PAPER • OPEN ACCESS

Predictive maintenance of wind turbine's main bearing using wind farm SCADA data and LSTM neural networks

To cite this article: Y Vidal *et al* 2023 *J. Phys.: Conf. Ser.* **2507** 012024

Predictive maintenance of wind turbine's main bearing using wind farm SCADA data and LSTM neural networks

Y Vidal [1,2], B Puruncajas [1,3], F Castellani[4] and C Tutivén[3,5]

[1] Control, Data, and Artificial Intelligence (CoDAlab), Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC), Campus Diagonal-Besós (CDB), Eduard Maristany, 16, 08019 Barcelona, Spain
[2] Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Pau Gargallo 14, 08028 Barcelona, Spain
[3] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, Faculty of Mechanical Engineering and Production Science (FIMCP), Mechatronics Engineering, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
[4] Department of Engineering, University of Perugia, Perugia, Italy
[5] Universidad Ecotec, Km. 13.5 Samborondón, Samborondón, EC092302, Ecuador.

E-mail: `yolanda.vidal@upc.edu`, `bpurunca@espol.edu.ec`, `francesco.castellani@unipg.it`, `cjtutive@espol.edu.ec`

Abstract. Field failures of wind turbine main bearings cause unwanted downtime and significant maintenance costs. Currently, this industry seeks to increase its reliability, for which condition monitoring and predictive maintenance systems have been adopted. In most industrial wind farms, the integrated Supervisory Control and Data Acquisition (SCADA) system provides data that is stored averaged every 10 minutes that can be used to quantify the health of a wind turbine (WT). This research presents a framework for the analysis of data collected from the SCADA system of an operating wind farm, aiming to early detect the main bearing failure using a Long-Short-Term Memory (LSTM) neural network. For prediction, SCADA variables of the temperature of turbine components near the main bearing, rotor speed, ambient temperature, and generated power are taken into account. The results show that the proposed methodology can detect the target failure up to 4 months in advance of the fatal breakdown. The results obtained confirm the applicability of the proposed model in real scenarios that can help the operator with enough time to make more informed maintenance decisions.

1. Introduction

Wind energy is one of the most widely used renewable energy resources with low environmental impact in the world for the generation of electricity. High growth is estimated worldwide, with 1000 GW installed by 2050 with a WT installation rate of 200 GW/year [1]. The main challenge of wind energy is to achieve low operating and maintenance costs (O&M). It is estimated that O&M in wind farms can represent 30% of the total cost of energy due to its remote location, difficult transportation, components and logistics costs, and downtime [2]. One of the fundamental elements of the wind turbine rotation mechanisms are bearings. These can be divided into pitch bearings, yaw bearings, transmission bearings (main bearing and gearbox bearings), and generator bearings [3]. Bearings are commonly subjected to harsh environments, including vibration and shock under varying wind speeds. This has led to bearings operating beyond their limits, which explains the high failure rate of these components [4]. Recently, data show that main bearing failure rates (over a 20-year life) can be as high as 30% [5]. The industry identified main bearing failure as the second largest reliability challenge after WT gearboxes [6].

In general, bearing failures can be classified into fatigue, wear, corrosion, plastic deformation, electrical erosion, cracking, and fracture failure modes. There are different causes and behaviors for these modes, such as deformation, tension, craters, and fractures [7]. Mechanical failures are often accompanied by increased heat loss. In bearings, failures occur with increasing temperature and indicate reduced performance or imminent failure [8]. Most industrial wind farms operate through supervisory control and data acquisition (SCADA) systems, which record monitoring information from different components of the WT, such as wind speed, generated power, and bearing temperature. Therefore, the temperature signals from the SCADA data can provide information on the performance of the main bearing. The recorded data create an opportunity to process the collected time series data for various applications, such as the diagnosis, detection, or prognosis of WT failures [9].

In the literature, several works address failure detection through the use of SCADA data in WTs in operation. For example, in [10], the prediction of the main bearing failure has been carried out on two wind farms, with a total of 84 WTs, where four WTs have the failure. Based on the principles of Ensemble Learning, the anticipation of failure is obtained in no less than one month. In [11], a study is carried out based on several machine learning techniques to predict bearing failure in the gearbox; a wind farm with 13 WTs is used, where three WTs have bearing failures. The best model obtained is an LSTM network with a false alarm rate of 50% and one month of anticipation of failure. In [12], a hybrid model is made to predict fatigue in the main bearings, where the part of the bearing that suffers fatigue consists of known physical formulations, and the unknown degradation of the grease is represented by deep neural networks. The tests are carried out on 10 WTs, proposing regreasing intervals to extend the useful life of the bearings. The studies mentioned above used and validated their methodology in real wind farms; however, the results do not show failure, if not less than a month after failures occur.

In this work, in contrast to the aforementioned works, it has been possible to detect a main bearing failure four months in advance to the fatal breakdown of the component. The proposed methodology works with a normality model based on the use of SCADA data from wind farms composed of 12 WTs. It should be noted that no additional information provided by the SCADA data is used. Additionally, this methodology has not worked with labeled data due to the difficulty in obtaining data from the failure under study and other aspects associated with possible errors in the process. Therefore, the model can be used in any WT that uses SCADA data, even if this type of fault never occurred in the past. The signals of external sensors provided by WT SCADA are used to capture hidden trends in the main bearing temperature. The LSTM neural network has been chosen for modeling due to its characteristic of capturing long-range dependencies and nonlinear dynamics.

The contribution of this research is as follows: i) propose a methodology for the prediction of main bearing failure based on the results obtained in a wind farm in operation, ii) design and develop an efficient deep learning model to predict long-term time series (the incipient fault has been detected 4 months in advance of the fatal breakdown of the component), iii) carry out a comparative analysis in a wind farm where one WT is affected by the failure under study, another turbine has another type of failure, and the rest are healthy.

The presented document has the following structure: Section 2 carries out an analysis of the SCADA data to be selected, as well as a complete statement of the proposed methodology, while Section 3 presents the results and discussion. The document ends in Section 4 with possible future directions.

2. The proposed methodology

The proposed approach for the early prediction of main bearing failure is based on the creation of a normality model by using an LSTM neural network. The SCADA data of the selected variables under healthy conditions are used to train the model. Then, when the model predicts the target variable from data that is faulty, this will end with greater errors in the predictions. Figure 1 shows the normality model approach.

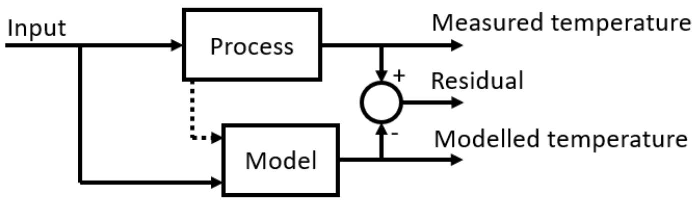

Figure 1. Normality model. When already trained (only with healthy data) the inputs are data that may or may not be healthy; the output shows the real measured SCADA temperature and the estimated (modelled) temperature by the LSTM network. In the middle is the residual (error that can be higher if the data is not healthy).

The methodology used to derive the proposed approach is the following: 1) SCADA data are described and analyzed to select the variables to be used for the creation of the normal behavior model; 2) treatment of outliers and empty data is considered; 3) data partition into train, validation and test datasets is performed; 3) data normalization and reshape is made; 4) the LSTM architecture design for the specific problem is created; 5) the fault indicator for the early prediction of the main bearing failure is developed and described; 6) test on a real SCADA dataset from an operational wind farm is performed.

2.1. SCADA data set

The SCADA data used in this study correspond to a wind farm commissioned in Spain in 2006. The wind farm has 12 operational WTs, where each wind turbine has a nominal power of 1500kW, a diameter of 77 m, IEC IIa class, a swept area of 4657 m^2, and wind speed range for energy production between 3.5 m/s and 25 m/s. WT SCADA data variables can be grouped into environmental, electrical, component temperature, hydraulic, and electrical variables. The maximum, minimum, and standard deviation values are available in 10-minute time periods for the SCADA data. For the study, continuous operational data from the wind farm were collected from February 2017 to November 2018.

2.1.1. Data analysis In the Introduction, it was described the various failure modes of the main bearings, which in many cases lead to a temperature increase. For this reason, the SCADA temperature variables are selected near the main bearing. In particular, the selected SCADA variables are the low-speed shaft temperature, the bearing noncoupling side temperature, the gearbox temperature, and the generator temperature. Finally, also the generated real power and rotor speed are selected, which provide information about the operating regions of the WT, and the ambient temperature due to its influence on all the subsystems of the WT. The selected variables are detailed in Table 1 (note that each variable has a range of operating values). Figure 2 shows the readings of the SCADA variables selected.

Table 1. Selected SCADA variables with their respective description and value ranges.

Variable	Description	Range	Units
Pot	Generated real power	[0, 2000]	kW
TempAmb	Ambient temperature	[−5, 40]	°C
TempCojLOA	Bearing non-coupling side temperature	[0, 120]	°C
TempEjeLento	Low-speed shaft temperature	[0, 120]	°C
TempGen	Generator temperature	[0, 175]	°C
TempRodamMultip	Gearbox temperature	[0, 120]	°C
VelRotor	Rotor speed	[0, 50]	rpm

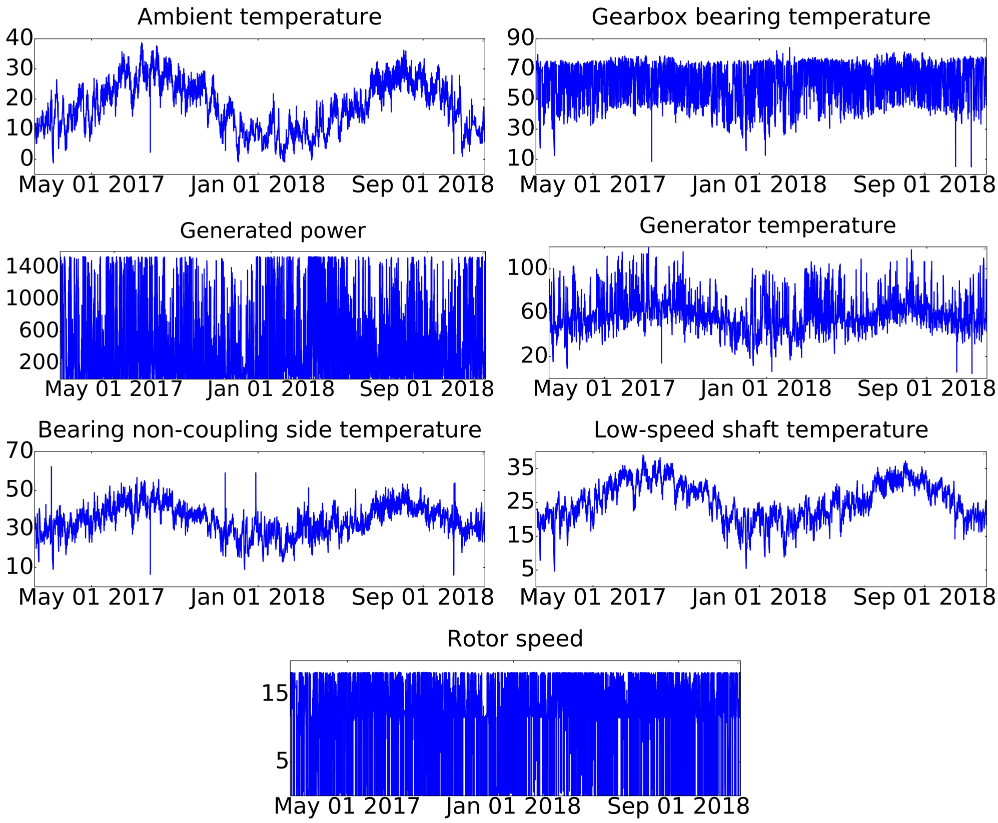

Figure 2. Example of the seven SCADA variables selected to develop the normality model.

2.1.2. Data cleaning Real SCADA data have unavoidable outliers from different WT readings. In this work, data cleaning is carried out in such a way that information is not lost, since it could cause loss of information for the prediction of failures. The strategy used filters the data from different sensors using the ranges defined for the different sensors (see Table 1).

Another problem in SCADA data is missing values due to sensor malfunction; when added to the filtered data range of the aforementioned outliers, missing values increase. The imputation strategy used to solve this problem is the use of a piecewise cubic Hermite interpolation polynomial (pchip). This type of interpolation works with known points and specific slopes at the interpolation points. The curve obtained from pchip preserves the trend of the data, respecting the monotonicity and guaranteeing at least the first derivative.

2.1.3. Train, validation, and test splitting For the development of the proposed early prediction model, the following steps will be followed: (1) divide the data for training, validation, and testing, (2) train the neural network with the training data and validate the model with the validation data, (3) evaluate the model with the test data. This section will focus in the first step; the remaining steps will be explained throughout the document.

The data have been divided so that its time series have not been broken in time (the information has not been mixed); consequently, the training and validation data correspond to the year 2017, while the test data correspond to the year 2018. Finally, the division is detailed in Figure 3; note that this division has been made for the entire wind farm.

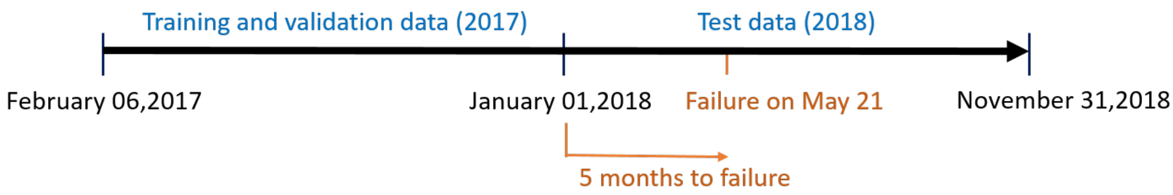

Figure 3. Data distribution for WT2.

Data from 2017 have been divided into 90% for training and 10% for validation, obtaining 42,481 samples for training and 4,609 samples for validation. The data from 2018 is used for testing, obtaining 47,953 samples.

2.1.4. Data normalization The SCADA data of the selected variables have different magnitudes due to the different nature of the measurements. In deep learning and machine learning applications, it is recommended to scale data to facilitate learning [13]. Therefore, the maximum-minimum scaling has been selected, guaranteeing data in the range [0, 1]. It is a simple technique, and its only drawback is dealing with outliers, which it has already resolved by range filtering the data. Normalization of the data has been carried out so that the maximum and minimum values have been taken from the training data and used to normalize the validation and testing data.

2.1.5. Data arrangement Data have been ordered so that a sample has a length of 144 time steps, corresponding to one day, with the intention that the network learns the temporal trend characteristics of one day of operation. The next step is to reshape the data to be ready as input for the LSTM neural network. For this, seven input data are defined, which are shown in Table 1. The table 2 shows the input data array of the LSTM neural network, where k represents the sample number and the instants of time are defined from $t-144$ to $t-1$.

For the output of the neural network, the target is the SCADA variable corresponding to the low speed shaft temperature at the time instant t (at the end of the day).

2.2. LSTM

In recent years, deep learning has proven its powerful hands-on learning ability in conjunction with the superiority of big data processing, attracting attention from various fields. LSTM networks have significant advantages over other artificial intelligence techniques due to its memory capacity over time series [14]. The LSTM neural network is a type of recurrent neural network (RNN) that solves the problem of gradient vanishing while preserving the quality of the RNN for processing sequential data [15]. The LSTM neural network decides the update of the information through the architecture of its cell. The reader is referred to the excellent reference

Table 2. Data matrix where columns represent the seven SCADA variables, rows represent 144 timestamps, and the value of k represents the sample number.

	SCADA variable 1	**SCADA variable 2**	**. . .**	**SCADA variable 7**
	$y^{(k)}_{t-1,1}$	$y^{(k)}_{t-1,2}$		$y^{(k)}_{t-1,7}$
$\mathbf{X}^{(k)} =$	$y^{(k)}_{t-2,1}$	${}^{(k)}_{t-2,2}$	$\cdots$	$y^{(k)}_{t-2,7}$
	$\vdots$	$\vdots$		$\vdots$
	$y^{(k)}_{t-144,1}$	$y^{(k)}_{t-144,2}$		$y^{(k)}_{t-144,7}$

[16] for a comprehensive explanation of the inference formulas for the LSTM network, as well as its fundamentals.

2.2.1. LSTM model A many-to-one LSTM neural network has been chosen due to the approach proposed in this work. For the input of the network, there is a sequence of 144 data and an output corresponding to the temperature of the low-speed shaft temperature at the instant of time t. To choose the best LSTM neural network model, tests were performed with a unidirectional LSTM neural network and a bidirectional LSTM neural network. The results showed similarities, so the LSTM unidirectional neural network was chosen because of its lower computational cost. The unidirectional LSTM network architecture has a hidden layer, and the hyperparameters are as follows.

The optimal number of units in terms of time, computational cost, and results in the hidden layers of the LSTM cells is 100. The batch size is selected to be 64, a learning rate of 0.01 is defined, and 200 epochs are used to train the network. Finally, the mean square error (MSE) has been selected as the loss function of the neural network. Figure 4 shows the architecture of the LSTM network.

2.2.2. Trained model The realization of the model has been done with the PyTorch programming language [17] on a laptop with a NVIDIA GeForce RTXTM2060 video card, a ninth generation Intel i7 processor, 16 GB RAM and a Windows 10 operating system. The training lasted 31 minutes for each wind turbine.

2.3. Fault Indicator (FI)

The early failure indicator is based on the generation of a nominal residual by comparing the measurements of the physical variable of the system y_t (temperature of the low-speed shaft at the time instant t) with their estimate $\hat{y}_t$ (output of the LSTM network). In this work, the residual value is $|y - \hat{y}|$. When considering the uncertainty of the model, the residual generated is not enough to establish an alarm signal through a threshold; there is a risk of generating false alarms.

To solve this problem, a simple moving average (SMA) filter is implemented to smooth the residual shape, removing sudden changes from the residual. The SMA filtering process has been implemented through a moving window of sizes 144 and 1008; that is, one day and one week of data. Fault detection is activated when the SMA of the residual is more significant than a threshold. In particular, the mean μ and the standard deviation σ of the SMA of the training residual are calculated. Finally, the threshold is defined as:

$$threshold = \mu + \kappa\sigma. \tag{1}$$

Note that by adjusting the value of κ it allows for threshold tuning.

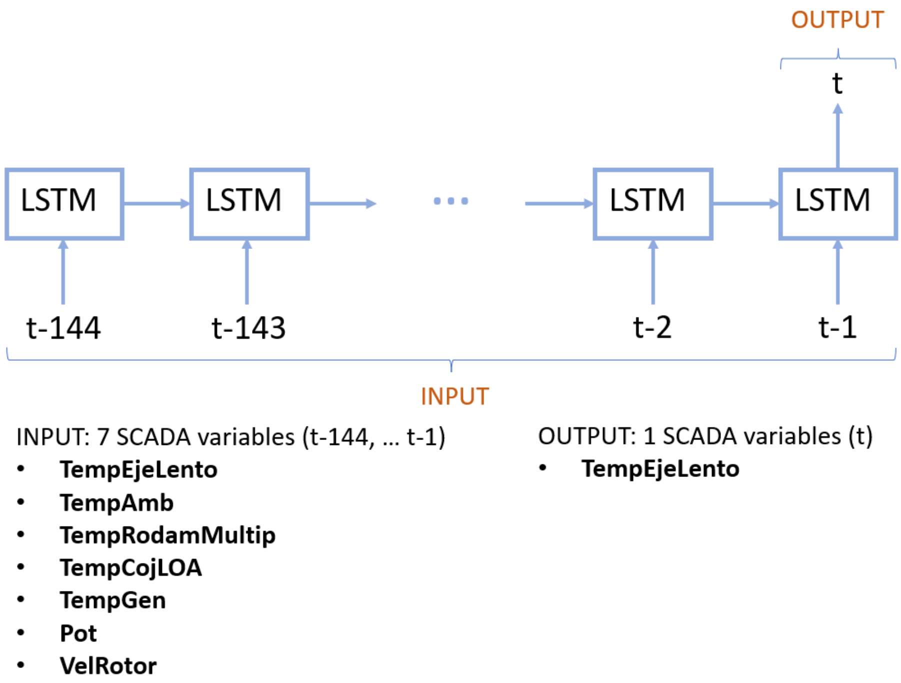

Figure 4. Many-to-one LSTM architecture.

3. Results and Discussion

The presented section describes the results obtained from a proposed methodology based on alarm activations when the FI exceeds a threshold. Table 3 summarizes the alarms activated in a wind farm with different values of SMA and thresholds. Only one wind turbine (WT2) has experienced a failure of interest that occurred on May 21, 2018.

To fine-tune the smoothing of the SMA and the decision threshold, the task of smoothing the FI is performed using only training and validation data to minimize the number of false alarms. Two SMA values are used, one with 144 data (a day) and the other with 1008 data (a week). With SMA values of 144 and κ of 3 ($\mu + 3\sigma$), eleven alarm activations are obtained, but to improve the results, the value of κ is changed to 6 ($\mu + 6\sigma$), which results in six alarms. The results are further improved by increasing the SMA to 1008, and tests are carried out with κ values of 3 and 6, with the latter providing the best results.

Figure 5 shows the best SMA result (weekly smoothing, 1008 data) of the residual curve in training, validation, and testing in the wind farm in detail. For the WT2, an anomaly is detected on February 18, receiving an alarm four months before the total breakdown of the component. The detected anomaly is due to a possible initial failure (initial crack, friction,...), which has been detected months in advance of the fatal breakdown of the component. There is also an alarm activation in WT8, which is related to a different component, i.e., a gearbox replacement carried out on the WT8 from March 22 to April 11. Therefore, it is not completely a false alarm. In fact, it is noteworthy that there are no false alarms in a 2-year period over a whole wind farm, which are excellent results.

In summary, the proposed methodology based on alarm activations when the FI exceeds a threshold shows promising results for detecting anomalies in wind farms. The fine-tuning of the smoothing of the SMA and the decision threshold is critical to reduce the number of false alarms. The results obtained from the methodology include an early detection of a failure of interest in a wind turbine, which may help to prevent significant financial losses due to unexpected downtimes.

Table 3. Experimental summary of SMA and threshold value over the training and validation datasets. The activated alarms are presented with an × mark.

WT	**SMA(144)**		**SMA(1008)**	
	$\mu + 6\sigma$	$\mu + 3\sigma$	$\mu + 6\sigma$	$\mu + 3\sigma$
WT1	-	×	-	×
WT2	×	×	×	×
WT3	×	×	-	-
WT4	-	×	-	×
WT5	-	×	-	-
WT6	-	×	-	×
WT7	×	×	-	×
WT8	×	×	×	×
WT9	-	×	-	×
WT10	×	×	-	×
WT11	-	-	-	-
WT12	×	×	-	-

4. Conclusions

The proposed methodology for the early prediction of main bearing failure using SCADA data from a wind farm yielded several key findings, which are summarized below.

Firstly, the analysis showed that the low-speed shaft temperature and the non-coupling side temperature were found to be indicative of the main bearing failure. This finding is significant as it can help the maintenance team to identify early warning signs of main bearing failure and take corrective actions before it becomes a more significant problem.

Secondly, the Long Short-Term Memory (LSTM) architecture design used in the proposed methodology proved to be effective, and the proposed fault indicator demonstrated a high level of accuracy in detecting the main bearing failure. LSTM is a type of recurrent neural network that can learn long-term dependencies in time-series data, making it particularly well-suited for analyzing SCADA data from wind turbines. The high accuracy of the fault indicator in detecting main bearing failure is significant, as it can help the maintenance team to focus their attention on the affected turbines and take corrective actions proactively.

Thirdly, the comparative analysis conducted on a wind farm with one WT affected by the failure, another turbine with a different type of failure, and the rest being healthy showed that the methodology was able to identify the affected WT accurately and early. This finding is crucial as it can lead to reduced downtime and maintenance costs, as corrective actions can be taken before the failure becomes more severe.

Fourthly, the results demonstrate that the stated approach is effective in detecting a main bearing fault that resulted in a significant increase in temperature. Although only one failure was available in the investigated wind park data, which is insufficient for statistical analysis, any bearing fault leading to heat release might be detectable by the proposed strategy. However, to more extensively investigate the performance of the model, it is necessary to apply the model to other wind parks with main bearing failure issues. Therefore, future work will test the model on a larger dataset to assess its performance in different scenarios and draw more generalizable conclusions

Finally, note that having few false alarms is crucial in any alarm system, especially in the context of wind turbines, as it can significantly affect the performance and productivity of the maintenance team. False alarms can lead to alarm fatigue, a condition where the maintenance team becomes desensitized to alarms due to their frequent occurrence. Alarm fatigue can lead to a decrease in the maintenance team's response time to real alarms, potentially resulting in an increase in downtimes, which can lead to significant financial losses for the wind farm. The proposed methodology, which is based on alarm activations when the FI exceeds a threshold, aims to minimize the number of false alarms by fine-tuning the smoothing of the SMA and the decision threshold. As described in this section, the methodology was tested on a wind farm with 12 wind turbines over a period of two years, and no real false alarms were detected during this period. The absence of false alarms in the proposed methodology is a significant advantage, as it ensures that the maintenance team is not bombarded with unnecessary alarms, which can lead to alarm fatigue. This enables the maintenance team to focus their attention on real alarms and respond to them promptly, minimizing downtimes and maximizing the productivity of the wind farm.

In summary, the proposed methodology for the early prediction of main bearing failure using SCADA data from a wind farm demonstrated several significant findings, including the identification of key indicators of main bearing failure, the effectiveness of the LSTM architecture design, and the ability to accurately identify affected turbines early. These findings are critical as they can help to improve the productivity and performance of wind turbines, reduce maintenance costs, and prevent downtime.

5. Acknowledgments

This work is partially funded by the Spanish Agencia Estatal de Investigación (AEI) - Ministerio de Economía, Industria y Competitividad (MINECO), and the Fondo Europeo de Desarrollo Regional (FEDER) through the research project PID2021-122132OB-C21; and by the Generalitat de Catalunya through the research project 2021 SGR 01044.

6. References

[1] IRENA I 2019 *Abu Dhabii*
[2] Crabtree C J, Zappalá D and Hogg S I 2015 *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy* **229** 727–746
[3] Liu Z and Zhang L 2020 *Measurement* **149** 107002
[4] Saidi L, Ali J B, Bechhoefer E and Benbouzid M 2017 *Applied Acoustics* **120** 1–8
[5] Hart E, Turnbull A, Feuchtwang J, McMillan D, Golysheva E and Elliott R 2019 *Wind Energy* **22** 1534–1547
[6] Keller J, Sheng S, Cotrell J and Greco A 2016 Wind turbine drivetrain reliability collaborative workshop: a recap Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States)
[7] Encalada-Dávila Á, Puruncajas B, Tutivén C and Vidal Y 2021 *Sensors* **21** 2228
[8] Feng Y, Qiu Y, Crabtree C J, Long H and Tavner P J 2013 *Wind Energy* **16** 728–740
[9] Marti-Puig P, Blanco-M A, Cárdenas J J, Cusidó J and Solé-Casals J 2019 *Energies* **12** 453
[10] Beretta M, Julian A, Sepulveda J, Cusidó J and Porro O 2021 *Sensors* **21** 1512
[11] Desai A, Guo Y, Sheng S, Phillips C and Williams L 2020 Prognosis of wind turbine gearbox bearing failures using scada and modeled data *Annual Conference of the PHM Society* vol 12 pp 10–10
[12] Yucesan Y A and Viana F A 2021 *Computers in Industry* **125** 103386
[13] Bhanja S and Das A 2018 *arXiv preprint arXiv:1812.05519*
[14] Choe D E, Kim H C and Kim M H 2021 *Renewable Energy* **174** 218–235
[15] Chung D W, Ko J H and Yoon K Y 2022 *Journal of Electrical Engineering & Technology* 1–15
[16] Sherstinsky A 2020 *Physica D: Nonlinear Phenomena* **404** 132306
[17] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L *et al.* 2019 *Advances in neural information processing systems* **32**

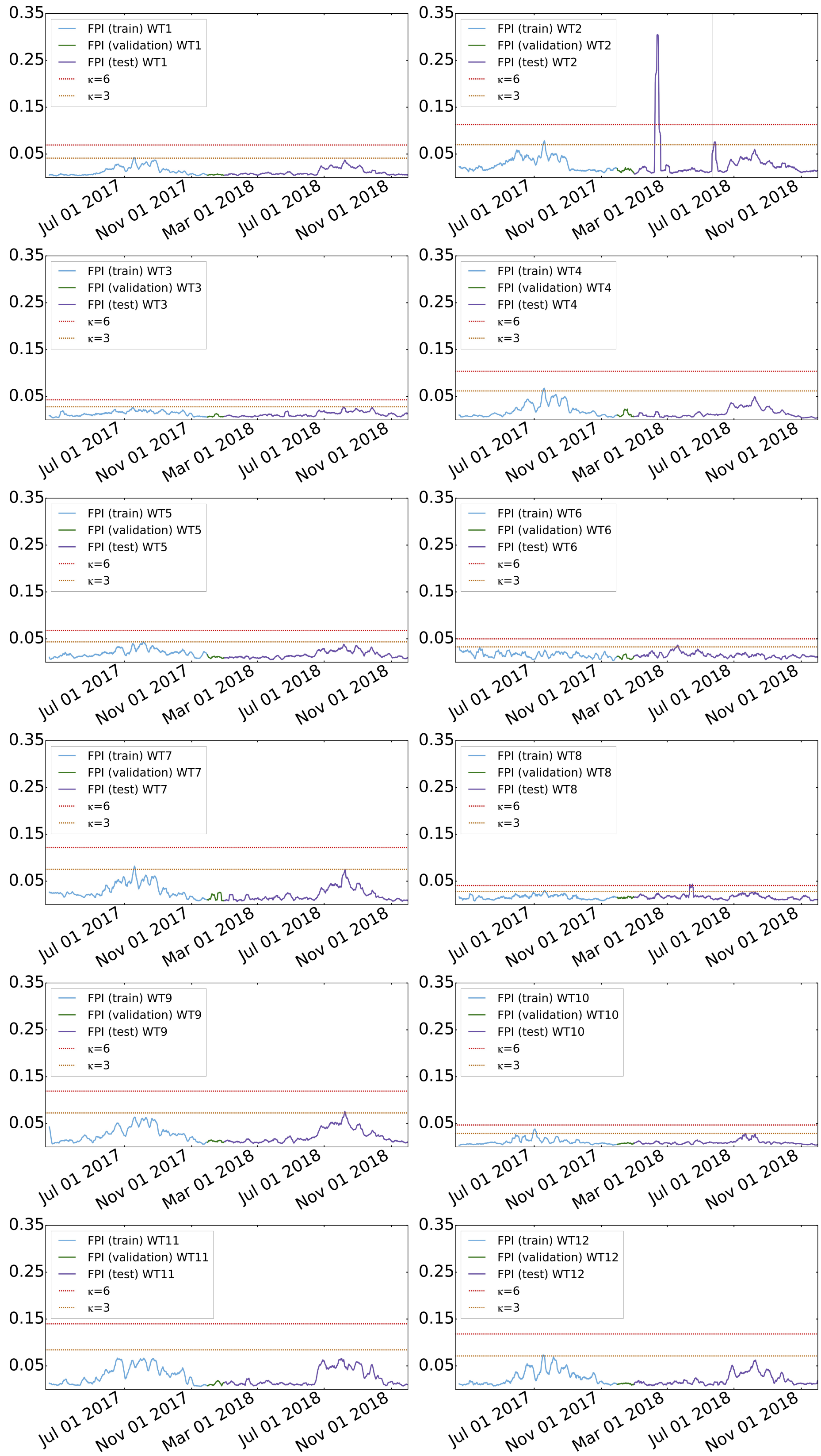

Figure 5. Results obtained over training, validation, and test datasets. The thresholds with values of κ equal to 6 and 3 are shown for each WT.