A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines

Yueqi Wu, Xiandong Ma*

Engineering Department, Lancaster University, Lancaster, United Kingdom, LA1 4YW

Abstract

With the increasing installation of the wind turbines both onshore and offshore, condition monitoring technologies and systems have become increasingly important in order to reduce the downtime and operations and maintenance (O&M) cost, thus maximising economic benefits. This paper presents a novel machine learning model-based data-driven approach to accurately evaluate the performance of the turbines and diagnose the faults. The approach is based on Long-short term memory (LSTM) incorporating a statistical tool named Kullback-Leibler divergence (KLD). The hybrid LSTM-KLD method has been applied to two faulty wind turbines with gearbox bearing fault and generator winding fault respectively for fault detection and identification. The proposed method is then compared with three other well-established machine-learning algorithms to investigate its superiority. The results show that the proposed method can produce a more effective detection with accuracy reaching 94% and 92% for the turbines, respectively. Furthermore, the proposed method can effectively distinguish the alarms from the faults, from which the distinguished alarms can be considered as an early warning of the fault occurrence.

Keywords: Wind turbine, Condition monitoring, Long-short term memory, Kullback-Leibler Divergence

1. Introduction

As a type of renewable energy, wind power has been continually growing over the past decades [1]. According to the report by Global Wind Energy Council, the global wind energy sector installed a record 60.4 GW of new capacity in 2019, which increased 19% from the installation in 2018 [2]. However, with the exponentially installed capacity of wind power, the high cost of operations and maintenance (O&M) has become a vital issue, especially for offshore wind farms. The annual O&M cost is on average 3% of the original cost of the turbines. For offshore wind farms, the O&M cost contributes 30% of the total income of the turbines over 20 years of operating time [3]. Since wind turbines (WTs) are usually installed in remote areas and moving further offshore, more effort and costs are required to ensure their reliable operation. Therefore, it is important to detect the WT faults at the early stage and thus improve their reliability. In order to reduce the O&M cost and minimise the economic loss caused by downtime, an appropriate condition monitoring (CM) technique that is able to detect the early faults before being developed into the catastrophic stage would be crucial.

As we know, there are various types of signals that can be used for WT condition monitoring (WTCM). The signals mainly include acoustic emission, torque, temperature, lubrication oil, vibration, strain, electrical powers, ranging from different signal domains. Monitoring data are usually acquired from the supervisory control and data acquisition (SCADA) system for modern WTs [4]. In terms of the CM models, the data-driven model-based methods, as compared to physical-based models, are focusing on constructing the relationship between inputs and outputs of the system. It does not require building a mathematical model of the physical system [5]. Numerous statistical, machine learning and hybrid data-driven approaches have been studied. In statistical approaches, the factors such as variance, kurtosis, skewness, mean value and root mean square (RMS) values are acquired from time-series data to monitor the

*Corresponding author.

Email addresses: y.wu31@lancaster.ac.uk (Yueqi Wu), xiandong.ma@lancaster.ac.uk (Xiandong Ma)

rotor performance, blade surface roughness and gearbox failures of the WTs [6–9]. Usually, the deviations of those statistical values are monitored online during the operation. The statistical methods are proven as mature and flexible to implement; however, inaccurate decisions might be made because of their high sensitivity to noise since the noise tolerance of the statistical methods is lower.

Machine learning based methods such as artificial neural network (ANN) and support vector machine (SVM) have been developed fast in the past decade. These nature-inspired methods have been widely applied in different WT subsystems such as gearbox, generator, and power transmission systems [10–12]. In literature [13], the author proposed a CM method based on fusion of spatial-temporal features of SCADA data by convolutional neural network (CNN) and gated recurrent unit (GRU) to predict the output for recognising the WT condition. Wang et al [14] developed a deep neural network (DNN) based framework to model lubricant pressure and the results showed that the DNN model has more accurate results when compared with k-nearest neighbour, Lasso, Ridge, SVM, and ANN. Hu et al. [15] proposed a deep belief network (DBN) based prognostic model to improve the prediction accuracy and facilitate prognostic uncertainty for bearing degradation detection. In order to improve the accuracy and robustness of the CM methods, hybrid methods have also been developed. A hybrid statistical-machine learning method based on fast spectral kurtosis and multi-branch CNN was proposed for identifying the complex fault in WT gearbox [16]. It has been proved that this method can diagnose the gearbox fault with over 97% accuracy. Pan et al. [17] combined DBN, self-organising map and particle filtering to evaluate the degradation process and predict the remaining useful life of WT gearbox effectively. Several machine learning algorithms such as CNN, RNN (recurrent neural network) and LSTM can also be combined together to provide a more accurate CM method at the cost of model complexity [18–20].

Efforts have also be made to develop CM methods for early fault detection. Bangalore et al. proposed an ANN based CM method that used SCADA data to produce the early fault warning on WT gearbox in order to arrange maintenance resources [21]. Wang et al. employed operational condition clustering and optimised DBN modelling for early fault detection of the main bearing [22]. However, most CM approaches have not yet utilised multiple variables being monitored, e.g., temperatures and pressures, from the same WT subsystem to improve the CM capability. Besides, only few researches have taken SCADA alarms into consideration. In addition to the operational and environmental parameters, SCADA system includes a detailed record of alarm logs, revealing the malfunction of particular parameters of subsystems and components of the turbines. There are two types of alarm signals. The first type is usually triggered when certain measurement value of the components exceeds the pre-set threshold, which is considered as true alarms [23]. Another type is triggered when the system has experienced transient changes due to such as acute changes and disturbances of wind speed, which is thus considered as false alarms [24]. Due to inherent correlation between alarms and fault, it would be sensible to investigate if the WT alarm signals can be verified as an early warning for performance monitoring. By analysing the alarms and using time-sequence probability-based optimisation, the maintenance schedule can be well-organised and thus the WT reliability is improved [25]. Alarm signals can also be used to find the root cause of a fault or stoppage because the early warning raised by the alarms can be related to the fault occurrence [26]. Hence, alarm signals can be used to crosscheck potential faults identified from the data against what was actually happening and thus can play a significant role in WTCM.

To address the issues mentioned above, in this paper, a novel CM approach is presented based on a deep learning algorithm incorporating a statistic tool to estimate the operating condition of WTs. Multiple monitoring variables that contribute to the specific subsystems are taken into consideration to improve the reliability of fault diagnosis. Besides, the alarms are also fully used as significant evidences to support diagnostic results. Specifically, the LSTM is adopted to achieve the behaviour prediction of the key subsystems and then the KLD is employed to detect the fault by comparing probability distributions of the variables over the time. In the end, the monitoring data can be classified as normal, fault, true alarm and false alarm while the severity of the fault is also evaluated.

The main contributions of this paper are given as follows:

• A novel data-driven model-based CM method based on LSTM with KLD is proposed. The proposed method aims to implement a CM that can analyse the WT operating conditions automatically and detect both alarms and faults simultaneously.

• LSTM is adopted to capture relation features in temporal dependencies among monitoring data in an iterative manner, thus improving the prediction capability. The KLD is used as fault indicator which measures the severity of the fault by comparing probability distributions. By adopting a COST function based on normal value probability and alarm value probability of the calculated KLD values, the optimised thresholds are determined to distinguish the

normal, alarm and fault conditions.

• The distinguished alarms can be cross-referenced with SCADA alarm logs in order to provoke early warning of the fault. The proposed approach is validated by two faulty WTs with different but representative faults. By comparing the results with other well-established methods, the effectiveness of the proposed method is demonstrated in terms of detection accuracy and time complexity.

The remainder of this paper is organised as follows. The variable description and data pre-processing are presented in Section II. Section III describes the hierarchy structure of the combined LSTM and KLD, where the working principle of the proposed method is also presented. Case studies are undertaken in Section IV with both gearbox and generator being targeted. The performance of the proposed approach is compared in Section V with other machine learning methods, followed by conclusions in Section VI along with the future improvements.

2. Variable description and data pre-processing

A SCADA system is a control system architecture that consists of graphical user interfaces, data communications, computers and sensors. It is often used for high-level supervisory management of machine industries and has been used for monitoring the industrial, infrastructure and facility processes [27]. The SCADA data used in our model building is from an operational wind farm over a one-year time period, which consists of 26 WTs where each turbine has 128 monitoring variables. According to the functionalities of the monitoring variables, those 128 monitoring variables are categorised into 14 subsystems including blades, yaw, generator, main bearing, gearbox, controller, operation states, grid states, cooling system, rotor, nacelle, transformer, environmental parameters and general parameters. Among them, temperature related variables play an important role in fault diagnosis. In the commercial WT SCADA system, the temperature related variables take 40% of the overall monitoring variables [28].

Due to data storage and transmission limitation, the SCADA system usually processes and stores data at ten minute intervals despite being sampled at a rate of seconds. Although subtle changes may be lost due to low resolution data, SCADA data provide valuable online information with depth and breadth regarding the performance and operational history of the WTs. Therefore, SCADA data have been employed widely by researchers as the basis for CM systems [3, 29, 30]. In this research, the turbine with the gearbox bearing fault and the turbine with the generator winding fault have been tested. Because the turbine can be inactive for some periods of time, the performance of these data cannot match the normal condition, which could bring unpredictable influences on the analysis. Hence, SCADA data cannot be used directly and it is necessary to remove these data when no power is generated before modelling [31]. Besides, the digital constants in the SCADA data also need to be removed in order to minimise the interference brought by them.

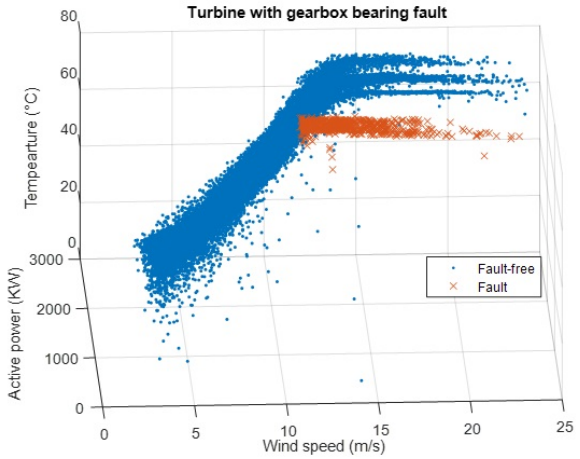

Fig.1 Gearbox bearing temperature changes with wind speed and active power

Fig.1 and Fig.2 illustrate examples of the target turbines, where Fig.1 shows the relationship among bearing temperature, wind speed and active power of the turbine with the gearbox bearing fault whereas Fig.2 shows the

relationship among generator winding temperature, wind speed and active power of the turbine with the generator winding fault. In the figures, blue dots represent fault-free data while red crosses represent fault data. It can be observed that the active power and the temperature of these specific subsystems gradually increase with the wind speed before the rated wind speed, which is 14 m/s. The active power and the temperature become stable after the rated wind speed because pitch controls adjust pitch angle of the rotor blades to maintain the rotation speed of the turbine. However, when the fault occurs at some points, the turbines operate with a reduced power output to prevent further damage. During these abnormal operations, the temperatures keep rising, showing abnormal behaviour compared with the normal operations. Clearly, there are strong correlations among wind speed, active power and temperatures, which are therefore selected for implementing the proposed approach.

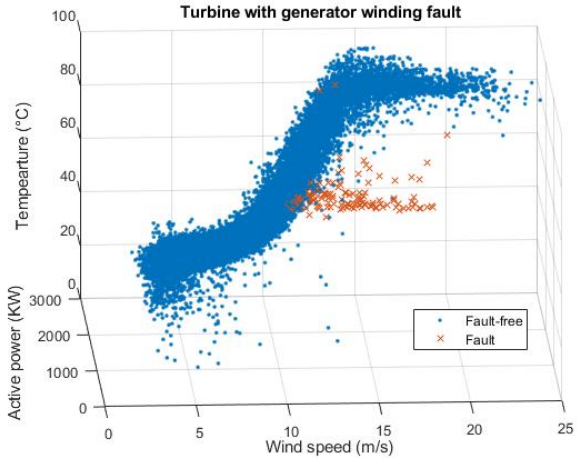

Fig.2 Generator winding temperature changes with wind speed and active power

3. Proposed hybrid condition monitoring approach

The LSTM algorithm was first proposed in 1995 by Sepp Hochreiter and Jürgen Schmidhuber [32]. It is a type of deep learning algorithm whose structure is implemented based on RNN. Unlike other regression based machine learning algorithms, the cell memory of LSTM can preserve the hidden state through time, while, in the meantime, adding new information. The KLD used in this paper is also called relative entropy, which has been applied as a measure of data representativeness. The KLD is a statistical tool developed to measure difference between two probability density distributions. It has been widely used in neuroscience and machine learning due to its ability in characterising relative entropy in information systems [33].

3.1. Long-short term memory

As mentioned, LSTM is developed based on RNN in order to solve the vanishing gradient problem. The feedback loops are embedded in every recurrent layer of the RNN, and thus the information can be preserved. However, with the increasing feedback loops in RNN, the gradient of the loss function decays exponentially with time. Compared to RNN, the LSTM has four interacting layers (cell state, forget gate, input gate and output gate) inside a LSTM cell [34], where, except for the standard units of RNN, a set of cell states and gates are added to control which memory should be stored. With this structure, the gradient decent problem of RNN is solved and "long-term memory" is achieved. The overall structure of the LSTM is illustrated in Fig.3.

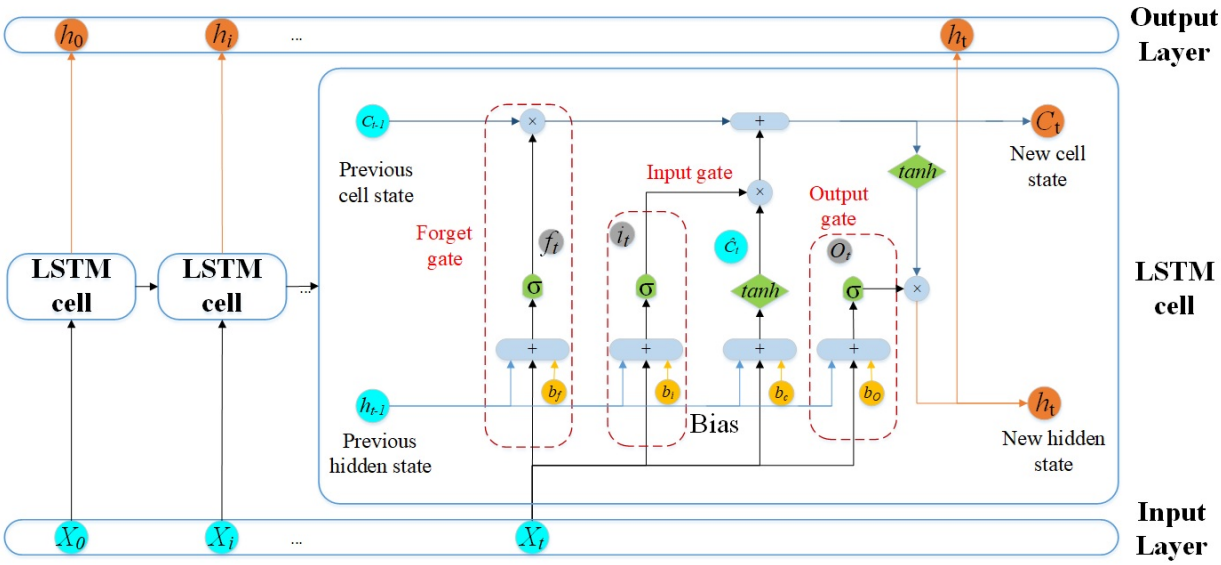

Fig.3 Schematic diagram of LSTM structure

At time step t, the LSTM cell has an input vector $[h_{t-1}, X_t]$. The new cell state, denoted by C_t, and the new output, denoted by h_t, inherit information from former states (C_{t-1} and h_{t-1}) and are transmitted to the next cell at time step $t+1$. The output h_t of the cell is determined by the current input X_t and the previous outputs h_{t-1}, h_{t-2}, while the cell state C_t can preserve information to flow forward without any losses. The gates in a LSTM cell include a forget gate (f), an input gate (i), and an output gate (O). The forget gate determines what information will be added (or forgot) to the cell state C_{t-1} when new input enters the network. The input gate decides what new information from the input will be updated into the cell state. The output gate decides what information updated in the cell state is sent to the network as input for next time step $t+1$, as represented by h_t.

Mathematically, the output f_t of the forget gate is given by

$$f_t = \sigma\left(W_f \odot [h_{t-1}, X_t] + b_f\right) \tag{1}$$

where σ is the sigmoid activation function; W_f and b_f are the weight and bias of the forget gate, respectively, and the operator $\odot$ denotes element-wise multiplication. The output of the sigmoid function is a value between 0 and 1, with 0 representing to forget this value and 1 representing to preserve this value.

Two activation functions are then used to decide which information to be stored in the cell. The sigmoid function σ at the input gate is used to decide which information to update while the hyperbolic tangent ($tanh$) function is used to decide what new information to add to the cell state denoted by $\hat{C}_t$.

$$i_t = \sigma\left(W_i \odot [h_{t-1}, X_t] + b_i\right) \tag{2}$$

$$\hat{C}_t = tanh(W_C \odot [h_{t-1}, X_t] + b_C) \tag{3}$$

where i_t is the output of the input gate; W_i and b_i are the weight and bias of the input gate whereas W_C and b_C are the weight and bias of the $tanh$ function, respectively.

This leads to the cell state C_t from C_{t-1} by computing the function below.

$$C_t = C_{t-1} f_t + i_t \hat{C}_t = C_{t-1} f_t + tanh\left(W_C \odot [h_{t-1}, X_t] + b_C\right) i_t \tag{4}$$

Then the LSTM cell is to compute the output O_t of the output gate by computing the following function.

$$O_t = \sigma\left(W_O \odot [h_{t-1}, X_t] + b_O\right) \tag{5}$$

where W_O and b_O are the weight and bias of the output gate, respectively. The sigmoid function σ in eq. (5) is used to decide which part of the cell state should be outputted.

The last step is to compute the output h_t of the cell, as described below.

$$h_t = O_t tanh(C_t) \tag{6}$$

where the *tanh* function is used to fit the output value in the range between 0 and 1.

The vanishing gradient problem in LSTM is solved by back-propagation. The gradient is calculated through back propagation along time using the chain rule. With all gradients calculated according to the corresponding error term (loss function), the weights associated with input gate, output gate, and forget gate are updated. More details about the back-propagation process of the LSTM can be refereed in the literature [35].

Compared to other machine learning algorithms such as CNN and DBN, one of the significant advantages of LSTM is that it can link the previous information to the current state. LSTM is a time related RNN, suitable for processing and predicting important events with relatively long time intervals and delays in time series. Many researches have employed LSTM and proved that it has advantages in solving time-series data especially for predicting performance of the rotating machinery components [36–38]. Hence, LSTM is chosen as the data modelling method in our study.

3.2. Kullback-Leibler Divergence

The KLD is used to compare the one probability distribution against another one [33]. For two discrete probability distributions P and Q in the same probability space X, the KLD from Q with respect to P can be expressed as:

$$\begin{aligned} KLD(P, Q) &= \sum_{x \in X} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \\ &= -\sum_{x \in X} P(x) \log\left(\frac{Q(x)}{P(x)}\right) \end{aligned} \tag{7}$$

If distributions P and Q are from continuous random variables, the KLD can be expressed as:

$$KLD(P, Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{8}$$

The KLD of the two distributions will be 0 if the two distributions are identical. The KLD value can be considered as an index to evaluate the discrimination between two probability density distributions. A larger KLD value indicates a larger difference between the two distributions. In KLD based condition monitoring, the probability density distribution of normal data is usually used as a reference whereas the one produced by the current/online data is used for comparison [39]. Xie et al. [40] proposed a KLD based fault detection method for dynamic systems such as gearbox, where, as compared to principal component analysis and statistical local approach, the KLD based method showed a better sensitivity in detecting the incipient fault. KLD can also be used to distinguish different types of faults based on different fault features [41–44].

3.3. Hybrid method

To implement the proposed model, four steps are required following pre-processing of the raw data, as illustrated in Fig.4. First, wind speed and active power output are selected as the model inputs, while temperature and pressure variables, which reflect the operation condition of the subsystem as a whole, are selected as the target output. The LSTM model is then built based on those variables. To train the LSTM model, the size of training dataset is vital, which is trade-off between the prediction performance and computation time. Small datasets can ease computation complexity but likely overfit the training data, resulting in poor performance. Larger datasets can help better learn model parameters but the dataset may be over-representative of the problem along with high computation demand. In our study, ten days' data (1440 points per variable considering ten-minute interval of SCADA measurements) are used to train the model. The data are then predicted on a daily basis, i.e., by sliding window with 24 hours, continuously for 30 days to ensure the modelling accuracy. It was found that increasing the training dataset size does not have significant improvement on the performance in our study. The probability density distributions of prediction data and original data are then calculated respectively. The third step is to evaluate the discrimination between the prediction

and the real data in terms of their KLD values. Note that daily data are considered to calculate the probability density distribution. The prediction values are considered as the condition at which the turbine supposes to operate while the measurement data represent the real operation condition of the turbine. Their difference can be evaluated by KLD and a larger KLD value represents a worse condition of the turbine. Finally, the normal, alarm and fault conditions are distinguished by introducing a two-level threshold strategy of the KLD values, which are defined as fault-free condition (H_0) and fault condition (H_1).

The probability density function of the KLD can be described as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-KLD_M)^2/2\sigma^2} \tag{9}$$

where KLD_M and σ are the mean value and standard deviation of the KLD, respectively.

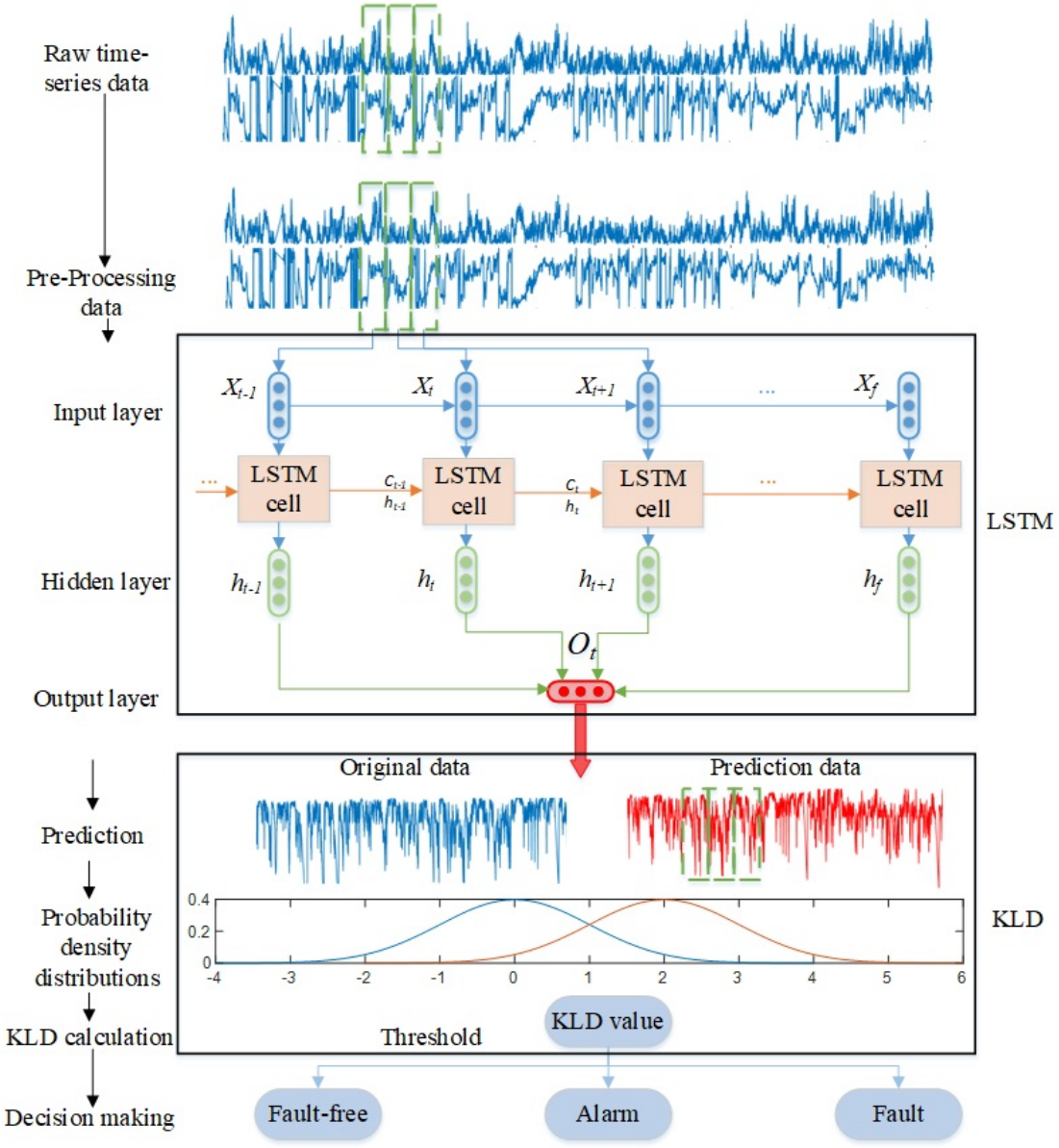

Fig.4 Structure of the proposed LSTM-KLD fault diagnosis method

Referring to [45], to determine the thresholds, a normal value probability (P_{NV}) and an alarm value probability (P_{AV}) are adopted here. The P_{NV} and P_{AV} can be calculated from the error function (10) of the KLD:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{10}$$

$$P_{NV} = 0.5\left(1 + \mathrm{erf}\left(\frac{h - KLD_{M0}}{\sigma_0 \sqrt{2}}\right)\right) \tag{11}$$

$$P_{AV} = 1 - 0.5\left(1 + \mathrm{erf}\left(\frac{h - KLD_{M1}}{\sigma_1 \sqrt{2}}\right)\right) \tag{12}$$

where KLD_{M0}, σ_0, KLD_{M1} and σ_1 are the mean and standard deviation values for two different conditions. By combining eqs. (11) and (12), a COST function can be formed to evaluate the performance of the thresholds.

$$\mathrm{COST} = P_{NV} + P_{AV} \tag{13}$$

The optimal thresholds are chosen when the COST reaches its minimum [46], at which both P_{NV} and P_{AV} are minimum as well. The threshold H_0 is determined when KLD_M=KLD_{M0} with a corresponding standard deviation of σ_0. It indicates the system is in a normal condition. When KLD_M=KLD_{M1} and σ=σ_1, the system is in an alarm condition. The data lower than H_0 are considered as normal condition while the data higher than H_1 are considered as faulty condition. The calculated KLD values between H_0 and H_1 are considered as alarm condition. Examples of these thresholds for specific components of the faulty turbines will be given in the subsequent section.

4. Case studies

4.1. Case 1: Gearbox fault

The first subsystem to be studied is the gearbox that is used to transmit kinetic power to the generator from the rotor. The WT torque control is made based on the gearbox by adjusting the rotation speed and torque accordingly. The two common faults associated with the gearbox are bearing and gear teeth faults. The unpredictable wind profiles can cause rapid changes of the torque, which may lead to misalignment of the gear teeth and uneven load for the bearing. Besides, the failure of the gearbox cooling system can also result in failures of bearing and gear teeth [32]. The gearbox under our study has 6 monitoring variables, including gearbox bearing temperature 1 at the main speed shaft bearing connected to the rotor, gearbox oil pressure, gearbox oil heat exchanger output temperature, gearbox oil sump temperature, gearbox oil pressure behind pump, and gearbox bearing temperature 2 at the high-speed shaft connected to the generator. To improve the diagnosis accuracy, all these six monitoring variables are analysed.

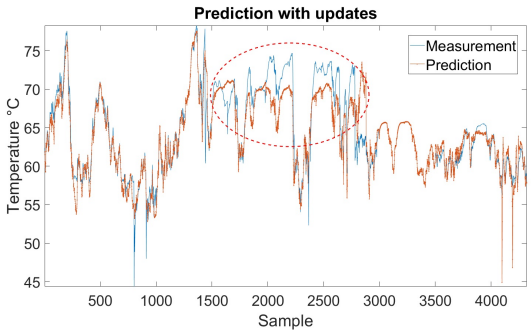

Fig.5 Prediction result during gearbox fault time

Fig.5 gives an example of the prediction result of the gearbox bearing temperature 1 as compared with the real measurements from 30 days, i.e., 4320 data samples, where the faulty time period of 10 days is circled in red. Because the prediction model is trained based on healthy data, the predicted results therefore can represent the performance of the target subsystem under fault-free condition. It can be observed from the figure that the predicted temperature is

Table 1. Thresholds of the gearbox components

Threshold	Gearbox bearing temperature 1 (°C)	Gearbox oil pressure (bar)	Gearbox oil heat exchanger output temperature (°C)	Gearbox oil sump temperature (°C)	Gearbox oil pressure behind pump (bar)	Gearbox bearing temperature 2 (°C)
H_0	0.3188	0.3477	0.5462	0.1197	0.1838	0.3381
H_1	1.9213	4.3833	3.0523	2.3068	4.2164	1.9404

lower than the measured temperature during the fault time. The probability density distributions of both predicted and measured data are shown in Fig.6, where the KLD value indicating overall divergence of the two probability density distributions is 7.2893. Note that only 10-days data are shown in the Fig. 6, focusing on the period of fault occurrence.

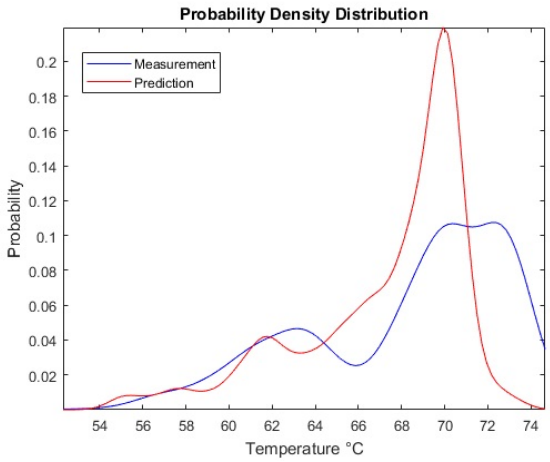

Fig.6 Probability density distributions of predicted and measured data of gearbox bearing temperature 1

The predicted data of all six variables are compared with measured data on a daily basis to calculate the KLD between them. The performance of the gearbox is shown in Fig.7, where the components 1-6 represent gearbox bearing temperature 1, gearbox oil pressure, gearbox oil heat exchanger output temperature, gearbox oil sump temperature, gearbox oil pressure behind pump, and gearbox bearing temperature 2, respectively, and the fault index is measured in terms of KLD. The corresponding thresholds H_0 and H_1 for each component are also displayed in the figure to determine whether there is a fault or just alarms. Table 1 gives the thresholds associated with each component of the gearbox.

Fig.7 shows that all the components are working normally for the first 5 days and alarms occur frequently since day 6. By checking the alarm logs, it is found that both false alarms and true alarms were triggered accordingly. The alarm named “Gearbox oil sump high temperature warning” and “Gearbox oil sump stop” were triggered before the fault occurrence. The two alarms clearly indicated the gearbox malfunction and therefore considered as true alarms. Thus the proposed method can give early warning of the fault 5 days in advance. After day 10, the gearbox bearing temperature 2 exhibited the first fault behaviour which were then propagated to other gearbox components in the following days. In day 16, all the gearbox components presented temporally normal behaviour due to the decreased wind speed. With the fault severity increased, the fault index reached its maximum on day 18. Overall, the fault lasted for ten days with a high fault index appearing on all components. Since all the faulty components related to temperature or oil pressure, it can be concluded that the fault occurred in the cooling system of the gearbox. The pressure sensors are usually installed at the end of the filter whereas the temperature sensors are installed in the oil sump to monitor the temperature of the lubricating oil. With the dysfunction of the cooling system, the heat cannot be dissipated actively, thus leading to further damage of the gearbox due to the fault.

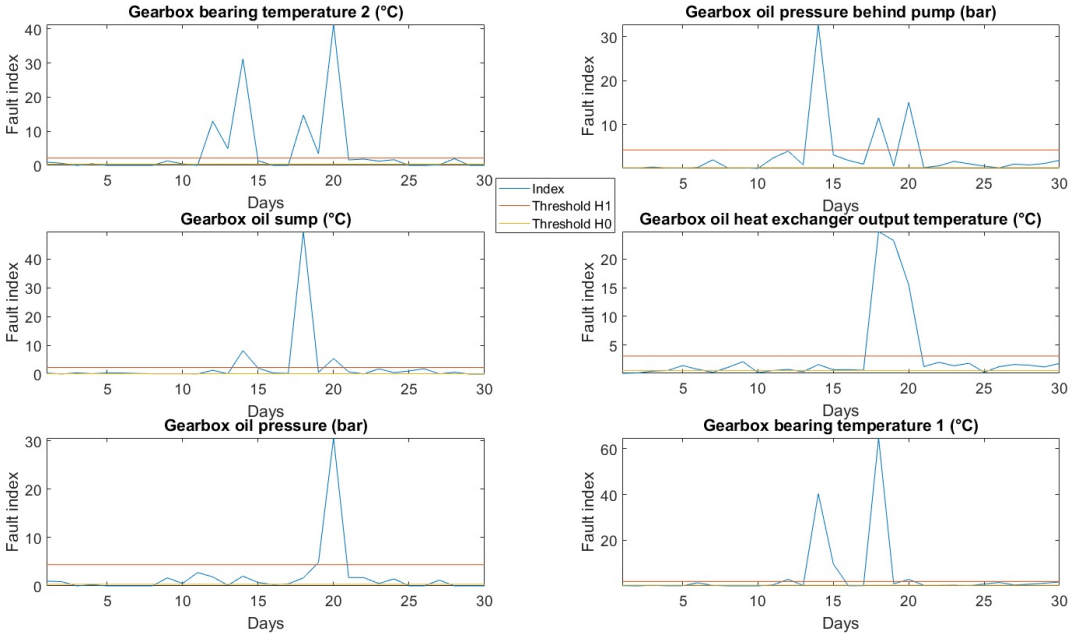

Fig.7 Fault index in terms of KLD of the gearbox variables

4.2. Case 2: Generator fault

The second subsystem to be studied is the generator that plays a key role in WT operation and produces electric power by converting the rotation mechanical power. The most common fault occurring in the generator is associated with the generator winding fault. This can be caused by phase-to-phase fault, phase to earth fault, or by additional thermal and mechanical stresses on the machine winding resulting from these electrical faults. The fault in the generator usually results in rising of generator temperatures. The generator under our study also has 6 monitoring variables, including generator stator top side temperature, generator stator bottom side temperature, generator bearing temperature phase A, generator bearing temperature phase B, generator cooling water advance temperature, and generator cooling water return temperature. Temperature sensors are embedded within the winding to monitor top and bottom side temperatures of the generator stator whereas bearing temperature sensors are normally fitted to the bearing cartridge to monitor the temperatures of the generator bearings. Temperature sensors are also installed on the two ends of the cooling water system to monitor the temperatures of cooling water.

As an example, Fig.8 shows predicted and measured values of the generator stator top side temperature from 30 days, i.e., 4320 data samples, where the faulty time period of 3 days is circled in red. Clearly the measured temperature is different from the predicted one during this faulty time period, indicating that the component experienced abnormal behaviour. As can be seen from the figure, the temperature is normally around 52° and there are several large spikes in the temperature. The occurrence of these temperature spikes are due to larger wind speeds and hence the higher power output during these time periods, as indicated from investigation of the SCADA data. The probability density distributions of measurement and prediction values are shown in Fig.9, where the KLD of these two distributions is 16.1747. Note that only 3-days data are shown in the Fig. 9, focusing on the period of fault occurrence.

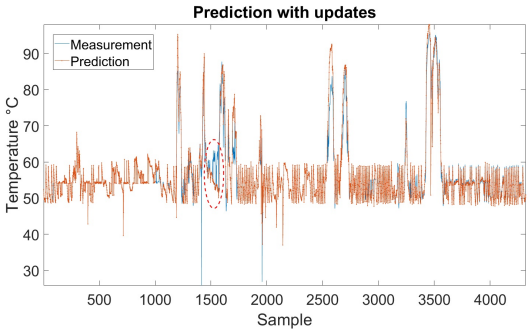

Fig.8 Prediction result during generator fault time

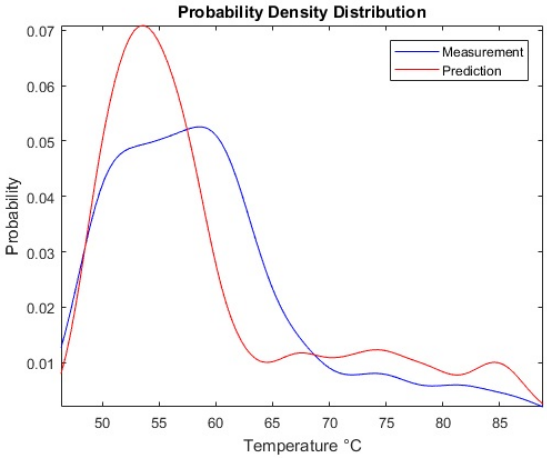

Fig.9 Probability density distributions of predicted and measured data for top side temperature of the generator stator

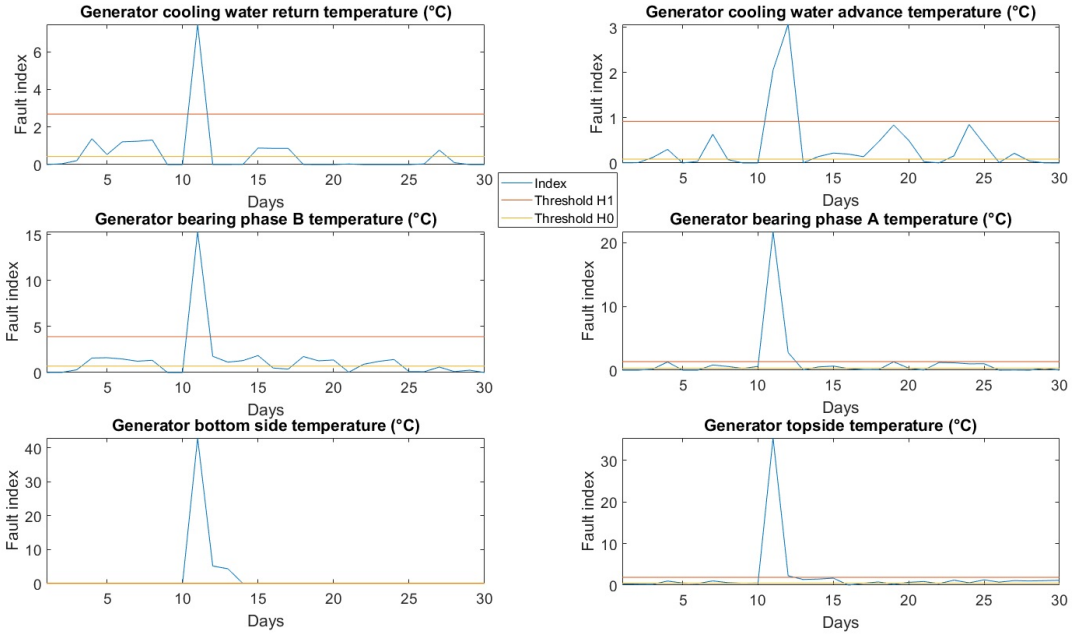

Fig.10 Fault index in terms of KLD of the generator variables

August 27, 2021

Table 2. Thresholds of generator components

Threshold	Generator topside temperature (°C)	Generator bottom side temperature (°C)	Generator bearing phase A temperature (°C)	Generator bearing phase B temperature (°C)	Generator cooling water advance temperature (°C)	Generator cooling water return temperature (°C)
H_0	0.4377	0.0088	0.2865	0.7004	0.0861	0.4322
H_1	1.3135	0.1199	1.3505	2.9033	0.9223	2.6908

Fig.10 demonstrates behaviour of these generator components during one month in terms of KLD values (i.e., fault index). Table 2 gives the thresholds associated with each component of the generator. For the first three days, the generator operated normally since all the KLD of the monitoring variables were below H_0. Alarms were triggered after day 3. Having checked SCADA alarm logs, it is found that the "Machine interface high temperature warning" alarm was triggered several times before the fault occurrence, which also confirm that the predicted alarms are correct. On day 10, all fault indices raised to a comprehensive high level, where the two highest indices represent generator stator top side temperature and generator stator bottom side temperature, respectively. This verifies that the fault happens on the generator winding, which lasted for three days before being fixed by maintenance and then returned to fault free condition.

5. Performance evaluation

It is clear that the proposed method can distinguish the normal, alarm and fault states of the operating turbines. To further test the accuracy and effectiveness of the hybrid method, dataset size will be extended to 100 days. The proposed LSTM-KLD will be compared with three other well-established machine-learning algorithms combined with KLD, which are SVM-KLD, CNN-KLD and DBN-KLD, respectively. The implementation of these hybrid methods are shown in Fig. 11, where the steps of these hybrid methods are essentially same. Only the SVM, CNN and DBN are used to replace the LSTM to produce the prediction results, which are then incorporated with KLD to calculate the fault index.

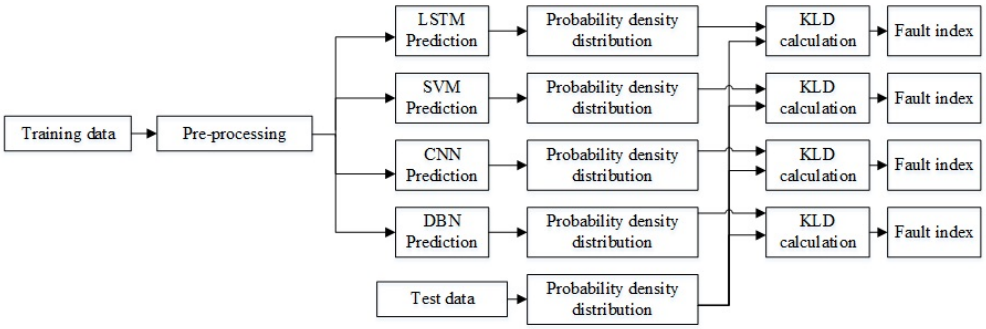

Fig.11 Schematic diagram of the hybrid method implementation

The performance of those four hybrid methods are evaluated by a confusion matrix to determine the accuracy rate when classifying normal, alarm and fault data. As is well known, the confusion matrix is also called as the error matrix, which is often used for solving classification issues [46]. It is essentially a visualised table showing the performance of a classification algorithm, where the rows of the confusion matrix show the prediction results while the columns show the actual data class [47]. For a confusion matrix, there are four main categories representing true positive, false positive, true negative and false negative. In this specific application, true positive indicates that the normal condition data are categorised correctly whereas false positive indicates that alarms and fault data are mistakenly considered as normal. True negative means the alarm and fault are classified precisely whereas the false negative shows how many

Table 3.Confusion matrices from four hybrid models for gearbox diagnosis
(a) LSTM-KLD (b) SVM-KLD (c) CNN-KLD (d) DBN-KLD

LSTM-KLD		Normal	Alarm	Fault
Actual Class	Normal	60 93.8%	4 6.2%	0 0%
	Alarm	0 0%	24 96%	1 4%
	Fault	1 9.1%	0 0%	10 90.9%
	Prediction Class			

(a)

SVM-KLD		Normal	Alarm	Fault
Actual Class	Normal	58 90.6%	6 9.4%	0 %0
	Alarm	0 0%	22 88%	3 12%
	Fault	0 0%	1 9.1%	10 90.9%
	Prediction Class			

(b)

CNN-KLD		Normal	Alarm	Fault
Actual Class	Normal	58 90.6%	6 9.4%	0 0%
	Alarm	2 8%	19 76%	4 16%
	Fault	0 0%	0 0%	11 100%
	Prediction Class			

(c)

DBN-KLD		Normal	Alarm	Fault
Actual Class	Normal	28 43.8%	34 53.1%	2 3.1%
	Alarm	0 0%	16 64%	9 36%
	Fault	0 0%	0 0%	11 100%
	Prediction Class			

(d)

Table 4. The overall accuracy of four hybrid methods for gearbox fault diagnosis

Accuracy	LSTM-KLD	SVM-KLD	CNN-KLD	DBN-KLD
TRUE	94%	90%	88%	55%
FALSE	6%	10%	12%	45%

normal data are mistakenly considered as alarm or fault data. By combing the true positive rate and true negative rate, the accuracy of the method is thus obtained. The decision of each class is made on a daily basis due to requirement of KLD calculation as mentioned previously. Hence, 100 decisions are made for confusion matrix evaluation.

Table 3 and Table 4 show confusion matrices and the overall accuracy of these hybrid methods for gearbox fault diagnosis, respectively. The true positive rate can be obtained by using the correct predictions divided by total samples available in the same row. For example, the true positive rate of the proposed LSTM-KLD can be calculated as 60/ (60+4) = 93.8% whereas the true negative rate for alarm and fault are 96% and 90.9%, respectively. Clearly, the LSTM-KLD outperforms three other hybrid methods.

Likewise, Tables 5 and 6 show confusion matrices and the overall accuracy of four hybrid methods for generator winding fault diagnosis, respectively. Again, the LSTM-KLD outperforms three other hybrid methods.

The computation time of the four hybrid methods is also compared to evaluate the efficiency of the proposed method. The programmes are run on MATLAB R2018a under the computer with Intel ® Core ™ i7- 6820 HK CPU @ 2.70GHz and 16.0 GB RAM. The GPU acceleration was disabled during the training process. The computation process includes training process and judgement process. The training processing is completed by the machine learning algorithms whereas the judgement process is completed by KLD. The four methods are run 10 times each on both faulty turbines and the average time is listed in Table 7. As can be seen from the table, the SVM-KLD consumes the shortest computation time since SVM relies only on the kernel function. The other three methods are essentially deep-learning algorithms and possess a much more complex structure in comparison to the SVM. Thus, those three methods consume a longer computation time, among which the LSTM-KLD clearly consumes shortest computation time.

Table 5. Confusion matrices from four hybrid models for generator winding diagnosis

(a) LSTM-KLD (b) SVM-KLD (c) CNN-KLD (d) DBN-KLD

LSTM-KLD		Normal	Alarm	Fault
Actual Class	Normal	64 90.1%	6 8.5%	0 1.4%
	Alarm	0 0%	25 96.2%	1 3.8%
	Fault	0 0%	0 0%	3 100%
		Prediction Class		

(a)

SVM-KLD		Normal	Alarm	Fault
Actual Class	Normal	62 87.3%	8 11.3%	1 1.4%
	Alarm	0 0%	25 96.2%	1 3.8%
	Fault	0 0%	0 0%	3 100%
		Prediction Class		

(b)

CNN-KLD		Normal	Alarm	Fault
Actual Class	Normal	64 90.1%	7 9.9%	0 0%
	Alarm	0 0%	23 88.5%	3 11.5%
	Fault	0 0%	0 0%	3 100%
		Prediction Class		

(c)

DBN-KLD		Normal	Alarm	Fault
Actual Class	Normal	40 56.3%	23 32.4%	8 11.3%
	Alarm	0 0%	18 69.2%	8 30.8%
	Fault	0 0%	0 0%	3 100%
		Prediction Class		

(d)

Table 6. The overall accuracy of four hybrid methods for generator winding fault diagnosis

Accuracy	LSTM-KLD	SVM-KLD	CNN-KLD	DBN-KLD
TRUE	92%	90%	90%	61%
FALSE	8%	10%	10%	39%

Table.7 Average computation time required from four hybrid methods for fault diagnosis

Computation time	LSTM-KLD	SVM-KLD	CNN-KLD	DBN-KLD
Training (s)	167.536	32.101	306.808	279.027
Judgement (s)	21.683	12.994	21.148	27.495

6. Conclusion

This paper presents a novel WTCM approach by a hybrid method of LSTM and KLD. The effectiveness of the proposed method have been evaluated and validated by SCADA data acquired from an operational wind farm. Two case studies are carried out to detect gearbox bearing fault and generator winding fault respectively from the turbines. In order to improve the reliability of diagnostic results, multiple monitoring variables relating to the specific WT subsystem are analysed. The probability density distributions of the measured data and the predicted data are compared and calculated in order to bring a first intuitive impression of the health condition of the turbine subsystem. The KLD values are calculated and used as a fault index to quantify the fault severity of the turbines. By cross-checking the predicted alarms and SCADA alarm logs, an early warning of the fault can be detected 5 days in advance. In order to further evaluate the performance and effectiveness, the proposed LSTM-KLD is compared with three other machine learning algorithms combined with KLD and the results prove superiority of the LSTM-KLD method.

The results demonstrate that the proposed data-driven model-based approach is accurate and sensitive for fault detection. It is capable of identifying different faults occurring in the different WT subsystems. With the proposed method, normal, alarm, fault conditions can be clearly distinguished, thus enhancing detection robustness. The results also show that the LSTM-KLD requires a shortest computation time in comparison to other deep learning approaches. Furthermore, by adopting the proposed method, the false alarm rate from the SCADA system is reduced, thus improving the diagnosis confidence. Future work will focus on the datasets with higher sampling rate to reveal fault dynamics and hence the fault mechanism. This method can also be extended to detect early faults of other WT subsystems and components.

7. Acknowledgement

Yueqi Wu gratefully acknowledges the support of the Engineering Department at Lancaster University in terms of his PhD Studentship. The permission to use SCADA data from Wind Prospect Ltd is also gratefully acknowledged.

References

[1] T. K. Mahmoud, Z. Y. Dong, J. Ma, A developed integrated scheme based approach for wind turbine intelligent control, IEEE transactions on sustainable energy 8 (2016) 927–937.
[2] A. M. Update, Global wind report, Global Wind Energy Council (2019).
[3] W. Zhang, X. Ma, Simultaneous fault detection and sensor selection for condition monitoring of wind turbines, Energies 9 (2016) 280.
[4] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault diagnosis Part II: Signals and signal processing methods, IEEE Transactions on Industrial Electronics 62 (2015) 6546–6557.
[5] P. Qian, X. Ma, P. Cross, Integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox, IET Renewable Power Generation 11 (2017) 1177–1185.
[6] J. J. Christensen, C. Andersson, S. Gutt, Remote condition monitoring of vestas turbines, in: Proceedings of the European Wind Energy Association conference, 2009.
[7] P. Caselitz, J. Giebhardt, Rotor condition monitoring for improved operational safety of offshore wind energy converters, J. Sol. Energy Eng. 127 (2005) 253–261.
[8] P. Guo, N. Bai, Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods, Energies 4 (2011) 2077–2093.
[9] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, M. Papaelias, Condition monitoring of wind turbines: Techniques and methods, Renewable Energy 46 (2012) 169–178.
[10] P. Bangalore, L. B. Tjernberg, An artificial neural network approach for early fault detection of gearbox bearings, IEEE Transactions on Smart Grid 6 (2015) 980–987.
[11] Y. Zhao, D. Li, A. Dong, D. Kang, Q. Lv, L. Shang, Fault prediction and diagnosis of wind turbine generators using SCADA data, Energies 10 (2017) 1210.
[12] H. Soliman, H. Wang, B. Gadalla, F. Blaabjerg, Condition monitoring for dc-link capacitors based on artificial neural network algorithm, in: 2015 IEEE 5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG), IEEE, 2015, pp. 587–591.
[13] Z. Kong, B. Tang, L. Deng, W. Liu, Y. Han, Condition monitoring of wind turbines based on spatio-temporal fusion of scada data by convolutional neural networks and gated recurrent units, Renewable Energy 146 (2020) 760–768.
[14] L. Wang, Z. Zhang, H. Long, J. Xu, R. Liu, Wind turbine gearbox failure identification with deep neural networks, IEEE Transactions on Industrial Informatics 13 (2016) 1360–1368.
[15] C. Hu, H. Pei, X. Si, D. Du, Z. Pang, X. Wang, A prognostic model based on DBN and diffusion process for degrading bearing, IEEE Transactions on Industrial Electronics 67 (2019) 8767–8777.
[16] J. Zhang, B. Xu, Z. Wang, J. Zhang, An FSK-MBCNN based method for compound fault diagnosis in wind turbine gearboxes, Measurement 172 (2021) 108933.

[17] Y. Pan, R. Hong, J. Chen, W. Wu, A hybrid DBN-SOM-PF-based prognostic approach of remaining useful life for wind turbine gearbox, Renewable Energy 152 (2020) 138–154.
[18] G. Yue, G. Ping, L. Lanxin, An End-to-End model based on CNN-LSTM for industrial fault diagnosis and prognosis, in: 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), 2018, pp. 274–278.
[19] D. Wei, B. Wang, G. Lin, D. Liu, Z. Dong, H. Liu, Y. Liu, Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report, Energies 10 (2017) 406.
[20] B. Zhao, C. Cheng, Z. Peng, X. Dong, G. Meng, Detecting the early damages in structures with nonlinear output frequency response functions and the CNN-LSTM model, IEEE Transactions on Instrumentation and Measurement 69 (2020) 9557–9567.
[21] P. Bangalore, M. Patriksson, Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines, Renewable Energy 115 (2018) 521–532.
[22] H. Wang, H. Wang, G. Jiang, J. Li, Y. Wang, Early fault detection of wind turbines based on operational condition clustering and optimized deep belief network modeling, Energies 12 (2019) 984.
[23] Y. Qiu, Y. Feng, D. Infield, Fault diagnosis of wind turbine with SCADA alarms based multidimensional information processing method, Renewable Energy 145 (2020) 1923–1931.
[24] H. Zhao, H. Liu, W. Hu, X. Yan, Anomaly detection and fault analysis of wind turbine components based on deep learning network, Renewable energy 127 (2018) 825–834.
[25] Y. Qiu, Y. Feng, P. Tavner, P. Richardson, G. Erdos, B. Chen, Wind turbine SCADA alarm analysis for improving reliability, Wind Energy 15 (2012) 951–966.
[26] K. Leahy, C. Gallagher, P. O'Donovan, D. T. O'Sullivan, Cluster analysis of wind turbine alarms for characterising and classifying stoppages, IET Renewable Power Generation 12 (2018) 1146–1154.
[27] S. A. Boyer, SCADA: supervisory control and data acquisition, volume 3, Isa Research Triangle Park, 1999.
[28] J. Dai, J. Cao, D. Liu, L. Wen, X. Long, Power fluctuation evaluation of large-scale wind turbines based on SCADA data, IET Renewable Power Generation 11 (2016) 395–402.
[29] P. Qian, X. Ma, D. Zhang, Estimating health condition of the wind turbine drivetrain system, Energies 10 (2017) 1583.
[30] P. Tavner, Offshore wind turbines: reliability, availability and maintenance, volume 13, IET, 2012.
[31] Y. Wang, X. Ma, P. Qian, Wind turbine fault detection and identification through PCA-based optimal variable selection, IEEE Transactions on Sustainable Energy 9 (2018) 1627–1635.
[32] G. Van Houdt, C. Mosquera, G. Napoles, A review on the long short-term memory model, Artificial Intelligence Review 53 (2020) 5929–5955.
[33] T. Van Erven, P. Harremos, Rényi divergence and Kullback-Leibler divergence, IEEE Transactions on Information Theory 60 (2014) 3797–3820.
[34] X. Le, H. V. Ho, G. Lee, S. Jung, Application of long short-term memory (LSTM) neural network for flood forecasting, Water 11 (2019) 1387.
[35] K. Yang, Y. Liu, Y. Yao, S. Fan, A. Mosleh, Operational time-series data modeling via LSTM network integrating principal component analysis based on human experience, Journal of Manufacturing Systems (2020).
[36] M. Jalayer, C. Orsenigo, C. Vercellis, Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, fast fourier and continuous wavelet transforms, Computers in Industry 125 (2021) 103378.
[37] Z. K. Abdul, A. K. Al Talabani, D. O. Ramadan, A hybrid temporal feature for gear fault diagnosis using the long short term memory, IEEE Sensors Journal 20 (2020) 14444–14452.
[38] R. Zhao, J. Wang, R. Yan, K. Mao, Machine health monitoring with LSTM networks, in: 2016 10th international conference on sensing technology (ICST), IEEE, 2016, pp. 1–6.
[39] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti, S. Longhi, Induction motor fault detection and diagnosis using KDE and Kullback-Leibler divergence, in: IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2013, pp. 2923–2928.
[40] L. Xie, J. Zeng, U. Kruger, X. Wang, J. Geluk, Fault detection in dynamic systems using the Kullback-Leibler divergence, Control Engineering Practice 43 (2015) 39–48.
[41] F. Zhang, Y. Liu, C. Chen, Y. Li, H. Huang, Fault diagnosis of rotating machinery based on kernel density estimation and Kullback-Leibler divergence, Journal of Mechanical Science and Technology 28 (2014) 4441–4454.
[42] J. Zeng, U. Kruger, J. Geluk, X. Wang, L. Xie, Detecting abnormal situations using the Kullback-Leibler divergence, Automatica 50 (2014) 2777–2786.
[43] A. Giantomassi, F. Ferracuti, S. Iarlori, G. Ippoliti, S. Longhi, Electric motor fault detection and diagnosis by kernel density estimation and Kullback-Leibler divergence based on stator current measurements, IEEE Transactions on Industrial Electronics 62 (2014) 1770–1780.
[44] J. Harmouche, C. Delpha, D. Diallo, Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II, Signal Processing 109 (2015) 334–344.
[45] A. Youssef, C. Delpha, D. Diallo, An optimal fault detection threshold for early detection using Kullback-Leibler divergence for unknown distribution data, Signal Processing 120 (2016) 266–279.
[46] S. Dasgupta, A cost function for similarity-based hierarchical clustering, in: Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 118–127.
[47] A. Luque, A. Carrasco, A. Martín, A. de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognition 91 (2019) 216–231.