IET Journals

The Institution of Engineering and Technology

Alarms-related wind turbine fault detection based on kernel support vector machines

Yueqi Wu[1] ✉*, Xiandong Ma[1]*

[1]*Engineering Department, Lancaster University, Lancaster, LA1 4YW, UK*

✉ *E-mail: y.wu31@lancaster.ac.uk*

Abstract: Wind power is playing an increasingly significant role in daily life. However, wind farms are usually far away from cities especially for offshore wind farms, which brought inconvenience for maintenance. Two conventional maintenance strategies, namely corrective maintenance and preventive maintenance, cannot provide condition-based maintenance to identify potential anomalies and predicts turbines' future operation trend. In this study, a model based data-driven condition monitoring method is proposed for fault detection of the wind turbines (WTs) with SCADA data acquired from an operational wind farm. Due to the nature of the alarm signals, the alarm data can be used as an intermedium to link the normal data and fault data. First, KPCA is employed to select principal components (PCs) to retain the dominant information from the original dataset to reduce the computation load for further modelling. Then the selected PCs are processed for normal-abnormal condition classification to extract those abnormal condition data that are classified further into false alarms and true alarms related to the faults. This two-stage classification approach is implemented based on the KSVM algorithm. The results demonstrate that the two-stage fault detection method can identify the normal, alarm and fault conditions of the WTs accurately and effectively.

1 Introduction

With the increasing of electricity usage, wind power has become the world's fastest-growing renewable energy source. The wind turbines (WTs) installed capacity has been rising exponentially in past decades. From 2001 to 2017, the worldwide wind power installed capacity has increased from 23,900 to 539,581 MW, and new installed capacity in 2017 was 52,573 MW [1]. Due to the rich and stronger winds in the offshore areas, the installation of the WTs has been moved from onshore to offshore. The location of WTs, especially for offshore WTs, drives the operation and maintenance (O&M) cost to rise significantly. For an offshore WT which has a 20-year lifelong time, the O&M costs can be about 25–30% of the overall energy generation or 75–90% of the investment cost on O&M [2, 3]. Besides, the harsh operating environment will bring more difficulties for maintenance. There are two conventional maintenance strategies for the WTs, namely corrective maintenance and preventive maintenance [4]. However, the O&M costs from these two conventional strategies tend to be high when either little failures or a large number of failures occur. Hence, development of a condition-based and intelligent maintenance strategy for wind turbines would be significant and necessary to ensure a reliable, safe and cost-effective operation of the wind power systems.

This paper presents research results of a model-based data-driven WT fault detection method, which creates a relationship to identify the false alarms and true alarms related to the faults. The model is performed using the KSVM incorporating the KPCA based on the historical SCADA data. The alarm of the WT system can be triggered when key component signals exceed the pre-defined threshold limits usually due to design defects, changing of WT running states and components malfunction [5]. Since the alarms could reveal the working conditions of the turbine's components, it can be regarded as a significant index to indicate an early warning of the vital faults. Firstly, the computation load can be reduced by choosing specific principal components (PCs). Secondly, the chosen PCs are used to build the normal–abnormal classification model. Finally, a classification model based on the extracted abnormal data is built to classify the alarms and faults.

2 Methodology

2.1 PCs analysis (PCA)

The PCA transforms a set of correlated variables into a set of linearly uncorrelated variables, which are the PCs of the original dataset. It has been widely used to visualise relatedness and genetic distance between variables. The process can be achieved by calculating the eigenvalues of the covariance matrix or singular values of non-orthogonal matrix condition [6, 7]. PCA has shown its strong capability in dimension reduction and been verified by researches in different fields [8]. By selecting the first few PCs, the major information can be maintained and the dimension of the original dataset is then dramatically reduced. Hence, this technique has been widely applied in feature extraction and incorporated with various machine learning algorithms such as artificial neuron network (ANN) to monitor and predict the performance of WTs [9].

To obtain the PCs from a dataset X with n-by-p dimensions, where p is the number of the variables and n is the number of the samples of each variable, eigenanalysis for the covariance matrix M of original dataset X needs to be performed. First, the dataset X need to be standardised:

$$z_j = \frac{x_j - \bar{x}_j}{\sigma_{x_j}} \quad (j = 1, 2, \ldots, p) \tag{1}$$

where $\bar{x}_j$ is the mean value of x_j, σ_{x_j} is the standard deviation of x_j, and $Z = [z_1 z_2, \ldots, z_p]$ is denoted as the standardised dataset with n-by-p dimensions. The covariance matrix M of Z is defined as

$$M_{i,j} = \text{cov}(Z_i, Z_j) = E[(Z_i - \mu_i)(Z_j - \mu_j)]$$
$$= E[Z_i Z_j] - \mu_i \mu_j \tag{2}$$

where $\mu_i = E(Z_i)$ is the mean value of the ith row of Z. The PCs can be derived from the covariance matrix by using singular value decomposition (SVD). The singular values of the matrix M can be calculated by

$$M = USW^{\text{T}} \tag{3}$$

where S is an n-by-p rectangular matrix contains the ith singular values of M. U is an n-by-n matrix called the left singular vectors

consists of the n largest eigenvalues of MM^T and W^T is a p-by-p matrix called right singular vectors associated with the orthonormalised eigenvectors of M^TM [10]. By sorting the singular values in descending order and finding their corresponding singular vectors in the same order, the ith PC can be obtained by the following equation:

$$Y_i = U_{i1}z_1 + U_{i2}z_2 + \cdots + U_{ip}z_p \quad (i = 1, 2, \ldots, p) \tag{4}$$

The singular values of M are the variances of their corresponding PCs. Hence, the magnitudes of each singular value represent the weighted information contained in the original dataset. To select the number of PCs, the accumulated variance contributions from each PC need to be calculated. The contribution a_i of the variance s_i for the ith PC is defined as

$$a_i = \frac{s_i}{\sum_{i=1}^{p} s_i} \tag{5}$$

To obtain the information from the original dataset, the selection of k PCs should be as large as possible while still satisfying $k < p$. However, the number of PCs must be compromised to achieve dimension reduction. In our study, the accumulated variance contribution is selected no smaller than 85%.

2.2 Support vector machine (SVM)

The SVMs are set of supervised learning models that could be applied for regression and classification analysis with associated learning algorithms [11].

Because the original problem might be in a finite-dimensional space and might not be linearly separable in that space, it needs to be mapped into a much higher-dimensional space to make the separation much easier. An n-by-k training dataset Y can be considered as n points in k dimensions, implying each point Y_i ($i = 1, 2, \ldots, n$) contains k PCs. The training process for Y_i and its pre-defined class c_i are given in the form below:

$$(Y_1, c_1), \ldots, (Y_n, c_n), \quad c \in \{-1, 1\} \tag{6}$$

where c_i is either -1 or 1, indicating the class of the point Y_i. If any alarms are triggered at time instant i, the class of the Y is assigned to $c_i = -1$; otherwise $c_i = 1$. A hyperplane needs to be found to divide the overall samples into two classes. To satisfy this condition, the hyperplane should follow:

$$\begin{aligned} w \cdot Y_i + b &\geq 1 \quad \text{if } c_i = 1 \\ w \cdot Y_i + b &\leq -1 \quad \text{if } c_i = -1 \end{aligned} \tag{7}$$

The inequality (7) can also be written as

$$c_i(w \cdot Y_i + b) \geq 1 \tag{8}$$

where w is weight to the hyperplane and b is the bias. Points Y_0 for which $c_i(w \cdot Y_0 + b) = 1$ are named support vectors [11]. Therefore, the optimal hyperplane is described as

$$w_0 Y + b_0 = 0 \tag{9}$$

This hyperplane is unique that separates the training data with a maximal margin. The distance $\rho(w, b)$ between the projections of the training vectors of two different classes is thus given by

$$\rho(w, b) = \min_{\{Y: c = 1\}} \frac{Y \cdot w}{|w|} - \min_{\{Y: c = -1\}} \frac{Y \cdot w}{|w|} \tag{10}$$

The optimal hyperplane (w_0, b_0) is the arguments that maximise the distance. It follows:

$$\rho(w_0, b_0) = \frac{2}{|w|} = \frac{2}{\sqrt{w_0 \cdot w_0}} \tag{11}$$

The w needs to be minimised to satisfy the constraint defined by (11). The weights w_0 for the optimal hyperplane in the feature space can be written as a linear combination of support vectors

$$w_0 = \sum_{i=1}^{n} c_i a_i^0 Y_0 \tag{12}$$

where a_i^0 is the Lagrangian multiplier, which is to be described in (14).

Thus, the classification of an unknown vector Y is made by transforming a vector to the feature space ($Y \rightarrow \phi(Y)$) and then classified by the sign function:

$$f(Y) = w \cdot \phi(Y) + b \tag{13}$$

To satisfy the constraints (8), the Lagrangian multiplier is constructed as a standard optimisation technique

$$L(w, b, \Lambda) = \frac{1}{2} w \cdot w - \sum_{i=1}^{n} a_i[c_i(w \cdot Y_0 + b) - 1] \tag{14}$$

where $\Lambda^T = (\alpha_1, \ldots, \alpha_n)$ is the vector of non-negative Lagrange multipliers which satisfy the constraints defined by (8).

With (12), the classification function $f(Y)$ for an unknown vector Y can be extended to

$$f(Y) = w \cdot \phi(Y) + b = \sum_{i=1}^{n} c_i a_i \phi(Y)\phi(Y_0) + b \tag{15}$$

2.3 KPCA and KSVM

Both PCA and SVM could only solve linear separable problems. Hence, to solve a larger dataset with a linear inseparable problem, the kernel function is introduced. By using the kernel, the linear operations of PCA are performed in a reproducing kernel Hilbert space. Therefore, the linear inseparable problem can be solved by using kernel function projecting to a higher dimension.

KPCA is an extension version of the PCA using the kernel function to perform the originally linear operations in a reproducing kernel Hilbert space. As introduced above, the calculation of PCA can be transferred into the eigenanalysis. By mapping the original data into the feature space using the RBF (radial basis function) kernel. It is defined as

$$K(Z) = K(Z, Z^T) = e^{(-\gamma|Z - Z^T|^2)} \tag{16}$$

where Z is the original input dataset and Z^T is its transpose [2]. $|Z - Z^T|^2$ is considered as the squared Euclidean distance between them. The γ is the width of the kernel, which cannot be predicted precisely and has to be constrained by the model or defined by the user [12].

By replacing the original dataset with the kernel, the covariance matrix of (2) can be rewritten as

$$c = \frac{1}{n} \sum_{i=1}^{n} (K(z_i) - \bar{K})(K(z_i) - \bar{K})^T \tag{17}$$

$$\bar{K} = \frac{1}{n} \sum_{i=1}^{n} K(Z_i) \tag{18}$$

Then, following the same procedures as described by (3) and (4), the singular values and vectors, the kernelised PCs can be obtained.

Similar to the KPCA, the solution of KSVM also involves the transformation of the input dataset. In this case, the selected PCs
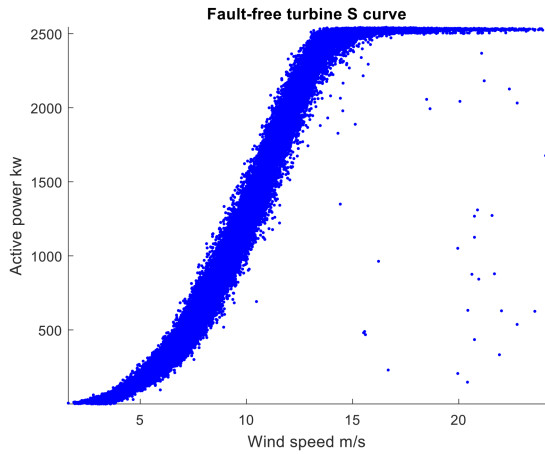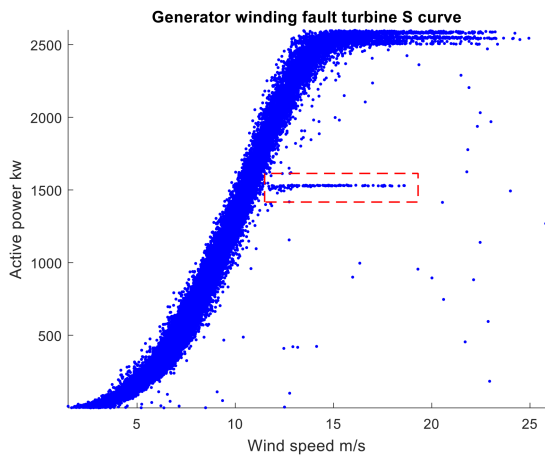
Fig. 1 *Power curve of the fault-free turbine*

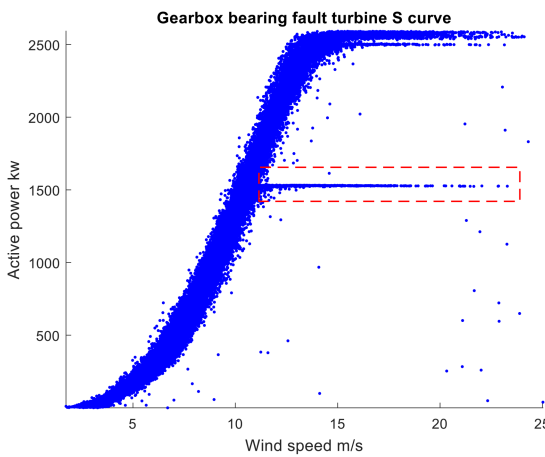**Fig. 2** *Power curve of the turbine with a generator winding fault*

Fig. 3 *Power curve of the turbine with gearbox bearing fault*

are employed, and the kernelised classification function, as derived from (15) can be written as

$$f(\boldsymbol{Y}) = \sum_{i=1}^{n} c_i \alpha_i K(\boldsymbol{Y}, \boldsymbol{Y}_0) = \sum_{i=1}^{n} c_i \alpha_i \mathrm{e}^{(-\gamma|Y-Y_0|^2)} \qquad (19)$$

where $\boldsymbol{Y}$ is the input data, which needs to be classified and $\boldsymbol{Y}_0$ are the support vectors.

3 SCADA data

The SCADA system is a data acquisition and control system that is used for high-level supervisory management through computers, graphical user interfaces and network data communications [13].

The SCADA data used in this paper were acquired from an operational wind farm which consists of 26 turbines over 12 months. To test and validate the proposed classification model, it is necessary to use historical data from an operational wind farm. Unlike the high-frequency condition monitoring data, SCADA data have a low sampling rate usually at 10 min/sample to reduce data storage amount while still maintaining the vital information about the operation and performance of the wind turbines [4]. The monitoring variables for each turbine consist of 128 readings among various types of physical and electrical signals, such as temperatures, pressures, power outputs and control signals. Pre-processing to the data is essential for further analysis due to the occasions that the turbines are in inactive during the periods of low and high speeds. Besides, the digital and constant data need to be removed to prevent inferences to the processing [14, 15].

As examples, Figs. 1–3 show the wind power curve of three different turbines. For wind turbines, the S-curve refers to the relationship between the output power and wind speed [16]. The output power would often be reduced when the fault occurs to prevent the fault from being developed into the detrimental one. The dashed box indicates the fault area. As can be observed from the figures, the turbine with a generator winding fault has a shorter time period of fault exposure compared to the turbine with a gearbox bearing fault.

To detect the faulty condition of the wind turbine, a two-stage classification method is proposed, as illustrated in Fig. 4. By checking time-series data, the original dataset includes data under the normal working condition and those alarm data. The alarm data also contain the fault data related to the alarms triggered during the fault period. Then abnormal data are further classified into the true positive signals, indicating the occurrence of a real fault, and false positive signals, which can be considered as a warning.

Three normal data selection methods are used in our study. The first one is to choose the first 5000 samples in the original dataset, which is referred as to method 1. The second method is to choose 2500 samples before and after the fault respectively, which is referred as to method 2. The last method is to choose 5000 samples randomly among the normal data, which is referred as to method 3. The fault detection method is then applied to both faulty turbines, as shown in Figs. 2 and 3. The results given in the next section are based on the turbine with a gearbox bearing fault with the normal data being selected using method 3.

4 Results and discussion

4.1 Monitoring variable selection

After pre-processing the original data by removing those control and DC signals, there are 78 variables in total remaining for further data dimension reduction. All the data samples relating to the fault are selected and processed with KPCA. To select the appropriate PCs, the variance contribution of each PC needs to be calculated, as given in Table 1. 16 PCs are therefore selected to meet the requirement of achieving 85% accumulated variance contribution.

4.2 Normal-abnormal condition classification

The selected PCs will be further processed by KSVM. Since KSVM is a supervised learning algorithm, the dataset needs to be divided into two groups, the data under normal conditions and the data under abnormal condition (formed by false alarms and true alarms related to the fault). Since it is impossible to plot 16-dimensional graph form the selected 16 PCs, all the results will be plotted in 2D space about wind speed and active power.

Fig. 5 gives an example of the data needing to be processed for normal–abnormal classification, where the blue dots represent the normal data and red crosses represent the abnormal data.

As mentioned above, to process the data using the KSVM algorithm, the linear inseparable data in a lower dimension can be projected into a higher dimension and thus differentiated by a hyperplane. As an example, Fig. 6 shows the working principle of the KSVM, where the blue dots represent the normal data and red dots represent the abnormal data. The support vectors are labelled by green circles while the fitted hyperplane is demonstrated in
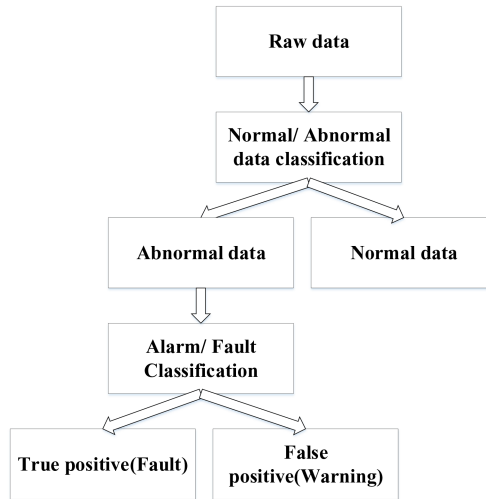
Fig. 4 *Overall modelling procedures*

Table 1 Variance contribution of the PCs

PCs	1	2	3	4	5	6
contribution%	6.59	6.56	6.34	6.34	6.17	5.64

PCs	7	8	9	10	11	12
contribution%	5.64	5.53	5.49	5.47	5.13	4.42

PCs	13	14	15	16	…	78
contribution%	4.41	4.40	3.72	3.68	…	0.000975
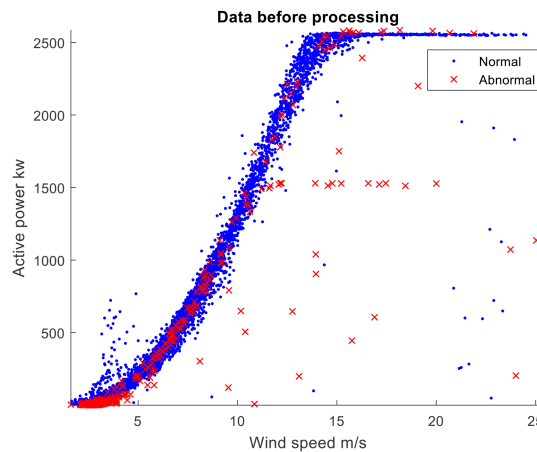
Fig. 5 *Power curve from normal and abnormal data*

gradient colour. The function of the fitted hyperplane is expressed as

$$z = 51.68 + 1.382x - 41.55y - 4.574x^2$$
$$+ 28.84xy + 17.82y^2 \tag{20}$$

where x, y are the wind speed and active power, respectively. The coefficient of determination r^2 is used to evaluate the accuracy of the fitting and the value of this fitted plane is 0.8605.

During this process, 70% of the data were used as the training set and 30% of the data were used for validation. The validation result is displayed in Figs. 7 and 8. In Fig. 7, the normal data classified as normal are shown in blue dots while the normal data classified as alarm are shown in blue crosses; the alarm data classified as alarm are shown in red dots and alarm data classified as normal data are shown in red crosses. Fig. 8 shows the confusion map of the normal-alarm classification result, which is used to evaluate the performance of the algorithm. The white areas show rates of both normal and alarm data were predicted correctly and the yellow areas show the misclassified data. As can be observed from the figure, the predicted normal data have reached 99.9% true and alarm data have reached 90.9% true, leading to a total accuracy of 99.4%.

4.3 Alarm-fault classification

After the procedure of normal-alarm classification, the alarm-fault classification is then processed.

Fig. 9 shows the alarm-fault classification in the relationship between wind speed and active power. The blue dots represent alarm signals and red dots represent for fault signals. The support
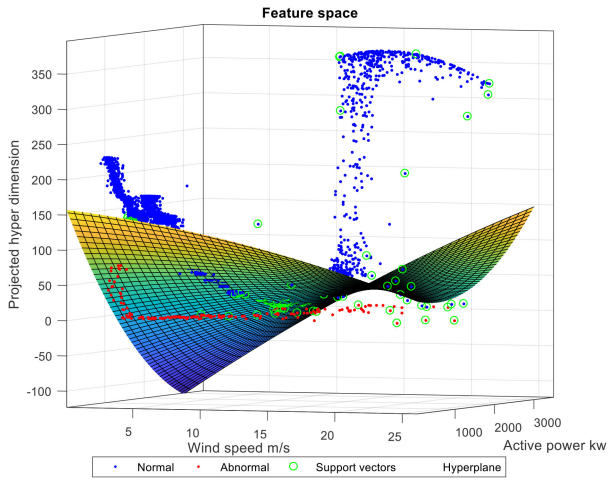
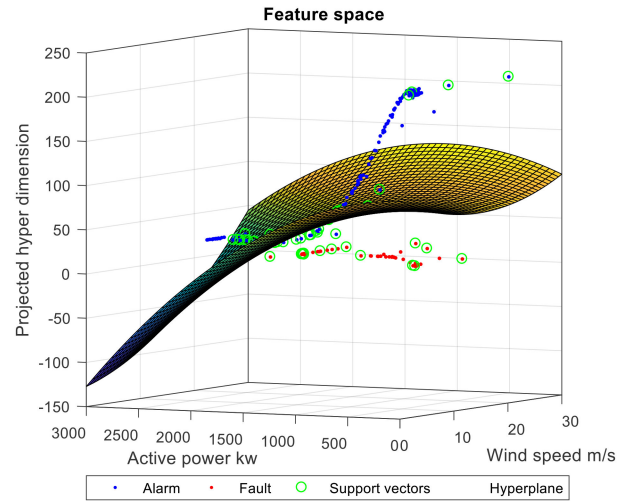**Fig. 6** *Normal–abnormal data classification using KSVM in the hyper dimension*

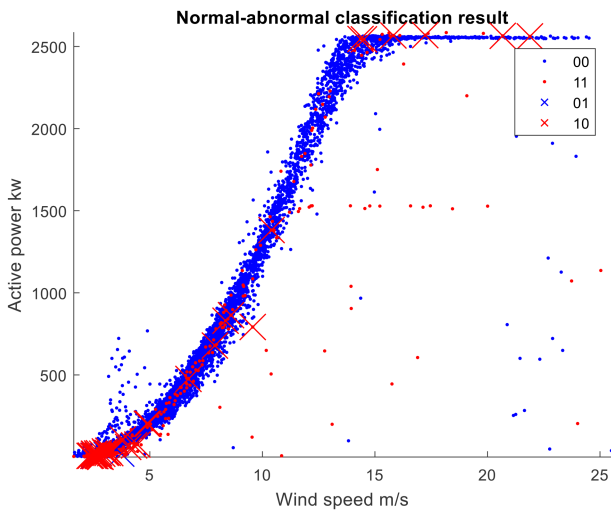**Fig. 9** *Alarm-fault classification using KSVM in hyper dimension*
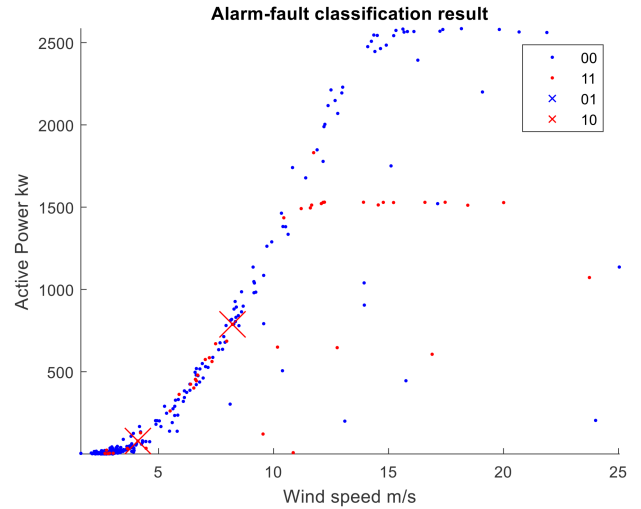

Fig. 7 *Normal–abnormal classification result*

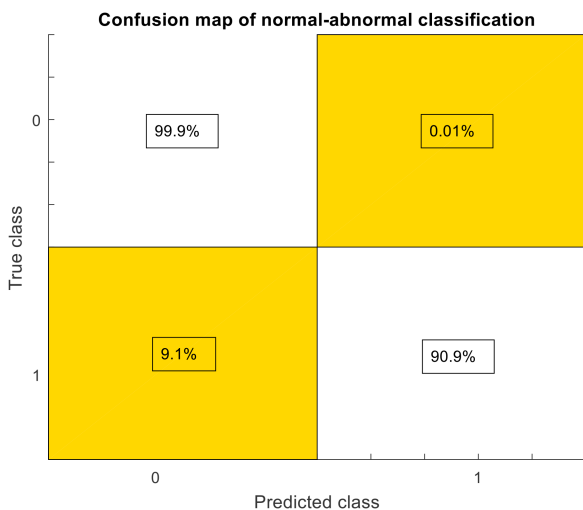**Fig. 10** *Alarm-fault classification result*

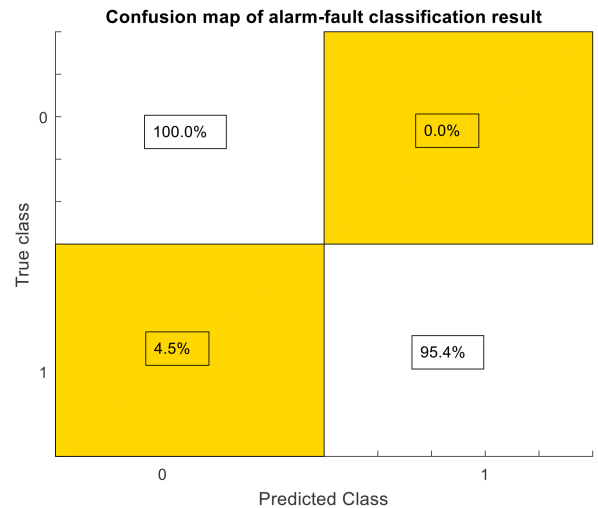**Fig. 8** *Confusion map of the normal–abnormal classification result*

Fig. 11 *Confusion map of the alarm-fault classification result*

vectors are labelled in the green circle and the classification hyperplane is fitted as

$$z = 68.2 + 14.47x - 16.24y + 6.274x^2 + 8.044xy - 8.719y^2 \quad (21)$$

where x, y are the wind speed and active power, respectively. The r^2 of the fitted plane is 0.7639.

Figs. 10 and 11 give the validation results of the alarm-fault classification model. The accuracy of an alarm-fault classification

model is also evaluated by the confusion map. It can be seen from the figure that it achieves 100.0% accuracy on alarm signal classification and 95.4% accuracy on fault signal classification. The total accuracy on alarm-fault classification has reached 99.3%.

To examine the robustness of the proposed methods, more turbines are tested with different SCADA data selection methods. It can be observed from Table 2 that the performances of the turbine with generator winding fault are not as good as the turbine with gearbox bearing fault. This might be due to the insufficient samples

Table 2 Classification results for the faulty turbines based on three different data selection methods

	Normal–abnormal classification, %	Alarm-fault classification, %
gearbox bearing fault turbine with method 1	99.7	96.3
gearbox bearing fault turbine with method 2	99.8	98.6
generator winding fault turbine with method 1	96.9	93.1
generator winding fault turbine with method 2	97.3	76.2
generator winding fault turbine with method 3	93.6	85.7

acquired from alarm and fault signals. It can be believed that if the amount of abnormal data increases, the classification model accuracy can be improved further.

5 Conclusion

With these alarm signals being identified, the fault can be warned at an early stage, which leaves sufficient time for maintenance scheduling. According to the results, several conclusions are drawn as follows:

- To select PCs of the monitoring variables, the accumulated variance of the PCs can be regarded as the most significant factor. However, to maintain the most information of the original dataset, the computation load needs to be compromised.
- Compared with other machine learning algorithms, the SVM has its strength in solving the two-group classification problem. Compared with the decision tree and discriminant analysis algorithms, the SVM demonstrates more accurate results.
- In terms of sample data selection, the turbine, which has a large amount of abnormal data, shows a better classification performance, indicating the influence of the sample selection.
- The KPCA can reduce the dimension in an acceptable range while the KSVM demonstrates excellent results for the two-stage classification.

Further work will be focused on the examination of the proposed approach incorporating with deep learning algorithms and verification of the results with more data from both simulations and physical test rig.

6 Acknowledgments

7 References

[1] Global Wind Energy Council: 'Global wind statistics', 2017, 4, (2018)
[2] Zhou, J., Liu, Y., Chen, Y.: 'Face recognition using kernel PCA and hierarchical RBF network'. 6th Int. Conf. on Computer Information Systems and Industrial Management Applications, 2007, CISIM'07, 2007, pp. 239–244
[3] García Márquez, F.P., Tobias, A.M., Pinar Pérez, J.M., *et al.*: 'Condition monitoring of wind turbines: techniques and methods', *Renew. Energy*, 2012, **46**, pp. 169–178
[4] Tavner, P.: 'Offshore wind turbines: reliability, availability and maintenance', 2012
[5] Qiu, Y., Feng, Y., Tavner, P., *et al.*: 'Wind turbine SCADA alarm analysis for improving reliability', *Wind Energy*, 2012, **15**, (8), pp. 951–966
[6] Taouali, O., Jaffel, I., Lahdhiri, H., *et al.*: 'New fault detection method based on reduced kernel principal component analysis (RKPCA)', *Int. J. Adv. Manuf. Technol.*, 2015, **85**, (5–8), pp. 1547–1552
[7] Abdi, H., Williams, L.J.: 'Principal component analysis', *Wiley Interdiscip. Rev., Comput. Stat.*, 2010, **2**, (4), pp. 433–459
[8] Skittides, C., Früh, W.: 'Wind forecasting using principal component analysis', *Renew. Energy*, 2014, **69**, pp. 365–374
[9] Ata, R.: 'Artificial neural networks applications in wind energy systems: a review', *Renew. Sustain. Energy Rev.*, 2015, **49**
[10] Puntanen, S., Styan, G.P.H., Isotalo, J.: 'Singular value decomposition', '*Matrix tricks for linear statistical models*' (Springer, Berlin, 2011), pp. 391–414
[11] Vapnik, V.: '*The nature of statistical learning theory*' (Springer, New York, 2010)
[12] Kline, R.: '*Principles and practice of structural equation modelling*' (Guilford Publications, New York, 2015)
[13] Boyer, S.A.: '*SCADA: supervisory control and data acquisition*' (International Society of Automation, Durham, 2009, 4th edn.)
[14] Qian, P., Ma, X., Cross, P.: 'Integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox', *IET Renew. Power Gener.*, 2017, **11**, (9), pp. 1177–1185
[15] Wang, Y., Ma, X., Qian, P.: 'Wind turbine fault detection and identification through PCA-based optimal variable selection', *IEEE Trans. Sustain. Energy*, 2018, **9**, (4), pp. 1627–1635
[16] Lydia, M., Kumar, S., Selvakumar, A., *et al.*: 'A comprehensive review on wind turbine power curve modeling techniques', *Renew. Sust. Energy Rev.*, 2014, **30**, pp. 452–460