Journal of Advanced Research in Applied Mechanics

Journal homepage:
https://semarakilmu.com.my/journals/index.php/appl_mech/index
ISSN: 2289-7895

IoT-based Machine Learning Comparative Models of Stream Water Parameters Forecasting for Freshwater Lobster

Abdelmoneim A. Bakhit[1], Nur Syahirah Mohd Sabli[1], Mohd Faizal Jamlos[1,2,*], Mohd Aminudin Jamlos[3], N.H. Ramli[3], Muhammad Aqil Hafizzan Nordin[1], Nura A. Alhaj[1], Ehtesham Ali[1]

[1] Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang Al Sultan Abdullah, 26600, Pekan, Malaysia
[2] Centre for Automotive Engineering, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600, Pekan, Malaysia
[3] Advanced Communication Engineering (ACE), Centre of Excellence, Faculty of Electronics Engineering Technology, Universiti Malaysia Perlis, 26600, Arau, Malaysia

ARTICLE INFO

Article history:
Received 7 February 2024
Received in revised form 25 March 2024
Accepted 8 April 2024
Available online 30 May 2024

Keywords:
Real-Time; Predictive Analytics; ARIMA; NNETAR; Naïve Bayes

ABSTRACT

Water quality parameters such as dissolved oxygen, pH, and mineral content are important factors for aquaculture. Predictive analytics can predict water conditions in aquaculture and significantly reduce the mortality probability of aquaculture products. This paper applied stream predictive analytics to the freshwater lobster farming dataset where its real-time data supplied by End Node Unit (ENU) which integrated with dissolved oxygen (DO), potential hydrogen (pH), electrical conductivity (EC), and total dissolved solids (TDS). The real-time data of ENU in Structured Query Language (SQL) is normally displayed for remote monitoring and the analytics will only be done after in different processing platform called batch analytics. Instead of batch, this paper demonstrates capability of stream analytics where the real-time data query from ENU streaming through Structured Query Language (SQL) right into R Studio and Autoregressive Integrated Moving Average (ARIMA) predictions executed on the query table simultaneously on the same processing platform. Previously, ARIMA, Neural Network Autoregressive (NNETAR), and Naïve Bayes, were run and evaluated in R Studio to identify the best algorithm for stream analytics. Prediction procedure in R studio start with importing real-time data stored in SQL database and stream into R Studio using command of "dbGetQuery(con,sql)". These three models evaluated the performance of freshwater lobster water conditions, dissolved oxygen (DO), potential hydrogen (pH), electrical conductivity (EC), and total dissolved solids (TDS). The data was collected for six months, and 70% was used as training data and 30% as test data. Compared to NNETAR and Naïve Bayes, ARIMA fits the entire data set well for 7 days; the ARIMA model exhibited lower absolute errors for pH and electrical conductivity, with errors ranging from 0.04 to 1.7 across days, while the NNETAR model had generally lower errors for TDS, with errors ranging from 0.3 to 0.7; however, the Naïve Bayes model's performance varied, with the lowest error for DO on day (5) 0.15 but higher errors for other parameters and days, including the highest error for electrical conductivity on day (6) 6.2. In conclusion, the average absolute errors for DO, pH, EC, and TDS are 0.163, 0.064, 0.705, and 0.498, respectively. Our findings underscore the efficacy of ARIMA for comprehensive water quality via stream prediction while highlighting the nuanced strengths and weaknesses of each model in forecasting specific parameters. This study contributes to the aquaculture literature by providing a nuanced comparative analysis of predictive models tailored to freshwater lobster farming, emphasizing the imperative role of stream predictive modelling. It enables real-time monitoring of water quality parameters, ensuring prompt interventions to maintain optimal conditions, thereby minimizing risks, enhancing aquaculture productivity, and ultimately contributing to sustainable and efficient freshwater lobster farming practices.

* *Corresponding author.*
E-mail address: mohdfaizaljamlos@gmail.com

https://doi.org/10.37934/aram.117.1.137149

1. Introduction

Aquaculture is an important source of income for rural Malaysian communities. Currently, freshwater aquaculture plays a role in improving rural communities' food security and standard of living. This is because freshwater fish farming requires less space and capital, making it more accessible to small-scale producers [1,2]. The cultivation of crayfish is one of the most common forms of freshwater aquaculture. Freshwater lobster is an excellent candidate for cultivation in Malaysia due to its high economic value, rapid growth rate, and favourable natural environmental conditions [3]. In freshwater lobster cultivation, the environment and water quality parameters play an essential role in ensuring the survival and growth of the lobster. Parameters of water quality must be monitored and analysed constantly. Therefore, accurate measurements of water quality parameters such as pH, TDS, EC, and DO must be taken continuously and in real-time [4,5]. Several studies [6,7] have underscored the significance of assessing pH, TDS, EC, and DO as a preventive measure against elevated water quality levels that could lead to severe detriments for lobsters, such as anoxia, hyperoxia, and hypoxia. The term "water quality monitoring" pertains to obtaining water samples and conducting analyses, with these parameters serving as indicators for assessing water quality. In aquaculture, the need for continuous 24 × 7 monitoring of water quality and precise feeding is evident. However, there's a potential downside, as persistent monitoring might disrupt the bacterial equilibrium in the aquaculture ecosystem, potentially diminishing the disease-resistance capabilities of lobsters [8]. Conventional aquaculture practices depended on the expertise of seasoned aquafarmers, who relied on observation and experiential judgment to detect and predict health risks in the farming environment. Monitoring shifts in water quality indicators, such as dissolved oxygen levels, pH, temperature, and salinity, alongside other factors recognized for their adverse effects on aquaculture, constitutes one crucial aspect. For optimal lobster cultivation, lakes and fisheries must maintain specific conditions, including a dissolved oxygen range of 5.60 ± 0.9 mg/L and a pH level of 7.12 ± 0.21 [9]. Continuous water quality monitoring is imperative to prevent substances beyond the acceptable limits for lobster tolerance. Deteriorating water conditions can result from multiple factors, including water sources, weather impact, sediment accumulation, and more. The application of traditional techniques may negatively impact lobster development, underscoring the need for meticulous care by lobster cultivators to ensure the production of high-quality lobsters. Neglecting to monitor water quality consistently can lead to the unfortunate demise of the lobster population [10].

Therefore, accurate anticipation of shifts in water quality trends using pertinent big data and intelligent algorithms can facilitate risk management through technological interventions. This, in turn, holds the potential to significantly enhance the operational efficiency of mariculture production [11]. Furthermore, aquaculture information technology serves a dual role: it enables acquiring highly accurate real-time data regarding aquatic environment parameters. It empowers farmers with timely alerts regarding any deviations from normal parameters. Moreover, it possesses the capability to address these anomalies autonomously. This highlights the emergence of intelligent aquaculture, resulting from the seamless integration of artificial intelligence, Internet of Things technology, and traditional aquaculture practices. This synergistic approach effectively manages risks associated with aquaculture, enhances the quality of aquatic products, and safeguards the survival rates of freshwater lobsters. Ultimately, this contributes to the sustainable advancement of the aquaculture sector [12].

There are many studies aimed at predicting water quality using machine learning models. Z. Ma *et al.,* [13] Utilizing a backpropagation neural network (BP-NN) framework, a prediction model was employed to estimate the water quality within high-density (300 PLs/m2) *Litopenaeus vannamei* shrimp tanks. Within this model, the hidden layer was equipped with a tangent sigmoid transfer function, while the output layer utilized a linear transfer function. This model was crafted using observed water quality data spanning 120 days, with weekly assessments carried out in four shrimp tanks. Among the input variables, nine parameters were chosen: water temperature, pH, total ammonia nitrogen, nitrite nitrogen, nitrate nitrogen, dissolved inorganic phosphorus, chlorophyll-a, chemical oxygen demand, and five-day biochemical oxygen demand. Through an iterative process, the number of nodes in the hidden layer was fine-tuned, identifying five optimal neuron nodes. The correlation coefficients between computed outcomes and experimental values for the training, testing, and combined training and testing datasets stood at 0.990, 0.979, and 0.992, respectively.

Authors In [14], A novel approach was undertaken to create a predictive model for estimating dissolved oxygen levels within water systems used in pond cultivation, specifically focusing on shrimp ponds. The method integrated K-means clustering and the Gated Recurrent Unit (GRU) neural network. During this study, the author formulated a model to forecast dissolved oxygen levels in pond culture water over one month. The dataset utilized for this research was sourced from the Ayue aquaculture farm in China. The outcomes revealed a Root Mean Square (RMS) value of 0.264 and an RMSE of 3.5%, signifying the model's favorable performance in predicting DO levels accurately.

A comparison was conducted between Long Short-Term Memory Networks (LSTM) and ARIMA methodologies to assess the effectiveness of error rate reduction for prediction. The research demonstrates the superiority of LSTM over ARIMA, with LSTM consistently outperforming the latter. The results indicate that LSTM yields an average reduction of 84–87 percent in error rates compared to ARIMA, showcasing the clear advantage and higher performance of the LSTM approach [15]. In this research paper, the ARIMA approach was specifically ARIMA, tailored to suit the dataset's characteristics under consideration and contrasted against the LSTM model. The project dealt with streaming data of diverse patterns, which the LSTM model might not optimally accommodate. The findings demonstrate that the ARIMA model yielded superior outcomes to regression models based on deep learning. Notably, the ARIMA model achieved the most favourable results, with a Mean Absolute Percentage Error (MAPE) of 2.76% and a Root Mean Square Error (RMSE) of 302.53. Intriguingly, among the deep learning models, the Gated Recurrent Unit (GRU) displayed better performance than the Long Short-Term Memory (LSTM) variant, exhibiting a MAPE of 3.97% and an RMSE of 381.34 [16]. The hybrid Autoregressive Integrated Moving Average-Nonlinear Autoregressive ARIMA-NAR model displayed enhanced modelling performance, showcasing reduced values in terms of mean square error, mean absolute error, and mean absolute percentage error, specifically measuring at 0.2209, 0.1373, and 0.0406. This improvement in accuracy rendered the ARIMA-NAR hybrid model more capable of forecasting Tuberculosis (TB) incidence with greater precision when contrasted with the performance of the standalone ARIMA model [17].

ARIMA and NAR neural networks incorporating four delays and a hidden layer with twelve layers were employed on the dataset. The data spanned from January 2007 to March 2016. While this model demonstrated strong predictive capabilities, its applicability to streaming datasets is limited due to the dynamic nature of the ARIMA's p,d,q variables. These variables need frequent adjustment to align with the dataset's characteristics, posing a challenge when dealing with streaming data [18].

Therefore, stream predictive analysis for pH, TDS, and EC in freshwater lobster cultivation revolutionizes aquaculture management, enabling proactive adjustments based on real-time data insights. This precision approach optimizes water quality, safeguards lobster health, and enhances productivity, significantly advancing sustainable aquaculture practices [19,20].

In the context of freshwater lobster aquaculture, this study seeks to address existing drawbacks and gaps in the literature related to water quality prediction. The prevailing literature highlights the critical role of continuous monitoring for parameters such as pH, TDS, EC, and DO in ensuring the well-being of lobsters. However, the potential disruption to the bacterial equilibrium and consequent impact on lobster health due to persistent monitoring, as indicated by prior studies [8], underscores a need for a refined approach. Our proposed methodology contributes by employing stream predictive analytics, allowing for real-time adjustments based on dynamic data insights. Integrating ARIMA, NNETAR, and Naïve Bayes models offers a nuanced comparative analysis, filling a gap in the literature that lacks comprehensive evaluations of predictive models tailored to freshwater lobster farming. Additionally, the literature emphasizes the significance of accurate anticipation of shifts in water quality trends [11], and our study directly addresses this by evaluating the efficacy of predictive models in a streaming data environment.

In reviewing existing papers, it's evident that various predictive models have been applied to water quality prediction in aquaculture. However, there is a notable gap in the literature regarding the specific application of stream predictive analytics to freshwater lobster farming. Our study aims to bridge this gap by providing insights into the performance of ARIMA, NNETAR, and Naïve Bayes models in predicting water conditions for freshwater lobster cultivation. Including a detailed comparative analysis, especially focusing on nuanced strengths and weaknesses, enhances the current understanding and guides future research in this domain.

The structure of this paper is as follows: The first section is the introduction. The second section is methodology. The third section results and discusses the sensors' accuracy and the water quality prediction. The last section is the conclusion.

2. Methodology

2.1 Prediction of Water Quality Parameters

The End Node Unit (ENU) mainly consists of two-layer structures; the lower layer is for sensing the water parameters, while the upper layer is for sending the data to GU via Lora module. The electronic circuit board of ENU consists of a microprocessor unit, Adafruit Lora RFM9x at 915 MHz, 900MHz Antenna, and sensor interfacing ports. Data acquisition for this experiment is gathered from wireless sensor datasets stored in a database on a server. were transmitted wirelessly through the Lora protocol to the central station in the server room. The central station then processed the received buffer data and saved it on the database server. The data sensors in the database are downloaded and sorted according to date, time, and day. Figure 1 shows the topology of data acquisition for this experiment, The sensors utilized in the freshwater lobster breeding tank include the DO sensor (DFROBOT SEN0237-A), pH sensor (pH meter-SEN0161-DFRobot), EC, and TDS sensor (Gravity Analog TDS/EC Sensor). The prototype is meticulously designed to adhere to operational requirements even in adverse environmental conditions such as hot, humid, and rainy weather. Housed within an IP56 enclosure box, the prototype exhibits robustness and waterproof characteristics, as illustrated in Figure 2 shows the End Node Unit (ENU) prototype and the ENU at the freshwater lobster breeding tank. Notably, the deployed sensors are of industrial-grade quality and hold standard certifications to ensure heightened precision, accuracy, and reliability in data collection. Furthermore, the sensors undergo thorough calibration procedures, as detailed in Table 1, guaranteeing the acquisition of accurate and dependable data.

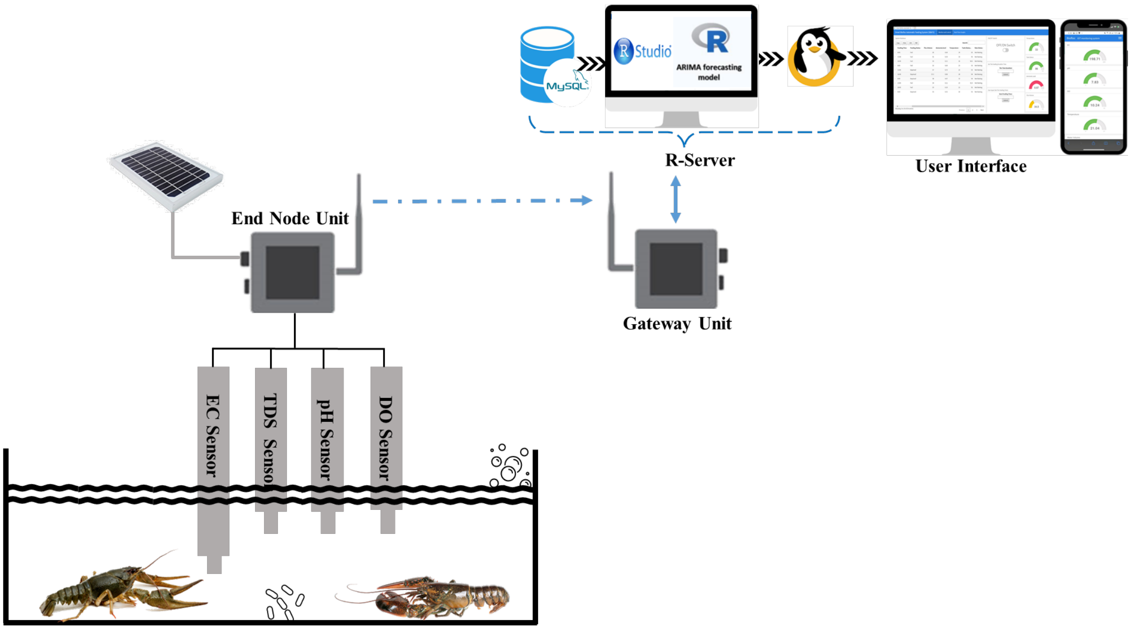
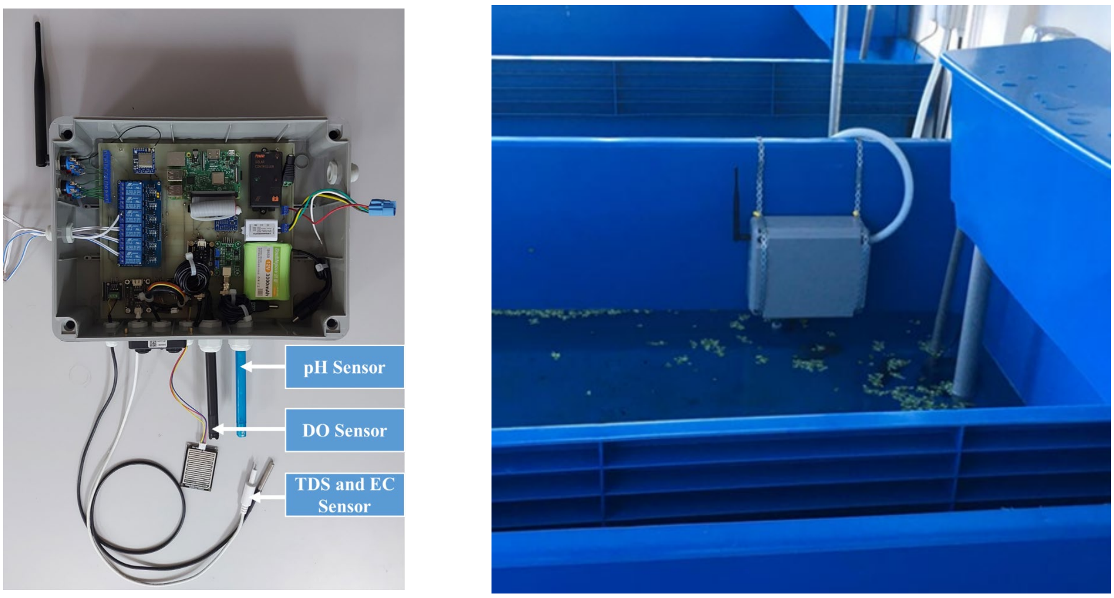


Fig. 1. The topology of the wireless system in freshwater lobster farming



(a) (b)

Fig. 2. (a) Prototype of ENU (b) The wireless system collects the water quality parameter at the breeding freshwater lobster tank

Table 1
End-node sensor device

Measured Parameter	Measuring range	Accuracy
pH [21]	0-14 pH	± 0.1pH
DO [22,23]	0–20.00 mg/L	± 0.2 mg/L
TDS & EC	0 ~ 1000 mg/L	± 10%

The data collection site is located in the southern part of Kuantan in Pahang State, near Sepat Beach, and has a lot of crustacean breeding ponds. This experiment used data from the freshwater lobster pond, and parameters such as oxygen level, mineral level, and temperature were used. The collected data were saved on a database server and sent to the query table of R-studio for pre-processing.

As previously mentioned, the proposed system able to execute streaming predictive analytics. The analytics executed as a standalone analytics server by importing data from SQL database to query table of R Studio software to perform the stream prediction. The R Studio version 1.4.1103 is installed to predict the water parameters of DO, pH, EC, and TDS. Several libraries such as DBI, ggplot2 and forecasting are installed as well in R studio for importing data set from SQL database, plotting graphs and performing prediction. Generally, the proposed system prediction procedure started from data collection which continuously being saved into SQL database, see Figure. 3. From database, the data query is streaming in real-time into R Studio environment and predicting will be performed.

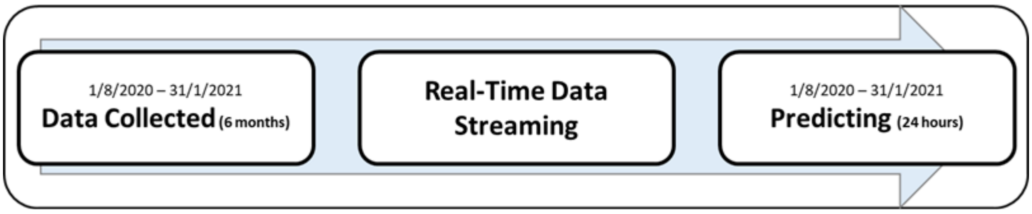


Fig. 3. Prediction stage flowchart

Prediction procedure in R studio start with importing real-time data stored in SQL database and stream into R Studio using command of "dbGetQuery(con,sql)". Since SQL database required password to access the data set, "con" command is used to grant the access by providing the host, user, password and dbname value. Next, "[c(pH)]" command is used to select the class to perform prediction in the data set. Prediction execution pseudo-code for pH parameter is shown in the Figure 4. The other parameters of DO, EC and TDS follow the exact steps but different in term of coding. Each ARIMA, NNETAR, and Naïve Bayes model have gone through the prediction flow.

```
con <- DBI::dbConnect(RMariaDB::MariaDB(),
                      dbDriver   = ("MySQL"),
                      host   = "127.0.0.1",
                      user      = "root",
                      password      = "830321fj",
                      dbname = "IOT_DB",
                      port     = 3306)


sql <- "SELECT PH FROM WaterQuality ORDER BY ID DESC LIMIT 24;"
dataSet <- dbGetQuery(con,sql)
getwd()

HazardAcid <- read.csv("HazardAcid.csv", header = TRUE, sep = ',')
HazardAlkali <- read.csv("HazardAlkali.csv", header = TRUE, sep = ',')
clock <- read.csv("clock.csv", header = TRUE, sep = ',')

table(unlist(lapply(dataSet,class)))
PH <- as.numeric(dataSet[, c("PH")])
plot(PH)

autoarima1<-auto.arima(PH, seasonal = FALSE, stepwise = FALSE, approximation = FALSE)
forecast1<-forecast(autoarima1, h=24)
```

Fig. 4. Pseudo code for pH parameter predictive analytics in R studio

As for the ARIMA model, auto-ARIMA was used as it worked efficiently to describe the values of p, d, and q. Auto-ARIMA uses a combination of unit root tests, minimization of Akaike Information Criterion (AIC), and maximum likelihood (MLE) to obtain an ARIMA model. P, d, and q values are chosen by minimizing AICs. Varying the ARIMA model is considered by varying p, d, and q from the current model test. In this experiment, the predictive models use 70% of 6 months of data for training, while 30% are used for testing. Absolute error was used in the perception of models by comparing the real data with the prediction model's result for 7 days. Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE) are commonly used to assess the predictive accuracy of a model [24]. Among them, MAE better reflects the actual situation of the prediction error value. Its calculation formula is:

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|z - z_j\right| \tag{1}$$

where z is the original sequence value and z_j is the predicted value, n is the number of samples.

The model NNETAR, with its neural network model, fits a time series. The vector value containing the legged values of series and f is a neural network with four hidden nodes in a single node. The error series is assumed to be homoscedastic. The forecast.nnetar command was used to control the simulation done in R-studio. As for Naïve Bayes, the forecast.naive command was used to predict the value of each parameter. An evaluation of absolute errors was done to compare which model fits well with this experiment. The flow of prediction model comparison in R is shown in Figure 5. A further explanation of the results of the prediction model comparison is given in section 3.

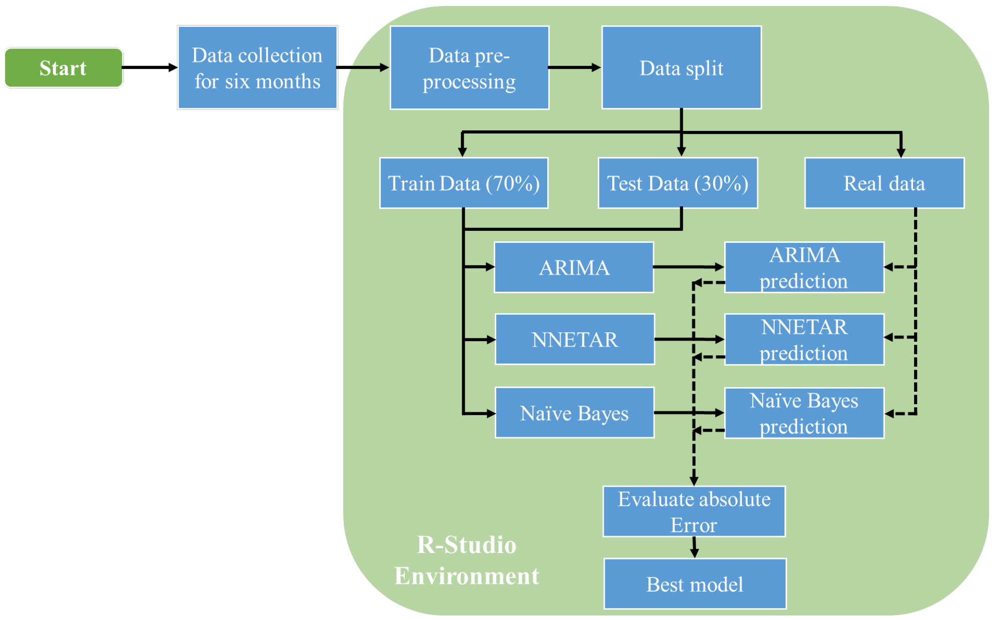


Fig. 5. Flow of prediction model experiment

3. Results

3.1 Sensors Accuracy Results

The proposed system was tested for six months under different environmental conditions. During this period, accuracy is tested and verified to ensure accurate predictive analytics based on real-time data. It was noticed that the DO and pH sensors cap probes had to be changed and filled with solution monthly for accurate results. To validate the accuracy and reliability of sensors and their prediction, an on-shelf independent system for TDS, EC, pH meter (Hanna 981,017, USA), and DO AZ8403 meter was taken as a reference to compare with the proposed system. Figure 6 shows the results of the accuracy percentage error for pH, TDS, EC, and Do. Table 2 shows summary statistics of the data, including the minimum and maximum values for each parameter. The maximum accuracy percentage error is ≈10%, which is acceptable [25].

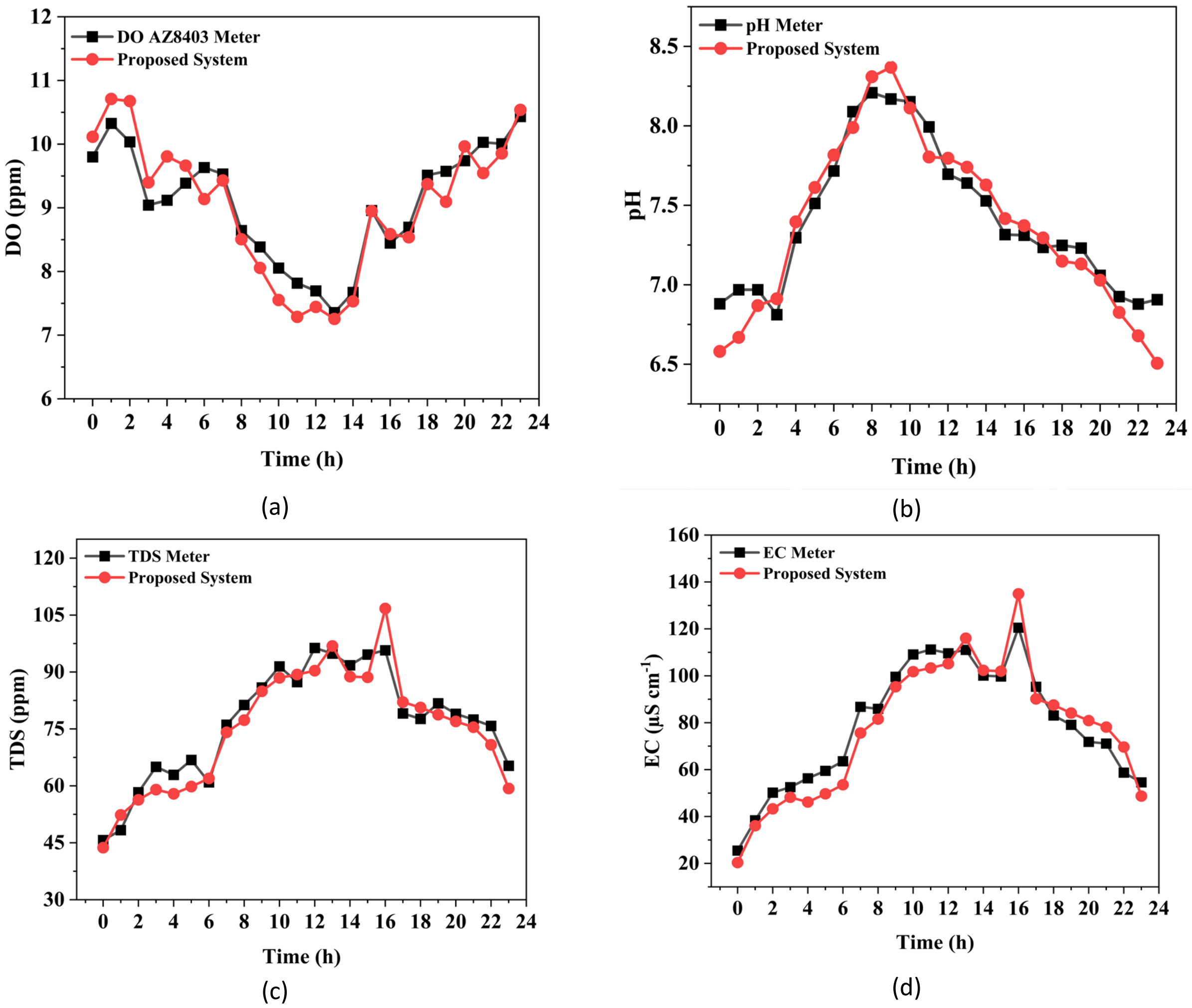


Fig. 6. (a) Manual measurement vs. Proposed system (a) DO (ppm), (b) pH, (c) TDS (ppm), (d) EC (µS cm^{-1})

Table 2
Minimum and Maximum Accuracy Percentage Error

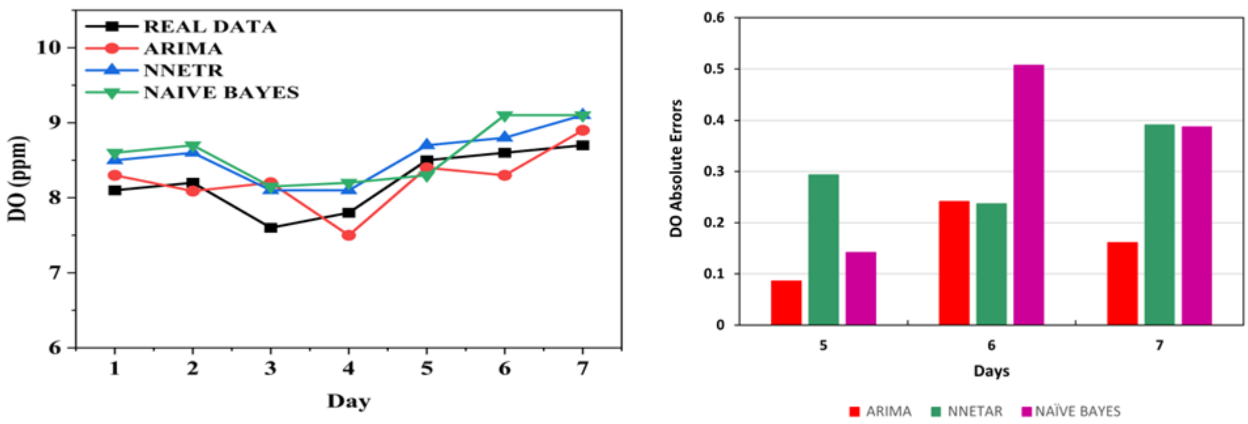
Parameter	Accuracy Percentage Error	Minimum %	Maximum%
DO	3.27	0.09	7.5
EC	2.95	0.88	5.34
pH	1.81	0.42	5.79
TDS	3.00	0.73	7.54

3.2 Validation of Water Quality Prediction

In this experiment, water quality for freshwater lobsters is evaluated using four variables: DO, EC, TDS, and pH. observable measures the level of oxygen diffused into the freshwater lobster's breeding environment. Regarding the parameter, the DO in the freshwater lobster's breeding tank for the testing period has shown a normal reading between 7 and 9 ppm.

During six months of data collection, it was observed that the DO value was deteriorating between 1 ppm and 4 ppm but within the normal range. The DO value in aquaculture water is susceptible to solar radiation, temperature, and another environmental factors. It is cyclical overtime. Using three models, ARIMA, NNETAR, and Naïve Bayes, the average value of DO for seven days was analyzed and the average of DO periods was predicted, as shown in Figure 7(a). Figure

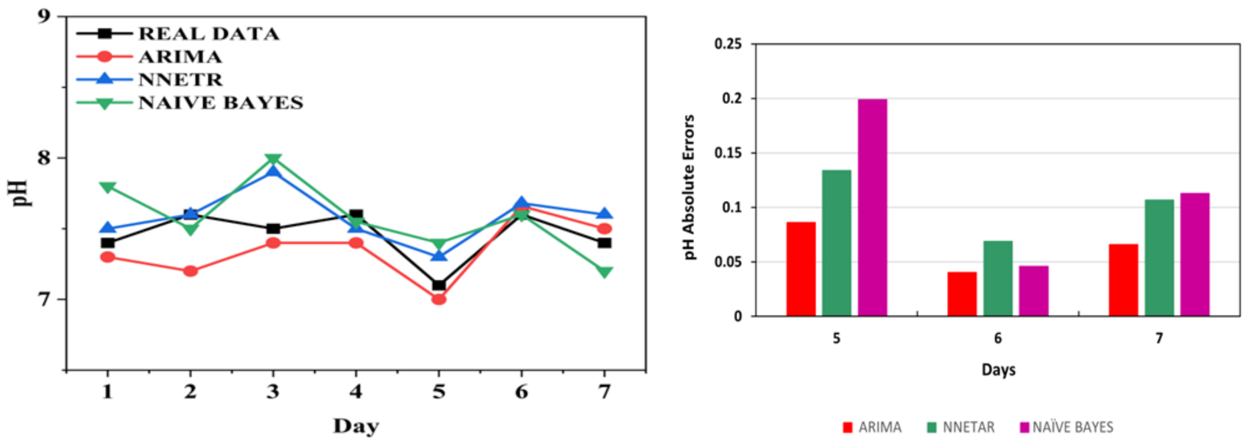
7(b) shows the error of all models except for Naïve Bayes for DO is less than 50%. On day five, NNETAR was slightly higher than ARIMA and Naïve Bayes with 29% error. On the sixth day, ARIMA and NNETAR have almost similar errors ranging from 20% to 30%, while Naïve Bayes has the highest for three days, which is 51%. On the seventh day, ARIMA prediction has a low error percentage below 20% which the lowest among others.



(a) (b)

Fig. 7. DO parameter (a) Actual Vs. predicted value (b) absolute error

Meanwhile Figure 8(a) depicted the pH parameter quite revealing in several ways. First, the real-time data changed across time, and the ARIMA model predicted a value almost identical to the actual value. Also, the pH data predicted by NNETAR and Naïve Bayes had quite a difference in predicted value at day 3, which differs from the ARIMA model. The correlation of real-time and prediction data can be seen in Figure 8(b). The graph shows that the Naïve Bayes model has the highest absolute error compared to NNETAR and ARIMA on the fifth day. The differences in minimum and maximum absolute errors between predictive models show that Naïve Bayes had the highest for all three days. From this graph, it can be concluded that the ARIMA model fits well with the pH data.



(a) (b)

Fig. 8. pH parameter (a) Actual Vs. predicted value (b) absolute error

Electrical conductivity (EC) and Total Dissolved Solid (TDS) to measure mineral levels in water. It is observed the data trend is fluctuated over the days as demonstrated by Figure 9(a) and (c). NNETAR and Naive Bayes readings differed compared to real-data. Figure 9(b) analysed Naïve Bayes model has contributed to the highest absolute error of EC between real-time and predicted while ARIMA shows the lowest absolute error. Meanwhile TDS dataset indicate the good fit of all three models which portrayed ununiform patterns. It is apparent from Figure 9(d) that ARIMA has better performance compared to NNETAR and Naive Bayes.

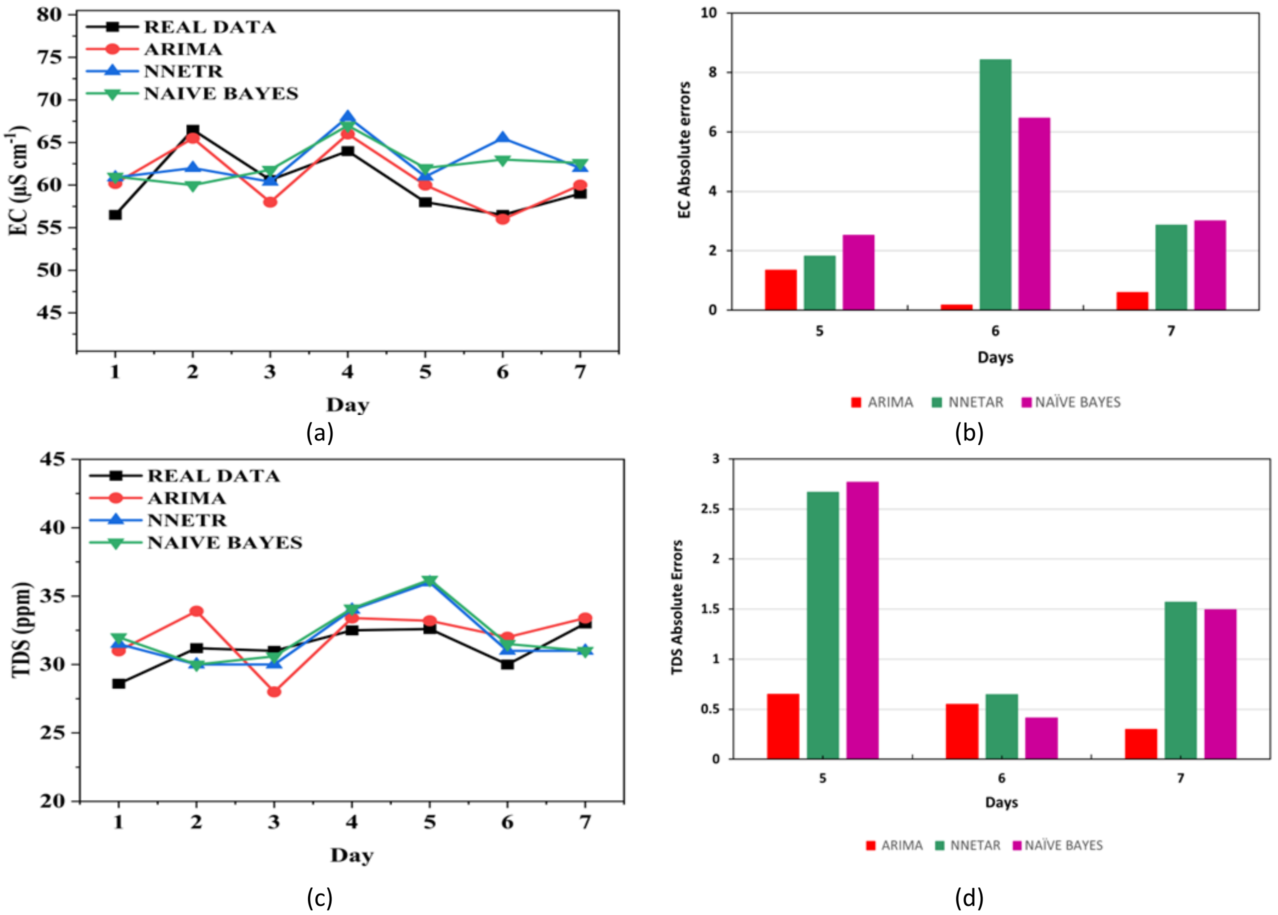


Fig. 9. (a) EC Actual Vs. predicted value (b) EC absolute error (c) TDS Actual Vs. predicted value (d) TDS absolute error

Table 3 presents the absolute error values depicting the disparity between actual and forecasted outcomes across various parameters – DO, pH, EC, and TDS – using three distinct prediction methods: ARIMA, NNETAR, and Naïve Bayes. These results provide insight into the performance of the three prediction methods across the considered parameters, offering a quantitative assessment of their accuracy and effectiveness in forecasting the studied values.

Table 3
The absolute error between actual and predicted values

Parameter	ARIMA			NNETAR			Naïve Bayes		
	Day 5	Day 6	Day 7	Day 5	Day 6	Day 7	Day 5	Day 6	Day 7
DO	0.09	0.22	0.17	0.29	0.22	0.18	0.15	0.51	0.39
pH	0.08	0.04	0.06	0.13	0.06	0.11	0.2	0.04	0.11
EC	1.7	0.5	0.9	1.9	8.2	2.5	2.2	6.2	2.6
TDS	0.6	2.6	2.7	0.6	0.7	0.4	0.3	1.6	1.5

3.3 Comparative Analysis and Limitations of Existing Approaches

Our study presents a comprehensive comparative analysis of water quality prediction models (ARIMA, NNETAR, and Naïve Bayes) applied to freshwater lobster farming. To contextualize our findings, we acknowledge the limitations of existing approaches, as summarized in Table 3.

i. Dissolved Oxygen (DO): Existing methods, including Naïve Bayes, tend to exhibit errors above 50%, especially on Day 5, indicating challenges in precise DO prediction during certain periods. While ARIMA and NNETAR demonstrate better performance, their errors are not negligible, suggesting a scope for refinement.
ii. pH Levels: The ARIMA model consistently outperforms others in predicting pH levels, showcasing lower absolute errors on all days. Naïve Bayes, on the other hand, exhibits higher errors, emphasizing its limitations in capturing the dynamic nature of pH changes.
iii. Electrical Conductivity (EC): ARIMA consistently presents lower absolute errors across all three days, highlighting its effectiveness in predicting mineral levels. NNETAR and Naïve Bayes, however, demonstrate substantial errors, particularly on Day 6, indicating potential challenges in accurately forecasting EC.
iv. Total Dissolved Solids (TDS): ARIMA consistently outperforms NNETAR and Naïve Bayes, with lower absolute errors across all three days. Notably, NNETAR and Naïve Bayes show notable discrepancies in predicted values on Day 6, suggesting potential limitations in capturing TDS variations.
v. General Limitations**:** While our proposed models enhance predictive accuracy, inherent challenges persist. For instance, the dynamic nature of streaming data may pose difficulties for certain models, as observed in the case of Naïve Bayes for DO and EC. Moreover, the need for meticulous calibration and adaptation to specific datasets is crucial, as illustrated by the varying errors across parameters and days.

4. Conclusions

The paper collected data on DO, pH, EC, and TDS for 6 months, which drives the prediction accuracy capability of the proposed system. Moreover, the proposed system cannot only predict, but its capability to send the data from My Structured Query Language (MySQL) to a query table in prompt has led it to predict in stream for the next seven days without any involvement of third-party software. The proposed system presents the method for predicting water quality in freshwater lobster aquaculture based on the ARIMA, NNETAR, and Naïve Bayes models in R Studio. The comparison of these three models shows the best-fit model based on the absolute error calculation for seven days. Overall, ARIMA fits well with four parameters: DO, pH, EC, and TDS. The average ARIMA absolute error for three consecutive days for DO, pH, EC, and TDS are 0.163, 0.064, 0.705, and 0.498. Future work concerns taking control actions such as (pH+, pH-, air pump, and water pump) in order to maintain the water parameters based on the prediction. In addition, utilize the prediction data for deeper analysis and trying different prediction methods.

Acknowledgement
The authors express their gratitude to the Malaysian Ministry of Higher Education for generously funding this research through the MTUN Matching Grant under Grant No. RDU212802 and UIC211503 and additional financial support from Universiti Malaysia Pahang Al-Sultan Abdullah internal grant UIC191205.

References

[1] Troell, Max, Barry Costa-Pierce, Selina Stead, Richard S. Cottrell, Cecile Brugere, Anna K. Farmery, David C. Little et al. "Perspectives on aquaculture's contribution to the Sustainable Development Goals for improved human and planetary health." *Journal of the World Aquaculture Society* 54, no. 2 (2023): 251-342. https://doi.org/10.1111/jwas.12946

[2] Marselina, S., Fakhrurroja, H., & Sari, B. (2023, January). Designing an IoT-Based Freshwater Lobster Cultivation Monitoring Dashboard. In *Proceedings of the 1st International Conference on Sustainable Engineering Development and Technological Innovation, ICSEDTI 2022, 11-13 October 2022, Tanjungpinang, Indonesia*. https://doi.org/10.4108/eai.11-10-2022.2326420

[3] TENGKU ZAINAL ABIDIN, T.Z.H., MAT ALI, N.N., ABU BAKAR, F.I., AHMAD BAKRI, F.A., ABU BAKAR, M.F., MALIK, N.H. and ZAINAL ABIDIN, M.U.N.I.R.A., 2023. Nutritional Composition and Antioxidant Activity of Freshwater Lobster in Malaysia: A Short Review. *Asian Journal of Chemistry*, *35*(2), pp.301-304. https://doi.org/10.14233/ajchem.2023.24050

[4] Bakhit, Abdelmoneim A., Mohd Faizal Jamlos, Nura A. Alhaj, and Rizalman Mamat. "Biofloc Farming with IoT and Machine Learning Predictive Water Quality System." In *2022 IEEE International RF and Microwave Conference (RFM)*, pp. 1-4. IEEE, 2022. https://doi.org/10.1109/RFM56185.2022.10065258

[5] Rosmawati, Mulyana, and Muhammad Azmi Rafi. "Pertumbuhan dan kelangsungan hidup benih Lobster air tawar (Cherax quadricarinatus) yang diberi pakan buatan berbahan baku tepung Keong mas (Pomacea sp.)." *Jurnal Mina Sains* 5, no. 1 (2019): 31-41. https://doi.org/10.30997/jms.v5i1.1771

[6] Photsathian, Thongchai, Thitiporn Suttikul, and Worapong Tangsrirat. "Design and improvement of wireless crayfish breeding system by controlling water temperature and monitoring pH via cloud system services." In *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, pp. 205-208. IEEE, 2018. https://doi.org/10.1109/ICCRE.2018.8376465

[7] Sharma, Arti, and Isha Devi. "A sustainable biosorption technique for treatment of industrial wastewater using snail shell dust (Bellamya bengalensis)." Environmental Monitoring and Assessment 195, no. 3 (2023): 389. https://doi.org/10.1007/s10661-023-10968-6

[8] Mursidi, Fathin-Amirah, M. Y. Ina-Salwany, Aslah Mohamad, Nurhidayu Al-saari, Aslizah Mohd-Aris, M. N. A. Amal, M. Zamri-Saad, Hisae Kasai, Sayaka Mino, and Tomoo Sawabe. "Vibriosis in Fish: A Review on Disease Development and Prevention." *Journal of Aquatic Animal Health* 31, no. 1 (2019). https://doi.org/10.1002/aah.10045

[9] Jin, Shiyu, Lisa Jacquin, Feng Huang, Mantang Xiong, Ruojing Li, Sovan Lek, Wei Li, Jiashou Liu, and Tanglin Zhang. "Optimizing reproductive performance and embryonic development of red swamp crayfish Procambarus clarkii by manipulating water temperature." *Aquaculture* 510 (2019): 32-42. https://doi.org/10.1016/j.aquaculture.2019.04.066

[10] Abdurrasyid, Abdurrasyid, Indrianto Indrianto, Meilia Nur Indah Susanti, and Yudhi S. Purwanto. "Detection of water quality in crayfish ponds with IoT." *Bulletin of Electrical Engineering and Informatics* 10, no. 2 (2021): 886-897. https://doi.org/10.11591/eei.v10i2.1968

[11] ZXiao, Zhong, Lingxi Peng, Yi Chen, Haohuai Liu, Jiaqing Wang, and Yangang Nie. "The dissolved oxygen prediction method based on neural network." *Complexity* 2017 (2017). https://doi.org/10.1155/2017/4967870

[12] Obado, Steve Agono. "IoT Based realtime fish pond water quality monitoring model." PhD diss., Strathmore University, 2019.

[13] Ma, Zhen, Xiefa Song, Rong Wan, Lei Gao, and Dawei Jiang. "Artificial neural network modeling of the water quality in intensive Litopenaeus vannamei shrimp tanks." *Aquaculture* 433 (2014): 307-312. https://doi.org/10.1016/j.aquaculture.2014.06.029

[14] Cao, Xinkai, Yiran Liu, Jianping Wang, Chunhong Liu, and Qingling Duan. "Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network." *Aquacultural Engineering* 91 (2020): 102122. https://doi.org/10.1016/j.aquaeng.2020.102122

[15] Yamak, Peter T., Li Yujian, and Pius K. Gadosey. "A comparison between arima, lstm, and gru for time series forecasting." In *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence*, pp. 49-55. 2019. https://doi.org/10.1109/ICMLA.2018.00227

[16] P. T. Yamak, L. Yujian, and P. K. Gadosey, “A comparison between ARIMA, LSTM, and GRU for time series forecasting,” 2019, doi: 10.1145/3377713.3377722. https://doi.org/10.1145/3377713.3377722

[17] Wang, K. W., C. Deng, J. P. Li, Y. Y. Zhang, X. Y. Li, and M. C. Wu. "Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network." *Epidemiology & Infection* 145, no. 6 (2017): 1118-1129. https://doi.org/10.1017/S0950268816003216

[18] Tudor, Cristiana, and Robert Sova. "EU net-zero policy achievement assessment in selected members through

automated forecasting algorithms." *ISPRS International Journal of Geo-Information* 11, no. 4 (2022): 232. https://doi.org/10.3390/ijgi11040232

[19] Tran-Quang, Vinh, and Anh Ha-Ngoc. "Aquaculture environment prediction based on improved lstm deep learning model." In *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, pp. 486-491. IEEE, 2021. https://doi.org/10.1109/NICS54270.2021.9701532

[20] Li, Xiaomei, Huan Liu, Rongli Gai, and Zumin Wang. "Prediction of water quality factors for sea cucumber farming based on Dual-Assisted prediction model." In *2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, pp. 306-313. IEEE, 2022. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics55523.2022.00076

[21] Susanti, Novita Dwi, Diang Sagita, Ignatius Fajar Apriyanto, Cahya Edi Wahyu Anggara, Doddy Andy Darmajana, and Ari Rahayuningtyas. "Design and implementation of water quality monitoring system (Temperature, pH, TDS) in aquaculture using IoT at low cost." In *6th International Conference of Food, Agriculture, and Natural Resource (IC-FANRES 2021)*, pp. 7-11. Atlantis Press, 2022. https://doi.org/10.2991/absr.k.220101.002

[22] Duangwongsa, Jomsuda, Thepparath Ungsethaphand, Pradchayaporn Akaboot, Supakit Khamjai, and Sayan Unankard. "Real-time water quality monitoring and notification system for aquaculture." In *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, pp. 9-13. IEEE, 2021. https://doi.org/10.1109/ECTIDAMTNCON51128.2021.9425744

[23] Nucita, A. A., Francesco De Paolis, and Gabriele Ingrosso. "X-rays from eclipsing cataclysmic variable systems: the eclipse profile." In *Journal of Physics: Conference Series*, vol. 354, no. 1, p. 012013. IOP Publishing, 2012. https://doi.org/10.1088/1742-6596/1517/1/012105

[24] Wu, Jiang, Jianjun Zhang, Wenwu Tan, Hao Lan, Sirao Zhang, Ke Xiao, Li Wang, Haijun Lin, Guang Sun, and Peng Guo. "Application of time serial model in water quality predicting." *Comput. Mater. Continua* 74, no. 1 (2023): 67-82. https://doi.org/10.32604/cmc.2023.030703

[25] Alahi, Md Eshrat E., Subhas Chandra Mukhopadhyay, and Lucy Burkitt. "Imprinted polymer coated impedimetric nitrate sensor for real-time water quality monitoring." *Sensors and Actuators B: Chemical* 259 (2018): 753-761. https://doi.org/10.1016/j.snb.2017.12.104