

## Article

# Design of an Intelligent Variable-Flow Recirculating Aquaculture System Based on Machine Learning Methods

Fudi Chen <sup>1,2,†</sup>, Yishuai Du <sup>1,2,†</sup>, Tianlong Qiu <sup>1,2</sup>, Zhe Xu <sup>3</sup>, Li Zhou <sup>1,2</sup>, Jianping Xu <sup>1,2</sup>, Ming Sun <sup>1,2,4</sup>, Ye Li <sup>1,2</sup> and Jianming Sun <sup>1,2,3,4,\*</sup>

- <sup>1</sup> CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; chenfudi@qdio.ac.cn (F.C.); duyishuai@qdio.ac.cn (Y.D.); oceanman@qdio.ac.cn (T.Q.); zhouli@qdio.ac.cn (L.Z.); xujianping@qdio.ac.cn (J.X.); sunming1@qdio.ac.cn (M.S.); liye@qdio.ac.cn (Y.L.)
  - <sup>2</sup> Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China
  - <sup>3</sup> Dalian Huixin Titanium Equipment Development Co., Ltd., Dalian 116039, China; Xuzhe\_DLHX@hotmail.com
  - <sup>4</sup> Dalian Key Laboratory of Conservation of Fishery Resources, Liaoning Province Key Laboratory of Marine Biological Resources and Ecology, Liaoning Ocean and Fisheries Science Research Institute, Dalian 116023, China
- \* Correspondence: jianming\_sun@outlook.com; Tel.: +86-532-82898031  
† These authors contributed equally to this work.

**Featured Application:** The proposed classification models could be adapted to develop a recirculating aquaculture system with continuous variable-flow control technology.



**Citation:** Chen, F.; Du, Y.; Qiu, T.; Xu, Z.; Zhou, L.; Xu, J.; Sun, M.; Li, Y.; Sun, J. Design of an Intelligent Variable-Flow Recirculating Aquaculture System Based on Machine Learning Methods. *Appl. Sci.* **2021**, *11*, 6546. <https://doi.org/10.3390/app11146546>

Academic Editor: Manuel Armada

Received: 12 June 2021

Accepted: 14 July 2021

Published: 16 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** A recirculating aquaculture system (RAS) can reduce water and land requirements for intensive aquaculture production. However, a traditional RAS uses a fixed circulation flow rate for water treatment. In general, the water in an RAS is highly turbid only when the animals are fed and when they excrete. Therefore, RAS water quality regulation technology based on process control is proposed in this paper. The intelligent variable-flow RAS was designed based on the circulating pump-drum filter linkage working model. Machine learning methods were introduced to develop the intelligent regulation model to maintain a clean and stable water environment. Results showed that the long short-term memory network performed with the highest accuracy (training set 100%, test set 96.84%) and F1-score (training 100%, test 93.83%) among artificial neural networks. Optimization methods including grid search, cuckoo search, linear squares, and gene algorithm were proposed to improve the classification ability of support vector machine models. Results showed that all support vector machine models passed cross-validation and could meet accuracy standards. In summary, the gene algorithm support vector machine model (accuracy: training 100%, test 98.95%; F1-score: training 100%, test 99.17%) is suitable as an optimal variable-flow regulation model for an intelligent variable-flow RAS.

**Keywords:** recirculating aquaculture system; variable-flow regulation model; circulating pump-drum filter linkage working technique; machine learning methods; gene algorithm support vector machine



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With global economic growth, consumer demand for seafood products is also increasing. However, fishery productivity is facing a massive challenge of declining resources due to environmental pollution and overfishing [1]. The recirculating aquaculture mode is an effective solution to maintain the supply of seafood products and support the modern and sustainable development of the aquaculture industry while decreasing ecological impact [2]. A recirculating aquaculture system (RAS) can offer a high degree of environmental control

and uses various technologies to carry out physical filtration, biofiltration, and disinfection for water recycling [3].

The core of an RAS is the water treatment system, which mainly includes micro-screen drum filters, biofilters, oxidation devices, and disinfection devices [4]. Suspended solids removal is a critical part of water treatment in the RAS. Suspended solid particles are composed mainly of feces, residual feed, and bacterial flocs [5–7]. Feed is the main source of suspended solids in the system, and studies have shown that 25% of feed is converted into suspended solids in an RAS [8]. Suspended solid particles have been proven to be the leading cause of high turbidity in aquaculture water, which can cause stress reactions and endanger the health of aquatic animals [9]. As residence time increases, the suspended solids block the breeding facilities and increase chemical oxygen demand. Organic solid waste can be mineralized and decomposed to increase ammonia and nitrite concentrations and increase the load on the nitrification function of the biofilter [10]. Suspended solids removal devices in an RAS can be roughly classified according to the particle size of the suspended matter: sedimentation separation devices, micro-mesh filtration devices, foam fractionators, and ozone generators. The micro-screen drum filter, which is a physical filter device widely used in RASs, has the characteristics of strong adaptability, minimal floor space, and a high level of automation [11]. In a drum filter, the screen is fixed on a rotating drum frame on the horizontal axis and partially submerged in water; water flows into the drum and radially through the straining cloth, which captures fine particles with a suitable mesh size [12]. The micro-screen is the central working part of the drum filter, and the mesh number can directly affect filtration performance. Gravdal Arve et al. [13] reported that the removal rate of particles larger than 60  $\mu\text{m}$  by the drum filter could reach more than 68%. Su et al. [14] found that the removal rate rapidly increased when the mesh number was increased from 150 to 200. The effect was apparent when the screen mesh was 200; the TSS removal rate reached 54.90%. Generally, 200 mesh is the principal mesh size used, as it is the one with the most outstanding technical and economic advantages [11].

A high-power centrifugal pump and an oversized drum filter are generally used to ensure sufficient circulation flow and filtration ability in an RAS [15]. The water in a traditional fixed-flow RAS is highly turbid when the breeding animals are fed and when they defecate. However, at other times, the water is relatively clean and does not require high-power pumps to recirculate it, resulting in wasting resources. Compared with the traditional fixed-flow RAS, the variable-flow RAS can increase the total water circulation to accelerate the water treatment process when organic particles increase, and the ammonia and nitrite then can be eliminated from the source [16]. In addition, the variable-flow RAS consumes a low amount of electricity when the water is relatively clean. However, manual operation is often used to adjust the circulation pump frequency to determine the appropriate total water circulation in the variable-flow RAS. The manual operation experience may cause the water treatment efficiency to not match the actual situation, resulting in insufficient water processing efficiency or waste of electricity. Hence, an intelligent variable-flow RAS for culturing *Litopenaeus vannamei* was developed in the present study. Machine learning, which has emerged with Big Data technologies and created new opportunities in multidisciplinary aquaculture, was used to develop the intelligent variable-flow model. Currently, machine learning is applied in related fields, including environmental assessment, water management, animal welfare, disease detection, feeding control, and species recognition [17–23]. More data-intensive machine learning approaches have been reported, but model- and technology-intensive approaches have been infrequent [24,25]. For industrial control in recirculating aquaculture, in particular, there is an urgent need to apply machine learning models to improve instrument efficiency and promote the development of intelligent equipment applications.

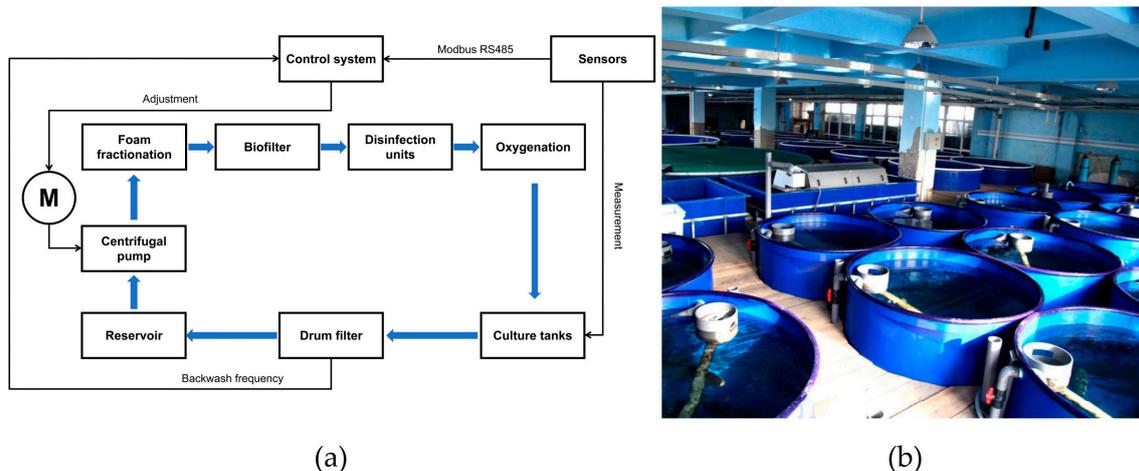
The primary purpose of the present study was to develop the circulating pump-drum filter linkage working technique using machine learning methods. Water quality indicators and the backwash frequency of the drum filter were used as primary indicators

in developing a variable-flow model. An intelligent variable-flow RAS can rapidly remove suspended solids and reduce ammonia and nitrite generation from the source.

## 2. Materials and Methods

### 2.1. Experimental RAS

The experimental RAS used the recirculating aquaculture system of Dalian Huixin Titanium Equipment Development Co., Ltd. (Dalian city, China) for breeding *L. vannamei*. Figure 1a shows the schematic of the experimental RAS control system. The control system collected the water quality indicators by connecting them with the sensors. Water quality changes can be monitored in real time, and the centrifugal pump was controlled by variable-frequency operation using a flow regulation model based on machine learning. The variable-flow circulation caused different trends in the drum filter backwash frequency during the unit period (0.5 h). The water quality indicators were used to train the regulation strategy model for variable-flow circulation. The types of water treatment equipment included biofilters, a micro-screen drum filter, an ultraviolet generator, ozone generators, foam fractionators, and oxygenation cones. Figure 1b shows the actual indoor workshop. The RAS contained 10 circular FRP tanks with a diameter of 1.8 m and a depth of 1.4 m, with a total water volume of 35 m<sup>3</sup>. Shrimp were fed five times a day during the culture period with a 36% protein commercial feed (Dale 2# shrimp commercial feeds, Dale, Inc., Yantai, China). During the early stage of shrimp culture, the amount of feed accounted for 5–8% of the total biomass of shrimp. The amount of feed was reduced over time and accounted for 3.7–5% of the total biomass by the end of the culture process. The whole culture process lasted for 90 days, with a culture density of 800 individuals/m<sup>3</sup> and a final yield of 525 kg of shrimp.



**Figure 1.** The experimental recirculating aquaculture system: (a) the schematic of the experimental RAS control process; (b) the RAS in Dalian Huixin Titanium Equipment Development Co., Ltd.

### 2.2. Variable-Flow Experiment Design

Turbidity (NTU) is mainly influenced by water flow fluctuations and can only reflect the instantaneous transparency of the water body. This study proposes a technique for detecting turbidity in an RAS based on a micro-screen drum filter. The backwash frequency of the drum filter within a unit period (0.5 h) was used to represent overall RAS turbidity, and the variable-flow regulation model was constructed using the backwash frequency and various water quality data. The variable-flow regulation model can determine the operating frequency of the centrifugal pump for the next period using real-time data from the current period. The intelligent variable-flow RAS technology is implemented by controlling the RAS circulation rate by changing the circulating pump flow rate. The primary purpose of the variable-flow RAS is to implement a linkage control technology to

model the relationship between the micro-screen drum filter backwash frequency and the circulation flow rate.

The total flow rate of the circulating pump was set to three levels: 55, 65, and 75 m<sup>3</sup>/h. The circulation rate was operated with a cycle of 24 h. A cycle started with a circulation rate of 55 m<sup>3</sup>/h and was adjusted to 65 m<sup>3</sup>/h after an interval of 24 h and then to 75 m<sup>3</sup>/h after the same interval (24 h). The drum filter controller collected backwash data every 0.5 h. Turbidity sensors were placed at the main return pipeline to monitor and record overall RAS water turbidity. Water quality indicators, including water temperature (T), dissolved oxygen (DO), pH, and salinity, were measured by sensors in real time using YSI ProPlus portable sensors. Total suspended solids (TSS), total ammonia nitrogen (TAN), and nitrite nitrogen (NO<sub>2</sub>-N) were measured daily with a Palintest 7500 water quality analyzer.

The circulating pump was set to three circulating levels: slow (55 m<sup>3</sup>/h), medium (65 m<sup>3</sup>/h), and fast (75 m<sup>3</sup>/h). In the variable-flow RAS, the circulation rate was maintained at a medium level, and the control system read water quality indicators and backwash times from sensors at every unit period. The circulation rate for the next period could be adjusted to slow or fast levels. The circulation adjustment process could be operated in two ways: upshift and downshift. In the drum filter controller program, the backwash frequency was recorded for 48 periods in a day, using 0.5 h as a period. The circulating pump was utilized to determine the upshift/downshift for the next period by reading the current water quality sensors, current backwash frequency, and current circulating level. A water gauge controlled the drum filter backwash frequency; the backwash frequency reflects water turbidity in the RAS. Downshifts (−1) and upshifts (+1) of circulating pump frequency were used as indicators of circulation levels. The water quality indicators, current circulating pump frequency, and the drum filter backwash frequency were chosen as independent variables, and the downshifts (−1)/upshifts (+1) data were considered as the dependent variable. As the whole culture process lasted for 90 days in the RAS, the total circulation rate was set to 55 m<sup>3</sup>/h for the first 30 days, 65 m<sup>3</sup>/h for the middle 30 days, and 75 m<sup>3</sup>/h for the last 30 days. Establishing a variable-flow circulation strategy was the core task of the experiment, and therefore the circulation rate regulation model was constructed using the optimal classification model based on machine learning to control the variable-flow circulation rate in the RAS.

As shown in Figure 2, the drum filter controller was used to collect the backwash frequency, circulation flow rate, and water quality data that were then uploaded to the industrial PC through the RS485 protocol. The embedded system was connected to the industrial computer. The dataset was processed with the optimal machine learning model in the industrial computer to regulate pump frequency for the next period and feed it back to the embedded system, so that the RAS circulation flow rate could be regulated intelligently.

### 2.3. Machine Learning Methods

#### 2.3.1. Artificial Neural Networks (ANNs)

ANNs are statistical learning algorithms that possess prediction and approximation abilities given sufficient and considerable inputs [26]. ANNs are derived from the biological neural networks in the human brain. Interconnected artificial neural networks are usually composed of neurons that can deal with the inputs and follow various situations. ANNs are suitable not only for machine learning but also pattern recognition. Therefore, ANNs have become a popular way of indicating a function by observation in the case of complex data. Figure 3a shows a typical ANN structure, including input, hidden, and output layers.

In this study, several ANN methods, including the backpropagation neural network (BPNN), extreme learning machine (ELM), probabilistic neural network (PNN), and long short-term memory (LSTM) neural network, were used to develop variable-flow models. The BPNN and ELM are feedforward neural networks with no cycles or loops. Information propagates in one direction, forward from the input layer, through the hidden layer, and then to the output layer, in a feedforward neural network.

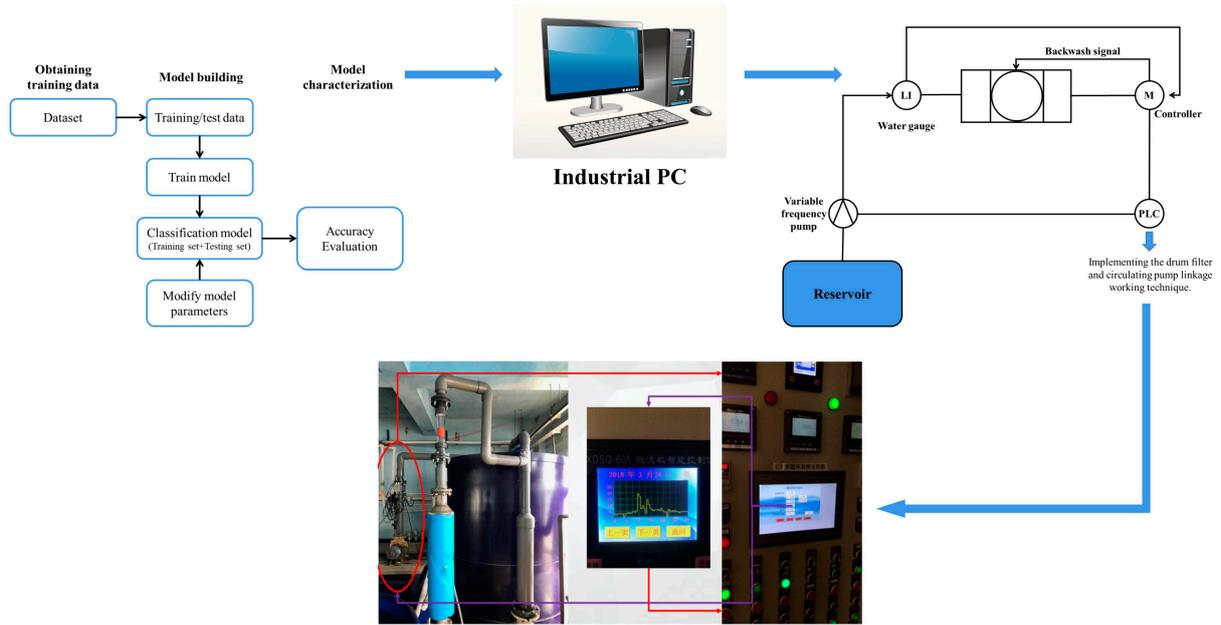


Figure 2. Design of control system in the variable-flow RAS.

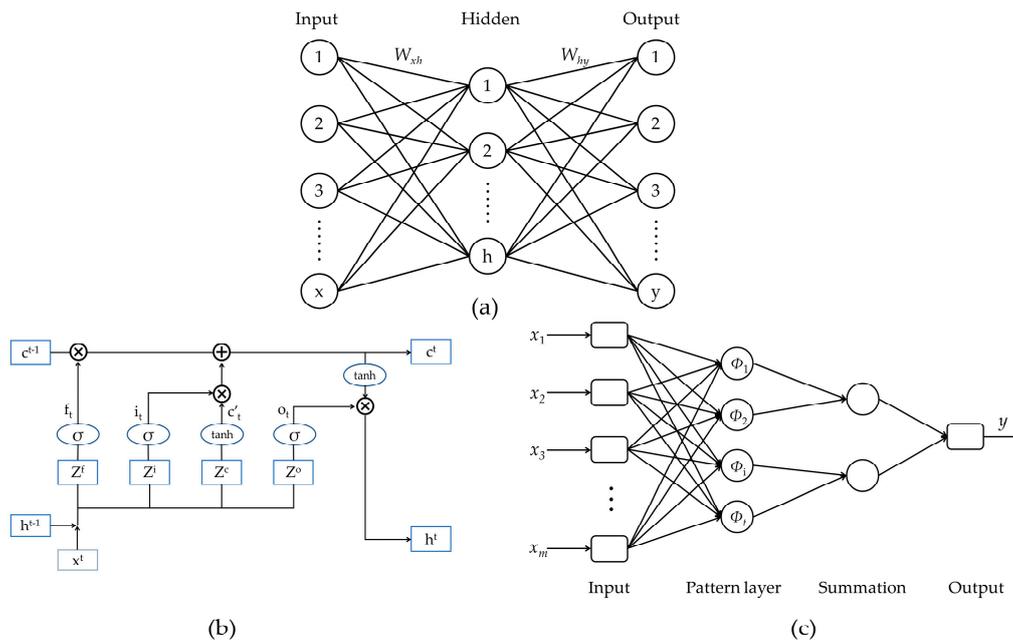


Figure 3. Artificial neural network architectures: (a) the typical structure of an ANN; (b) the LSTM architecture; (c) the PNN architecture.

The activation function can introduce a nonlinear factor to the neuron so that the ANN can approximate any nonlinear function. In the present study, a sigmoid function was adopted in the BPNN model and ELM model. For the sigmoid activation function, it holds that

$$f(z) = \frac{1}{1 + \exp(-z)}, \tag{1}$$

where the output of the sigmoid function is between 0 and 1. For the binary classification task, the output of the sigmoid is divided into a positive class/negative class when the output satisfies a certain probability condition.

Figure 3b shows the schematic of the LSTM network. The LSTM network is a special

RNN focusing on long sequences of data [27]. A standard LSTM unit comprises a cell, an input gate, an output gate, and a forget gate to solve the long-term dependency problem. Long-term memory information is stored during three steps (forgetting, remembering, and outputting) in an LSTM. In the present study, a rectified linear unit (ReLU) function was applied in the LSTM model. The ReLU function is described as

$$f(x) = \max(0, x), \quad (2)$$

which means that

$$\text{ReLU}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}. \quad (3)$$

The convergence rate of the stochastic gradient descent obtained by the ReLU function is much faster than the tanh/sigmoid function. However, the learning rate should be set appropriately to prevent neurons in the network from losing their activation ability. In this study, the parameters of the LSTM training process were set as follows: sequence input layer = 9, initial learning rate = 0.01, learning rate drop factor = 0.1, batch size = 128, number of training epochs = 200, hidden layer = 1 (with 32 hidden units). Adaptive moment estimation (Adam) was chosen as the optimization method. The fully connected layer was set as 2 for the binary classification task.

Figure 3c shows the architecture of a typical PNN, which was first proposed by Dr. D.F. Specht [28]. As a branch of a radial basis network, PNN has the advantages of a simple learning process and fast training time. Therefore, PNN models can be well implemented in hardware since the neuron number in each layer is fixed. Generally, a PNN network contains four layers: input layer, pattern layer, summation layer, and output layer. The input layer simply distributes the input to the neurons in the pattern layer. The pattern layer neuron may compute its output by Gaussian function when receiving  $x$  from the input layer. It holds that

$$y_g(x; \sigma) = \frac{1}{l_g (2\pi)^{n/2} \sigma^n} \sum_{i=1}^{l_g} \exp\left(-\sum_{j=1}^n \frac{(x_{ij}^{(g)} - x_j)^2}{2\sigma^2}\right), \quad (4)$$

where  $l_g$  denotes the total number of samples,  $n$  is the input feature, sigma represents the smoothing parameter, and  $x_{ij}$  represents the  $j$ -th data of the  $i$ -th neuron of the class  $g$ . The summation layer connects the pattern layer units of each class, and then the output layer is responsible for outputting the category with the highest score in the summation layer. K-fold cross-validation is useful for preventing models with small datasets from overfitting but is not used too frequently in deep learning. The dataset is equally divided into  $k$  parts. Every time a unique fold is used as a validation subset, the remaining pattern examples train the ANN. In this study, we introduced 4-fold cross-validation to evaluate the machine learning models. The evaluation indicators were all calculated by averaging the 4-fold cross-validation results.

### 2.3.2. Support Vector Machine (SVM)

An SVM has excellent generalization ability between model complexity and learning ability when dealing with limited sample information [29]. In SVM applications, choosing the appropriate kernel function and suitable parameters is crucial for prediction accuracy. As for the linear separable binary classification, finding the optimal hyperplane that divides all samples with maximum margin is the principal function of an SVM. For linear problems, the optimal classification hyperplane in separating two classes of training vector sets  $D$  is

$$D = \left\{ (x^1, y^1), \dots, (x^l, y^l) \right\}, \quad x \in R^n, y \in (-1, 1). \quad (5)$$

The plane was assumed as

$$\langle w, x \rangle + b = 0, \tag{6}$$

When the optimal classification surface is generated, the vectors are classified without error, and when redundancy occurs, a typical hyperplane is assumed where  $w$  and  $b$  are constrained:

$$\min_i |\langle w, x^i \rangle + b| = 1. \tag{7}$$

The classification hyperplane in the regular form must satisfy the following constraints:

$$y^i [\langle w, x^i \rangle + b] \geq 1, \quad i = 1, \dots, l. \tag{8}$$

The coordinate of the point  $x$  in the hyperplane at a distance  $d(w, b; x)$  is

$$d(w, b; x) = \frac{|\langle w, x^i \rangle + b|}{\|w\|}, \tag{9}$$

The final hyperplane that can satisfy the separated samples is the hyperplane that minimizes the data:

$$\Phi(w) = \frac{1}{2} \|w\|^2. \tag{10}$$

For nonlinear classification, the idea of SVM is to map the samples to a high-dimensional space, where the nonlinear problem is transformed into a linear solution using a kernel function, at which point the weight  $w$  is expressed as

$$w = \sum_{i=1}^l \alpha_i y_i \Phi(x_i). \tag{11}$$

Introducing the relaxation variable  $\xi (\xi \geq 0)$  describing the function interval, the optimization equation under the kernel approach is expressed as

$$\min \alpha + C \sum_{i=1}^l \xi. \tag{12}$$

The model is described as

$$\left\{ \begin{array}{l} y_l \left( \sum_{j=1}^l \alpha_j y_j K(x_j, y_i) + b \right) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ \alpha \geq \alpha_j, \quad j = 1, \dots, l \\ \alpha \geq -\alpha_j, \quad j = 1, \dots, l \\ \alpha, b \in R, \xi_i \geq 0, \quad i = 1, \dots, l \end{array} \right. . \tag{13}$$

In the present study, the SVM model was adopted to control the inverter frequency to improve circulating pump operating efficiency under different water quality conditions. The SVM is a kind of machine learning algorithm with a high generalization ability to classify and predict small samples. As upshifting and downshifting of the circulating pump is a binary problem, water quality indicators as variables can provide good generalization ability for the model. Support vector classification (SVC) can be used as the core algorithm for developing drum filter-circulating pump linkage technology. However, there is no international standard for selecting optimal parameters, and the parameter selection principles are based on dataset performance and the construction of a more reliable solution through

cross-validation methods [30,31]. Here, we used the Gaussian kernel function in resolving the nonlinear support vector classification task:

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right), \quad (14)$$

For the SVM model, the penalty parameter  $C$  and RBF kernel parameter  $g$  need to be decided to improve the classification accuracy. In the present study, several optimizing algorithms, including grid search (GS), least squares method (LS), genetic algorithm (GA), and cuckoo search (CS) algorithm, were applied to improve the classification performance of the SVM model. The parameters of GA were set as follows: max generation = 300, population size = 50, generation gap = 0.9, range of parameter  $c = (0, 100)$ , range of parameter  $g = (0, 1000)$ . For the CS algorithm, the parameters were set as follows: iteration = 300, number of nests = 20, probability = 0.25. The best parameters of GS and LS methods were obtained through the traversal method; the ranges of  $c$  and  $g$  were set as  $(0, 100)$  and  $(0, 1000)$ , respectively. K-fold cross-validation was utilized in the SVM models to prevent overfitting, and the evaluation indicators were calculated using averaging. The optimal SVM model can be determined by comparing the evaluation indicators of classification results from different algorithms.

### 3. Results

#### 3.1. Data Processing for Variable-Flow Regulation

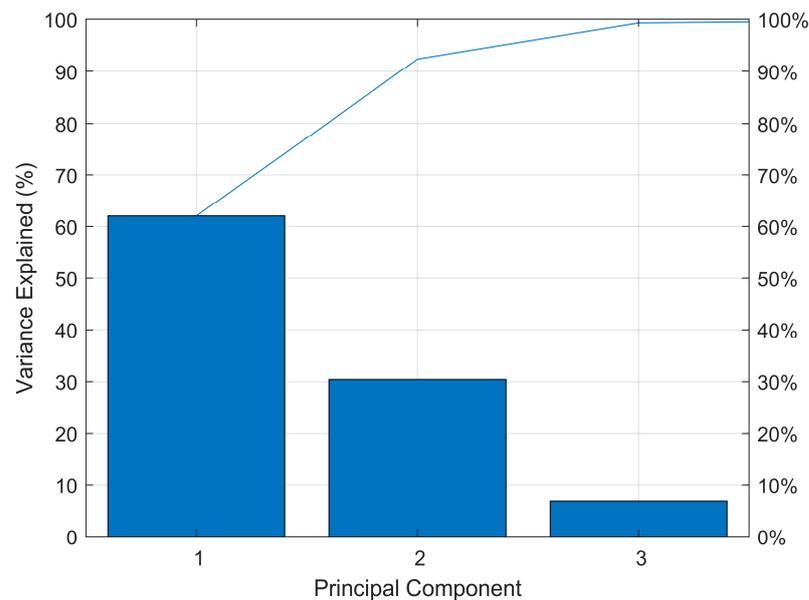
Ranges of the water quality data and backwash frequency from the measurements at three total circulation rates in RAS are shown in Table 1. The variable-flow regulation was decided by the frequency of the circulating pump. The upshifting and downshifting of the circulating pump inverter as two indicators of the classifier were labeled as 1 (upshift) and  $-1$  (downshift) in the dataset. In order to develop the variable-flow regulation models based on the machine learning methods, water quality indicators, current circulation flow rate, and current backwash frequency were used as input variables, and regulating data (upshift/downshift) for the next period (0.5 h) were used as output variables. Upshift/downshift data were labeled by manual marking. The marking principal was decided from the variable-flow experiments under three circulation rates in RAS. The binary classification models can be applied for variable-flow regulation strategy, and the current data for water quality indicators and backwash frequency can be used to determine the total circulation rates for the next period through the classification models.

**Table 1.** Ranges of water quality indicators at three total circulation rate levels.

Indicator	55 m <sup>3</sup> /h	65 m <sup>3</sup> /h	75 m <sup>3</sup> /h
Temperature (°C)	27.30~27.70	27.40~27.70	27.40~27.80
pH	6.90~7.50	6.91~7.50	6.91~7.49
DO (mg/L)	7.00~8.95	7.00~8.92	7.01~8.97
Salinity (‰)	30.50~30.90	30.50~30.90	30.50~30.90
TAN (mg/L)	0.18~2.68	0.20~1.36	0.19~1.13
Nitrite nitrogen (mg/L)	0.06~1.91	0.07~0.81	0.05~0.72
TSS (mg/L)	11.60~33.54	12.06~42.40	8.85~26.25
Turbidity (NTU)	3.52~8.98	3.63~13.68	2.83~8.87
Backwash frequency (Times per 0.5 h)	1~9	4~18	3~22

A total of 375 datasets were collected in the experiment, of which 280 were used as the training set and 95 as the test set. The training data were normalized after data pre-processing. The first step in developing the machine learning models was to simplify the explanatory variables by principal component analysis (PCA). PCA can reduce the complexity of the dataset and reveal hidden structures. The simplified principal components

can be used as valid indicators to develop models. Figure 4 illustrates that the simplified variables reduced the original dataset from nine dimensions (water quality indicators) to three dimensions and could reflect 99% of the information in the original independent variables. However, the key components extracted from the original data were compressed and mapped to another space, and the simplified variables were not directly related to the original data [32]. Hence, in the present study, PCA successfully provided the optimal reduced representation for the data. The new dataset could then be used to develop machine learning models to reduce the complexity of the computation process.



**Figure 4.** PCA process for reducing the complexity of water quality indicators.

### 3.2. Intelligent Variable-Flow Models

#### 3.2.1. Results of the ANN Models

ANN classification models, including GA-BP, ELM, PNN, and LSTM, were used to adjust the circulating pump's frequency. The upshifting operation of circulating pump frequency was labeled as 1, and downshifting operation was labeled as  $-1$ . The classification process was regarded as a binary classification problem. The classification accuracy of both training set and test set data was calculated. For the BPNN model, the GA algorithm was applied to optimize the model performance. Models were tested by cross-validation to prevent the overfitting problem. ANN models were implemented by programming in Python 3.8.5 [33]. For the BPNN model, the maximum epoch was set to 1000 iterations, and the learning rate was set to 0.01 during the training process. The GA method optimized the BPNN model with the lowest error rate (2.59%) at 25 generations. The GA-BP model had the best validation performance (0.12) at epoch 142. For the LSTM model training process, loss and accuracy gradually converged after 350 iterations. The accuracy of the training set reached 100% when the loss was below 0.05.

Table 2 presents the evaluations of the ANN classification models. Results showed that the training accuracy of all the ANN models was beyond 90%. PNN and LSTM achieved the most accurate classification (100%). For the test set, the LSTM model had a 96.84% accuracy rate; however, the accuracy rates of other models were less than 90%. Thus, the optimal model was identified as the LSTM model, with the highest accuracy for both the training set (100%) and test set (96.84%) among the ANN models.

#### 3.2.2. Results of the SVM Models

The SVM models were developed in Python 3.8.5. As classification accuracy is directly related to the optimal parameters of the SVM model, we used several optimizing

methods to determine penalty parameter  $c$  and the kernel parameter  $g$  in the present study. Table 3 shows the optimizing methods for SVM models. The optimized parameters were determined by the grid search, least squares, cuckoo search, and gene algorithm.

**Table 2.** Classification accuracy of training sets and test sets from ANN models.

ANNs	Training Accuracy	Test Accuracy
GA-BP	92.14%	86.32%
ELM	95.00%	89.47%
PNN	100.00%	71.58%
LSTM	100.00%	96.84%

**Table 3.** Classification accuracy of training sets and test sets from SVM models.

Optimizations	Training Accuracy	Test Accuracy	Best Gamma	Best $c$
Grid Search	100.00%	97.89%	0.0039	48.50
Least Square	94.29%	96.84%	250.00	20.00
Cuckoo Search	100.00%	97.89%	0.56	25.51
Genetic Algorithm	100.00%	98.95%	0.33	66.43

As Table 3 shows, the accuracy rates of classification results of the SVM models were maintained at relatively high levels. The least squares method had 94.29% accuracy, and other methods all had 100% accuracy rates for the training set. The test set from the gene algorithm optimized support vector machine (GA-SVM) model had one set of data classified with the wrong label among 95 groups (accuracy 98.95%). The grid search optimized support vector machine (GS-SVM) and the cuckoo search optimized support vector machine (CS-SVM) both had two error sets (97.89%). For the least squares support vector machine (LS-SVM), the test set results exhibited lower accuracy (96.84%) than other methods. Thus, the GA-SVM was identified as the optimal SVM classification model through comprehensive comparison. Table 3 shows that the four searching algorithms optimized the parameters (penalty  $c$  and kernel radius  $g$ ). Although the accuracy could be maintained at a high level, the ranges of the optimized parameters of the SVM models were quite different. Therefore, it was necessary to further select the SVM model through evaluation indicators.

### 3.3. Model Evaluation

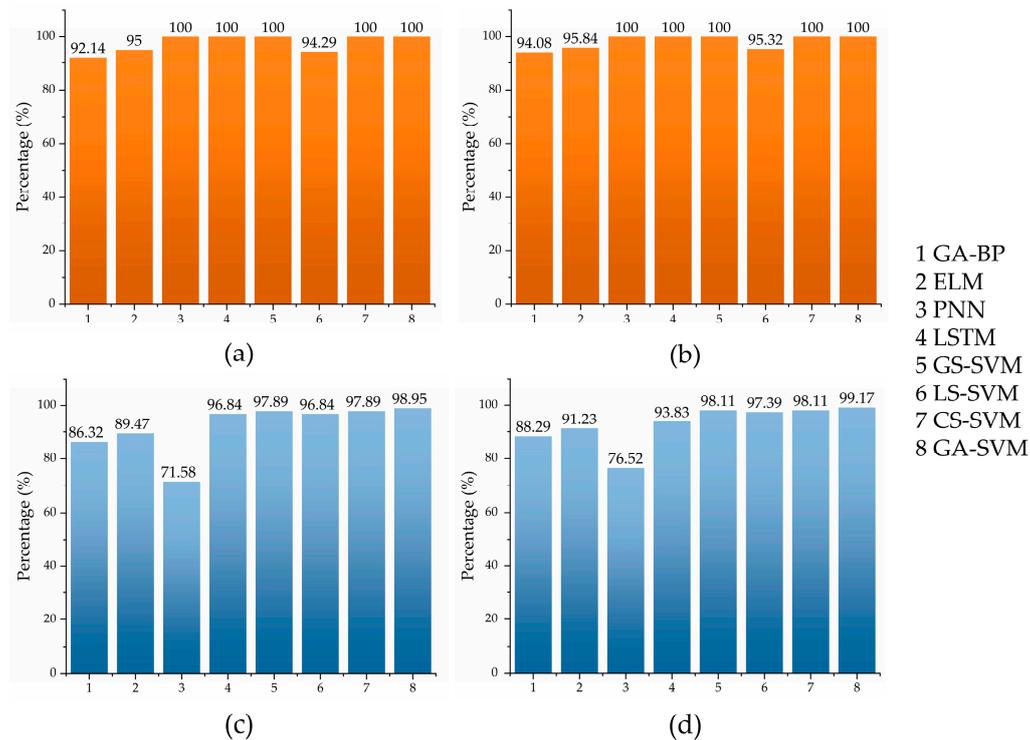
The confusion matrix, which comprehensively reflects the performance of the classifiers, can derive many evaluation indicators. Here, the calculated evaluation indicators, including accuracy, precision, recall, and F1-score, were used to evaluate classification performance for the binary classifier. The SVM model was estimated by 4-fold cross-validation, and the indicators were computed by averaging the folds. Accuracy represents the ratio of correct samples to the total samples without considering the positive and negative. Recall refers to the ratio of the correctly classified positive samples to the total true positive samples, and precision refers to the ratio of correctly classified positive samples to all classified positive samples. The F1-score indicator is proposed based on precision and recall to evaluate the indicators as a whole. The F1-score can be used to comprehensively consider the pros and cons of the classification models.

Table 4 shows the results of model evaluation indicators for machine learning classifiers. Figure 5 shows the histograms of the evaluation indicators (accuracy and F1-score) of the training set and test set from machine learning classification models. According to the summaries of the model evaluation indicators, GA-SVM shows both higher accuracy and F1-score than other machine learning methods. Accuracy can reflect the classification correctness of the global results of the model. The F1-score can reflect the weighted average between precision and recall, and the results show that the GA-SVM classifier can be

considered as an optimal model for drum filter-circulating pump linkage technology in a variable-flow RAS because the model indicators satisfied the criteria.

**Table 4.** Results of model evaluation indicators for machine learning classifiers.

Methods	Accuracy	Precision	Recall	F1-Score
GA-BP	92.14% (Train)	95.21% (Train)	92.98% (Train)	94.08% (Train)
	86.32% (Test)	90.74% (Test)	85.96% (Test)	88.29% (Test)
ELM	95.00% (Train)	96.41% (Train)	95.27% (Train)	95.84% (Train)
	89.47% (Test)	91.23% (Test)	91.23% (Test)	91.23% (Test)
PNN	100.00% (Train)	100.00% (Train)	100.00% (Train)	100.00% (Train)
	71.58% (Test)	81.48% (Test)	72.13% (Test)	76.52% (Test)
LSTM	100.00% (Train)	100.00% (Train)	100.00% (Train)	100.00% (Train)
	96.84% (Test)	94.34% (Test)	100% (Test)	93.83% (Test)
GS-SVM	100.00% (Train)	100.00% (Train)	100.00% (Train)	100.00% (Train)
	97.89% (Test)	98.11% (Test)	98.11% (Test)	98.11% (Test)
LS-SVM	94.29% (Train)	94.22% (Train)	96.45% (Train)	95.32% (Train)
	96.84% (Test)	96.55% (Test)	98.25% (Test)	97.39% (Test)
CS-SVM	100.00% (Train)	100.00% (Train)	100.00% (Train)	100.00% (Train)
	97.89% (Test)	98.11% (Test)	98.11% (Test)	98.11% (Test)
GA-SVM	100.00% (Train)	100.00% (Train)	100.00% (Train)	100.00% (Train)
	98.95% (Test)	98.36% (Test)	100% (Test)	99.17% (Test)



**Figure 5.** Evaluation indicators of classification models based on machine learning: (a) accuracy of the training sets; (b) F1-score of the training sets; (c) accuracy of the test sets; (d) F1-score of the test sets.

#### 4. Discussion

Feces and residual feed may decompose to organic suspended solids, which further generate TAN and nitrite, harming breeding animals' health. Suspended solids in the RAS also provide surface area that can be colonized by bacteria. As circulation intensity

increases, more particles accumulate, which may increase the bacterial carrying capacity of the system. Hence, rapid removal of solid waste is the most critical unit process in an RAS [34]. The traditional method of water quality regulation in an RAS is to act when water quality deteriorates. This approach leads to large fluctuations in the water environment, and the cost of water quality regulation becomes very high, often requiring many water exchanges to control water quality. This study proposes regulation of RAS circulation based on process control technology, relying on the microfilter backwash times in a unit period (0.5 h) as the main parameter to reflect the overall turbidity of the water body. The variable-flow RAS circulation strategy was designed to form microfilter-circulating pump linkage technology based on water quality parameters and backwash times at different flow rates. An intelligent variable-flow regulation model was developed to keep the water clean and quickly and dynamically remove suspended solids.

Related research has proven the significant differences in water quality between the high and low makeup water exchange treatment groups [35]. One study has shown that increasing RAS water circulation can effectively reduce ammonia and nitrite [36]; the higher the circulation level, the lower the ammonia and nitrite mass concentrations became. Moreover, the conversion of nitrite revealed a certain hysteresis, and the ammonia peak appeared earlier than the nitrite peak after feeding was stopped.

RAS solids come mainly from uneaten feed and fecal solids, and the decomposition and mineralization of these solids lead to elevated ammonia and nitrite levels in the RAS [10]. Data such as TAN, NO<sub>2</sub>-N, and TSS must be obtained by manual measurement and are challenging to obtain by sensors. According to Vinatea et al. [37], TSS tended to accumulate in the intensive *L. vannamei* culture and was eventually reflected in an increase in NTU. As both turbidity and TSS can reflect the clarity of a liquid, the turbidity parameter was used for modeling in this study. The principal component analysis (PCA) results for dimensionality reduction showed that turbidity, dissolved oxygen, pH, and temperature could be used as the leading indicators for modeling. The variable-flow regulation model obtains the current water quality indicators in real time and then applies these indicators to predict and classify the circulation rate for the next period. The turbidity sensor in turbulent flow had a measured data fluctuation that was too large, and the sensor arrangement position also caused measurement errors. An innovative point of this study is that the drum filter backwash frequency over a certain period was used as one of the critical factors for modeling instead of the momentary RAS water turbidity. Backwash times can effectively replace turbidity reading to reflect overall RAS water turbidity, avoiding the instability of the data collected by the turbidity sensor.

The application of machine learning methods in aquaculture-related research is focused mainly on the prediction, classification, and evaluation of water quality indicators such as dissolved oxygen, salinity, pH, ammonia, and nitrite [25]. In the present study, machine learning was used to model the variable-flow regulation strategy. Sensors collected water indicators, including DO, pH, temperature, and turbidity. In order to implement the variable-flow principle, the machine learning methods were introduced in the present study to develop the optimal variable-flow regulation model for RAS. The water quality indicators, the backwash frequency, and the circulating pump frequency were obtained through continuous monitoring. For the ANN methods, the LSTM model was identified as the optimal regulation model, since the accuracy and F1-score indicators reflected the strong ability of the LSTM classifier. The modeling data based on time series were collected from the continuously running RAS in the present study. The water quality indicators, backwash frequency, and total circulation rates were recorded through the fixed time interval during the whole rearing period. Research has shown that LSTM can indeed perform well in processing long time series sequences of data [38]. The optimal classification model needs to be relatively simple in order to be applied in the embedded devices. The variable-flow adjustment strategy in RAS also needs to respond quickly and satisfy the high standard of classification accuracy. All the evaluated indicators of the SVM models demonstrated better results compared with the LSTM model. The gene algorithm contributed the highest

accuracy and F1-score among the four optimization algorithms in the classification task. As a supervised algorithm, GA-SVM can be applied to effectively adjust water refreshment in RAS.

In future work on variable-flow RAS regulation, the data-driven model needs to be improved to establish continuous variable-flow control technology by adjusting circulating pump frequency. A larger quantity of data from the running RAS can ensure higher availability and robustness for optimizing the intelligent variable-flow strategy. The continuous variable-flow control technology prerequisite is required for the indicators (water quality, backwash frequency, and rearing cycle) to correspond to the ideal circulation volume. Furthermore, the interaction effects between various indicators need to be revealed through experiments and analysis. The ultimate goal of the study is to achieve a precise circulation control strategy in the RAS and execute rapid water treatment without affecting the health of the reared animals.

## 5. Conclusions

A variable-flow regulation model was established in the present study to implement the circulating pump-drum filter linkage working technique. Classification models based on machine learning methods between the explanatory variables and the regulation strategy were developed based on experimental data. ANN models including GA-BP, LSTM, PNN, and ELM were established. The LSTM model had the highest accuracy (training set 100%, test set 96.84%) and F1-score (training 100%, test 93.83%) and was regarded as the best classification model among ANN methods. SVM models were developed and optimized using linear squares, grid search, cuckoo search, and gene algorithm. Results showed that SVM models required less training time and exhibited higher accuracy compared with ANN models. Finally, the optimal model was GA-SVM, with the highest classification accuracy (training 100%, test 98.95%) and F1-score (training 100%, test 99.17%). The model was tested under cross-validation with precise classification performance and used for the circulating pump-drum filter intelligent linkage working technique.

**Author Contributions:** F.C. conducted the experiments, analyzed the data, developed the models, and wrote the paper. Y.D. and T.Q. reviewed the manuscript. Z.X. designed the RAS. L.Z., J.X., M.S. and Y.L. conducted the shrimp rearing experiments. J.S. conceived and designed the experiments. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key Program for International Cooperation on Scientific and Technological Innovation, Ministry of Science and Technology of the People's Republic of China, grant number 2017YFE0118300, and also funded by the National Key R&D Programs of China, grant numbers 2019YFD0900800 and 2019YFD0900502.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Acknowledgments:** We are grateful to the reviewers and the editors for their valuable and insightful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. FAO. *The State of World Fisheries and Aquaculture 2016*; Publications of Food and Agriculture Organization of the United Nations: Rome, Italy, 2016; p. 200.
2. Zhang, S.Y.; Li, G.; Wu, H.B.; Liu, X.G.; Yao, Y.H.; Tao, L.; Liu, H. An integrated recirculating aquaculture system (RAS) for land-based fish farming: The effects on water quality and fish production. *Aquac. Eng.* **2011**, *45*, 93–102. [[CrossRef](#)]
3. Ebeling, J.M.; Timmons, M.B. *Recirculating Aquaculture*, 3rd ed.; Ithaca Publishing Company: New York, NY, USA, 2010; pp. 171–474.

4. Badiola, M.; Mendiola, D.; Bostock, J. Recirculating Aquaculture Systems (RAS) analysis: Main issues on management and future challenges. *Aquac. Eng.* **2012**, *51*, 26–35. [[CrossRef](#)]
5. Chen, S.; Stechey, D.; Malone, R.J.D. Suspended solids control in recirculating aquaculture systems. *Dev. Aquac. Fish. Sci.* **1994**, *27*, 61.
6. Noble, A.C.; Summerfelt, S.T. Diseases encountered in rainbow trout cultured in recirculating systems. *Annu. Rev. Fish. Dis.* **1996**, *6*, 65–92. [[CrossRef](#)]
7. Wedemeyer, G. *Physiology of Fish in Intensive Culture Systems*, 1st ed.; Springer Science & Business Media: Boston, MA, USA, 1996; pp. 36–80.
8. Cripps, S.J.; Bergheim, A.J.A. Solids management and removal for intensive land-based aquaculture production systems. *Aquac. Eng.* **2000**, *22*, 33–56. [[CrossRef](#)]
9. Alabaster, J.S.; Lloyd, R.S. *Water Quality Criteria for Freshwater Fish*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2013; pp. 4–6.
10. Chiam, C.k.; Sarbatly, R.J.S.; Reviews, P. Purification of aquacultural water: Conventional and new membrane-based techniques. *Sep. Purif. Rev.* **2011**, *40*, 126–160. [[CrossRef](#)]
11. Xiao, R.; Wei, Y.; An, D.; Li, D.; Ta, X.; Wu, Y.; Ren, Q. Review on the research status and development trend of equipment in water treatment processes of recirculating aquaculture systems. *Rev. Aquac.* **2019**, *11*, 863–895. [[CrossRef](#)]
12. Vilbergsson, B.; Oddsson, G.V.; Unnthorsson, R. Taxonomy of means and ends in aquaculture production—Part 2: The technical solutions of controlling solids, dissolved gasses and pH. *Water* **2016**, *8*, 387. [[CrossRef](#)]
13. Gravdal, A. Process and Means for the Treatment of Water in an Aquaculture System. U.S. Patent 7,052,601, 30 May 2006.
14. Su, M.; Liu, H.; Song, H.; Hu, B. Study on the TSS removal efficiency and energy consumption of micro-screen drum filter. *Fish. Mod.* **2008**, *35*, 9–12.
15. Malone, R.J.R. *Recirculating Aquaculture Tank Production Systems*; Southern Regional Aquaculture Center: Stoneville, MS, USA, 2013; p. 12.
16. Prabhu, P.A.J.; Kaushik, S.; Geurden, I.; Stouten, T.; Fontagne-Dicharry, S.; Veron, V.; Mariojouis, C.; Verreth, J.; Eding, E.; Schrama, J. Water exchange rate in RAS and dietary inclusion of micro-minerals influence growth, body composition and mineral metabolism in common carp. *Aquaculture* **2017**, *471*, 8–18. [[CrossRef](#)]
17. Di Nunno, F.; Granata, F.; Gargano, R.; de Marinis, G. Forecasting of extreme storm tide events using NARX neural network-based models. *Atmosphere* **2021**, *12*, 512. [[CrossRef](#)]
18. Di Nunno, F.; Granata, F.; Gargano, R.; de Marinis, G. Assessment, Prediction of spring flows using nonlinear autoregressive exogenous (NARX) neural network models. *Environ. Monit. Assess.* **2021**, *193*, 1–17. [[CrossRef](#)] [[PubMed](#)]
19. López-Cortés, X.A.; Nachtigall, F.M.; Olate, V.R.; Araya, M.; Oyanedel, S.; Diaz, V.; Jakob, E.; Ríos-Momberg, M.; Santos, L.S. Fast detection of pathogens in salmon farming industry. *Aquaculture* **2017**, *470*, 17–24. [[CrossRef](#)]
20. Zhou, C.; Lin, K.; Xu, D.; Chen, L.; Guo, Q.; Sun, C.; Yang, X. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. *Comput. Electron. Agric.* **2018**, *146*, 114–124. [[CrossRef](#)]
21. Chen, Y.; Fang, X.; Yang, L.; Liu, Y.; Gong, C.; Di, Y. Artificial Neural Networks in the Prediction and Assessment for Water Quality: A Review. *J. Phys. Conf. Ser.* **2019**, *1237*, 042051. [[CrossRef](#)]
22. Zhou, C.; Xu, D.; Chen, L.; Zhang, S.; Sun, C.; Yang, X.; Wang, Y. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture* **2019**, *507*, 457–465. [[CrossRef](#)]
23. Li, D.; Wang, Z.; Wu, S.; Miao, Z.; Du, L.; Duan, Y. Automatic recognition methods of fish feeding behavior in aquaculture: A review. *Aquaculture* **2020**, *528*, 735508. [[CrossRef](#)]
24. Li, H.; Zhang, Z.; Zhao, Z.-Z. Data-mining for processes in chemistry, materials, and engineering. *Processes* **2019**, *7*, 151. [[CrossRef](#)]
25. Chen, Y.; Song, L.; Liu, Y.; Yang, L.; Li, D. A review of the artificial neural network models for water quality prediction. *Appl. Sci.* **2020**, *10*, 5776. [[CrossRef](#)]
26. Li, H.; Liu, Z.; Liu, K.; Zhang, Z.J. Predictive power of machine learning for optimizing solar water heater performance: The potential application of high-throughput screening. *Int. J. Photoenergy* **2017**, 2017. [[CrossRef](#)]
27. Granata, F.; Di Nunno, F. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric. Water Manag.* **2021**, *255*, 107040. [[CrossRef](#)]
28. Specht, D.F. Probabilistic neural networks. *Neural Netw.* **1990**, *3*, 109–118. [[CrossRef](#)]
29. Vapnik, V.; Guyon, I.; Hastie, T. Support vector machines. *Mach. Learn.* **1995**, *20*, 273–297.
30. Chen, F.; Li, H.; Xu, Z.; Hou, S.; Yang, D.J. User-friendly optimization approach of fed-batch fermentation conditions for the production of iturin A using artificial neural networks and support vector machine. *Electron. J. Biotechnol.* **2015**, *18*, 273–280. [[CrossRef](#)]
31. Li, H.; Chen, F.; Cheng, K.; Zhao, Z.; Yang, D. Prediction of zeta potential of decomposed peat via machine learning: Comparative study of support vector machine and artificial neural networks. *Int. J. Electrochem. Sci.* **2015**, *10*, 6044–6056.
32. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [[CrossRef](#)] [[PubMed](#)]
33. Bowles, M. *Machine Learning in Python: Essential Techniques for Predictive Analysis*, 1st ed.; John Wiley & Sons: Indianapolis, IN, USA, 2015; pp. 255–319.
34. Summerfelt, R.C.; Penne, C.R. Solids removal in a recirculating aquaculture system where the majority of flow bypasses the microscreen filter. *Aquac. Eng.* **2005**, *33*, 214–224. [[CrossRef](#)]

35. Davidson, J.; Good, C.; Welsh, C.; Summerfelt, S. The effects of ozone and water exchange rates on water quality and rainbow trout *Oncorhynchus mykiss* performance in replicated water recirculating systems. *Aquac. Eng.* **2011**, *44*, 80–96. [[CrossRef](#)]
36. Fivelstad, S.; Binde, M. Effects of reduced waterflow (increased loading) in soft water on Atlantic salmon smolts (*Salmo salar* L.) while maintaining oxygen at constant level by oxygenation of the inlet water. *Aquac. Eng.* **1994**, *13*, 211–238. [[CrossRef](#)]
37. Vinatea, L.; Gálvez, A.O.; Browdy, C.L.; Stokes, A.; Venero, J.; Haveman, J.; Lewis, B.L.; Lawson, A.; Shuler, A.; Leffler, J.W. Photosynthesis, water respiration and growth performance of *Litopenaeus vannamei* in a super-intensive raceway culture with zero water exchange: Interaction of water quality variables. *Aquac. Eng.* **2010**, *42*, 17–24. [[CrossRef](#)]
38. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1394–1401.