Wind Turbine Gearbox Failure Identification With Deep Neural Networks

Long Wang, *Student Member, IEEE*, Zijun Zhang, *Member, IEEE*, Huan Long, *Student Member, IEEE*, Jia Xu, and Ruihua Liu

***Abstract*—The feasibility of monitoring the health of wind turbine (WT) gearboxes based on the lubricant pressure data in the supervisory control and data acquisition system is investigated in this paper. A deep neural network (DNN)-based framework is developed to monitor conditions of WT gearboxes and identify their impending failures. Six data-mining algorithms, the *k*-nearest neighbors, least absolute shrinkage and selection operator, ridge regression (Ridge), support vector machines, shallow neural network, as well as DNN, are applied to model the lubricant pressure. A comparative analysis of developed data-driven models is conducted and the DNN model is the most accurate. To prevent the overfitting of the DNN model, a dropout algorithm is applied into the DNN training process. Computational results show that the prediction error will shift before the occurrences of gearbox failures. An exponentially weighted moving average control chart is deployed to derive criteria for detecting the shifts. The effectiveness of the proposed monitoring approach is demonstrated by examining real cases from wind farms in China and benchmarked against the gearbox monitoring based on the oil temperature data.**

***Index Terms*—Condition monitoring, data mining, deep neural network (DNN), lubricant pressure, wind turbine gearbox.**

I. Introduction

THE aging of wind farms produces emerging needs of novel approaches for offering more efficient health management functions. Advanced condition monitoring techniques for reducing the wind farm operations and maintenance cost caused by performance degradations and system failures are highly demanded [1], [2]. The predictive maintenance with early identifications of wind turbine (WT) malfunctions can prevent failure occurrences and minimize unscheduled downtime [3]. Major subsystems of a WT including the gearbox, generator, and bearing were vulnerable and subject to frequent failures [4]. Studies of their condition monitoring and fault diagnosis attracted great attentions from the wind energy industry [5], [6]. WT gearbox transmits mechanical energy into the generator with high speeds, meanwhile suffers heavy loads, transient impulses of brakes, and dust corrosions. Therefore, it is one of the most fragile components of WT. According to [4], the gearbox failure occupied 59% of the total WT failures and resulted in unexpected downtime as well as economic losses. Hence, the research of developing efficient and effective WT gearbox condition monitoring and fault identification approaches is meaningful and valuable.

Vibration signals have been widely considered in the fault diagnosis of WT gearboxes. Monitoring approaches based on vibration signals have been presented in many studies [7], [8]. One category of approaches focused on extracting useful features for indicating anomalies from noisy vibration signals. Lei *et al.* [9] proposed an adaptive stochastic resonance (SR) method to obtain weak characteristics from noisy vibration signals. Li *et al.* [10] introduced a noise-controlled second-order enhanced SR method based on the Morlet wavelet transform to extract fault features from WT vibration signals. Antoniadou *et al.* [11] utilized the empirical mode decomposition method to decompose the vibration signals. Luo *et al.* [12] integrated the spectral analysis and the acceleration enveloping technique to extract gear damage features. Feng and Liang [13] proposed a time–frequency analysis method to discover constituent frequency components of nonstationary signals for monitoring WT gearboxes. Teng *et al.* [14] applied the complex Gaussian wavelet to obtain the multiscale enveloping spectrogram for extracting weak features. In addition to signal processing methods, data-driven approaches for analyzing vibration signals are also presented in the literature. Rafiee *et al.* [15] modeled different gear conditions with neural networks (NNs) based on the standard deviation of wavelet packet coefficients. Zhang *et al.* [16] introduced a data-driven framework for monitoring WT gearbox conditions through a continuous examination of their vibration excitements. The practical value of monitoring gearboxes with vibration analyses is limited due to following drawbacks: 1) commercial WTs are not equipped with vibration sensors mounted on gearboxes; 2) the installation of extra vibration sensors requires additional investments; 3) the measurement accuracy of vibration signals

Manuscript received June 27, 2016; revised August 26, 2016; accepted September 2, 2016. Date of publication September 8, 2016; date of current version June 1, 2017. This work was supported in part by the Early Career Scheme Grant from the Research Grants Council of the Hong Kong Special Administrative Region under Project CityU 138313 and in part by the CityU Strategic Research Grant under Project 7004551. Paper no. TII-16-0574. *(Corresponding author: Z. Zhang.)*

L. Wang, Z. Zhang, and H. Long are with the Department of Systems Engineering and Engineering Management, College of Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: long.wang@my.cityu.edu.hk; zijzhang@cityu.edu.hk; hlong5-c@my.cityu.edu.hk).

J. Xu and R. Liu are with the Centre of Wind Farm Data Analysis and Performance Optimization, China Longyuan Power Group Corporation Ltd., Beijing 10034, China (e-mail: xujia@clypg.com.cn; liuruihua@clypg.com.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2016.2607179

might be impaired because the gearbox is coupled with other components.

Supervisory control and data acquisition (SCADA) systems have been deployed in most of commercial wind farms for collecting WT conditional data [17]. Monitoring WT gearboxes with SCADA data has been studied and the oil temperature was frequently considered as the monitoring target. Feng *et al.* [18] derived robust relationships among temperature, efficiency, rotational speed, as well as power and predicted gearbox failures by using the SCADA oil temperature. Garcia *et al.* [19] modeled the normal bearing temperature, the thermal difference, as well as the cooling oil temperature with NN and applied developed NN models to detect the incipient anomalies of gearboxes. Wang and Infield [20] employed a nonlinear state estimation method to develop the cooling oil temperature model for detecting gearbox faults. As the oil temperature is sensitive to the environmental conditions, its SCADA data can be noisy. Compared with the oil temperature, the gearbox lubricant pressure, which is less sensitive to external factors, can be considered as an alternative target for monitoring WT gearboxes. The lubricant is applied to smooth the gearbox operation and its pressure can vary with the accumulation of swarf produced by the mechanical wear.

This research develops a data-driven framework for monitoring WT gearboxes and identifying their failures based on the lubricant pressure SCADA data. To accurately model the lubricant pressure of healthy gearboxes with SCADA data, the deep neural network (DNN), which has been considered as a powerful approach for modeling higher complexities [21]–[25], is applied. Compared with shallow NNs, the DNN employs multiple hidden layers to better organize hidden units. The appropriate configuration of the DNN allows it to approximate complicated functions with a linear number of hidden units, while NNs might require an exponential number of hidden units to achieve the similar performance [26]. To prevent the overfitting, a dropout training procedure [27] is utilized to develop the DNN model. To validate the capability of DNN in modeling the lubricant pressure, the DNN model is compared with data-driven models developed by five famous algorithms, the *k*-Nearest Neighbors (kNN) [28], least absolute shrinkage and selection operator (Lasso) [29], ridge regression (Ridge) [30], support vector machines (SVM) [31], and shallow NN [32]. SCADA data of real gearbox failure cases are analyzed to show that the absolute percentage error (APE) of DNNs will shift before the failure occurrences. To identify the shift of APE, the exponentially weighted moving average (EWMA) chart is applied. Computational experiments are conducted to validate the effectiveness of the proposed monitoring approach. The advantage of monitoring WT gearboxes with lubricant pressure data rather than the oil temperature data is also investigated.

The remaining parts of this paper are organized as follows: In Section II, the proposed data-driven monitoring framework including the lubricant pressure modeling and the monitoring approach is introduced. Meanwhile, metrics for evaluating the performance of data-driven models are defined. In Section III, computational experiments are conducted and their results are analyzed. A conclusion of the proposed research is presented in Section IV.

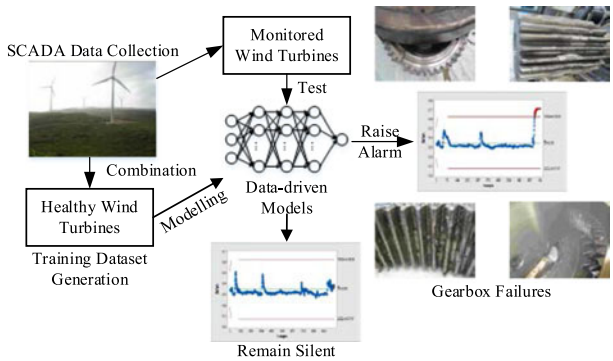


Fig. 1. Process of developing the data-driven monitoring framework.

II. Data-Driven Monitoring Framework

The data-driven monitoring framework is composed of two parts, the data-driven lubricant pressure model and the monitoring approach. A schematic diagram describing the process of developing the proposed data-driven monitoring framework is presented in Fig. 1.

As shown in Fig. 1, the proposed data-driven framework for monitoring WT gearboxes with the lubricant pressure data is developed by following steps:

Step 1. Generate the training dataset: Collect SCADA data of healthy WTs in a wind farm, clean invalid and missing values, and group processed SCADA data as a training dataset.

Step 2. Model the lubricant pressure: Apply data-driven algorithms to develop models for predicting the lubricant pressure based on the training dataset. To optimize parameter settings of algorithms, the grid search approach is employed.

Step 3. Model selection: Compute prediction errors of the lubricant pressure based on data-driven models and select the most accurate one for monitoring WT gearboxes.

Step 4. Online monitoring: Utilize the EWMA control chart to derive the criteria for identifying impending failures based on prediction errors of the selected model. Apply them into the online monitoring.

A. Data-Driven Models

The prediction model of the lubricant pressure described in (1) is built with three parameters, the gearbox oil temperature, T_o, power output, O_p, and shaft temperature, T_s, according to the suggestion of the domain expert

$$\hat{P} = \hat{f}(T_o, O_p, T_s). \tag{1}$$

In (1), $\hat{P}$ is the predicted lubricant pressure and $\hat{f}(\cdot)$ describes the developed data-driven model.

Six data-driven methods, the DNN, kNN, Lasso, Ridge, SVM, and NN, are applied to develop models for predicting the lubricant pressure based on the training dataset. Two metrics, the mean absolute percentage error (MAPE) and standard deviation of APEs (SDAPE), defined in (2)–(4), are employed to evaluate

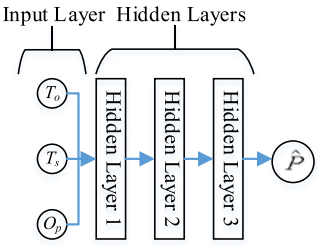


Fig. 2. Schematic diagram of DNN-based prediction model.

the prediction performance of data-driven models and select the most accurate one for monitoring

$$\text{APE} = \frac{\left|\hat{P} - P\right|}{P} \times 100\% \tag{2}$$

$$\text{MAPE} = \frac{1}{n_t}\sum_{i=1}^{n_t} \text{APE}_i \tag{3}$$

$$\text{SDAPE} = \sqrt{\frac{1}{n_t - 1}\sum_{i=1}^{n_t} (\text{APE}_i - \text{MAPE})^2}. \tag{4}$$

In (2)–(4), P is the actual lubricant pressure and n_t is the sample size of the test dataset.

1) Deep Neural Networks: In this research, a DNN with three hidden layers described in Fig. 2 is applied to learn the $\hat{f}(\cdot)$ in (1) from the training dataset.

The training process of DNN is to estimate parameters $\mathbf{w}^{(l)}$ and $\mathbf{b}^{(l)}$, which are weights and bias of the lth layer, by minimizing the average of squared errors as shown in

$$\{\mathbf{W}^{(l)},\mathbf{b}^{(l)}\} = \underset{\mathbf{W}^{(l)},\mathbf{b}^{(l)}}{\text{argmin}} \sum_{i=1}^{n} \frac{1}{2}(\hat{P}_i - P_i)^2, \quad l = 0, 1, \ldots, L. \tag{5}$$

In (5), L is the number of layers in DNN and n is the number of samples in the training dataset. The activation function considered in the DNN is the hyperbolic tangent function, described in

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{6}$$

To tackle the overfitting of DNN, one feasible approach is to ensemble all DNN models with suitable parameter settings [33]. However, such approach is computationally expensive as DNN models with different combinations of parameter settings need to be trained. The dropout method is an alternative and more efficient option for addressing DNN overfitting [27]. In many reported studies, it outperformed other regularization methods [34]. In a DNN developed with the dropout method, hidden units as well as their incoming and outgoing connections are temporarily excluded in the network during the training as shown in Fig. 3. Hidden units to be excluded are randomly selected with a fixed probability p set as 0.5 according to [35]. The application of the dropout method in training a DNN is equivalent to sample a simplified NN, which is composed of all retained hidden units

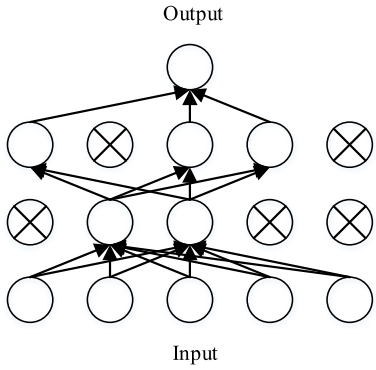


Fig. 3. Dropout neural network.

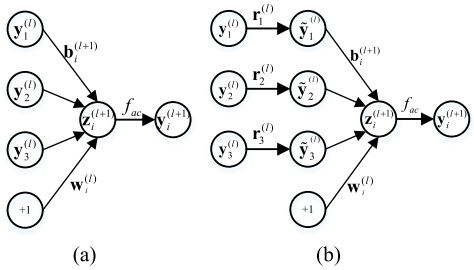


Fig. 4. Feed-forward operation without and with dropout. (a) Standard NN, (b) Dropout NN

(see Fig. 3) from its full map. Therefore, a DNN with u hidden units can be regarded as a collection of 2^u possible simplified versions that share weights. In the training, simplified DNNs are constructed based on the training data. Next, in the testing, 2^u possible simplified DNNs are assembled into one DNN by scaling-down outgoing weights of hidden units with multiplying p if those units are retained in each simplified DNN.

The details of the dropout method are demonstrated as follows.

Let $\mathbf{z}^{(l)}$ and $\mathbf{y}^{(l)}$ denote the input and output vector of the layer l, the standard feed-forward operation applying to construct the hidden unit i in the layer l can be described by

$$\begin{aligned} \mathbf{z}_i^{(l+1)} &= \mathbf{w}_i^{(l)}\mathbf{y}^{(l)} + \mathbf{b}_i^{(l+1)}, \\ \mathbf{y}_i^{(l+1)} &= f_{\text{ac}}(\mathbf{z}_i^{(l+1)}) \end{aligned} \tag{7}$$

where f_{ac} is the activation function.

In the training with a dropout method, a revised feed-forward operation is presented in (8) and illustrated in Fig. 4.

$$\begin{aligned} \mathbf{r}_j^{(l)} &\sim \text{Bernoulli}(p), \\ \tilde{\mathbf{y}}^{(l)} &= \mathbf{r}^{(l)} * \mathbf{y}^{(l)}, \\ \mathbf{z}_i^{(l+1)} &= \mathbf{w}_i^{(l)}\tilde{\mathbf{y}}^{(l)} + \mathbf{b}_i^{(l+1)}, \\ \mathbf{y}_i^{(l+1)} &= f_{\text{ac}}(\mathbf{z}_i^{(l+1)}) \end{aligned} \tag{8}$$

where $*$ denotes the element-wise product and $\mathbf{r}^{(l)}$ is a vector of independent Bernoulli random variables. In the training, the

element-wise product of generated $\mathbf{r}^{(l)}$ and the output vector of the layer l, $\mathbf{y}^{(l)}$, is considered as the input of the next layer. In the testing, weights, $\mathbf{W}_t^{(l)}$, are scaled by $\mathbf{W}_t^{(l)} = p\mathbf{W}^{(l)}$.

In this study, hidden layers with a maximal number of hidden units in each layer = 100 are employed to develop the DNN. To facilitate the training process, the parallelized stochastic gradient descent (SGD) algorithm [36] is applied to update parameters of DNN and its details are described in the Pseudo Code, ParallelSGD().

ParallelSGD $(\mathbf{T}, \alpha)$

```
{
  initialize W and b
  distribute T across nodes
  while not convergence criterion:
    for nodes m with training subset T_m, do in parallel:
      let W_m, b_m = W, b
      partitioin T_m into T_mc by cores m_c
      for cores m_c onnode m, do in parallel:
          get training example i ∈ T_mc
          update all weights w_jk ∈ W_m, biases b_jk ∈ b_m
              w_jk = w_jk − α ∂L(W,b|j)/∂w_jk
              b_jk = b_jk − α ∂L(W,b|j)/∂b_jk
      end for
    end for
    let W, b = Avg_m W_m, Avg_m b_m
  end while
}
```

In ParallelSGD(), $\mathbf{T}$ is the training dataset, α is the learning rate, and Avg_m is the final average weight of parameters in the node m.

2) Benchmarking Data-Driven Methods: The DNN is compared with five benchmarking data-driven methods, the kNN, Lasso, Ridge, SVM and NN, in modeling the lubricant pressure.

In kNN [28], the Euclidean distance is considered. To determine the optimal k value, kNN models with $k = 1, 2, \ldots, 15$ are evaluated through a 10-fold cross validation.

The Lasso model [29] for predicting the lubricant pressure is written as (9), in this study, and the estimation of its parameters is as follows (10):

$$\hat{P} = \mathbf{x}^T \hat{\boldsymbol{\beta}}^{\mathrm{Lasso}} + \hat{\beta}_0 \tag{9}$$

$$\{\hat{\boldsymbol{\beta}}^{\mathrm{Lasso}}, \hat{\beta}_0\} = \arg\min_{\boldsymbol{\beta}, \beta_0} \frac{1}{n} \sum_{i=1}^{n} \left\| P_i - (\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta}) \right\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \tag{10}$$

where $\boldsymbol{\beta}$ and β_0 are scalar parameters, while λ is a positive regularization parameter.

The optimal λ value is selected from a set $\{0.001, 0.002, \ldots, 0.5\}$ and the Bayes Information criterion (BIC) [37] defined in (11) is considered as the selection criteria

$$\mathrm{BIC} = -2 \cdot \ln\mathcal{L} + v \cdot \ln(n) \tag{11}$$

where $\mathcal{L}$ and v are two BIC parameters, which can be estimated according to [38]. The λ yields the smallest BIC considered in training the Lasso model.

The Ridge regression model [30] for predicting the lubricant pressure is presented in (12). A transformed input $\mathbf{v}$ as defined in (13) is considered by the Ridge regression model:

$$\hat{P} = \mathbf{v}^T \hat{\boldsymbol{\gamma}}^{\mathrm{Ridge}} + \hat{\gamma}_0$$

$$\{\hat{\boldsymbol{\gamma}}^{\mathrm{Ridge}}, \hat{\gamma}_0\} = \arg\min_{\boldsymbol{\gamma}, \gamma_0} \frac{1}{n} \sum_{i=1}^{n} \left\| P_i - (\gamma_0 + \mathbf{V}_i^T \boldsymbol{\gamma}) \right\|_2^2 + \kappa \|\gamma\|_2^2 \tag{12}$$

$$\mathbf{v} = \Phi(\mathbf{x}) = [\mathbf{x}, \mathbf{x}^2] \tag{13}$$

where $\boldsymbol{\gamma}$ and γ_0 are scalar parameters, while κ is the positive regularization parameter. A 10-fold cross validation is employed to select the optimal κ from a set $\{0.001, 0.002, \ldots, 0.5\}$.

The SVM [31] with a Gaussian kernel described in (14) is considered in this study:

$$K(\mathbf{x}, \mathbf{x}') = \exp(\xi \|\mathbf{x} - \mathbf{x}'\|^2) \tag{14}$$

where $\xi \in [0.0001, 0.001]$ is a model parameter. Besides ξ, a capacity factor with values = 1, 10, 100, and 1000 are also considered in training the SVM model. During the training, configurations of parameters are exhaustively searched and the one with the smallest 10-fold cross-validation error is selected.

The shallow NN model is composed of three layers, the input layer, hidden layer, and output layer. The activation function of the hidden layer is also a hyperbolic tangent function as shown in (6).

To improve the model performance, a L_1–norm regularization is applied to the objective function. By minimizing the average squared errors described in (15), parameters of NN are estimated:

$$\{\mathbf{W}, \mathbf{b}\} = \underset{\mathbf{W}, \mathbf{b}}{\mathrm{argmin}} \sum_{i=1}^{n} \frac{1}{2} (\hat{P}_i - P_i)^2 + \theta \|\mathbf{W}\|_1. \tag{15}$$

In (15), $\mathbf{W}$ describes the weights and $\mathbf{b}$ describes the biases. The number of hidden units is selected from a set $\{50, 80, 100\}$, and the θ value is selected from a set $\{0.001, 0.002, \ldots, 0.1\}$ in the training. The best combination of them is determined by the 10-fold cross validation. The parallelized SGD algorithm is also applied to accelerate the training of NN models.

B. Monitoring Approach

The variation of gearbox health conditions will induce an increase of prediction errors of the selected data-driven lubricant pressure model and it is meaningful to identify such patterns earlier to prevent the forthcoming gearbox failures. An EWMA control chart is applied to derive the criteria for continuously monitoring prediction errors and identifying their abnormal variations [39]. The EWMA statistic smooths out the uncontrollable noise based on a weighted average of recent and historical

TABLE I
DESCRIPTION OF FAILURE CASES AND SCADA DATA

Wind Farm	Index of WT_{gf} s	Indexes of healthy WTs	Failure Date	Collection period of SCADA data
Liaoning	49	35, 40, 42, 51, 53, 54, 62, 63	May 17th, 2015	April 1st to May 17th, 2015
Hebei	64	33, 50, 78	June 2nd, 2015	April 1st to June 2rd, 2015
Shaanxi	18	1, 6, 19, 22	June 6th, 2015	April 1st to June 6th, 2015
Shanxi[1]	14	41, 100	November 22nd, 2015	February 21st to November 22nd, 2015
Shanxi[2]	27	07, 09, 14, 16	November 21st, 2015	February 21st to November 21st, 2015
Shandong	None	1-66	None	January 1st to May 1st, 2015

Note: Two wind farms in Shanxi are denoted as Shanxi[1] and Shanxi[2].

observations. Moreover, the EWMA is sensitive to the small process shift. Lower and upper bounds for monitoring gearboxes are derived based on lubricant pressure prediction errors of normal WTs. An alarm of impending gearbox failures will be released in the online monitoring if prediction errors exceed either one of two boundaries.

To derive the lower and upper bound of the EWMA chart, a statistic, s_t, defined in (16) needs to be computed

$$s_t = \psi \mathrm{APE}_t + (1-\psi)s_{t-1} \tag{16}$$

where t is the time index, $\psi \in (0, 1)$ is the weight of historical APEs contributing to the estimation of s_t at t, and s_0 is set to the mean of historical APEs. In this study, ψ is set to 0.2 according to [39].

Based on (16), the mean and variance of s_t are further obtained by

$$\mu_{s_t} = \mu\mathrm{APE}, \quad \sigma^2_{s_t} = \frac{\sigma^2\mathrm{APE}}{n_m}\left(\frac{\psi}{2-\psi}\right)\left[1-(1-\psi)^{2t}\right] \tag{17}$$

where μAPE and σAPE are the mean and SDAPEs of all healthy WTs in the same wind farm, and n_m is the sample size.

Based on (17), the upper and lower bounds of EWMA control chart at t written in (18) and (19) can be constructed. The L is commonly set to 3 according to [40]:

$$\mathrm{UCL}(t) = \mu\mathrm{APE} + L\sigma\mathrm{APE}\sqrt{\frac{\psi[1-(1-\psi)^{2t}]}{(2-\psi)n_m}} \tag{18}$$

$$\mathrm{LCL}(t) = \mu\mathrm{APE} + L\sigma\mathrm{APE}\sqrt{\frac{\psi[1-(1-\psi)^{2t}]}{(2-\psi)n_m}}. \tag{19}$$

Besides recommended settings of L and ψ, a grid search can be applied to iteratively identify most suitable settings of L and ψ based on the APEs obtained in training.

III. COMPUTATIONAL STUDIES

The proposed data-driven framework is applied into the condition monitoring of gearboxes of total 92 WTs from six commercial wind farms in China. Five gearbox failure incidences were reported during the period covered by the considered SCADA data.

A. Data Description and Preprocessing

Commercial wind farms considered in this study are located in five provinces of China, Liaoning, Hebei, Shanxi, Shaanxi, and Shandong. All considered wind farms are equipped with SCADA systems and the sampling interval of the SCADA data was 10 min. The description of five failure incidences and applied SCADA data are provided in Table I.

A data pre-processing is conducted to remove invalid values from the SCADA data based on rules described in the following. As suggested by industrial experts, T_o should be less than or equals to 75 °C while P should be within an interval from 4 to 6. Therefore, two rules, $T_o \leq 75$ and $4 \leq P \leq 6$, are applied to filter the SCADA data first.

B. Comparative Analysis of Data-Driven Algorithms

After data preprocessing, all data of healthy WTs in the same wind farm are combined as the training dataset (Training data points range from 3641 to 76500 over wind farms) and lubricant pressure models are trained by six algorithms for each wind farm. SCADA data of WTs having impending gearbox failures ($\mathrm{WT_{gf}s}$), except data older than 20 days before failure occurrences, are considered as the test dataset to evaluate developed models (Similarly, test data points range from 659 to 8041). Since Shandong wind farm does not have gearbox failures, data of one randomly selected WT is treated as the test dataset (3600 data points). In practice, the number of data points for training DNN models can be set to at least ten times of the number of hidden unites contained in the DNN according to [41]. To better configure the training dataset, a validation dataset and a predefined target of the prediction error can be applied. More training data points are included until the prediction error of the developed DNN model is lower than a predefined target. To compare the performance of different models, MAPE and SDAPE are firstly computed by (3) and (4) and next reported in Tables II and III.

Table II indicates that DNN is better than other models as it offered lowest MAPEs over all wind farms. In Table III, the average of testing results further confirms the capability of DNN models. This superiority might be contributed by the dropout algorithm that provides an ensemble of many DNNs to overcome the bias-variance tradeoff. Computational results prove that the DNN algorithm is more effective than classical data-driven algorithms in developing the lubricant pressure prediction model.

TABLE II
MAPE AND SDAPE OF DIFFERENT ALGORITHMS

	Liaoning		Hebei		Shaanxi		Shanxi[1]		Shanxi[2]		Shandong	
Algorithm	MAPE	SDAPE	MAPE	SDAPE	MAPE	SDAPE	MAPE	SDAPE	MAPE	SDAPE	MAPE	SDAPE
Lasso	8.74	5.39	9.74	6.33	16.10	9.37	6.37	3.74	7.83	4.58	8.27	5.23
Ridge	7.88	5.73	10.63	8.03	15.59	9.12	6.90	4.07	8.20	4.83	7.68	5.56
kNN	4.85	6.51	9.72	6.21	16.78	10.97	7.97	4.57	9.29	5.53	5.47	6.32
SVM	4.42	3.04	8.61	5.96	14.60	9.87	12.30	6.56	12.65	7.06	6.00	3.39
NN	4.40	3.22	9.67	6.24	15.03	10.83	3.84	3.49	5.52	4.59	4.29	3.25
DNN	3.47	3.51	7.45	5.93	14.25	8.27	2.90	2.42	4.60	3.30	3.36	3.40

TABLE III
AVERAGE MAPE AND SDAPE OF MODELS OVER SIX WIND FARMS

Models	MAPE	SDAPE
Lasso	9.51	5.77
Ridge	9.48	6.22
kNN	9.01	6.69
SVM	9.76	5.98
NN	7.13	5.27
DNN	6.01	4.47

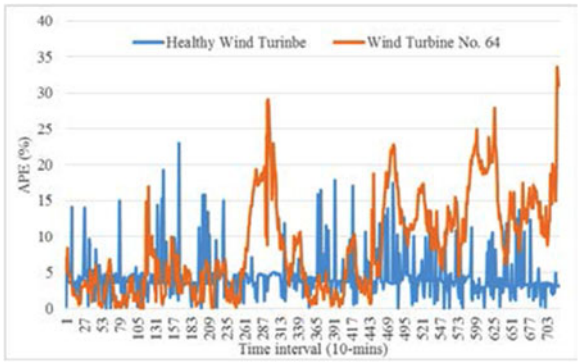


Fig. 5. APEs of turbine No. 64 in Hebei (May 13–May 17, 2015).

Besides the modeling accuracy, DNN requires the least computational time compared with other algorithms with tuned parameter settings. More impressively, the computational cost of the DNN with the parallelized SGD is less than the tuned SVM algorithm. Considering both the accuracy and computational efficiency, the DNN algorithm is considered more suitable to develop the lubricant pressure model in the monitoring framework.

By comparing APEs of the $\mathrm{WT}_{\mathrm{gf}}$ and healthy WTs, we can observe that the APE of $\mathrm{WT}_{\mathrm{gf}}$ will shift prior to occurrences of failures. Fig. 5 illustrates such pattern based on data collected from the Hebei wind farm and similar patterns are obtained for other wind farms. Hence, the APE of the DNN model can be utilized to indicate the impending gearbox failure.

C. Monitoring Results

The EWMA control chart is constructed to monitor APEs of each WT. To demonstrate the effectiveness of the proposed framework, APEs over one week prior to the gearbox failure occurrence are monitored by EWMA charts. The UCL and LCL of the EWMA chart are determined according to (18) and (19). Selected examples are presented in Figs. 6 and 7.

Fig. 6(a) demonstrates the EWMA chart of a healthy WT selected from the Liaoning wind farm, while Fig. 6(b) plots that of a problematic WT in the same wind farms. All EWMA statistics in Fig. 6(a) fall within control limits while outliers are detected in Fig. 6(b). In addition, the first outlier in Fig. 6(b) is observed two days prior to the occurrence of the gearbox failure so that an early alarm can be issued. Fig. 7 provides similar results for the Hebei wind farm.

According to computational results, all five gearbox failures are successfully identified at least one day ahead of the failure occurrence by the proposed monitoring framework. Such early detections offer sufficient time for wind farm operators to take reactions. If the abnormal gearboxes can be timely repaired or replaced, the unnecessary cost and downtime can be avoided.

To further examine the performance of the proposed framework in making false alarms, the data of each WT are examined week by week and the EWMA charts are iteratively applied. A summary of the monitoring results is presented in Table IV. Besides the successful identification of impending gearbox failures, the proposed monitoring framework does not generate false alarms.

APEs can be potentially affected by the shift of gearbox conditions and the model prediction capability. Through examining all APEs obtained based on data of healthy WTs, we discover that more than 99.50% of APEs are located within an interval $[\mu\mathrm{APE} - 3\sigma\mathrm{APE},\ \mu\mathrm{APE} + 3\sigma\mathrm{APE}]$ for all healthy WTs. Hence, detected outliers are majorly caused by the shift of gearbox conditions.

D. Benchmarking Analysis

Since T_o was often considered as the monitoring target in previous studies [18], the proposed framework is also applied to monitor the T_o and generated monitoring results are benchmarked against those presented in Section III-C. A DNN-based model for predicting T_o is provided in

$$\hat{T}_o = \hat{g}(P, O_p, T_s). \tag{20}$$

In (20), $\hat{T}_o$ is the predicted gearbox oil temperature and $\hat{g}(\cdot)$ describes the DNN model.

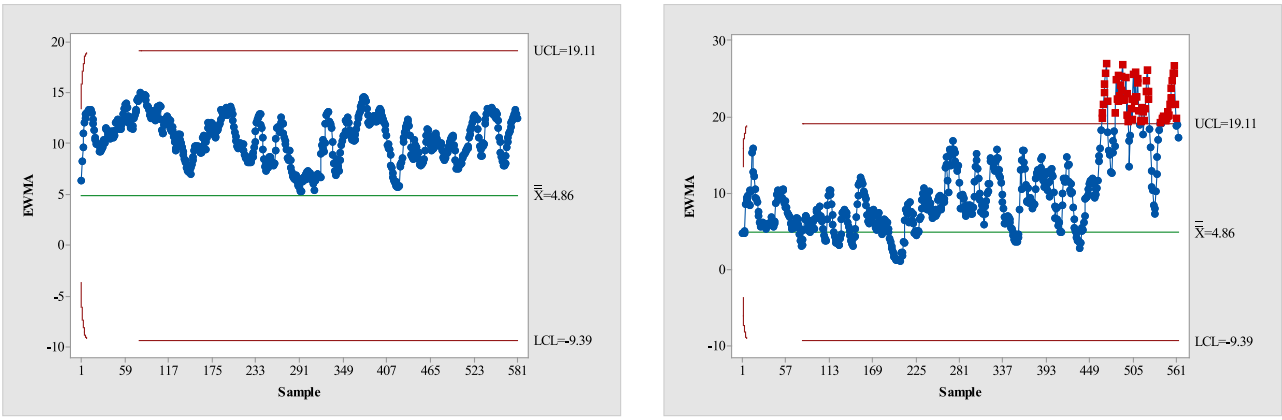


Fig. 6. EWMA chart of wind turbines in Liaoning (May 11–May 17). (a) Wind turbine No. 35, (b) Wind turbine No. 54.

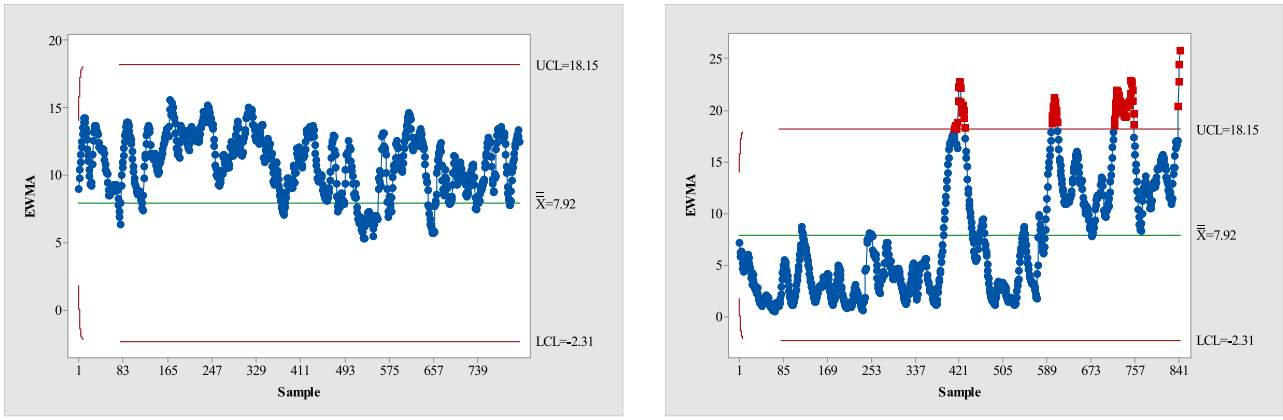


Fig. 7. EWMA chart of wind turbines in Hebei (May 27–June 2). (a) Wind turbine No. 50, (b) Wind turbine No. 64.

TABLE IV
CONFUSION MATRIX OF THE PROPOSED APPROACH

Actual \ Diagnosed	Unhealthy	Healthy
Unhealthy	5	0
Healthy	0	87

TABLE V
CONFUSION MATRIX OF THE ALTERNATIVE APPROACH

Actual \ Diagnosed	Unhealthy	Healthy
Unhealthy	1	4
Healthy	3	84

As similar as the monitoring procedure described in Section II, EWMA charts based on APEs of predicting T_o are developed over weeks stated in Table I for each wind farm. Monitoring results are presented in Table V.

According to Table V, we observe that only one failure case is successfully identified prior to its occurrence based on monitoring the gearbox oil temperature. Three false alarms are produced. Hence, the effectiveness of using the lubricant pressure in the monitoring is further examined.

In summary, the proposed framework can successfully identify all reported failures and does not cause false alarms by monitoring the lubricant pressure. In addition, the parallelized SGD algorithm makes the training process of DNN fast, even with the dropout method. Based on a computer with a CORE I5 CPU and an 8-GB memory, the training of the DNN model with three hidden layers for all six wind farms takes less than 5 mins. Therefore, the proposed framework is applicable to the real implementation.

IV. CONCLUSION

A DNN-based framework for monitoring the WT gearbox health and identifying impending failures based on SCADA data of the gearbox lubricant pressure was presented. The proposed framework was constructed with two phases, modeling and monitoring the lubricant pressure. The first phase targeted on the development of an accurate lubricant pressure prediction model based on the SCADA data of WTs with healthy

gearboxes. A DNN algorithm was employed to develop the prediction model and compared with five benchmarking data-driven models including kNN, Lasso, Ridge, SVM, and NN in terms of MAPE and SDAPE. In the second phase, an EWMA control chart was applied to derive criteria for monitoring the lubricant pressure. The proposed framework was validated with real gearbox failure incidences. In addition, monitoring the lubricant pressure was benchmarked against monitoring the gearbox oil temperature based on the proposed framework.

In the comparative analysis of prediction models, the DNN model offered more accurate prediction results. Thus, it was considered as the more suitable model in monitoring the lubricant pressure. Based on APEs of the DNN model, the EWMA control charts were developed for each WT and utilized to identify impending gearbox failures.

Based on computational results, we observed that APEs of all healthy WTs fell within the monitoring criteria. We also discovered that the impending gearbox failure could be identified two or three days ahead. In addition, monitoring the gearbox oil temperature with the proposed approach was less effective than monitoring the lubricant pressure in detecting impending gearbox failures.

Computational results supported that the lubricant pressure was effective to monitor WT gearboxes based on the proposed framework. A DNN with parallelized SGD training could offer more accurate predictions without significant computational costs.

REFERENCES

[1] J. Nilsson and L. Bertling, "Maintenance management of wind power systems using condition monitoring systems—Life cycle cost analysis for two case studies," *IEEE Trans. Energy Convers.*, vol. 22, no. 1, pp. 223–229, Mar. 2007.

[2] I. Staffell and R. Green, "How does wind farm performance decline with age?" *Renewable Energy*, vol. 66, pp. 775–786, 2014.

[3] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009.

[4] P. Tavner *et al.*, "Study of weather and location effects on wind turbine failure rates," *Wind Energy*, vol. 16, no. 2, pp. 175–187, 2013.

[5] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renewable Sustainable Energy Rev.*, vol. 13, no. 1, pp. 1–39, 2009.

[6] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proc. IEEE Power Electron. Mach. Wind Appl.*, 2009, pp. 1–7.

[7] A. A. Salem, A. Abu-Siada, and S. Islam, "Condition monitoring techniques of the wind turbines gearbox and rotor," *Int. J. Elect. Energy*, vol. 2, no. 1, pp. 53–56, 2014.

[8] W. Liu *et al.*, "The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review," *Renewable Sustainable Energy Rev.*, vol. 44, pp. 466–472, 2015.

[9] Y. Lei, D. Han, J. Lin, and Z. He, "Planetary gearbox fault diagnosis using an adaptive stochastic resonance method," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 113–124, 2013.

[10] J. Li, X. Chen, Z. Du, Z. Fang, and Z. He, "A new noise-controlled second-order enhanced stochastic resonance method with its application in wind turbine drivetrain fault diagnosis," *Renewable Energy*, vol. 60, pp. 7–19, 2013.

[11] I. Antoniadou, G. Manson, W. Staszewski, T. Barszcz, and K. Worden, "A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions," *Mech. Syst. Signal Process.*, vol. 64, pp. 188–216, 2015.

[12] H. Luo *et al.*, "Effective and accurate approaches for wind turbine gearbox condition monitoring," *Wind Energy*, vol. 17, no. 5, pp. 715–728, 2014.

[13] Z. Feng and M. Liang, "Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time–frequency analysis," *Renewable Energy*, vol. 66, pp. 468–477, 2014.

[14] W. Teng, X. Ding, X. Zhang, Y. Liu, and Z. Ma, "Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform," *Renewable Energy*, vol. 93, pp. 591–598, 2016.

[15] J. Rafiee, F. Arvani, A. Harifi, and M. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mech. Syst. Signal Process.*, vol. 21, no. 4, pp. 1746–1754, 2007.

[16] Z. Zhang, A. Verma, and A. Kusiak, "Fault analysis and condition monitoring of the wind turbine gearbox," *IEEE Trans. Energy Covers.*, vol. 27, no. 2, pp. 526–535, Jun. 2012.

[17] G. Smith, "SCADA in wind farms," in *Proc. IEE Colloquium Instrum, Electric. Supply Ind.*, 1993, pp. 11/1–11/2.

[18] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, "Monitoring wind turbine gearboxes," *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013.

[19] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, "SIMAP: Intelligent system for predictive maintenance: Application to the health condition monitoring of a windturbine gearbox," *Comput. Ind.*, vol. 57, no. 6, pp. 552–568, 2006.

[20] Y. Wang and D. Infield, "Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring," *IET Renewable Power Generation*, vol. 7, no. 4, pp. 350–358, 2013.

[21] Y. Park and M. Kellis, "Deep learning for regulatory genomics," *Nature Biotechnol.*, vol. 33, no. 8, pp. 825–826, 2015.

[22] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning," *Nature Biotechnol.*, vol. 33, pp. 831–838, 2015.

[25] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[26] Y. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2011, pp. 18–36.

[27] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *J Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[28] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statistical Soc. Series B*, vol. 73, pp. 267–288, 1996.

[30] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[31] A. Smola and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process Syst.*, 1997., vol. 9, pp. 155–161.

[32] B. Kröse, B. Krose, P. van der Smagt, and P. Smagt, *An Introduction to Neural Networks*. Amsterdam, the Netherlands: Univ. Amsterdam Press, 1993.

[33] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Berlin, Germany: Springer, 2000, pp. 1–15.

[34] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 8609–8613.

[35] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 437–478.

[36] B. Recht, C. Re, S. Wright, and F. Niu, "Hogwild: A lock-free approach to parallelizing stochastic gradient descent," in *Proc. Adv. Neural Inf. Process Syst.,* 2011, vol. 24, pp. 693–701.

[37] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[38] R. J. Tibshirani and J. Taylor, "Degrees of freedom in lasso problems," *Ann. Statist.*, vol. 40, no. 2, pp. 1198–1232, 2012.

[39] D. C. Montgomery, *Introduction to Statistical QualityControl*. New York, NY, USA: Wiley, 2007.

[40] S. S. Prabhu and G. C. Runger, "Designing a multivariate EWMA control chart," *J. Quality Technol.*, vol. 29, no. 1, pp. 8–15, 1997.

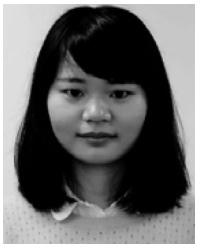
[41] Y. Abu-Mostafa, M. Magdon-Ismail, and H. Lin, *Learning from Data*. Berlin, Germany: AMLBook, 2012.

Long Wang (S'16) received the M.S. degree in computer science with distinction from the University College London, London, U.K., in 2014. He is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong.

His research interests include wind turbine monitoring, electricity price forecasting, computer vision, and reinforcement learning.

Zijun Zhang (M'12) received the B.Eng. degree in systems engineering and engineering management from the Chinese University of Hong Kong, Hong Kong, in 2008, and the M.S. and Ph.D. degrees in industrial engineering from the University of Iowa, Iowa City, IA, USA, in 2012 and 2009, respectively.

He is currently an Assistant Professor in the Department of Systems Engineering and Engineering Management, City University of Hong Kong. His research focuses on data mining and computational intelligence with applications in wind energy, HVAC, and wastewater processing domains.



Huan Long (S'16) received the B.Eng. degree in automation from Huazhong University of Science and Technology, Wuhan, China, in 2013. She is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong.

Her research interests include the wind farm layout planning and analysis, wind turbine monitoring, and wind farm performance robust optimization.

Jia Xu received the M.S. degree in instrument science and technology from Xi'an Jiaotong University, Shaanxi, China, in 2010.

He is currently the Head of the Centre of the Wind Farm Data Analysis and Performance Optimization at China Longyuan Power Group Co. Ltd., Beijing, China.

Ruihua Liu received the B.Eng. degree in Mechatronic Engineering and the M.S. degree in mechanical manufacturing and automation from the North China Electric Power University.

He is currently an Engineer at China Longyuan Power Group Co. Ltd., Beijing, China.