IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Transforming Dissolved Oxygen Prediction For Optimal Water Quality In Intensive Aquaculture

R. Chandupriya[1], Mr. Dharmaiahvari Prasad[2]

[1]PG Student, VEMU Institute of Technology, P. Kothakota.

[2]Assistant Professor, VEMU Institute of Technology, P. Kothakota.

ABSTRACT:

This project introduces a novel approach to forecasting dissolved oxygen levels in aquaculture settings, overcoming complexities in conventional methods. Integrating Light Gradient Boosting Machine (LightGBM) with Bidirectional Simple Recurrent Unit (BiSRU), the model effectively identifies pertinent parameters while minimizing irrelevant variables through linear interpolation and smoothing. LightGBM accurately predicts dissolved oxygen content, while the attention mechanism optimizes BiSRU's hidden states, enhancing predictive accuracy. Outperforming existing models, this hybrid model offers crucial insights for regulating aquaculture water quality. Additionally, an Ensemble model combining Bidirectional LSTM, GRU, Simple RNN, and Attention mechanisms demonstrates further improvement in Mean Squared Error (MSE) compared to individual algorithms, expanding the project's potential impact.

Keywords: Water Quality, Light GBM

INTRODUCTION:

Aquaculture plays a pivotal role in global food production, with China leading the sector. However, ensuring optimal water quality is essential for healthy aquatic life and sustainable yields. Dissolved oxygen (DO) levels serve as a crucial indicator, directly impacting aquatic organisms' survival and growth. Existing predictive models often struggle with computational speed and global contextual awareness, hindering accurate forecasts. Addressing these challenges, this project introduces a hybrid approach integrating Bidirectional Simple Recurrent Unit (BiSRU) and attention mechanisms from machine learning. BiSRU's capacity to capture past and future information, alongside its parallelized architecture, enhances sequence modeling. Additionally, attention mechanisms optimize data processing, focusing on pertinent information for precise predictions. By amalgamating these techniques, this study aims to revolutionize aquaculture management, ensuring sustainable and efficient production practices.

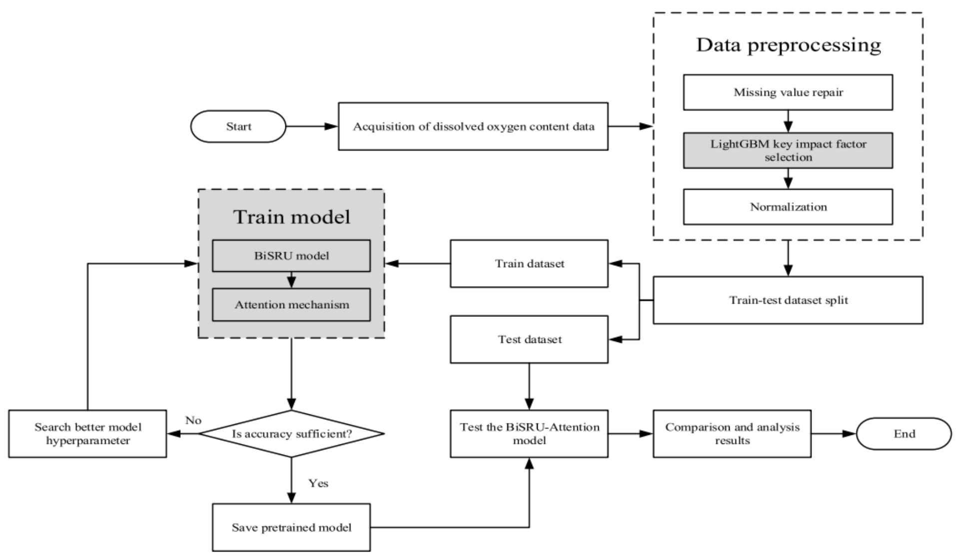
PROBLEM STATEMENT:

The limitations of traditional methods for predicting dissolved oxygen content in aquaculture environments, which are delayed by the nonlinearity, dynamics, and complexity of the system. These challenges result in reduced accuracy and slower prediction speeds, affecting the effective regulation of water quality and the sustainable development of intensive aquaculture.

PROPOSED METHOD:

In past many existing algorithms were introduced like XGBOOST, CNN, LSTM and many more but its prediction error rate is not good enough so author of the paper employing combination of many algorithms such as LIGHTGBM for features selection and remove of irrelevant features from training data, BI-SRU (Bi-directional Simple RNN) and Attention. LIGHTGBM helps in getting relevant features and Bidirectional Simple RNN will filter model by optimizing training features from both backward and forward position. Learning and weighting parameters was updated using ATTENTION algorithm. Model with best MSE (mean square error) will have high weight. MSE, MAE or RMSE refers to difference between original and predicted values so the lower the MSE the better is the model.

ARCHITECTURE:



WATER QUALITY DATASET:

	Sample ID	pH	Temperature (°C)	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Conductivity (µS/cm)
0	1	7.25	23.1	4.5	7.8	342
1	2	7.11	22.3	5.1	6.2	335
2	3	7.03	21.5	3.9	8.3	356
3	4	7.38	22.9	3.2	9.5	327
4	5	7.45	20.7	3.8	8.1	352
...	...	...	...	...	...	...
495	496	7.01	20.8	4.6	7.1	327
496	497	7.31	22.5	3.8	9.4	361
497	498	7.02	21.2	4.7	7.5	334
498	499	7.25	23.0	3.9	8.7	359
499	500	7.12	20.9	4.4	8.2	339

500 rows × 6 columns

In above screen loading and displaying dataset values

METHODOLOGY:

Data Preprocessing:

Import Necessary Libraries: Begin by importing essential Python libraries and packages required for data preprocessing, including pandas, numpy, lightgbm, matplotlib, scikit-learn, keras, os, and others.

Read and Display Dataset: Read the dataset values from the provided CSV file and display them to understand the structure and contents of the data.

Handle Missing Values: Handle missing values in the dataset by replacing them with zeros or using appropriate imputation techniques.

Exploratory Data Analysis (EDA):

Plot Graphs: Perform exploratory data analysis by plotting graphs for various features such as pH, Turbidity, and Conductivity to understand their distributions and patterns.

Feature Selection using LightGBM:

Implement LightGBM Algorithm: Implement the LightGBM algorithm to select relevant features from the dataset.

Fit Model and Calculate Feature Importances: Fit the LightGBM model to the dataset and calculate feature importances. Remove features with less importance (below 20%) from the dataset.

Visualize Feature Importances: Visualize feature importances using bar graphs to understand the significance of each feature.

Dataset Splitting:

Split Dataset: Split the dataset into features (X) and the target variable (Y). Scale the features using MinMaxScaler and reshape them for compatibility with LSTM models.

Further Split Dataset: Further split the dataset into training and testing sets with an 80:20 ratio to prepare for model building and evaluation.

Model Building and Training:

Implement Deep Learning Models: Implement various deep learning models including LSTM, GRU, and BISRU with Attention using the Keras Sequential API. Specify the number of layers, dropout rates, and other parameters for each model architecture. Compile Models: Compile the models using appropriate optimizers and loss functions to prepare them for training. Train Models: Train the models using the training dataset. Utilize Model Check point to save the best-performing model weights during training for future use.

Model Evaluation:

Evaluate Trained Models: Evaluate the trained models using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

Calculate and Visualize Metrics: Calculate and visualize the performance metrics for each model to compare their effectiveness.

Dissolved Oxygen Prediction:

Perform Prediction: Perform dissolved oxygen prediction on a separate test dataset. Read the test data from a CSV file and transform and reshape it for prediction. Predict Levels: Predict dissolved oxygen levels using the trained ensemble model and inverse transform the predictions to obtain actual dissolved oxygen values. Print Predictions: Print the predicted dissolved oxygen levels for each test data point to assess the model's performance.

EVALUATION:

Mean Squared Error (MSE):

MSE is calculated by taking the average of the squared differences between the predicted values and the actual values.

$$\mathrm{MSE} = \frac{1}{n} \quad {}^{n}_{i=1}(Y_i - \hat{Y}_i)^2$$

n is the number of data points.

Yi is the actual value of the target variable for data point i.

Y^I is the predicted value of the target variable for data point i.

 # Calculate Mean Squared Error (MSE)

```
defmean_squared_error(actual, predicted):
mse = np.mean((actual - predicted)**2)
returnmse
```

Root Mean Squared Error (RMSE):

RMSE is the square root of the MSE and provides a measure of the average magnitude of the errors in the predictions.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

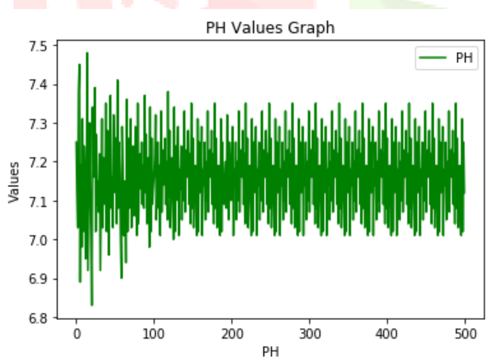
 # Calculate Root Mean Squared Error (RMSE)

```
defroot_mean_squared_error(actual, predicted):
mse = mean_squared_error(actual, predicted)
rmse = np.sqrt(mse)
return rmse
```

Mean Absolute Error (MAE)

```
from sklearn.metrics import mean_absolute_error
# Calculate Mean Absolute Error (MAE)
mae_value = mean_absolute_error(y_test, predict)
print("MAE: ", mae_value)
```

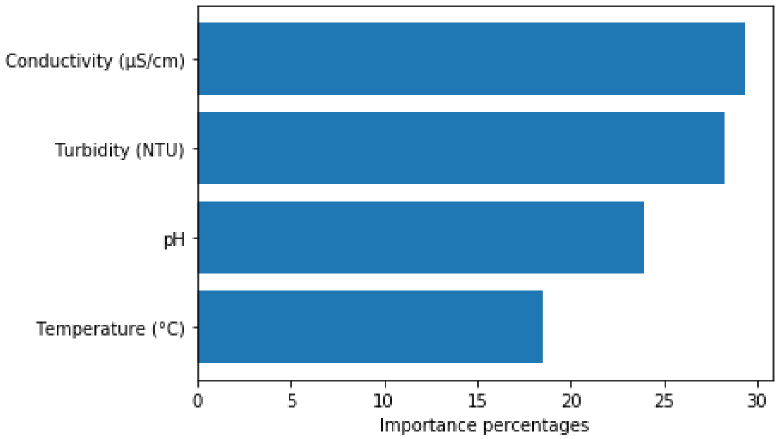
RESULTS:



In above

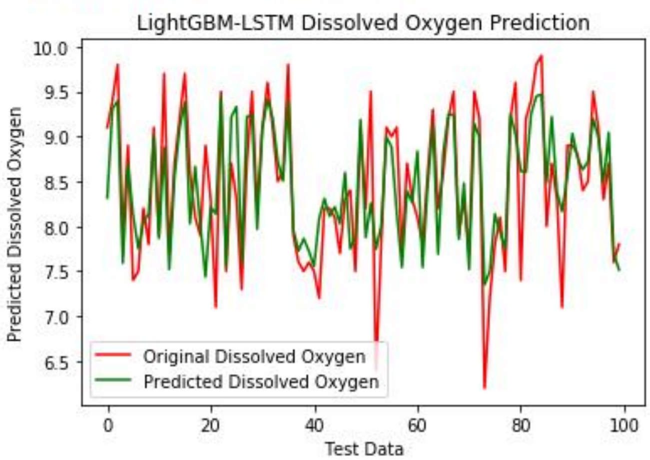
screen displaying PH and turbidity graph where X-axis represents record number and y-axis represents values

```
Temperature having less importance value so it will be removed out
              Features  Importances
1      Temperature (°C)    18.478261
0                    pH    23.913043
2       Turbidity (NTU)    28.260870
3  Conductivity (µS/cm)    29.347826
```



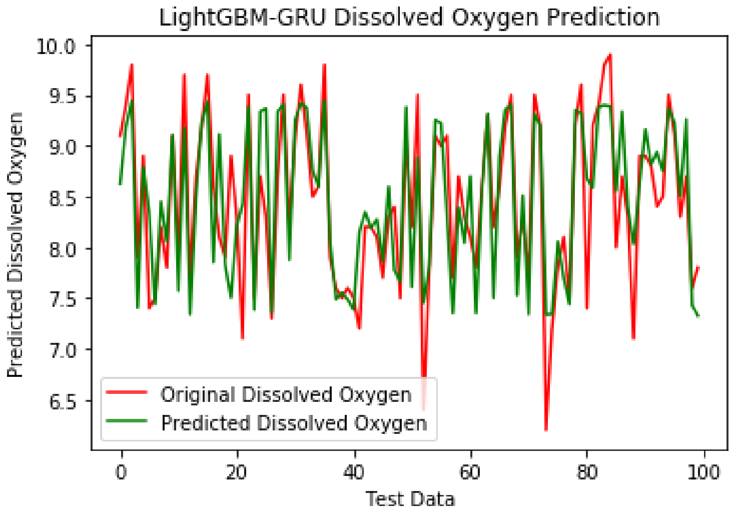
In above we can see importance value for each features obtained from LIGHTGBM and in all 'Temperature' got less value so it will be removed out and remaining 3 features will be used for training. In graph also we can Features Name and importance value.

```
LightGBM-LSTM MSE   : 0.2240669957306242
LightGBM-LSTM RMSE  : 0.47335715451509147
LightGBM-LSTM MAE   : 0.3459676418304442
```



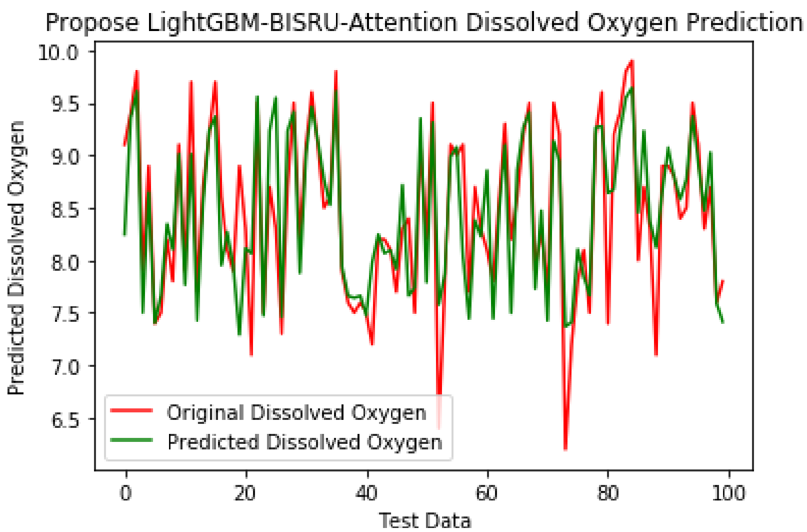
In above with LSTM we got 0.22 as the MSE value and we can see other metrics also (just divide 0.22 / 100). In graph x-axis represents Number of test Data and y-axis represents OXYGEN value and red line indicates TEST DATA OXYGEN LEVEL and green line indicates Predicted OXYGEN level and we can see both lines are overlapping with little GAP so LSTM is good but not accurate

```
LightGBM-GRU MSE  : 0.2282927704043715
LightGBM-GRU RMSE : 0.4777999271707474
LightGBM-GRU MAE  : 0.35497156810760494
```



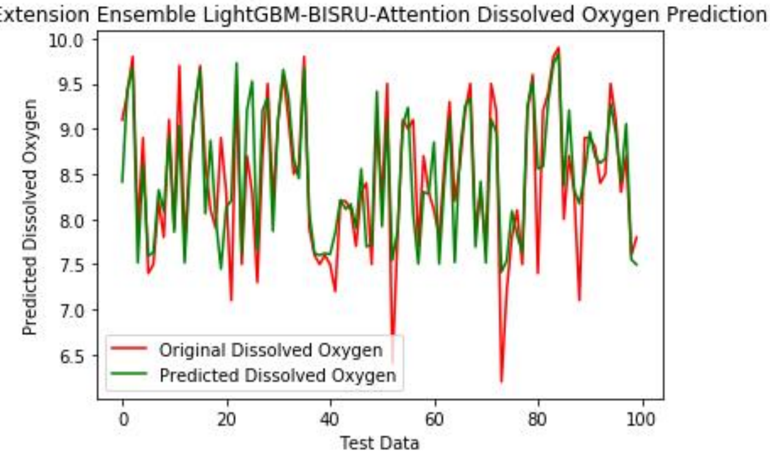
In above GRU output and its MSE values as 0.22

```
Propose LightGBM-BISRU-Attention MSE  : 0.20123957106935095
Propose LightGBM-BISRU-Attention RMSE : 0.44859733734090634
Propose LightGBM-BISRU-Attention MAE  : 0.3125025243759155
```

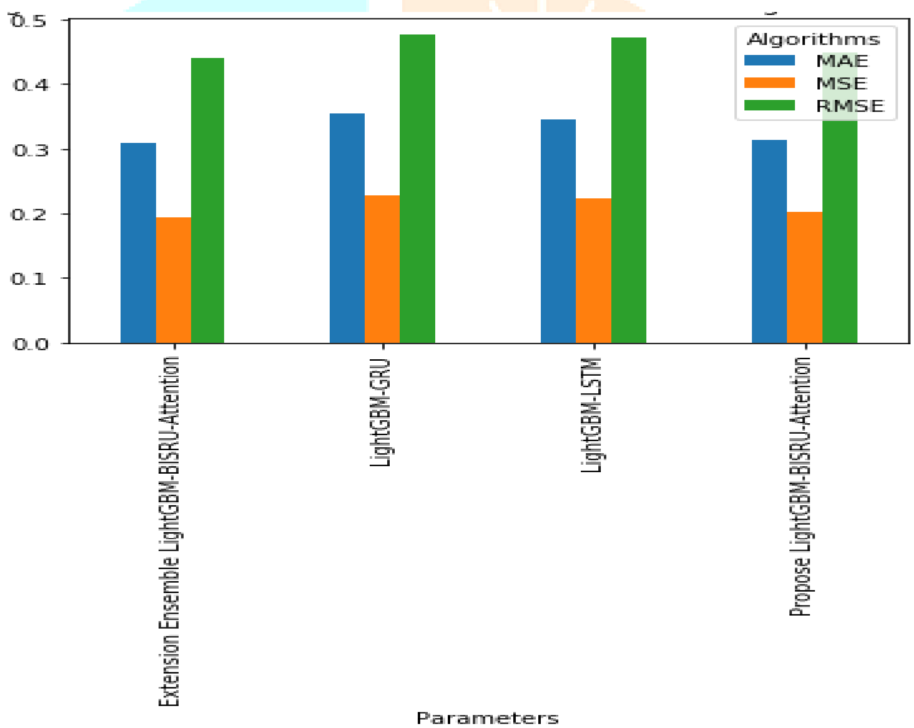


In propose algorithm got 0.20 MSE which is lower than existing algorithm

```
Extension Ensemble LightGBM-BISRU-Attention MSE  : 0.1941879293409357
Extension Ensemble LightGBM-BISRU-Attention RMSE : 0.44066759506564096
Extension Ensemble LightGBM-BISRU-Attention MAE  : 0.30817756175994876
```



Extension model we got 0.19 as the MSE value



In above graph 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

	Algorithm Name	MSE	RMSE	MAE
0	Existing LSTM	0.224067	0.473357	0.345968
1	Existing GRU	0.228293	0.477800	0.354972
2	Propose LightGBM-BISRU-Attention	0.201240	0.448597	0.312503
3	Extension Ensemble LightGBM-BISRU-Attention	0.194188	0.440668	0.308178

Displaying all algorithms performance

Prediction:

```
Test Data : [  7.12   4.1  343.  ] =====> Predicted Oxygen : 8.310683
Test Data : [  7.16   3.9  339.  ] =====> Predicted Oxygen : 8.309686
Test Data : [  7.29   4.5  364.  ] =====> Predicted Oxygen : 9.470409
Test Data : [  7.01   4.2  327.  ] =====> Predicted Oxygen : 7.5044622
Test Data : [ 23.2   9.1 351. ] =====> Predicted Oxygen : 8.766305
Test Data : [  7.35   3.7  369.  ] =====> Predicted Oxygen : 9.705333
Test Data : [  7.08   4.6  340.  ] =====> Predicted Oxygen : 7.9193616
Test Data : [  7.26   3.9  360.  ] =====> Predicted Oxygen : 9.252319
Test Data : [  7.03   4.2  330.  ] =====> Predicted Oxygen : 7.621204
```

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

CONCLUSION

This study introduces a novel hybrid model, Light-GBM-BISRU-Attention, to predict dissolved oxygen levels in intensive aquaculture, crucial for maintaining water quality. Existing algorithms like XGBOOST, CNN, and LSTM have limitations in prediction accuracy. Our model combines LIGHTGBM for feature selection, Bidirectional Simple RNN for optimizing training features, and Attention for learning and weighting parameters, resulting in improved accuracy.

Comparison with LSTM and GRU demonstrates the efficacy of our approach. Additionally, the Ensemble model, integrating Bidirectional LSTM, GRU, Simple RNN, and Attention, further reduces the mean square error (MSE), enhancing predictive capabilities. This comprehensive model holds promise for efficient water quality management in aquaculture.

REFERENCES:

[1] F. Hu, ‘‘Development of fisheries in China,’’ Reprod. Breeding, vol. 1, no. 1, pp. 64–79, 2021.

[2] S. Ayesha Jasmin, P. Ramesh, and M. Tanveer, ‘‘An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques,’’ Expert Syst. Appl., vol. 199, Aug. 2022, Art. no. 117160.

[3] M. H. Ahmed and L.-S. Lin, ‘‘Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique,’’ J. Hydrol., vol. 597, Jun. 2021, Art. no. 126213.

[4] R. Dehghani, H. TorabiPoudeh, and Z. Izadi, ‘‘Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques,’’ Model. Earth Syst. Environ., vol. 8, no. 2, pp. 2599–2613, Jun. 2022.

[5] X. Cao, N. Ren, G. Tian, Y. Fan, and Q. Duan, ‘‘A three-dimensional prediction method of dissolved oxygen in pond culture based on attention-GRU-GBRT,’’ Comput. Electron. Agricult., vol. 181, Feb. 2021, Art. no. 105955.

[6] W. Li, H. Wu, N. Zhu, Y. Jiang, J. Tan, and Y. Guo, ‘‘Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU),’’ Inf. Process. Agricult., vol. 8, no. 1, pp. 185–193, Mar. 2021.

[7] H. Liu, R. Yang, Z. Duan, and H. Wu, ‘‘A hybrid neural network model for marine dissolved oxygen concentrations time-series forecasting based on multi-factor analysis and a multi-model ensemble,’’ Engineering, vol. 7, no. 12, pp. 1751–1765, Dec. 2021.

[8] Q. Ren, X. Wang, W. Li, Y. Wei, and D. An, ‘‘Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network,’’ Aquacultural Eng., vol. 90, Aug. 2020, Art. no. 102085.

[9] X. Cao, Y. Liu, J. Wang, C. Liu, and Q. Duan, ‘‘Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network,’’ Aquacultural Eng., vol. 91, Nov. 2020, Art. no. 102122.

[10] P. Shi, G. Li, Y. Yuan, G. Huang, and L. Kuang, ‘‘Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine,’’ Comput. Electron. Agricult., vol. 157, pp. 329–338, Feb. 2019.

[11] J. Huang, S. Liu, S. G. Hassan, L. Xu, and C. Huang, ‘‘A hybrid model for short-term dissolved oxygen content prediction,’’ Comput. Electron. Agricult., vol. 186, Jul. 2021, Art. no. 106216.

[12] W. Cao, J. Huan, C. Liu, Y. Qin, and F. Wu, ‘‘A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction,’’ Aquacultural Eng., vol. 84, pp. 50–59, Feb. 2019.

[13] X. Nong, C. Lai, L. Chen, D. Shao, C. Zhang, and J. Liang, ‘‘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,’’ Ecol. Indicators, vol. 146, Feb. 2023, Art. no. 109845.

[14] J. Ling, Z. Zhu, Y. Luo, and H. Wang, ‘‘An intrusion detection method for industrial control systems based on bidirectional simple recurrent unit,’’ Comput. Electr. Eng., vol. 91, May 2021, Art. no. 107049.

[15] S. Ding, Y. Wang, and L. Kou, ‘‘Network intrusion detection based on BiSRU and CNN,’’ in Proc. IEEE 18th Int. Conf. Mobile Ad Hoc Smart Syst. (MASS), Oct. 2021, pp. 145–147.

[16] P. Ding, J. Li, M. Wen, L. Wang, and H. Li, ‘‘Efficient BiSRU combined with feature dimensionality reduction for abnormal traffic detection,’’ IEEE Access, vol. 8, pp. 164414–164427, 2020.

[17] B. Jiang, H. Gong, H. Qin, and M. Zhu, ‘‘Attention-LSTM architecture combined with Bayesian hyperparameter optimization for indoor temperature prediction,’’ Building Environ., vol. 224, Oct. 2022, Art. no. 109536.

[18] Q. Zhang, C. Qin, Y. Zhang, F. Bao, C. Zhang, and P. Liu, ‘‘Transformerbased attention network for stock movement prediction,’’ Expert Syst. Appl., vol. 202, Sep. 2022, Art. no. 117239.