ORIGINALARBEITEN/ORIGINALS

Early stage white etching crack identification using artificial neural networks

Xiaodi Liu[1] · Baher Azzam[1] · Freia Harzendorf[1] · Johann Kolb[2] · Ralf Schelenz[1] · Kay Hameyer[2] · Georg Jacobs[1]

Abstract

White Etching Cracks (WEC) in gearbox bearings is a major concern in the wind turbine industry, which can lead to a premature failure of the gearbox. Though many hypotheses regarding the generation of WEC have been proposed over the decades, the answer is still disputable. To trace back the failures to earlier stages before they occur, an innovative sensor-set has been utilized on a test rig to monitor the influencing factors that lead to WEC. This paperwork seeks to recognize abnormal patterns from recorded sensor data and derive statements of sensible sensor combinations in WEC early detection. A Long Short Term Memory (LSTM) network-based autoencoder is proposed for the anomaly detection (AD) task. Employing an auto-associative sequence-to-sequence predictor, a model is trained to reconstruct the normal time series data without WEC. The reconstruction error of testing time series data is evaluated for the determination of its anomaly. The results show that the specified LSTM autoencoder framework can qualitatively distinguish anomalies from collected multivariate time series data. Moreover, the anomaly score evaluated via reconstruction-error-based metrics can discriminate normal and abnormal behaviors in the study. This investigation's results entail a significant step towards early WEC risk detection and more cost-efficient wind turbine technology if this approach can be further applied on stream data with plausible thresholds in monitoring system.

✉ Freia Harzendorf
   freia.harzendorf@gmail.com

[1] Center for Wind Power Drives, RWTH Aachen University, Campus-Boulevard 61, 52074 Aachen, Germany

[2] Institute of Electrical Machines (IEM), RWTH Aachen University, Schinkelstraße 4, 52062 Aachen, Germany

White-Etching-Crack-Früherkennung unter Verwendung künstlicher neuronaler Netzwerke

Zusammenfassung

Das Wälzlagerschadensphänomen White Etching Cracks (WEC) stellt eine Herausforderung für die Windbranche dar. Es kann zu frühzeitigem Ausfall von Lagern in Getrieben von Windenergieanlagen führen. Obwohl bereits viele Jahre zu diesem Fehlermechanismus geforscht und viele Hypothesen aufgestellt wurden, ist der Entstehungsprozess immer noch nicht umfassend verstanden. Um frühe Stadien dieses Schadensphänomens zu detektieren, wurde ein innovatives Sensorsystem zur Überwachung von möglichen Einflussfaktoren entwickelt. Zur Erprobung wurde es auf einem Wälzlagerprüfstand appliziert. Im Rahmen dieser Veröffentlichung wird eine Methodik entwickelt, mithilfe derer abnormale Muster in den aufgezeichneten Messdaten identifiziert werden können. Auf dieser Basis ist es im zweiten Schritt möglich, eine sinnvolle Sensorkombination zur WEC-Früherkennung abzuleiten. Für die Anomaliedetektion wird ein netzwerkbasierter Autoencoder mit Long Short Term Memory (LSTM) gewählt. Unter Verwendung eines autoassoziativen Sequenz-zu-Sequenz-Prädiktors wird das Modell trainiert, um Zeitreihendaten ohne WEC zu rekonstruieren. Zur Identifizierung von Anomalien in Testdatensätzen wird der Rekonstruktionsfehler herangezogen. Die Ergebnisse zeigen, dass das angegebene LSTM-Autoencoder-Framework Anomalien qualitativ bei der Analyse von multivariaten Zeitreihendaten identifizieren kann. Darüber hinaus kann ein entwickelter Anomalie-Score, als eine Metrik basierend auf der Auswertung von Rekonstruktionsfehlern, normales und abnormales Verhalten unterscheiden. Die Ergebnisse dieser Untersuchung stellen einen bedeutenden Schritt in Richtung einer frühzeitigen WEC-Risikoerkennung und damit einer kostengünstigeren Windturbinentechnologie dar. Hierfür ist der entwickelte Ansatz auf Echtzeitmessungen mit plausiblen Schwellenwerten in das Getriebe Condition-Monitoring-System eine Windenergieanlage zu integrieren.

1 Introduction

A common challenge in wind turbine drivetrain technology is unexpected bearing failures caused by the so-called White Etching Cracks (WEC) [1]. A lot of research effort has been put in addressing a universal agreement on the formation mechanism of WEC over the years. Lubricant composition, presence of additional electrical exposure and mechanical loading are potential drivers of WEC formation summarized in previous literatures [2–4]. Recent research furthermore suggests, a specific time history of applied contact stresses might be the primary driver of early bearing failure due to WEC [5]. The shear band, which generated under compression loading, in interpreting the WEC mechanism is also observed [3, 6]. Electric current has been reported as one of the triggers of early WEC in bearings [7]. More recent research concludes that electric current plays an indirect role in the failure mechanism while an interaction with lubricant accelerate the formation of WEC [8]. Despite considerable effort in root cause research, the failure mechanism is due to the complex interaction of parameters not entirely understood.

This research tackles the bearing reliability problem due to WEC from another perspective. An elementary aspect of improving wind turbines' reliability is to detect these failures or their preliminary stages as early as possible. Therefore, an innovative sensor-set has been utilized to monitor the influencing factors that lead to WEC failure. The sensors-set was implemented in several measurement campaigns at a bearing component test bench [7]. Following each campaign, the bearing was closely examined to deter-mine whether WEC failure took place or was initiated. By analyzing multivariate data collected from the conducted measurement campaigns, a premature detection of the onset of WEC is hypothesized.

Analyzing multivariate data with the aim of detecting anomalies in time series captured by multi-sensors in a mechanical device is a popular task in literature these days. Many techniques have been developed within diverse research areas and application domains [9]. Gaussian-mixture-model based probabilistic approaches, similarity measurements such as Euclidean distance or Mahala-Nobis distance, novelty detection technique using Support Vector Machine (SVM) are frequently employed techniques [9–11] Modelling and detecting anomalies in rolling bearing elements of these mechanical devices is a major topic in practice. In the following the most important sources are presented. First sources having a bearing application are introduced. A novel method applying transition probabilities between system states on rolling bearing vibration data has been proposed to detect vibrational anomalies in rolling bearings, where those probabilities are provided by a markov chain [12]. In another recent research, an AD method employed a classification technique to discriminate between defect examples of rolling bearings using kurtosis and non-gaussianity score, and the proposed method increased the general sensibility in bearing fault diagnosis[13]. Other sources referred to failures' detection in industrial units within components like engines or turbines are also investigated. Multivariate time series assembled in these scenarios are mostly temporal dependent and highly correlated to generated structure damage; thus, normal

data are more robust over time, and reconstruction-based anomaly detection techniques benefit in dealing with relevant situations. Jinwon and Sungzoon [14] recommend using a variational autoencoder and utilize its generative characteristics to address anomalies, which outperforms the conventional principal components method. Instead of reconstruction error, reconstruction probability is used. Similarly, an autoencoder-model based method for condition monitoring of components such as bearings, rotors in a rotating machine is proposed to detect anomalies. Characteristics of machines are learned via normal vibration signals [15]. According to Sakurada [16], autoencoders detect anomalies in the spacecraft's telemetry data via learning normal state and manifest differently when input is anomalous. Autoencoder still suggests a superior performance without complex kernel computations requirements as in principle components analysis (PCA). Joao and Margarida's [17] investigations show the ability of a recurrent neural network and the Long-Short-Term-Memory (LSTM) network-based autoencoder detecting anomalous behavior in time series data collected from a smart sensor system for solar energy. Malhotra [18] proposed a similar sequence to sequence (Seq2Seq) method with LSTM encoder-decoder framework and computed the anomaly score via distributions of reconstruction errors. Park also concludes in [19], that an LSTM based variational autoencoder reconstructs sensory signals in their expected distributions. The anomalies are detectable using either score or state-based threshold.

According to the presented literature, various anomaly detection techniques have been employed on multi-sensor data to detect the abnormal patterns of industrial structural units. Nevertheless, self-adaptive models using deep learning networks for anomaly detection have not yet been applied in the WEC diagnosis. Although a condition monitor prototype in WEC early detection has been investigated and proposed in previous research, the exact sensor concept is not claimed yet. Thus, a gap exists between the state of the art and this research's goal. This research intends to utilize an LSTM network-based autoencoder appliance to detect anomalous patterns in a data set sampled with an innovative sensor set-up for WEC initiation.

In this paper, two models are developed and applied to the proposed autoencoder anomaly detector. The sensor data is collected through a component test rig. Temporal dependencies within input are considered for both encoder and decoder using serially connected LSTM layers. The autoencoder model is initially trained to reconstruct normal behaved time series data. In a next step, anomalies are identified based on specified metrics on reconstruction errors of testing time series data. Sensitivity analysis (SA) regarding the influential sensors in detecting the anomalies is included in the research. Specified reconstruction error met-rics are identified to determine thresholds that distinguish healthy and unhealthy states. This investigation's results entail a significant step towards early WEC risk detection and more cost-efficient wind turbine technology.

The remainder of the paper is organized as follows. The employed multi-sensor data and the reasons for measuring them are described in Chap. 2. Chap. 3 presents the methodology parts, including the LSTM autoencoder scheme and evaluation metrics. In Chap. 4, a workflow for the anomaly detection task is introduced. Then, modeling results are discussed in both qualitative and quantitative perspectives in Chap. 5. Finally, conclusions are drawn and summarized in Chap. 6.

2 Data description

The employed data for the proposed anomaly detector is collected through campaigns on a component test rig described in [7]. In the test rig, two rolling element bearings of type 6203 are tested [7, 20]. The test runs are conducted under conditions, which are claimed critical for the initiation of WEC. All experiments used in this paper are performed at a speed of 4500 rpm. For all tests, an axial pre-load of 1800 N was applied on the bearings. According to Evans et al. a transient load force condition on bearings is a potential driver leading to WEC [2]. Examples are load reversals or high loads. The presence of additional electrical exposure besides mechanical loading is mentioned as another possible WEC driver in literature [2] and is varied with the help of an external current source in the test runs. Furthermore, lubricant composition can influence the occurrence of WEC as oxidative decomposition of lubricants has been claimed to be a hydrogen generation mechanism in several studies [2, 3]. As hydrogen embrittlement was quoted as one big factor in WEC formation. According to authors in [5], the cracks are initiated due to the local plasticity promoted by hydrogen. Therefore, two lubricants with different compositions are used in the test runs (A & B). Both lubricant oils are specifically formulated to cause WEC, however due to confidentiality reasons involving IP rights, it is not possible to describe the exact oil compositions; however, the oil compositions used in a previous investigation [21] provide insights into the used lubricants. A high lubricant volume flow rate helps in increasing the thickness of generated lubrication film. Low lubrication film thickness leads to mixed friction and the possible formation of a tribo-chemical reaction layer. The distribution as well as chemical makeups of the tribo-film has a significant influence in friction and subsurface stress distribution. Some researches imply that presence of such tribo-film drives hydrogen in to high stress zone and eventually leads to WEC [2, 22]. However, the relationship

Table 1 Operational conditions of nine employed test runs collected from test rig

Test run	Volume flow rate [ml/min]	Rotational speed [rpm]	Axial load [N]	Voltage [V]	WEC [y/n]
1A	40 → 15	4500	1800	15	n
2A	40 → 8	4500	1500	15	n
3A	4	4500	1800	15	n
1B	4	4500	1800	0	n
2B	4	4500	1800	15	n
3B	4	4500	1300	15	n
4B	4 → 0 → 4	4500 → 0 → 4500	1800 → 0 → 1800	15 → 0 → 15	y
5B	10	4500	1800	15	y
6B	9 → 28 → 9 → 38 → 4	4500	1800	15	y

Table 2 Relevant sensory variables for WEC detection analysis

Variables	Description
Temp. bearing shaft 1	*Measured temperature of outer ring at bearing 1 through entire running time*
Temp. bearing shaft 2	*Measured temperature of outer ring at bearing 2 through entire running time*
Mean Temp. of bearings	*Calculated mean temperature value of both bearing outer rings through entire running time*
Axial load	*Applied pre-load on bearings through entire running time*
Rotational speed	*Applied rotational speed of the motor power through entire running time*
Volume Flow Rate	*Varied lubricant flow rate in the recirculating system*
Bearing Voltage	*Measured discharges between two bearings by voltmeter value through entire running time*
Resistor Voltage	*Measured regulated resistor voltage by voltmeter value through entire running time*
Pressure at storage container	*Environmental pressure set up around the storage container in bars*
Temp. storage container	*Measured temperature of lubricants in the recirculating system when it's passing back to the chamber*
Temp. chamber input	*Measured temperature of lubricants in the recirculating system when it's coming into the chamber*

between tribo-film and WEC formation mechanisms is not yet determined [22, 23]. Therefore, lubricant volume flow is varied during the test runs. Table 1 gives an overview of the operational conditions applied during the nine test runs as well as an indication of whether this test run leads to WEC in the end. The arrows in the table indicate, that within the test run the operational conditions have been varied from the value on the left of each arrow to the value on the right of the same arrow.

In this work, the following recorded sensor signals (please compare Table 2) are specified as input data for the anomaly detection analysis. They are chosen as either their unusual pattern might potentially give an insight on the generation of WEC or they describe the operational conditions of the bearing. Temperatures of two tested rolling bearings are monitored because a significant rise of temperature can indicate a bearing failure, in this case, WEC. The axial load force, rotational speed of the motor and volume flow rate are typical operational conditions of a gearbox. The measured discharge in bearings is another relevant sensory signal which is measured by a voltmeter, a serial connected resistor of 75 kΩ is applied to regulate the simulated current below certain value. Additionally, the environmental pressure of the test rig chamber and lubricant's temperatures before and after flowing into the chamber are respectively measured and recorded. The sen-

sory dataset contains 11 variables overall, they are relevant for the WEC detection analysis and showed in Table 2 below. Apart from those, sensory signals recording acoustic emissions energy and oil conditions were recorded as well, however the corresponding analysis results will not be presented in this paper.

3 Methodolog ment to characterize h y

3.1 LSTM autoencoder scheme

The reconstruction-based approaches have proved to deliver satisfying results in detecting industrial damaged units according to the literature review in Chap. 2. Autoencoder, as one of the commonly used reconstruction-based methods, enables a Seq2Seq reconstruction modeling structure. In this study, healthy test runs map normal pattern while failed test runs cover biased pattern. Fig. 1 displays the structure of the proposed model. Both encoder and decoder are composed of a few stacked LSTM> layers as depicted. The encoder learns a compressed vector representation c with a fixed length of the input data sequence x. Simultaneously, the decoder predicts an output data sequence $\widetilde{x}$ with the compressed representation and parameters updated in the current hidden state of the decoder LSTM layer. Hence,

Fig. 1 Reconstruction model structure with encoder and decoder consists of stacked LSTM layers
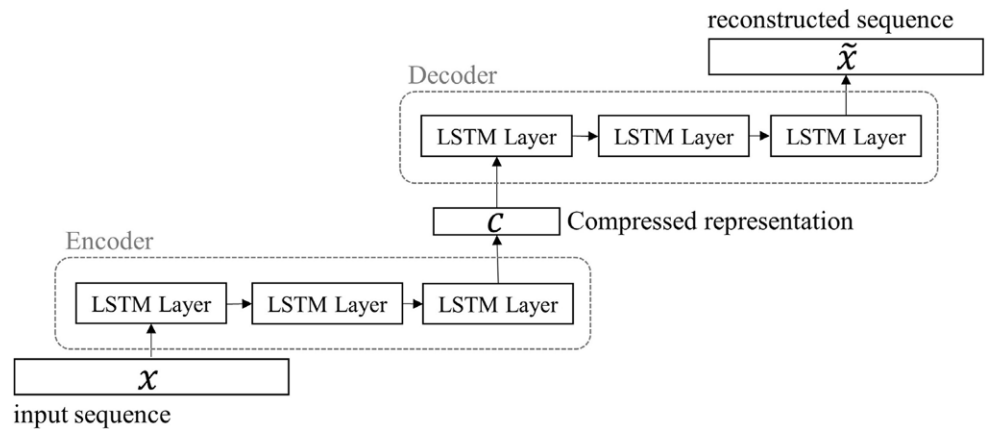
this predicted output is a reconstructed version of the original input data sequence.

In both encoder and decoder, LSTM deep learning networks are applied to memorize long term dependent relationships in time series and avoid vanishing gradients problem [24, 25]. It is realized through dedicated gates control g_i for input constraint, g_f for forgetting less important information and g_o for output constraint over hidden unit states h, namely the short-term memory unit. For large time series data, LSTM cells can capture temporal ordering information by keeping important previous states in short term memory units. The calculation of LSTM units for input of next timepoint based on current information and temporal information remembered from previous inputs is implemented along [26] and summarized in the following equations (1–6). With x^t being the current input and h^{t-1} being the output of previous hidden unit respectively:

$$\widehat{c} = \tanh(W\left[h^{t-1}, x^t\right] + b) \tag{1}$$

$$g_i = \sigma(W_i\left[h^{t-1}, \ x^t\right] + b_i) \tag{2}$$

$$g_f = \sigma(W_f\left[h^{t-1}, \ x^t\right], + b_f) \tag{3}$$

$$C_t = g_f \cdot C_{t-1} + g_i \cdot \widehat{C} \tag{4}$$

$$g_o = \sigma(W_o\left[h^{t-1}, \ x^t\right] + b_i) \tag{5}$$

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

$\widehat{c}$ represents the new information at the current moment, W represents single weight matrix for both input vector and weight vector, and b a biased term. In this study, a healthy test run sequence $X = \{x^1, x^2, x^3...x^N\}$ of length N where each x^i is the vector with the same number of variables at time instance t_i, as input. On the encoder side, the network is aimed at memorizing important information from previous moments. A current input vector x^i and previous short-term hidden state h^{i-1} can determine the hidden state h^i for time instance t_i follow the equations (1) to (6). A sequence of forward hidden states $h^E = \{h^1, h^2, h^3..., h^N\}$ is step-by-step calculated. The last hidden state h^N is assigned to the compressed vector c shown in Fig. 1 and works as the initial state of decoder h^D. Analog is the output reconstructed in reverse order $\{x^N, x^{N-1}, x^{N-2}...x^1\}$. On the last decoder LSTM layer, the time series is reconstructed through weight vectors at each hidden unit and the final output gate control: $X\prime = \{x^{1\prime}, x^{2\prime}, x^{3\prime}...x^{N\prime}\}$.

The model performance is optimized by minimizing the reconstruction error between output $X\prime$ and input X. One or more healthy test runs are merged as a training set. Before training starts, 10 percentage of training data is taken out as a validation set, and a hyper-parameter search is carried out to find each model's best configuration. Seven hyperparameters have been variated for the configuration search: hidden layer number, hidden unit number, batch size, epoch number, activation function, dropout rate and kernel regularizer [27, 28]. They influence the model performance in under-fitting and over-fitting, convergence speed, computational cost, and stability. The result is evaluated by the root mean squared reconstruction error of the validation set. The model is thereafter trained with an optimized configuration set. Trained models are evaluated by convergency and training time. Afterward, remained test run sequences data are predicted by the trained model in the testing phase. Reconstruction error computed after testing is evaluated by metrics discussed in the following section.

3.2 Evaluation metrics

In this work, both root mean squared reconstruction error (7) and mean absolute reconstruction error (8) between the input and output are computed for a qualitative evaluation.

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \widetilde{x}_n)^2} \tag{7}$$

$$MAE = \frac{1}{N} \cdot \sum_{n=1}^{N} |x_n - \widetilde{x}_n| \tag{8}$$

Moreover, the corresponding error distributions are determined and analyzed on characteristics of skewness S_{sk}, kurtosis S_{ku} and mean values with formulas defined in Equation (9) and (10).

$$S_{sk}(x) = \frac{\frac{1}{N} \cdot \sum_{n\_1}^{N} (x_n - \overline{x}_n)^3}{s^3} \qquad (9)$$

Here, s is the standard deviation of reconstruction error and N is the number of data instances of one time series data. The above formula is referred to the Fisher-Pearson coefficient of skewness [29]. Kurtosis is a measurement to characterize how heavy-tailed a distribution is compared to a normal distribution and the applied formula is [30]:

$$S_{ku}(x) = \frac{\frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \overline{x}_n)^4}{s^4} \qquad (10)$$

4 Implementation anomaly detection workflow

In this work, two models are defined for the anomaly detection task. The sensory signals are 10 operational conditions variables described in Chap. 2. One model is trained with a single healthy run while the other is trained with multiple healthy and failed test runs.

Fig. 2 shows the anomaly detection task's workflow: One starts with historical data acquisition, then a proper preprocessing on it. The split of training and testing set is coming afterward. The validation set is used for a hyper-parameter search before training, as discussed in Sect. 3.1. Training and testing phases come directly after the search process. The last step is the evaluation of the autoencoder with metrics on the reconstruction error. In addition is a sensitivity analysis on the relative variable importance in the learned neural network. The results of the two models are discussed in the following sections.
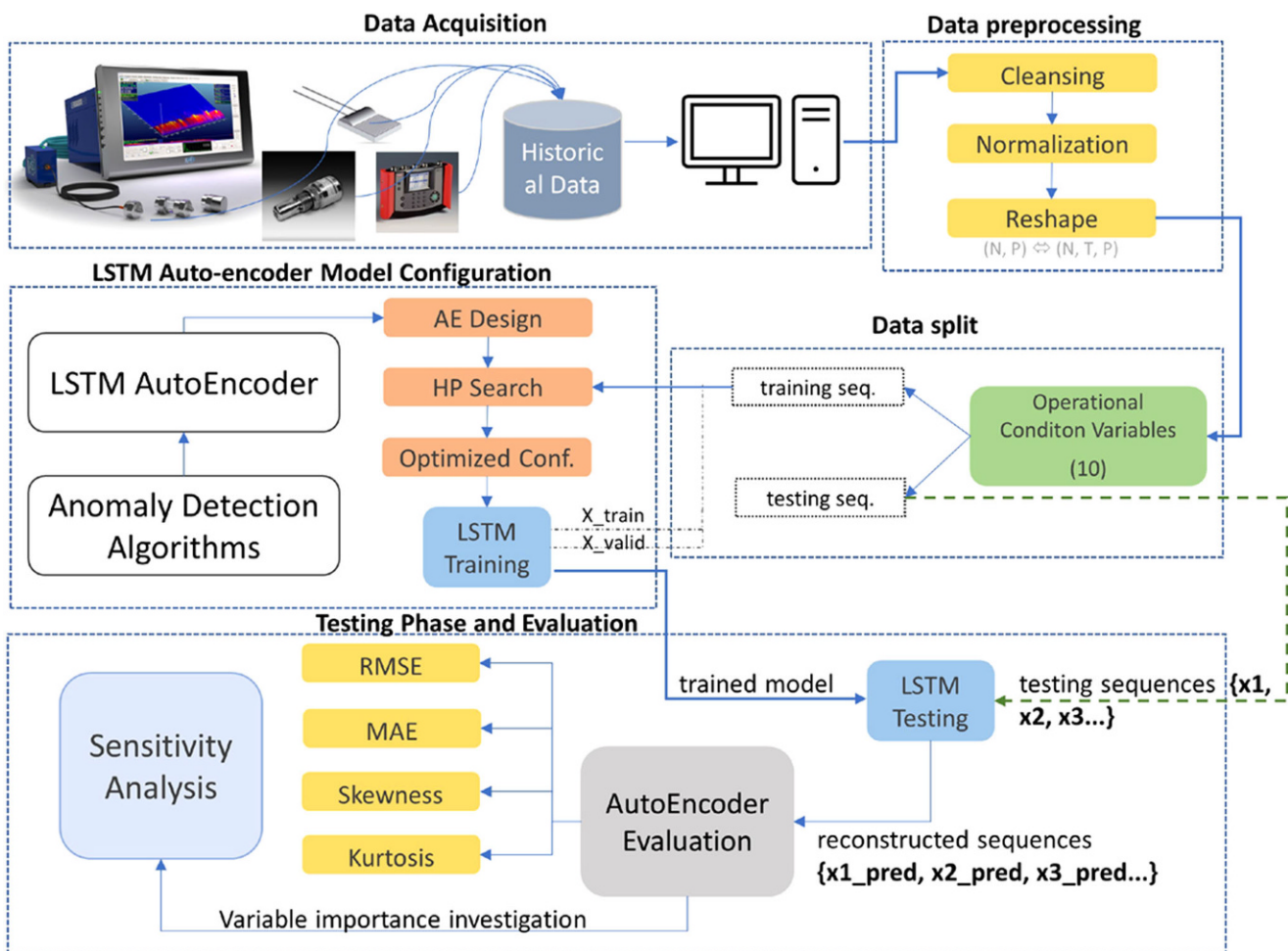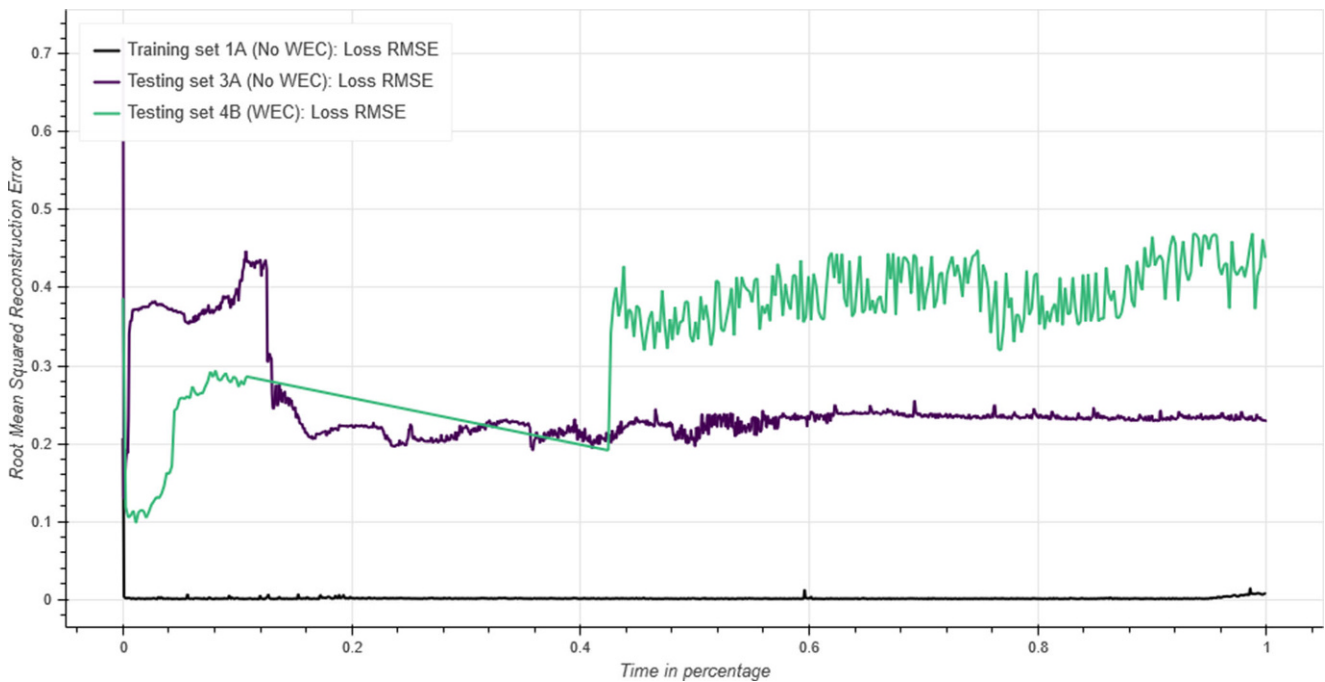

Fig. 2 The workflow of the anomaly detection task in the research

Fig. 3 Averaged reconstruction error (ARE) of training set 1A (healthy), testing set 3A (healthy) and 4B (WEC)

5 Results and discussions

5.1 Qualitative analysis

Fig. 3 shows the averaged reconstruction error (ARE) of the model trained with a single healthy test run 1A. Except for the ARE of training set 1A, this figure shows the ARE of two testing sets: one healthy test run 3A and one failed test run 4B.

With a normalized time-axis, the reconstruction processes can be compared employing the error magnitude. One can demonstrate that the reconstruction error of training set 1A is low and flatten around zero; this suggests a well-performed reconstruction model learned normal pattern.

Fig. 3 demonstrates that ARE of the healthy test run 3A is higher in the beginning compared with remained time; it decreases around 17% of its process. This possibly lies with a switched volume flow rate in training set 1A from 40 to 15 ml/min at 24 h over 140 h described in Table 1. Those reconstruction error curves yield expected results since 3A is a healthy test run, thus an overall low error curve implies a potential normal pattern of the bearing's behavior. When comparing ARE curves of the failed test run 4B in light green and the healthy test run 3A, a significant error rise around 41% percentage of the processing time is observed as predicted conditions strongly deviated by then. Firstly, it suggests that a failed test run is detectable via considerable higher reconstruction error. It further indicates a potential

start point of such anomalous behavior, which can be sought confirmation from experts for more rigorous inferences.

As test runs are performed with two lubricants A and B, it is of natural interest to study the model's ability in learning the difference. Therefore, multiple healthy test runs consist of either oil A or oil B are merged as one training set, then similarly tests with remained test runs. Fig. 4 shows though ARE is generally higher than the model trained with smaller dataset, failed test runs still have higher ARE than healthy test run. Also, significant increasing trends of ARE are observed in both failed test runs. From the qualitative analysis, one can conclude that the specified auto-associative model can distinguish failed test runs from healthy ones by interpreting reconstruction error curves.

5.2 Quantitative analysis

Furthermore, ARE distribution and its static characteristics are computed so that a quantitative analysis of the model performance is achieved. Fig. 5 first compares the error distribution histograms of models discussed in the last section. One can perceive from Subfigure (a): That the healthy test run 3A shows a prominent peak around 0.15 and narrowed to a small interval compared to the others. On the contrary, the histograms of failed test run 4B, 5B and 6B are wider spread and suggest multimodal histograms. On Subfigure (b), test runs' ARE are higher while the healthy test run 3A still delivers the least error. These findings tie with the previous qualitative analysis. In an anomaly detection task, the determination of a reasonable anomaly score for

Fig. 4 ARE of testing set 3A (healthy) and 4B (WEC), 5B (WEC), trained with dataset consist of test runs conditioned with oil A and oil B
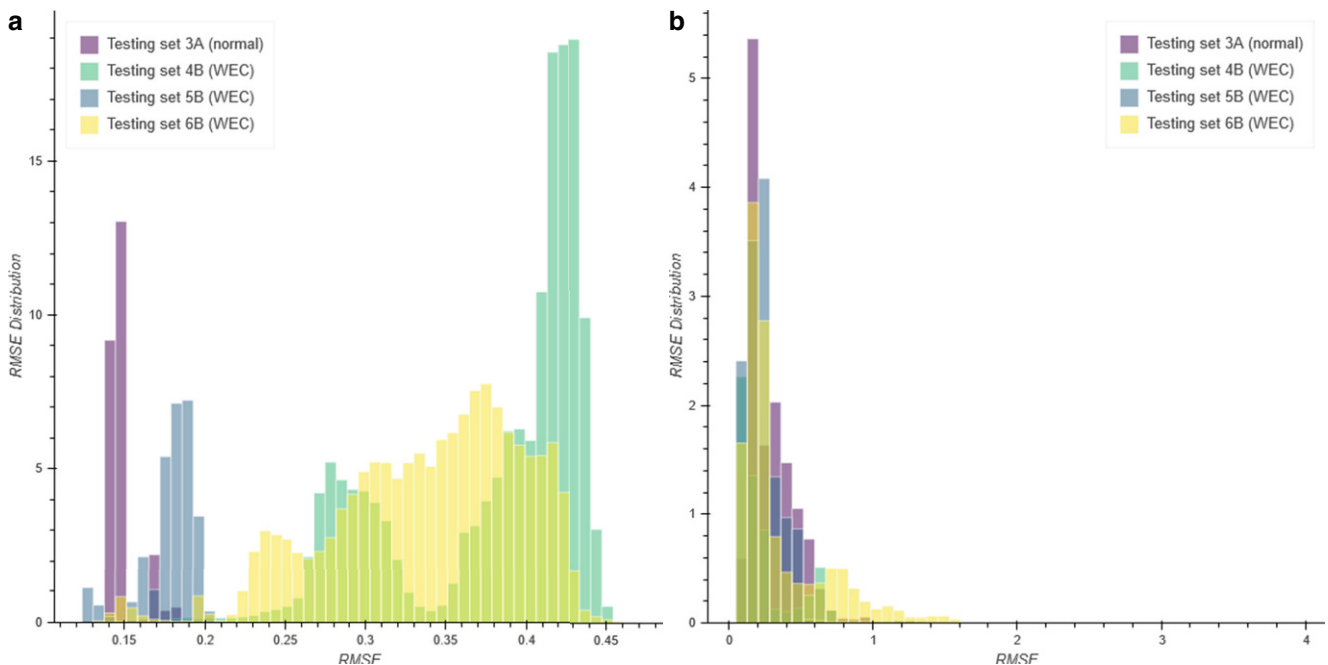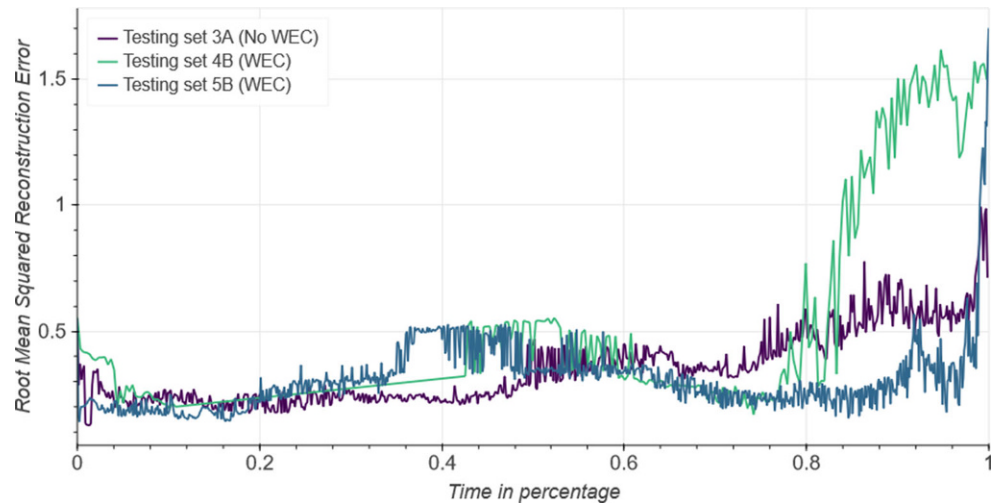
Fig. 5 Root Mean Squared Reconstruction Error distribution of model trained with a single healthy test run (**a**) and model trained with merged test runs (**b**)

data instance is the goal. In this study, an investigation on reconstruction error distribution characteristics is utilized to achieve this purpose. Mean value, as well as the kurtosis and skewness values of ARE distribution are computed and presented in bar graph Fig. 6.

Among all 8 testing sets, 4 healthy test runs yielded larger positive skewness than 3 failed ones. The reconstructed healthy data sequences are more right-skewed than failed ones. This healthy test run 1B is a false-positive case, as the corresponding skewness is less than other healthy test runs. Similarly, the kurtosis values, which indicate the tail heaviness of a distribution, strengthens the interpretation: healthy test runs except for the false-positive case 1B are

lighter in tails and peakier. Fig. 6 illustrated that healthy test runs (shown in blue) have demonstrated apparently higher kurtosis and skewness values than the failed test runs (shown in orange). One can assume that superior analysis results could be brought about when more test runs are available.

5.3 Sensitivity analysis

Considering the computational efficiency, one-factor-at-a-time (OAT) based SA is performed to determine how sensitive the auto-encoder is to each sensor channel input during learning. The principle of an OAT method is to study the
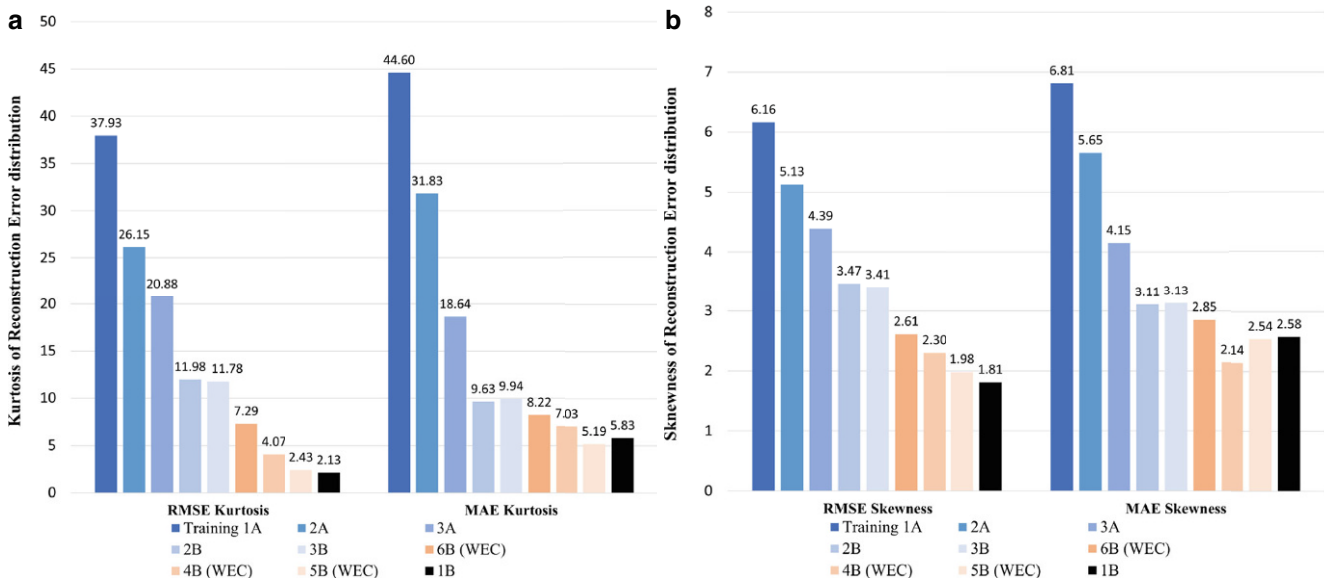
Fig. 6 Kurtosis (**a**) and skewness (**b**) in bar charts for all 8 test run sequences

contribution of input variables one by one [31]. A change in the error function when the specific input variable is removed from the network measures its predictive importance directly [32]. In this work, an OAT SA is applied for generated models.

One sensor variable is first clamped to its mean value, and then the model will be retrained. The varied RMSE before and after replacing each sensor variable is computed. Subsequently, changes in RMSE are ranked for all sensor variables. The corresponding ranking of the second model discussed in Sects. 4.1 and 4.2 is depicted in Fig. 7. This analysis implies the increasing sensitivity of relevant variables in the LSTM autoencoder. Accordingly, the tempera-

ture at bearing shaft 2 is the most influential factor in the reconstruction procedure with an RMSE difference at 0.032. Rotational drive speed is the least considered variable and changed in RMSE at 0.0019.

Given that the resulting ranking is based on constrained considerations of input and merely takes the error function into account, the complex parameters computations and updates in the network remained a black box; the analysis results should be treated with the utmost caution.
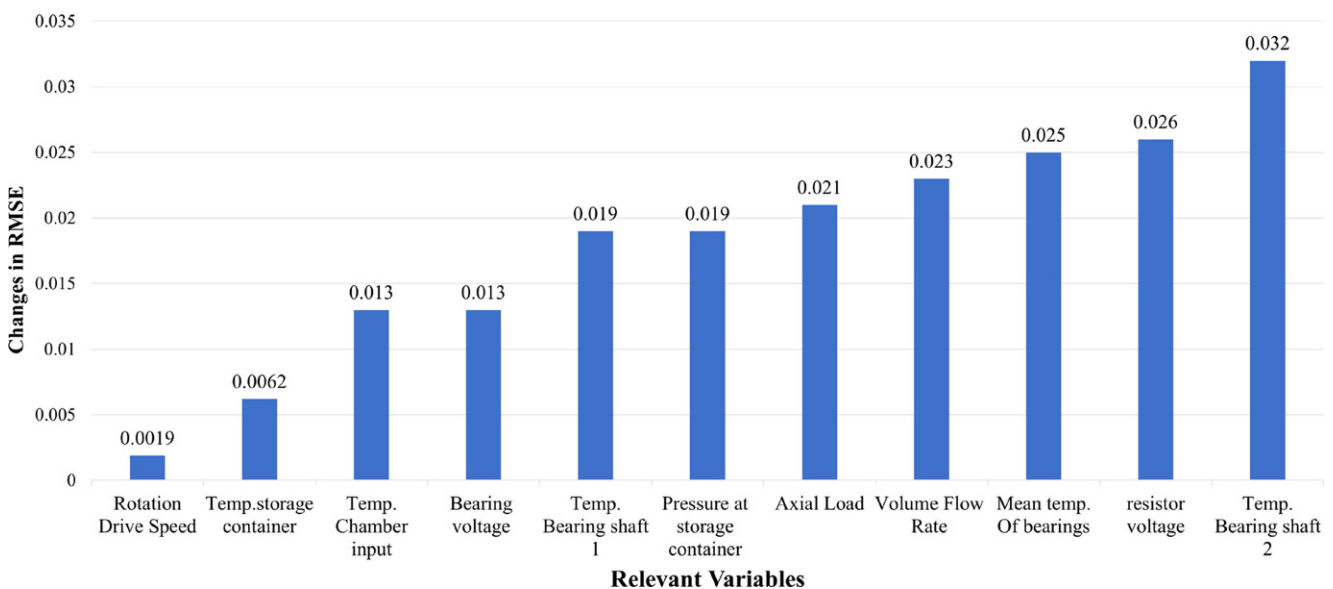

Fig. 7 Increasing RMSE changes regarding sensor variables

6 Conclusions

In this paper, an anomaly detection approach based on LSTM neural network is proposed to identify abnormally behaved data indicating an occurrence of WEC in gearbox bearings. By modeling multivariate time series data collected from the sensor-set described in Chap. 2, the proposed anomaly detector distinguishes tests resulting in WEC from tests without WEC via evaluations over the model performances. In this work, the autoencoder reconstruction models were first trained on normal patterns through healthy time series data before being tested on previously unseen data from both healthy test runs, and test runs which led to failure. The models along with the subsequent analyses of their results as outlined in the proposed method allowed the detection of anomalies in test runs which eventually reached the point of failure.

Additionally, a sensitivity analysis examined the influence of each input variable on the performance of the developed methodology. The results of this analysis may reduce the number of required sensors which would reduce the implementation costs of the proposed method to a new condition monitoring system for WEC detection in operational wind turbines. The proposed methods require significantly more time and computational effort in the development phase than in the implementation phase. This is due to the required training of the autoencoder models during development. Though the proposed method was trained and tested solely on historical data in this investigation, the algorithms used may also be applied in real-time on streamed data.

Before field deployment, the authors recommend the implementation of the required sensor setup to operational wind turbines in the field and using the accumulated data to further refine the thresholds used in the proposed method. As additional sources of noise may be present on the field, performing this step would introduce the autoencoder models to such noise in the development phase which is likely to limit the occurrence of false alarms due to noise. Additionally, a potential drawback of the study is that the test runs do not consider other failure modes, making it problematic if one intends to address the abnormality detected by model to WEC solely without considering other failure modes or even the interaction of them. This limitation remained unsolved as the test rig was set up for WEC initiation and the collected data sets are, therefore, primarily WEC-relevant. Despite of this shortcoming, the proposed method provides a promising tool for early detection of WEC failures in future wind turbine condition monitoring systems which address this costly failure type.

References

1. Evans M-H (2012) White structure flaking (WSF) in wind turbine gearbox bearings: effects of 'butterflies' and white etching cracks (WECs). Mater Sci Technol 28(1):3–22. https://doi.org/10.1179/026708311X13135950699254

2. Evans M-H (2016) An updated review: white etching cracks (WECs) and axial cracks in wind turbine gearbox bearings. Mater Sci Technol 32(11):1133–1169. https://doi.org/10.1080/02670836.2015.1133022

3. Holweger W, Wolf M, Merk D, Blass T, Goss M, Loos J, Barteldes S, Jakovics A (2015) White etching crack root cause investigations. Tribol Trans 58(1):59–69. https://doi.org/10.1080/10402004.2014.942938

4. Ščepanskis M, Jakovičs A, Kaldre I, Holweger W, Nacke B, Diederichs AM (2015) The numerical model of electrothermal deformations of carbides in bearing steel as the possible cause of white etching cracks initiation. Tribol Lett. https://doi.org/10.1007/s11249-015-0564-8

5. Manieri F, Stadler K, Morales-Espejel GE, Kadiric A (2019) The origins of white etching cracks and their significance to rolling bearing failures. Int J Fatigue 120:107–133. https://doi.org/10.1016/j.ijfatigue.2018.10.023

6. Su Y-S, Li S-X, Yu F, Lu S-Y, Wang Y-G (2021) Revealing the shear band origin of white etching area in rolling contact fatigue of bearing steel. Int J Fatigue 142:105929. https://doi.org/10.1016/j.ijfatigue.2020.105929

7. Zuercher M, Heinzler V, Schlücker E, Esmaeili K, Harvey TJ, Holweger W, Wang L (2018) Early failure detection for bearings in electrical environments. Int J Cond Monit 8(1):24–29. https://doi.org/10.1784/204764218822441997

8. Tung P-Y, McEniry E, Herbig M (2021) The role of electric current in the formation of white-etching-cracks. Philos Mag 101(1):59–76. https://doi.org/10.1080/14786435.2020.1819578

9. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection. ACM Comput Surv 41(3):1–58. https://doi.org/10.1145/1541880.1541882

10. Keogh EJ, Smyth P (eds) (1997) A Probabilistic approach to fast pattern matching in time series databases. AAAI Technical Report WS-98-07, Department of Information and Computer Science. University of California, Irvine

11. Davy M, Desobry F, Gretton A, Doncarli C (2006) An online support vector machine for abnormal events detection. Signal Process 86(8):2009–2025. https://doi.org/10.1016/j.sigpro.2005.09.027

12. Harris BW, Milo MW, Roan MJ (2016) A general anomaly detection approach applied to rolling element bearings via reduced-dimensionality transition matrix analysis. Proc Inst Mech Eng Part C J Mech Eng Sci 230(13):2169–2180. https://doi.org/10.1177/0954406215592439

13. Purarjomandlangrudi A, Ghapanchi AH, Esmalifalak M (2014) A data mining approach for fault diagnosis: an application of anomaly detection algorithm. Measurement 55:343–352. https://doi.org/10.1016/j.measurement.2014.05.029

14. Jinwon An SC (ed) (2015) Variational autoencoder based anomaly detection using reconstruction probability. , SNU Data Mining Center. http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf

15. Ahmad S, Styp-Rekowski K, Nedelkoski S, Kao O (2021) Autoencoder-based condition monitoring and anomaly detection method for rotating machines. Distributed and Operating Systems. TU Berlin. https://www.researchgate.net/publication/348832846_Autoencoder_based_Condition_Monitoring_and_Anomaly_Detection_Method_for_Rotating_Machines

16. Sakurada M, Yairi T (2014) Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Rahman A, Deng J, Li J (eds) Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis MLSDA'14, New York. ACM Press, New York, pp 4–11

17. Pereira J, Silveira M (2018) Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In: 17th IEEE International Conference 17 Dec 2018, pp 1275–1282

18. Malhotra P, Ramakrishnan A, Anand G, Vig L, Agarwal P, Shroff G (2016) LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv:1607.00148v2

19. Park D, Hoshi Y, Kemp CC (2017) A Multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. arXiv:1711.00614v1

20. Esmaeili K, Zuercher M, Wang L, Harvey T, Holweger W, White N, Schluecker E (eds) (2017) Advanced signal processing techniques for wind turbine gearbox bearing failure detection. British Institute of Non-Destructive Testing, Northampton, UK

21. Gould B, Demas NG, Pollard G, Rydel JJ, Ingram M, Greco AC (2019) The effect of lubricant composition on white etching crack failures. Tribol Lett. https://doi.org/10.1007/s11249-018-1106-y

22. Haque T, Korres S, Carey JT, Jacobs PW, Loos J, Franke J (2018) Lubricant effects on white etching cracking failures in thrust bearing rig tests. Tribol Trans 61(6):979–990. https://doi.org/10.1080/10402004.2018.1453571

23. Stadler K, Lai J, Vegter RH (2014) A review: the dilemma with premature white etching crack (WEC) bearing failures. In: Beswick JM (ed) Advances in steel technologies for rolling bearings. Bearing steel technologies, vol 10. ASTM International, West Conshohocken, pp 1–22

24. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

25. Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. Neural Comput 12(10):2451–2471. https://doi.org/10.1162/089976600300015015

26. Graves A (2012) Supervised sequence labelling with recurrent neural networks, 1st edn. Springer, Berlin, Heidelberg

27. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. IEEE Trans Neural Netw Learning Syst 28(10):2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

28. Reimers N, Gurevych I (2017) Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. arXiv:1707.06799v2

29. Brown S (2020) Measures of shape: skewness and kurtosis. https://brownmath.com/stat/shape.htm. Updated 26 Oct 2020

30. Westfall PH (2014) Kurtosis as peakedness, 1905–2014. R.I.P. Am Stat 68(3):191–195. https://doi.org/10.1080/00031305.2014.917055

31. Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. Environ Model Softw 25(12):1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

32. Sarle WS (2000) How to measure importance of inputs. SAS Institute, Cary