LSTM BASED SOFT-SENSOR FOR ESTIMATING NITRATE CONCENTRATION IN AQUAPONICS POND

A. Dharshan
Department of Artificial Intelligence and Data Science
Saranathan College of Engineering
Trichy, India
dharshanathi21@gmail.com

Purushottam Kumar
CoE in Artificial Intelligence
Machine Learning & Data Analytics Lab
National Institute of Technology
Trichy, India
kumarpurushottamnitt@gmail.com

Dr. S. Ravimaran
Department of Artificial Intelligence and Data Science
Saranathan College of Engineering
Trichy, India
ravimaran-aid@saranathan.ac.in

Dr. U Srinivasulu Reddy
Department of Computer Applications
Machine Learning & Data Analytics Lab
National Institute of Technology
Trichy, India
usreddy@nitt.edu

***Abstract*—In the field of aquaponics, where fish and plants coexist in a symbiotic environment, closely monitoring nitrate levels in the water is crucial due to their profound impact on aquatic and plant well-being. Traditional nitrate measurement methods are often time-consuming and costly. Various approaches, including first principles, IoT-based sensors, and machine learning-based soft sensors, have been attempted to address this challenge. However, these efforts face challenges such as expensive sensors, infrequent data collection, multi-stage data processing using limited sensor types, and the need for regular maintenance like cleaning and calibration. Additionally, varied environmental conditions affect sensor suitability for different water environments, and even some machine learning-based soft sensors have proven inaccurate. In response, soft sensors, especially deep learning-based ones, have gained prominence in industrial applications for their adaptability and accuracy. These sensors provide real-time insights into complex processes without requiring expensive hardware. In this study, an innovative solution was introduced using Long Short-Term Memory (LSTM) technology, a neural network architecture in deep learning known for capturing complex temporal patterns. LSTM is well-suited for modeling and predicting nitrate concentration changes in aquaponics, trained with extensive data collected from various aquaponic ponds. Through rigorous evaluation, a remarkable MSE value of 0.00074 and an impressive R-squared score of 0.98 were achieved, holding potential for scaling up to commercial applications, benefiting aquaponics operations, supporting researchers, and enhancing sustainability and productivity in aquaponic systems.**

Keywords—aquaponics, nitrate estimation, LSTM, deep learning, sustainability

1. INTRODUCTION

Aquaponics, a sustainable and innovative agricultural system, integrates aquaculture (the cultivation of fish or other aquatic creatures) with hydroponics (the cultivation of plants in water). This symbiotic ecosystem offers a dynamic synergy between aquatic life and plant growth creating a closed-loop environment where waste from fish becomes a valuable nutrient source for plants, and in return, plants purify the water for the fish [1]. As a result, aquaponics has garnered significant attention for its potential to address the growing global demand for efficient, eco-friendly food production. In the provided **Fig 1**, the interconnected components are depicted to illustrate the functioning of the aquaponic system. The process begins with a pump that circulated water from the fish tank to the biofilter. The biofilter is a crucial element that converts the ammonia-rich fish waste into nitrate, a vital nutrient for plant growth. The nutrient-rich water then flows to the plant bed, where plants uptake the nutrients, helping to purify the water. A siphon mechanism directs the cleaned water back to the fish tank, closing the loop. This carefully designed arrangement ensures a sustainable and efficient cycle, where well-being of both fish and plant is well maintained within the aquaponic environment. Central to the success of aquaponic system is the precise management of water quality parameters, with nitrate concentration being a critical factor. Nitrate, a form of nitrogen, serves as a primary nutrient for plant growth but can be harmful to aquatic life in excess. Thus, maintaining optimal nitrate levels is crucial to ensure the well-being of both aquatic organisms and cultivated plants within these systems [2]. Traditional methods such as adaptive filtering based soft sensor for estimating nitrogen concentration in aquaponics ponds [3], machine learning-based soft sensor models [4], novel LSTM-based soft sensors for wastewater treatment forecasting [5], edge computing-based smart aquaponics monitoring systems using deep learning in IoT environments [6][7], and development of soft sensors for water quality prediction[8] for monitoring nitrate concentrations as well as monitoring the water quality in aquaponics have often proven cumbersome, time-consuming, and cost-inefficient [9]. This challenge has triggered an exploration for innovative solutions that not only streamline nitrate estimation but also enhance its accuracy and real-time monitoring capabilities. In response to these imperatives, our research presents an innovative nitrate estimation system designed to revolutionize the way nitrate levels are managed in aquaponic systems. At the heart of our solution lies Long Short-Term Memory (LSTM) technology, a neural network architecture derived from deep learning. LSTM is renowned for its ability to capture complex temporal patterns [10], making it particularly well-suited for modelling and predicting nitrate concentration changes in aquaponic environments. The key differentiator of our system is its adaptability and accuracy, which enable it to provide real-time insights into nitrate dynamics without the need for expensive hardware or extensive maintenance.

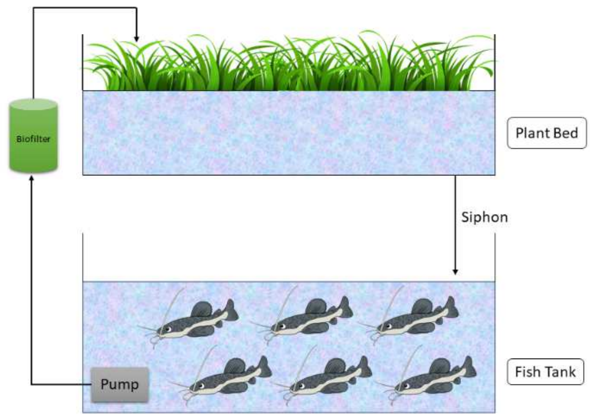

***Fig 1**. A schematic of small-scale aquaponic system.*

2. DATASET

The dataset provides comprehensive information about the freshwater catfish ponds in an aquaponics system. Automated data collection occurred via an ESP 32 microcontroller equipped with a range of water quality sensors, including the, DF Robot Dissolved Oxygen sensor, DF Robot Turbidity sensor, MQ135 Nitrate sensor, MQ-137 Ammonia sensor, Dallas Instrument Temperature sensor (DS18B20), and DF Robot pH sensor V2.2, logging data at 5-second intervals. This data collection spanned from June to mid-October 2021 and formed the basis for creating twelve distinct datasets, each originating from a different catfish pond within the aquaponics system. In each of these twelve ponds, six sensors were strategically placed to record essential IoT measurements encompassing temperature, turbidity, dissolved oxygen, pH, ammonia levels and nitrate concentration. Initially, the dataset considered a total of nine predictors, including timestamps, temperature, turbidity, dissolved oxygen levels, pH, ammonia levels, nitrate levels, population, fish length and fish weight.

3. CONTRIBUTION

The contribution of this research lies in the successful development and implementation of nitrate estimation system for aquaponic systems, addressing the critical need for precise monitoring of nitrate levels. Leveraging Long Short-Term Memory (LSTM) based Soft Sensor, our proposed solution demonstrates remarkable adaptability and accuracy in capturing complex temporal patterns associated with nitrate concentration changes. The system utilizes a dataset derived from freshwater catfish ponds equipped with a variety of water quality sensors, enabling real-time monitoring without the need for expensive hardware or extensive maintenance. The achieved Mean Squared Error (MSE) value of 0.00074 and an impressive R-squared score of 0.98 underscore the system's robust performance. This research not only contributes to nitrate estimation in aquaponic environments but also holds promise for scalability to commercial applications, thereby benefiting aquaponics operations, supporting researchers, and advancing sustainability and productivity in aquaponic systems.

4. DATA PREPROCESSING

- Initially, 12 datasets were collected, each representing a different catfish pond in the aquaponics system.
- Addressed missing values, Nan values, and Inf values by replacing them with the mean values of their respective columns
- All 12 datasets were merged into a single unified dataset for comprehensive analysis
- To ensure uniformity, feature names throughout the dataset were standardized.
- Negative values in the "Turbidity" and "pH" features were replaced with their respective column means.
- Outliers were detected and removed from the dataset using the z-score method to maintain data quality.
- Descaling the values of "Ammonia" and "Nitrate" to their original units involved using the following equations 4.1 and 4.2:

$$Ammonia = \frac{ln[Ammonia]}{10} \tag{4.1}$$

$$Nitrate = \sqrt{[Nitrate]} \tag{4.2}$$

- Features like "created_at" and "entry_id" were deemed non-essential to our objectives and were consequently excluded.
- To ensure uniformity and mitigate the impact of varying scales, feature scaling was performed using Min-Max scaling, which scales all features to a common range.

5. DATASET DESCRIPTION

The dataset employed in this study plays a pivotal role in the development and evaluation of an LSTM-based soft-sensor designed explicitly for estimating nitrate concentration in aquaponic ponds. It serves as the primary source of information for training, testing, and validating the performance of this novel monitoring system.

A. Key Parameters for Nitrate Estimation

1) Temperature (C): The water temperature, measured in degrees Celsius, is a critical parameter within our LSTM-based soft-sensor. Temperature variations significantly influence nitrate dynamics and the biological processes governing nitrate conversion.

2) Turbidity (NTU): Turbidity, quantified in Nephelometric Turbidity Units (NTU), holds relevance as it impacts water quality, which, in turn, can affect nitrate solubility and distribution within the aquaponics ecosystem.

3) Dissolved Oxygen (g/ml): The concentration of dissolved oxygen, expressed in grams per milliliter (g/ml), is intricately connected to nitrate dynamics. Adequate oxygen levels support nitrification processes that influence nitrate levels.

4) pH: The pH level, measuring water acidity or alkalinity, plays a role in nitrate transformations, It can affect the rates of nitrification and denitrification processes, directly influencing nitrate concentrations.

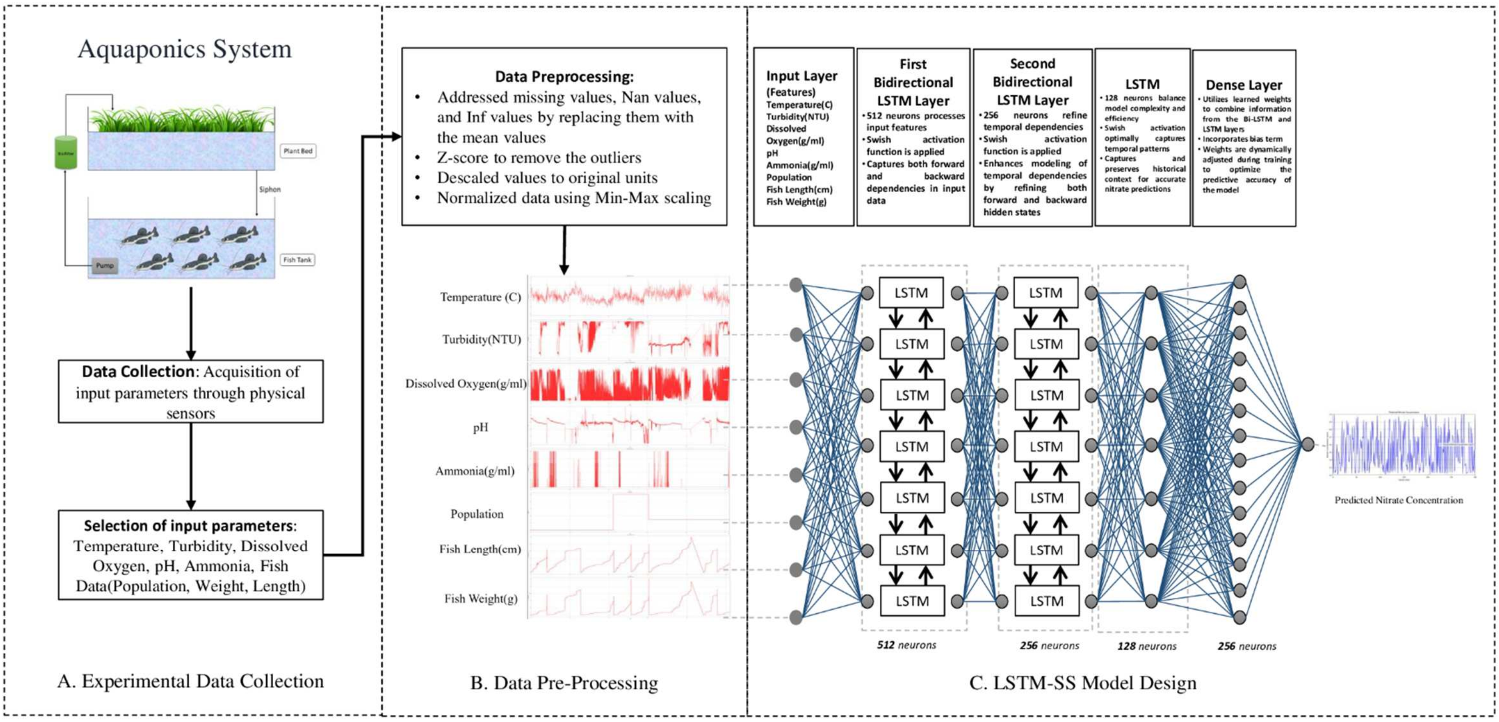

Fig 2. *Design methodology of LSTM-SS for Nitrate Concentration: This diagram illustrates the Aquaponics Experimental Data Collection, Data Pre-Processing and LSTM-SS Model Design phases*

5) *Ammonia (g/ml):* Ammonia concentration, in grams per milliliter (g/ml), is interlinked with nitrate due to their shared involvement in the nitrogen cycle. Understanding ammonia levels aids in comprehending nitrate dynamics.

6) *Nitrate (g/ml):* Nitrate concentration, measured in grams per milliliter (g/ml), is the central focus of our soft-sensor. It serves as both the target variable for prediction and the parameter of interest for monitoring and control.

B. Ecosystem Characteristics:

1) *Population:* The population of aquatic organisms within the aquaponics system is considered as a contextual factor. It can influence nitrate levels through biological processes and nutrient cycling.

2) *Fish Length(cm) and Fish Weight(g):* Fish metrics, including length in centimeters (cm) and weight in grams (g), are integrated into our soft-sensor's framework. These metrics provide insights into fish growth and activity, which can indirectly affect nitrate dynamics [11].

6. ESTIMATING NITRATE CONCENTRATION WITH SOFT SENSOR

In aquaponics, monitoring nitrate levels is crucial, yet traditional methods are slow and labor-intensive. To address this issue, a soft sensor based Nitrate Concentration Estimator is proposed in the present work. Leveraging deep learning with LSTM networks, LSTM-SS offers a practical solution for aquaponics, promising efficient water quality management and enhanced sustainability.

5.1. LSTM-based Soft Sensor (LSTM-SS)

LSTM-Based Soft Sensors (LSTM-SS) have emerged as a transformative force in industrial applications [12]. These innovative systems harness the power of Long Short-Term Memory(LSTM) technology, which excels at capturing temporal patterns in sequential data [13]. The LSTM-SS represents a significant innovation in aquaponic management. By continuously monitoring nitrate levels in real-time, LSTM-SS draws insights from historical data to forecast nitrate concentrations. This precision-driven approach revolutionizes aquaponics management, enhancing operational efficiency and environmental sustainability.

5.2. Design of LSTM-SS for Nitrate Concentration

The architecture of LSTM-SS comprises several pivotal components, as illustrated in **Fig 2**, In the Experimental Data Collection phase, data is gathered from the Aquaponic system, including the data collection process and the selection of input parameters. This information feeds into the Data Pre-processing stage, where the collected data is prepared for analysis. The model architecture for LSTM-SS includes the input layer, first and second bidirectional layers,

Table 1. Parameters used for Nitrate Estimation

S.No	Parameters	Non-Null Count	Datatype
1.	Temperature (C)	957860	Float64
2.	Turbidity(NTU)	957860	Float64
3.	Dissolved Oxygen(g/ml)	957860	Float64
4.	pH	957860	Float64
5.	Ammonia(g/ml)	957860	Float64
6.	Nitrate(g/ml)	957860	Float64
7.	Population	957860	Float64
8.	Fish_Length(cm)	957860	Float64
9.	Fish_Weight(g)	957860	Float64

the LSTM layer, and the dense layer. These components collectively form the intricate structure of the LSTM-SS model, facilitating the accurate prediction of nitrate concentration in the aquaponic system.

Input Data Preprocessing

Before input data is fed into the model, it undergoes preprocessing such as reshaping. This preparation phase ensure that the input features related to the aquaponics system are appropriately formatted for LSTM modeling.

Bidirectional LSTM Layers

The LSTM-SS incorporates two Bidirectional LSTM layers, enhancing its ability to capture not only past information but also future dependencies within the time series data. This bidirectional approach significantly contributes to the model's capacity to make accurate nitrate concentration predictions.

First Bidirectional LSTM Layer (512 Neurons):

In the model, the first Bidirectional LSTM layer, as represented in the equation (*6.1*), consists of 512 neurons and utilizes the Swish activation function. It processes the input sequence, capturing both forward and backward dependencies, as demonstrated by the equation (*6.2*).

$$h^{f1} = \text{swish}(W_x^{f1}x + W_h^{f1}h^{f1} + b_h^{f1}) \tag{6.1}$$

$$h^{b1} = \text{swish}(W_x^{b1}x + W_h^{b1}h^{b1} + b_h^{b1}) \tag{6.2}$$

Where:

- h^{f1} represents Forward hidden state in the first Bidirectional LSTM layer.
- h^{b1} represents Backward hidden state in the first Bidirectional LSTM layer.
- *x* represents the input features such as Temperature, Turbidity, Dissolved Oxygen, pH, Ammonia, Population, Fish Weight, Fish Length.
- W_x^{f1} represents the weight matrix for the input features(*x*) in the forward hidden state.
- W_h^{f1} represents the weight matrix for the previous forward hidden state (h^{f1}).
- b_h^{f1} represents the bias term for fine-tuning the forward hidden state.
- W_x^{b1} represents the weight matrix for the input features(*x*) in the backward hidden state.
- W_h^{b1} represents the weight matrix for the previous backward hidden state (h^{b1}).
- b_h^{b1} represents the bias term for fine-tuning the backward hidden state.

Second Bidirectional LSTM Layer (256 Neurons):

Following the first layer, the second Bidirectional LSTM layer, also with 256 neurons and swish activation, further refines the modeling of temporal dependencies. The utilization of bidirectional processing in this layer enhances its ability to capture intricate patterns in both forward and backward directions, contributing to the comprehensive temporal understanding represented in the equations (*6.3*) and (*6.4*).

$$h^{f2} = \text{swish}(W_x^{f2}x + W_h^{f2}h^{f2} + b_h^{f2}) \tag{6.3}$$

$$h^{b2} = \text{swish}(W_x^{b2}x + W_h^{b2}h^{b2} + b_h^{b2}) \tag{6.4}$$

Where:

- h^{f2} represents the forward hidden state in the second Bidirectional LSTM layer.
- h^{b2} represents the backward hidden state in the second Bidirectional LSTM layer.
- *x* represents the input features such as Temperature, Turbidity, Dissolved Oxygen, pH, Ammonia, Population, Fish Weight, Fish Length.
- W_x^{f2} represents the weight matrix for the input features(x) in the forward hidden state in the second Bidirectional LSTM layer.
- W_h^{f2} represents the weight matrix for the previous forward hidden state (h^{f2}).
- b_h^{f2} represents the bias term for fine-tuning the forward hidden state in the second Bidirectional LSTM layer.
- W_x^{b2} represents the weight matrix for the input features(*x*) in the backward hidden state in the second Bidirectional LSTM layer.
- W_h^{b2} represents the weight matrix for the previous backward hidden state (h^{b2}).
- b_h^{b2} represents the bias term for fine-tuning the backward hidden state in the second Bidirectional LSTM layer.

LSTM Layer (128 Neurons):

The LSTM (Long Short-Term Memory) layer in this aquaponics nitrate concentration estimation plays a vital role in modeling temporal dependencies within sensor data. As depicted in the equation (*6.5*), This layer significantly contributes after the Bidirectional LSTM layers by further capturing and preserving historical context in the data. This additional layer excels at retaining long-term patterns, allowing the neural network to make precise nitrate concentration predictions.

$$h^3 = \text{swish}\ (W_x^3 x + W_h^3 h^3 + b_h^3) \tag{6.5}$$

Where,

- h^3 represents the hidden state of LSTM layer.
- W_x^3 represents the weight matrix for input features, adjusting their impact on the LSTM layer hidden state.
- W_h^3 represents the weight matrix associated with the previous hidden state in the LSTM layer, it determines how the previous hidden state impacts the current hidden state.
- b_h^3 represents the bias term for LSTM layer, which allows the model to capture offset or shifts in the data.

Dropout Layers

To mitigate overfitting and enhance the generalization capabilities of LSTM-SS, dropout layers with a dropout rate of 0.2 are strategically placed within the model architecture. These layers prevent the model from relying too heavily on specific data points, resulting in a more robust predictor.

Dense Layer

The final estimation of nitrate concentration is calculated by combining the information learned from both Bidirectional LSTM layers, along with the Dense layers and bias terms, as expressed in the equation (*6.6*). This dense layer integrates the features extracted from the preceding layers, providing a comprehensive understanding of the temporal dependencies and patterns within the aquaponic system's sensor data. The inclusion of bias terms contributes to the precision of the nitrate concentration prediction, signifying the completion of the LSTM-SS model's computational phases.

$$y = W_{hy}^{f1}h^{f1} + W_{hy}^{b1}h^{b1} + W_{hy}^{f2}h^{f2} + W_{hy}^{b2}h^{b2} + W_{hy}^{3}h^{3} + b_y \qquad \textbf{6.6}$$

Where:

- W_{hy}^{f1} , W_{hy}^{b1} , W_{hy}^{f2} , W_{hy}^{b2} , W_{hy}^{3} represents weights matrices that determine the impact of each hidden state on the estimated nitrate concentration. These weights are learned during training.
- $h^{f1}, h^{b1}, h^{b2}, h^{3}$ represents the hidden states from the Bidirectional LSTM layers and LSTM layer. They represent learned features and relationships within the input data.
- b_y represents the bias term that allows for adjustments to the final estimated nitrate concentration.

5.3. Hyperparameter Selection, Tuning, and Optimization for LSTM-SS

The performance of LSTM-SS relies on the careful selection, tuning, and optimization of various hyperparameters.

Activation Function

An essential component of the LSTM-SS is the activation function. In this work, a custom 'swish' activation function is employed, introducing non-linearity into the model and aiding in its capacity to capture complex relationship.

Number of LSTM units

The choice of the number of LSTM units in each layer is a critical hyperparameter that influences the model's ability to capture temporal dependencies. This parameter is optimized to strike a balance between model complexity and generalization.

Learning Rate

The learning rate for the Nadam optimizer is fine-tuned to control the convergence speed and accuracy of LSTM-SS, with an appropriate learning rate value of 0.001. This adjustment ensures that the model converges efficiently to a solution.

Batch Size

During the training phase, the batch size plays a crucial role. A batch size of 32 was selected with consideration for its impact on both the convergence speed and memory requirements of LSTM-SS. This careful selection ensures efficient training while managing memory consumption effectively.

5.4. Training LSTM-SS

The LSTM-based Soft-Sensor (LSTM-SS) was trained using a vast aquaponics dataset comprising nine key parameters, totally 9,57,860 data points each. To ensure robust evaluation, the dataset was split into an 80% training set (7,66,288 points) and a 20% validation set (1,91,572 points). The training process was conducted with five epochs, carefully selected to balance model convergence and computational efficiency.

5.5. Evaluation metrics of the proposed LSTM-SS

The performance of the proposed LSTM-SS was evaluated by using following standard statistical metrics:

5.5.1. Mean Squared Error (MSE)

Mean Squared Error (MSE) is a fundamental metric used to quantify the overall accuracy of LSTM-SS predictions. It measures the average squared difference between the model's predicted nitrate concentrations and the actual values from the validation dataset. The MSE is calculated as follows in the equation (*6.7*):

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \qquad \textbf{6.7}$$

Where:

- N represents the total number of validation data points.
- Y_i is the actual nitrate concentration for the *i*-th data point.
- $\hat{Y}_i$ is the nitrate concentration predicted by LSTM-SS for the *i*-th data point.

Lower MSE value indicated better predictive accuracy, with values closer to zero signifying minimal prediction error.

5.5.2. R-squared Score (R^2)

The R-squared score (R^2), also known as the coefficient of determination, assesses how well the LSTM-SS predictions fit the actual data. As shown in the formula (*6.8*), It measures the proportion of the variance in nitrate concentration that is explained by the model. The R^2 score ranges from 0 to 1, with higher values indicating better model performance. It is calculated as follows:

$$R^2 = 1 - \frac{MSE}{Var(Y)} \qquad \textbf{6.8}$$

Where:

- MSE is the Mean Squared Error.
- Var(Y) is the variance of the actual nitrate concentrations.

An R^2 score of 1 signifies a perfect fit, while a score of 0 indicates that the model's predictions are no better than simply using the mean of the observed data.

5.5.3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a metric that quantifies the average absolute difference between LSTM-SS predictions and actual nitrate concentrations. As depicted in the formula (*6.9*), MAE provides insight into the magnitude of errors made by the model. It is calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |Y_i - \hat{Y}_i| \qquad \textbf{6.9}$$

Where:

- N represents the total number of validation data points.
- Y_i is the actual nitrate concentration for the i-th data point.
- $\hat{Y}_i$ is the nitrate concentration predicted by LSTM-SS for the i-th data point.

A lower MAE indicates that the model's predictions closely align with the actual data.

7. Results and discussions

The LSTM-based Soft-Sensor (LSTM-SS) impressively estimates nitrate concentration in aquaponics system with a MSE of 0.00074, an R-squared score (R^2) of 0.98, and a MAE of 0.0170, showcasing its exceptional accuracy. Beyond these numerical evaluations, the model's strength lies in its adaptability to a diverse dataset, encompassing various environmental and system parameters, enabling it to discern intricate temporal patterns. This temporal awareness, facilitated by bidirectional LSTM layers, is particularly crucial for precise nitrate concentration predictions. Moreover, LSTM-SS's robust training ensures its generalization to various environmental conditions. In practice, this translates to real-time nitrate concentration monitoring and management, revolutionizing aquaponics system control while reducing the labor and time required for manual measurements. Ultimately, LSTM-SS emerges as a powerful tool for enhancing aquaponics system efficiency and sustainability, with implications that extend far beyond numerical metrics.

Actual vs Predicted Graph

The performance assessment of the LSTM-based nitrate estimation system begins with the examination of actual vs predicted graphs. These graphical representations serve as essential tools in evaluating the model's accuracy and reliability in forecasting nitrate concentration changes. The visual comparison between the model's predictions and actual nitrate concentration values provides valuable insights into the system's predictive capabilities.

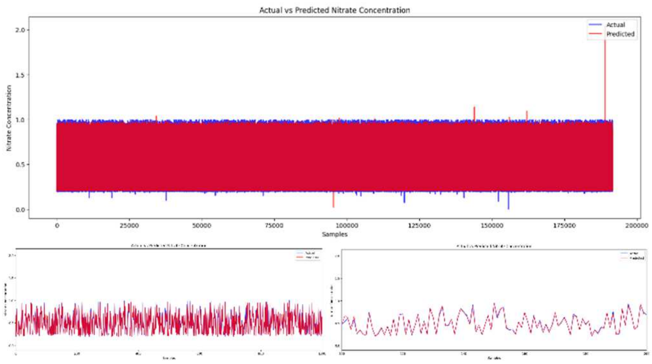

Fig 3. *Actual vs Predicted Graph*

In **Fig 3**, showing the actual vs predicted graph in three different zoomed-in views. Upon close inspection, the graphical representation reveal a remarkable alignment, to the extent that the predicted and actual curves appear to converge into a singular line. This convergence underscores that the model is very good at making accurate predictions. However, the slight deviations observed at certain points in the graph can be attributed to inherent uncertainties in real-world systems. Factors such as environmental fluctuations, measurement errors, or variations in aquaponics conditions may contribute to these minor divergences. It's crucial to recognize that the model's ability to closely track the actual values, despite these small discrepancies, showcases its robust performance.

Residual Plot

In addition to actual vs predicted graphs, the residual plot visually represents the discrepancies between the actual nitrate concentrations and the corresponding predictions generated by the LSTM-SS model, As depicted in **Fig 4**, the x-axis denotes the actual nitrate concentrations, while the y-axis represents the residuals, defined as the differences between the observed values and the model predictions.

The resulting residual plot provides a visually compelling representation of the model's performance in estimating nitrate concentrations. The distribution of residuals, centered and scattered around the zero-line, signifies a well-fitted model. Notably, the plot showcases a prominent concentration of residuals aligning closely with the actual nitrate concentrations. Furthermore, a subtle downward trend in the outer regions implies that the model tends to slightly underestimate nitrate concentrations. This subtle observation, far from indicating a significant discrepancy, suggests a conservative bias, highlighting the model's cautious approach in predicting nitrate levels. Overall, the residual plot underscores the model's effectiveness in capturing the underlying patterns within the aquaponics data, instilling confidence in its predictive capabilities.

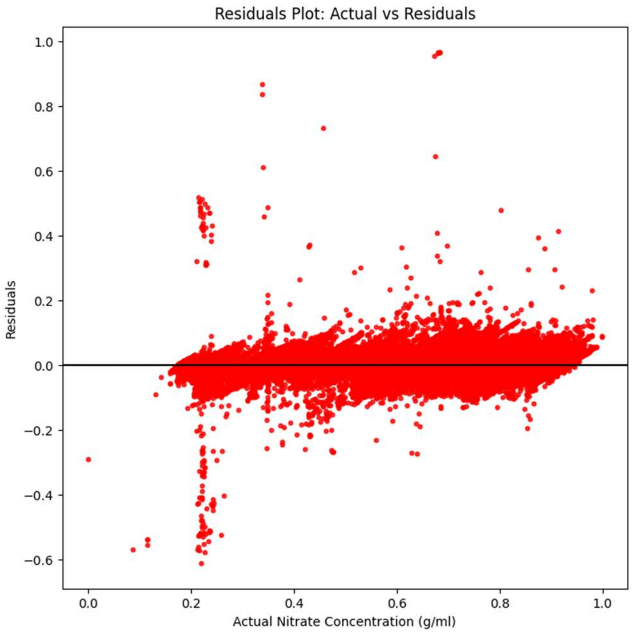

Fig 4. Residual Plot

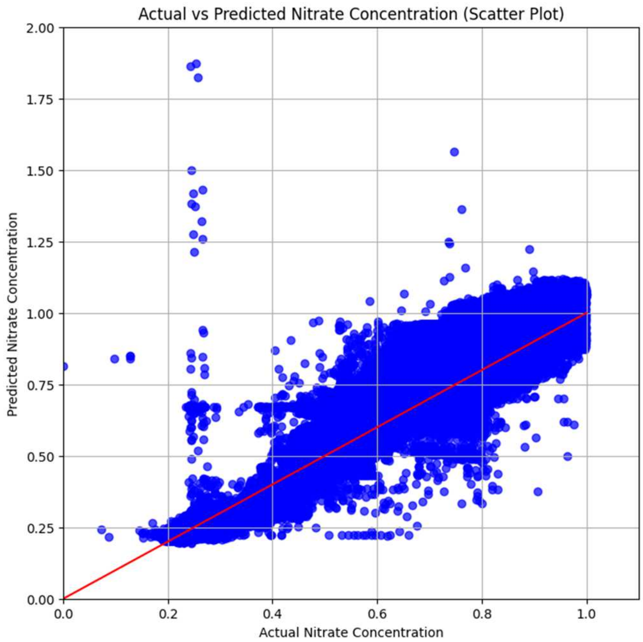

Fig 5. Scatter Plot

Scatter Plot

In addition to residual plot, the employment of scatter plot offers a comprehensive view of the LSTM model's predictive abilities. The scatter plot, as presented in **Fig 5**, provides a clear representation of the relationship between actual nitrate concentrations and corresponding LSTM-predicted values. The close clustering of data points around the regression line indicates a robust positive correlation, affirming the model's accuracy. The thick scattering of points around this line demonstrates the model's precision in capturing variations in nitrate levels, offering reliable real-time insights. While the majority of points align closely with the line, a few scattered outliers suggest minor deviations in prediction accuracy. These outliers, though sparse, prompt consideration for potential areas of model improvement. Overall, this visualization assures the model's effectiveness in predicting nitrate dynamics, emphasizing its practically for real-world applications without the need for extensive resources or maintenance. Moreover, the model's ability to accurately predict nitrate concentrations across various conditions positions it as a valuable tool for aquaponics management. The real-time insights provided by the scatter plot empower practitioners to make informed decisions, with the model's minimal maintenance requirements, underscored its practically as a reliable and accessible solution for enhancing aquaponic system control and sustainability.

Error Distribution Plot

To deepen the understanding of the model's performance, an error distribution plot was constructed, as illustrated in **Fig 6**. This plot serves as a valuable tool for evaluating the distribution of prediction errors, providing insights into the extent of disparities between predicted and actual values.

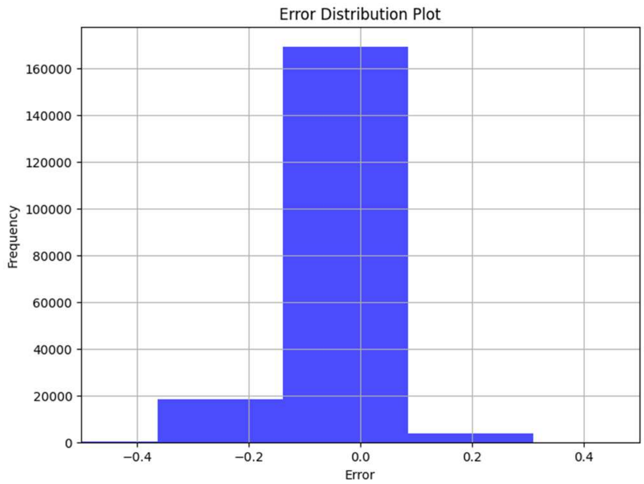

Fig 6. Error Distribution Plot

An ideal scenario is reflected by a symmetrical distribution centered around zero error, signifying model accuracy. Upon scrutinizing the error distribution plot, observing a nearly symmetrical distribution with minimal outliers, underscores the model's consistency and reliability in estimating nitrate concentrations. The model's performance, as explained by the error distribution plot, aligns seamlessly with the overall accuracy and adaptability of our innovative nitrate estimation system. This comprehensive analysis further strengthens our confidence in the model's ability to deliver precise and reliable predictions across diverse scenarios.

In summary, the results from this study, supported by the actual vs predicted graphs, residual plot, scatter plot and error distribution plot, provide substantial evidence of the critical significance of accurate nitrate estimation in safeguarding the water quality for both plants and fish in aquaponics systems. The LSTM-based system not only optimizes nitrate estimation but also elevates its precision, establishing itself as an invaluable instrument for ensuring the purity and well-being of aquatic life and cultivated plants within this ecosystem, thereby promoting sustainable and efficient food production in aquaponics.

8. CONCLUSION

In conclusion, the research highlights the LSTM-based Soft-Sensor (LSTM-SS) as a versatile and precise tool for estimating nitrate concentration in aquaponics systems. The model achieves noteworthy numerical metrics, including a mean squared error (MSE) of 0.00074, an R-squared score (R^2) of 0.98, and a mean absolute error (MAE) of 0.0170, showcasing its exceptional accuracy. However, the significance of LSTM-SS goes beyond these quantitative measures. LSTM-SS's robust training procedures contribute to its applicability across different environment conditions, ensuring its reliability in various aquaponic settings. The practical impact of this adaptability becomes evident in real-time nitrate concentration monitoring and management. By automating these processes, LSTM-SS transforms aquaponics system control, offering more efficient and sustainable alternative to manual measurements that typically demand significant labor and time investments. LSTM-SS stands out as a powerful and versatile tool that surpasses numerical metrics, positioning itself as a catalyst for improving aquaponics system efficiency and sustainability.

9. FUTURE WORK

For future work, exploring the integration of LSTM-based Soft-Sensor (LSTM-SS) into aquaponics systems presents a promising direction. Further research can focus on optimizing the model for diverse environmental conditions, scaling its application to commercial settings, and refining its real-time monitoring capabilities. Additionally, investigating potential enhancements, such as incorporating additional parameters or expanding the dataset, will contribute to the ongoing evolution of efficient and sustainable aquaponic systems.

REFERENCES

[1] T. Y. Kyaw and A. K. Ng, "Smart aquaponics system for urban farming," Energy procedia, vol. 143, pp. 342–347, 2017.

[2] S. López-Tejeida, L. Felix-Cuencas, J. J. De León-Ramirez, L. B. Flores-Tejeida, M. Villegas-Villegas and J. F. García-Trejo, "Methods to measure nitrogen content in fish culture: A review," *2021 XVII International Engineering Congress (CONIIN)*, Queretaro, Mexico, 2021, pp. 1-5.

[3] Hui Li, Weijun Li, Matthew McEwan, Daoliang Li, Guoping Lian, Tao Chen, "Adaptive filtering-based soft sensor method for estimating total nitrogen in aquaponic systems", Computers and Electronics in Agriculture, Volume 186,2021,106175, ISSN 0168-1699

[4] Pattnaik, B.S., Pattanayak, A.S., Udgata, S.K. *et al.* Machine learning based soft sensor model for BOD estimation using intelligence at edge. *Complex Intell. Syst.* **7**, 961–976 (2021).

[5] Boyan Xu, Ching Kwek Pooi, Kar Ming Tan, Shujuan Huang, Xueqing Shi, How Yong Ng, "A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance", Journal of Water Process Engineering, Volume 54, 2023, 104041, ISSN 2214-7144.

[6] C. S. Arvind, R. Jyothi, K. Kaushal, G. Girish, R. Saurav and G. Chetankumar, "Edge Computing Based Smart Aquaponics Monitoring System Using Deep Learning in IoT Environment," *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020, pp. 1485-1491.

[7] M. F. Taha, G. ElMasry, M. Gouda, L. Zhou, N. Liang, A. Abdalla, D. Rousseau, and Z. Qiu, "Recent advances of smart systems and internet of things (iot) for aquaponics automation: A comprehensive overview," Chemosensors, vol. 10, no. 8, p. 303, 2022.

[8] Hsiang-Yang Shyu, Cynthia J. Castro, Robert A. Bair, Qing Lu, and Daniel H. Yeh, "Development of a Soft Sensor Using Machine Learning Algorithms for Predicting the Water Quality of an Onsite Wastewater Treatment System",ACS Environmental Au **2023** 3 (5), 308-318.

[9] Karimanzira, Divas and Rauschenbach, Thomas. "An intelligent management system for aquaponics" *at - Automatisierungstechnik*, vol. 69, no. 4, 2021, pp. 345-350.

[10] Alex Sherstinsky. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network, 2018. cite arxiv:1808.03314

[11] P. Kumar, P. Tiwari and U. S. Reddy, "Estimating fish weight growth in aquaponic farming through machine learning techniques," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-7.

[12] Gilbert Chandra D., Vinoth B., Srinivasulu Reddy U., Uma G., Umapathy M., "Recurrent Neural Network based Soft Sensor for flow estimation in Liquid Rocket Engine Injector calibration", Flow Measurement and Instrumentation, Volume 83, 2022, 102105, ISSN 0955-5986.

[13] W. Ke, D. Huang, F. Yang and Y. Jiang, "Soft sensor development and applications based on LSTM in deep neural networks," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 2017, pp. 1-6