Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

An Innovative Hybrid Approach to Forecasting Soluble Oxygen for Optimal Water Purification in Highly Concentrated Aquaculture

Mr. Shaik Mohammed Imran [1], D.Yashaswini[2],Ch.Sahasra[3], B.Gayathri[4]

[1] Assistant Professor,Department of CSE-AIML, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad

[2],[3],[4] Student, Department of CSE-AIML, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad

ABSTRACT:

An important measure of the water's quality in an aquaculture setting is the concentration of dissolved oxygen. Disintegrating oxygen content prediction using conventional techniques is slow and inaccurate due to the complexity, nonlinearity, and dynamics of the process. This research develops a hybrid model that addresses these problems by combining The radiation gradient enhancement Machine (LightGBM) with This relationship Simple Rechargeable Unit (Biru). The first step was to find the important parameters by using linear interpolation and smoothing. After removing superfluous variables, the LightGBM algorithm predicts dissolved oxygen in highly intensive aquaculture and establishes its relevance. Lastly, the attention approach was used to map the learning parameter matrices and weighting matrices, allowing various weights to be applied to the Biru's hidden states. The results shown that the given prediction model can capture the upward trend of oxygen dissolution fluctuations over a 10-day period with a rate of accuracy reaching 96.28% in only 122 seconds. It takes the least amount of time to compare the model impacts of Biru - Attention, LightGBM - GRU, LightGBM-LSTM, as well as LightGBM - Biru. The improved accuracy of its predictions makes it a valuable tool for controlling the water quality in intensive aquaculture.

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

INTRODUCTION

In terms of aquaculture production, China dominates the global stage; the nation is accountable for more than 70% of the global total. China produced 53.88 million tons of aquaculture products in 2021, or 80.50% of the world's total. Due to the fact that many physiological processes, including breathing, eating, excretion, and reproduction, take place in water, the quality of the water in an aquatic environment has a direct impact on the quantity and quality of aquatic commodities. Spot on! No such reference was found. Because aquatic organisms rely on dissolved oxygen (DO) for metabolic processes, its presence is indicative of high water quality. Disease outbreaks and mass death are possible outcomes of either too much or too little dissolved oxygen (DO), which may impact the proper development of farmed shrimp, fish, and other creatures, leading to substantial financial losses for businesses. Ensuring the growth of aquatic products in an optimal environment, anticipating trends in dissolved oxygen concentrations, and promptly regulating them are all crucial for preventing water quality degradation, reducing aquaculture risks, and promoting the sustainable and healthful growth of intensive aquaculture. Academics from all around the world have been studying dissolved oxygen forecasting systems that use machine learning, and their work has yielded a lot of discoveries. One example is Liu's forecasting model, which incorporates the particle swarm optimum gravity search approach, empirical wavelet transforms, and grey correlation degree. Through the use of variable-mode decomposition (VMD) to separate and denoise the raw data, followed by feeding the decomposed data through a deep belief networks (DBN) for prediction, Ren was able to create an improved model for analyzing the dissolved oxygen trend, as shown in the experiments. After applying PCA, or principal component analysis, to identify the most important factors influencing dissolved oxygen levels, Cao built a prediction model using K-means clustering and GRU. Shi offered an alternative model that relied on clustering and an improved extreme learning technique, while Huang put forward a hybrid model that combined CEEMDAN-LZC and GOBLPSO to increase accuracy. An innovative feature-extraction-based model was suggested, which enhanced predictive performance and yielded precise dissolved

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

oxygen level forecasts. Nong suggests a model for predicting dissolved oxygen levels that combines support vector regression with optimization and multi-feature engineering techniques; testing confirms that this model significantly improves prediction accuracy. Difficulty in meeting the demand for accurate and productive aquaculture production is caused by the fact that the oxygen dispersion prediction models mentioned earlier can only forecast oxygen levels at future moments and are not fast enough to capture global contextual information. With its highly parallelized design and two-layer reverse-stacked network topology, the Bidirectional Simple Rechargeable Unit (Biru) is able to gather both past and future knowledge. Biru is used extensively in several domains due to its capability of sequence modeling and its enhancement of the gradient disappearance issue. To obtain high accuracy forecasts for intrusion identification for industrial control systems, Jie employed a Biru-based technique, for instance. For reliable prediction of network intrusion, Ding employed an intrusion detection model that included CNN and Biru. Ding put forth a model for protecting networks from intruders that makes use of Biru and feature reduction to spot suspicious data. Machine learning has wide-ranging applications since it processes data utilizing attention mechanisms to identify the positive impact size between input and output data. For instance, Jiang suggested a model for indoor temperature predicting that combines LSTM, transformer, and attention mechanisms; this model efficiently and accurately predicts room temperature changes. An attentive network framework based on Alternator Encoder was successfully developed by Zhang by integrating the transformer model with numerous attention mechanisms. This framework allowed for reliable prediction of stock patterns. In order to produce accurate predictions of water quality, Mei introduced a hybrid model that combines convolutional neural networks (CNNs), gradient recurrent units (GRUs), and attention mechanisms. The attention layer may modify the weights of individual neurons. Through the integration of stack structure, multi-attention mechanism, and TCN, Li put forward a dissolved oxygen prediction model. This model has the potential to enhance the accuracy of water quality parameter predictions in marine pastures and contribute positively to the growth of marine fisheries. By using a hybrid attention model grounded on parallel deep learning, Duan was able to successfully forecast the state of tool wear. With an attention mechanism that

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

adaptively learns the proportion of weights, the model may focus on data that is more significant to the current task among the multiple inputs, thereby addressing the disadvantages of traditional techniques. The water column undergoes complicated biochemical processes in intense aquaculture, and the variables' interplay is intricate. Feeding all factors into the dissolved oxygen prediction model for water quality the parameters in intensive aquaculture suffers from an inaccurate and computationally complex prediction, leading to a complex prediction model network structure and overlapping or redundant information. In large sample applications, LightGBM's performance is dramatically improved, memory consumption is reduced, and data processing efficiency is maximized. LightGBM is a tree centric boosting method. To increase the accuracy of risk prediction, Wang developed the LightGBM model for business finance. To help people make informed decisions about investing in cryptocurrencies, Sun used LightGBM to create a model that can anticipate future price trends. For short-term the power of wind forecasting, Ren built an attention mechanism based on CNNLSTM-LightGBM. This model successfully predicts the future. In order to address the drawbacks of conventional methods for dissolved oxygen prediction, this work suggested a hybrid model that incorporates a the LightGBM algorithm (Light Gradient Boosting Machine), a Biru (Bidirectional Simple Recurrent Unit), and an Attention mechanism. The important factors influencing the concentration of dissolved oxygen in intensive aquaculture were identified using the LightGBM. In order to simplify the network design and provide dissolved oxygen predictions, a nonlinear combination model was suggested. This model makes use of an attention mechanism and a continuous simple loop unit. The experimental findings demonstrated that this model's forecasting capabilities may provide technical assistance in the precise regulation of ecological parameters in extensive aquaculture operations.

RELATED WORK

"Development of fisheries in China,"

Hu, Hangzhou, et al. 2021

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

The fishing industry is a subset of the food production sector that involves the commercial cultivation of fish and other aquatic plants and animals for their edible flesh. Upstream departments include things like making and supplying fishing gear, boats, machinery, and instruments; downstream departments include things like storing, processing, transporting, and selling aquatic products; and auxiliary departments include things like building fishing ports. Chinese fisheries are vital to the country's economy since they provide not only a vast quantity of human food but also a wealth of industrial raw materials. The paper provides a comprehensive overview of how to develop fisheries in China, covering topics such as the history of fisheries, the present state of the submerged seed industry, aquaculture, capture, aquatic product processing, aquatic food for animals and feed, and leisure fishery. It also assesses the value of Chinese fisheries to global fisheries, present problems and strategies for fisheries development in China, and more. For China's fisheries development, this article has significant guiding relevance.

"An intelligent framework for prediction and forecasting of dissolved oxygen level and bioflick amount in a shrimp culture system using machine learning techniques,"

Jasmin, S. Ayesha, Pradeep Ramesh, and Mohammad Tanveer,2022

In a shrimp culture system, this work uses state-of-the-art machine learning algorithms to predict and forecast bioflick quantity and dissolved oxygen (DO). In an aquatic shrimp culture system, the investigation was conducted using average DO as well as bioflick quantity as goal parameters. Twelve distinct data subsets were generated for the purpose of developing the model, with seventeen numbers of cultural and climatic characteristics taken into account and three distinct feature selection methods used. Random Forest, Ad Boost, and the deep neural network are three well-known machine learning techniques that were used to build the model. We gathered 36 unique models, ran eight model validation tests to see how accurate they were, and then analyzed the findings. The best predictive model out of 36 models was developed using the Random Forest approach for dissolved oxygen prediction using integrated cultural and meteorological characteristics. It had a R^2value of 0.709, a prediction accuracy of 98.26%, and a

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

score of 0.7381. In addition, a framework for predicting DO and bioflick quantity in a shrimp-based bioflick cultivation structure was established, and exploring data analysis was performed for prediction. When developing the prediction model, dissolved oxygen was determined to be more robust. A scientific understanding of the agricultural system and its prediction processes led to the development of the Intelligent Framework. Both experienced shrimp farmers and those just starting out in the industry may use the built framework to create their own models of prediction that operate best in their specific farming environments.

"Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique,"

Ahmed, Mohammad Hafez, and Lian-Shin Lin.2021

The presence of dissolved oxygen (DO) within moving water is difficult to describe employing process-centered water quality models because of the many processes that influence the amount present and the complicated connections between them. This work used the quantified regression forest (QRF) machine learning methodology to create data-driven models for forecasting dissolved oxygen (DO) levels in three rivers. The rivers drain watersheds in diverse locations with varied land use and land cover characteristics. To create and verify the models, we consulted water quality data from 2007 to 2019. After identifying key drivers of DO using the variable significance index, models were built with various combinations of these drivers as input variables. After utilizing 80% of the data to calibrate the models for each input scenario, the remaining 20% was used to verify the models by anticipating the DO concentrations. Across all stations, the best model performance was achieved by using water temperatures and pH values as input variables, together with specific conductance and COD, or chemical oxygen demand, as top predictors. When compared to models from the United States Environmental Protection Agency and the multilayered perceptron neural network (MLPNN), the created models performed better in terms of both explaining data variation and providing reduced prediction errors. It may be possible to construct parsimonious models with just a handful of predictors over in-stream DO forecasts, given that the top-ranked forecasters for the three rivers,

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

which are geographically distant from each other, are similar. These predictors are easily accessible for model construction and are frequent physio-chemical water quality indicators in current environment water quality monitoring systems.

"A three-dimensional prediction method of dissolved oxygen in pond culture based on Attention-GRUGBRT,"

Cao, Inaki, et al. 2021

Because a pond is an open body of water, the oxygen concentration in it may be varied in all three dimensions. Products made by aquatic organisms that inhabit various levels of water have distinct oxygen consumption requirements. Dissolved oxygen levels in various parts of the pond cannot be accurately predicted using the conventional, one-dimensional method based on data collected at a single monitoring site. The authors of this study provide a solution to these issues by proposing a dissolved oxygen prediction approach that is based on a gradient boosted regression tree (GBRT) and an Attention-Gated Recurrent Unit (GRU). To begin, data on the environmental variables impacting the dispersion of dissolved oxygen was gathered, and then, using Attention-GRU, a model for predicting dissolved oxygen at the central monitoring site was built. The multifaceted coordinate system was then set up with the core observing point as the origin. To estimate the dissolved oxygen in any place of the pond water, the GBRT algorithm was optimized using the Random Search algorithm (RS). When compared to the LSTM, ELM, and CNN models, the Attention-GRU model presented in this study achieved significantly better results in the a single-dimensional prediction of oxygen content at the central monitoring point (MSE:0.121, MAE:0.219, and RMSE: 0.348). An RS-GBRT model with an MSE of 0.097, MAE of 0.191, and RMSE of 0.313 was suggested for the three-dimensional prediction of pond dissolved oxygen. Each assessment index was noticeably better than the models like Extra Tree, Random Forest, and Bagging. According to the experimental findings, the suggested approach is capable of making precise predictions about the dissolved breathable air in the pond's three-dimensional environment.

METHODOLOGY

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

Various modules have been developed to carry out this project. We will use this module to upload, read, and show datasets to the application. 1) Upload Dataset

Second, we'll use this module to prepare the dataset for use in the application by removing any missing values, normalizing and shuffling the values, and then dividing the dataset in half. We'll then use 80% of the information to train the application and 20% for testing.

3) train LSTM using LIGHTGBM: this module instructs users on how to train an LSTM model using the LIGHTGBM method. After that, they can apply the learned model to test data and evaluate the predictions' accuracy.

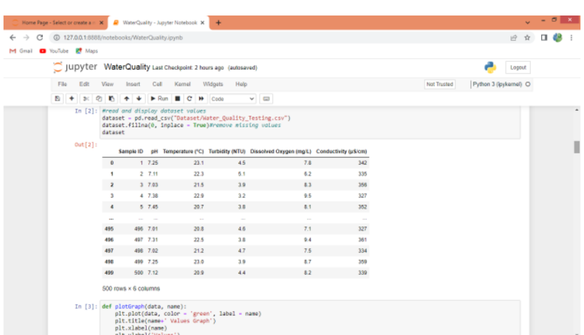
This module trains the LIGHTGBM, Bidirectional Simple RNN, and Attention algorithms using train data. Then, using test data, the trained model may be utilized to assess the prediction accuracy.

5) Train The bidirectional LSTM, GRU, Easy RNN, and Attention: This module is used to train the The opposite direction LSTM, GRU, Simple RNN, and Attention algorithms. Test data is then used to determine the prediction accuracy of the trained model.

6. Accuracy Comparison Graph: All methods may be shown on this graph.

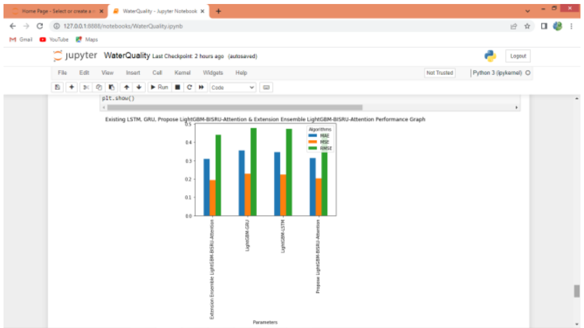
The seventh option is to "Upload Test Data," which allows users to submit test results and have the extension model utilize those results to make an oxygen level prediction.

RESULT AND DISCUSSION

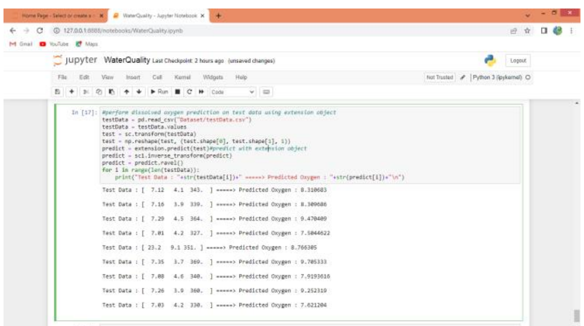


Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

In above result loading and displaying dataset values



In above result x-axis represents algorithm names and y-axis represents MSE, MAE and RMSE values in different colour bars and in all algorithms Extension has got less MSE error



In above screen predicting Oxygen level in test data using extension object

CONCLUSION

Here is what this research adds: In order to increase the accuracy of the predictions, the input information is first processed as well as filled using linear interpolation. Any abnormal data is then corrected using smoothing. In order to determine the impact of various water quality factors on dissolved oxygen, LightGBM was used, bearing in mind the strength of their association. We provide LightGBM-Biru-Attention, an innovative hybrid oxygen absorption prediction model. To accurately estimate dissolved oxygen levels, the unidirectional organization (Biru bidirectional simple recurrent unit) converts future data into present-time point predictions and uses an attention mechanism to pick out crucial points. This paper proposes a new hybrid model

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

called LightGBM-Biru-Attention to solve the problems that traditional oxygen levels prediction methods have with non-linear and non-smooth intensive farming water quality parameters, such as slow operation speed, complicated network structure, and inadequate capture of global contextual information. The three methods—LightGBM, Biru, and Attention—were combined to form the LightGBM-Biru-Attention model. The findings demonstrate that the suggested technique surpasses LightGBM-GRU, LightGBM-LSTM, Biru-Attention, and LightGBM-Biru when RMSE, MAE, MSE, and R2 are used. When it comes to forecasting dissolved oxygen time series, the hybrid methodology of LightGBM-Biru-Attention is now the gold standard in intensive aquaculture due to its superior predictive ability. Additional research is needed since this study has several limitations. More sophisticated algorithms like swarm spider optimization, particle swarm optimization, and the bat algorithm will be studied in the future to see whether they can be integrated with Biru to enhance the accuracy and efficiency of dissolved oxygen level predictions.

REFERENCES

[1] Huang, Juande, et al. "A hybrid model for short-term dissolved oxygen content prediction." Computers and Electronics in Agriculture 186 (2021): 106216.

[2] Cao, Weija, et al. "A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multiscale feature extraction." Aquacultural Engineering 84 (2019): 50- 59.

[3] Nong, Xizhi, et al. "Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China." Ecological Indicators 146 (2023): 109845.

[4] Ling, Jie, et al. "An intrusion detection method for industrial control systems based on bidirectional simple recurrent unit." Computers & Electrical Engineering 91 (2021): 107049. [5] Ding, Shinshu, Yangqin Wang, and Liang Kou. "Network intrusion detection based on Biru and

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

CNN." 2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS). IEEE, 2021.

[6] Ding, Peng, et al. "Efficient Biru combined with feature dimensionality reduction for abnormal traffic detection." IEEE Access 8 (2020): 164414-164427.

[7] Jiang, Ben, et al. "Attention-LSTM architecture combined with Bayesian hyperparameter optimization for indoor temperature prediction." Building and Environment 224 (2022): 109536.

[8] Zhang, Qiuyue, et al. "Transformer-based attention network for stock movement prediction." Expert Systems with Applications 202 (2022): 117239.

[9] Mei, Peng, et al. "Prediction model of drinking water source quality with potential industrial-agricultural pollution based on CNN-Preattentions." Journal of Hydrology 610 (2022): 127934.

[10] Li, Dashe, et al. "An interpretable hierarchical neural network insight for long-term water quality forecast: A study in marine ranches of Eastern China." Ecological Indicators 146 (2023): 109771.

[11] Duan, Jian, Xi Zhang, and Tie-in Shi. "A hybrid attention-based paralleled deep learning model for tool wear prediction." Expert Systems with Applications 211 (2023): 118548.

[12] Wang, Di-in, Lang Li, and Da Zhao. "Corporate finance risk prediction based on LightGBM." Information Sciences 602 (2022): 259-268.

[13] Sun, Miaoli, Mingxi Liu, and Zeina Sima. "A novel cryptocurrency price trend forecasting model based on LightGBM." Finance Research Letters 32 (2020): 101084.

[14] Ren, Juan, et al. "A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism." Energy Reports 8 (2022): 437-443.

[15] Liang, Weiz hang, et al. "Predicting hard rock pillar stability using GBDT, Boost, and LightGBM algorithms." Mathematics 8.5 (2020): 765.

[16] Li, Yiheng, and Weidong Chen. "A comparative performance assessment of ensemble learning for credit scoring." Mathematics 8.10 (2020): 1756.

[17] Lei, Tao, et al. "Simple recurrent units for highly parallelizable recurrence." arrive preprint arXiv:1709.02755 (2017).

Vol.15 No.3(2024),132-143
DOI: https://doi.org/10.61841/turcomat.v15i3.14784

[18] Li, Yuanjiang, et al. "Clothing sale forecasting by a composite GRU– Prophet model with an attention mechanism." IEEE Transactions on Industrial Informatics 17.12 (2021): 8335-8344.

[19] Liu, Juntao, et al. "Accurate prediction scheme of water quality in smart mariculture with deep Bi-S-SRU learning network." IEEE Access 8 (2020): 24784-24798.

[20] Chen, Zijin, et al. "DA-Bi-SRU for water quality prediction in smart mariculture." Computers and Electronics in Agriculture 200 (2022): 107219.