IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 6, 2017, accepted January 17, 2018, date of publication February 9, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2800530

Fault Diagnosis of Water Quality Monitoring Devices Based on Multiclass Support Vector Machines and Rule-Based Decision Trees

SHUANGYIN LIU[1], LONGQIN XU[1], QIUCHENG LI[2,3], XUEHUA ZHAO[4], DAOLIANG LI[3]
[1]College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
[2]Advanced System Technology (AST) Branch of Fraunhofer IOSB, 98693 Ilmenau, Germany
[3]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[4] School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China

Corresponding author: Qiucheng Li (qiucheng.li@iosb-ast.fraunhofer.de)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471133, Grant 61571444, and Grant 61473331, in part by the Guangdong Science and Technology Plan of Project under Grant 2017B010126001, Grant 2017A070712019,Grant 2016A040402043, Grant 2015A070709015, Grant 2015A020209171, Grant 2016B010125004, and Grant 2014B040404070, in part by the Foundation for High-level Talents in Higher Education of Guangdong Province under Grant 2016KZDXM001 and Grant 2016GCZX001, in part by the Guangdong Natural Science Foundation under Grant 2016A030310072, and in part by the China Scholarship Council.

ABSTRACT Preventing faults of sensors, wireless transmitters, and gateways are essential for water quality management in intensive aquaculture. It remains a challenging task to achieve high fault diagnostic accuracy of water quality monitoring and controlling devices. This paper proposes a hybrid water quality monitoring device fault diagnosis model based on multiclass support vector machines (MSVM) in combination with rule-based decision trees (RBDT). In the modeling process, an RBDT is used to diagnose the gateway fault and wireless transmitter fault at the same time as a feature selection tool to reduce the number of features. A multiclass support vector machine classifier is employed to diagnose the faults of water quality sensors due to its robustness and generalization. We adopted an RBDT-MSVM algorithm to construct a fault diagnosis model. The diagnostic results indicate that RBDT-MSVM model has great potential for fault diagnosis of online water quality devices. RBDT-MSVM was tested and compared with other algorithms by applying it to diagnose faults of water quality monitoring devices in river crab culture ponds. The diagnostic results indicate that the model has great potential for fault diagnosis of online water quality devices. The experimental results show that the proposed model RBDT-MSVM can achieve classification accuracy as high as 92.86%, which is superior to the other three fault diagnosis methods. The results clearly confirm the superiority of the developed model in terms of classification accuracy, and that it is a suitable and effective method for fault diagnosis of water quality monitoring devices in intensive aquaculture.

INDEX TERMS Fault diagnosis, MSVM, rule-based decision tree, wireless sensor networks, aquaculture, water quality.

I. INTRODUCTION

In China, more and more aquaculture industries are equipped with water quality monitoring systems, such as dissolved oxygen sensors, pH sensors and water temperature sensors in river crab culture [1], [2]. The accuracy and reliability of sensors is the key to realize the optimal rearing strategies and to ensure the safety of river crabs. Although the water quality monitoring devices are well constructed and robust, the possibility of faults is inherent due to stresses involved because online water quality sensors stay below the water, and the solar-powered wireless transmitters are exposed in the wild. Some devices will inevitably malfunction after a relative long-term operation [3]. The faults of devices in the system may not only cause the misinterpretation of operations, but also increase costs, decrease product quality and affect the profit of farmers. Early detection of incipient faults can minimize break down loss and reduces maintenance time. Furthermore, the availability and reliability of devices will also be increased. More and more water quality monitoring systems in aquaculture are used in China. Early detection of device faults can not only ensure the safety of river crab, but also save a lot of maintenance cost [4]. Consequently, there has been an increasing demand for automated predictive maintenance and fault diagnosis system.

The most common faults of water quality monitoring devices are gateway faults, wireless transmitter faults, and sensor faults [5], [6]. Different approaches for fault diagnosis of wireless sensor networks have been successfully proposed. Most of these techniques involve comparisons between neighboring nodes and using distributed algorithms. Most of current fault detection and diagnosis systems are based on the DFD (Distributed Fault Detection) algorithm, which checks out the failed nodes by data exchanging and mutual testing among intra—network neighbor nodes [7], [8].

The DFD algorithm has proved to be a reliable technique to diagnose the condition of wireless sensor networks which has the property of neighboring similarity. In this paper, because different nodes are in different river crab culture ponds whose water conditions vary a lot, the quantities of different dissolved oxygen nodes are different which doesn't qualify the requirement of DFD. However, different nodes share the same change flow of the quantity of dissolved oxygen. Experts systems have been developed for fault diagnosis, for example Qian *et al.* [9] used an expert system for real-time fault diagnosis of complex chemical processes. However, if a system is very complex, many fault diagnosis rules are generated and it is difficult to manage them.

Intelligent methods, like artificial neural network and SVM, have also been developed for fault diagnosis [5], [10]–[17]. Artificial neural networks have been widely used, for example, Jahromi *et al*. [10] applied a sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis. Xiao *et al*. [11] proposed a fault detection and diagnosis method of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model.However, the neural network approach suffers from a number of weaknesses, including the requirement of a large number of controlling parameters, difficulty in obtaining a stable solution, and the danger of over-fitting. The support vector machines (SVM) technique can overcome these weaknesses. For example, Yang *et al*. [5] proposed a fault diagnosis method forwater quality monitoring and control equipment in aquaculturebased on multiple SVM combined with D-S evidence theory. Chang *et al*. [15] applied afault diagnosis of a mine hoist based on PCA and support vector machines. Gao and Hou [16] proposed an improved support vector machines and GS-PCA for fault diagnosis approach of Tennessee Eastman process. SVM was developed for recognizing patterns or discriminating between two groups by Vapnik [18]. SVM represents some of the most advanced pattern recognition platforms today. SVM achieves an optimum network structure by striking a proper balance between empirical error and the Vapnik-Chervonenkis (VC) confidence interval [15]–[17]. In general, the performance of SVM is better than ANN, and the solutions reached by SVM areunique, optimum, and absent from local minima. In this study, SVM is used fordiagnosis of sensor fault to overcome weaknesses of artificial neural network.

Online water quality sensors need wireless transmitters and gateways to supply power or send out the data, so the fault diagnosis of wireless transmitters and gateways are fundamental to the fault diagnosis of sensors. Although the quantities of different dissolved oxygen nodes are different, different nodes shared the same kind of wireless transmitters which provide sensor nodes with power supply and communication ability. In this study, rule-based decision trees (RBDT) is used to detect gateway fault and wireless transmitter fault, at the same time, RBDT is usedas a feature selection procedure to remove irrelevant features for reducing the amount of data needed to achieve good learning, classification accuracy, and a reduction in computational time for diagnosis of sensor faults [5].

The proposed approach consists of two stages. First, the rule-based decision trees (RBDT) diagnose the gateway fault and wireless transmitter fault in the meantime as a feature selection tool for reducing the number of features. Secondly, the MSVM classifier is used to diagnose the faults of water quality sensors.

The structure of this paper is organized as follows. Section 2 introduces decision tree, support vector machine, and multiclass classification. Section 3 presents the background of fault diagnosis of water quality monitoring devices. Section 4 presents the fault diagnosis model of water quality monitoring devices based on RBDT-MSVM. In section 5, details of the experimental environment, data acquisition, feature extraction, and results are given. Conclusions and proposals for future work are summarized in section 6.

II. RELATEDMETHODOLOGY

A. DECISION TREE

Decision tree, a popular tool in machine learning, is an efficient tool for classificationproblems [19]. Unlike other classification approaches that use a set of features (or bands) jointly toperform classification in a single decision step, the decisiontree is based on a multistage or hierarchical decision schemeor a tree like structure. Thetrees consist of internal nodes (with two children) and terminal nodes or leaf nodes (withoutchildren). Each internal node is associated with a decisionfunction to indicate which node to visit next, while teach terminal node shows the output of a given input vectorthat leads the visit to this node [20]. Each node of the decision treestructure makes a binary decision that separates either oneclass or some of the classes from the remaining classes. Theprocessing is generally carried out by moving down the treeuntil the leaf node is reached. This is known as a top–down approach.

Traditionally, Decision trees consistof two phases, the first phase is called tree building, and the other is tree pruning [21], which is constructed from existing data to classify future data.

The construction of a decision tree is based on binary recursive partitioning, which is an iterative process that splits the datainto partitions.Initially, all the training samples are used todetermine the structure of the tree [20]. The algorithm then breaks the data using every possible binary split and selects the split that partitions the data into two parts such that it minimizes the sum of the squared deviations from the meanin the separate parts. The splitting process is then applied to each of the new branches. The process continues until each node reaches a user-specified minimum node size (i.e., the number of training samples at the node) and becomes aterminal node.

Since the tree is constructed from training samples, it may suffer from over-fitting when the full structure isreached. This may deteriorate the classification accuracy of the tree when applied on unseen data and hence may lead to less generalization ability. Therefore, a pruning process is generally adopted using a validation data set and the user specified cost complexity factor [22], [23].

In this paper, a set of expert rules were used in supporting the tree building and substituting the tree pruning for detecting and diagnosing faults. Those rules and thresholds are from experienced maintenance technicians, and are utilized to detect faulty gateways and faulty wireless transmitters. A decision tree is also utilized as feature selection procedure to remove irrelevant features for reducing the amount of data needed to achieve good learning, classification accuracy, and a reduction in computational time for diagnosis of sensor fault. It will be described in details in the next section.

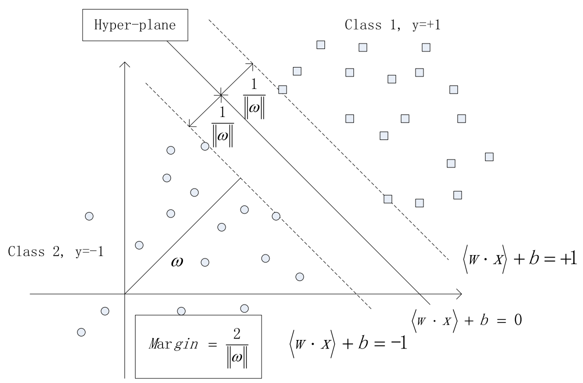

FIGURE 1. Optimal separating hyper-plane with the largest margin.

B. SUPPORT VECTOR MACHINE

Support vector machine (SVM) was firstly proposed by Vapnik and Lerner at 1995 to solve non-linear model [18]. The SVM algorithm is based on the statistical learning theory and the Vapnik–Chervonenkis (VC) dimension [16], [17]. SVM analysis seeks to find an optimal separating hyper-plane by maximizing the margin between the separating data, as illustrated in Fig. 1. Let $\{x_i, y_i\}_{i=1}^{n}$ be the training sample set, where $X_i \in R^p$ is the input vector, $y_i \in \{-1, 1\}$ is the class labels. SVMs try to find an optimal hyper-plane $< w \cdot x > +b = 0$, where w is the normal to the optimal hyper-plane, and b is a scalar threshold [24].

The optimal hyper-plane can be found by solving the following constrained optimization problem:

$$\begin{aligned}&\text{Minimize } \frac{1}{2}\|w\|^2\\ &\text{Subject to } y_i\cdot < w\cdot x + b > \geq 1, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{1}$$

SVM is trained as a quadratic optimization problem[18]:

$$R = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j \langle x_i \cdot x_j\rangle \tag{2}$$

Is maximized subject to $\sum_{i=1}^{n} a_i y_i = 0$ and $C \geq a_i \geq 0$, where C is the penalization parameter, which is used to control trade-off between the training error and the margin.

When the above optimization problem is solved, the normal to the optimal hyper-plane, W, can be computed by

$$W = \sum_{i=1}^{n} a_i y_i x_i \tag{3}$$

Thus, by solving the dual optimization problem, the non-linear decision function is gained as follows:

$$f(x) = sign(\sum_{i=1}^{n} y_i a_i \langle x_i \cdot x\rangle + b \tag{4}$$

There exists a kernel function $k(x_i, x)$ which can make SVM realize nonlinear classification. The value of $k(x_i, x)$ equals to $\varphi(X_i)\cdot\varphi(\mathrm{X})$, where $\varphi(\cdot)$ is the transformation function, which makes the input data into higher-dimensional feature space. Then the non-linear decision function of SVM is described as below:

$$f(x) = sign(\sum_{i=1}^{n} y_i a_i k(x_i, x) + b \tag{5}$$

In Eq. (5), $k(x_i, x)$ is the kernelfunction which satisfies Mercer's condition corresponding to a dot product in some feature spaces. Four common Mercer kernel functions [25], where d, γ, and σ are constants, are listed in Table 1.

TABLE 1. Common kernel functions.

Name	Function Expression
Linear Kernel	$K(x_k, x) = x_k^T x$
Polynomial Kernel	$K(x_k, x) = (x_k^T x / \sigma^2 + \gamma)^d$
RBF Kernel	$K(x_k, x)=\exp(-\|x_k - x\|^2/\sigma^2)$
Sigmoid Kernel	$K(x_k, x) = \tanh(\gamma x_k^T x + \gamma)$

C. MULTICLASS CLASSIFICATION

SVM was originally designed for binary classification. Nevertheless, fault diagnosis of water quality monitoring devices usually falls into multiclass classification. Several methods are proposed to effectively extend it for multi classification such as "one-against-all" and "one-against-one" [26], [27], while more binary SVMs and training amount are required. Li *et al*. indicated that the one-against-other (OAO) approach requires a shorter training time [26]; hence, in our study, OAO is used for fault diagnosis of water quality sensor. For a data set with k different classes, the step of method constructs k (k - 1)/2 binary SVMs as follows. Step 1: For any new test data X, put it into each classifier to obtain a result of classification. When a binary classifier (Class i and class j) indicates the data set X should be class i, the class i gets one vote. Step 2: Make a decision with the classwhich has the greatest number of votes.Step 3: Repeat these steps till last binary SVM classifier is constructed. In this way, multiclass classification based on SVM may be constructed for a case of k class classification problem.

III. PROBLEM STATEMENT AND DATA PREPARATION

In this study, we developed a fault diagnosis system and diagnosed newly appearing faults precisely and rapidly. A water quality monitoring system is selected for study. The water quality monitoring system proposed in this paper is used in outdoor river crab aquaculture to ensure the safety of river crabs by monitoring the quantity of dissolved oxygen (DO) in the river crab pond and according to the quantity of DO automatically controlling the aerator. If the water quality monitoring system fails for a long time without any warning, the river crab may be in danger, and one year's effort of river crab culture farmers may be in vain. The sensors, actuators, wireless transmitters, gateways and remote servers compose the water quality monitoring and controlling system. The sensor is paired with a wireless transmitter which contains two main blocks: (a) a power unit consisting of a solar power panel and a rechargeable battery; it is for powering both the sensor and the transmitter and (b) a wireless communication module; it is for communicating with the gateway (see functional diagram shown in Fig. 2, the system typology structure diagram is shown in Fig.3).

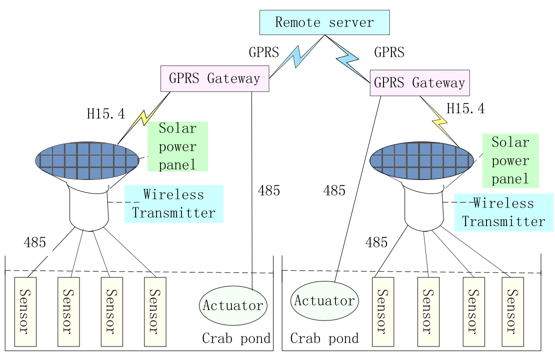

FIGURE 2. Functional diagram of water quality monitoring system.

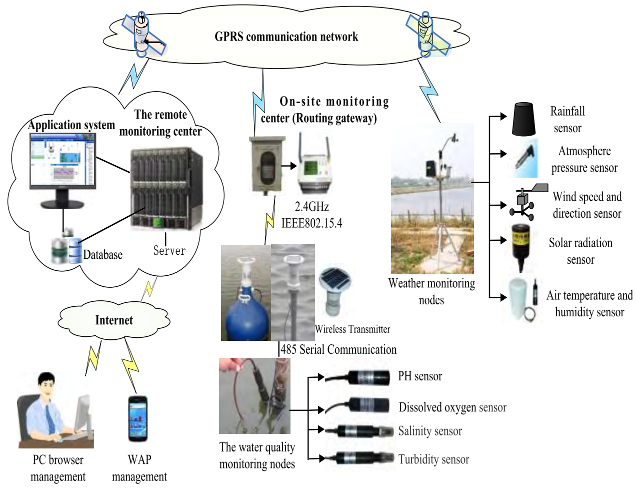

FIGURE 3. The topology structure diagram of the water quality monitoring system.

TABLE 2. Data channels of water quality monitoring devices.

Channel Number	Channel Name	Acquisition Equipment
1	City power supply	Gateway
2	Current one	Gateway
3	Current two	Gateway
4	Switch signal	Gateway
5	Analog signal one	Gateway
6	Analog signal two	Gateway
7	Device voltage	Gateway
8	Device information	Wireless transmitter
9	Device voltage	Wireless transmitter
10	Signal strength	Wireless transmitter
11	Dissolved oxygen	Water quality sensor
12	water temperature	Water quality sensor
13	DO engineering value	Water quality sensor

When the sensor wants to sense the water quality parameters in the river crab pond, actions of the water quality monitoring and controlling system proceed as follows: (a) power supply unit of wireless transmitter supplies power for the sensor; (b) sensor probe obtains power; (c) sensor probe senses the water quality parameters; (d) wireless transmitter acquires those parameters through 485 serial communication; (e) wireless communication module of wireless transmitter sends those data to gateway via H15.4 wireless communication protocol; (f) gateway processes those data and control the actuator if necessary; (g) gateway send those data and control history to the remote server. To keep a suitable water

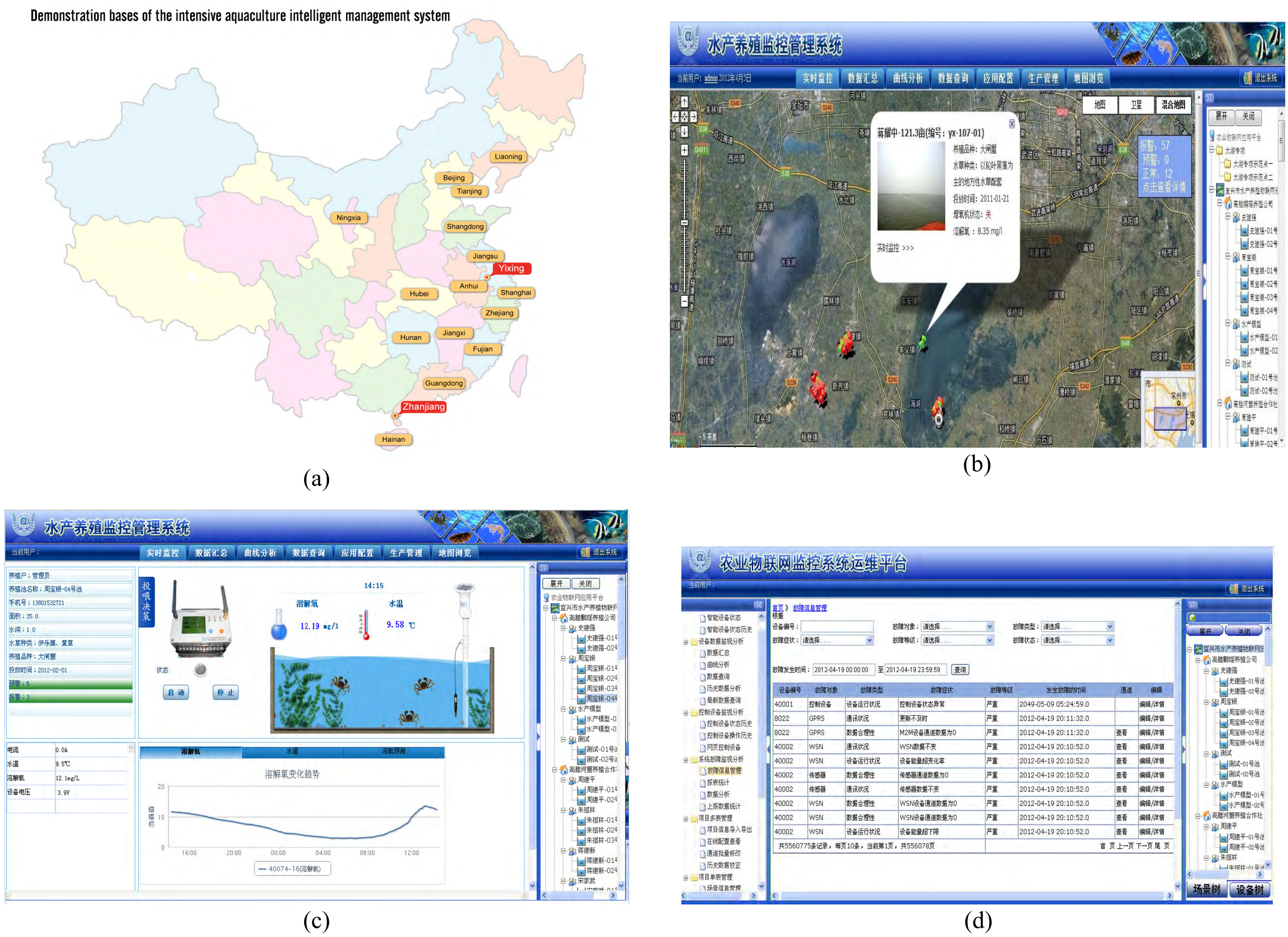

FIGURE 4. Digital Wireless Monitoring and Controlling System of Aquaculture Water Quality: (a) the distribution map of Digital Wireless Monitoring System of Aquaculture Water Quality, (b) aerial view of experiment base, (c) operation system interface, and (d) maintenance systems interface.

quality for the river crab in the pond, it's important that each operational step described above should proceed normally. When a failure occurs, the value of the sensor couldbe out of normal range or not acquired. Traditionally, the technician uses an electronic meter and a portable computer to obtain parameters of operational equipment, reads the maintenance handbook, and then diagnoses reason(s) for the fault(s). This process involves considerable time, and the maintenance efficiency depends on the ability and experience of the technician. In order to preclude human error and take advantage of expert domain knowledge in fault diagnosis, rule-based decision tree and multiclass SVM are introduced as reliable means for fault diagnosis in this paper.

Thirteen common faults are selected from the operational processes described above, classified as three broad types, including ‘‘gateway fault’’,‘‘wireless transmitter fault,’’ and ‘‘sensor fault.

The data used in this study have been produced by the water quality monitoring system. The system is installed in 69 river crab culture ponds in County of Yixing, Jiangsu Province, China. The software system is shownin Fig. 4. In this system, the water quality sensor has three channels, including dissolved oxygen, water temperature, and dissolved oxygen engineering value. The wireless transmitter has three channels, include device information, device voltage, and signal strength. The gateway has seven channels, including city power supply, current one, current two, switch signal, analog signal one, analog signal two, and device voltage. That information is the source of the fault diagnosis system (Table 2).

In this study, 4 towns and 69 river crab farms were monitored. Each experiment pond was about 20,000 m^2, and the water level is1 to 2 m. The original data was collected to the remote server every ten minutes. The actual picture of the digital wireless monitoring system is shown in Fig.4.

IV. FAULT DIAGNOSIS OF WATER QUALITY MONITORING DEVICES BASED ON MULTICLASS SUPPORT VECTOR MACHINES AND RULE-BASED DECISION TREE MODEL BUILDING

Rule-Based Decision Tree (RBDT) detects gateway faults and wireless transmitter faults as a reduction in computational

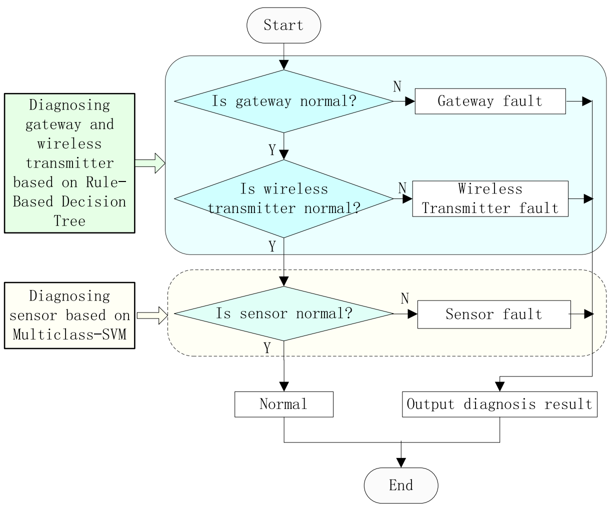

FIGURE 5. Flow chart of fault diagnosis system for water quality monitoring devices.

time for diagnosis of sensor faults. If the diagnosis result of RBDT is that both gateway and wireless transmitter are normal, Multiclass SVM will be utilized to diagnosethe sensors. The flow chart is shownin Fig. 5.

A. FAULT DIAGNOSIS OF GATEWAY AND WIRELESS TRANSMITTER BASED ON RULE-BASED DECISION TREE

As shownin Fig. 2 and Fig. 3, the structure of the water quality monitoring and controlling system is a tree, and the data flow is pretty clear, an up-down approach, from sensor to wireless transmitter to gateway. Decision Tree is used to diagnosethe gateway, if the gateway is normal, then diagnosethe wireless transmitter, if the wireless transmitter is normal, then diagnose the sensor using SVM. Domain knowledge of experienced maintenance technicians has been used in every binary split to build the decision tree (Fig. 6).

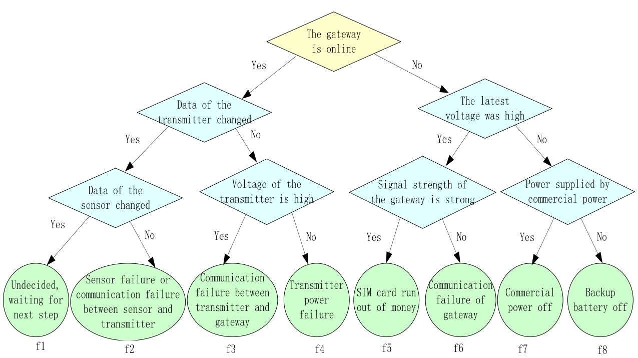

FIGURE 6. Decision Tree.

The rules generated from the decision tree in Fig. 6 are as follows.

Rule 1: IF (Gateway status = Offline) AND (The latest Gateway power status = high) AND (The latest Gateway communication status = strong)

THEN Fault type is f5, SIM card inside of gateway runs out of money

Rule 2: IF (Gateway status = Offline) AND (The latest Gateway power status = high) AND (The latest Gateway communication status = weak)

THEN Fault type is f6, communication failure of gateway communication module

Rule 3: IF (Gateway status = Offline) AND (The latest Gateway power status = low) AND (The power supplier = commercial power)

THEN Fault type is f7, commercial power of gateway is off

Rule 4: IF (Gateway status = Online) AND (The latest Gateway power status = low) AND (The power supplier = backup battery)

THEN Fault type is f8, backup battery runs out of power

Rule 5: IF (Gateway status = Online) AND (The wireless transmitter power status = high) AND (The wireless transmitter communication status = strong)

THEN Fault type is f1, Undecided, waiting for next step

Rule 6: IF (Gateway status = Online) AND (The wireless transmitter power status = high) AND (The wireless transmitter communication status = strong)

THEN Fault type is f2, sensor fault

Rule 7: IF (Gateway status = Online) AND (The wireless transmitter power status = high) AND (The wireless transmitter communication status = weak)

THEN Fault type is f3, communication failure of the wireless transmitter communication module

Rule 8: IF (Gateway status = Online) AND (The wireless transmitter power status = low) AND (The wireless transmitter communication status = strong)

THEN Fault type is f4, power failure of the wireless transmitter.

B. FAULTDIAGNOSIS OF WATER QUALITY SENSOR BASED ON MULTICLASS SVM

In this paper, the approach based on multi-SVM classifier is proposed for fault diagnosis of water quality sensor.The implementation process of fault diagnosis of water quality sensor based on multiclass SVM is shown in Fig. 7. The fault diagnosis system consists of data acquisition, pretreatment data, feature extraction, fault diagnosis model, testing results, and application.

The implementation process of fault diagnosis of water quality senor based on MSVM can be described as follows:

Step 1: Data acquisition and normalization. Data acquired from water quality monitoring system. The data are normalized to improve the treatment effect according to the formula:

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{6}$$

where $x_{\max}$ is the maximum in the series data, $x_{\min}$ is the minimum in the series data.

Step2: Feature Extraction. Statistical methods are employed to data dimension reduction and extract feature of the data. Redundantinformation in the original data space can be excluded by statistical methods.

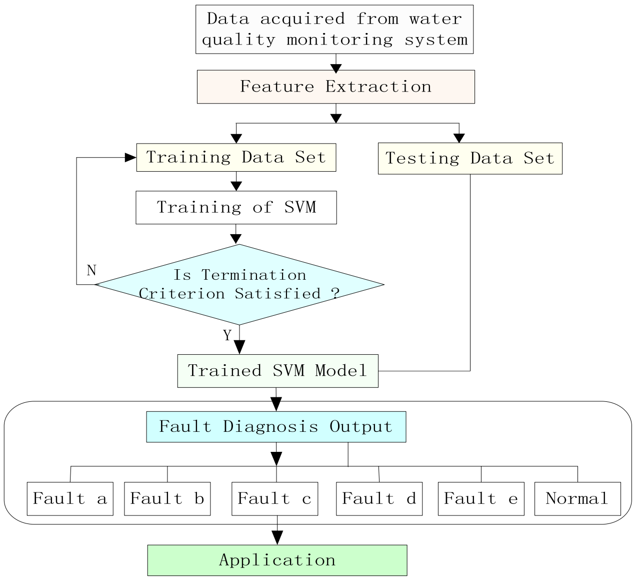

FIGURE 7. Flow chart of fault diagnosis of senor based on MSVM.

Step 3: Construct training sample set and test sample set, then form training patterns.

Step 4: Train MSVM on the training set, the parameter for MSVM-based classifier is determined by adopting cross validation method, and obtain the fault diagnosis model, and test the performance of the fault diagnosis model with test sample.

Step 5: For a new application of fault diagnosis task, extract fault diagnosis index of water quality sensor and form a set of input variables x. Then compute the estimation result $\hat{y}$ using Eq. (5).

V. EXPERIMENTS

A. DATA ACQUISITION

In our experiments, the multi-parameter sensor fusion which combines water temperature and dissolved oxygen is employed as research object, and five fault types of the sensor including (a) water temperature unchanged, (b) water temperature abnormal change, (c) rapid water temperature variation, (d) dissolved oxygen consistent high, (e) dissolved oxygen abnormal change. The local weather station supplied the weather parameterevery ten minutes, including air temperature and humidity, solar radiation, wind speed and direction, atmosphere pressure, and rainfall (Fig.3). In this paper, solar radiation was used to discrete the data flow collected by sensors. The original data was collected by devices in Fig.2 and Fig.3, every ten minutes from 69 river crab ponds, in County of Yixing, Jiangsu Province, China, from June to November. In this paper, we manually select faulty and normal data sets from that six-month test period. The 22 normal data sets are selected from monitoring points where the sensor, wireless transmitter and gateway all work normally more than one month. The 73 faulty data sets are collected from different monitoring points where faults happen, and the duration for different fault types varies from two days to one month. The measurement from the monitoring system updates every ten minutes. For example, for one month or 30 days, we have 4320 measurements in the data set.

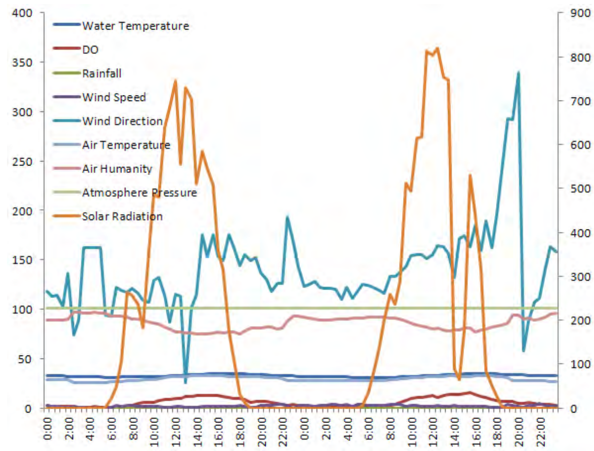

FIGURE 8. Values of the 9 measurements in the same time domain.

B. FEATURE EXTRACTION

The original data sets were recorded every ten minutes, it looks like that they were continuous in time series. They have to be cut into smaller meaningful data sets, so they can be used to perform fault diagnosis by analyzing the feature of different fault types. Fig. 8 shows a plot of the values of nine variables over the time, while Table 3 shows the correlation coefficient for each pair of variables. As showed in Table 3, the solar radiation had the highest correlation of coefficients, so the solar radiation was used to cut the origin data set to discrete smaller data sets.

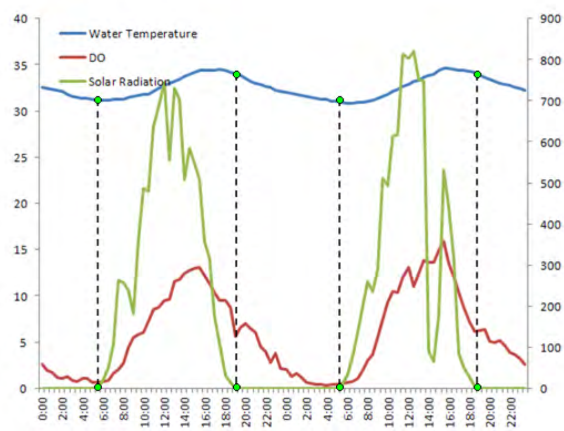

FIGURE 9. Time domain plots of DO, water temperature and solar radiation values acquired by normal sensors.

As showed in Fig. 9, the original time series was cut into smaller time series according to whether the value of solar radiation is zero or not. Each cut section of time series was used a minimal unit to extract feature. Fig. 10 shows

TABLE 3. Correlation of coefficients between the 7 weather variables and 2 sensor variables.

	Rainfall	Wind Speed	Wind Direction	Air Temperature	Air Humanity	Atmosphere pressure	Solar Radiation
Water Temperature	-0.0735	-0.2153	0.4203	0.8005	-0.7678	-0.3681	0.6075
Dissolved Oxygen	-0.1318	-0.1712	0.2572	0.8957	-0.8557	0.0335	0.9162

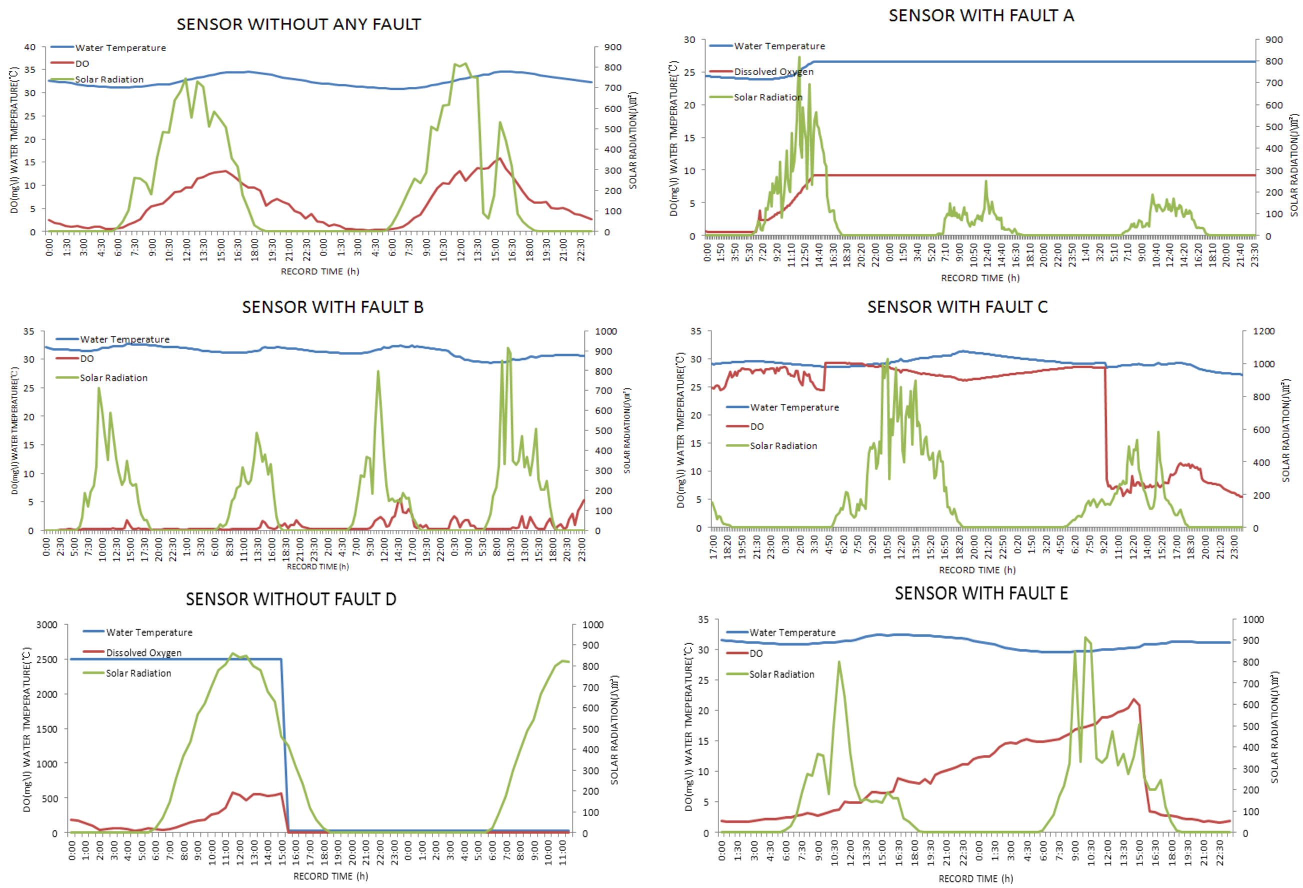

FIGURE 10. Time domain plots of water quality parameters combined with Solar Radiation Sequences.

the time domain sequences taken from water quality sensors for different conditions. The time domain signal can be used to perform fault diagnosis by analyzing sensor signals obtained from the experiment. Statistical methods used can provide the characteristics of time domain series. Statistical analysis of time series yields different descriptive statistical parameters. So we chose measures including standard error, standard deviation, sample variance, kurtosis, skewness and derivativeas criteria for extracting features from each cut section of time series. The definitions of the performance metrics and their calculations are summarized in Table 4.

The experimental input data is required to pre-treat to make the entries suitable for the multiclass SVM models. There are 10 necessary features selected to describe DO and water temperature respectively, hence there are 20 parameters selected in order to classify 5 kinds of sensor faults and one kind of normal situation that correspond to each kind of fault during training or testing. The pretreatment of input data is shown in Table 5.

TABLE 4. Performance indice and their calculations.

Statistical methods	Calculation
Standard error	$\sqrt{\frac{1}{\text{n-2}}[\sum(y-\bar{y})^2-\frac{[\sum(x-\bar{x})(y-\bar{y})]^2}{(x-\bar{x})^2}]}$
Standard deviation	$\sqrt{\frac{\sum x^2-(\sum x^2)}{\text{n(n-1)}}}$
Sample variance	$\sqrt{\frac{\sum x^2-(\sum x^2)}{\text{n(n-1)}}}$
Kurtosis	$\{\frac{n(n+1)}{\text{(n-1)}\ (n-2)(n-3)}\sum(\frac{x_i-\bar{x}}{s})^4\}-\frac{3(n-1)^2}{(n-2)(n-3)}$
Skewness	$\frac{n}{\text{(n-1)}}\sum(\frac{x_i-\bar{x}}{s})^3$
Derivative	$dp(n)=p(n)-p(n-1)$

TABLE 5. Pretreatment of input data for multiclass SVM.

X_1	X_2	X_3	X_4	X_5	X_6	…	X_{17}	X_{18}	X_{19}	X_{20}	Fault type
31..232	0.0937	31.733	30.781	31.251	1.917	…	0.079	0.952	343.560	0.306	no fault
30.445	0.0269	30.720	30.235	30.425	1.828	…	0.321	0.485	304.452	0.186	no fault
23.991	0.0032	24.0766	24.093	23.985	1.850	…	-0.060	0.183	407.841	0.060	no fault
23.592	0.0021	23.689	23.894	23.603	2.501	…	-0.028	0.176	424.664	0.466	no fault
…	…	…	…	…	…	…	…	…	…	…	…
26.560	0	26.560	26.560	26.560	…	…	…	0	956.154	0	(a)Water temperature unchanged
26.560	0	26.560	26.560	26.560	…	…	…	0	929.594	0	…
26.559	0	26.559	26.559	26.559	…	…	…	0	1009.274	0	…
…	…	…	…	…	…	…	…	…		…	…
2499.39	0	2499.39	2499.39	2499.39		…		0	14996.34	0	(b)Water temperature unchanged
2499.39	0	2499.39	2499.39	2499.39		…		0	12496.95	0	…
2499.39	0	2499.39	2499.39	2499.39		…		0	12496.97	0	…
…	…	…	…	…	…	…	…	…	…	…	…
25.703	9.0129	31.586	21.011	25.065	2.713	…	0.517	10.572	282.736	3.002	(c)rapid water temperature variation
25.849	3.8196	30.480	23.486	25.549	3.892	…	1.117	6.994	284.341	1.954	…
21.835	7.7874	27.177	19.403	20.633	2.393	…	1.032	7.774	240.182	2.791	…
19.908	8.6575	26.496	17.856	18.701	3.742	…	1.565	8.640	218.989	2.806	…
24.144	11.0147	28.628	20.130	23.089	1.778	…	0.231	8.498	120.718	3.416	…
…	…	…	…	…	…	…	…	…	…	…	…
29.422	0.0098	29.556	29.221	29.438	2.009	…	-0.384	0.335	853.247	0.099	(d) dissolved oxygen consistant high
28.864	0.03541	29.198	28.587	28.851	1.809	…	0.084	0.611	837.068	0.196	…
30.732	0.1066	31.299	30.189	30.731	1.792	…	-0.082	1.110	921.949	0.326	…
29.746	0.0586	30.136	29.351	29.746	1.709	…	0.067	0.784	862.642	0.251	…
24.880	0.2037	25.731	24.424	24.649	2.159	…	0.797	1.308	298.562	0.451	…
…	…	…	…	…	…	…	…	…	…	…	…
23.131	0.1686	23.705	22.565	23.154	1.504	…	-0.008	1.139	277.574	0.411	(e) change law of dissolved oxygen abnormal
22.327	0.0092	22.514	22.203	22.309	2.264	…	0.525	0.311	267.929	0.113	…
31.343	0.0157	31.515	31.181	31.337	1.725	…	0.097	0.334	188.056	0.125	…
31.024	0.0050	31.115	30.938	31.023	1.656	…	0.064	0.177	155.121	0.090	…
31.925	0.0789	32.248	31.354	31.989	2.546	…	-0.717	0.893	351.185	0.281	…
30.329	0.3086	31.173	29.673	30.110	1.549	…	0.363	1.499	333.623	0.607	…

After collecting and arranging the maintenance data available in a six-month period, there are 95 available samples for this study. We randomly assign these data sets to training and testing groups.

C. RESULTS AND DISCUSSION

The proposed RBDT-MSVM algorithm was implemented in the Matlab7.13.0564 programming language. The experiments are made on a 2.50 GHz Core (TM) i5-3210M CPU personal computer (PC) with 4.0G of memory running Microsoft Windows 7.

Classification is a two phase process: training and testing. Training is the process of learning to label from the examples. Training can be supervised mode or unsupervised mode. Here, supervised mode is used for training. Testing is the process of checking how well the classifier has learnt to label the unseen examples. The four different kernel functions, such as polynomial function, sigmoid functionand RBF of MSVM

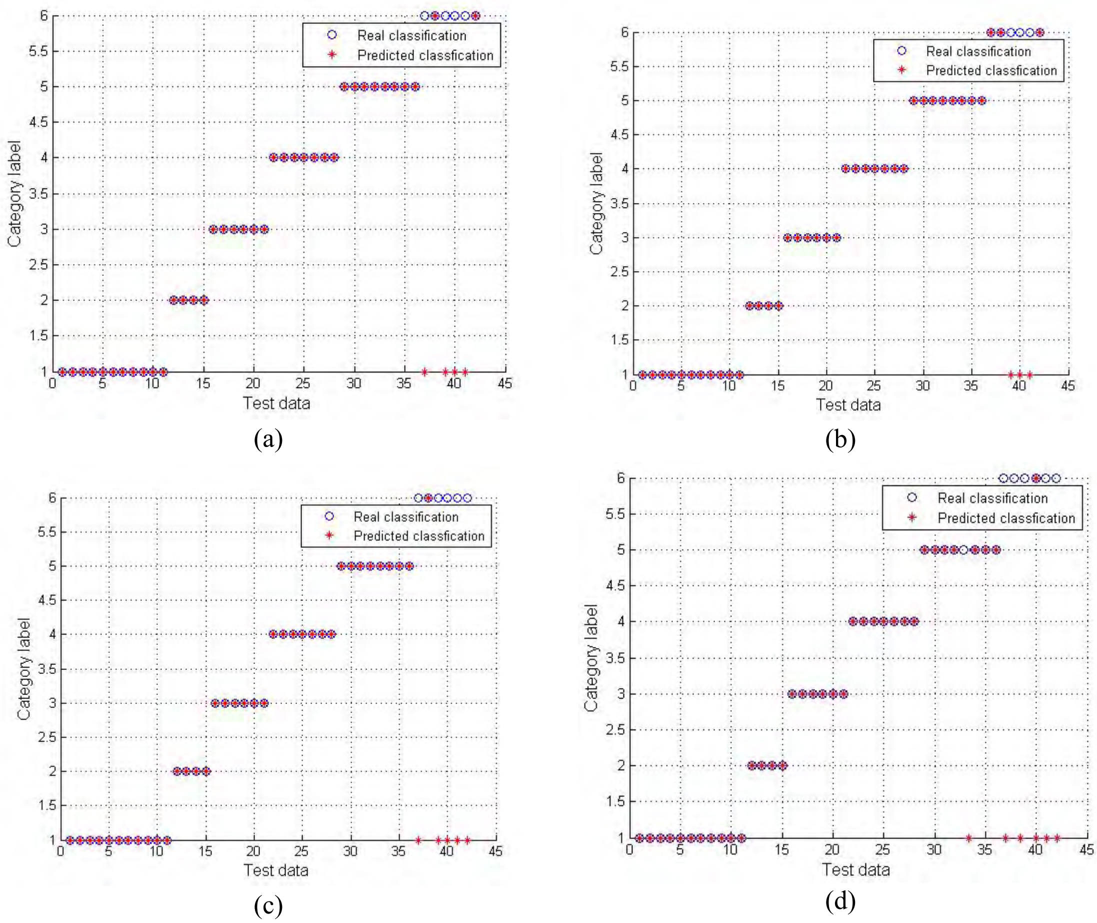

FIGURE 11. **Test result of SVM with four different kernel functions. (a) polynomial. (b) RBF. (c) sigmoid. (d) linear.**

TABLE 6. **Diagnostic accuracy statistical analysis on three models.**

No.	Classifier	Diagnostic accuracy
1	MSVM - sigmoid	88.0952
2	MSVM - polynomial	90.4762
3	MSVM - RBF	92.8571
4	MSVM -1inear	86.9814

were used for classification. The parameters of MSVM [26] were given as follows: penalty parameter $c = 10$, kernel parameter $g = -1$. The classification efficiency for the four different functions of MSVM is given in Table 6.

From Table 6, the MSVM models of all the kernel function tested yielded classification accuracy in excess of 86.9814%; RBF kernel functions provided accuracy of 92.8571% in fault classification. The testing results of the four MSVM classifiers for sensor faults are presented in the form of confusion matrix in Fig. 11.

The interpretation of the confusion matrix is as follows: The diagonal elements in the confusion matrix show the number of correctly classified instances. Output results are numbered directly from 1 to 6. For example, output "1" indicates the sensor is normal; output "2, 3, 4, 5, 6"respectively indicates fault"a, b, c, d, e". In Fig. 11, the RBF kernel function MSVM model classified can acquire good performance of classification and requires a little operation. The other three kernel function MSVM models classified the normal situation and fault "a, band c" correctly, and misclassified fault "dand e" as normal situation. Those MSVM models find that it was difficult to discriminate between fault e and normal situation.

To analyze and compare fault diagnosis performance, a standard BPNN model andstandard least square support vector regression (standard-LSSVR) were also tested to diagnosethe sample, and the results are shown in Table 7, which are not so good. The reason is that the number of fault samples is not large enough for the BPNN to be well trained to exploit all of its potential.

This study presents a fault diagnosis of water quality monitoring devices based on RBDT-MSVM. The results of application in water quality monitoring devices fault diagnosis demonstrate that the prediction method based on

TABLE 7. Survey of classifier performances in respect of kernel functions.

Classifier	Diagnostic accuracy (%)
MSVM - RBF	92.8571
BPNN	60.3468
standard-LSSVR	90.6237

RBDT-MSVM is both effective and feasible; this fault diagnosis is important for decision making regarding water quality management in river crab pondsso that the testing costs and production schedule can be optimized.

VI. CONCLUSION AND FUTURE WORK

Fault diagnosis of wireless sensor networks is one of the newest research areas in the field of water quality monitoring. Many researchers reported the fault diagnosis of sensors, but here sensors,wireless transmitters and gateways have been all considered.

Implementation of some other measures may improve the performance of RBDT-MSVM and accordingly the results of fault detection. For instance, rule-based domain knowledge has been employed to build the decision tree quickly and effectively, and has helped remove the irrelevant data for the diagnosis of sensor fault.

The six senor situations were classified using four different MSVM kernel functions of MSVM model in support vector machine. The RBF kernel model gives the best classification efficiency for six sensorsituations. The classification results and statistical measures were used for evaluating the RBDT-MSVM model. The total classification accuracy was 92.8571%. From the above result we can conclude that MSVM classifier with RBF kernel function combined with RBDT is a good candidate for fault diagnosis of water quality monitoring devices.

In future work, we intend to integrate the fault diagnosis algorithms of water quality monitoring devices based RBDT-MSVM within the intelligent water quality monitoring and controlling software platform, so that the alerting system will be able to issue early warnings based on the diagnosis results.

REFERENCES

[1] S. Liu, L. Xu, and D. Li, "Multi-scale prediction of water temperature using empirical mode decomposition with back-propagation neural networks," *Comput. Elect. Eng.*, vol. 49, pp. 1–8, Jan. 2016.

[2] S. Y. Liu *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. Agricult.*, vol. 95, pp. 82–91, Jul. 2013.

[3] L. Xu, S. Liu, and D. Li, "Prediction of water temperature in prawn cultures based on a mechanism model optimized by an improved artificial bee colony," *Comput. Electron. Agricult.*, vol. 140, pp. 397–408, Aug. 2017.

[4] X. Zhu, D. Li., D. He, J. Wang, D. Ma, and F. Li, "A remote wireless system for water quality online monitoring in intensive fish culture," *Comput. Electron. Agricult.*, vol. 71, pp. S3–S9, Apr. 2010.

[5] H. Yang, S. G. Hassan, L. Wang, and D. Li, "Fault diagnosis method for water quality monitoring and control equipment in aquaculture based on multiple SVM combined with D-S evidence theory," *Comput. Electron. Agricult.*, vol. 141, pp. 96–108, Sep. 2017.

[6] A. Lindhe, L. Rosén, T. Norberg, and O. Bergstedt, "Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems," *Water Res.*, vol. 43, no. 6, pp. 1641–1653, Apr. 2009.

[7] M.-J. Peng *et al.*, "Real-time simulations to enhance distributed on-line monitoring and fault detection in pressurized water reactors," *Ann. Nucl. Energy*, vol. 109, pp. 557–573, Nov. 2017.

[8] F. Boem, R. M. G. Ferrari, and T. Parisini, "Distributed fault detection and isolation of continuous-time non-linear systems," *Eur. J. Control*, vol. 17, nos. 5–6, pp. 603–620, 2011.

[9] Y. Qian, X. Li, Y. Jiang, and Y. Wen, "An expert system for real-time fault diagnosis of complex chemical processes," *Expert Syst. Appl.*, vol. 24, no. 4, pp. 425–432, May 2003.

[10] A. T. Jahromi, M. J. Er, X. Li, and B. S. Lim, "Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis," *Neurocomputing*, vol. 196, pp. 31–41, Jul. 2016.

[11] H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, "Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model," *Chemometrics Intell. Lab. Syst.*, vol. 161, pp. 96–107, Feb. 2017.

[12] H. A. Illias, X. R. Chai, and A. H. A. Bakar, "Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis," *Measurement*, vol. 90, pp. 94–102, Aug. 2016.

[13] J. C. M. Oliveira, K. V. Pontes, I. Sartori, and M. Embiruçu, "Fault detection and diagnosis in dynamic systems using weightless neural networks," *Expert Syst. Appl.*, vol. 84, pp. 200–219, Oct. 2017.

[14] Z. Gao, C. Ma, D. Song, and Y. Liu, "Deep quantum inspired neural network with application to aircraft fuel system fault diagnosis," *Neurocomputing*, vol. 238, pp. 13–23, May 2017.

[15] Y.-W. Chang, Y.-C. Wang, T. Liu, and Z.-J. Wang, "Fault diagnosis of a mine hoist using PCA and SVM techniques," *J. China Univ. Mining Technol.*, vol. 18, no. 3, pp. 327–331, Sep. 2008.

[16] X. Gao and J. Hou, "An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process," *Neurocomputing*, vol. 174, pp. 906–911, Jan. 2016.

[17] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.

[18] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[19] S. Sardari, M. Eftekhari, and F. Afsari, "Hesitant fuzzy decision tree approach for highly imbalanced data classification," *Appl. Soft Comput.*, vol. 61, pp. 727–741, Dec. 2017.

[20] Y. Cui, J. Shi, and Z. Wang, "Analog circuit fault diagnosis based on quantum clustering based multi-valued quantum fuzzification decision tree (QC-MQFDT)," *Measurement*, vol. 93, pp. 421–434, Nov. 2016.

[21] V. T. Tran, B.-S. Yang, M.-S. Oh, and A. C. C. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1840–1849, 2009.

[22] M. Xu, P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, "Decision tree regression for soft classification of remote sensing data," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 322–336, Aug. 2005.

[23] K. Kim and J.-S. Hong, "A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis," *Pattern Recognit. Lett.*, vol. 98, pp. 39–45, Oct. 2017.

[24] V. Kecman, "Support vector machines—An introduction," in *Support Vector Machines: Theory and Applications* (Studies in Fuzziness and Soft Computing), vol. 177, L. Wang, Ed. Berlin, Germany: Springer, 2005.

[25] V. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Comput.*, vol. 12, no. 9, pp. 2013–2036, Sep. 2000.

[26] D. Li, W. Yang, and S. Wang, "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine," *Comput. Electron. Agricult.*, vol. 74, no. 2, pp. 274–279, Nov. 2010.

[27] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si, "Fish species classification by color, texture and multi-class support vector machine using computer vision," *Comput. Electron. Agricult.*, vol. 88, pp. 133–140, Oct. 2012.

SHUANGYIN LIU received the Ph.D. degree from the College of Information and Electrical Engineering, China Agricultural University, in 2014. He is currently a Professor with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering. His current research interests are in the areas of intelligent information system of agriculture, artificial intelligence, software engineering, and computational intelligence.

LONGQIN XU received the M.S. degree from the Faculty of Computer, Guangdong University of Technology, in 2006. She is currently an Associate with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering. Her main research interests are related to intelligent information system of agriculture, artificial intelligence, machine learning, data mining, and computational intelligence.

QIUCHENG LI received the M.Sc. degree from the College of Information and Electrical Engineering, China Agricultural University, in 2012. He is currently pursuing the Ph.D. degree with the Faculty of Computer Science and Automation, Technical University of Ilmenau, Germany. At the same time, he works for the Fraunhofer IOSB-AST, Germany. His current research interests are in the areas of autonomous robotics, artificial intelligence, and computational intelligence.

XUEHUA ZHAO received the Ph.D. degree from the College of Computer Science and Technology, Jilin University, in 2014. He is currently a Lecturer with the School of Digital Medium, Shenzhen Institute of Information Technology. His main research interests are related to machine learning and data mining.

DAOLIANG LI received the Ph.D. degree from the College of Information and Electrical Engineering, China Agricultural University, in 1999.He is currently a Professor and a Ph.D. Supervisor with the College of Information and Electrical Engineering, China Agricultural University. His current research interests are in the areas of data mining, and knowledge engineering, with applications to agricultural engineering.