The Journal of Supercomputing (2024) 80:2718–2747
https://doi.org/10.1007/s11227-023-05580-x

A deep learning-enabled IoT framework for early hypoxia detection in aqua water using light weight spatially shared attention-LSTM network

Peda Gopi Arepalli[1] · **K. Jairam Naik**[1]

Accepted: 3 August 2023 / Published online: 19 August 2023

Abstract
Dissolved oxygen (DO) is a critical factor in maintaining healthy aquatic ecosystems, including aquaculture ponds. Low DO levels can lead to hypoxia conditions, which are detrimental to fish health and productivity. To deal with this issue, we intend for a smart monitoring system that predicts hypoxia conditions due to low DO levels in aquaculture ponds. The proposed system collects water quality data using Internet of things (IoT) devices and segments it into different categories based on water quality parameters, with a particular focus on low DO levels. By detecting hypoxia conditions early, fish farmers can take corrective measures to prevent fish mortality and improve fish health. To achieve this, our proposed system uses a lightweight Spatially Shared Attention Long Short-Term Memory (SSA-LSTM) model that captures both temporal and spatial dependencies of DO content in water, enabling accurate prediction of hypoxia conditions. Our model outperforms traditional LSTM models and other existing state-of-the-art models, achieving 99.8% accuracy. The proposed system provides a reliable and efficient solution to monitor hypoxia conditions in aquaculture systems and help fish farmers make informed decisions for optimal fish health and productivity.

Keywords Hypoxia detection · Aquaculture · Dissolved oxygen (DO) · Water quality · Deep learning · IoT

✉ Peda Gopi Arepalli
apgopi.phd2020.cse@nitrr.ac.in

K. Jairam Naik
Jnaik.cse@nitrr.ac.in

[1] Department of Computer Science & Engineering, National Institute of Technology Raipur, Raipur, India

1 Introduction

The aquaculture industry has gained increasing importance in recent years, as it provides a sustainable source of protein for a growing global population. However, maintaining optimal water quality [1] is crucial for the growth and health of fish in aquaculture systems [2]. Hypoxia conditions caused by low DO levels can have a particularly negative impact on fish health, leading to reduced growth, increased mortality rates, and decreased profitability. Traditional methods of monitoring DO levels, such as manual sampling and laboratory tests are time-consuming and may not detect hypoxia early enough to prevent fish mortality [9]. Fortunately, the development of IoT technologies has revolutionized the aquaculture industry by enabling the live monitoring of water parameters such as DO, nitrate ammonia etc. [1]. The lively collected data can be analysed to predict hypoxia conditions, allowing farmers to take corrective measures before the situation becomes dire. With this approach, farmers can stay ahead of potential hypoxia events and implement necessary interventions in a timely manner, ultimately improving the sustainability of aquaculture.

While the availability of real-time data provides new opportunities for insight into the complex relationships between water quality parameters and fish health, the high volume and complexity of the data present significant challenges [5]. Machine learning approaches have been intended to take up this challenge, as they can automatically learn intricate relationships in the data and make precise predictions based on historical data. State-of-art deep learning models [4, 18, 19] and 28 have limitations in hypoxia prediction due to the availability and quality of data, model complexity, generalization, and interpretability issues. It is important to carefully evaluate their performance before relying on them for critical decisions. Among these approaches, LSTM-based models have shown promising results in capturing the temporal dependencies in water quality data.

However, traditional LSTM models have limitations in terms of computational efficiency, as they require many parameters and can be slow to train and evaluate. To address this issue, lightweight variants of LSTMs have been proposed, and one such variant is the Spatially Shared Attention LSTM (SSA-LSTM). The proposed system collects water quality data using IoT devices and uses the aquatic quality index (AWQI) [26] to label the data and segment it into different categories based on water quality parameters, with a particular focus on low DO levels. The categorized data has been supplied to the SSA-LSTM model. The proposed model can learn both spatial and temporal properties of DO content in water, enabling accurate prediction of hypoxia conditions. Additionally, the SSA-LSTM model has fewer parameters, making it more computationally efficient than traditional deep-learning models [7, 13]. The advantages of the SSA-LSTM model are significant, as it provides accurate predictions of hypoxia conditions in aquaculture systems, which can help farmers make informed decisions regarding water quality [15, 16] management, feeding, and harvesting. By detecting hypoxia conditions early, farmers can take corrective measures to prevent fish mortality and improve fish health. The use of smart water quality [1] management systems such as the SSA-LSTM model is essential for the sustainable growth of aquaculture systems and to ensure a steady supply of protein to an ever-growing global population.

The key objectives of the study are:

1. To design and implement a cutting-edge IoT system incorporating multiple sensors for measuring the composition of water bodies, by determining real-time data from multiple fishponds. (Sect. 3.1)
2. Utilize the AWQI as a tool to categorize and assess the water quality for hypoxic conditions to evaluate the suitability of the aquatic habitat for different aquatic species. (Sect. 3.3.1)
3. Design a lightweight and optimized hypoxia forecasting model for aqua ponds to improve accuracy and efficiency, while promoting sustainability and productivity in aquaculture operations. (Sect. 4.1)
4. Assess the performance of the anticipated hypoxia forecasting model and current models in handling input data, to determine their impact on accuracy and efficiency in hypoxia prediction. (Sect. 4.3)
5. Assess the effect of different water quality parameters on the anticipated model's ability to predict the occurrence of hypoxia and non-hypoxia situations in a given water body. (Sect. 4.2)
6. Examine the ecological impact of existing methodologies on aquatic animals and assess the effectiveness and significance of the proposed solutions that have been incorporated. (Sect. 4.6)

The rest portion of this article are structured as follows: In Sect. 2, we investigate related work in the field of aquaculture and machine learning for DO prediction and hypoxia situations in water bodies. Section 3 presents materials and proposed methods that describes the proposed SSA-LSTM. In Sect. 4, we demonstrate the results of experimental evaluation. Section 5 concludes with the implications of these findings and future directions for the researchers.

2 Literature survey

In the literature review section, the existing research on hypoxia and low DO conditions prediction using deep learning will be analysed. An overview of the field, including prediction methods and techniques, will be provided. The section will primarily focus on the use of deep learning and analyse the strengths, weaknesses, gaps, and limitations in the literature. Additionally, potential solutions or alternative approaches to improve prediction model accuracy and efficiency will be suggested.

[29] projected a water quality predictive model fusing kPCA and RNN to forecast dissolved oxygen concentration using noisy sensory data. kPCA is used to reduce noise, while RNN maintains temporal information for accurate predictions. Results show the model outperforms comparative methods by 8%. [21] used machine learning to predict hypoxia in a lagoon by analysing dissolved oxygen

and environmental variables. They identified key drivers and synergies using SHAP and emphasized the importance of daily time scales for accurate modelling. SHAP summary plots were used for easy-to-understanding variable attributions. [24] proposed the CSELM method for accurately predicting dissolved oxygen change in aquaculture. CSELM uses k-medoids on Dynamic Time Warping space to group related time slots and separate datasets into distinct groups. The new Soft plus ELM algorithm with partial least squares optimization improves accuracy and efficiency, tolerating data deficit, ambiguous outliers in sensors, and nonlinear, continuously generated data streams that challenge accurate predictions. [6] proposed a hybrid method to foretell DO variation in aquaculture using k-means and an adjusted Soft plus ELM with PSO. The model reduces operation costs by accurately predicting water quality parameters. PSO optimizes model parameters for better prediction performance and accuracy compared to previous models.

[20] conducted a study on the efficacy of attention-based RNN on short and long-term extrapolation of DO in aquaculture for sensible administration and control. They proposed two new RNN structures to capture historical or spatiotemporal interactions separately or simultaneously, achieving similar functioning with earlier methods. However, predicting dissolved oxygen remains challenging due to external factor interference and irregularity in its changes, especially for long-term prediction. [14] anticipated a collective forecasting model built on EEMD and LSSVM to improve dissolved oxygen prediction accuracy. DO time series are decomposed into stable subsequence's using EEMD and reconstructed by PSR. The prediction model for each sub-sequence, improves the generalization ability of the overall forecasted results. However, the study was conducted in a specific location, and the method's performance in other locations with different environmental conditions is unclear. [3] propose a stacked ensemble machine learning model to enhance the precision of DO forecast in aquaculture. The model combines three different base learners with one meta-learner to overcome nonlinearity and complexity associated with dynamic changes in DO levels. While this approach improves DO prediction accuracy compared to standalone models, there may still be other factors affecting water quality in aquaculture production that are not accounted for by this model.

[7] proposed a DO prediction model using K-means and GRU for aquaculture water quality management. Key dynamics affecting changes in DO were selected by PCA, and similarity measures based on Euclidean distance and dynamic time-warping distance were combined to improve clustering accuracy. Experimental results showed higher prediction accuracy compared to conventional approaches, but the limited dataset used in the study may limit its generalizability to other scenarios. (JOEL et al. 2018) found that time-dependent models, specifically the LSTM, are more accurate than non-time-dependent models for predicting dissolved oxygen levels. However, the generalizability of the model to different ponds was not explored. Despite this limitation, the proposed models performed well with the LSTM achieving the highest accuracy. (QIN [23] intended a model that combines Variational Mode Decomposition (VMD) and Deep Belief Network. The model can separate and denoise raw data, perfectly predict DO content, and provides up to 10% higher accuracy and stability than

similar regression algorithms. However, data collection for prediction may be difficult due to the complex high-dimensional nonlinear data space of recirculating aquaculture systems.

[27] has anticipated a prognostication model created on LSTM NN. The intended method based on LSTM NN for water quality forecast has the advantages of nonlinearity, abstraction, and higher predictive accuracy compared to other methods. However, its limitations include needing more training data for better results. The study by [17] presents an innovative approach to water quality prediction in aquaculture systems. By incorporating a classification model and box-plot analysis, the authors can identify the most influential water quality parameters. The results demonstrate remarkable accuracy, with a 95% prediction accuracy for PCA output and outstanding classification accuracy using the Gradient Boosting Classifier Method. Despite these successes, the study acknowledges limitations such as the lack of a dynamic approach to account for varying numbers of specimens, and less precise results compared to the proposed methods.

The literature reviewed in this section has several confines that need to be addressed in the imminent study.

1. In the existing literature, many studies have utilized IoT components for real-time data collection, but it would be favourable to perform additional research to optimize the integration of IoT components and water quality sensors for accurate data collection of various water quality parameters.
2. It is suggested that more research be conducted to explore the effective use of AWQI in assessing hypoxia conditions and improving the management of aquatic habitats for different aquatic species.
3. There appears to be a lack of research that focuses on developing lightweight and optimized hypoxia forecasting models specifically for aqua ponds, which require real-time monitoring and management. Therefore, it would be valuable to conduct further research in this area.
4. While there are studies that evaluate the performance of hypoxia forecasting models, there is a need for more research that compares the intended model's performance with current models and identifies the factors that impact their accuracy and efficiency.
5. While there are studies that investigate the impact of water quality parameters on hypoxia events, there is a need for more research that focuses on the optimal selection of input features and modelling techniques to improve the model's sensitivity to these parameters.
6. While there are studies that assess the ecological impact of aquaculture operations, there is a need for more research that evaluates the effectiveness of proposed solutions in promoting sustainability and productivity while minimizing negative impacts on aquatic ecosystems. Additionally, there is a need for more research on the economic feasibility and stakeholder engagement required for the successful implementation of these solutions.

3 Materials and methods

Firstly, this section provides details on data collection, storage on cloud, information about the publicly available datasets, and discussion on data pre-processing techniques used. Next, it focus on the methodology used for analysing hypoxia forecasting in ponds. By linking these topics, understand the overall process and the specific steps to predict hypoxia conditions in aquatic ecosystems.

3.1 Data collection

This section provides details on the process of data collecting from water bodies using IoT sensors and storing those in the cloud storages and analysing them. Real-time data is collected using an Arduino board equipped with various water quality sensors, including pH, dissolved oxygen (DO), temperature, turbidity, ammonia, nitrate, and manganese sensors. The data is collected from three distinct fishponds located in Guntur, Andhra Pradesh, India, each containing different fish species, including murrel, catla, and a multispecies assortment. The proposed SSA-LSTM method is used to analyse the data. Figure 1 shows the IoT-based SSA-LSTM model for hypoxia analysis in fishponds. To enable internet connectivity and communication between the sensors and the Arduino board, a NODEMCU board is utilized. This setup efficiently collects accurate data on water quality parameters.

Table 1 represents the hardware configuration details of sensors, ARDUINO board, and NODE MCU. The data collection process involved gathering 74,759 data records over one year (February 2022 to January 2023) using these sensors. The data was collected through IoT and manual measurements from fishponds. Subsequently, the data was collectively stored in cloud storage and analysed using an SSA-LSTM model to forecast hypoxia in fishponds. The updated data is publicly available on [22], January 24).

Figure 2 represents a violin plot of the water quality parameters in three different fishponds located in Guntur, Andhra Pradesh, India. The nitrate parameter

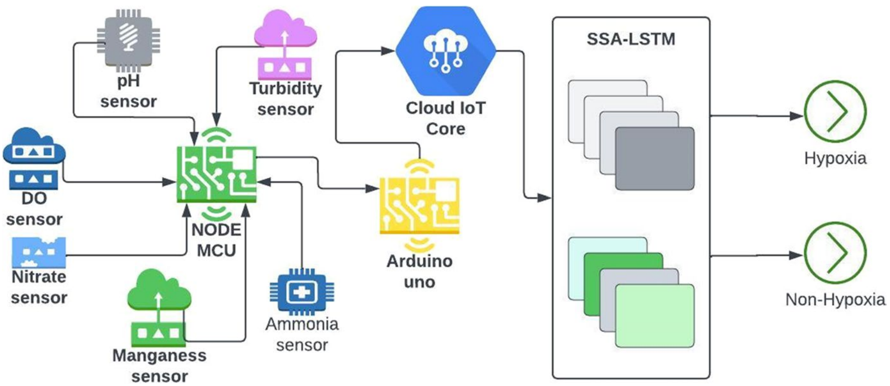


Fig. 1 An IoT-based SSA-LSTM model for hypoxia analysis in fishponds

Table 1 Specifications of hardware used for data collection

Device	Model of the device	Input Voltage	Output Voltage
Arduino	Arduino Uno	+55 V	7 to 12 V
Node MCU	ESP8266	4.5 V-10 V	3.3 V
NITRATE(PPM)	Lab-On-Chip (LOC)	*50 PPM*	10 and 800
PH	Colorimetric	59.16 mV	0 mV signal at 7.0 pH
AMMONIA (mg/l)	GS06	1 V	*0 V to 5 V*
TEMP	LM35	35 V and -2 V	10 mV
DO	Amperometry	230 V	0 to3.0 V
TURBIDITY	Nephelometers	0 to 4.5 V	3.9994 V
MANGANESE (mg/l)	2S Water	5.1	*12 V*

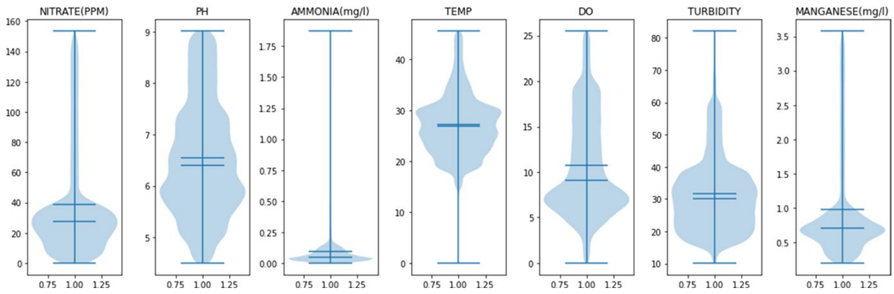


Fig. 2 Shape of each water quality parameter collected from fishponds

data points mostly fall within the range of 18 to 40 ppm, with a few outliers reaching up to 160 ppm. The pH parameter data points range from 5.8 to 7.4. For the ammonia parameter, most of the data points are between 0.001 and 0.025, with a few outliers reaching up to 1.75. The temperature ranges between 25 °C and 31 °C, while the DO ranges between 6 and 15 ppm. Turbidity falls between 22 and 40 ppm, and manganese is between 0.5 and 3.5.

The study utilizes a publicly available dataset called "WaterData" [10, 3], which the Government of India collected. This dataset comprises 9624 samples from various historical sites across different Indian states. The Indian Government initiated this enterprise to assess water quality and safety in the subcontinent. The dataset includes different essential parameters that determine water quality, such as DO, BOD, pH, conductivity, temperature, coliform, nitrate, and total coliform (total number of bacteria).

Figure 3 presents a violin plot representing the water quality parameters of the public dataset. The DO ranges from 0 to 10 ppm, temperature from 16 °C to 24 °C, pH from 7.0 to 8.4, turbidity below 250 ppm, conductivity below 40,000 ohms, BOD from 0 to 500 mg/l, and nitrate from 0 to 25 mg/l.

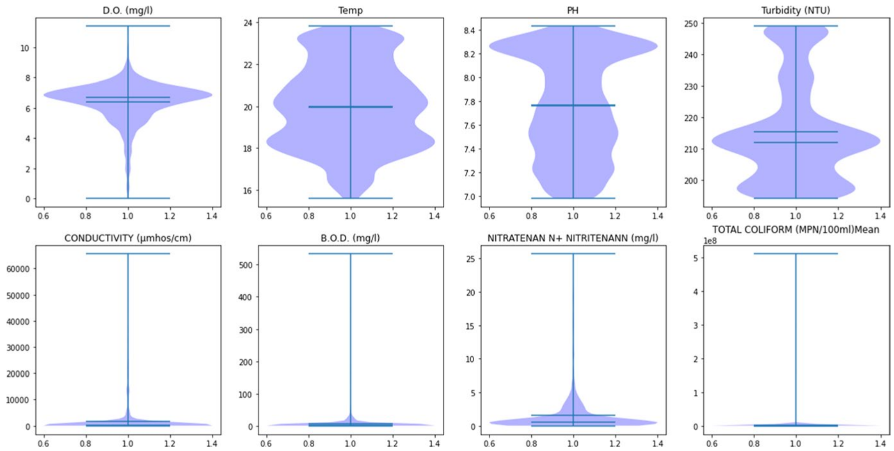


Fig. 3 Shape of each water quality parameter from the public dataset [10; 3]

3.2 Data pre-processing

Data obtained manually from the field through devices often contains irregularities, including missing or inconsistent data, which environmental factors like weather variability, mechanical issues such as faulty sensors, and data storage problems can cause. These irregularities can result in significant differences between expected and actual measured values. Developing accurate prediction models requires high-quality data input. The dataset used for model building must be of excellent quality, consistent, non-redundant, and well-organized. Since irregularities are common in all datasets, it is essential to format, reduce, and normalize the acquired dataset as precisely as possible.

Threshold checking was used to normalize the out-of-bound data obtained from the water quality sensors. Equation1 can be utilized for the normalization of out-of-bound data [25].

$$X_T = \frac{(X - \min(X))}{(\max(X) - \min(X))} \tag{1}$$

where, X_T Represents threshold normalization, and X represents the data vector.

Another data anomaly that affects the forecasting model's performance is the presence of 'Null' values in the dataset. Null as well as NAN values were managed using the mean value method as described in Eq. 2 [25].

$$\overline{Y} = \frac{\sum_{i}^{n} Y}{n} \tag{2}$$

$\overline{Y}$ represents the mean of the data column, n represents the number of data points in a queue, and $\sum_{i}^{n} Y$ represents the sum of all data points in a column.

Using the mean value, null values were normalized while missing values were added to the dataset.

3.3 Methodology

In this study, we collect real-time water quality data using various sensors for effective aquaculture management. Figure 4 illustrates the proposed architectural model of SSA-LSTM in aquaculture management. We consider two datasets for analysis [10, 22, 26] June 22)/, January 24)]. Then, pre-process the data in stages, including threshold processing, data filing, filtering, and error correction. After preprocessing the data, we calculate AWQI (Aquatic Water Quality Index) based on hypoxia conditions for aquatic animals. We segregate the data based on AWQI values, distinguishing between hypoxia and non-hypoxia conditions. We apply the SSA-LSTM model and evaluate its performance, specifically its accuracy and precision in predicting results.

3.3.1 Computation of AWQI for data segregation

The AWQI represents a single value that holistically reflects the water quality at a specific time and location. We calculate it by using several different water quality parameters that we individually measure. In this study, we employ the arithmetic mean approach for WQI calculation. Below Table 2 represents a comparison of important indicators of the estimated parameters in two datasets. And shows the unit weight [2, 26], assigned to the calculation of AWQI.

Assign a weightage (w_i) to each water quality parameter based on its relative importance, Where the sum of all weights is equal to 1.

$$w_1 + w_2 + \cdots + w_i = 1 \tag{3}$$

Standardize each parameter's value (p_i) between 0 and 100 using the following equation:

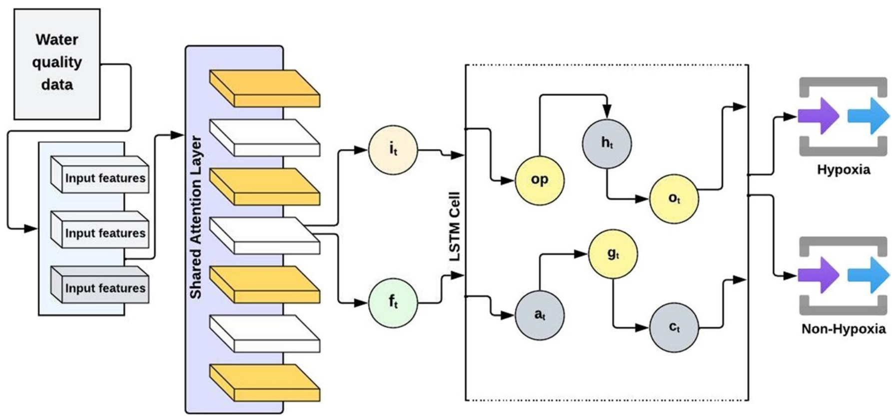


Fig. 4 Architecture of SSA-LSTM for predicting hypoxia

Table 2 Important indicators of the estimated water quality parameters in two datasets and their unit weights

Parameters	Water quality parameters acceptable ranges for aquaculture	Water data	Weight assigned	Unit weight	Ponds	Weight assigned	Unit weight
D.O(Mg/I)	5.0 to 7.0 mg/L	6.29 ± 1.34	5	0.1836	5.64 ± 2.45	5	0.1724
Temp	25 °C to 32 °C	20.005 ± 2.05	4	0.1224	24.55 ± 8.41	3	0.1034
PH	6.5 to 8.5	7.76 ± 0.44	5	0.1632	6.67 ± 1.50	4	0.13793
Turbidity	< 30 NTU	215.48 ± 16.18	–	–	125.90 ± 88.35	4	0.1379
CONDUCTIVITY	200 to 800 µS/cm	1795.3 ± 5592.5	4	0.1428	–	–	–
BOD(mg/I)	< 6 mg/L	6.80 ± 28.41	5	0.1632	–	–	–
NITRARENAN	< 50 mg/L	1.55 ± 2.88	4	0.1224	–	–	–
COLIFORM	–	7577.13 ± 170,734	3	0.102	–	–	–
NITRATE	–	52.23 ± 39.09	5	0.1724	52.23 ± 39.09	5	0.1724
AMMONIA	< 0.02 mg/L	–	4	0.13793	0.34 ± 0.38	4	0.13793
MANGANESE	< 0.05 mg/L	–	4	0.1379	1.08 ± 0.57	4	0.1379

$$p_{i_{\text{standardize}}} = \left[(p_i - p_{i_{\min}}) / (p_{i_{\max}} - p_{i_{\min}}) \right] * 100 \tag{4}$$

where $p_{i_{min}}$ and $p_{i_{max}}$ are the min and max standards of the parameter, correspondingly.

$$S_i = \left[(P_{i_{\text{standerdized}}} / T_i) \right] * 100 \tag{5}$$

(Where T_i is the target value for the parameter).

$$AQI = \left[(W_1 * S_1) + (W_2 * S_2) + \cdots + (W_n * S_n) \right] \tag{6}$$

where W_i signifies each parameter's unit weight, q_i signifies each variable's 0–100 sub-index rating, and n represents the number of sub-indices combined. Water quality parameters that significantly affect overall water quality and vary based on legal limits, such as WHO/IS-10500, must be considered when measuring stability. Parameters with low permissible limits significantly affect water quality [8].

Table 3 enlists sub-index ratings of each water quality variable at 0, 40, 60, and 100 and n signifies the number of sub-indices combined with their rating. Different water parameters have been considered here, and their impact stemmed at different intervals based on fish conditions. Considering that a sub-index rating between 100 and 60 is best suited for fish, a sub-index rating below 60 and the water quality is inadequate for fish farming.

Table 3 Rating of each water quality parameter for fish (Tallar et al., 2016)

Parameter	Unit	100	60	40	0
NITRATE	PPM	$0 \le x \le 40$	$40 < x \le 80$	> 80	< 0
PH	–	$4 \le x \le 6$	$6 < x \le 7$	> 7	< 4
AMMONIA	mg/l	$0 \le x \le 0.1$	$0.1 < x \le 0.5$	> 0.5	< 0
TEMP	^{o}C	$15 \le x \le 20$	$20 < x \le 30$	> 30	< 15
DO	mg/l	$x > 6.5$	$3 \le x \le 6.5$	$0 \le x < 3$	< 0
TURBIDITY	NTU	$0.5 \le x \le 100$	$100 < x \le 210$	> 210	< 0.5
CONDUCTIVITY	µmhos/cm	$150 \ge x \ge 0$	$225 \ge x \ge 150$	$300 \ge x \ge 225$	< 0 or > 300
BOD	Mg/l	Six $\ge x \ge 0$	$80 \ge x \ge 6$	$125 \ge x \ge 80$	< 0 or > 125
TOTAL COLIFORM	MPN/100 ml	$50 \ge x \ge 0$	$500 \ge x \ge 50$	$10{,}000 \ge x \ge 500$	< 0 or $> 10{,}000$
MANGANESE	mg/l	$0.6 \le x \le 0.75$	$0.75 < x \le 1$	> 1	< 0.6

Algorithm-1: Computation of AWQI

Input: Water parameters pH, DO, BOD, EC, NN, and TC.
Output: Aquatic WQI.

1. Assign weightage for each parameter using Table-2
2. Apply standardization using Eq-1.
3. Loop: For $c = 1 \rightarrow C$;
4. Assign w_i
5. Standardize each p_i using Eq-2.
   # Calculate the sub-indices
6. Loop: For p= $1 \rightarrow P$;
7. calculate S_i using Eq-3.
   # Calculate the AWQI
8. Compute AWQI using Eq-4.
9. End

Algorithm-1 elucidates the process of calculation of water quality indices [2, 26]; Nong et al. (2020). The inputs for the algorithm are water quality parameters: pH, DO, BOD, EC, NN, and TC. Water quality parameters are selected in the first step followed by the development of a rating scale through step-2 and step 3, based on fish conditions represented in Table 3 Sub- indices are calculated in step 4, and the unit weight is assigned using step 5. In step 6, the quality range is calculated to finally compute the WQI.

The SSA-LSTM method is a lightweight deep learning model used in this study to predict hypoxia conditions in aquaculture ponds based on dissolved oxygen (DO) levels. This method is chosen because it effectively captures both temporal and spatial dependencies of DO content in water, making it suitable for accurately predicting hypoxia conditions. The relevance of the SSA-LSTM model lies in its ability to handle time-series data with spatial variations, which is essential for monitoring water quality in aquaculture ponds. The model's attention mechanism allows it to focus on specific regions of interest in the data, such as areas with low DO levels, improving the accuracy of hypoxia prediction. Figure 4 shows the architecture of SSA-LSTM for predicting hypoxia, x_t represents the input features at time step t which are public dataset features, real-time dataset features, and class labels, which are passed through the shared attention layer to compute the attention weights a_t. The attention weights and input features are then passed to the LSTM cell, which updates its hidden statics h_t and cell state c_t using the input gate i_t, forget gate f_t, output gate o_t, and cell input g_t. The updated hidden state h_t is then passed to the output layer to compute the output probabilities y_t.

The shared attention layer uses a weight matrix to compute a weighted sum of the input features at each time step, which produces the attention weights a_t. The attention weights determine the relative importance of each input feature for the current prediction and are used to modulate the inputs to the LSTM cell. The LSTM cell is responsible for updating the hidden state h_t and cell state c_t at each time step. The input gate i_t controls the amount of new information that is added to the cell state, while the forget gate f_t determines how much of the previous cell state is retained. The cell input g_t

represents the new input at the current time step, which is modulated by the attention weights a_t. The output gate o_t controls how much of the cell state is used to compute the hidden state h_t.

Our proposed SSA-LSTM lies in the incorporation of Spatially Shared Attention mechanisms. Unlike the original LSTM, which processes sequential data independently, our SSA-LSTM introduces attention mechanisms that allow the model to focus on specific parts of the input sequence that are most relevant for making predictions. This attention mechanism enables the model to capture more meaningful and contextually important information, leading to enhanced accuracy and performance in forecasting hypoxia conditions in fishponds. By leveraging Spatially Shared Attention, our SSA-LSTM model can effectively analyse the spatial relationships between water quality parameters and their impact on hypoxia conditions. This added capability empowers the model to make more informed predictions and contributes to the novelty and efficacy of our proposed approach.

3.4 Working of SSA-LSTM

This section represents the working of SSA-LSTM. Below the SSA-LSTM with spatial attention can be defined as follows:

- $W_a \in R^{D*A}, U_a \in R^{H*A}, V_a \in R^{H*A}$—are the weight matrix for the attention mechanism.
- $b_a \in R^A, b_i \in R^H, b_f \in R^H, b_o \in R^H, b_c \in R^H$ – are the bias terms for the attention mechanism, input gate, forget gate, output gate, and memory cell.
- $W_i \in R^{D*H}, U_i \in R^{H*H}, W_f \in R^{D*H}, U_f \in R^{H*H}, W_o \in R^{D*H}, U_o \in R^{H*H}$ – are the weight matrix for the input gate, forget gate, and output gate.
- $W_c \in R^{D*H}, U_c \in R^{H*H}$ – are the weight matrix for the memory cell

Note that the attention weights a_t represent the importance of each input feature at each time step, allowing the model to focus on the most relevant features for the current prediction. The weight sharing between the attention mechanism and the SSA-LSTM gates and memory cell allows the model to efficiently process spatial data with fewer parameters.

$$a_t = \mathrm{softmax}(W^T \tanh(U_a h_{t-1} + V_a x^T + b_a)) \tag{7}$$

where, x^T is the transpose of the input sequence X, $W \in R^A$ is a weight vector for the attention mechanism.

Equation 7 represents the computation of the attention mechanism in an RNN-based model for sequence-to-sequence learning. Here, x^T is the transpose of the input sequence X, and W is a weight vector for the attention mechanism.

$$\begin{aligned} i_t &= \sigma\left(W_i\left(a_t * x_t\right) + U_i h_{t-1} + b_i\right) \\ f_t &= \sigma\left(W_f\left(a_t * x_t\right) + U_f h_{t-1} + b_f\right) \\ o_t &= \sigma\left(W_o\left(a_t * x_t\right) + U_o h_{t-1} + b_o\right) g_t = tanh\left(W_c\left(a_t * x_t\right) + U_c h_{t-1} + b_c\right) \end{aligned} \tag{8}$$

where, * represents element-wise multiplication.

Equation 8 represents the computation of the input, forget, output, and cell state gates in an LSTM. Here, a_t is the attention mechanism output at time t, x_t is the input at time t, and $h_{(t-1)}$ is the hidden state vector at the previous time step. i_t, f_t, and o_t are sigmoid functions representing the activation of the input, forget, and output gates, respectively. g_t is the new candidate cell state value, computed using a hyperbolic tangent function.

Algorithm-2: Hypoxia prediction using SSA-LSTM

1. Input: water quality variables
2. Output: Forecasting "hypoxia"
   # Initialization of input, and output variables
3. Input vectors $X = [x_1, x_2, \ldots., x_t]$ with dimensions D
4. Target sequence $Y = [y_1, y_2, \ldots., y_t']$ with C classes
5. Weight matrix for the attention mechanism (W_a, U_a, V_a)
6. Bias terms for the attention mechanism $(b_a, b_i, b_f, b_o, b_c)$
7. input gate, forget gate, output gate, and memory cell.
   $(W_i, U_i, W_f, U_f, W_o, U_o)$
8. Weight matrix for the memory cell. (W_c, U_c)
9. LSTM hidden state h_0 and cell state c_0 to zero.
   # Processing of input data using the attention mechanism
10. Compute the attention weights a_t : using Eq-7
    $a_t \leftarrow \{W^T, x^T, U_a, V_a, b_a, h_{t-1}\}$
    # Processing of data using SSA-LSTM gates and memory cell
11. For each input (x_t, a_t)
12. Calculate the i_t , f_t, o_t, and g_t using Eq-8
13. $x_t, a_t = \{ i_t , f_t, o_t, g_t\}$
14. Computer $c_0$and h_0: using Eq-9
    $c_t \leftarrow \{f_t, c_{t-1}, i_t, g_t\}$
15. Compute the new hidden state h_t: using Eq-10
    $h_t \leftarrow \{o_t, c_t\}$
16. Compute the output probabilities y_t using Eq-11
    $y_t \leftarrow \{W_y, h_t\}$
17. Compute the loss: using Eq-12
18. $L = \{y_t', y_t\}$
19. Update the weights using backpropagation.

$$c_t = f_t * c_{(t-1)} + i_t * g_t \tag{9}$$

Equation 9 computes the new cell state value c_t at time t, earlier cell $c_{((t-1))}$ and the latest aspirant cell g_t. The input and forget gates control the contributions of the new candidate value and previous cell state, respectively.

$$h_t = o_t * \tanh(c_t) \tag{10}$$

Equation 10 computes the latest hidden state vector h_t at moment t, which is a function of the current cell state value c_t and the output gate o_t.

$$y_t = \mathrm{softmax}(W_y h_t) \tag{11}$$

Equation 11 computes the output probability distribution over possible output symbols at time t, given the currently hidden state vector h_t. W_t is a weight matrix that maps h_t to the output probability distribution.

$$L = -\sum_t {}' log(y_t) \tag{12}$$

Equation 12 is a function for loss value. y_t is the anticipated output probability distribution at time t, and y_t' is the one-hot encoded correct output symbol at time t. The goal is to minimize this loss function over the entire sequence.

The algorithm-2 presented is a deep learning model that uses an attention mechanism and SSA-LSTM architecture to forecast hypoxia in aqua ponds based on water quality variables. The input is a sequence of water quality variables represented as vectors, and the output is a target sequence that represents the forecasted hypoxia with multiple classes. The attention mechanism helps the model to focus on the relevant features in the input sequence, while the SSA-LSTM architecture combines LSTM with the attention mechanism to provide better modelling of spatial and spectral information. The model updates the hidden state and memory cell using the input, forget, and output gates, and the new hidden state is used to compute the output probabilities using the weight matrix.

Finally, the model calculates the loss using the cross-entropy loss function and updates the weights using backpropagation to improve the model's accuracy in detection of hypoxia presents in the water or not.

4 Results and discussions

The experimental section describes the performance of the proposed SSA-LSTM model with the public dataset [10]) and real-time dataset [22], Jan 24). Illustrate the performance of anticipated models with existing models. Evaluate the proposed classifier with existing classifiers using statistical tests. Analyse the impact of each feature on the proposed model classification with the label (growth and mortality). Finally, provided the ecological impact analysis of proposed and state-of-art models.

4.1 Performance of proposed SSA-LSTM model with public dataset and real-time dataset

Here, we created two hypotheses, a null hypothesis (H_0) and an alternative hypothesis (H_1) based on our problem (performance of proposed SSA-LSTM model with public dataset and real-time dataset).

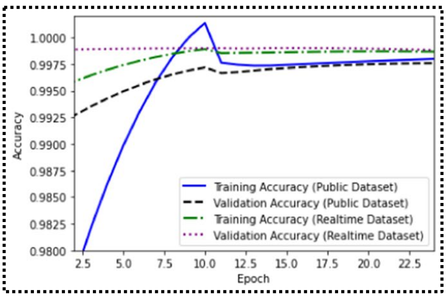
H$_0$ The proposed SSA-LSTM is not performing better with two datasets (Public and Real-time data).

H$_1$ The proposed SSA-LSTM is performing better with two datasets (Public and Real-time data).

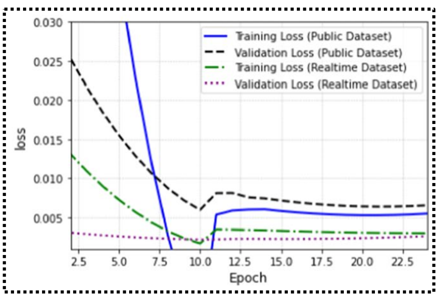
Figure 5a showcases the outstanding performance of the proposed Spatially Shared Attention LSTM (SSA-LSTM) model in predicting hypoxia conditions in water bodies. Specifically, the model achieves remarkable accuracy scores on both the public and real-time datasets, with training and testing accuracies of 99.67% and 99.64% on the public dataset, and 99.87% and 99.45% on the real-time dataset, respectively. The success of the SSA-LSTM model in this application is of great importance, as hypoxia conditions in water bodies can have severe ecological consequences.

Figure 5b demonstrates the loss of the intended model on public and real-time data. The evaluation of the hypoxia detection model based on the SSA-LSTM approach yielded promising results. Specifically, the model achieved a loss value of 0.0038 on the public dataset and a lower loss value of 0.0029 on the real-time dataset, indicating its effectiveness in accurately identifying the hypoxia condition in water bodies. To additionally enhance the performance of the SSA-LSTM model, researchers developed a variation called the Spatially Shared Attention SSA-LSTM model. This model incorporates an attention mechanism that allows the model to selectively focus on relevant spatial locations in the input data, thereby enabling it to effectively handle spatial variations in hypoxia levels within a water body. Accurate prediction of these conditions can aid in the implementation of effective mitigation measures to prevent further damage to the ecosystem.

The SSA-LSTM's effectiveness in this application can be attributed to several key features. Firstly, the model can efficiently process high-dimensional spatial–temporal data, which is prevalent in water quality monitoring. This allows the model to learn and identify relevant features for accurate classification and prediction. Secondly, the SSA-LSTM's shared attention mechanism allows the model to focus on important features while disregarding irrelevant ones, leading to improved accuracy,



(a) Accuracy

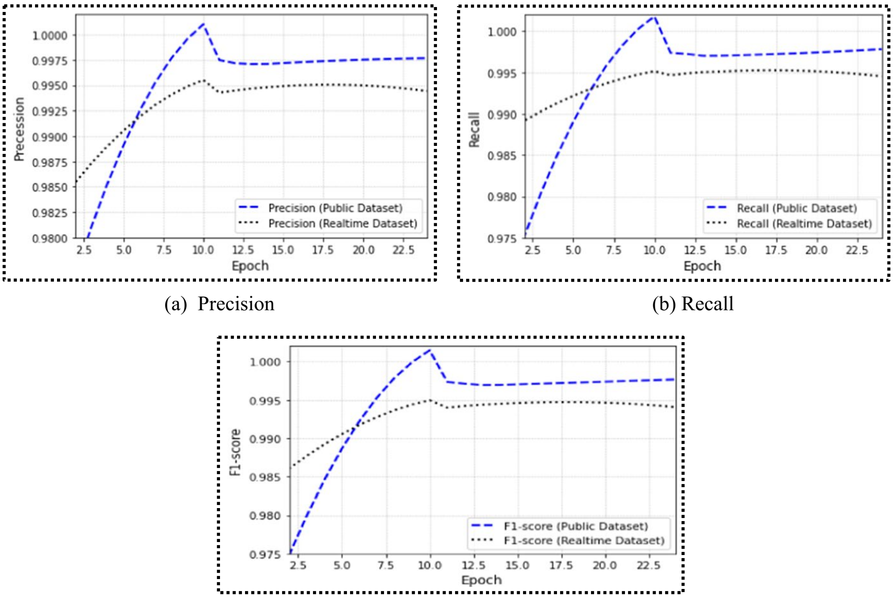


(b) Loss

Fig. 5 Result analysis of proposed SSA-LSTM model on public and real-time data and reduced computational costs. This is particularly useful in scenarios where the spatial distribution of features is non-uniform or varies across different samples. Lastly, the SSA-LSTM's LSTM component enables the model to capture temporal dependencies, which is essential in predicting hypoxia conditions that develop over time. The ability of the LSTM to model long-term dependencies allows the model to make accurate predictions based on historical information.

Figure 6a shows the precision scores of the proposed SSA-LSTM model in detecting hypoxia conditions in water bodies. The model achieves a precision score of 0.9978 on the public dataset, indicating that 99.78% of the predicted hypoxia conditions were correctly classified. On the real-time dataset, the model achieves a slightly lower precision score of 0.9951, indicating that 99.51% of the predicted hypoxia conditions were correctly classified. Figure 6b shows the recall scores of the proposed model in detecting hypoxia conditions. The recall is the proportion of actual hypoxia conditions that were correctly identified by the model. The model achieves a recall score of 0.9981 on the public dataset, indicating that 99.81% of the actual hypoxia conditions were correctly identified. On the real-time dataset, the model achieves a recall score of 0.9955, indicating that 99.55% of the actual hypoxia conditions were correctly identified.

Figure 6c shows the F-score, which is the harmonic mean of precision and recall. The F-score provides a measure of the model's overall performance in detecting hypoxia conditions. The model achieves an F-score of 0.9980 on the public dataset, indicating a good balance between precision and recall. On the real-time dataset, the model achieves a slightly lower *F*-score of 0.9948, indicating that the model



(a) Precision

(b) Recall

(c) F1-Score

Fig. 6 Precision, recall and f1-score curve of SSA-LSTM on public and real-time data may have more difficulty achieving a good balance between precision and recall in real-world scenarios. The high precision, recall, and F-score scores on both datasets demonstrate the robustness and effectiveness of the proposed SSA-LSTM model in detecting hypoxia conditions in water bodies. The model's ability to accurately identify hypoxia conditions is critical for effective management and mitigation, highlighting the importance of the SSA-LSTM in water quality monitoring and related fields requiring accurate detection of spatial–temporal data. the spatial limitation of the proposed SSA-LSTM model trained on South Indian Pond data highlights the importance of careful interpretation and consideration of its applicability to other regions. Efforts should be made to improve the model's generalizability through data augmentation, transfer learning, and validation to ensure its utility in broader geographical contexts.

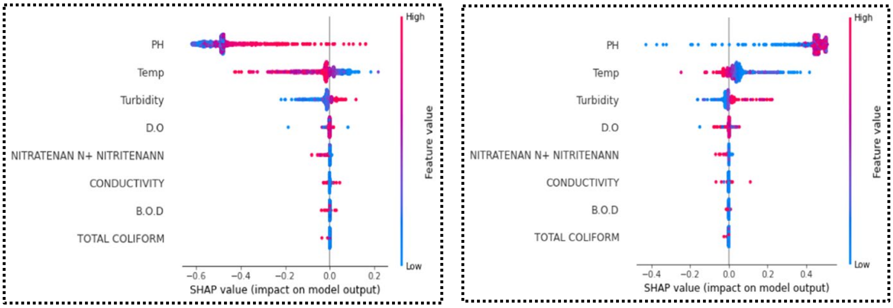
Through systematic experimentation, we optimized our SSA-LSTM model by increasing the learning rate, reducing the batch size from 128 to 64, and decreasing the number of hidden units from 4 to 2. These adjustments led to highly promising results, significantly improving the model's overall performance. The increased learning rate enhanced the model's adaptability during training, while the smaller batch size and reduced hidden units facilitated improved convergence and prevented overfitting, respectively. As a result, our SSA-LSTM now exhibits substantially enhanced predictive capabilities, making it more proficient in analysing time series data and delivering accurate predictions. The rigorous parameter tuning process not only boosted the model's performance but also deepened our understanding of its behaviour and effectiveness in handling complex temporal data.

Both experiments used two different datasets (public and real-time datasets). Both experimental results H_0 are rejected (H_1 is accepted). That confirms that the proposed SSA-LSTM is performing better with two datasets (Public and Real-time data).

4.2 Impact of water quality parameters on hypoxia and non-hypoxia situation in ponds

The SHAP (SHapley Additive exPlanations) library is a popular library for interpreting the output of deep learning models. In the context of fish mortality and growth, the SHAP library can help identify the water quality parameters that have the most impact on these outcomes. The below SHAP values are used to identify the water quality parameters that have the most impact on fish mortality and growth.

Figure 7 shows the impact of various water quality parameters on the SSA-LSTM model's ability to predict hypoxia and non-hypoxia in fish using a public dataset. The model considered multiple features such as DO, pH, temperature, ammonia, manganese, nitrate, BOD, conductivity, total hardness, and turbidity. For predicting hypoxia, pH had a positive impact of 0.2, but a negative impact of 0.6, indicating its complex role. The temperature had a moderately positive impact of 0.1 but also a negative impact of 0.4. Turbidity showed a mixed impact, with a negative impact of 0.2 and a positive impact of 0.1. DO had both positive and negative impacts of 0.1 and 0.12, respectively, while the impact of other features was minor. On the other



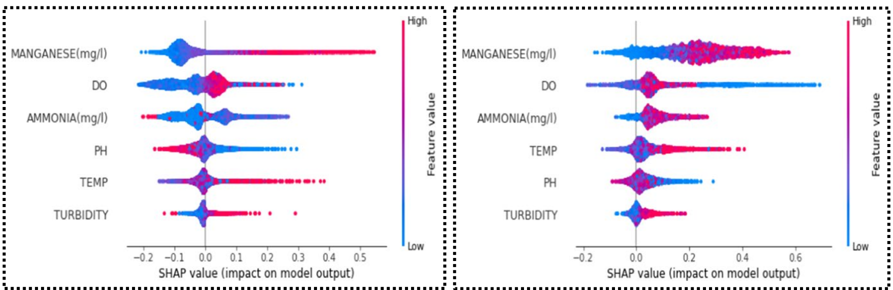
(a) Hypoxia (b) Non-Hypoxia

Fig. 7 Impact of water quality parameters using SSA-LSTM model on predicting labels hypoxia and non-hypoxia conditions with a public dataset

hand, for predicting non-hypoxia, pH showed a strong positive impact of 0.5 but also a negative impact of − 0.4, indicating a complex relationship. The temperature had a similarly strong positive impact of 0.5 but also a negative impact of − 0.24. Turbidity had a negative impact of -0.16. DO showed equal positive and negative impacts of 0.1, indicating its mixed role.

The impact of other features was minor in both cases. These results suggest that pH, temperature, and turbidity play critical roles in determining hypoxia and non-hypoxia in fish, while DO has a slight positive impact. Understanding the influence of these factors can enhance the accuracy of the model's predictions and provide valuable insights into the factors that affect fish health.

The results presented in Fig. 8 provide valuable insights into the impact of various water quality parameters on the SSA-LSTM model's ability to predict hypoxia and non-hypoxia of fish using the real-time dataset. The SHAP library was employed to conduct an in-depth analysis of the model's performance and determine the influence of each feature on the predictions. The features included in the model were Dissolved Oxygen (DO), pH, temperature, ammonia, manganese, and turbidity. For



(a) Hypoxia (b) Non-Hypoxia

Fig. 8 Impact of water quality parameters using SSA-LSTM model for predicting labels hypoxia and non-hypoxia conditions with the real-time dataset the hypoxia label, the SHAP values indicated that ammonia and DO have significant impacts with an equal magnitude of 0.3. Manganese had a very strong positive impact of 0.5 and a negative impact of 0.2. Meanwhile, pH, turbidity, and temperature showed good positive and negative impacts, respectively. For the non-hypoxia label, the SHAP values revealed that ammonia had a positive impact of 0.3 and a negative impact of 0.1. DO have a strong positive impact of 0.7 and a strong negative impact of 0.2. Manganese showed strong positive and negative impacts of 0.6 and 0.1, respectively, and the temperature had a strong positive impact of 0.4 and a slightly negative impact. These findings underscore the continued importance of monitoring ammonia and DO levels in aquatic environments to predict hypoxia and non-hypoxia of fish accurately. Additionally, the model's ability to provide insights into the impact of specific water quality parameters on fish health can be used to develop strategies for improving water quality and ensuring sustainable fish populations. Overall, the results presented in Fig 9 demonstrate the significance of various water quality parameters in predicting hypoxia and non-hypoxia of fish, with ammonia and DO being strong predictors for both labels. By providing a more in-depth understanding of the impact of different factors on fish health, this study can inform more effective management strategies for maintaining healthy aquatic ecosystems.

4.3 Performance illustrate of proposed models with state-of-art models.

Created two hypotheses a null hypothesis (H_0) and an alternative hypothesis (H_1) based on our problem (comparing the performance of the proposed SSA-LSTM model with state-of-art models).

H_0 The proposed SSA-LSTM performance is poor compared to state-of-art techniques.

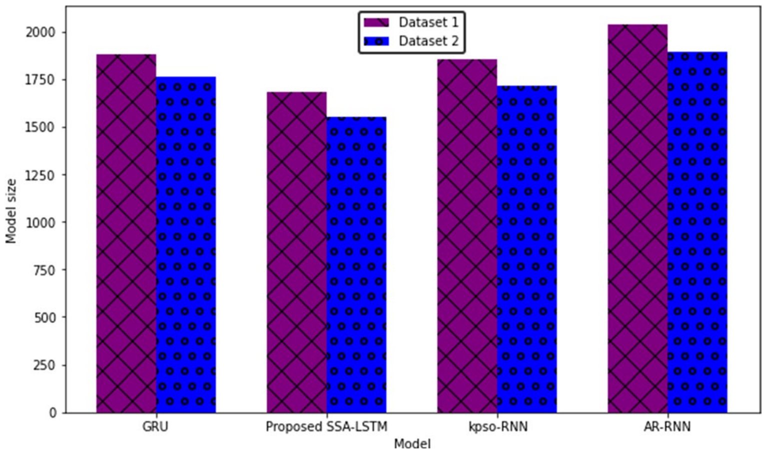


Fig. 9 Proposed and state-of-art models' sizes after training with both public and real-time datasets

H_1 The proposed SSA-LSTM performance is better than state-of-art techniques.

Table 4 presents the findings of a proportional analysis of the intended Spatially Shared Attention LSTM (SSA-LSTM) and other state-of-the-art models in handling hypoxia situations in aqua ponds using a public dataset. The proposed model has demonstrated excellent performance with an accuracy of 99.83%, the precision of 99.78%, recall of 99.81%, F-score of 99.80%, and a low loss value of 0.0038. The SSA-LSTM is well-suited for time series analysis tasks due to its ability to effectively capture long-range dependencies and patterns in sequential and spatial data. In comparison, other state-of-the-art models such as GRU, proposed by [7], only achieve 56% accuracy due to limited interpretability and overfitting problems. Similarly, RNN, proposed by, achieves 97% accuracy but is a complex model, and KPCA-RNN, proposed by [29], achieves 98% accuracy but produces more parameters, making it a complex model. Other models such as CSELM proposed by [6], SEML proposed by [19], PSO-SELM proposed by [24], and EEMD-LSSVM proposed by (Huan et al. in 2018), have also shown good accuracy in time series analysis tasks. However, these models often suffer from handling spatial and temporal features effectively. The proposed model is also compared with the standard LSTM but proposed model performs well over standard LSTM.

When it comes to real-time data time series analysis tasks, the SSA-LSTM model has some advantages over other models. Table 5 displays the comparative results of various models, including the proposed Spatially Shared Attention LSTM (SSA-LSTM), on a public dataset for detecting hypoxia situations in aqua ponds. The SSA-LSTM, a lightweight and efficient neural network architecture, shows exceptional performance with an accuracy of 99.67%, the precision of 99.51%, recall of 99.55%, F-score of 99.48%, and low loss of 0.0029. The SSA-LSTM's ability to capture long-range dependencies and patterns in sequential and spatial data makes it suitable for time series analysis tasks. Compared to other state-of-the-art models, such as GRU, proposed by (Cao et al. in 2020), that only achieves 56% accuracy due to limited interpretability and overfitting, and RNN, proposed by, which has a high accuracy of 97% but is a complex model. KPCA-RNN, proposed by [29], has 98% accuracy but produces more parameters, making it a complex model. Additionally, other models, such as CSELM proposed by [6], SEML proposed by [19], PSO-SELM proposed by [24], and EEMD-LSSVM proposed by (Huan et al. in 2018), has shown good accuracy in time series analysis tasks. However, they often face challenges in effectively handling spatial and temporal features. The proposed model is also compared with the standard LSTM but proposed model performs well over standard LSTM.

Both experiments used two different datasets (public and real-time datasets). Both experimental results H_0 are rejected (H_1 is accepted). That confirms that the proposed SSA-LSTM is performing better with two datasets (Public and Real-time data). The performance of the proposed SSA-LSTM classifier performs better than state-of-art techniques.

Table 4 Performance of proposed SSA-LSTM with state-of-art models on a public dataset

Metrics	Models								
	GRU [7]	RNN	CSELM [6]	SEML [3]	PSO-SELM [24]	KPCA-RNN [29]	EEMD-LSSVM [14]	Proposed model	Standard LSTM
Accuracy	0.5646	0.9715	0.9706	0.9795	0.5387	0.9868	0.9656	**0.9967**	0.9891
Precision	0.5792	0.9647	0.9710	0.9572	0.4343	0.9774	0.9230	**0.9978**	0.9911
Recall	0.5291	0.9792	0.9775	0.9235	0.3879	0.9861	0.9879	**0.9981**	0.9914
F-score	0.5527	0.9687	0.9871	0.9694	0.4872	0.9767	0.9549	**0.9980**	0.9917
Loss	0.6786	0.0656	0.0328	0.0294	0.4613	0.0122	0.0253	**0.0038**	0.0042

The significance of the bold values represents the results generated by the proposed model. These values indicate the outcomes achieved by the model in the respective columns

Table 5 Performance of proposed SSA-LSTM with state-of-art models on a real-time dataset

Metrics	Models								
	GRU [7]	RNN	CSELM [6]	SEML [3]	PSO-SELM [24]	KPCA-RNN [29]	EEMD-LSSVM [14]	Proposed model	Standard LSTM
Accuracy	0.9781	0.9847	0.9871	0.9893	0.8216	0.9819	0.9897	**0.9987**	0.9921
Precision	0.9521	0.9877	0.9810	0.9885	0.0255	0.9881	0.9892	**0.9951**	0.9915
Recall	0.9519	0.9856	0.9785	0.9686	0.0461	0.9874	0.9862	**0.9955**	0.9931
F-score	0.9437	0.9734	0.9673	0.9781	0.0282	0.9877	0.9895	**0.9948**	0.9928
Loss	0.0046	0.0456	0.0298	0.0382	0.1784	0.0185	0.0057	**0.0029**	0.0039

The significance of the bold values represents the results generated by the proposed model. These values indicate the outcomes achieved by the model in the respective columns

4.4 Computational cost comparison of proposed classifier with existing classifiers

Created two hypotheses a null hypothesis (H_0) and an alternative hypothesis (H_1) based on our problem (comparing classifiers).

H_0 The proposed SSA-LSTM is complex on both public and real-time data compared to state-of-art classifiers.

H_1 The proposed SSA-LSTM is simple and lightweight on both public and real-time data compared to state-of-art classifiers.

Fig 9 shows the comparison of proposed and state-of-art model sizes after training with both public and real-time datasets. The proposed SSA-LSTM model, trained with both public and real-time datasets, has a smaller number of parameters compared to the state-of-the-art GRU, KSPO-RNN, and AR-RNN models while achieving comparable or better performance according to evaluation metrics. This is a significant advantage as it makes the proposed model more computationally efficient and easier to train. On the other hand, existing models may require many parameters, making them more computationally expensive and difficult to optimize.

The comparison Table 6 showcases the computational cost of four deep learning models (GRU, Proposed SSA-LSTM, AR-RNN, and KPSO-RNN) trained on two datasets (Real-time Data and Public Data). Upon analysis, it is evident that the Proposed SSA-LSTM model exhibits competitive computational efficiency. It outperforms the other models in terms of training time on both datasets, taking around 180.3 s on the Real-time Data and 150.8 s on the Public Data. Additionally, the Proposed SSA-LSTM model demonstrates superior memory management, with peak memory usage of 176 MB and 132 MB on the respective datasets, which is notably lower than the other models. the Proposed SSA-LSTM model exhibits faster training times and lower memory usage compared to other models (GRU, AR-RNN, and KPSO-RNN) on both the Real-time Data and Public Data datasets. These advantages can be attributed to the efficient architecture and algorithm design of the

Table 6 Computational cost of proposed and existing models

Model	Dataset	Training time (seconds)	Peak memory usage (MB)
GRU	Real-time data	190.5	216
	Public data	120.2	182
Proposed SSA-LSTM	Real-time data	180.3	176
	Public data	150.8	132
AR-RNN	Real-time data	157.0	188
	Public data	95.2	163
KPSO-RNN	Real-time data	210.2	226
	Public data	130.5	199

Proposed SSA-LSTM, which is specifically tailored for the given datasets. Additionally, the model's optimization techniques and well-suited hyperparameter settings contribute to its computational efficiency. The reduced computational cost of the Proposed SSA-LSTM makes it a practical and cost-effective choice for real-world applications, particularly for low-scale fishers or scenarios with limited computing resources. By providing valuable insights into the model's performance and resource requirements, this research enables informed decision-making, ensuring the selection of the most suitable model for aquatic ecosystem management and hypoxia prediction. Experiments used two different datasets (public and real-time datasets) the results H_0 are rejected (H_1 is accepted). That confirms that the proposed SSA-LSTM is performing better with two datasets (Public and Real-time data). The performance of the proposed SSA-LSTM classifier performs better than state-of-art techniques.

4.5 Statistical tests for evaluating the performance of proposed classifier with existing classifiers

In this research, we use two significance statistical tests namely Friedman test and Nemenyi test used for analysis of classifiers. These two are valuable tools for researchers comparing multiple classifiers. The Friedman test assesses overall differences among the classifiers, while the Nemenyi test further refines the analysis by pinpointing specific pairwise differences. Their non-parametric nature makes them versatile and robust in various scenarios, helping researchers make informed decisions about model selection and performance evaluation.

Created two hypotheses a null hypothesis (H_0) and an alternative hypothesis (H_1) based on our problem (comparing classifiers).

H_0 The classifiers are equal.

H_1 The classifiers are different.

4.5.1 Friedman test

The Friedman [11, 12] test is a non-parametric statistical test used to determine whether there are significant differences between the performances of multiple classifiers on a given dataset. The compassion of different classifiers using the Friedman test is taking inputs from Tables 4 and 5. These tables are used as inputs for the Friedman test. Table 7 shows the ranking and average of each classifier. The assignment of rank for each classier for different metrics like better model assign with rank 1 and continue. After assigning the ranks calculate the average of each row taken as $R_{j.}$

$$X^2{}_f = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} R^2{}_j - \frac{k(k+1)2}{4}\right] \tag{13}$$

The $X^2{}_f$ value is 9.72

Table 7 Rank and average rank of each metrics for different classifiers

Metrics	Models			
	GRU	Proposed SSA-LSTM	AR-RNN	KPSO-RNN
Accuracy	4	1	2	3
Precision	4	1	3	2
Recall	4	1	3	2
F-score	4	1	3	2
Loss	2	1	4	3
Avg (R_j)	3.6	1	3	2.4

$$F_f = \frac{(N-1){x_f}^2}{N(k-1) - {x_f}^2} \tag{14}$$

F_f value is 7.36.

Calculated test statistic: ${X^2}_f = 9.72$.

Degrees of freedom (d_f): 3 (since there are 4 classifiers, d_f = k−1 = 4−1 = 3).

Critical value (from statistical tables or software) for d_f = 3 and α = 0.05: F_f = 7.81 (approximate value).

Since the calculated ${X^2}_f$ (9.72) is greater than the critical value F_f (7.81), we can reject H0. The p-value, which represents the probability of observing a test statistic as extreme or more extreme than 9.72 under the assumption of H0, is less than $\alpha = 0.05$.

4.5.2 Nemenyi test

The Nemenyi test is a post hoc test that is commonly used in conjunction with the Friedman test to determine which classifiers are significantly different from each other in terms of their performance.

$$\mathrm{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{15}$$

q_α=2.569 {k = 4, N = 5}.

Critical Difference (CD) = 2.09536.

To prove that H0 should be rejected in both the Friedman test and the Nemenyi test, we need to compare the calculated test statistics (${X^2}_f$ and F_f) with their respective critical values and determine if the p-values are less than the chosen significance level (α). Remember, the significance level (α) is typically set to 0.05.

Calculated test statistic: CD = 2.09536.

There is no direct p-value for the Nemenyi test; instead, we compare the difference in rankings between classifiers with the CD value.

In the Nemenyi test, classifiers are considered significantly different from each other if the difference in rankings (average ranks) is greater than the CD value. Since the calculated CD (2.09536) is used to compare the differences in rankings, and all the classifiers obtained a difference higher than CD, we can say that they are significantly different.

In both cases, the null hypothesis (H0) is rejected, indicating that there are significant differences between the classifiers' performances based on the chosen metrics and the analysis conducted. This means the classifiers are not equal, and their performances are different from each other. The proposed SSA-LSTM model has shown to be better than the other models in this study.

4.6 Ecological impacts on aquatic animals with proposed and state-of-art models

Here we aim to analyse the ecological impact of water quality on aquatic animals by considering both existing and proposed solutions. Our holistic approach includes assessing the accuracy of a deep learning model and assigning weights to each water quality parameter used in the model's training. These weights are determined by the parameters' importance in determining the ecological impact on fish populations. To calculate the overall impact, we take the average of these weights using Eq. 16.

$$OI_m = \frac{\sum W_i}{n} \tag{16}$$

where OI_m is the overall impact of model m, w_i weight of i$^{\text{th}}$ parameter, n Number of parameters.

Then, we use the accuracy of the deep learning model as an additional factor in our calculations. By multiplying the accuracy of the model with the overall impact factor, we can obtain a more comprehensive and accurate assessment of the ecological impact of water quality on fish populations.

$$E = A_m \times OI_m \tag{17}$$

E Ecological impact, A_m accuracy of model m

Table 8 focuses on the comparison of proposed and state-of-art deep learning models used in water quality analysis to predict the ecological impact on fish populations. These models consider various parameters such as pH, temperature, and dissolved oxygen, to forecast fish growth and mortality rates accurately. The SSA-LSTM model has been proposed as a promising model for ecological impact analysis, with an ecological impact value of 79.89%, which outperforms other existing

Table 8 Ecological impact on aquatic animals with proposed and state-of-art models

Authors	No. of parameters	Model	A_m	OI_m	E (%)
[7]	11	GRU	0.9781	0.54	52.81
	9	RNN	0.9841	0.67	65.93
[29]	6	KPCA-RNN	0.9891	0.64	63.30
Proposed (SSA-LSTM)	7	SSA-LSTM	0.9987	0.80	79.89

models like GRU, RNN, and KCPA-RNN. It is worth noting that the ecological impact of water quality on fish populations can occur under both hypoxic and non-hypoxic conditions. Hypoxia, or oxygen depletion, is a significant environmental stressor that can cause fish mortality and negatively impact aquatic ecosystems. Non-hypoxic conditions, on the other hand, can also have adverse effects on fish populations, such as altering their growth, reproduction, and behaviour. Therefore, it is essential to develop models that can predict the ecological impact of water quality under both hypoxic and non-hypoxic conditions. Deep learning-based models, like the SSA-LSTM model, can be trained to account for both hypoxic and non-hypoxic conditions and accurately predict their ecological impact on fish populations. These models provide crucial insights into managing and preserving aquatic ecosystems and can aid in making informed decisions about water management and conservation efforts.

5 Conclusion

The application of IoT and AI technologies to monitor water quality in the aquaculture industry presents a significant opportunity to transform fish farming operations. The proposed smart monitoring system, which employs the Spatially Shared Attention LSTM (SSA-LSTM) model, represents a significant improvement in the accuracy of hypoxia condition prediction, leading to better management practices that improve fish growth and reduce mortality rates. The system collects real-time and accurate water quality data using IoT devices and utilizes the aquatic quality index (AWQI) to segment the data into different categories based on water quality parameters, with a focus on low DO levels. The advanced predictive capability of the SSA-LSTM model improves the accuracy and efficiency of hypoxia condition forecasting, enabling timely corrective measures to prevent fish mortality and enhance fish health. The proposed model produces precision, recall, and F-score values of 99.2, 99.4, and 99.5 on public datasets and 99.23, 99.45, and 99.8 on real-time data, outperforming existing models by 2 to 5%. These results demonstrate the high performance and practical application of the proposed model in real-world aquaculture systems.

The study highlights the influence of water quality parameters on fish growth and mortality prediction using the proposed model and investigates the impact of spatial and temporal characteristics of water quality parameters on aquatic animals. This research provides valuable insights into the potential benefits of utilizing IoT and AI technologies in aquaculture systems, which can lead to more sustainable and efficient operations that benefit both the environment and the economy. However, further research is needed to validate the proposed SSA-LSTM model's performance in different aquaculture scenarios and optimize the IoT system's design and implementation for real-world applications. Overall, the proposed system represents a promising and innovative approach to addressing the challenges of low dissolved oxygen levels and hypoxia conditions in aquaculture systems, with the potential to transform the industry and improve fish health and productivity while minimizing environmental impact.

Author contributions PGA contributed to concepts, development of methodologies, Sensor and Arduino board Design & assembling, dataset collection & creation, experimentation, results analysis, and writing of the original draft; KJN contributed to concepts, experimentation, results analysis, writing, document review, editing, and overall supervision. All authors read before submission and approved the final manuscript for submission.

Funding Not applicable.

Data availability The data that support the findings of this study are available on request from the corresponding author. The data is not publicly available due to privacy or ethical restrictions.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests.

Consent to participate All authors gave explicit consent to participate in this work.

Consent for publish All authors gave explicit consent to publish this manuscript.

Ethical approval All authors have seen and agreed with the contents of the manuscript and are looking forward to publishing this paper on this journal.

References

1. Aghel B, Rezaei A, Mohadesi M (2019) Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach. Int J Environ Sci Technol 16(8):4823–4832
2. Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, García-Nieto J (2019) Efficient water quality prediction using supervised machine learning. Water 11(11):2210
3. Anbarivan. (n.d.). Indian water quality data (Version 1). https://www.kaggle.com/datasets/anbarivan/indian-water-quality-data
4. Arepalli PG, Khetavath JN (2023) An IoT framework for quality analysis of aquatic water data using time-series convolutional neural network. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-023-27922-1
5. Bisht AK, Singh R, Bhatt A, & Bhutiani R. (2017). Development of an automated water quality classification model for the River Ganga. In International Conference on Next Generation Computing Technologies. Springer, Singapore, pp 190–198.
6. Cao S, Zhou L, Zhang Z (2021) Prediction of dissolved oxygen content in aquaculture based on clustering and improved ELM. IEEE Access 9:40372–40387
7. Cao X, Liu Y, Wang J, Liu C, Duan Q (2020) Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network. Aquacult Eng 91:102122
8. CPCB | Central Pollution Control Board. (2019, October 23). CPCB | Central Pollution Control Board; cpcb.nic.in. https://cpcb.nic.in/wqstandards/ Accessed on 2021, May 5.
9. Dabrowski JJ, Rahman A, & George A. (2018, December). Prediction of dissolved oxygen from pH and water temperature in aquaculture prawn ponds. In: Proceedings of the Australasian joint conference on artificial intelligence-workshops (pp. 2–6).
10. Data.gov.in. (2017). https://data.gov.in/catalog/water-quality-india-2013?filters%5Bfield_catalog_reference %5D=2914901&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc. Accessed on 2021, May 5
11. Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 32(200):675–701

12. Friedman M (1940) A comparison of alternative tests of significance for the problem of m rankings. Ann Math Stat 11(1):86–92
13. Gopi AP, & Naik KJ. (2022). An IoT model for Fish breeding analysis with water quality data of pond using Modified Multilayer Perceptron model. In 2022 International Conference on Data Analytics for Business and Industry (ICDABI) (pp. 1–6), Sakhir, Bahrain, 2022, pp 1–6, doi: https://doi.org/10.1109/ICDABI56818.2022.10041617.
14. Huan J, Cao W, Qin Y (2018) Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. Comput Electron Agric 150:257–265
15. Imani M et al (2021) A novel machine learning application: Water quality resilience prediction Model. Sci Total Environ 768:144459
16. Kar D (2019) Wetlands and their fish diversity in Assam (India). Transylvanian Rev of Syst and Ecol Res 21(3):47–94
17. Khan MSI, Islam N et al (2022) Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. J of King Saud Univ-Comput Inf Sci 348:4773–4781
18. Khullar S, Singh N (2022) Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. Environ Sci Pollut Res 29(9):12875–12889
19. Kozhiparamban RAH, Swetha P, Harigovindan VP (2023) Accurate dissolved oxygen prediction for aquaculture using stacked ensemble machine learning model. National Academy Sci Lett 46(3):203–207
20. Liu Y, Zhang Q, Song L, Chen Y (2019) Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. Comput Electron Agric 165:104964
21. Politikos DV, Petasis G, Katselis G (2021) Interpretable machine learning to forecast hypoxia in a lagoon. Eco Inform 66:101480
22. Pondsdata (2023). Pondsdata | Kaggle. Retrieved January 24, 2023, from https:///datasets/d39674fc577e40a74fe614ac3288c7a4d0dc47b2e51b7aaebedf7851c77414a4
23. Ren Q, Wang X, Li W, Wei Y, An D (2020) Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. Aquacult Eng 90:102085
24. Shi P, Li G, Yuan Y, Huang G, Kuang L (2019) Prediction of dissolved oxygen content in aquaculture using Clustering-based Softplus Extreme Learning Machine. Comput Electron Agric 157:329–338
25. Subasi A (2020) Practical machine learning for data analysis using python. Academic Press
26. Tallar RY, Suen JP (2016) Aquaculture Water Quality Index: a low-cost index to accelerate aquaculture development in Indonesia. Aquacult Int 24(1):295–312
27. Wang Y, Zhou J, Chen K, Wang Y, & Liu L (2017). Water quality prediction method based on LSTM neural network. In 2017 12th international conference on intelligent systems and knowledge engineering (ISKE) (pp. 1–5). IEEE
28. Zhang Q et al (2022) A watershed water quality prediction model based on attention mechanism and Bi-LSTM. Environ Sci and Pollut Res 2950:75664–75680
29. Zhang YF, Fitch P, Thorburn PJ (2020) Predicting the trend of dissolved oxygen based on the kPCA-RNN model. Water 12(2):585