MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 18, No. 1(2026), pp 9-18 DOI : 10.18860/mat.v18i1.37617
Received : November 19th 2025; **Accepted** : March 14th 2026 ; **Avalaible Online** : March 28th 2026

A Robust Framework for Dissolved Oxygen Forecasting in Precision Aquaculture: A LightGBM Approach with Advanced Feature Engineering

Nyoman Wira Prasetya, Richard Wijaya Harianto

***Abstract*— Accurate prediction of necessary water quality parameters such as Dissolved Oxygen (DO) is very critical in precision aquaculture and is essential for performance-based decision-making. This thesis fills the gap between reactive monitoring and predictive intelligence through the construction of a solid machine learning infrastructure. We convert high frequency multivariate time series data into a supervised learning problem by an advanced feature engineering process that generates temporal predictions including lag features and rolling window statistics. A Light Gradient Boosting machine (LightGBM) algorithm trained using the above-mentioned engineered dataset has an extreme predictive power and achieves a Mean Absolute Percentage Error (MAPE) of just 6.25% on the unknown data. Results of single-variable interpretation analysis showed that short term data, especially the 5-minute rolling statistics of DO and turbidity variability, are the main driving factors for the model prediction. This research confirms that a feature-engineered LightGBM approach is a computationally efficient, but highly accurate approach to supporting the development of early warning systems in modern aquaculture as a computationally scalable approach.**

***Index Terms*—Aquaculture, Dissolved Oxygen (DO), Feature Engineering, LightGBM, Time Series Forecasting.**

I. INTRODUCTION

The aquaculture industry plays an increasingly important role in helping to guarantee global food security, in meeting the demands of the growing world population for protein during the stagnation of the capture fisheries [1]. According to the Food and Agriculture Organization (FAO), in 2022, for the first time, aquaculture production exceeded the production of traditional fisheries, pointing to its importance in the future of aquatic food systems. To meet this growing demand in a sustainable way, the industry is experiencing a major transformation towards more intensive, controlled, and data-based production systems, a paradigm called Precision Aquaculture [3], [4]. This approach combines modern technologies like the Internet of Things (IoT), automation, and Artificial Intelligence (AI), to optimise production processes and improve resource efficiency and environmental sustainability [5]. A cornerstone of precision aquaculture is the careful management of water quality, as the stability of critical environmental parameters is one of the primary determinants of success and has a direct impact on the health, growth rate, and survival of aquatic organisms [6]. Among these, Dissolved Oxygen (DO) is well known to be one of the most important and most time-varying variables [7]. Adequate concentrations of DO are crucial for respiration and metabolic processes, and a sharp drop in DO concentrations below critical levels, a situation referred to as hypoxia, may result in severe stress, poor appetite, disease susceptibility, and, in acute cases, mass mortality events, causing catastrophic economic losses to farmers [8], [9].

With the emergence of IoT devices, the

Manuscript received October 17, 2025. This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University.
NW Prasetya Author is with the Computer Science Department, School of Computer Science, Bina Nusantara University, Indonesia, email nyoman.wira@binus.ac.id
RW Harianto Author is with the Computer Science Department, School of Computer Science, Bina Nusantara University, Indonesia, email richard.harianto@binus.ac.id

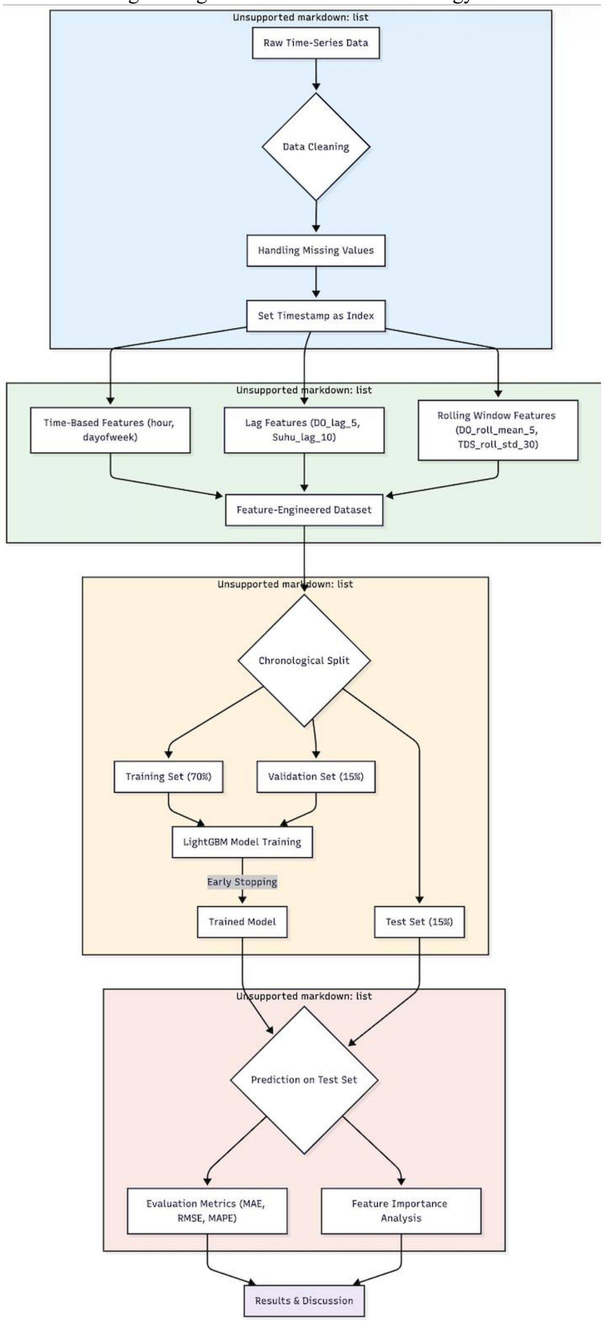
monitoring of water quality in aquaculture has seen a revolution as it has allowed the acquisition of parameters like temperature, turbidity, Total Dissolved Solids (TDS), and DO among others, through continuous monitoring and remotely [10], [11]. The term paper by Syauqy et al. and Prasetya et al. is top-notch in the current state of the art and provides good IoT architectures that were successfully able to capture water quality data and present it [12]. These systems have successfully overcome the drawbacks of the traditional manual monitoring, which is sporadic, labor-intensive, and inaccurate [13]. However, the development in terms of data acquisition has brought about the next challenge: data interpretation. Most available IoT systems are descriptive in nature; they simply display real-time or historical information on a dashboard and accurately answer the question, "what is happening now?" [14]. This leaves the farm manager in a constant reactive management mode, where corrective measures are usually taken after some critical conditions have been noticed. This latency, even of a few hours, can be harmful to the cultured stock. A major gap in the research literature, therefore, relates to the paradigm shift from the reactive to the proactive paradigm in management. The need to develop intelligent systems that go beyond just data logging to predictive understanding of information has become an urgent need to answer the critical question - "What will happen next?" [15]. The real-time DO forecasts made a few hours ahead of time would allow farmers to be proactive, to reduce the risks before extreme conditions occur, and to use their resources like energy for the aeration at the most appropriate times.

To remedy this major loophole, this study proposes and validates a strong methodology for short-term prediction of Dissolved Oxygen in aquaculture ponds using a machine learning technique. Specifically, it bases its research on the Light Gradient Boosted Machine (LightGBM), a highly scalable and efficient LightGBM algorithm implementation of gradient boosting decision trees, which recently has shown high performance on tabular data [16]. While tree-based models are inherently non-sequential in nature, the main body of our methodology is due to a sophisticated feature engineering process. We work on a tabular representation of the raw, multivariate time-series data, by systematically generating features which explicitly capture the temporal dynamics of the system (historical values - lag features, short-term trend and volatility - rolling window features, periodicity - time-based features). The main contributions of this research are fourfold: (1) A computationally efficient and highly accurate DO forecasting model is designed and validated using LightGBM, which shows that the DO forecasting model is suitable to effectively be applied in practical applications for precision aquaculture [17]. (2) It demonstrates systematically an effective feature engineering pipeline that converts a complicated time-series forecasting problem to a supervised regression task [18], [19]. (3) It performs a rigorous quantitative assessment of the model performance, with the low Mean Absolute Percentage Error (MAPE) of 6.25% on unseen test data, which shows the strong prediction ability. (4) It can provide useful information about the dynamics of the aquaculture environment from the fitness of the feature importances of the model, so it marks the main historical and environmental stimuli that will have a strong impact on future DO levels. This framework can be not only used as an operational data-driven solution for replacing a reactive approach to pond management with a more proactive one, but can also serve as a baseline tool for the implementation of smart decision support systems for the aquaculture industry.

II. DESIGN AND METHOD

This paper uses a systematic approach in building and validating a predictive model for dissolved oxygen (DO) forecasting. The workflow, shown in Figure 1, includes some important steps: data acquisition and preprocessing, an elaborate feature engineering, model selection and training, and a careful evaluation of the performance of the model. This paper describes in detail each step in order to ensure the transparency and reproducibility of the study.

Fig 1. Diagram for Research Methodology



2.1 Dataset Description

The data set used for this study was taken from a continuous water quality monitoring system that was deployed in an aquaculture pond. The collection data was ready at around one-minute intervals from 23 September 2021 to 29 September 2021 as a high-frequency multivariate time series data set. Each record in the data set includes a main Timestamp reading and four accompanying sensor readings: Temperature (Suhu) is the reading of the water temperature in degrees Celsius (°C), Turbidity is a measure of the cloudiness of the water, which is indicative of the presence of suspended particles in the water, Total Dissolved Solids (TDS) is the concentration of solid particles in the water that have dissolved with the water, and Dissolved Oxygen (DO) is the concentration of dissolved oxygen which is used as the target variable in our forecasting model. This high-resolution data set serves as a rich basis for recording the complex and dynamic interactions of environmental parameters in an aquaculture setting.

2.2 Data Preprocessing

In order to assure the quality and integrity of the data before the modeling process, a series of data preprocessing were carefully performed as they are instrumental in the performance of any machine learning model [20]. First of all, a temporal indexing process was performed in which the Timestamp column was transformed into a datetime format and made an index in the DataFrame. This is a basic step for time series analysis, which allows for chronological sorting, as well as time-based manipulations. Following this, a strategy for dealing with missing values was considered. Although the first data set contained no missing data, a proactive method has been performed to allow for possible gaps in the real world when dealing with sensor data (e.g. due to transmission errors, sensor malfunction, etc.) [21]. We used the method of time-based linear interpolation (interpolate(method='time')), which is a powerful method to process sequential sensor data that interpolates missing values by taking the straight line between the previous and following data values [22].

2.3 Feature Engineering

The main idea of our methodology is to cast the time series forecasting problem into a supervised learning regression problem. This is done through an extensive feature engineering process, which is a necessity for the tree-based models like LightGBM to identify and learn from temporal patterns [23]. Concerning the four initial sensor variables, 53 features were constructed in a systematic manner. These characteristics can be divided into three different categories as follows. First, to make the cyclical pattern such as diurnal (daily), weekly seasonality, etc., to be captured in the model, a number of time-based variables were derived directly from the timestamp index, including the hour of the day, day of the week, month, and day of the year [24]. Second, lag features were developed to put the model directly in historical context and also make it possible to learn autoregressive patterns [25]. These characteristics are the values of the four sensor variables (Suhu, Turbidity, TDS, DO) of 1, 5, 10, 30, and 60 previous time steps. Third, rolling window characteristics were computed in order to capture the recent behavior of each variable as a summary characteristic of local trend and volatility for that variable [26]. For each sensor variable, the rolling mean was calculated to detect the local trend, and the rolling standard deviation was calculated to assess the recent volatility for window sizes of 5, 10, and 30 time steps. After creating these attributes, the first rows of the dataset, which had NaN values, were dropped in order to make sure that the model was trained on full records.

2.4 Model Design: Light Gradient Boosting Machine (LightGBM)

For the forecasting task, we chose Light Gradient Boosting Machine (LightGBM) model. LightGBM is a state-of-the-art implementation of the popular Gradient Boosting Decision Tree (GBDT) algorithm, which is known for its high performance in terms of accuracy and efficiency, especially on large-scale tabular data [27]. Unlike the level-wise tree growth strategy of traditional GBDT implementations, LightGBM uses a leaf-wise growing strategy, which makes it converge faster and usually produces models of lower loss [28]. Its efficiency is improved further by two new techniques, namely GOSS (Gradient-based One-Side Sampling) and EFB (Exclusive Feature Bundling), which drastically reduce the computational overhead during training and do not decrease the accuracy [28]. Its implementation for a large number of features and its success in major forecasting competitions make it a perfect candidate for this study.

2.5 Experimental Setup

The experimental protocol was thoroughly developed to train the model and measure the generalization effect on unseen data. For time series data, random cross-validation is not suitable; it is necessary to use different splits to avoid data leakage where the model is trained on the future data in order to make predictions in the past, which means an optimistic assessment of the performance [29]. In order to avoid this, we used a strict chronological division and divided the dataset into three non-overlapping and consecutive sets: a Training Set (the first 70% of the data), a Validation Set (the following 15%), and a Test Set (the last 15%). The LightGBM model was trained on the training set based on regression_l1 as its loss function (Mean Absolute Error). The aim was to prevent overfitting and to acquire the optimal number of boosting rounds, for which an early stopping mechanism was adopted [30]. The performance of the model was evaluated on the validation set at each iteration, and if the performance of the validation error plateaued for 50 successive iterations, the training was automatically terminated. This way, the model can generalize well for the new data and is not affected by learning the noise of the training set [31].

2.6 Evaluation Metrics

To provide a comprehensive and multi-faceted assessment of the model's predictive accuracy on the test set, three standard regression metrics were employed [32]. The first, Mean Absolute Error (MAE), measures the average absolute difference between the predicted ($\ddot{y}_i$) and actual (y_i) values and is easily interpretable as it is in the same unit as the target variable. The second, Root Mean Squared Error (RMSE), represents the square root of the average of the squared differences and penalizes larger errors more heavily than MAE. The third, Mean Absolute Percentage Error (MAPE), expresses the average absolute error as a percentage of the actual values, providing a relative measure of error that is independent of the scale of the data. The formulas for these metrics are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \ddot{y}_i\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \ddot{y}_i)^2}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \ddot{y}_i}{y_i}\right|$$

In these equations, n is the number of samples in the test set, y_i is the actual DO value, and $\ddot{y}_i$ is the DO value predicted by the model.

III. Result and Discussion

This chapter presents and discusses in detail the empirical research findings and their scientific and practical implications. The section is divided into two sections: First, we will explain the results of the feature engineering after which a quantitative analysis of the result of the LightGBM model's predicting power follows. It is then followed by an interpretative analysis of the feature importances of the model, which is used to explain the main drivers of DO dynamics. Finally, a general discussion is presented for the positioning of the results and the limitations of the study in the context of precision aquaculture and for the identification of future research directions.

3.1 Feature Engineering Transformation

The first basic step in our analytical framework was the transformation of raw time series data into a high dimensional feature space that can be used for the LightGBM model. This process unfolded from the four sensor variables of the initial set into a comprehensive set of 53 predictors. To create more transparency and highlight the result of this critical transformation, Table I shows a sample of the feature-engineered dataset. This table highlights the original target variable (DO) and some example engineered time-based (hour), lag (DOlag5), and rolling window (DOrollmean5, Turbidityrollstd5) features that comprise the input for the predictive model.

Table 1. Example of the feature-engineered dataset

Timestamp	DO	hour	DO_lag_5	DO_roll_mean_5	Turbidity_roll_std_5
2021-09-23 19:06:16	0.3906	19	0.3418	0.35986	0.004322
2021-09-23 19:07:21	0.3760	19	0.3467	0.36279	0.004322
2021-09-23 19:08:28	0.3711	19	0.3467	0.36621	0.004776
2021-09-23 19:09:33	0.3564	19	0.3418	0.36816	0.004322
2021-09-23 19:10:39	0.3564	19	0.3613	0.37011	0.003832

3.2 Model Performance Evaluation

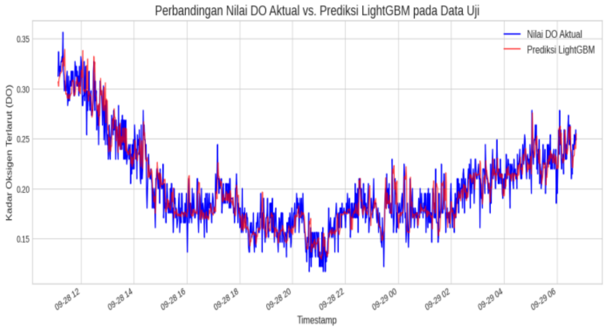
The LightGBM model was trained and later tested on the hidden test set following the approach outlined in Chapter 2. Three standard regression measures were used in order to estimate the predictive accuracy of the model and the findings of the analysis are summarized in Table II. Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) of the model were 0.0123 and 0.0163 respectively. More importantly, the Mean Absolute Percentage Error (MAPE) was found to be very low at only 6.25. This is an indication of high degree of predictive fidelity meaning that the model forecasts on average are quite far away in the real values of DO, which is quite acceptable to operational use in the aquaculture sector.

Table 2. Model Performance on the Test Set

Metric	Value
Mean Absolute Error (MAE)	0.0123
Root Mean Squared Error (RMSE)	0.0163
Mean Absolute Percentage Error (MAPE)	6.25%

To make a qualitative evaluation of the model forecasting performance, the model predicted values of DO were compared to ground-truth values in the test set as shown in Figure 2. The visualization clearly shows that the model is skillful in capturing the non-linear dynamics of the DO time series which are complex. The calculated trend (red line) shows a very high level of correspondence to the real data (blue line), capturing not only the strong diurnal variations, but also the more subtle and high-frequency ones. Such good visual correlation confirms the quantitative measures in Table II and confirms the strength of the model in practical forecasting functions in the real world.

Fig 2. Comparison of Actual DO Values vs. LightGBM Predictions on the Test Set



3.3 Feature Importantance Analysis

One of the main benefits of using tree-based ensemble models like LightGBM is that they have an inherent interpretability characteristic (via feature importance analysis). The analysis will enable the deconstruction of the black box, where the most important predictors are the ones that make the most significant contribution to the forecasting decision of the model. Figure 3 shows the 20 most significant predictors according to the trained LightGBM model in terms of their cumulative contribution (F-score) to the predictive power of the model.

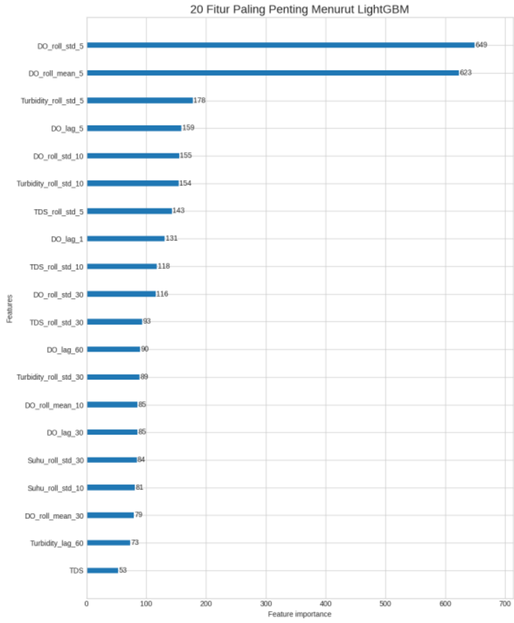


Figure 3. The 20 Most Influential Predictors as Determined by the LightGBM Model.

As the analysis shows, the predictive ability of the model is highly compelling and logical: the latest historical data dominates in the predictive ability. It has five most influential features, which are DO_roll_std_5 (5-minute rolling standard deviation of DO), DO_roll_mean_5 (5-minute rolling mean of DO), Turbidity_roll_std_5 (5-minute rolling standard deviation of Turbidity), DO_lag_5 (DO value in 5 minutes ago), and DO_roll_std_10 (10-minute rolling standard deviation of DO).

This hierarchy highlights the fact that the most potent predictors of the near-term state of DO are the immediate trend, volatility and lagged states of the market. The dominance of DO_roll_std_5 indicates that the anionic acute instability or high-speed change in the level of DO is a key antecedent that the model has been programmed to measure with a lot of weight. Moreover, Turbidity_roll_std_5 being ranked in the third position as a feature of maximum significance indicates a strong and dynamic relationship between the clarity and dissolved oxygen in water which may fluctuate quickly and is probably facilitated by biological processes like algal photosynthetic processes and respiration.

3.4 Discussion

The empirical findings from this paper are very interesting and have important implications for precision aquaculture development. The superiority of the feature-engineered LightGBM model with a

MAPE of 6.25% verifies the hypothesis that this methodological approach can successfully predict a key and highly volatile water quality parameter. This performance goes beyond academic legitimisation and is an empirical step towards the operation of strategies of proactive management in aquaculture. A forecasting tool with such a level of accuracy can serve as the analytical engine of an early-warning system which will be capable of informing farm managers of an imminent hypoxic event with sufficient time to carry out preventative measures, such as the strategic initiation of aeration systems.

The feature importance analysis provides us with deep insight into the biophysical dynamics of the pond ecosystem. The overwhelming presence of short-term temporal features (i.e. lag and rolling window statistics over 5 and 10-minute intervals) is conclusive evidence that the DO concentration is under the control of high-frequency dynamics in this system. This observation has a very important methodological implication: a high resolution of data acquisition, at minute-level intervals, is not only useful but necessary to build accurate predictive models. It is highly likely that models that are trained on sparser, hourly data would not be able to capture these pivotal short-term fluctuations and therefore would result in substantially degraded performance. Moreover, the high importance of the volatility of turbidity (Turbidityrollstd_5) implies that, for predictive purposes, monitoring the changing rate of turbidity (water clarity) is almost as informative as measuring the level of DO itself. This is presumably due to rapid changes in the quantity of phytoplankton, where blooms or die-off of large numbers of phytoplankton can directly and immediately affect the rate of oxygen production and consumption.

In machine learning terms, this is a successful study for proving that a time-series problem can be successfully converted into a supervised regression problem by applying careful feature engineering. With this paradigm, powerful, computationally efficient, and interpretable algorithms can be used within it, such as LightGBM, that otherwise would not be favored over computationally more expensive sequential architectures such as LSTMs or Transformers. For many actual deployment use cases, especially ones that involve field or edge computing devices with limited computational resources, this approach offers the best combination of predictability and performance.

Although the findings of this study are strong, there are some limitations that, in turn, serve as promising avenues for future research. Firstly, despite the high-frequency nature of the dataset, it only covers a very short period of time (one week). This is adequate for diurnal cycle modeling but excludes the ability to capture seasonal or operational length cycle patterns. Therefore, the model performance could be affected in varying climatic conditions. Secondly, the model was derived from data in a single aquaculture pond, and the generalizability of the models to other ponds with different hydrodynamics or stocking densities or management protocols was not investigated. Third, the current model is endogenous as it only depends on the relations among the measured sensor data. The incorporation of exogenous data such as meteorological data (e.g. solar irradiance, barometric pressure) or operational data (e.g. feeding time, aeration activating points) could further advance predictive accuracy.

Future research should therefore focus especially on collecting longitudinal data over longer periods of time to produce more robust, seasonally sensitive models. Of critical future importance will be to apply the proposed framework to a diverse portfolio of aquaculture sites to rigorously test the transferability. Finally, a good contribution to the field would be to compare this feature-engineered LightGBM approach to state-of-the-art deep learning architectures, in a quantitative way that describes the trade-offs between model complexity, computational cost, and prediction accuracy within the domain of this particular application.

IV. Conclusion and Suggestions

4.1 Conclusion

This research was initiated with the key objective to drive the shift from reactive and proactive management of precision aquaculture. While the existing IoT does a commendable job of collecting real-time data it's limited by the amount of information available to them which leaves the farm manager to decipher the swirl of information displayed and respond to it after the fact. One of the most important water quality parameters, Dissolved Oxygen (DO), was successfully predicted in this study to show that a robust machine learning framework has been established and validated to give accurate short-term predictions.

By using an advanced approach utilizing feature engineering to transform raw time-series into a feature-rich tabular representation, we proved that the Light Gradient Boosting Machine (LightGBM) algorithm is quite good and computationally light as a forecasting algorithm. The proposed methodology is empirically validated with very convincing results. The model was able to achieve a very high level of predictive accuracy on the unseen test set, with a Mean Absolute Percentage Error (MAPE) of 6.25%. The robust visual correlation between simulated and observed values further supported the contention that the model was capable of reflecting the various high-frequency and diurnal behavior of DO concentrations.

Furthermore, the feature importance analysis added a great diagnostic element of the model, which demonstrated that the model's predictions are highly driven by the short term historical data, in particular, the rolling mean and standard deviation of DO and Turbidity on 5-minute intervals. This result not only reveals the high frequency of the environment of the pond, but also emphasizes the indispensable value of the high resolution data collection for constructing well-founded predictive models.

In conclusion, this study has successfully shown that a feature engine LightGBM model is an excellent tool that offers a robust, explainable and efficient solution to DO forecasting in aquaculture. In addition to providing a validated proof of concept for a reliable early-warning system, the framework described here is practically feasible and is a step towards empowering the farmer for acting proactively based on predicted information.

4.2 Suggestions for Future Research

The study results are robust and provide a number of promising directions for further research that can help expand the utility and application of predictive modelling in aquaculture. Within the scope and limitations of this work, the following recommendations are advanced:

1. Data and Model Robustness: The present model was trained on one week of longitudinal data. The necessity of longitudinal data over several seasons and a wide range of operational conditions clearly needs to be an invaluable focus of future investigations. This would allow more robust models to be created which can accommodate longer seasonality and be trained to fit wider ranges of environmental conditions, thus providing higher reliability for all year round.
2. Incorporation of Exogenous Variables: The present model is endogenous; relying only on the sensor information provided within the pond. The combination with exogenous data such as meteorological information (e.g., solar irradiance, barometric pressure, wind speed) and operational data (e.g., feeding times, aeration on/off times) might offer further contextual information, and potentially provide further gains in forecasting skill.
3. Cross-Site Generalizability Study: This study was done on data of one aquaculture site. A crucial next step would be to confirm the proposed framework in different and diverse aquaculture systems (e.g. different species, stocking densities, pond geometries and geographical locations). Such a study would prove extremely useful in testing the generalizability of the model and determining the steps that may be required for transfer learning or site-specific calibration.
4. Comparative Benchmarking with Deep Learning Models: It would be a valuable contribution to the field if this feature

engineered LightGBM method had a comparative study comparing the performance/computational efficiency of this method v/s state of the art Deep Learning Architectures like pro returning Long Short Term Memory (LSTM) and Transformer. This would lead to a better understanding of the tradeoffs between model complexity, interpretability and predictive power in this domain of applications. Advancement to Integrated Decision Support System: A further potential endeavour for future development is taking the forecasting model and incorporating it into an integrated decision support system. The system would be able to make recommendations automatically for actions to be taken (e.g., "Activate Aerator 2 in 60 minutes") or possibly developed as a closed-loop control system that automatically adjusts aeration equipment based on predicted DO levels to further optimize resource inputs with minimal human intervention.

ACKNOWLEDGMENT

The authors would like to thank the School of Computer Science, Bina Nusantara University that have provided financial support for this research scheme.

REFERENCES

[1] Gatta, P. P. (2022). The State of World Fisheries and Aquaculture 2022. In FAO eBooks. https://doi.org/10.4060/cc0461en.

[2] Kumar, A., & Saini, M. (2023). A Comprehensive Review of Sustainable Aquaculture Practices. Advanced Research in Agriculture Science and Technology, 6(2), 1-10.

[3] Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J. A., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L. M., Schellewald, C., Skøien, K. R., Alver, M. O., & Berckmans, D. (2017). Precision fish farming: A new framework to improve production in aquaculture. Biosystems Engineering, 173, 176–193. https://doi.org/10.1016/j.biosystemseng.2017.10.014

[4] Narsale, S. A., Prakash, P., Mohale, H. P., Baraiya, R., Sheikh, S., Kirtikumar, P. B., Mansukhbhai, C. R., Kadam, R. V., & Tekam, I. (2024). Precision Aquaculture: A way forward for Sustainable agriculture. Journal of Experimental Agriculture International, 46(5), 83–97. https://doi.org/10.9734/jeai/2024/v46i52360.

[5] *O'Donncha, Fearghal & Grant, Jon. (2019). Precision Aquaculture. IEEE Internet of Things Magazine. 2. 26-30. 10.1109/IOTM.0001.1900033.*

[6] *Mohammed, Imran & Blesslene, Hepzibah & Faizullah, Mohamed & Nicy, Brita & devi, Vimala & Pandiyan, Praveenkumar. (2025). Revolution of AI in Aquaculture and Fish Processing: A Review https://ojs.pphouse.org/index.php/IJBSM Natural Resource Management International Journal of Bio-resource and Stress Management. International Journal of Bio-resource and Stress Management. 16. 1-14. 10.23910/1.2025.6194.*

[7] Isla, Mercedes. (2008). Water quality in recirculating aquaculture systems (ras) for arctic charr (salvelinus alpinus L.) culture..

[8] Ali, Bulbul & Anushka & Mishra, Abha. (2022). Effects of dissolved oxygen concentration on freshwater fish: A review. International Journal of Fisheries and Aquatic Studies. 10. 113-127. 10.22271/fish.2022.v10.i4b.2693.

[9] Ali, Bulbul & ., Anushka & Mishra, Abha. (2022). Effects of dissolved oxygen concentration on freshwater fish: A review. International Journal of Fisheries and Aquatic Studies. 10. 113-127. 10.22271/fish.2022.v10.i4b.2693.

[10] Syauqy, Dahnial & Hanggara, Buce & Purnomo, Welly & Putra, Widhy Hayuhardhika Nugraha & Prasetya, Nyoman. (2022). Automated Continuous IoT-based Monitoring System for Vaname Shrimp Cultivation Management. Computer Engineering and Applications Journal. 11. 89-100. 10.18495/comengapp.v11i2.402.

[11] Prasetya, Nyoman & Harahap, Arya & Aulady, Fadhli & Wulandari, Inayah. (2023). Design of Water Monitoring System in Aquaponics Based on Arduino Nano and Raspberry Pi. MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology). 15. 29-42. 10.18860/mat.v15i1.23005.

[12] Rodríguez, M. C. B., González, C. E. M., Almanza, E. D. C., Rodríguez, C. G., Regino-Vergara, J. Á., & López-Padilla, A. (2025). Benefits and challenges of the internet of things in aquaculture production: a literature review. Frontiers in Sustainable Food Systems, 9. https://doi.org/10.3389/fsufs.2025.1590153.

[13] Papolonias, Juffil & Lavilles, Rabby & Miano, Joel. (2025). Development of water quality monitoring system for fish farming. Bulletin of Electrical Engineering and Informatics. 14. 2962-2974. 10.11591/eei.v14i4.7673.

[14] Saha, Sajal & Rajib, Rakibul & Kabir, Sumaiya. (2018). IoT Based Automated Fish Farm Aquaculture Monitoring System. 201-206. 10.1109/ICISET.2018.8745543.

[15] Yang, H., Sun, M., & Liu, S. (2023). A hybrid intelligence model for predicting dissolved oxygen in aquaculture water. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1126556

[16] Zhou, S., Song, C., Zhang, J., Chang, W., Hou, W., & Yang, L. (2022). A Hybrid Prediction Framework for Water Quality with Integrated W-ARIMA-GRU and LightGBM Methods. Water, 14(9), 1322. https://doi.org/10.3390/w14091322.

[17] Hartanto, Anggit & Kholik, Yanuar & Pristyanto, Yoga. (2023). Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model. JOIV : International Journal on Informatics Visualization. 7. 2270. 10.30630/joiv.7.4.01740.

[18] Porto, B. M., & Fogliatto, F. S. (2024). Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. BMC Medical Informatics and Decision Making, 24(1). https://doi.org/10.1186/s12911-024-02788-6.

[19] Shukla, M. (2025). Interpreting Time Series Forecasts with LIME and SHAP:A Case Study on the Air Passengers Dataset. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-7358158/v1

[20] Lee, A. H. S., Shankararaman, V., & Ouh, E. L. (2023). Extending the Horizon by Empowering Government Customer Service Officers with ACQAR for Enhanced Citizen Service Delivery. 2021 IEEE International Conference on Big Data (Big Data), 1952–1958. https://doi.org/10.1109/bigdata59044.2023.10386189

[21] Petkovski, Aleksandar & Shehu, Visar. (2023). Anomaly Detection on Univariate Sensing Time Series Data for Smart Aquaculture Using Deep Learning. SEEU Review. 18. 1-16. 10.2478/seeur-2023-0030.

[22] Al-Kassab-Córdova, A., & Soto-Becerra, P. (2025). Causal Inference in Public Health: A Call to Stop Causal Fishing Expeditions. Public health reports (Washington, D.C. : 1974), 333549251342034. Advance online publication. https://doi.org/10.1177/00333549251342034

[23] Li, Mingjie & Li, Zeyan & Yin, Kanglin & Nie, Xiaohui & Zhang, Wenchi & Sui, Kaixin & Pei, Dan. (2022). Causal

Inference-Based Root Cause Analysis for Online Service Systems with Intervention Recognition. 10.1145/3534678.3539041.

[24] Villarroel, D. & Mayer, Miguel Angel & Leis, Angela & Karkaletsis, Vangelis & Stamatakis, Konstantinos & Metsis, Vangelis & Nasikas, P. & Labsky, Martin & Ruzicka, M. & Svátek, Vojtěch & López-Ostenero, F. & Peinado, V.. (2010). Assisting Quality Assessment (AQUA) – a system based on semantic web and information extraction technologies to support medical quality labelling.

[25] ProRCA: a causal python package for actionable root cause analysis in real-world business scenarios. (n.d.). https://arxiv.org/html/2503.01475v1

[26] Su, L., Zuo, X., Li, R., Wang, X., Zhao, H., & Huang, B. (2023, October 31). A Systematic review for transformer-based long-term series forecasting. arXiv.org. https://arxiv.org/abs/2310.20218

[27] Paialunga, P. (2025, July 15). Hands-On Attention Mechanism for Time Series Classification, with Python. Towards Data Science. https://towardsdatascience.com/hands-on-attention-mechanism-for-time-series-classification-with-python/

[28] Nuanmeesri, S., Poomhiran, L., Kadmateekarun, P., & Tarasawatpipat, C. (2023). Optimizing Predictive Modeling for Water Quality in Farms with Blended Artificial Neural Network. Journal of System and Management Sciences, 14(1). https://doi.org/10.33168/jsms.2024.0118

[29] Mohan, S., Kumar, B., & Nejadhashemi, A. P. (2025). Integration of Machine Learning and Remote Sensing for Water Quality Monitoring and Prediction: A Review. Sustainability, 17(3), 998. https://doi.org/10.3390/su17030998

[30] Zambrano, A. F., Giraldo, L. F., Quimbayo, J., Medina, B., & Castillo, E. (2021). Machine learning for manually-measured water quality prediction in fish farming. PLoS ONE, 16(8), e0256380. https://doi.org/10.1371/journal.pone.0256380

[31] Ma, Y., Fang, Q., Xia, S., & Zhou, Y. (2024). Prediction of the Dissolved Oxygen Content in Aquaculture Based on the CNN-GRU Hybrid Neural Network. Water, 16(24), 3547. https://doi.org/10.3390/w16243547

[32] Eze, E., & Ajmal, T. (2020). Dissolved Oxygen Forecasting in Aquaculture: A Hybrid Model Approach. Applied Sciences, 10(20), 7079. https://doi.org/10.3390/app10207079