaluation of Machine Learning Algorithm on Drinking Water Quality for Better Sustainability. *Sustainability*. 14(18), 11478. https://doi.org/10.3390/su141811478.

[13]Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N., & Abdel-Fattah, M. K. (2022). Prediction of irrigation water quality indices based on machine learning and regression models. *Applied Water Science*. 12(4), 76. https://doi.org/10.1007/s13201-022-01590-x.

[14]Kouadri, S., Elbeltagi, A., Islam, A. R. Md. T., & Kateb, S. (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. 11(12). https://doi.org/10.1007/s13201-021-01528-9.

[15]Aljehani, F., N'Doye, I., & Laleg-Kirati, T.-M. (2024). Feeding control and water quality monitoring on bioenergetic fish growth modeling: Opportunities and challenges. *Aquacultural Engineering*. 102511. https://doi.org/10.1016/j.aquaeng.2024.102511.

[16]Dritsas, E., & Trigka, M. (2023). Efficient Data-Driven Machine Learning Models for Water Quality Prediction. *Computation*. 11(2), 16. https://doi.org/10.3390/computation11020016.

[17]Chen, L., Wu, T., Wang, Z., Lin, X., & Cai, Y. (2023). A novel hybrid BPNN model based on adaptive evolutionary Artificial Bee Colony Algorithm for water quality index prediction. *Ecological Indicators*. 146, 109882. https://doi.org/10.1016/j.ecolind.2023.109882.

[18]Nasir, N., & others. (2023). Deep learning detection of types of water-bodies using optical variables and ensembling. *Intelligent Systems with Applications*. 18, 200222. https://doi.org/10.1016/j.iswa.2023.200222.

[19]Arepalli, P. G., & Naik, K. J. (2023). A deep learning-enabled IoT framework for early hypoxia detection in aqua water using lightweight spatially shared attention-LSTM network. *The Journal of Supercomputing*. 1–30. https://doi.org/10.1007/s11227-023-05580-x.

[20]Xu, J., Wang, K., Lin, C., Xiao, L., Huang, X., & Zhang, Y. (2021). FM-GRU: A time series prediction method for water quality based on seq2seq framework. *Water*. 13(8);1031. https://doi.org/10.3390/w13081031.