Expert Systems With Applications 199 (2022) 117160

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

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

S. Ayesha Jasmin, Pradeep Ramesh, Mohammad Tanveer [*],[1]

Department of Aquacultural Engineering, College of Fisheries Engineering, Tamil Nadu Dr. J. Jayalalithaa Fisheries University, Nagapattinam, 611 002, India

ARTICLE INFO

Keywords:
Prediction testing
Brackish water aquaculture system
Biofloc system
Machine learning methods
Dissolved oxygen and biofloc amount

ABSTRACT

The present study approaches towards the feasibility of prediction and forecasting of dissolved oxygen (DO) and biofloc amount using the state of art machine learning algorithms in a shrimp culture system. The study was carried out considering; average DO and biofloc amount as target parameters in a shrimp culture system. There were seventeen numbers of culture and meteorological parameters considered and three different feature selection techniques used to create twelve different data subsets for model development. The model development was carried out using three popular machine learning algorithms viz., Random Forest, Adaboost and Deep neural network. The totals of thirty-six different models were obtained and their accuracies were evaluated with seven model validation tests and results were obtained and discussed. Out of thirty-six models, Random Forest technique applied model for prediction of dissolved oxygen with combined culture and meteorological parameters (R^2value – 0.709, prediction accuracy – 98.26%, score – 0.7381) was found to be the best one for predictive model development. Moreover, exploratory data analysis was carried out for prediction and a framework for prediction of DO and biofloc amount in a shrimp based biofloc culture system was developed. The dissolved oxygen was found to be more robust in the predictive model development. The Intelligent Framework was developed based on the study conducted to understand and carryout prediction in farming system in a scientific manner. The developed framework can help the literate farmers and new entrants of shrimp farming to devise their own prediction models suitable for the farming conditions.

1. Introduction

Land-based aquaculture plays an important role in providing nutritious food, increasing economic growth, expanding employment opportunities, and reducing poverty by incorporating intensive culture practises to produce a variety of aquatic species to meet demand due to the depletion of marine fish populations and the increasing human population. Inland fisheries has a pivoted role in achieving eight Sustainable Development Goals (SDGs) among seventeen SDGs which were adopted by United Nation member states in 2015 with an aim of creating a better and more sustainable future for all (Lynch et al., 2020). Moreover, aquaculture has been the primary source of fish for human

Abbreviations: Adj R^2, Adjusted R^2; ANN, Artificial Neural Network; APHA, American Public Health Association; API, Application Programming Interface; AUC, Area Under the Curve; Bayes Net, Bayesian Network; BFT, Biofloc Technology; BP, Back Propagation; CAA, Coastal Aquaculture Authority; CNN, Convolutional Neural Network; DA, Discriminant Analysis; DBSCAN, Density-Based Spatial Clustering of Application with Noise; DCGANs, Deep Convolutional Generative Adversarial Networks; DFA, Daily Feed Amount; DNN, Deep Neural Network; DO, Dissolved Oxygen; DOC, Days of Culture; DT, Decision Tree; EDA, Exploratory Data Analysis; FAO, Food and Agriculture Organization; Gas, Genetic Algorithms; GRU, Gated Recurrent Unit neural network; HP, Horse Power; k-NN, k-Neighbour Network; LgR, Logistic Regression; LnR, Linear Regression; M.I, Mutual Information; MAE, Mean Absolute Error; mIOU, Mean Intersection Over Union; ML, Machine Learning; MLP, Multi-Layer Perceptron; MLR, Multiple Linear Regression; MSE, Mean Square Error; NASA, National Aeronautics and Space Administration; NASAPOWER, NASA-Prediction of Worldwide Energy Resources; NN, Neural Network; OSA, Obstructive Sleep Apnea; RBF, Radial Basis Function; ReLU, Rectified Linear Unit; RF, Random Forest; RH, Relative Humidity; RMSE, Root Mean Square Error; SDG, Sustainable Development Goals; SDR, Secchi Disc Reading; SVM, Support Vector Machine; SVR, Support Vector Regression; TLU, Threshold Logic Unit; TPA, Target Parameter Analysis; TSS, Total Suspended Solids.

* Corresponding author.
E-mail address: nobletanveer@gmail.com (M. Tanveer).
[1] ORCHID ID: https://orcid.org/0000-0002-8676-7178.

https://doi.org/10.1016/j.eswa.2022.117160
Received 27 August 2021; Received in revised form 22 March 2022; Accepted 30 March 2022
Available online 4 April 2022

Table 1
Recent ML studies in aquaculture sector.

Author(s), Year	Study location	Algorithm(s) used	Emphasis of the study
Rahman & D'Este, 2015	Australia	DT, SVM, BayesNet, MLP, RBF	Shellfish farm closure prediction and cause identification
Shahriar, Rahman, & McCulloch, 2014	Australia	SVM, k-NN, MLP, REPTree, BayesNet	Prediction of shellfish farm closure
Shahriar & Rahman, 2013	Australia	SVM, RBF,LnR	Algal bloom prediction
Rahman, 2013	Australia	SVM	Benthic habitat mapping
Razman et al., 2019	Malaysia	DA, SVM, k-NN, DT, LgR, RF and NN	Hunger classification in Lates calcarifer
Raman et al., 2016	Malaysia	MLP neural network	Counter system for larvae and juvenile fish
Rossi et al., 2016	Italy	ANN	Fish falsification detection
Kang et al., 2017	Korea	k-means clustering	Optimal aquaculture site selection
Liu et al., 2013	China	GAs, SVR	Aquaculture water quality prediction
Zhao et al., 2018	China	CNN,DCGANs	Live fish identification
Zhou et al., 2019	China	CNN	Evaluation of fish feeding intensity
Yan et al., 2014	China	BP neural network	Prediction model for dissolved oxygen in crab culture pond
Cao et al., 2020	China	K-means clustering and GRU	Prediction of dissolved oxygen in pond culture water
Li et al., 2018	China	MLP, ANN, SVM	Prediction of dissolved oxygen

consumption at a global level since 2016, accounting for fifty two percent of all fish consumed (2018). Though intensive aquaculture achieves maximum production in a minimum quantity of water, it has its own rising concerns such as excessive use of feed, more water usage, disease outbreaks, effluent discharge and environmental impact such as eutrophication (Piedrahita, 2003). Besides high yield and profit, lower impact on surrounding environment and enhancement of feed nutrient utilization should be considered for a sustainable aquaculture system (Bossier & Ekasari, 2017).

Hence, to rectify the above-mentioned constraints, advanced aquaculture system such as Biofloc Technology (BFT) system was developed to reduce excessive water usage, minimize effluent discharge, make complete utilization of feed nutrient and improve the farm bio-security (Avnimelech, 2007; Burford et al., 2003). In BFT system, microorganisms such as bacteria, fungi, microalgae and zooplankton can often floc together into a consumable biomass called 'biofloc' which are utilized by the culturing animal inside, thus contributing increase in growth & survival by enhancing digestive enzyme activities (Long et al., 2015) and by improving immunity of the animal (Kim et al., 2014) making the BFT system environmentally friendly and economically viable.

The aquatic animals perform all their physiological activities in water such as breathing, eating, excreting, maintaining salt balance and reproduction and hence it is important to maintain the water quality which ultimately determines the success and failure of the aquaculture. Water quality parameters viz., dissolved oxygen (DO), water temperature, pH, salinity, alkalinity, solids (Total Suspended Solids (TSS) and settling solids), and orthophosphate should be continuously monitored in the aquaculture ponds. Among that dissolved oxygen (DO) and biofloc level are important indicators of biofloc culture water quality. DO is the main limiting factor of intensive aquaculture, because it causes hypoxia, affects fish growth, food conversion level and feeding efficiency (Pollock et al., 2007). DO level below 3 mg/L cause poor intake of feed and affect the health of the fish (Abdel-Tawwab et al., 2019). Schveitzer et al. (2013) identified that DO in shrimp ponds is related to salinity, temperature, density, and shrimp size. Avnimelech (2012) reported that the biofloc level of 1–40 ml/L is sufficient for growth of shrimp and if the biofloc level exceeds 50 ml/L, it will lead to the depletion of DO during morning time (Samocha et al., 2007).

Moreover, meteorological factors viz., surface temperature (°C), wind speed (m/s), precipitation (mm/day), relative humidity (%), surface pressure (kPa)also influence the water quality in the pond. Climate change has both direct and indirect impacts on fisheries and aquaculture. The direct impact of climate change leads to change in physiology and behavior of fish that affect fish growth, reproduction, mortality and distribution. The indirect impact leads to change in the structure and composition of ecosystem in which the fish depends on food (Allison et al., 2009; 2010). Harley et al. (2006) reported that inland (freshwater) fisheries and aquaculture are affected by changing rainfall patterns and water use. Kapetsky (2000) reported that solar radiation and air temperature influence water temperature, which affects natural productivity of inland and marine waters as well as growth of fish species.

Hence, establishing an accurate and practical prediction model that considers climatic factors is important for smart aquaculture practise, as climate is one of the undervalued features that is rarely considered. Therefore, an accurate forecast model can be developed using the state-of-art artificial intelligence techniques such as; machine learning techniques. Several studies have been conducted with the successful application of machine learning in aquaculture. The synopsis of different ML application studies conducted in aquaculture sector highlighting the algorithms used and emphasis of the study is presented in Table 1.

The present study focuses on the development of a predictive model framework for shrimp culture using daily culture and meteorological data, with biofloc and dissolved oxygen as experimentally identified target parameters. Three different feature selection methods were used to select important feature combinations as subsets for developing different models from the seventeen features collected.

Each feature selection method resulted in two different datasets, resulting in six different datasets with different features. For each model, machine learning algorithms such as ensemble methods and deep neural network techniques were applied to the respective datasets, and each dataset was used as input for model development. The accuracies of each model were evaluated and compared; the performance of the models was evaluated using seven model validation tests. The best approach for framework development was chosen after a comparative analysis of the two different approaches, namely target parameter analysis 1 and 2.The application of these state-of-art techniques paves the way for intelligent aquaculture practises, aids precision farming, and reduces dependence on conventional aquaculture practises. Hence, the intelligent framework was developed based on understandings and findings of the study to develop machine learning prediction model.

2. Materials and methods

2.1. Study location and conditions

The experimental data for the present study were collected from HiTide Sea Farms (11°21′34.1″N, 79°48′45.5″E), Kattur, Nagapattinam dt, Tamil Nadu, India with the study period of 100 days starting from March-June 2018. The typical commercial shrimp farming takes 90–120 days to culture, based on climatic conditions in the prevailing area. In India, the first cycle of culture in a year starts from March to June, following a break season from November to January. The suitable weather for healthy growth of shrimps can be derived during this period. The culture data were taken from two uniform embankment ponds with an area of 0.28 ha (40 m × 35 m). The ponds were prepared for culture and the stocking density was maintained with 120 numbers /m^2 as per the guidelines of Coastal Aquaculture Authority (CAA) for biofloc culture (2005), India. In each pond, total of twenty HP aerators (2HP – 4 nos., 3HP – 4 nos.) were placed in such a manner to maintain clockwise

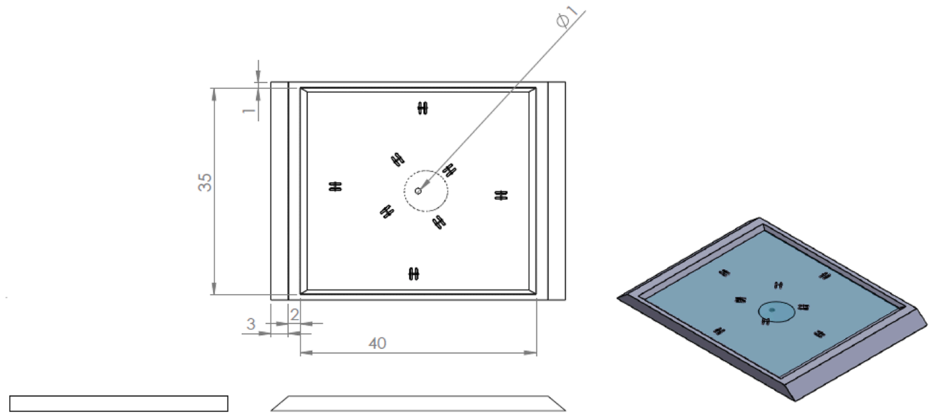

Fig. 1. Layout of the pond and aerator distribution.

Table 2
Standard equipment for measuring different parameters.

S. No	Parameter	Frequency (per day)	Measuring technique/ equipment
1	Daily feed amount (Kg)	6 slots (4-hour duration per slot)	Manually
2	pH change	Morning @ 8.00 AM Evening @ 4.00 PM	pH meter (Nunes instrument, Coimbatore)
3	Water colour	Visual Observation @ 12.00 PM and 5.00 PM	Manually
4	Water temperature (°C)	Morning @ 8.00 AM Evening @ 4.00 PM	Thermometer (Rexson product India, Delhi)
5	Salinity(ppt)	Once in a week	Refractometer (True sense technologies, Rohini, New Delhi)
6	Biofloc level (ml/L)	Measurement @ 12.00 PM	Imhoff cone
7	Secchi disc reading (cm)	Morning @ 8.00 AM Evening @ 4.00 PM	Secchi disc
8	Carbon essence (Kg)	Input to the pond @ 8.00 AM	Manually
9	Water level (cm)	Morning @ 8.00 AM Evening @ 4.00 PM	Measuring scale fixed at pond
10	Operated horse power	6 slots (4-hour duration per slot)	Watt meter(Crown electronics system, New Delhi)
11	Average Dissolved Oxygen level(ppm)	Average of 6 slots (4-hour duration per slot)	DO meter(YSI Ohio, USA)

water circulation and for drain collection at central drain structure (Fig. 1).

The culture parameters such as daily feed amount (Kg), days of culture (DOC), pH change, water colour, biofloc amount (ml/L), secchi disc reading (cm), carbon essence (Kg), Water level (cm), water temperature (°C), salinity (ppt), operated HP, average DO (ppm) were taken into consideration and measured daily at regular intervals using standard equipment (Table 2) and standard methods (American Public Health Association (APHA) (APHA), 2005). The surface temperature (°C), wind speed (m/s), precipitation (mm/day), relative humidity (%), surface pressure (kPa) data were collected daily for the study period and location from NASA Prediction of Worldwide Energy Resources (NASAPOWER) (https://power.larc.nasa.gov/data-access-viewer/). The culture data were collected for all the 100 days of culture from two ponds, which served as the data source for the prediction possibility analysis.

The workflow of the present study is as follows:

(1) Finding out decisive parameters of the biofloc culture.
(2) Exploring relationship between and among different meteorological parameters and culture parameters.
(3) Development of accurate predictive models for the decisive parameters chosen and to evaluate the applicability of predicted models and.
(4) Selection of best model and development of framework.

2.2. Culture and meteorological parameters in biofloc culture practice

The culture and meteorological parameters were collected for the analysis in the present study. It is vital to choose parameters carefully to improve the accuracy and reliability of the analysis process. Therefore, proper care has been taken to consider 17 most important culture (12 numbers) and climatic parameters (5 numbers). The basic description and the role of a particular parameter in the study are given as follows:

2.2.1. Daily feed amount (Kg)

It is the daily feed amount in Kg required for the culture animal. It varies linearly as the animal's size grows on a daily basis. In the present study, the feed was given four times a day in the culture pond, with a six-hour interval between each meal. The daily feed amount is the significant parameter which is associated with water quality and growth responses of culture animal (Tacon et al., 2002). Since shrimp experience more feeding stress than any other animal when they are deprived of food (Lara et al., 2017), it is important to provide enough and precise feed based on their biomass.

2.2.2. Days of culture (DOC)

In the present study, consideration has been given to DOC as one of the critical parameters, which specially denotes the change in culture condition with respect to its value. It is obvious that the culture condition will not remain same throughout the culture period. Therefore, by knowing the value of DOC, it is possible to interpret the intensity of other parameters like; salinity, turbidity, shrimp weight, DO requirement etc. at the given conditions.

2.2.3. pH change

It is a measure of difference in morning pH and evening pH in the culture water. The pH change is the significant parameter to consider because it influences the growth of bacteria and algae (Timmons et al., 2002), affects the physiological conditions of shrimp and controls almost all chemical reaction occur in the water (Lemonnier et al., 2004).

2.2.4. Water colour

Water colour is yet another parameter that is subjective in nature. It indicates the quality of water at the given time. It is the visual measure of water quality and is usually analysed by experienced people. In the present study, it was measured in categorical value and then converted to ordinal values for the convenience and ease of handling.

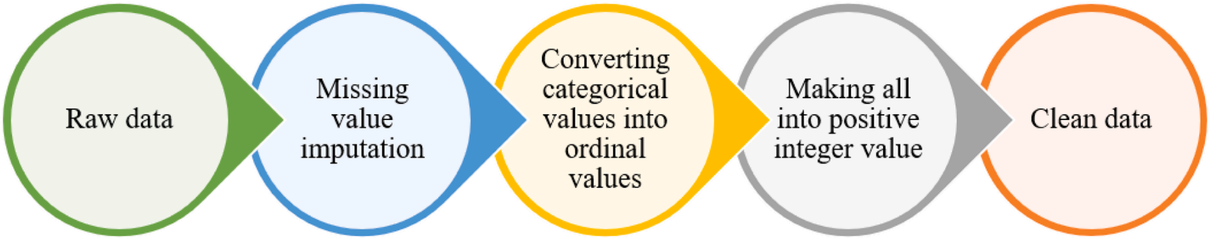

Fig. 2. Steps involved in Data pre-processing.

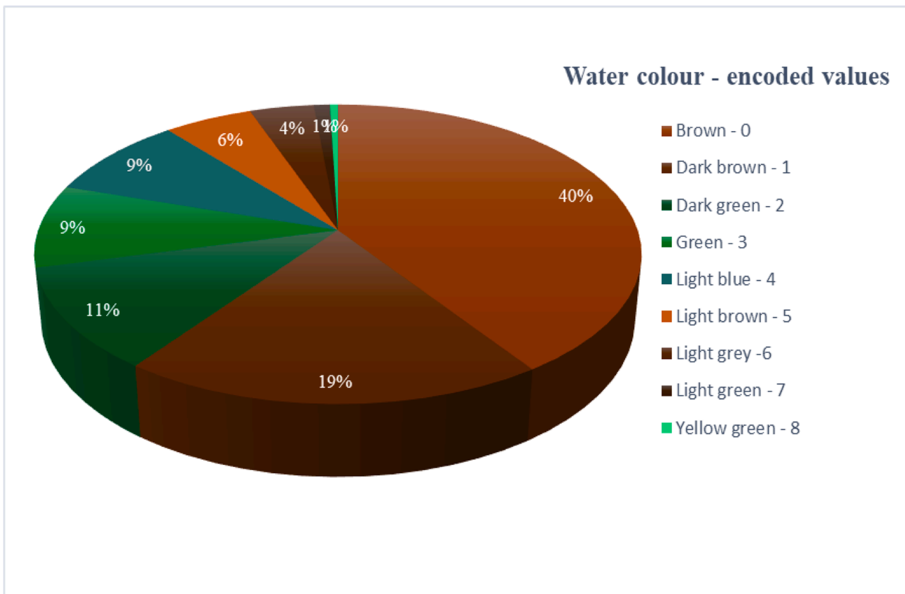

Fig. 3. Water colour and its encoded values.

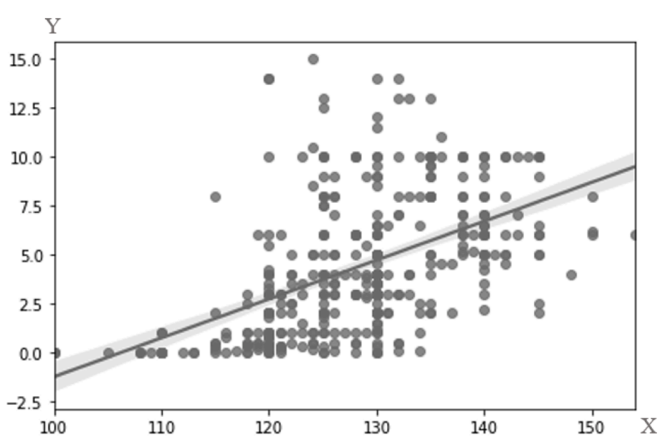

Fig. 4. linear correlation between X and Y variables.

2.2.5. *Biofloc amount (ml/L)*

It is an approximate amount of biofloc present in the culture pond, as determined by the Imhoff cone. Biofloc plays important role in converting nitrogen compounds into microbial protein thus maintaining the water quality and improving the growth rate of the culture animal (Emerenciano et al., 2017). Hence, this parameter helps to understand the amount of microbial biomass present in the culture pond. Therefore, both theoretically and practically, the amount of biofloc was considered to be one of the most important decisive parameters in the present study.

2.2.6. *Secchi disc reading (cm)*

It is the measure used to quantify the turbidity of pond water. It is an important parameter to consider since turbidity influences the penetration of light and reduced light penetration causes increase in the phytoplankton growth which leads to oxygen deficit (Vinatea et al., 2010). Higher biofloc concentration and total suspended solids, both of which can be detrimental to fish growth, are indicated by a lower secchi disc value.

2.2.7. *Carbon essence (Kg)*

It indicates the amount of carbon source in Kg added to the pond. The choice of carbon source is important since it determines the composition of floc and amount of energy storage polymers (Hollender et al., 2002). In the present study, molasses was given as primary carbon source in order to maintain the required C: N ratio in the pond.

2.2.8. *Water level (cm)*

The water level in the pond was measured using standard scale installed inside the pond so as to determine the volume of water present in the pond. The daily water level values were recorded in the present study, allowing for indirect quantification of evaporation loss.

2.2.9. *Water temperature (°C)*

Water temperature is an important parameter considered in the present study. It affects the floc characteristics and is normally associated with ambient condition. In cold (4 °C) as well as cool conditions (18–20 °C), deflocculation of bioflocs occurs (Wilén et al., 2003) whereas at higher temperature (30–35 °C), bulking of sludge happens. At intermediate water temperature (20–25 °C), biofloc will remain

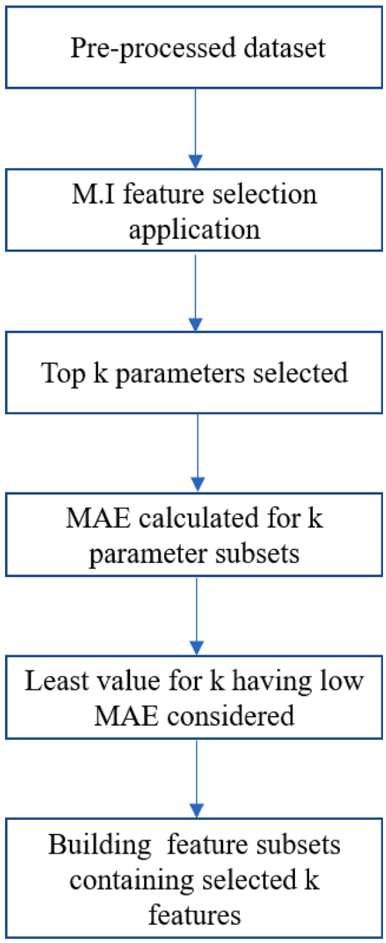

Fig. 5. M.I feature selection procedure.

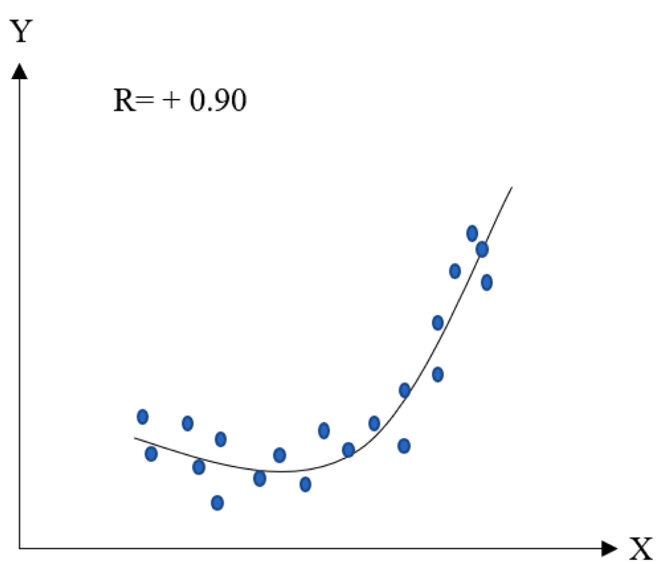

Fig. 6. Spearman's correlation between X and Y variables.

stable (Krishna & Van Loosdrecht, 1999).

2.2.10. Salinity (ppt)

Salinity is the concentration of dissolved salts usually expressed in parts per thousand (ppt) and is measured using a salinity refractometer. Though salinity has no effect on shrimp survival (Li et al., 2017), studies

Table 3
Feature selection methods used in various studies.

S. no	Title of the study	Authors & Year	Feature selection methods employed
1	Daily activity feature selection in smart homes based on Pearson correlation coefficient	Liu et al., 2020	Pearson's correlation
2	Sentiment analysis on movie reviews using ensemble features and pearsoncorrelation-based feature selection	Rangkuti et al., 2018	Pearson's correlation
3	Suitable feature selection for OSA classification based on snoring sounds	Temrat et al., 2018	Pearson's correlation
4	Mutual information-based feature selection for intrusion detection systems	Amiri et al., 2011	Mutual information
5	Network anomaly detection: methods, systems and tools	Bhuyan et al., 2013	Mutual information
6	Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier	Paul et al., 2017	Spearman's correlation
7	Machine learning in DNA microarray analysis for cancer classification	Cho & Won, 2003	Spearman'scorrelation
8	Speech emotion recognition based on feature selection and extreme learning machine decision tree	Liu et al., 2018	Spearman's correlation

show that the growth rate of pacific shrimp is compromised at salinities below the isosmotic point (Ray & Lotz, 2018).

2.2.11. Operated HP

Operated Horsepower is the measure of energy required for aeration in the pond. Aeration devices ranging from 0.1 to 100 W/m^3 are used for aeration in aquaculture ponds (Boyd, 1998).Increased horsepower results in higher aeration mixing intensity, which raises the shear rate and causes more floc breakage (Spicer & Pratsinis, 1996). Therefore, it was considered in the present study.

2.2.12. Average DO (ppm)

It is the average of DO measured every four hours of the day in the pond. It is a significant factor that influences shrimp growth and health in culture ponds. The oxygen intake by sediments, oxygen production by photosynthesis, oxygen consumption by aquatic animals, biochemical reactions in water, and pond aeration all play a role in determining the DO in the pond water (Datta, 2012). Hence, average DO was considered as one of the decisive target parameters in the present study.

2.2.13. Surface temperature (°C)

It is the temperature of the atmosphere at a height of 2 m above ground level. The atmospheric temperature varies with the presiding climate. The growth rate of shrimp in the cold season is obviously lower than in the warm season, resulting in extending the culture period during the cold season to produce marketable size shrimps (Ponce-Palafox et al., 2019). Hence, surface temperature was considered in the present study.

2.2.14. Wind speed (m/s)

It is the wind speed measured at a height of 2 m above ground level. Strong winds improve the oxygenation and diffusion of various gaseous metabolites such as carbon dioxide, hydrogen sulphide, and ammonia from the water by accelerating up the rate of air exchange in the atmosphere. The mixing of wind and water also breaks the thermal stratification and helps oxygenated water penetration through the bottom of the pond.

Table 4
Predictive algorithms used in the study and its successful application in the recent research studies.

S. no	Title of the study	Authors & Year	ML algorithms	Algorithms compared	Obtained accuracy	Remarks
1	Estimation of biomass in wheat using random forest regression algorithm and remote sensing data	Zhou et al., 2016	Random forest regression	SVR and ANN	R^2 – Regression for the three stages were 0.533, 0.721 and 0.79	RF regression works best for biomass prediction – analogical to biofloc in the present study
2	Modelling of impact of water quality on infiltration rate of soil by random forest regression	Singh et al., 2017	Random forest regression	ANN	Training – MAE (mm/h) – 1.550Testing MAE (mm/h) – 3.173	RF provided better results in prediction of water quality
3	High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm	Mutanga et al., 2012	Random forest regression	–	RMSE Pred _ 0.441 kg/m^2	RF regression works best for biomass prediction – analogical to biofloc in the present study
4	Comparison of AdaBoost and Logistic regression for detecting colorectal cancer patients with synchronous liver metastasis	Wen et al., 2009	AdaBoost regression	Logistic regression	Accuracy – 0.82 AUC – 0.854	Adaboost works for highly sensitive and intensive applications
5	House price prediction using regression techniques: A comparative study	Madhuri et al., 2019	AdaBoost regression	Multiple regression, Ridge regression, Lasso regression, Elastic net, Grading boosting regression	R^2 score – 0.78 RMSE – 0.179 MSE – 0.321	Adaboost works well for numerical regression analogical to DO values in present study
6	A DNN-based semantic segmentation for detecting weed and crop	You et al., 2020	Deep neural network	–	mIOU accuracy – 89.01 – 90.22%	DNN provided suitable results in farming condition
7	Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application	Mohan & Patil, 2018	Deep neural network	KNN, Learning vector quantisation, Random Forest	Accuracy 1.For 90% Training, 10 % testing – 81.23% 2. For 80% Training, 20 % testing – 78.98% 3.For 70% Training, 30% testing – 80.06%	DNN works well in agriculture and allied sectors

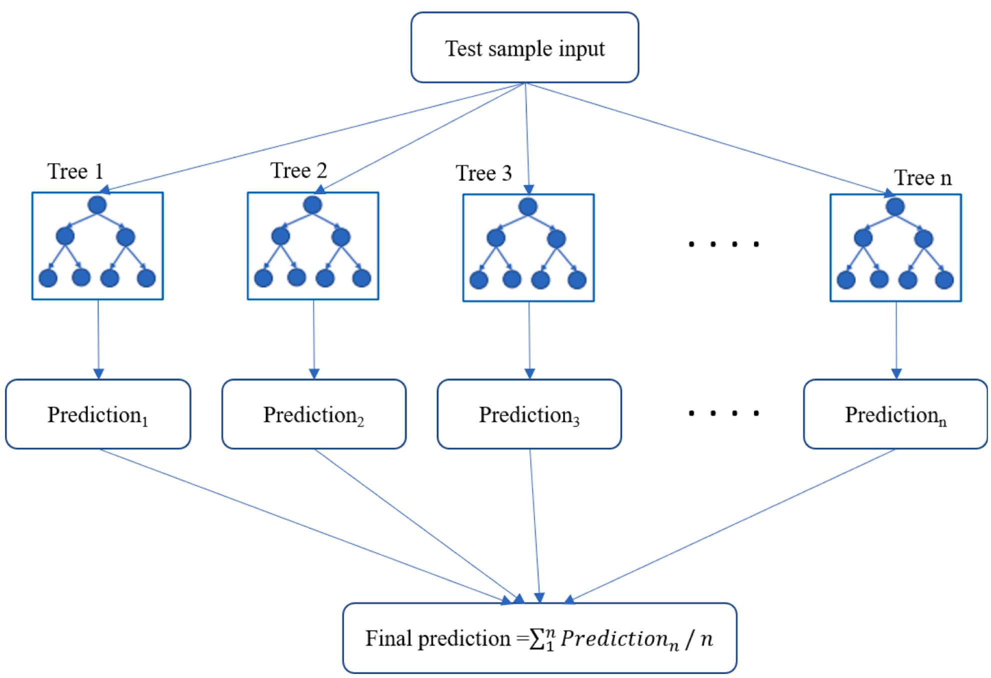

Fig. 7. Random forest algorithm.

2.2.15. *Precipitation (mm/day)*

Precipitation was also considered in the present study. It affects the pH of the pond water and the feed intake of the culture animals. Heavy rainfall may result in flooding, invasion of unwanted species, infilling of pond water and pond wall erosion (Rutkayová et al., 2018).

2.2.16. *Surface pressure (kPa)*

It is the measure of atmospheric pressure at a height of 2 m above the ground level. It is also an indicator of weather, where low-pressure system cause cloudiness, precipitation and high-pressure system make clear weather (Hurley & Boos, 2015).

2.2.17. *Relative humidity (%)*

It refers to the amount of moisture content present in the atmosphere at a given temperature and pressure. It is an important parameter because it influences the pond water evaporation; the higher the humidity level, the lower the evaporation rate (Trenberth, 2011).

Table 5
Functions of hyperparameter and its values used in RF and AdaBoost.

Hyper-parameter	Function	Hyper-parameter valuesoptimised
n_estimators	Number of decision trees being built (default value is 100 in sklearn)	100,600,700,1000
max_features	Maximum number of features selected for a node splitting (types: 'sqrt','log2')	'sqrt', 'auto'
max_depth	Maximum number of levels set in a decision tree (if none selected, it continues to split until reaching purity)	20,25
min_samples_split	Minimum number of samples required to split an internal node (default value = 2)	2,15
min_samples_leaf	Minimum number of data points requirements in a node of Decision tree	1
random_state	Provided to control random number generator used. Model with specific random_state produce similar results and model with 0 as random_state produce different results by reusing same instance	0

2.3. Data pre-processing

The collected raw data was pre-processed into clean usable data (Fig. 2), in order to develop machine learning models using state of the art algorithms. The study was carried out in Jupyter notebook using python 3.8 programming language and anaconda libraries. The data pre-processing steps used in the present study are,

Missing value imputation -The collected data were checked for imputation to fill missing values.
Converting categorical features – Then, based on the dataset's counts, a feature called water colour with categorical values was converted into ordinal values (Fig. 3).
Making all into positive values – Moreover, data with negative difference values were modulated to positive integer values in order to make the dataset more suitable for machine learning algorithms.

Then, Exploratory data analysis (EDA), a preliminary analysis to derive relationship and correlation between different parameters was carried out. It often employs data visualization methods to understand the insights about different features and their correlations. It is important to identify features that are significant and highly influential on target parameters which in turn help to improve the model accuracy.

2.4. Selection of target parameter for prediction and validation

Features that are more important for a successful culture were chosen from the parameters selected, and models were developed for them. In the present study, two features were experimentally chosen as target parameters and the analysis carried out.

2.4.1. Target parameter analysis 1: Prediction of dissolved oxygen level in the pond

The dissolved oxygen is the most important limiting factor affecting productivity, a meteorology-based prediction model for DO is required. Therefore, the operational cost of the culture can be reduced and make better prediction about future events. Many studies have been conducted that demonstrate the importance of DO in biofloc culture (Boyd et al., 2018; Emerenciano et al., 2017; Mallya, 2007), which assisted in the selection of average DO as one of the target parameters.

2.4.2. Target parameter analysis 2: Prediction of biofloc amount in the pond

Biofloc amount, in addition to average DO, is an important parameter in biofloc culture, as it affects productivity. The higher the biofloc level, the more aeration and mixing power is required, resulting in higher operational costs. Therefore, maintaining the optimum biofloc level of 10–15 ml/L in the shrimp pond is essential for proper functionality of the system. Many other studies (Burford et al., 2003; Suita et al., 2016; Thompson et al., 2002; Wasielesky et al., 2006) have shown that biofloc is more important in increasing productivity, which led to the selection of biofloc amount as one of the target parameters for which a prediction model was developed. Hence, the biofloc prediction model will help us in forecasting future events and taking strategic precautions based on the predicted values, while also reducing our reliance on traditional methods such as Imhoff cone measurements.

2.5. Data sampling framework

Three different feature selection methodologies were applied to the dataset with the aforementioned parameters in order to select the subset of input features that were highly related to the target variable. The feature selection methods are useful for identifying irrelevant features, reducing memory usage and computational time, and improving prediction accuracy.

The feature selection methods utilized in the present study are,

a) Pearson's correlation feature selection
b) Mutual Information feature selection
c) Spearman's rank correlation feature selection

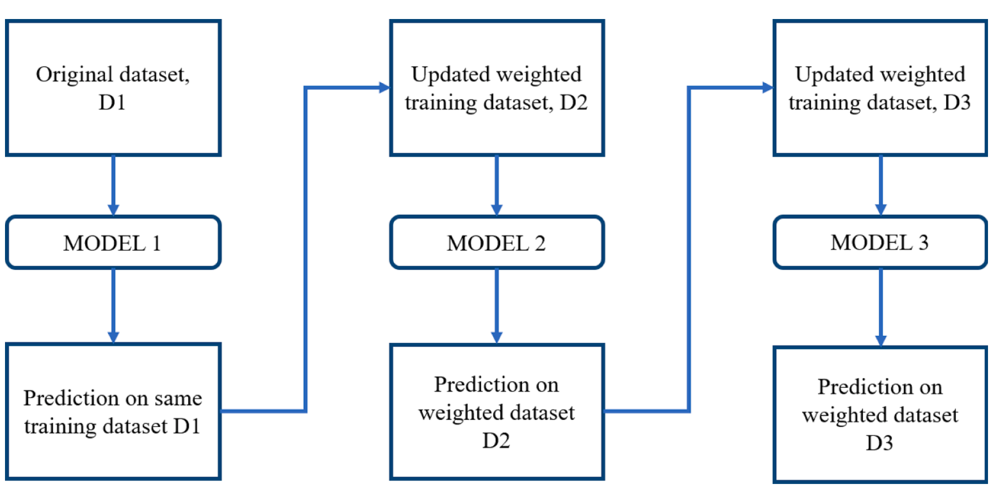

Fig. 8. AdaBoost algorithm.

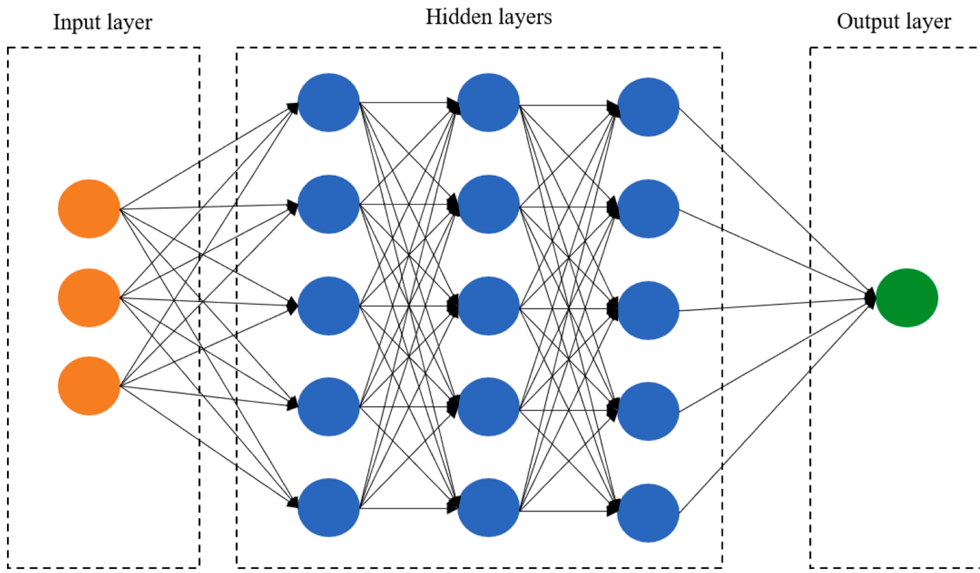

Fig. 9. Multi-layer perceptron -DNN.

Table 6
Typical regression MLP architecture.

Hyper-parameter	Typical value	Hyper-parameter valuesoptimised
Number of input neurons	One per input feature	128
Number of hidden layers	Depends on the problem, typically 1 to 5	3
Number of neurons per hidden layer	Depends on the problem, typically 10 to 100	256
Number of output neurons	1 per prediction dimension	1
Hidden activation	ReLU	RELU
Output activation	None/ ReLU/Logisitic/Tanh	Linear
Loss function	MSE/MAE/Huber	MAE
Epoch	10,100,500,1000 and larger	500
Batch size	32,64,128 etc	32

Table 7
Feature subset selection criteria.

TPA 1		TPA 2	
Feature subset	Selection criteria	Feature subset	Selection criteria
D_{I1}	Features having Pearson's correlation more than 50%	D_{II1}	Features having Pearson's correlation more than 50%
D_{I2}	D_{I1} and all meteorological parameters included	D_{II2}	D_{II1} and all meteorological parameters included
D_{I3}	Top k Mutual information features with least MAE	D_{II3}	Top k Mutual information features with least MAE
D_{I4}	D_{I3} and all meteorological parameters included	D_{II4}	D_{II3} and all meteorological parameters included
D_{I5}	Features having Spearman's correlation more than 50%	D_{II5}	Features having Spearman's correlation more than 50%
D_{I6}	D_{I5} and all meteorological parameters included	D_{II6}	D_{II5} and all meteorological parameters included

2.5.1. *Pearson's correlation feature selection*

It is the most common correlation method which measures the linear correlation between two variables. The concept behind selecting Pearson's correlation feature is that good variables are highly correlated with the target variable. Pearson's correlation is obtained by dividing slope obtained by linear regression of variable Y by variable X, by the ratio of standard deviations (Fig. 4). Let $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ be the means of two variables X and Y respectively and$n$ = number of observations then Pearson's correlation coefficient $\rho_{Pearson}$ is defined as follows:

$$\rho_{Pearson}(\mathrm{X},\mathrm{Y}) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}}$$

The linear correlation score ranges from −1 to 1, with a negative value representing that the factors are not relevant in terms of linearity, a positive value representing that the parameters are extensively matching with some proportionality in linear terms, and a zero representing that the factors have no relation in both linear and non-linear scales. Fixing a threshold value for feature selection is an important step that can help to eliminate lower correlation features. Features with values greater than the threshold value were chosen for further analysis because they can provide desirable results for machine learning prediction analysis, as the larger the value, the stronger the relationship. For the present study, scikit-learn library was implemented to provide the correlation statistic using the *corr ()* functionality.

2.5.2. *Mutual information feature selection*

Mutual information (MI) is a decision tree method that uses information gain to measure the reduction in uncertainty for one variable when the other variable's value is known. Since it calculates the reduction in entropy from transformation of a dataset, it can be used to select relevant features to predict the target parameter. If X and Y are two variables, then MI can be defined as follows:

$$\mathrm{MI}(\mathrm{X},\mathrm{Y}) = \iint f_{X,Y}(x,y)\log_2\left(\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)}\right)dxdy$$

MI can consider only positive values of the dataset and it is a vector quantity. High MI value indicates larger reduction in uncertainty and vice versa, whereas zero MI indicates that the variables are independent. The mutual information feature selection procedure is explained in Fig. 5.

2.5.3. *Spearman's rank correlation feature selection*

Unlike Pearson's correlation coefficient, Spearman's rank correlation does not take linearity into account when measuring the strength and direction of a monotonic association between two variables (Fig. 6). It is the application of Pearson's correlation in which the data is converted into ranks before the co-efficient is calculated. This method is

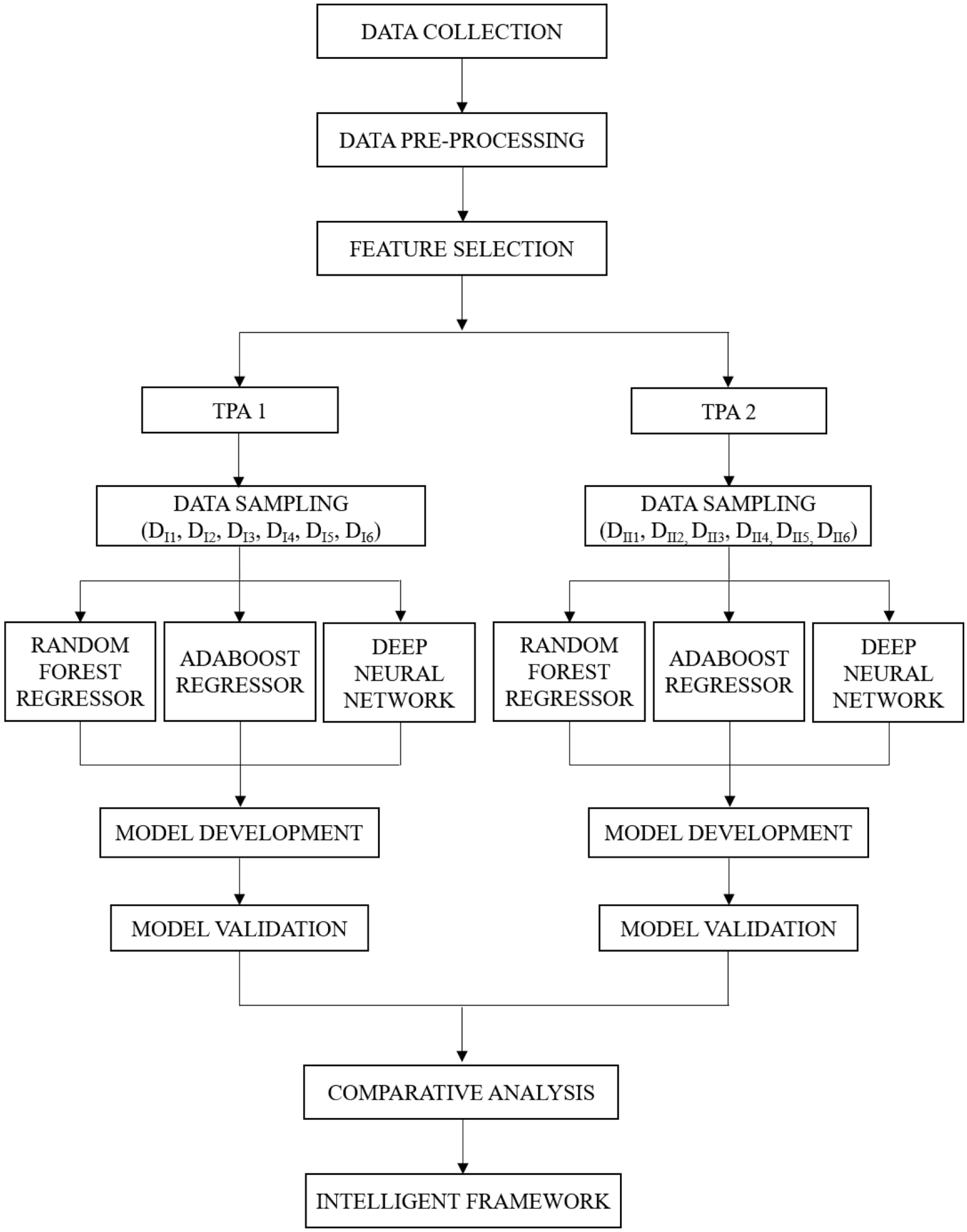

Fig. 10. Overall process flow chart.

used to find features that are relevant to the target variable that were not found by considering the linear relationship between them because it measures the monotonic relation between variables.

Spearman's correlation for the variables X and Y is calculated as follows,

$$\rho_{Spearman}(X, Y) = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where d_i = difference between the ranks of x_i and y_i and$n$ = number of observations.

Like Pearson's correlation, $\rho_{Spearman}$ assumes values between -1 and 1, where $+1$ indicates perfect positive monotonic relation, -1 indicates perfect negative monotonic relation and 0 indicates that the variables are monotonically independent. Also, $-1<\rho_{Spearman}$ greater than 1 indicates imperfect monotonic relationship. The present study utilises this selection technique to sort of decisive parameters which are suitable for the analysis and prediction testing. Previous studies employed the feature selection techniques are shown in Table3.

2.6. Prediction algorithm implementation for different cases

In this study, the prediction of the target parameter is a regression problem with a continuous variable as the output., popular machine learning techniques which are proven to be effective in many agriculture and allied farming applications (Ale et al., 2019; Chen et al., 2020; Sirsat et al., 2017) such as ensemble methods, Random forest and AdaBoost algorithms and Deep neural network algorithm were applied using Jupyter notebook on Anaconda environment using python as

Table 8
Selected feature subsets.

Features	TPA 1: Average DO as target parameter						TPA 2: Biofloc amount as target parameter					
	D_{I1}	D_{I2}	D_{I3}	D_{I4}	D_{I5}	D_{I6}	D_{II1}	D_{II2}	D_{II3}	D_{II4}	D_{II5}	D_{II6}
DOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Daily feed amount (Kg)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
pH change												
Water colour											✓	✓
Biofloc amount (ml/L)	✓	✓			✓	✓						
SDR (cm)	✓	✓			✓	✓	✓	✓			✓	✓
Carbon essence (kg)												
Water level (cm)												
Water temperature (°C)							✓	✓			✓	✓
Salinity (ppt)	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓
Operated HP	✓	✓	✓	✓			✓	✓	✓	✓		
Average DO (ppm)							✓	✓			✓	✓
Surface temperature (°C)	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind speed (m/s)		✓		✓		✓		✓		✓		✓
Precipitation (mm/day)		✓		✓	✓	✓		✓		✓		✓
Surface pressure (kPa)	✓	✓		✓	✓	✓		✓	✓	✓		✓
R.H (%)		✓		✓		✓		✓		✓		✓

programming language. The literature study has been carried out to find out suitable algorithms for the present study. Hence, the successful algorithms from literature related to present study were critically analysed (Table 4) and chosen as model development algorithms. It involves the application of an algorithm directly to the data set, and development of predictive models with specific accuracy.

2.6.1. Random forest regression

Random forest is an ensemble machine learning technique which performs both classification and regression tasks using hundreds and thousands of decision trees splits and a statistical technique called bagging. The final output of the RF regression algorithm is the average of all decision tree outputs (Fig. 7), and the voting classifier is used for the same purpose in classification. Each decision tree consists of leaf nodes and decision nodes, which use a test function to evaluate the given input and then pass it on to different branches based on the features of the sample. In the first step, the decision tree must make the best split among all variables, and the process must be repeated recursively until the terminal node is reached. If X be the input vector containing m features with X= $\{x_1,x_2,\ldots,x_n\}$, Y be the output scalar and S_n be the training set containing n observations, then at the end of the training process, a prediction function formed as follows:

$$\widehat{y} = \widehat{h}\left(X, S_n^{\Theta}\right)$$

where $\widehat{y}$ = output of the decision tree, $\widehat{h}$= prediction function and Θ = randomized variable.

In RF, all decision trees are combined using bagging technique (Breiman, 1996) so as to reduce the variance associated with prediction thereby improving the prediction performance. The bagging algorithm randomly selects training data subset samples sand applies decision trees to these samples to get q number of prediction trees. The final result is the average of output of individual decision trees, the output can be obtained as follows:

$$\widehat{Y} = \frac{1}{q}\sum_{l=1}^{q}\widehat{Y}_l = \frac{1}{q}\sum_{l=1}^{q}\widehat{h}\left(X, S_n^{\Theta_l}\right)$$

where $\widehat{Y}_l$ = output of the l -th tree, and $l = 1, 2\ldots\ldots, q$.

Further, the hyper parameter tuning helps to increase the predictive power of the model which are available in sklearn's built-in random forest function. In this study, some of the hyperparameters such as n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf are tuned at the stage of model development before taking the averages of prediction and the functions of each hyper parameter used are discussed in Table5.

2.6.2. Adaboost regressor

AdaBoost also known as Adaptive Boosting, is a boosting technique in ensemble method of machine learning in which it makes n number of decision trees in a sequential manner, except for the first, each decision tree is grown from previously grown decision trees (Fig. 8). In this technique, weights are assigned to each instance, with higher weights assigned to incorrectly predicted instances. Adaboost's core concept is to fit a series of weak learners to repeatedly updated versions of the data. It begins with fitting a regressor to the original dataset, then fitting additional regressors to the same dataset, but with the weights of instances adjusted according to the current prediction's error, and finally focusing on difficult cases with the subsequent regressors.

In this study, Scikit learn built-in Ada Boost Regressor was implemented which uses the algorithm known as AdaBoost.R^2 (Drucker, 1997).The algorithm was written using the following steps:

a) Consider a sequence of input with m number of samples $(x_1,y_1)\ldots\ldots(x_m,y_m)$ where $y \in$ R.
b) Initialize the first iteration ($t = 1$) with Distribution D_t (i) = $1/m$ for all i and average loss function.$\overline{L_t} = 0$
c) Iterate while average loss function $\overline{L_t} < 0.5$, Call Weak learning algorithm and provide it with distribution D_t.
d) Build the regression model $f_t\ (x) \rightarrow$ y and calculate the loss for each training sample as:

$$|f_t(x_i) - y_i|$$

e) Calculate the loss function L_t $(i)$for each training sample using three different functional forms as:

- Linear: $L_t\ (i) = l_t\ (i)\ /\ \mathrm{Den}_t$
- Square law: $L_t\ (i) = (l_t\ (i)\ /\ \mathrm{Den}_t)^2$
- Exponential: $L_t\ (i)$ = 1-exp (-$l_t\ (i)\ /\ \mathrm{Den}_t$)

where Den_t =.max ($l_t\ (i)$)

i = 1…m.Calculate an average loss as:.$\overline{L_t} = \sum_{i=1}^{m} L_t(i)D_t(i)$

g) Set $\beta_t = \overline{L_t}/(1 - \overline{L_t})$ and update distibution D_t as.

$$D_{t+1}(i) = \frac{D_t(i)\beta_t^{(1-L_t(i))}}{Z_t}$$

whereZ_t = normalization factor chosen for such that D_{t+1} will be a distribution.

Table 9
Inferences on results of EDA.

S. no	Target parameter	Another parameter	Trend following	Correlation value	Inference
1	Average DO (ppm)	Surface pressure (kPa)	Linear increment	0.60	DO in the pond decreases with decrease in surface pressure; it proves the atmospheric inclusion of oxygen has taken considerable part for the aid of oxygen supply in the present culture.
2	Average DO (ppm)	Biofloc amount (ml/L)	Linear decrement	0.63	Increase in biofloc amount decreases the DO level which represents the consumption of oxygen by flocs, benthos, feed oxidation and shrimp respiration.
3	Biofloc amount (ml/L)	Surface temperature (°C) & water temperature (°C)	Linear increment	0.71 & 0.55	In this pond, water temperature and atmospheric temperature found to be greater than 27 °C, hence higher temperature influences the bulking of Biofloc by supporting mesophilic heterotrophs to grow.
4	Average DO (ppm)	Salinity (ppt)	Linear decrement	0.63	When salinity increases, DO decreases. This is due to the fact of salting out effect i.e. increasing salt concentration leads to decrease in oxygen solubility.
5	Biofloc amount (ml/L)	Operated HP	Linear increment	0.78	Increase in biofloc level needs more aeration to suspend them so that Operated HP increases gradually with increase in biofloc. In this pond Operated HP is more influenced by Biofloc.
6	Biofloc amount (ml/L)	Daily feed amount (Kg)	Linear increment	0.70	Since daily feed amount is given based on the biomass, increase in shrimp size necessitates the increase in biofloc amount causes linear increment of daily feed amount.
7	Biofloc amount (ml/L)	SDR (cm)	Non-linear decrement	0.57	Biofloc increases with decrease in SDR value. Lower SDR value means higher the turbidity i.e. concentration of suspended solids which indicates the increase in Biofloc level.

h) Set $t = t + 1$ and output the final hypothesis:

$$f_{fin}(x) = \inf\left[y \in Y : \sum_{t:f_t(x)\leq y} \log\left(\frac{1}{\beta_t}\right) \geq \frac{1}{2}\sum_t \log\left(\frac{1}{\beta_t}\right)\right]$$

In the present study, hyper parameters such as random_state and n_estimators were tuned before making prediction. The functions of the hyper parameters are described in Table 5. Thus, it helped to optimize the model prediction.

2.6.3. Deep neural network (DNN)

DNN model is a Multi-layer perceptron (MLP) deep neural network containing deep stack of hidden layers. One input layer, one or more hidden layers with threshold logic unit (TLU), and one TLU output layer make up a MLP as shown in (Fig. 9). Every layer except output layer has bias neuron and fully connected to next layer. The algorithm works by handling one mini-batch at a time and goes through full training datasets multiple times. Each pass is called an epoch and each mini-batch is passed into internal first hidden layer. The algorithm then computes the outputs of all the neurons and passes the results on to the next layer, which computes the outputs and passes them on to the next layer, and so on until the final output layer's outputs are computed. This is the forward pass. The algorithm then uses the chain rule to calculate error contribution from each connection in the layer below until it reaches the input layer, where it measures the network's output error. This reverse pass efficiently measures error gradient over all connection weights by propagating backwards, and it then modifies the connection weights to lower the error. The step function is replaced with an activation function, where gradient descent makes some progress at each step, in order to make the algorithm function properly. The following are some of the activation functions applied in the study:

a) The hyperbolic tangent function, tanh (z) = 2 σ (2z) −1.
b) Rectified Linear Unit function, ReLU (z) =.$\max(0, z)$
c) Logistic function, σ (z) = 1 / (1 + exp (-z)).

The MLP requires only one output neuron to predict a single value in regression tasks. Any activation function can also be used for output neurons. To obtain positive values, the ReLU activation function was used, and to obtain the range of values, the hyperbolic tangent or logistic

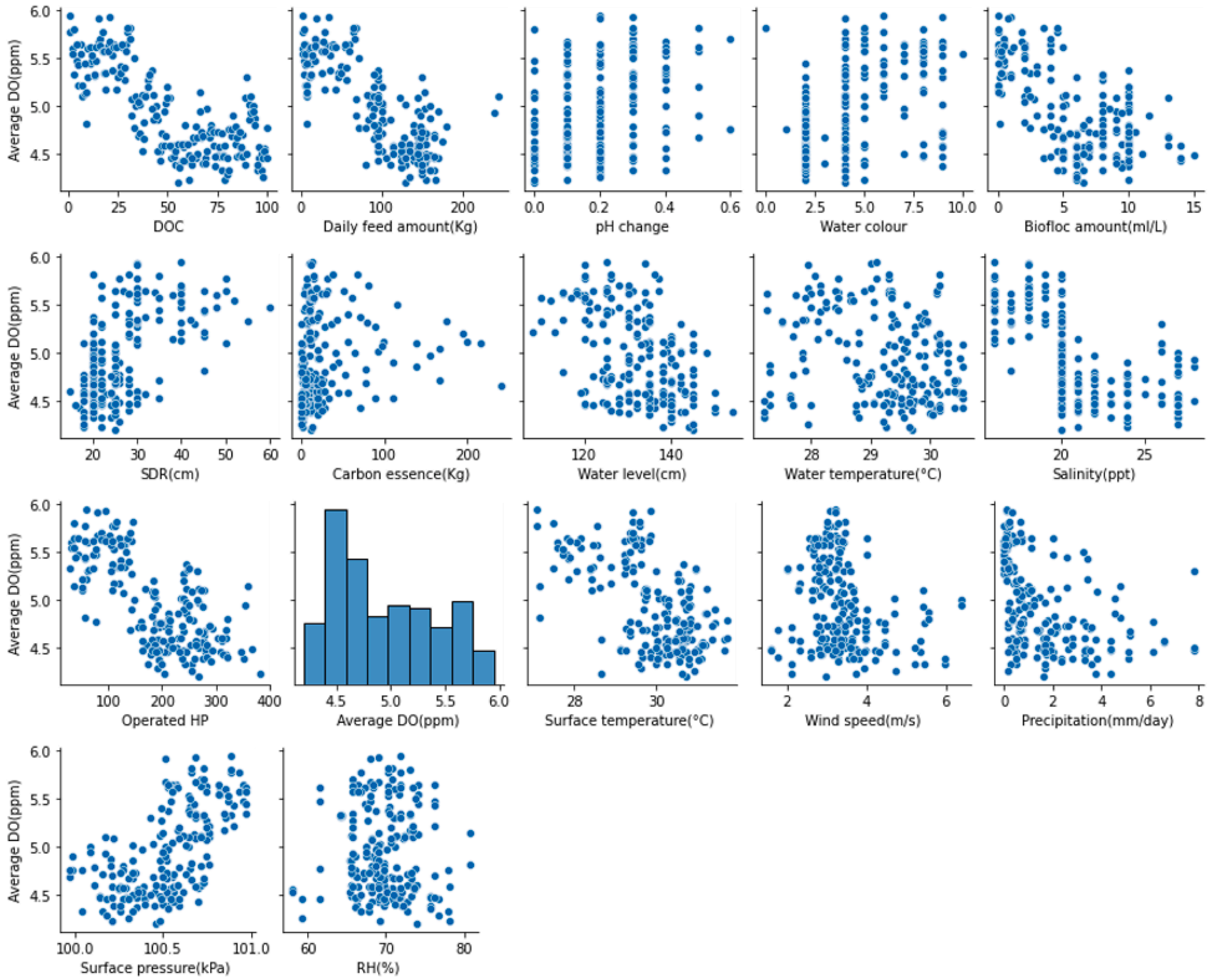

Fig. 11. Average DO vs other parameters.

function was used. The loss function to use during training includes mean squared error or mean absolute error if outliers are present. The typical regression MLP architecture is given in the Table 6. In the present study, Sequential API was employed which is the simplest kind of keras model for neural networks, composed of a single stack of layers, connected sequentially.

2.7. Predictive algorithm implementation

Using the previously mentioned feature selection methods and the selection criteria listed in Table7, feature subsets with features strongly related to selected target parameters were selected in each approach. The aforementioned machine learning algorithms were applied to all of the feature subsets using Jupyter notebook and Python 3.8 as the programming language, and the developed models were evaluated for accuracy and other performance metrics in subsequent operations. The total of 36 models was developed with two approaches and three machine learning algorithms on 12 different datasets. The flow chart showing overall work process is illustrated in Fig. 10.

2.8. Performance evaluation

The performance of the developed model was evaluated using common metrics viz., mean square error (MSE), root mean Square Error (RMSE), Mean Absolute Error (MAE), R^2 and adjusted R^2.

2.8.1. Mean square error (MSE) and Root mean square error (RMSE)

Mean square error is used to find the squared difference between actual and the predicted value of the target parameter. The larger the difference between predicted and the actual values, the larger the mean square error. Root mean square error is the square root of MSE, helps to find the difference between actual and the predicted values where the presence of outliers is exaggerated. The RMSE indicates how closely observed values aligned to the model's predicted value. The unit of MSE score is the squared of unit of target value whereas the unit of RMSE score match the unit of target value. The equations of MSE and RMSE are as follows:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}$$

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

where y_i = actual value, $\widehat{y}_i$ = predicted value and n = no. of instances.

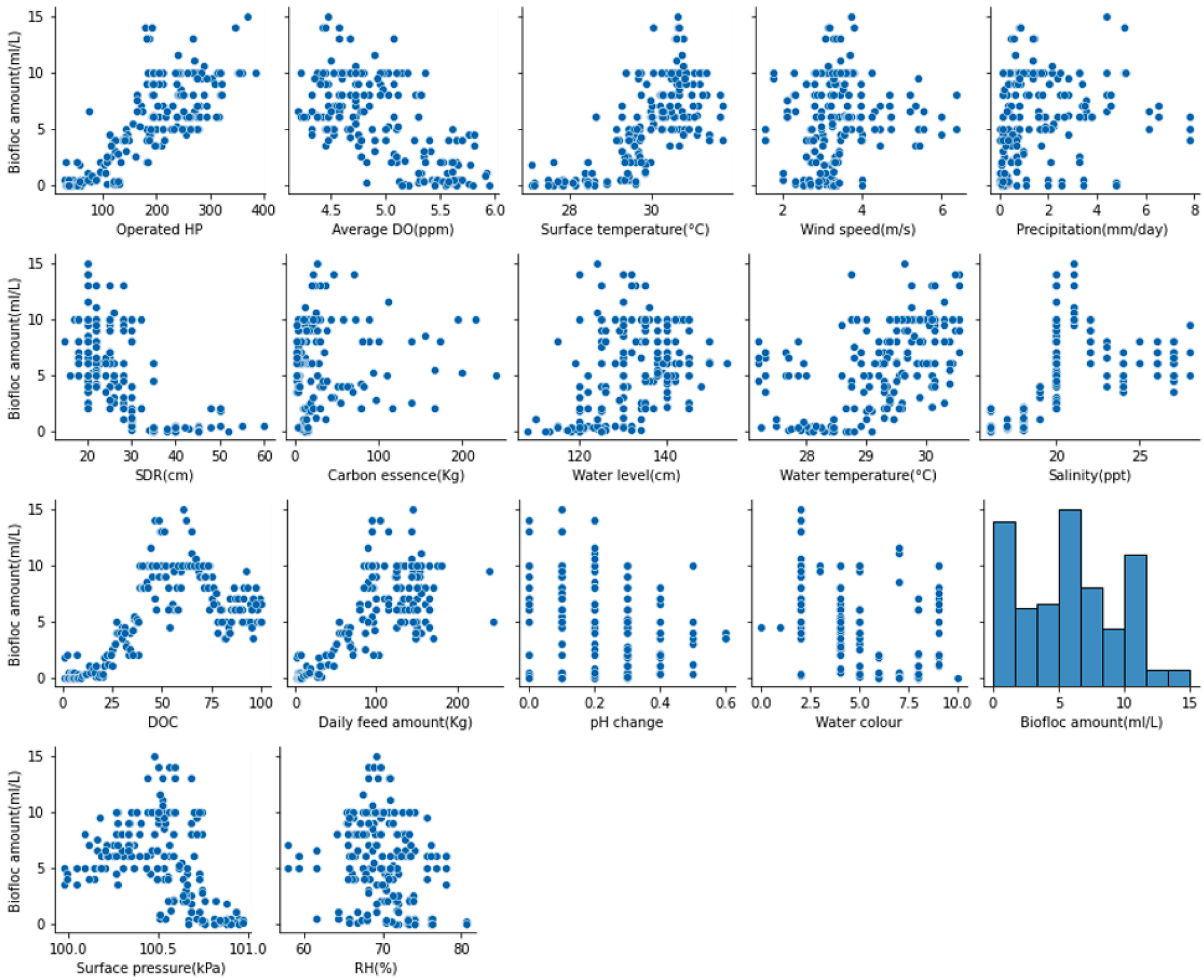

Fig. 12. Biofloc amount vs other parameters.

2.8.2. *Mean absolute error (MAE)*

Mean absolute error (MAE)is the mean of absolute difference between predicted and the actual value. Unlike RMSE and MSE, the MAE doesn't give more or less weights to individual difference but score increases linearly with increase in error. MAE is the best to measure the average magnitude of errors in prediction set without factoring in their direction. The equation of MAE is as follows:

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

where y_i = actual value, $\hat{y}_i$ = predicted value and n = no. of instances.

2.8.3. *R^2 and adjusted R^2 values*

Correlation was used to explain the strength of the relationship between the dependent and independent variables, while R^2 was used to explain the variance of the dependent and independent variables. It represents the goodness of fit of the regression model developed. R^2 was calculated by the following formula:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_{avg}\right)^2}{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

where SS_{res} = residual sum of squares (sum of squares of perpendicular distance between data points and average line),SS_{tot} = total sum of squares (sum of squares of perpendicular distance between data points and best fitted line).

R^2 value nearer to 1 represents better is the model. Adjusted R^2 (R^2_{adj}) is the modification of R^2in which its value increases when the new variable improves the model is added and its value decreases when the added new variable deteriorates the model. The R^2_{adj} is calculated by following formula:

$$R^2_{adj} = 1 - \left[\frac{\left(1 - R^2\right)(n - 1)}{n - k - 1}\right]$$

where n = no. of instances, k = no. of independent variables.

3. Results and discussion

The feature selection methods for the two approaches were carried out and feature subsets (Table 8) were selected according to the selection criteria discussed earlier. The results pertaining to exploratory analysis of datasets, model development and its indices, validations are discussed in this section.

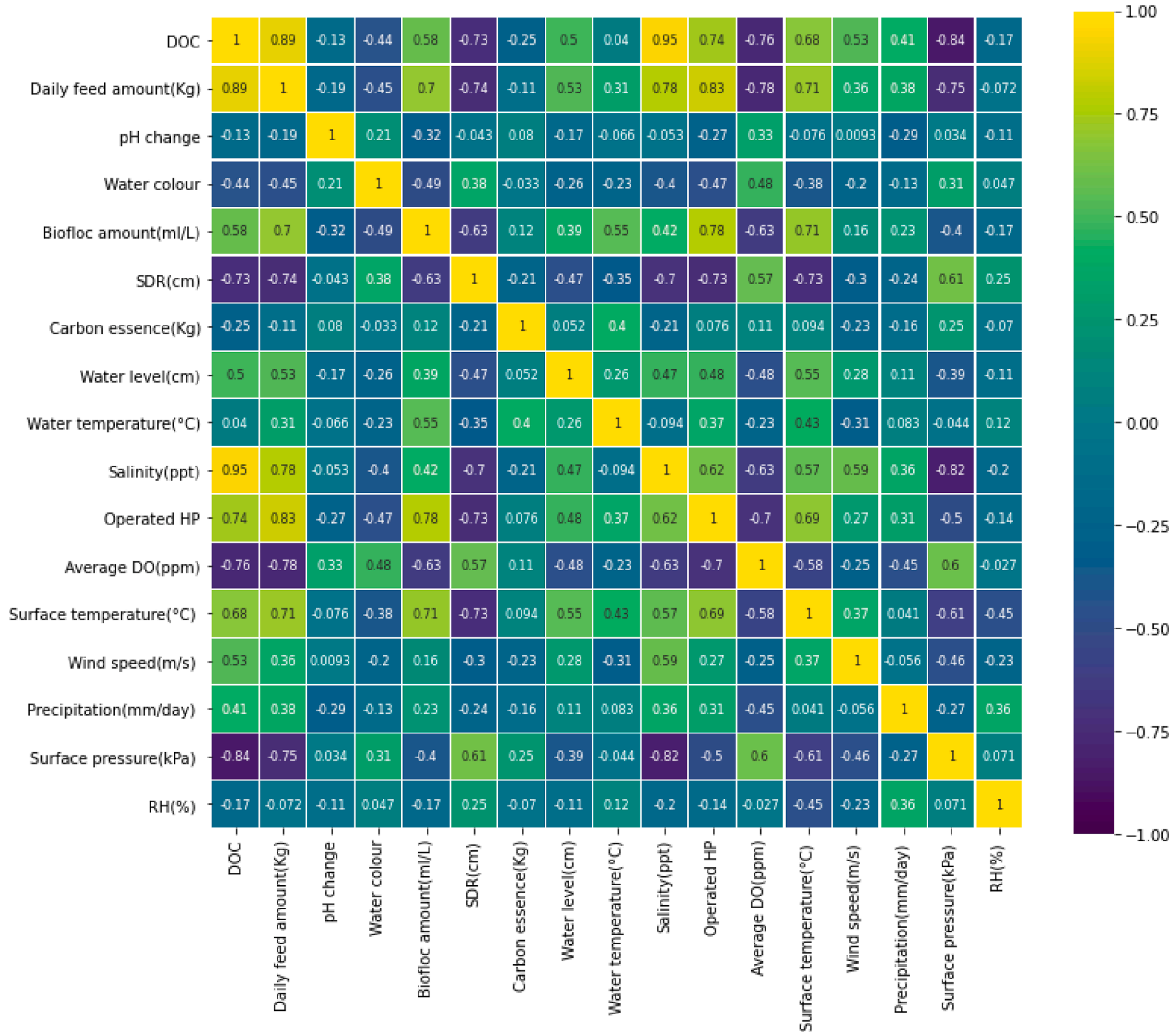

Fig. 13. Heatmap showing linear correlation between different parameters.

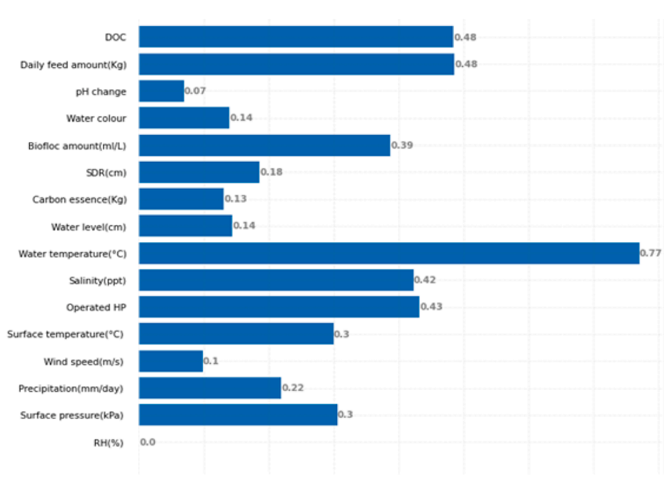

Fig. 14a. M.I between different input parameters(y) and target parameter Average DO (x).

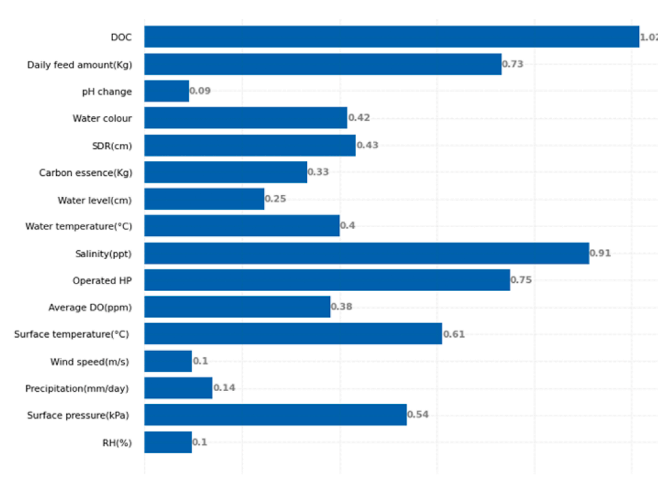

Fig. 14b. M.I between different input parameters(y) and target parameter Biofloc amount (x).

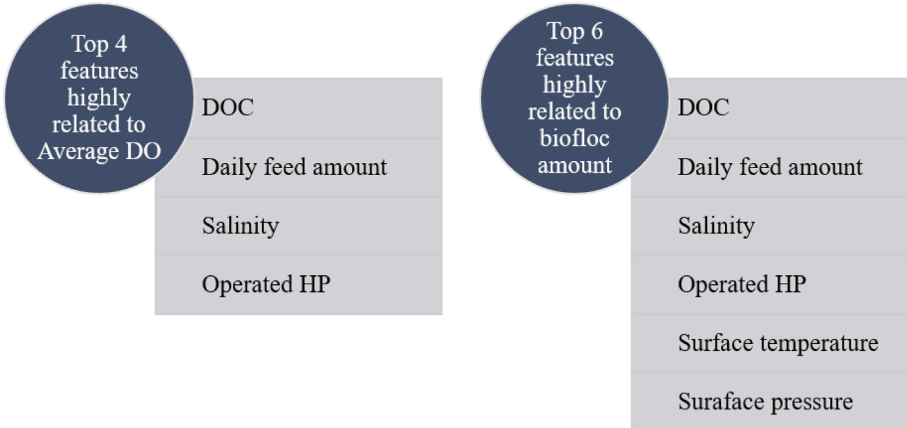

Fig. 15. Features selected in M.I selection methods.

Table 10
Results of Spearman's correlation test.

Spearman's correlation with Average DO		Spearman's correlation with Biofloc amount	
DOC	−0.735	DOC	0.587
Daily feed amount(Kg)	−0.744	Daily feed amount(Kg)	0.652
pH change	0.337	pH change	−0.341
Water colour	0.474	Water colour	−0.502
SDR(cm)	0.560	SDR(cm)	−0.576
Carbon essence(Kg)	0.266	Carbon essence(Kg)	0.158
Water level(cm)	−0.467	Water level(cm)	0.383
Water temperature(°C)	−0.227	Water temperature(°C)	0.574
Salinity(ppt)	−0.688	Salinity(ppt)	0.549
Operated HP	−0.288	Operated HP	−0.288
Biofloc amount(ml/L)	−0.596	Average DO(ppm)	−0.596
Surface temperature(°C)	−0.541	Surface temperature(°C)	0.715
Wind speed(m/s)	−0.289	Wind speed(m/s)	0.260
Precipitation(mm/day)	−0.569	Precipitation(mm/day)	0.372
Surface pressure(kPa)	0.601	Surface pressure(kPa)	−0.437
Relative humidity(%)	−0.126	Relative humidity(%)	−0.126

Table 11
R^2 and adjusted R^2 values of datasets.

TPA 1 (Average DO)			TPA 2 (Biofloc amount)		
Datasets	R^2 value	Adjusted R^2 value	Datasets	R^2 value	Adjusted R^2 value
D_{I1}	0.685	0.672	D_{II1}	0.728	0.718
D_{I2}	0.709	0.692	D_{II2}	0.738	0.723
D_{I3}	0.681	0.675	D_{II3}	0.718	0.710
D_{I4}	0.706	0.692	D_{II4}	0.720	0.707
D_{I5}	0.697	0.684	D_{II5}	0.722	0.710
D_{I6}	0.707	0.692	D_{II6}	0.749	0.733

3.1. Results of EDA

Exploratory data analysis helps to find out the pattern and hidden relation between the various parameters in the system. Some interesting relation and insights of different parameters (Table 9) were found in results of EDA using comparison with all the study parameters (Figs. 11 and 12).

3.2. Results of feature selection

3.2.1. Feature selection using Pearson's correlation selection method

It can be inferred from the correlation heat map (Fig. 13) that SDR and Surface pressure are positively correlated with the target parameter, average DO. Meanwhile, average DO is negatively correlated to DOC, daily feed amount, biofloc amount, salinity, operated HP, and surface temperature. It implies that farming using the biofloc culture technique is predicted to be complex, requiring the farmer to carefully maintain and adjust the important aforementioned parameters in a controlled level for proper dissolved oxygen level maintenance. Another target parameter, biofloc, has a positive linear correlation with the other parameters, like; operated HP, surface temperature, daily feed amount, DOC, and water temperature. It indicates that biofloc level can be easily controlled with proper adjustment of these positively correlated parameters, and SDR and average DO have a negative linear correlation, demonstrating its reliability and significance..

3.2.2. Feature selection using M.I selection method

For TPA 1, from the Fig. 14(a), top 4 features were selected and for TPA 2, from the Fig. 14b, top 6 features were selected which are shown in Fig. 15.

Most features were neglected in these subsets during M.I selection because they don't provide much information about the target variable; therefore, their removal allows for more robust model development with fewer but more relevant features. The variation of the result from the correlation result can be seen prominently because M.I measures the reduction in uncertainty. Salinity, for example, is found to be more related to the amount of biofloc, which was not found using Pearson's correlation test. It adds variety and uniqueness to the current study, allowing for better prediction testing and analysis.

3.2.3. Feature selection using Spearman's correlation selection methods

Table 10 shows that another factor influencing average DO, precipitation, is non-linearly correlated with 56 percent in a negative direction, which is a significant finding that was not found in other linear correlation tests. Moreover, water colour is found to be negatively related to the amount of biofloc. Hence, this method demonstrates variability and reliability in obtaining various datasets for analysis, model development, and suggesting a better framework.

3.3. Algorithm comparison and performance metrics:

Table 11 infers that adding meteorological features to feature subsets obtained using three feature selection methods significantly increases the R^2value. It means that the prediction would be better with the inclusion of meteorological parameters of the particular location. Furthermore, Adj R^2 value also increased in the dataset (D_{I2},D_{I4},D_{I6},D_{II2}, D_{II4},D_{II6})compared to (D_{I1},D_{I3},D_{I5},D_{II1},D_{II3},D_{II5}), indicating that adding features didn't degrade the model accuracy and performance but instead improved it.

The prediction accuracies of all three algorithm models are compared in TPA 1 (Table 12), and their order of performance based on

Table 12
Results of different models and their performance metrics.

Prediction test results for average dissolved oxygen												
Datasets	Random Forest regressor					AdaBoost regressor					Deep neural network	
	Score	Prediction accuracy	MSE	RMSE	MAE	Score	Prediction accuracy	MSE	RMSE	MAE	Prediction accuracy	MAE
D_{I1}	0.7405	0.9824	0.058	0.242	0.195	0.7299	0.9817	0.060	0.246	0.206	0.9782	0.135
D_{I2}	0.7381	0.9826	0.055	0.235	0.187	0.7368	0.9824	0.057	0.239	0.192	0.9771	0.137
D_{I3}	0.7448	0.9823	0.057	0.239	0.191	0.7164	0.9816	0.062	0.250	0.207	0.9766	0.155
D_{I4}	0.7443	0.9826	0.057	0.239	0.188	0.7126	0.9810	0.067	0.260	0.213	0.9792	0.144
D_{I5}	0.7415	0.9824	0.057	0.240	0.193	0.7348	0.9818	0.060	0.246	0.205	0.9809	0.138
D_{I6}	0.7322	0.9822	0.058	0.241	0.193	0.7438	0.9827	0.055	0.235	0.190	0.9797	0.123

Prediction test results for biofloc amount												
Datasets	Random Forest regressor					AdaBoost regressor					Deep neural network	
	Score	Prediction accuracy	MSE	RMSE	MAE	Score	Prediction accuracy	MSE	RMSE	MAE	Prediction accuracy	MAE
D_{II1}	0.8148	0.8981	2.401	1.549	1.170	0.7957	0.8592	2.642	1.625	1.346	0.8871	0.421
D_{II2}	0.8176	0.8992	2.373	1.540	1.122	0.8042	0.8752	2.541	1.594	1.310	0.8770	0.784
D_{II3}	0.8066	0.8931	2.525	1.589	1.174	0.8012	0.8784	2.566	1.601	1.294	0.9006	0.668
D_{II4}	0.8331	0.9013	2.348	1.532	1.120	0.7862	0.8719	2.774	1.665	1.330	0.8658	0.706
D_{II5}	0.8583	0.9018	1.840	1.356	0.998	0.7739	0.8758	2.953	1.718	1.371	0.9025	0.455
D_{II6}	0.8312	0.8979	2.204	1.484	1.067	0.7934	0.8684	2.695	1.641	1.322	0.8945	0.492

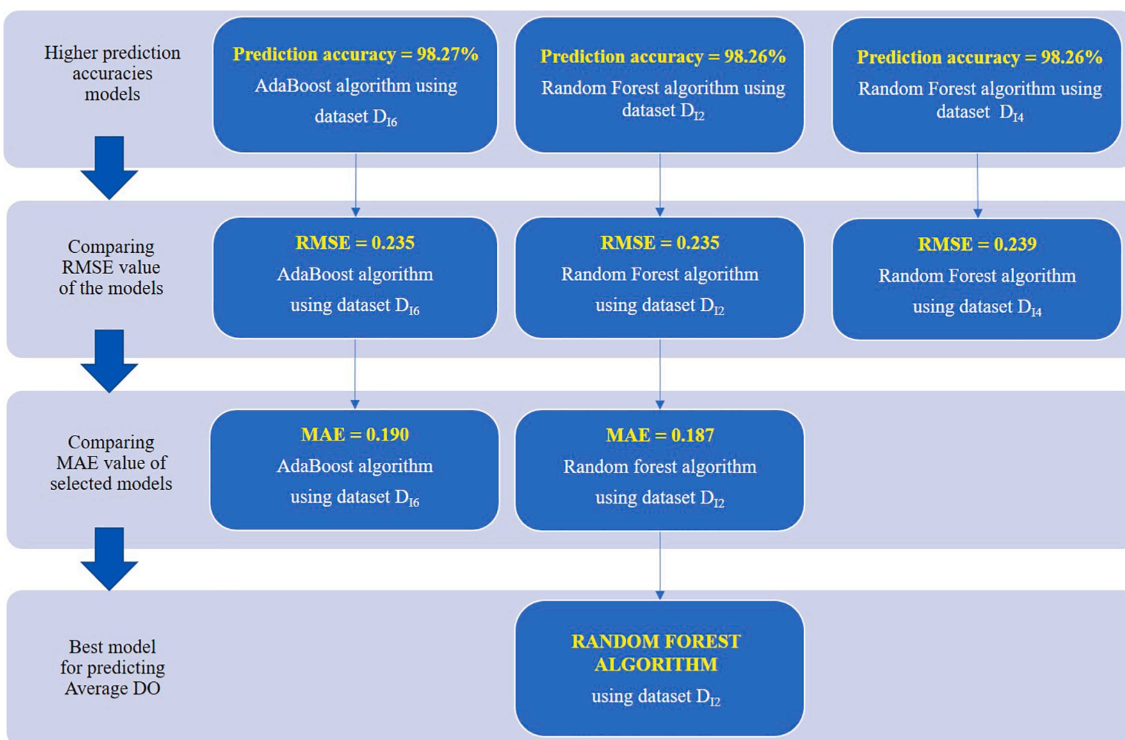

Fig. 16. Best model selection procedure for prediction of Average DO.

prediction accuracy is determined as follows:

RF > Adaboost > DNN

The best model selection procedure for prediction of average DO is illustrated in Fig. 16. For TPA 2 (Table 12), prediction accuracies of all the three algorithm models are compared and found their order of performance based on prediction accuracy as.

RF > DNN > Adaboost

The best model selection procedure for prediction of biofloc amount is illustrated in Fig. 17.In terms of model score and performance metrics, average DO prediction models outperform biofloc prediction models when comparing prediction accuracies of the two approaches. Hence, this proves the applicability and best suitability of ML in the intensive aquaculture system for DO prediction. Though, biofloc prediction provided results with a low score, it had a prediction accuracy of less than 91 percent. Hence, using ML to predict biofloc in a shrimp pond is not recommended, and it may need to be improved further by incorporating other non-culture parameters. While considering the ML algorithms, Random Forest and AdaBoost both suit well to predict average DO because they have provided best prediction accuracy of greater than 98 percent where, in particular Random Forest outperforms AdaBoost with lower errors. However, when considering the input parameters, feature subsets with and without climatic parameters produced the same result and had no special impacts on results with the inclusion of climatic parameters as one of the criteria for model development, but also feature subsets having climatic parameters improves the model accuracy to a considerable amount, though they have negligible effects on

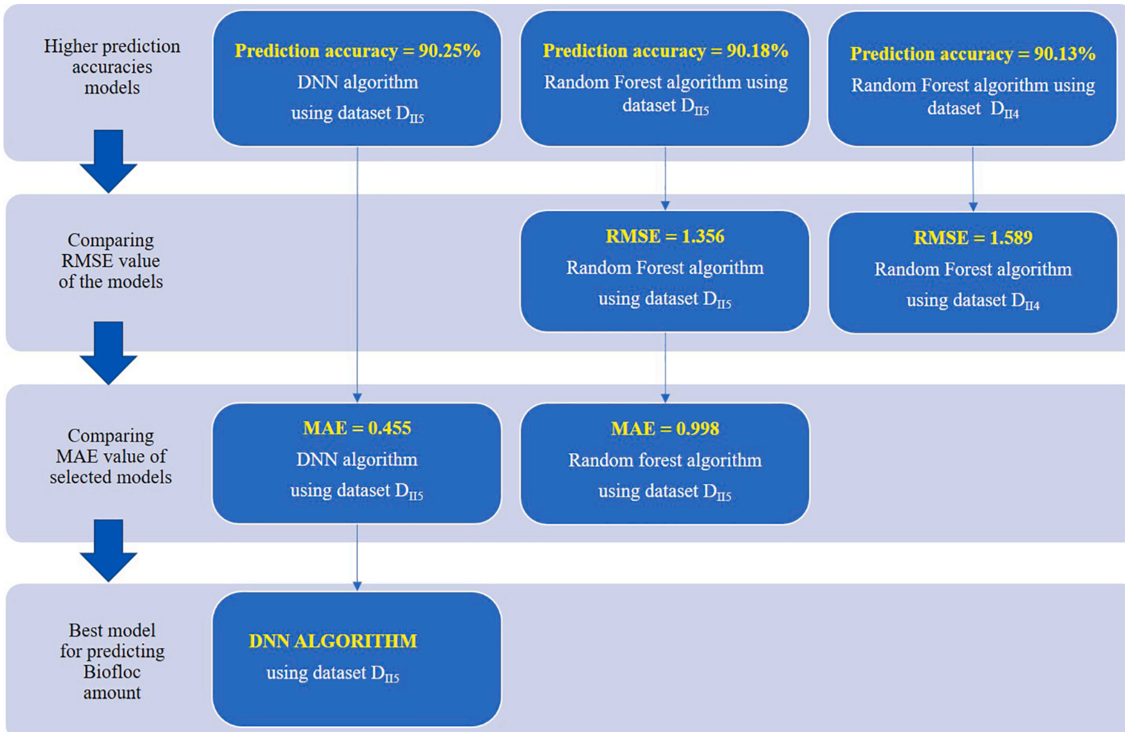

Fig. 17. Best model selection procedure for prediction of Biofloc amount.

Table 13
Comparison with other studies.

Algorithms used for prediction of DO	Performance metrics	Ayesha et al. (2022) (An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques)	Ren et al. (2020) (Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network)	Remarks
Bagging techniques	R^2	0.709	0.901	Justifiable value obtained
	RMSE	0.235	0.858	Reduced error value
	MAE	0.187	0.466	Reduced error value
Boosting techniques	R^2	0.709	0.926	Justifiable value obtained
	RMSE	0.235	0.860	Reduced error value
	MAE	0.190	0.541	Reduced error value
Neural network techniques	R^2	0.709	0.933	Justifiable value obtained
	MAE	0.138	0.255	Reduced error value

prediction accuracy.

Further, the successful DO models in the present study were compared with the similar study conducted in recirculating aquaculture systems using the same algorithms (Ren et al., 2020). The results were numerically compared and tabulated in Table 13. The Root mean square error and mean absolute error obtained for the developed models in the present study are significantly lower than those found in the literature. The biofloc prediction was carried out for the first time, and justifiable results were also obtained for the same. The model significance in the present study was justified both technically and statistically.

3.4. Suggested framework

The suggested framework (Fig. 18) was developed with all the understanding and the results obtained in the present study. The suggested framework can assist future researchers in the field of culture parameter prediction and forecasting in selecting the appropriate method for consideration and analysis, as well as advancing the research. The method of feature selection should be chosen according to the flow chart for different study parameters based on their count and varsity (Fig. 18). The framework – Flow chart then depicts the algorithm selection, accuracy evaluation, and return to feature selection step if low accuracy is obtained, as well as other critical procedures.

The framework (Fig. 18) begins with the number of parameters

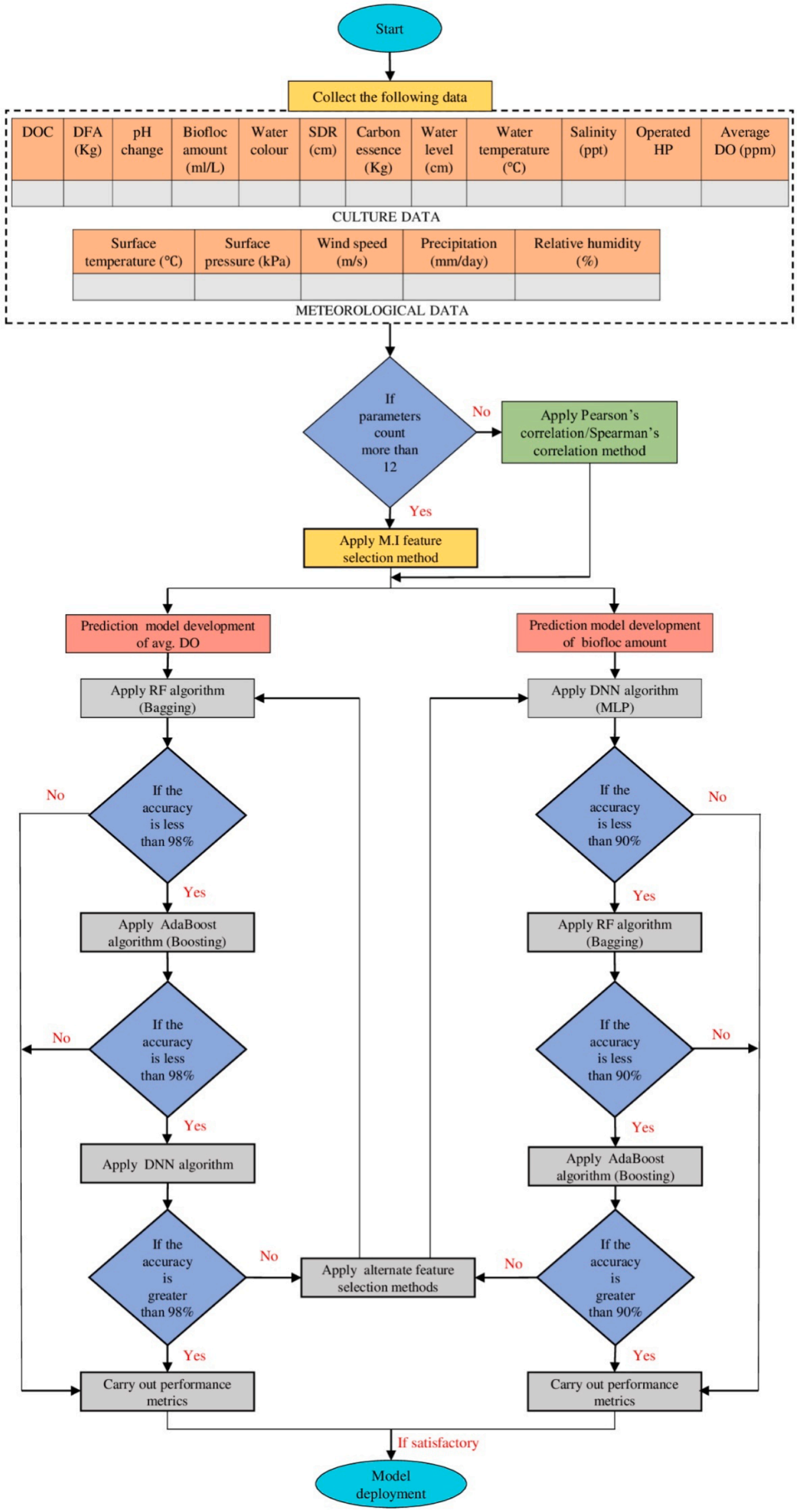

Fig. 18. Intelligent framework.

Table 14
Research Findings obtained from the study and its relevance with literature.

S. No	Subject to analysis (Factor/ Parameters)	Research findings	Similar results obtained by previous researchers
1	Farm parameters relations	The linear incremental relationship obtained from Temperature and biofloc amount helps in the adjustment of floc levels.	Huang et al., 2021
		Maintaining an optimal flocculation level for healthy growth of shrimp, lowers the operational costs of aeration (HP) and feeding (Kg of feed).	Azhar et al., 2016
2	Performance of future selection techniques	Mutual Information feature selection technique selects fewer but more important constraints.	Bennasar et al., 2015
		The inclusion of meteorological parameters in the dataset for the prediction analysis – Improve the model accuracy.	Exclusive finding
3	Algorithm performance	Random Forest algorithms are similar to other popular algorithms in terms of prediction accuracy, rather outperforms in model score and reduction in error.	Zhou et al., 2016
		Since, relation of biofloc amount with some other parameters viz., pH change, water color and salinity reveal its non-linearity. The obtained resultwas also suggested that non-linear algorithm namely DNN performs well for biofloc amount prediction.	Exclusive finding
4	Comparative finding – TPA	The Dissolved Oxygen (DO) predictive model is more robust and reliable than biofloc amount prediction model.	Exclusive finding

chosen for the prediction model development. Based on the observation from the study, if the total number of parameters for the study is less than 12, the Pearson correlation method can be used to extract the more decisive parameters that need to be included in model development. In other case if there are more than 12 parameters, MI feature selection helps in obtaining the most important cum essential parameters for the prediction model development. Then, the model development approaches to the application of algorithm stage, the dissolved oxygen prediction model needs to be tested first with random forest algorithm, since its accuracy and performance in terms of DO prediction is well and good as observed from the study, followed by AdaBoost and DNN. The DNN is the first choice in Biofloc amount prediction model, as the problem statement itself is a non-linear trend followed by random forest and Adaboost. The study compared different outcomes and research findings (Table 14) obtained upon development of a constructive intelligent framework to existing studies of similar results in order to obtain seamless prediction in the farming environment.

4. Summary and conclusion

The precise management of modern biofloc culture systems requires real-time prediction of dissolved oxygen and biofloc quantity. Therefore, choosing appropriate feature selection methods and suitable algorithms for the development of a precise prediction model for dissolved oxygen level and biofloc amount, the intelligent framework has been developed in the present study. Daily feed amount (Kg), days of culture (DOC), pH change, water colour, secchi disc reading (cm), carbon essence (Kg), Water level (cm), water temperature (°C), salinity (ppt), operated HP, surface temperature (°C), wind speed (m/s), precipitation (mm/day), relative humidity (%), surface pressure (kPa) were collected as raw data and data pre-processing techniques were performed to make data suitable for prediction. Three different feature selection techniques viz., Pearson's correlation, Mutual Information, Spearman's rank correlation feature selections were applied to find most decisive data subsets suitable for development of different predictive models. The models were developed using three popular algorithms viz., Random Forest, Adaboost and Deep neural network. The results of the study indicated that Random Forest algorithm performs well for the prediction of Average DO and DNN performs well for the prediction of biofloc amount according to the R^2 (0.7381), MAE (0.187), RMSE (0.235) values indicating that the prediction accuracy and reliability of the model are higher. Based on the study conducted, an intelligent framework has been developed in the present study which helps the developers to choose the appropriate data processing methods, feature selection methods and algorithm appropriate for their area of research. The developed framework also suggests the flawless pipeline for working with machine learning algorithm in aquaculture farming related studies. The same procedure on development of intelligent framework can be followed for the development of intelligent framework for any system of interest, would help in ultimate development of successful and reliable prediction tool for researchers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the College of Fisheries Engineering for kind support and motivation towards research. The authors also acknowledge HiTide sea farms, Sirkazhi, Tamil Nadu for resource assistance and support towards the research.

References

American Public Health Association (APHA). (2005). *Standard Method for Examination of Water and Wastewater* (21st edn.). AWWA, WPCF, Washington: APHA.

Abdel-Tawwab, M., Monier, M. N., Hoseinifar, S. H., & Faggio, C. (2019). Fish response to hypoxia stress: Growth, physiological, and immunological biomarkers. *Fish Physiology and Biochemistry, 45*, 997–1013.

Ale, L., Sheta, A., Li, L., Wang, Y., Zhang, N. (2019). *Deep learning based plant disease detection for smart agriculture*, in: 2019 IEEE Globecom Workshops (GC Wkshps). IEEE, pp. 1–6.

Allison, E. H., Perry, A. L., Badjeck, M.-C., Neil Adger, W., Brown, K., Conway, D., … Andrew, N. L. (2009). Vulnerability of national economies to the impacts of climate change on fisheries. *Fish and Fisheries, 10*, 173–196.

Amiri, F., Yousefi, M. R., Lucas, C., Shakery, A., & Yazdani, N. (2011). Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications, 34*, 1184–1199.

Avnimelech, Y. (2007). Feeding with microbial flocs by tilapia in minimal discharge bio-flocs technology ponds. *Aquaculture, 264*, 140–147.

Avnimelech, Y. (2012). *Biofloc-based aquaculture systems*. New Delhi, India: Aquaculture Production Systems.

Azhar, M. H., Supriyono, E., Nirmala, K., & Ekasari, J. (2016). Organic carbon source and C/N ratio affect inorganic nitrogen profile in the biofloc-based culture media of Pacific white shrimp (*Litopenaeus vannamei*) ILMU KELAUTAN. *Indonesian Journal of Marine Sciences, 21*(1), 23. https://doi.org/10.14710/ik.ijms.21.1.23-28.

Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using Joint Mutual Information Maximisation. *Expert Systems with Applications, 42*(22), 8520–8532. https://doi.org/10.1016/j.eswa.2015.07.007

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys & Tutorials, 16*, 303–336.

Bossier, P., & Ekasari, J. (2017). Biofloc technology application in aquaculture to support sustainable development goals. *Microbial Biotechnology, 10*, 1012–1016.

Boyd, C. E. (1998). Pond water aeration systems. *Aquacultural Engineering, 18*, 9–40.

Boyd, C. E., Torrans, E. L., & Tucker, C. S. (2018). Dissolved oxygen and aeration in ictalurid catfish aquaculture. *Journal of the World Aquaculture Society, 49*, 7–70.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123–140.

Burford, M. A., Thompson, P. J., McIntosh, R. P., Bauman, R. H., & Pearson, D. C. (2003). Nutrient and microbial dynamics in high-intensity, zero-exchange shrimp ponds in Belize. *Aquaculture, 219*, 393–411.

Cao, X., Liu, Y., Wang, J., Liu, C., & Duan, Q. (2020). Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network. *Aquacultural Engineering, 91*, Article 102122.

Chen, H., Huang, J. J., & McBean, E. (2020). Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random Forest and support vector regression, for a cabbage farmland. *Agricultural Water Management, 228*, Article 105923.

Cho, S.-B., Won, H.-H. (2003). *Machine learning in DNA microarray analysis for cancer classification*, in: Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 2003-Volume 19. pp. 189–198.

Datta, S. (2012). Management of water quality in intensive aquaculture. *Respiration, 6*, 602.

Drucker, H. (1997). Improving regressors using boosting techniques, in: *ICML. Citeseer*, (pp. 107–115).

Emerenciano, M. G. C., Martínez-Córdova, L. R., Martínez-Porchas, M., & Miranda-Baeza, A. (2017). Biofloc technology (BFT): A tool for water quality management in aquaculture. *Water Quality, 5*, 92–109.

FAO (2018). *The State of World Fisheries and Aquaculture 2018*-Meeting the sustainable development goals. Rome. Licence: CC by-NC-SA 3.0 IGO.

Guidelines for Regulating Coastal Aquaculture (2005). CAA (Coastal Aquaculture Authority), Chennai, India, (pp. 102-124).

Harley, C. D., Randall Hughes, A., Hultgren, K. M., Miner, B. G., Sorte, C. J., Thornber, C. S., … Williams, S. L. (2006). The impacts of climate change in coastal marine systems. *Ecology Letters, 9*, 228–241.

Hollender, J., van der Krol, D., Kornberger, L., Gierden, E., & Dott, W. (2002). Effect of different carbon sources on the enhanced biological phosphorus removal in a sequencing batch reactor. *World Journal of Microbiology and Biotechnology, 18*, 359–364.

Huang, H. H., Luo, T., Lei, Y. J., Kuang, W. Q., Zou, W. S., & Yang, P. H. (2021). Water quality, shrimp growth performance and bacterial community in a reusing-water biofloc system for nursery of *Penaeus vananmei* rearing under a low salinity condition. *Aquaculture Reports, 21*, Article 100894. https://doi.org/10.1016/j.aqrep.2021.100894

Hurley, J. V., & Boos, W. R. (2015). A global climatology of monsoon low-pressure systems. *Quarterly Journal of the Royal Meteorological Society, 141*, 1049–1064.

Kang, M.-S., Jung, Y.-G., & Jang, D.-H. (2017). A Study on the Search of Optimal Aquaculture farm condition based on Machine Learning. *The Journal of The Institute of Internet, Broadcasting and Communication, 17*, 135–140.

Kapetsky, J. M. (2000). Present applications and future needs of meteorological and climatological data in inland fisheries and aquaculture. *Agricultural and Forest Meteorology, 103*, 109–117.

khoshnevisYazdi, S., Shakouri, B. (2010). The effects of climate change on aquaculture. *International Journal of Environmental Science and Development 1*, 378.

Kim, S.-K., Pang, Z., Seo, H.-C., Cho, Y.-R., Samocha, T., & Jang, I.-K. (2014). Effect of bioflocs on growth and immune activity of Pacific white shrimp, *Litopenaeus vannamei* postlarvae. *Aquaculture Research, 45*, 362–371.

Krishna, C., & Van Loosdrecht, M. C. (1999). Effect of temperature on storage polymers and settleability of activated sludge. *Water Research, 33*, 2374–2382.

Lara, G., Hostins, B., Bezerra, A., Poersch, L., & Wasielesky, W., Jr (2017). The effects of different feeding rates and re-feeding of *Litopenaeus vannamei* in a biofloc culture system. *Aquacultural Engineering, 77*, 20–26.

Lemonnier, H., Bernard, E., Boglio, E., Goarant, C., & Cochard, J.-C. (2004). Influence of sediment characteristics on shrimp physiology: pH as principal effect. *Aquaculture, 240*, 297–312.

Li, E., Wang, X., Chen, K., Xu, C., Qin, J. G., & Chen, L. (2017). Physiological change and nutritional requirement of Pacific white shrimp *Litopenaeus vannamei* at low salinity. *Reviews in Aquaculture, 9*, 57–75.

Li, X., Ai, J., Lin, C., & Guan, H. (2018). Prediction model of dissolved oxygen in ponds based on ELM neural network. In *in: IOP Conference Series: Earth and Environmental Science* (p. 22003).

Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Mathematical and Computer Modelling, 58*, 458–465.

Liu, Y., Mu, Y., Chen, K., Li, Y., & Guo, J. (2020). Daily activity feature selection in smart homes based on Pearson correlation coefficient. *Neural Processing Letters*, 1–17.

Liu, Z.-T., Wu, M., Cao, W.-H., Mao, J.-W., Xu, J.-P., & Tan, G.-Z. (2018). Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing, 273*, 271–280.

Long, L., Yang, J., Li, Y., Guan, C., & Wu, F. (2015). Effect of biofloc technology on growth, digestive enzyme activity, hematology, and immune response of genetically improved farmed tilapia (Oreochromis niloticus). *Aquaculture, 448*, 135–141.

Lynch, A. J., Elliott, V., Phang, S. C., Claussen, J. E., Harrison, I., Murchie, K. J., … Stokes, G. L. (2020). Inland fish and fisheries integral to achieving the Sustainable Development Goals. *Nature Sustainability, 3*, 579–587.

Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019). House price prediction using regression techniques: A comparative study. In *in: 2019 International Conference on Smart Structures and Systems (ICSSS)* (pp. 1–5). IEEE.

Mallya, Y. J. (2007). *The effects of dissolved oxygen on fish growth in aquaculture*. Final Project: The United Nations University Fisheries Training Programme.

Mohan, P., & Patil, K. K. (2018). Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application. *International Journal of Intelligent Engineering and Systems, 11*, 167–176.

Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation, 18*, 399–406.

Paul, D., Su, R., Romain, M., Sébastien, V., Pierre, V., & Isabelle, G. (2017). Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. *Computerized Medical Imaging and Graphics, 60*, 42–49.

Piedrahita, R. H. (2003). Reducing the potential environmental impact of tank aquaculture effluents through intensification and recirculation. *Aquaculture, 226*, 35–44.

Pollock, M. S., Clarke, L. M. J., & Dubé, M. G. (2007). The effects of hypoxia on fishes: From ecological relevance to physiological effects. *Environmental Reviews, 15*, 1–14.

Ponce-Palafox, J. T., Pavia, Á. A., López, D. G. M., Arredondo-Figueroa, J. L., Lango-Reynoso, F., del Refugio Castañeda-Chávez, M., … Castillo-Vargasmachuca, S. G. (2019). Response surface analysis of temperature-salinity interaction effects on water quality, growth and survival of shrimp *Penaeus vannamei* postlarvae raised in biofloc intensive nursery production. *Aquaculture, 503*, 312–321.

Rahman, A. (2013). Benthic Habitat Mapping from Seabed Images using Ensemble of Color, Texture, and Edge Features. *International Journal of Computational Intelligence Systems, 6*, 1072–1081.

Rahman, A., & D'Este, C. (2015). Shellfish farm closure prediction and cause identification using machine learning methods. *Computers and Electronics in Agriculture, 110*, 241–248.

Raman, V., Perumal, S., Navaratnam, S., & Fazilah, S. (2016). Computer assisted counter system for larvae and juvenile fish in malaysian fishing hatcheries by machine learning approach. *Journal of Computers, 11*, 423–431.

Rangkuti, F. R. S., Fauzi, M. A., Sari, Y. A., & Sari, E. D. L. (2018). Sentiment analysis on movie reviews using ensemble features and pearson correlation based feature selection. In *in: 2018 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 88–91). IEEE.

Ray, A.J., and Lotz, J.M. (2018). *Comparing salinities in intensive, commercial-scale biofloc for Pacific white shrimp*. Global Aquaculture advocate. Retrieved from https://www.globalseafood.org/advocate/comparing-salinities-in-intensive-commercial-scale-biofloc-for-pacific-white-shrimp/?headlessPrint=AA Accessed September 7, 2021.

Razman, M. A. M., Susto, G. A., Cenedese, A., Majeed, A. P. A., Musa, R. M., Ghani, A. S. A., … Mukai, Y. (2019). Hunger classification of Lates calcarifer by means of an automated feeder and image processing. *Computers and Electronics in Agriculture, 163*, Article 104883.

Ren, Q., Wang, X., Li, W., Wei, Y., & An, D. (2020). Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. *Aquacultural Engineering, 90*, Article 102085. https://doi.org/10.1016/j.aquaeng.2020.102085

Rossi, F., Benso, A., Di Carlo, S., Politano, G., Savino, A., & Acutis, P. L. (2016). FishAPP: A mobile App to detect fish falsification through image processing and machine learning techniques. In *in: 2016 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). IEEE* (pp. 1–6).

Rutkayová, J., Vácha, F., Maršálek, M., Beneš, K., Civišová, H., Horká, P., … Šulista, M. (2018). Fish stock losses due to extreme floods–findings from pond-based aquaculture in the Czech Republic. *Journal of Flood Risk Management, 11*, 351–359.

Samocha, T. M., Patnaik, S., Speed, M., Ali, A.-M., Burger, J. M., Almeida, R. V., … Brock, D. L. (2007). Use of molasses as carbon source in limited discharge nursery and grow-out systems for *Litopenaeus vannamei*. *Aquacultural Engineering, 36*, 184–191.

Schveitzer, R., Arantes, R., Costódio, P.F.S., do Espírito Santo, C.M., Arana, L.V., Seiffert, W.Q., Andreatta, E.R. (2013). Effect of different biofloc levels on microbial activity, water quality and performance of *Litopenaeus vannamei* in a tank system operated with no water exchange. *Aquacultural Engineering 56*, 59–70.

Shahriar, M.S., Rahman, A. 2013. *Spatial-temporal prediction of algal bloom*, in: (2013) Ninth International Conference on Natural Computation (ICNC). IEEE, pp. 973–977.

Shahriar, M. S., Rahman, A., & McCulloch, J. (2014). Predicting shellfish farm closures using time series classification for aquaculture decision support. *Computers and Electronics in Agriculture, 102*, 85–97.

Singh, B., Sihag, P., & Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment, 3*, 999–1004.

Sirsat, M. S., Cernadas, E., Fernández-Delgado, M., & Khan, R. (2017). Classification of agricultural soil parameters in India. *Computers and Electronics in Agriculture, 135*, 269–279.

Spicer, P. T., & Pratsinis, S. E. (1996). Shear-induced flocculation: The evolution of floc structure and the shape of the size distribution at steady state. *Water Research, 30*, 1049–1056.

Suita, S. M., Braga, A., Ballester, E., Cardozo, A. P., Abreu, P. C., & Wasielesky, W. (2016). Contribution of bioflocs to the culture of *Litopenaeus vannamei* post-larvae determined using stable isotopes. *Aquaculture International, 24*, 1473–1487.

Tacon, A. G. J., Cody, J. J., Conquest, L. D., Divakaran, S., Forster, I. P., & Decamp, O. E. (2002). Effect of culture system on the nutrition and growth performance of Pacific white shrimp *Litopenaeus vannamei* (Boone) fed different diets. *Aquaculture Nutrition, 8*, 121–137.

Temrat, P., Jiraraksopakun, Y., Bhatranand, A., Wea-asae, K. (2018). *Suitable feature selection for OSA classification based on snoring sounds*, in: 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE, (pp. 1–4).

Thompson, F. L., Abreu, P. C., & Wasielesky, W. (2002). Importance of biofilm for water quality and nourishment in intensive shrimp culture. *Aquaculture, 203*, 263–278.

Timmons, M. B., Ebeling, J. M., Wheaton, F. W., Summerfelt, S., & Vinci, B. J. (2002). *Recirculating aquaculture systems*. Ithaca (New York): Northeastern Regional Aquaculture Center.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research, 47*, 123–138.

Vinatea, L., Gálvez, A. O., Browdy, C. L., Stokes, A., Venero, J., Haveman, J., … Leffler, J. W. (2010). Photosynthesis, water respiration and growth performance of *Litopenaeus vannamei* in a super-intensive raceway culture with zero water exchange: Interaction of water quality variables. *Aquacultural Engineering, 42*, 17–24.

Wasielesky, W., Jr, Atwood, H., Stokes, A., & Browdy, C. L. (2006). Effect of natural production in a zero exchange suspended microbial floc based super-intensive culture system for white shrimp *Litopenaeus vannamei*. *Aquaculture, 258*, 396–403.

Wen, J., Zhang, X., Xu, Y., Li, Z., & Liu, L. (2009). Comparison of AdaBoost and logistic regression for detecting colorectal cancer patients with synchronous liver metastasis. In *in: 2009 International Conference on Biomedical and Pharmaceutical Engineering* (pp. 1–6). IEEE.

Wilén, B.-M., Jin, B., & Lant, P. (2003). The influence of key chemical constituents in activated sludge on surface and flocculating properties. *Water Research, 37*, 2127–2139.

Yan, J. X., Yu, L. J., Mao, W. W., & Cao, S. Q. (2014). Study on prediction model of dissolved oxygen about water quality monitoring system based on BP neural network. *Advanced Materials Research, 912–914*, 1407–1411.

You, J., Liu, W., & Lee, J. (2020). A DNN-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture, 178*, Article 105750.

Zhao, J., Li, Y., Zhang, F., Zhu, S., Liu, Y., Lu, H., & Ye, Z. (2018). Semi-supervised learning-based live fish identification in aquaculture using modified deep convolutional generative adversarial networks. *Transactions of the ASABE, 61*, 699–710.

Zhou, C., Xu, D., Chen, L., Zhang, S., Sun, C., Yang, X., & Wang, Y. (2019). Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture, 507*, 457–465.

Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal, 4*, 212–219.