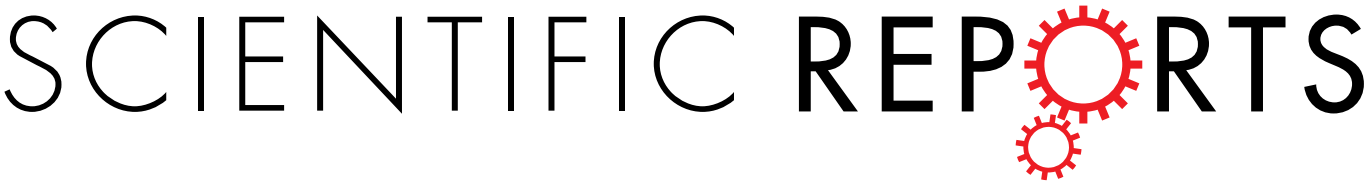

OPEN

Dissolved oxygen content prediction in crab culture using a hybrid intelligent method

Received: 02 February 2016
Accepted: 17 May 2016
Published: 08 June 2016

Huihui Yu[1,2,3,*], Yingyi Chen[1,2,3,*], ShahbazGul Hassan[1,2,3] & Daoliang Li[1,2,3]

A precise predictive model is needed to obtain a clear understanding of the changing dissolved oxygen content in outdoor crab ponds, to assess how to reduce risk and to optimize water quality management. The uncertainties in the data from multiple sensors are a significant factor when building a dissolved oxygen content prediction model. To increase prediction accuracy, a new hybrid dissolved oxygen content forecasting model based on the radial basis function neural networks (RBFNN) data fusion method and a least squares support vector machine (LSSVM) with an optimal improved particle swarm optimization(IPSO) is developed. In the modelling process, the RBFNN data fusion method is used to improve information accuracy and provide more trustworthy training samples for the IPSO-LSSVM prediction model. The LSSVM is a powerful tool for achieving nonlinear dissolved oxygen content forecasting. In addition, an improved particle swarm optimization algorithm is developed to determine the optimal parameters for the LSSVM with high accuracy and generalizability. In this study, the comparison of the prediction results of different traditional models validates the effectiveness and accuracy of the proposed hybrid RBFNN-IPSO-LSSVM model for dissolved oxygen content prediction in outdoor crab ponds.

Dissolved oxygen is one of the most important physical properties in crab ponds because it has great influence on the overall health and growth status of the aquatic ecosystem[1]. Proper control and management of dissolved oxygen in crab pond aquaculture is crucial for the developing crabs and has a significant impact on the quality and quantity of the final product[2].Thus, the efficient and accurate prediction of the dissolved oxygen content in modern aquaculture can provide a basis for water quality control and management, reducing aquaculture risk and financial losses and optimizing operation[3,4]. Liu *et al.*[2] have built dissolved oxygen content prediction models using machine learning methods that use just the complete valid data to build the forecasting model without considering invalid data samples. In a practical application, however, the data provided by a single sensor may lack accuracy or have limits[5], e.g., missing data or extreme data. Hence, this study presents a new hybrid dissolved oxygen content forecasting model that first uses the data fusion method to improve information accuracy and provide trustworthy training samples and then builds the dissolved oxygen content prediction model.

Irregularities in the data from multiple sensors may be due to incomplete or partial data or human activity[6]. The accuracy of the prediction model relies on the accuracy of the training data. Moreover, the data collected by a single sensor maybe inaccurate or limited[7]. Hence, a good method must be adopted to improve the accuracy of the sensor data. In recent years, different data fusion[5] strategies such as feature selection, activity recognition, fault detection and precise information provision[7–10] have been developed to integrate information from multiple sensors. Data fusion techniques are mathematical techniques used to combine multiple values into a single precise value. The radial basis function neural network (RBFNN) method is one of the artificial neural network methods used for multi-sensor data fusion that has high accuracy[11]. The goal of using data fusion in this research is to obtain more precise data for the dissolved oxygen content prediction model.

In the last few years, different approaches have been applied to water quality prediction. The typical water quality simulation and prediction models can be divided into physical models and black box models. The physical methods are based on mathematical theory[12,13]. Thus, it is a difficult task to determine the parameters or

[1]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China. [2]Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, P.R. China. [3]Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing, 100083, P.R. China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.L. (email: dliangl@cau.edu.cn)

extrapolate from the sub-models[2]. In contrast to physical models, 'black box' prediction models do not rely on situation-specific details and do not need to determine many parameters. Therefore, there have been many attempts made to use artificial intelligence techniques, such as time series methods, artificial neural network methods, and support vector machine methods[1,14–17].

Time series methods have been used for linear water quality prediction models, for example, the auto regressive moving average (ARMA) model and the auto regressive integrated moving average (ARIMA) model[18–20]. However, the dissolved oxygen content in crab aquaculture is complicated, changes nonlinearly and is influenced by many factors, hence, some time series methods cannot provide precise prediction accuracy[21]. Artificial neural networks (ANN) comprise a general purpose model that has been used to develop forecasting models with nonlinear series[22]. Hatzikos *et al.* develop neural network models to predict seawater quality indicators such as water pH, temperature and dissolved oxygen content[23]. Ma *et al.* use a back propagation neural network (BP-NN) model to predict water quality for aquaculture water management[24]. Although the neural network models have lower mean absolute error in their predictions and react more evenly to the indicators than their logistic counterparts[25], ANN also suffers from drawbacks, such as no exact rule for setting the neural network parameters and the time complexity of the learning process[21]. To overcome these disadvantages, a new approach should be explored.

A least squares support vector machine (LSSVM) is a robust regression technique used to solve with few samples and perform nonlinear function regression[26], so it has been applied in prediction methods for various fields[25,27,28]. In this study, a least squares version of SVM (LSSVM) is considered, in which the training is expressed in terms of solving a set of linear equations in the dual space instead of quadratic equations, as for the standard SVM case[29]. Moreover, the least squares support vector machine (LSSVM) improves on SVM by applying linear least squares criteria to the loss function[4,30]. In addition, the kernel parameter σ and the regularization parameter C in the LSSVM training procedure significantly influence forecasting accuracy[31]. To achieve a high level of performance with LSSVM models, the key parameters have to be tuned. To date, an exact method of obtaining an optimal set of LSSVM hyper parameters has not been determined.

Particle swarm optimization (PSO) is a heuristic global optimization method introduced by Kennedy and Eberhart in 1995[32]. It is widely used in fields such as function optimization, parameter training, model classification due to its many advantages, including its simplicity and easy implementation[33–35]. However, basic particle swarm optimization does not ensure convergence to an optimal solution and is also prone to partial optimization, which reduces precision in the regulation of its speed and direction[36]. Due to comparatively poor efficiency, a number of studies have been conducted on improving the performance of PSO algorithms, which are used in parameter optimization[2,37]. This study presents an improved particle swarm optimization algorithm for simultaneously optimizing the LSSVM parameters.

Results and Discussion

Dissolved oxygen content fusion result analysis. All computation for the prediction model was performed in MATLAB by coding in M files. In the proposed algorithm, all the experimental data were obtained from the same crab pond. The meteorological data were collected by the weather station installed on the shore of the pond, and the water quality data were collected from sensors installed in the same pond. All these experimental data were simultaneously transferred to the Digital Wireless Monitoring System of Aquaculture Water Quality and used by the proposed algorithm. The experimental data include water temperature, solar radiation, wind speed, rainfall, humidity, and four dissolved oxygen content sensor readings, which were used for data fusion. Because the data obtained from a single dissolved oxygen sensor is more unreliable, dissolved oxygen content values from dissolved oxygen sensors in four locations were used for data fusion along with other factors to obtain a relatively accurate dissolved oxygen content value. Then, the fused dissolved oxygen content value was used with the water temperature, solar radiation, wind speed, rainfall, and humidity data as the input sample for machine learning.

In the fusion method, we applied the RBF algorithm to the four DO sensors' data to obtain a better forecast-model training sample. The dissolved oxygen content value obtained by sensor1 was accepted as the real dissolved oxygen content value. For this part of the study, we used the first 500 data values from each of the four DO sensors as the fusion training samples and the next 200 as the test samples. Then, the water temperature, solar radiation, wind speed, rainfall, air humidity, and dissolved oxygen content values (from the 200 test sample groups) used for fusion were used as the prediction training and test samples for the dissolved oxygen content forecasting model.

After development, the RBFNN method was used to fuse or integrate the data; Fig. 1 shows the result of the data fusion. According to Fig. 1, there are many invalid (zero) and distorted (much higher than those nearby) dissolved oxygen content values in the original data. There is even one dissolved oxygen sensor (sensor 4) reporting data that is too low. However, fusion on the dissolved oxygen content data can effectively eliminate low credibility data. The data plots show that the fusion data discounts the bad sample data and generates more credible samples (Fig. 1).

Table 1 shows several invalid and distorted data values among the original dissolved oxygen content values obtained by the four DO sensors. As shown by the table, the results indicate that the fusion method can train with the other input factors to fuse data to obtain more trustworthy data for the prediction model. For example, at 20:20, the dissolved oxygen content value from sensor 1 is valid, and at 10:20, the dissolved oxygen content values from both sensor 2 and sensor 3 are invalid. The fusion values are more accurate than the invalid values. As the accuracy of the prediction model is dependent on the training samples, the fusion method is suitable to obtain more reliable data.

Forecast results analysis with fusion data. In the dissolved oxygen content forecasting model, the water temperature, solar radiation, wind speed, rainfall, air humidity, and previous fusion dissolved oxygen content value (test sample data) were combined for use as the prediction training and test values. The first 120 groups of

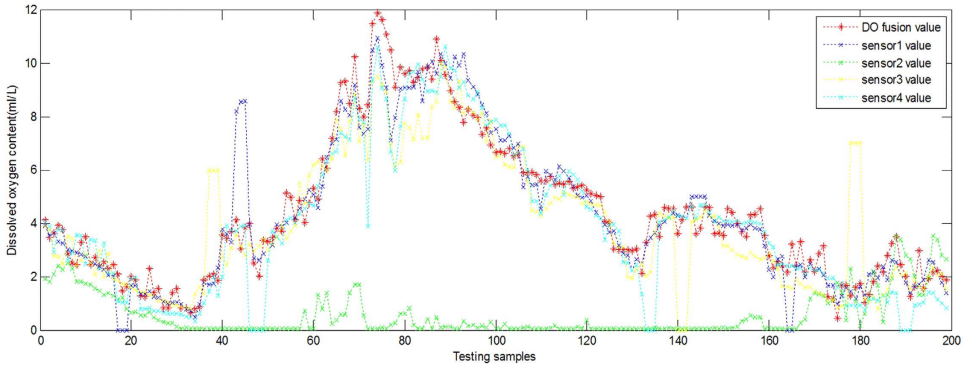

Figure 1. Fusion data and the original sensors' dissolved oxygen content value.

Time	Dissolved oxygen content				
	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Fusion value
29-06-2015 20:20	0.00	1.21	1.72	1.05	1.4547
30-06-2015 02:40	1.72	0.06	6.00	1.97	2.0888
30-06-2015 03:00	2.34	0.06	6.00	1.28	1.8274
30-06-2015 14:00	7.43	0.06	6.38	3.87	8.4344
01-07-2015 09:20	3.20	0.06	2.03	0.00	3.2813
01-07-2015 10:20	4.28	0.06	0.00	4.25	4.1231
01-07-2015 18:40	0.00	0.06	1.64	2.40	2.1536
02-07-2015 03:00	1.40	1.40	1.51	0.00	1.2531
02-07-2015 05:40	1.16	1.16	7.00	0.89	1.6136

Table 1. Error data fusion results obtained with the four DO sensor value-based RBF neural network data fusion methods.

data were used as training data and the last 80 groups of data were used as test data. To compare and evaluate the regression results of the RBFNN-IPSO-LSSVM dissolved oxygen content prediction model, we also used the BP neural network and standard LSSVM methods to predict the dissolved oxygen content. To analyse and compare prediction performance, the BP neural network, standard LSSVM and the optimized LSSVM all used the fusion data as training samples and forecast the sequence fusion test samples for the same time period. The BP neural network method consists of six input variables and one output variable, a hidden layer with five initial neurons, and a maximum training step value of 10^4. The standard LSSVM model parameters were selected by a 5-fold cross-validation method.

In the optimized LSSVM, the parameters were optimized by the improved PSO algorithm. The population size of the traditional PSO and the improved PSO were set to 50, the maximum evolution generation was set to 200, the particle dimension is 2, the initial inertia weight $u = 0.6$, and in the improved PSO the mutation probability $p_m = 0.3$ and $c_1 = c_2 = 1.5$. The fitness performance graph in Fig. 2 shows the fitness curves decrease rapidly at the outset of a generation but soon level off. The optimal parameters selected for the LSSVM by the improved IPSO were $C = 151.46, \sigma = 1.89$.

In this step, the fused dissolved oxygen content data were used as the real data for the three comparison methods' samples. Figure 3 contains the forecasting results of the BP, LSSVM and IPSO-LSSVM methods. The results show that the proposed hybrid model is more suitable and effective for dissolved oxygen content prediction. It has a strong ability to learn using a small nonlinear sample to achieve excellent generalizability.

As different algorithms employ different experimental methodologies, this study uses different standard statistical performance evaluation criteria. The standard statistical criteria include the root mean square error (RMSE), the mean absolute percentage error (MAPE), the Nash-Sutcliffe efficiency coefficient (NSC), the coefficient of determination (R^2) and the running time (T). Table 2 shows the error index for the three models. It indicates that the IPSO-LSSVM model combining the improved particle swarm optimization algorithm with the least squares support vector machine hybrid model is more adequate than the standard LSSVM and the BP neural network. The index (MAE, RMSE, MSE, NSC and T) of the IPSO-LSSVM is better than those of the other models. The running time (T) of the IPSO-LSSVM model is less than the LSSVM model, indicating the improved PSO method effectively selects the parameters of the LSSVM. Using the same test data, the relative MAE, RMSE and MSE differences between the IPSO-LSSVM model and the LSSVM model were 34.63%, 47.62% and 60.14% in the testing period. The relative MAE, RMSE and MSE differences between the IPSO-LSSVM model and the BP neural network model were 51.28%, 54.69%, and 67.32% in the testing period. So the IPSO-LSSVM model is more able to solve the solar greenhouse temperature prediction problem than the SVM and BP neural networks. Moreover, the

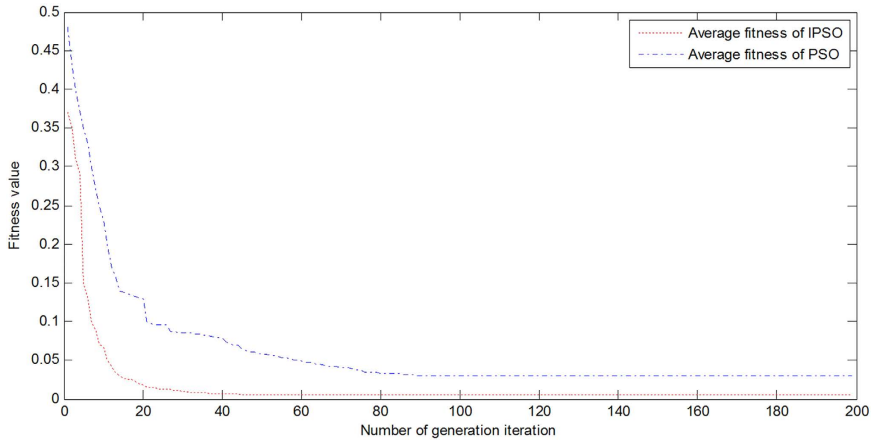

Figure 2. Fitness performance for the proposed IPSO and the traditional PSO.

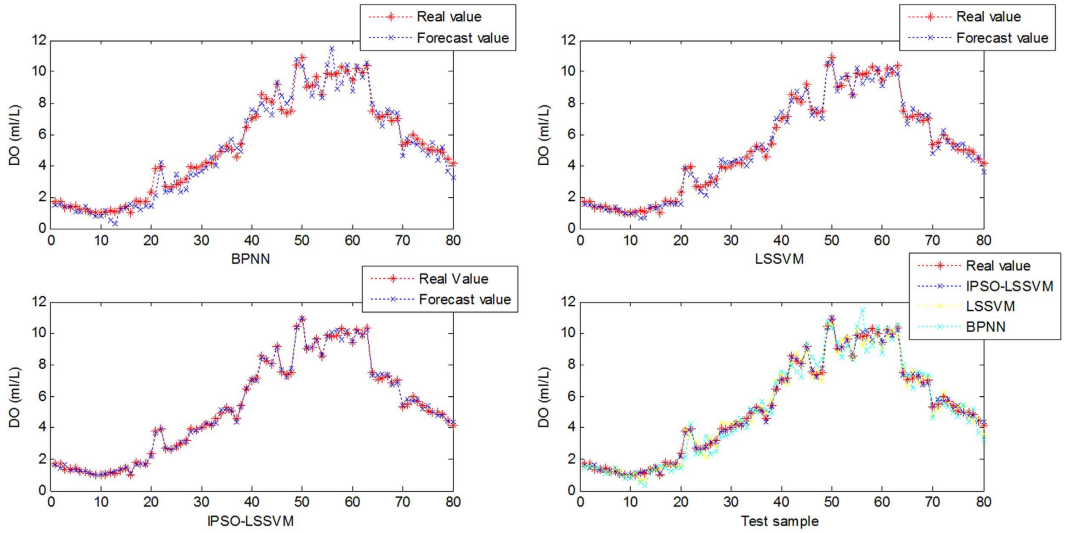

Figure 3. The dissolved oxygen content forecasting value of the RBFNN-IPSO-LSSVM in contrast with the comparison models.

Model	MAE	RMSE	MSE	NSC	T
IPSO-LSSVM	0.2814	0.4057	0.1085	0.9531	3.2143
LSSVM	0.4305	0.7745	0.2722	0.9187	3.1265
BPNN	0.5776	0.8954	0.3320	0.9002	4.3298

Table 2. Error statistics of four forecasting models.

NSC of the IPSO-LSSVM is higher than that of the other models. It is obvious that the IPSO-LSSVM model has significantly more reliable performance and generalizability and a higher prediction accuracy than other models.

Conclusions

In this study, a novel hybrid algorithm has been proposed for dissolved oxygen content prediction in outdoor crab ponds. To remove redundant and erroneous data from the original data, the RBF neural network method is used to fuse the original data to prepare training samples for the prediction model. Then, improved particle swarm optimization is used for the selection of the parameters for least squares support vector regression. Obtained results show that the proposed model yields better prediction accuracy in comparison with several other machine learning methods. Looking at the fusion result, we can see the fusion method removes erroneous data, increasing the original data's reliability, leading to a more trustworthy training sample for the prediction learning machine. The forecasting results show the IPSO-LSSVM model predicts more accurately than the traditional models.

The proposed hybrid model is used to predict the dissolved oxygen content in outdoor crab ponds. Our results demonstrate that the RBFNN-IPSO-LSSVM prediction model is effective and feasible.

For further study, the influence of some factors such as water temperature, wind speed, and solar radiation on the dissolved oxygen content is unclear. Hence, a method that can determine which factors should be used as input is important for improving the accuracy of the prediction model. Additionally, to obtain more precise hybrid predictor methods, the combination of different signal processing tools, feature selection and learning machines may be examined.

Materials and Methods

Data preparation. The data used in this study were collected by the Digital Wireless Monitoring System of Aquaculture Water Quality. Figure 4(a) shows the system structure diagram. The system is applied in more than 1000 river crab farming ponds, approximately 10000 acres. The system is made up of three major parts: the data acquisition layer; the information transport layer; and the application layer. The data acquisition layer is comprised mainly of the water quality monitoring sensors, such as the pH sensor, DO sensor, salinity sensor, and the weather monitoring station for temperature, solar radiation, atmospheric humidity, and wind speed. All of the data are moved via the transport layer to the application layer for data apperception, intelligent information processing, and logical operations.

In this study, the water quality and meteorological data were obtained from 21 June to 12 July in intervals of twenty minutes, totalling more than 1000 samples. In this study, the experimental details are as follows: (a) The size and population density of the crabs. During late June and early July, crabs complete their third shelling and begin the growth process for the fourth shelling. During this period, the stocking density of crabs is approximately 2000 per acre, their average weight is approximately 75 grams, and they reach a diameter of approximately 5.0 cm. (b) The area of the pond. The crab pond length is 130 meters and the width is 45 meters, for a total area of 5850 square meters. (c) The location of the sampling sites. The dissolved oxygen content collection sites are all in the same crab pond. The four dissolved oxygen sensors are evenly distributed in the crab pond; the specific locations are shown in Fig. 4(b). The meteorological data are collected by the weather station installed on the shore of the pond. These collection data include water temperature, solar radiation, wind speed, rainfall, humidity, and the dissolved oxygen content values from the four dissolved oxygen sensors. All these experimental data are simultaneously transferred to the Digital Wireless Monitoring System of Aquaculture Water Quality and used for the proposed algorithm.

The structure of the prediction model. Before training the machine learning model, we need to pre-process the original dissolved oxygen content data using the RBF neural network. Then, the improved particle swarm algorithm is used to determine the kernel parameter σ and the regularization parameter C for the least squares support vector machine. Finally, the dissolved oxygen content prediction model is built by training the forecasting method. As depicted in Fig. 5, the proposed method consists of three main parts:

- Data fusion.
- Least squares support vector regression parameters selection.
- Training of the learning machine.

Multi-sensor Data fusion by RBF neural network. The radial basis function (RBF) neural network is one of the neural network models that learn by measuring Euclidean distance data[38]. The RBF neural network model has a three-layer structure: the input layer, the hidden layer, and the output layer[39]. Movement from input layer to hidden layer is nonlinear, and that from hidden layer to output layer is linear. Determining the number of hidden nodes is an important issue that has a substantial impact on the neural model quality[40]. The input X is an M-dimensional vector, $X = [x_1, x_2, \ldots, x_m]$. The input layer units are only distributed to the hidden layer[41].

In the RBFNN method, each neuron in the hidden layer has a Gaussian function described as:

$$\varphi_i = \exp\left(-\frac{\|X - C_i\|^2}{2b_i^2}\right) \tag{1}$$

where $X = (x_1, x_2 \ldots x_m)$ is the i th input element, C_i is th centre of the i th hidden unit, b is the width of the receiving field, and n is the number of input elements.

The output layer is activated by the linear combination of the hidden layer units, which can expressed as:

$$y = \sum_{i=1}^{n} \varphi_i \omega_i \tag{2}$$

where φ_i is the weight of the connection from the hidden layer to the output layer. The RBF neural networks adopt the K-means clustering algorithm to improve the fusion result. The RBF neural network process for multi-sensor data fusion is shown in Fig. 6 and can be described as follows:

Step 1: Select the input vectors x_i as the training sample, $x_i \in R^m$ represents each group of data coming from all m detectors;

Step 2: Train the RBF neural networks by the K-means clustering algorithm and select the parameters;

Step 3: Set the error value and run the algorithm until the termination criterion is satisfied;

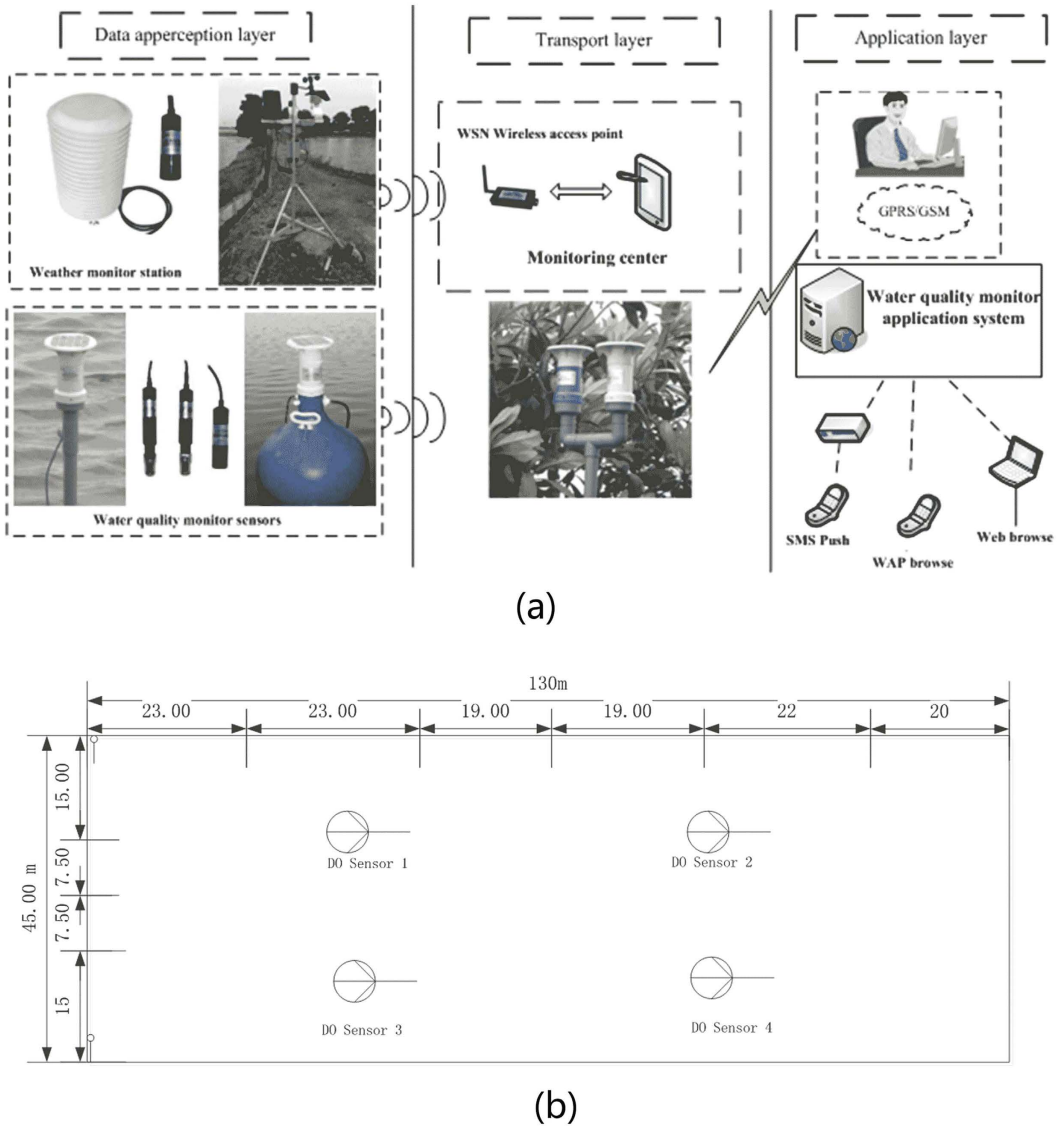

Figure 4. Experimental data collection system and location: (**a**)The structure diagram of the digital wireless monitoring system; (**b**) The sensor layout in the river crab pond. The photographs of the man, computer systems and web browser in Fig. 4 were taken by first author Huihui Yu in Gaocheng town, Yixing city, Jiangsu province. The drawing of the man in the top right of the figure, and the drawing of the equipment next to the "Water quality monitor application system" were created by Yingyi Chen. And, the whole figure was designed and drawn by author Huihui Yu and Yingyi Chen.

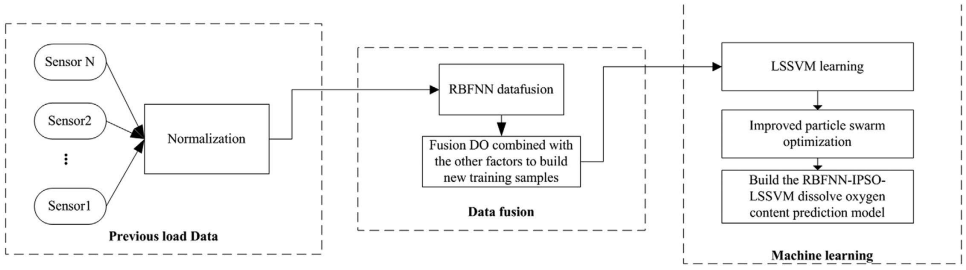

Figure 5. Proposed algorithm for dissolved oxygen content forecasting.

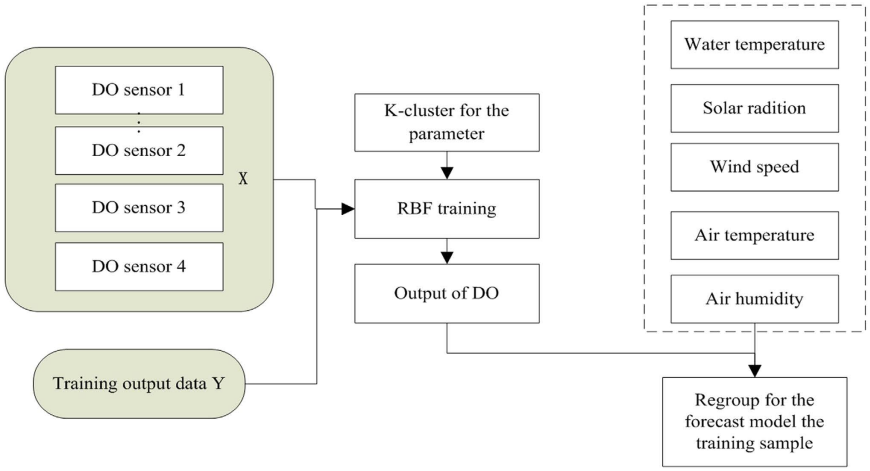

Figure 6. Process of multi-sensor data fusion by the K-cluster RBF method.

Step 4: Get the fused dissolved oxygen content value from the RBF neural network, then combine the environmental data and the fused dissolved oxygen content value into a new sample for the IPSO-LSSVM.

Least squares support vector machine (LSSVM). The least squares support vector machine (LSSVM) has been introduced as a reformulation of the standard support vector machine (SVM), which has simple techniques[42,43]. The LSSVM can address linear and nonlinear systems with structured risk minimization, and it has been successfully applied in many fields. In an LSSVM model, the training dataset is assumed to be $\{x_k, y_k\}$ $k = 1, 2, \ldots, l$, where $x_k \in R^n$ is an input vector and $y_k \in R$ is its corresponding target vector. The regression problem can be transformed into the following optimization problem:

$$\min_{w,b,e} J(w, j, e) = \frac{1}{2} w^T w + C\frac{1}{2}\sum_{k=1}^{l} {e_k}^2 \tag{3}$$

$$s.t.\ \ y_k = w^T \varphi(x_k) + b + e_k (k = 1, 2, \ldots, l) \tag{4}$$

where w is the weight vector, C is the regularization parameter, e_k is the error between the predicted and actual values of the system and $\varphi(.): R^n \rightarrow R^f$ is a function used to map the input space to a higher dimensional space. Using the Lagrangian function, the LSSVM model is written as follows:

$$\hat{y} = f(x) = \sum_{k=1}^{l} \alpha_k K(x, x_k) + b \tag{5}$$

where $K(x, x_k)$ is the kernel function. In this study, the RBF was selected as the kernel function because the RBF ably handles nonlinear relationships and its overall performance is excellent. This method maps the sample to a high dimensional space in a nonlinear fashion and it has few required parameters; therefore, it is the most popular option for kernel function. The kernel function is shown in Eq. (6):

$$K(x, x_k) = \exp\left(-\frac{\|x - x_k\|^2}{2\sigma^2}\right) \tag{6}$$

The model can be written as follows:

$$\hat{y} = f(x) = \sum_{k=1}^{l} \alpha_k \times \exp\left(-\frac{\|x - x_k\|^2}{2\sigma^2}\right) + b \tag{7}$$

Improved particle swarm optimization. The algorithm for particle swarm optimization (PSO) is an evolutionary optimization algorithm. It is initialized with a group of N random particles and simulates social behaviour among individuals[44,45]. The position of particle$i$ is represented as $X_i = (x_{i1}, x_{i2}, \ldots x_{ij}, \ldots, x_{iR})$. The velocity of particle i is represented as $V_i = (v_{i1}, v_{i2}, \ldots v_{ij}, \ldots, v_{iR})$, where $i = 1, 2, \ldots N$ and $j = 1, 2, \ldots, R$. The best previous position of particle i is represented as $pbest_i = (pbest_{i1}, pbest_{i2}, \ldots pbest_{ij}, \ldots, pbest_{iR})$ and the best particle among all particles is represented as $pgbest_i = (pgbest_{i1}, pgbest_{i2}, \ldots pgbest_{ij}, \ldots, pgbest_{iR})$. $pbest_{ij}$ is the local best position of the particle i in the j th dimension, and the $pgbest_{ij}$ is the global best position of the swarm in the j th dimension. During the search process, the direction of each particle is adjusted by dynamically altering its velocity

according to both its own movement and that of neighbouring particles[46,47]. The particle position and velocity are updated according to the following equation:

$$v_{ij}(k+1) = u \cdot v_{ij} + c_1 r_1(pbest_{ij}(t) - x_{ij}(t)) + c_2 r_2(pgbest_{ij}(t) - x_{ij}(t)) \tag{8}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{9}$$

where i is the iteration counter, u is the inertia weight, which is used to control the impact of the previous history of the current particle, c_1 is the cognition learning factor, c_2 denotes the social learning factor, and r_1 and r_2 are uniformly distributed random variables within the range [0,1].

The original particle swarm optimization has a slow convergence and may only find the traditional local optimum. In the study, an improved particle swarm optimization is presented to find a suitable inertia weight to balance the local and global search abilities. The dynamic adjustment method of the inertia weight u is written as Eq. (10):

$$u = \begin{cases} u \cdot (1 - p_m), fitness(x_{ij}) \le fitness(pgbest_j) \\ u \cdot (1 + p_m), fitness(pgbest_j) < fitness(x_{ij}) \le fitness(pbest_{ij}) \end{cases} \tag{10}$$

The p_m is introduced in this study as a mutation probability used to change the inertia weight$u$. When $fitness(x_{ij}) \le fitness(pgbest_j)$, the particles have a status which is close to the global optimum, so they are given a smaller inertia weight than the current one. The smaller inertia weight can help the particles to reach optimum status more quickly. When $fitness(pgbest_j) < fitness(x_{ij}) \le fitness(pbest_{ij})$, the particles are not close to the global optimum, and therefore, a bigger inertia weight is needed to change their position and velocity. The improved particle swarm optimization increases the convergence rate and also improves the accuracy of the solution.

Parameter selection. For least squares support vector regression, the kernel parameter σ and the regularization parameter C in the LSSVM training procedure have significant influence on forecasting accuracy. The improved particle swarm optimization is devoted to optimizing the kernel parameter σ and the regularization parameter C. Each particle represents a potential solution of the vector $d = (C, \sigma)$. The fitness function represents the performance of each particle and the fitness function is defined in the model as follows:

$$fitness = \sqrt{\frac{1}{N}\sum_{i}^{N}(\hat{y}_i - y_i)} \tag{11}$$

where the $\hat{y}_i$ represent the predicted values, the y_i represent the actual values, and N represents the size of the predicted value subset. The particle with a minimal fitness value is the global extreme point. The process of optimizing the LSSVM parameters with IPSO can be described as follows:

Step 1: Particle initialization and IPSO parameter setting: generate a population of initial particles that consists of parameter C and kernel parameter σ. Set the maximum number of iterations $k_{\max}$, the particle population number N, and the minimum fitness value for error limitation.

Step 2: Set the iteration variable $k = k + 1$.

Step 3: Calculate the fitness function value of each particle using Eq. (11): use the current particle as the individual extreme point of every particle and the particle with the minimal fitness value as the global extreme point.

Step 4: Calculate the weight $u$using Eq. (9), then update the velocity and position of the particles according to Eqs (8)–(9).

Step 5: Stop the algorithm if the termination criterion is satisfied and the best LSSVM model is produced. Otherwise, return to Step 2.

References

1. Carbajal-Hernández, J. J., Sánchez-Fernández, L. P., Carrasco-Ochoa, J. A. & Martínez-Trinidad, J. F. Immediate water quality assessment in shrimp culture using fuzzy inference systems. *Expert Syst Appl* **39,** 10571–10582 (2012).
2. Liu, S. *et al.* Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. *Comput Electron Agr* **95,** 82–91 (2013).
3. Han, H., Chen, Q. & Qiao, J. An efficient self-organizing RBF neural network for water quality prediction. *Neural Networks* **24,** 717–725 (2011).
4. Liu, S. *et al.* A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. *Eng Appl Artif Intel* **29,** 114–124 (2014).
5. Zhang, C., Hu, Y., Chan, F. T. S., Sadiq, R. & Deng, Y. A new method to determine basic probability assignment using core samples. *Knowl-Based Syst* **69,** 140–149 (2014).
6. Kabir, G., Demissie, G., Sadiq, R. & Tesfamariam, S. Integrating failure prediction models for water mains: Bayesian belief network based data fusion. *Knowl-Based Syst* **85,** 159–169 (2015).
7. Mouazen, A. M., Alhwaimel, S. A., Kuang, B. & Waine, T. Multiple on-line soil sensors and data fusion approach for delineation of water holding capacity zones for site specific irrigation. *Soil and Tillage Research* **143,** 95–105 (2014).
8. Lin, Y., Li, J., Lin, P., Lin, G. & Chen, J. Feature selection via neighborhood multi-granulation fusion. *Knowl-Based Syst* **67,** 162–168 (2014).
9. Banerjee, T. P. & Das, S. Multi-sensor data fusion using support vector machine for motor fault detection. *Inform Sciences* **217,** 96–107 (2012).
10. Kushwah, A., Kumar, S. & Hegde, R. M. Multi-sensor data fusion methods for indoor activity recognition using temporal evidence theory. *Pervasive and Mobile Computing* **21,** 19–29 (2015).

11. Jing, X., Yao, Y., Zhang, D., Yang, J. & Li, M. Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometric recognition. *Pattern Recogn* **40,** 3209–3224 (2007).
12. Steven D. Culberson R.H.P. Aquaculture pond ecosystem model: temperature and dissolved oxygen prediction mechanism and application. *Ecol Model* **89,** 231–258 (1996).
13. Hathurusingha, P. I. & Davey, K. R. A predictive model for taste taint accumulation in Recirculating Aquaculture Systems (RAS) farmed-fish – demonstrated with geosmin (GSM) and 2-methylisoborneol (MIB). *Ecol Model* **291,** 242–249 (2014).
14. Ferreira, N. C., Bonetti, C. & Seiffert, W. Q. Hydrological and Water Quality Indices as management tools in marine shrimp culture. *Aquaculture* **318,** 425–433 (2011).
15. Preis, A. & Ostfeld, A. A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. *J Hydrol* **349,** 364–375 (2008).
16. Mosley, L. M. *et al.* Predictive modelling of pH and dissolved metal concentrations and speciation following mixing of acid drainage with river water. *Appl Geochem* **59,** 1–10 (2015).
17. Johnsson, O., Sahlin, D., Linde, J., Lidén, G. & Hägglund, T. A mid-ranging control strategy for non-stationary processes and its application to dissolved oxygen control in a bioprocess. *Control Eng Pract* **42,** 89–94 (2015).
18. Cassidy, R. & Jordan, P. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. *J Hydrol* **405,** 182–193 (2011).
19. Li, C. & Hu, J. A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting. *Eng Appl Artif Intel* **25,** 295–308 (2012).
20. Hatvani, I. G., Kovács, J., Kovács, I. S., Jakusch, P. & Korponai, J. Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. *Ecol Eng* **37,** 629–635 (2011).
21. Abdoos, A., Hemmati, M. & Abdoos, A. A. Short term load forecasting using a hybrid intelligent method. *Knowl-Based Syst* **76,** 139–147 (2015).
22. Holger R Maier, G. C. D. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ Modell Softw* **15,** 101–124 (2000).
23. Hatzikos, E. V., Tsoumakas, G., Tzanis, G., Bassiliades, N. & Vlahavas, I. An empirical study on sea water quality prediction. *Knowl-Based Syst* **21,** 471–478 (2008).
24. Ma, Z., Song, X., Wan, R., Gao, L. & Jiang, D. Artificial neural network modeling of the water quality in intensive Litopenaeus vannamei shrimp tanks. *Aquaculture* **433,** 307–312 (2014).
25. Wu, Q. & Law, R. An intelligent forecasting model based on robust wavelet ν-support vector machine. *Expert Syst Appl* **38,** 4851–4859 (2011).
26. Lu, C. *et al.* Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif Intell Med* **28,** 281–306 (2003).
27. Kisi, O. & Cimen, M. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J Hydrol* **399,** 132–140 (2011).
28. Wang, D., Wang, M. & Qiao, X. Support vector machines regression and modeling of greenhouse environment. *Comput Electron Agr* **66,** 46–52 (2009).
29. Suykens, J. A. K., J, D. B. L., Lukas, L. & Vandewalle, J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* **48,** 85–105 (2002).
30. Liao, R., Zheng, H., Grzybowski, S. & Yang, L. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electr Pow Syst Res* **81,** 2074–2080 (2011).
31. Long, B., Xian, W., Li, M. & Wang, H. Improved diagnostics for the incipient faults in analog circuits using LSSVM based on PSO algorithm with Mahalanobis distance. *Neurocomputing* **133,** 237–248 (2014).
32. Kennedy, J. & C, E. R. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*. Perth, Australia. **4**, 1942–1948 (1995).
33. Selakov, A., Cvijetinović, D., Milović, L., Mellon, S. & Bekut, D. Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. *Applied Soft Computing* **16,** 80–88 (2014).
34. Abdi, M. J. & Giveki, D. Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. *Eng Appl Artif Intel* **26,** 603–608 (2013).
35. Subasi, A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Comput Biol Med* **43,** 576–586 (2013).
36. Kundu, R., Das, S., Mukherjee, R. & Debchoudhury, S. An improved particle swarm optimizer with difference mean based perturbation. *Neurocomputing* **129,** 315–333 (2014).
37. He, S., Wu, Q. H., Wen, J. Y., Saunders, J. R. & Paton, R. C. A particle swarm optimizer with passive congregation. *Biosystems* **78,** 135–147 (2004).
38. Kim, J. S. & Jung, S. Implementation of the RBF neural chip with the back-propagation algorithm for on-line learning. *Applied Soft Computing* **29,** 233–244 (2015).
39. Shen, W., Guo, X., Wu, C. & Wu, D. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowl-Based Syst* **24,** 378–385 (2011).
40. Chai, W. & Qiao, J. Passive robust fault detection using RBF neural modeling based on set membership identification. *Eng Appl Artif Intel* **28,** 1–12 (2014).
41. Lin, G. & Chen, L. A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. *J Hydrol* **288,** 288–298 (2004).
42. Vapnik, V. The nature of statistical learning theory. Springer Science & Business Media, *New York* (2000).
43. Suykens, J. A. K., De Brabanter, J., Lukas, L. & Vandewalle, J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* **48,** 85–105 (2002).
44. Nieto, P. J. G. *et al.* A hybrid PSO optimized SVM-based method for predicting of the cyanotoxin content from experimental cyanobacteria concentrations in the Trasona reservoir: A case study in Northern Spain. *Appl Math Comput* **260,** 170–187 (2015).
45. Kennedy, J., Eberhart, R. C. & Shi, Y. (eds.) *Swarm Intelligence*. Morgan Kaufmann Publshers Inc, San Francisio,CA (2001).
46. García Nieto, P. J., García-Gonzalo, E., Alonso Fernández, J. R. & Díaz Muñiz, C. A hybrid PSO optimized SVM-based model for predicting a successful growth cycle of the Spirulina platensis from raceway experiments data. *J Comput Appl Math* **291,** 293–303 (2016).
47. Gholgheysari Gorjaei, R., Songolzadeh, R., Torkaman, M., Safari, M. & Zargar, G. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. *Journal of Natural Gas Science and Engineering* **24,** 228–237 (2015).

Acknowledgements

The authors thank native English-speaking expert Dr. S. G. Hassan from China Agricultural University for reviewing the language fluency of our paper. The work in this paper was supported by the National Natural Science Foundation Framework Project (No. 61571444) and the National Natural Science Foundation of China (61471133).

Author Contributions

H.Y. and Y.C. collected the image data. H.Y., Y.C. and D.L. designed the algorithm for processing images. H.Y. and Y.C. wrote the paper. H.Y., Y.C. and D.L. contributed to the analysis of the experimental data. Dr. S.G. Hassan reviewed the language in the paper.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Yu, H. *et al.* Dissolved oxygen content prediction in crab culture using a hybrid intelligent method. *Sci. Rep.* **6**, 27292; doi: 10.1038/srep27292 (2016).