Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Wind Turbine Main Bearing Failure Prediction using a Hybrid Neural Network

To cite this article: Karen Bermúdez *et al* 2022 *J. Phys.: Conf. Ser.* **2265** 032090

Journal of Physics: Conference Series **2265** (2022) 032090 doi:10.1088/1742-6596/2265/3/032090

Wind Turbine Main Bearing Failure Prediction using a Hybrid Neural Network

Karen Bermúdez[1], Eduardo Ortiz-Holguin[1], Christian Tutivén[2], Yolanda Vidal [3,4] and Carlos Benalcázar-Parra[5]

[1]ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, Faculty of Electrical and Computer Engineering (FIEC), Computer Engineering, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
[2]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
[3]Control, Modeling, Identification and Applications (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
[4]Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Pau Gargallo 14, 08028 Barcelona, Spain
[5]Universidad ECOTEC, Km. 13.5 Vía a Samborondón - Guayaquil, Ecuador

E-mail: `leortiz@espol.edu.ec, kbermude@espol.edu.ec, cjtutive@espol.edu.ec, yolanda.vidal@upc.edu, carlosbenalcazarparra@gmail.com`

Abstract.
Energy is necessary for economic growth and improved well-being, but it poses a great challenge to be generated without increasing costs and avoiding pollution. A viable option is wind energy because it is a clean and renewable. However, continuous monitoring and maintenance of wind turbines is required for the further development of wind farms. Main bearing failures were identified by the European Academy of Wind Energy as a critical issue in terms of increasing the availability and reliability of wind turbines. In this work, it is proposed a hybrid neural network for main bearing failure prognosis. This network consists of a two-dimensional convolutional neural network (to extract spatial-temporal characteristics from the data) sequentially connected with a long short-term memory network (to learn sequence patterns) to predict the slow-speed shaft temperature (the closest temperature to the main bearing). The mean square error between its real measurement and its prediction gives a failure indicator. When it is greater than a defined threshold, then an alarm is triggered and gives the maintenance staff time to check the component. The advantage of this strategy is that it does not need faulty data to be trained, since it is based on a normality model, that is, it is trained with a single class of data (healthy) and does not require incurring high costs per acquisition of new sensors since SCADA data is used (which comes in all industrial size turbine). The results show that the use of a hybrid network can identify failures around four months before a fatal failure occurs.

1. Introduction

Energy is a fundamental piece in the development of societies. However, due to climate change and global warming that has been evident in recent years, the interest of governments in generating clean energy has grown. An alternative to the common use of fossil-fuel energy sources to generate energy is the use of natural resources, which are clean and exists in a wide geographical area around the world. Wind energy is one of the natural alternatives that has had great interest and growth in recent years. In 2020, the wind industry showed a year-on-year growth of 53%, making it the best year in history for this sector. Installing more than 93 GW wind power [1]. This growth goes hand in hand with the continued increase in the turbine's size, which now comes with average rotor diameters greater than 150 meters and turbine capacity greater than 7.5 MW. This increase in the size of the turbines also presents challenges that these projects must face. One of the main challenges is the cost of components maintenance, which can make projects more expensive.

The European Academy of Wind Energy (EAWE) identified main bearing failures as a critical problem in terms of increasing the availability and reliability of WTs [2]. Therefore, it is important to have an effective maintenance plan for this component. The ways to perform maintenance are preventive, corrective and, predictive. The first two methods generate high maintenance costs in wind power plants. This factor has accelerated the interest in research of better condition monitoring (CM) systems. It is the crux of the matter when moving from time-based preventive maintenance, which remains the current core practice for WTs, to predictive maintenance, as it relies on the actual condition of the equipment rather than the average or expected life statistics. Most of the CM systems use vibration analysis, acoustic emission sensors or oil analysis during the operation of turbine components to determine the probability of a future failure [3]. These techniques, however, are usually very expensive, due to the price of additional sensors or other mechanical components that need to be installed in the turbines. One, alternative is to use data obtained from the supervisory control and data acquisition (SCADA) system that comes with every industrial-size WT. In recent years there has been a growing interest in using these data not only for the proper control of the turbines but also for CM since it avoids the increase in costs by not having to buy additional sensors. For example, in [4] is proposed a method that only requires healthy WT SCADA data to be collected to train an artificial neural network with a Bayesian's regularization to predict main bearing failures. In [5], an extreme learning machine strategy is used to monitor and evaluate the health status of WTs. A new ensemble approach based on Mahalanobis distance is proposed in [6] to detect anomalies in WT generators and then diagnose their failure modes.

There are certain limitations in forecasting failures using real SCADA data just by using the linear time series model or the neural network model. Therefore, in this work, the combination of several method advantages and several best algorithms is used. This method based on convolutional neural network and a long short-term memory (CNN-LSTM) is able to predict the WT slow-speed shaft temperature in advance, thus giving maintenance personnel time to coordinate a maintenance without incurring high costs. This new technique combines the advantages of convolutional neural networks that can extract effective spatial characteristics from the data and long short-term memory that can not only find the interdependence of the data in time series, but also automatically detect the best mode. Suitable for relevant data, this method can effectively improve the accuracy of the main bearing failure prognosis. The hybrid network is trained using only normal (healthy) SCADA data and then when inference is made with future data, abnormal changes in slow-speed shaft temperature can be detected, which will be an alarm indication.

The remainder of this paper is organized as follows. A brief description of the used WT and of the SCADA data are provided in Section 2. In Section 3 the proposed methodology and the hybrid network model are explained. The obtained results are given and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Wind turbine and data descriptions

In this work, two WTs (one healthy and one with main bearing failure) located in Spain are analized. Some specifications are that the WTs can generate a power of 1500 kW and have a rotor diameter of 77 m. (class IEC IIa). Figure 1 shows the major components of this WT model. The WT power production begins at wind speed of 3.5 m/s and can go to 25 m/s where occurs an automatic stop [7]. The optimal performance is reached at a low wind speed of 11.1 m/s. The main WT model technical specifications are described in Table 1.

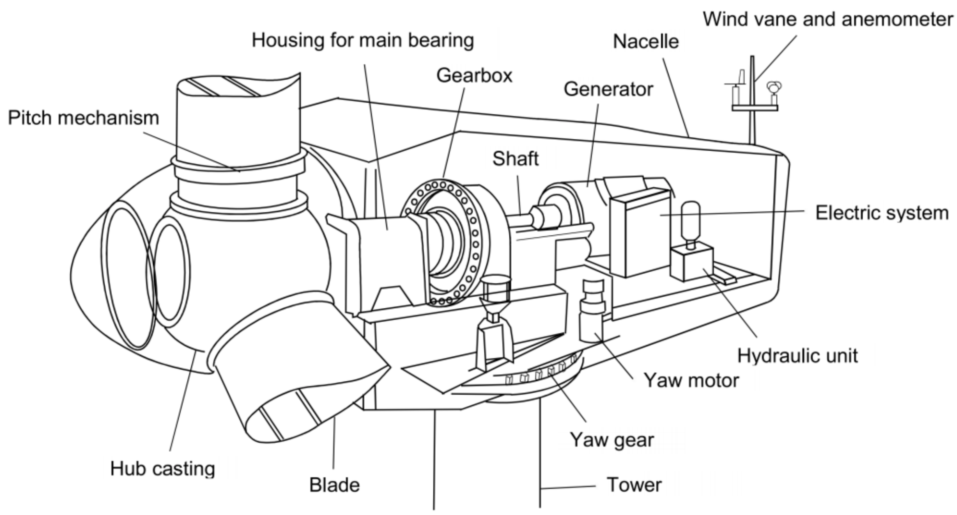

Figure 1. Main components of the (WT) [8].

Table 1. Technical specifications of the WT.

Number of blades	3
Nominal power	1500 kW
Rotor diameter	77 m
Wind class	IEC IIa
Swept area	4657 m2
Nominal rotation speed	18.3 rpm
Cult-in wind speed	3.5 m/s
Cult-out wind speed	25 m/s
Bearings	Double spherical roller bearings
Power regulation	Independent pitch (variable speed)

These WTe uses a double spherical main roller bearing, which compensates misalignment by allowing low to medium speeds and large radial loads. The component of interest, in this work, is the main bearing which can be affected by different types of failures as indicated in the Table. 2. When bearing failure initiates (e.g., initial crack), it is usually accompanied by a momentary release of frictional heat, but then the bearing temperature goes back to normal (crack is stabilized and not growing). The importance of this methodology is to detect this heat release months before the bearing is completely damaged. For more detailed information about the mentioned failure modes, see [4] and, [9].

Table 2. Bearing Failure modes (SKF classification adapted from ISO 15243:2004) [9].

Failure mode	Sub-failure
Fatigue	Subsurface-initiated, Surface-initiated
Wear	Abrasive wear, Adhesive wear
Corrosion	Moisture corrosion, Frictional corrosion
Electrical erosion	Excessive current erosion, Current leakage erosion
Plastic deformation	Overload deformation, Indentations from debris
Fracture and cracking	Forced fracture, Fatigue fracture, Thermal cracking

The SCADA data is obtained from February 06, 2017, to November 30, 2018. The SCADA system is measuring different sensor measurements and, for each one, the mean, maximum, minimum, and standard derivation of every 10 minutes samples are collected. In addition to the SCADA data, a file that contains the work orders with failures and maintenance information is available. For example, maintenance dates, maintenance duration and repair action information. From this information, it is found that there exists a main bearing fault on 21 May 2018. Thus, this information is used to split the dataset. In particular, to test if the proposed methodology can predict the failure months in advance.

3. Methodology

In this section, the proposed methodology is described. It is composed by the following steps: variable selection, data cleaning, data split, data normalization, a feature reshape stage, the definition of the model to be trained and, finally, the definition of a failure prognosis threshold to alert when a main bearing pre-failure exists. Each of these stages are detailed below.

3.1. SCADA sensors measurements selection

The original dataset is composed of many sensor measurements, but when it is desired to study a specific failure, data analysts must be skilled in choosing the most important measures for the component to be studied [10]. One of the challenges of this work is to select the variables of the SCADA data to be used. This is because if variables closely related to other components of the turbine (and not only to the component to be studied) are used, this methodology could not only detect the fault of interest but also faults in other components that have a high coupling with the variables used. For example, if the blade position signal is used as one of the input variables, the model could detect faults related to pitch. This problem limits the identification of failures in a specific component. For that reason, in this work is proposed to use the environmental temperature and wind speed, which are external sensors, and only three WT component measurements: slow-speed shaft temperature (feature of interest), active power, and the gearbox bearing temperature, see Table 3.

3.2. Data cleaning

Before training a model, it is important to clean the data to determine and correct any missing or outlier values. In this stage, the aim is to treat the data so that the dataset has coherence and thus help the model to have better predictions. There are different ways of detecting outliers. In this study, a method based on the realistic values range filter is used (see [4]). That is, any

Table 3. Selected featured for normality model.

Feature	Description	Range	Units	Location
Pot	Generated power	[0,2000]	kW	Internal
TempAmb	Environmental temperature	[-5,40]	°C	External
VelViento	Wind speed	[0,60]	rpm	External
TempRodamMultip	Gearbox bearing temperature	[0,100]	°C	Internal
TempEjeLento	Slow-speed shaft temperature	[0,120]	°C	Internal

value outside a real range of behavior of the variables is considered an outlier. Table 3 shows the defined sensor ranges. Any measure outside these ranges is considered an outlier. These outliers are not eliminated directly, they were labeled as missing values to be processed later.

There are several options to deal with missing values, which include: replace them with the mode, median, average, and other techniques. However, they can introduce a bias in the mean and standard deviation [11]. In this work, the Piecewise cubic Hermite interpolation polynomial is employed to replace missing values with new values that are consistent with the dataset [12]. This technique preserves the shape of the data, respects monotonicity, and ensures that at least the first derivative is continuous. To fill in the missing values at the beginning of the time series, the first value found is used [13]. To fill in the values at the end of the time series, the last value found is used. Figure 2 shows a part of the original slow-speed shaft temperature feature (with missing values) time series, and the feature with data imputation.

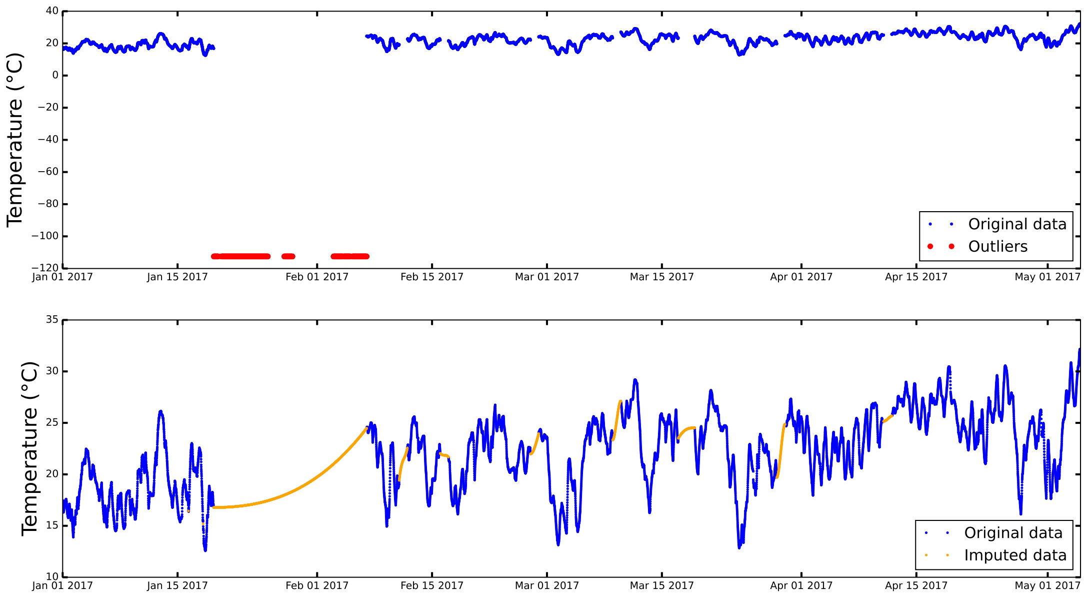

Figure 2. Time series of slow-speed shaft temperature feature with outliers and interpolated data.

3.3. Data split: Train, validation and test set

For the data split, the work orders' information is used. The document indicates that the WT with a failure has a repair on the main bearing on May 21, 2018. It is important to emphasize that this work is based on a normality model, which means that the model is trained only with one-class data (healthy). One of the normality model advantages is that there is no data unbalance (more data of one class), what usually happens in classification models where it is difficult or almost impossible to find data with failures in real applications. For example, for a new WT there is no data with failures. With this proposed methodology, there is no need to have data with failures. Data from February 06, 2017, to September 24, 2017, is used as training data. In the work orders, no fault related to main bearing is found between these dates. The validation data goes from September 24, 2017, to January 1, 2018. Finally, as the main challenge is to predict the main bearing failure several months in advance, the test dataset is defined from January 1, 2018, to November 30, 2018. The figure 3 details the split into train, validation and, test datasets.

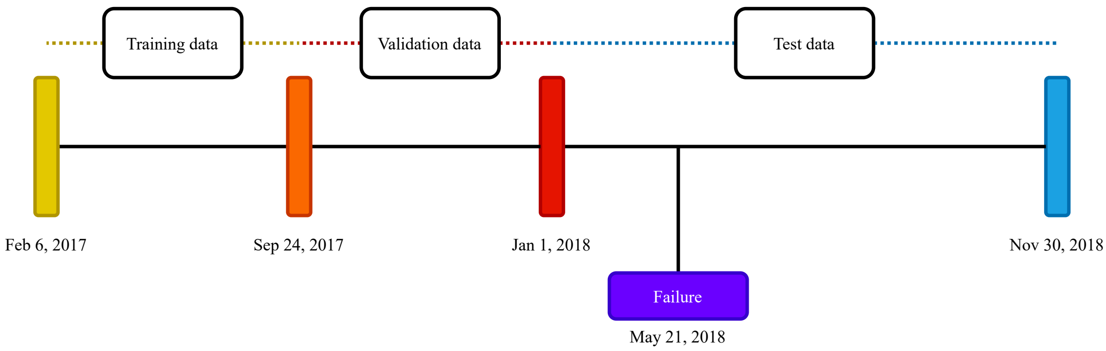

Figure 3. Data split: train, validation and test set.

Note that for the healthy turbine, the split of the data is carried out taking the same dates as the faulty turbine, since as there are no work orders that record failures, any range of data can be used to train your model. This allows us to show and compare the results of using this methodology in a healthy turbine and in a faulty turbine.

3.4. Data Normalization

Due to the fact that values of selected features have different magnitude ranges, it is important to normalize the data. For example, as can be seen in Table 3 the generated power has values from 0 to 2000 and the wind speed has values from 0 to 60. The goal of normalization is to change the values of numeric columns in the dataset to a common scale. Normalization helps the output model to not be biased towards the large-scale inputs [14]. Here, the min-max normalization is calculated by:

$$y = \frac{x - min}{max - min}, \tag{1}$$

where y is the normalized dataset, min and max are the training minimum and maximum values and, x is the original set of input values. So, the entire range of values of x are mapped to a new range from 0 to 1.

3.5. Image creation

As in this work, the first part of the model is composed by a two-dimensional CNN that needs a matrix as input, not a vector, feature reshape is developed. This consists of converting the time-series information into a matrix structure that is accepted by the CNN. The transformation is carried out individually for each of the selected features used as inputs. The matrix size used is 12×12 (144 values) with 4 channels, and it is constructed as follows. Every 144 samples (this corresponds to one day of data) are transformed to one matrix of size 12×12 with 4 channels (one channel per selected feature), similarly to RGB images (with 3 channels). Figure 4 shows the result of this approach.

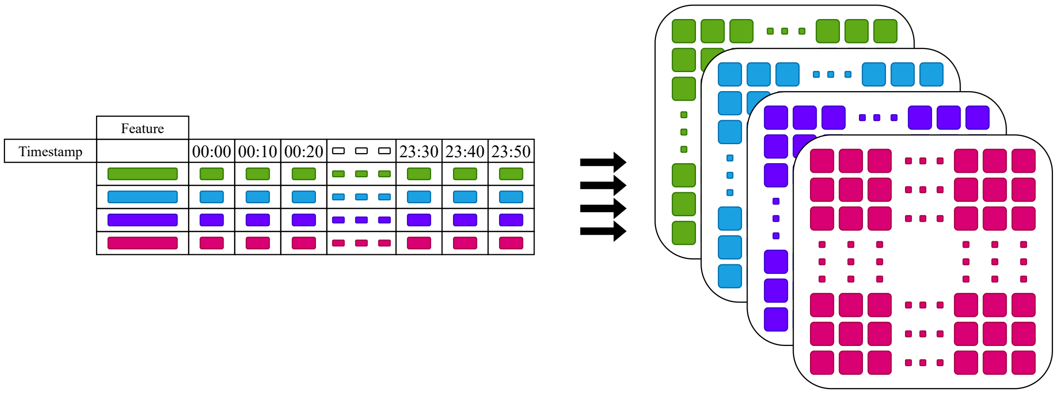

Figure 4. Image creation process.

3.6. Hybrid neural network (CNN + LSTM)

A hybrid model structure is proposed in this section, and it is based on the five selected features shown in Table 3. The model inputs are the generated power, environmental temperature, wind speed, and gearbox bearing temperature. The output model is the slow-speed shaft temperature (the variable most related to the studied fault).

The proposed model, as shown in Figure 6, consists of a CNN and a LSTM networks. A two-dimensional CNN is used to extract spatial information for every one day matrix. Spatial information refers to data having location-based relation with other data [15]. Remember that in the previous Section 3.5 images composed of 144 measurements corresponding to one day are constructed, maintaining their temporal order. This means that each cell or value in this matrix has a temporal relation with the value that is in the cell to the left (value ten minutes before) and right (value ten minutes after). In CNNs, when the kernel analyzes every value in the matrix, its purpose is to find the spatial relationship (in this case temporal) between the different values of the matrix. At the end of the convolutional layer spatial-temporal features are obtained. Since the output of the CNN is fed to an LSTM network for analysis of spatio-temporal features, it is necessary to reshape (flatten) them into a sequential structure (vector). This vector is the input to the LSTM network, which has significant advantages over other machine learning techniques due to its memory capacity over time series proven useful in learning sequences that contain long-term patterns [16]. The architecture of the LSTM neural network's cell determines when the information is updated. Figure 5 shows the basic component (LSTM cell) of a LSTM network; note the information flow control gates within the cell referenced as input gate (i_t), output gate (O_t), and forget gate (f_t). In the LSTM cell architecture, h_t represents the hidden input states of the current time, Y_t represents the hidden states of exit at the current time step and h_{t-1} is the hidden states of the previous time step. The LSTM cell functions are σ (sigmoid function) and tanh (hyperbolic tangent function). C_t is a memory cell used for information preservation, and the flow of information to or from C_t is regulated by three gates.

The architecture of the proposed hybrid neural network is described in Figure 6 and the description of each layer is described in Figure 7.

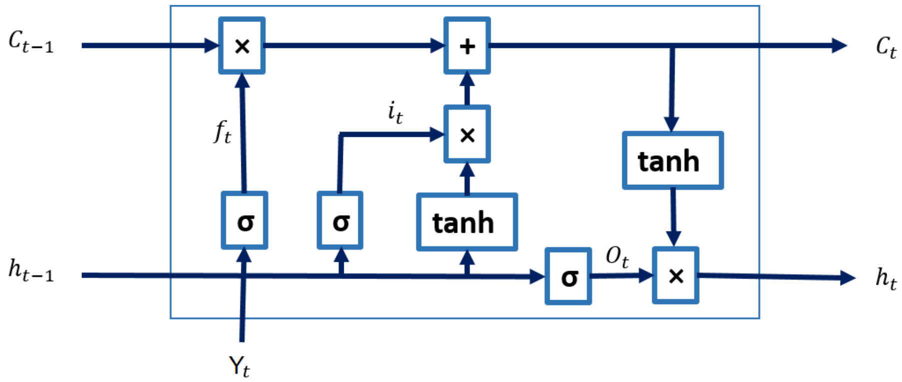

Figure 5. LSTM cell unit.

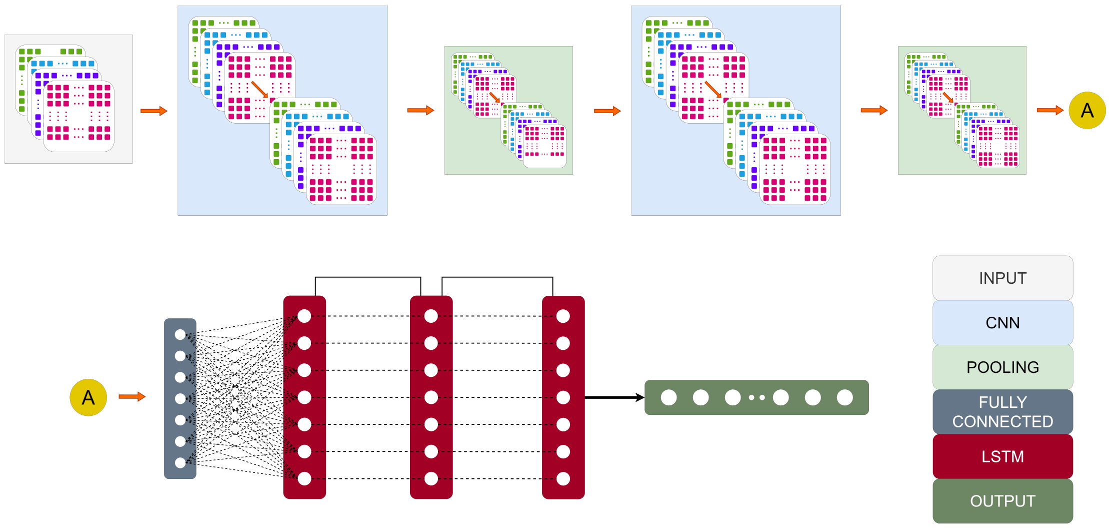

Figure 6. Hybrid neural network proposed.

3.7. Fault prognosis threshold

In this work, a threshold is defined to indicate that a prefailure exists. When the mean square error (MSE) between the real slow-speed shaft temperature and its model prediction is higher than the defined threshold, an alarm is triggered. However, before defining this threshold, it is necessary to group the MSE results by weeks and then apply an exponential weighted moving average (EWMA) to them to reduce the number of false positive alarms. Finally, to define the threshold, first the training dataset is passed through the model. The mean (μ) and the standard deviation (σ) of the obtained output are then calculated. Finally, the threshold is defined as:

$$\text{threshold} = \mu + 4\sigma. \tag{2}$$

The methodology stages are described in Figure 8.

4. Results

This section details and analyzes the results of the proposed methodology. The figure 9 shows that for the faulty WT from February 4 to February 11, 2018, the processed MSE signal exceeds the defined threshold, which represents a constant alert signal for maintenance personnel and gives them time to prepare for a preventive maintenance. Therefore, the system guarantees the

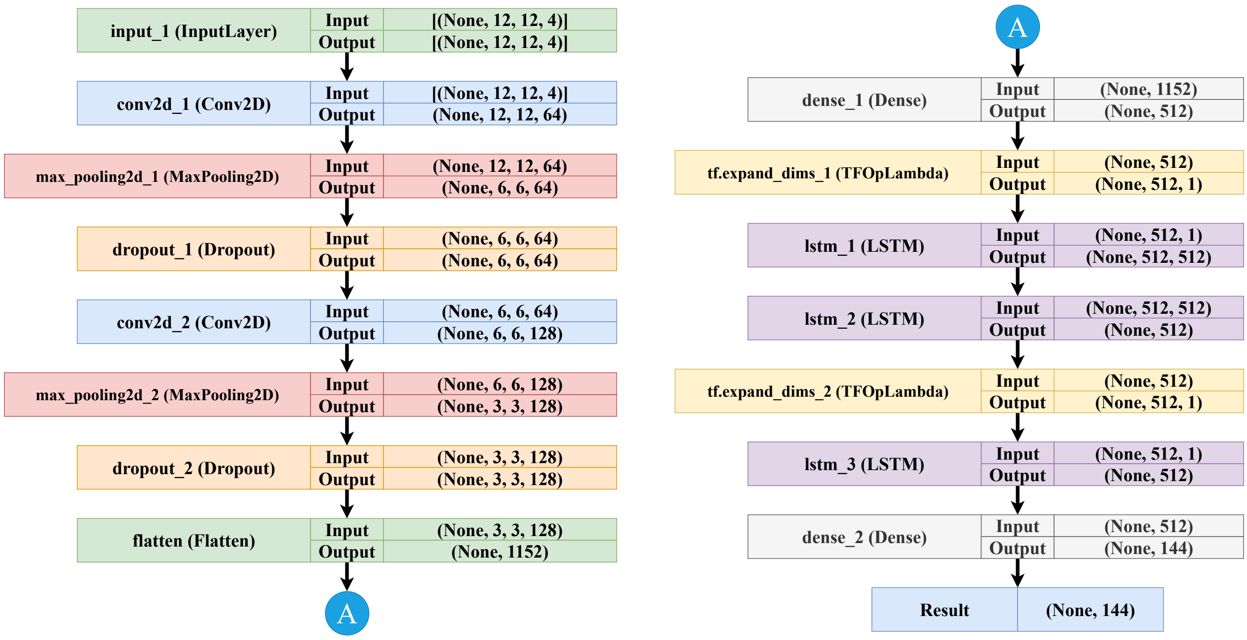

Figure 7. Hybrid neural network architecture.

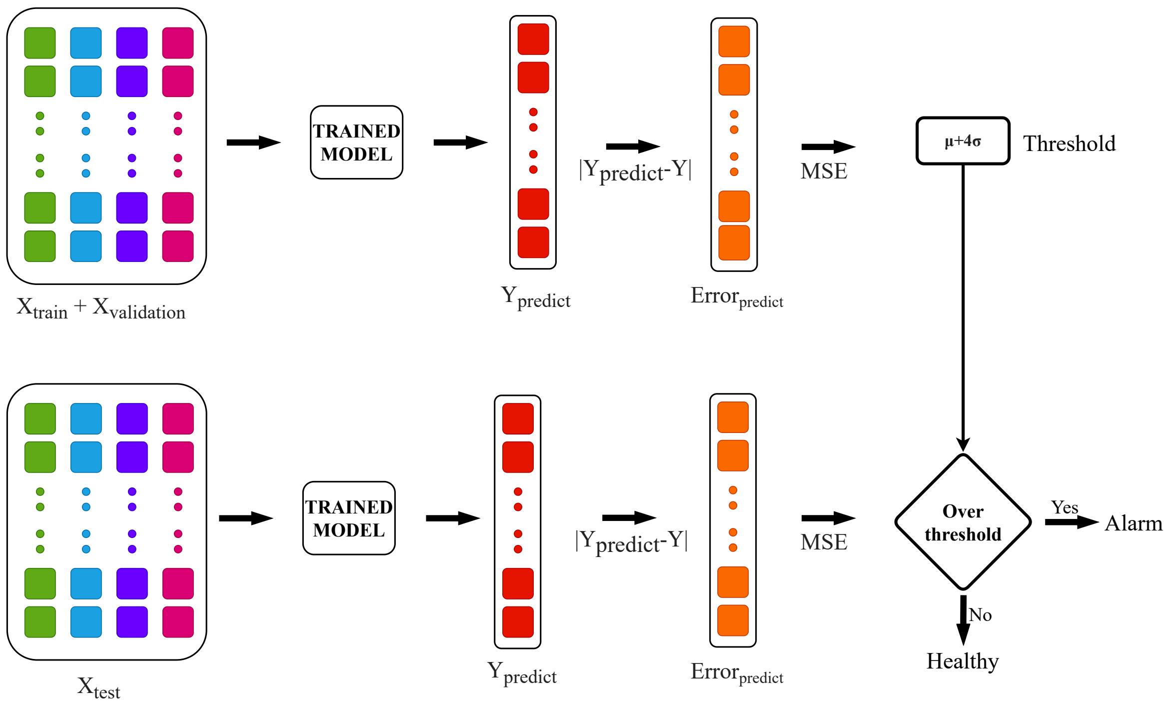

Figure 8. Normality model.

failure alarm months before it occurs. As previously stated, when a bearing failure initiates, frictional heat is typically released briefly and then the bearing temperature stabilizes. This is why the error falls below the threshold before maintenance In addition, it is observed that once the maintenance has been carried out, the processed MSE signal no longer exceeds the threshold.

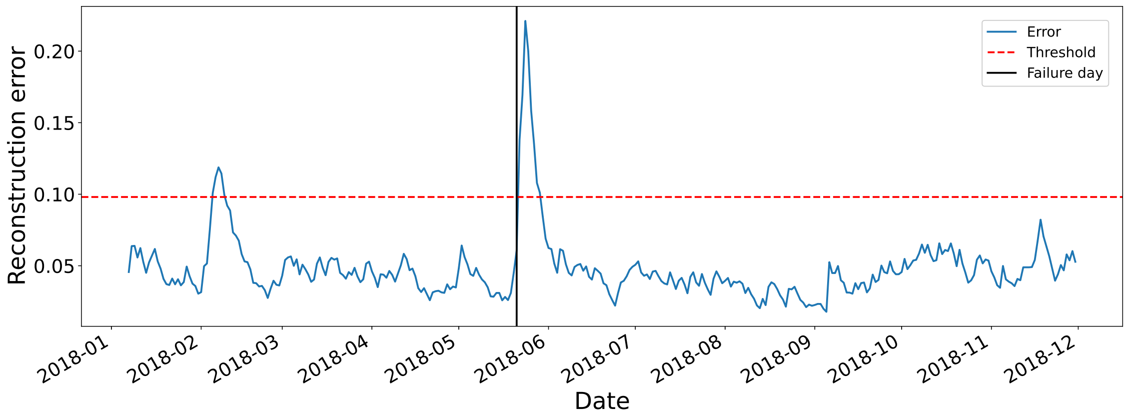

Figure 9. Reconstruction error on the test data for the wind turbine that suffered the fault of interest.

On the other hand, Figure 10 shows the result of using the same methodology in a turbine that does not have the studied failure. As can be seen, the error signal never exceeds the defined threshold.

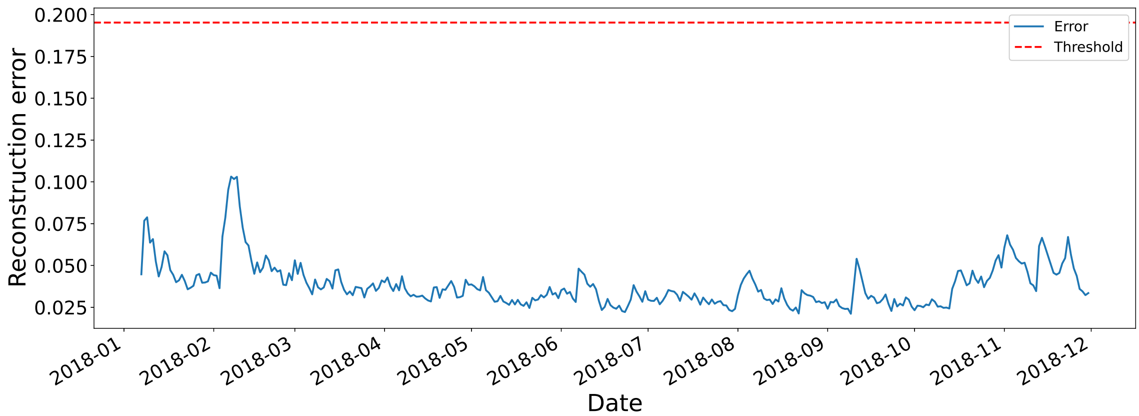

Figure 10. Reconstruction error on the test data for a healthy wind turbine.

5. Conclusions

In this work, a hybrid neural network model is proposed and trained to predict months in advance when a main bearing failure is about to occur. The results demonstrate that the strategy can predict the main bearing failure around four months before it occurs, and thus let turbine operators plan a preventive maintenance. It must be taken into account that this model is only used for the turbine under study. In case it is desired to deploy main bearing failure prognosis in other turbines, it is recommended that each one should have its own model. But the advantage of this methodology is that it uses solely SCADA data and requires only healthy data to be deployed. Therefore, not failure data (which are very difficult or almost impossible to obtain in real applications) is needed to train the model. Therefore, the strategy validated on this turbine is applicable to any other turbine.

Acknowledgments
This work has been 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 DPI2017-82930-C2-1-R; and by the Generalitat de Catalunya through the research project 2017 SGR 388. The authors thank a lot to Smartive company, as this work would not have been possible without their support in the ceding of wind farm data.

References

[1] G. W. E. Council, "Gwec— global wind report 2021," 2021.
[2] E. Hart, B. Clarke, G. Nicholas, A. Kazemi Amiri, J. Stirling, J. Carroll, R. Dwyer-Joyce, A. McDonald, and H. Long, "A review of wind turbine main bearings: design, operation, modelling, damage mechanisms and fault detection," *Wind Energy Science*, vol. 5, no. 1, pp. 105–124, 2020.
[3] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renewable energy*, vol. 133, pp. 620–635, 2019.
[4] Á. Encalada-Dávila, B. Puruncajas, C. Tutivén, and Y. Vidal, "Wind turbine main bearing fault prognosis based solely on scada data," *Sensors*, vol. 21, no. 6, p. 2228, 2021.
[5] P. Marti-Puig, M. Serra-Serra, J. Solé-Casals *et al.*, "Wind turbine prognosis models based on scada data and extreme learning machines," *Applied Sciences*, vol. 11, no. 2, p. 590, 2021.
[6] X. Jin, Z. Xu, and W. Qiao, "Condition monitoring of wind turbine generators using scada data analysis," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 202–210, 2020.
[7] E. Radzka, K. Rymuza, and A. Michalak, "Wind power as a renewable energy source," *Journal of Ecological Engineering*, vol. 20, no. 3, 2019.
[8] Z. Jiang, W. Hu, W. Dong, Z. Gao, and Z. Ren, "Structural reliability analysis of wind turbines: A review," *Energies*, vol. 10, no. 12, p. 2099, 2017.
[9] "Bearing damage and failure analysis," https://www.skf.com/binaries/pub12/Images/0901d1968064c148-Bearing-failures—14219_2-EN_tcm_12-297619.pdf, 2017, accessed: 2021-07-08.
[10] P. Marti-Puig, A. Blanco-M, J. J. Cárdenas, J. Cusidó, and J. Solé-Casals, "Effects of the pre-processing algorithms in fault diagnosis of wind turbines," *Environmental modelling & software*, vol. 110, pp. 119–128, 2018.
[11] Z. Zhang, "Missing data imputation: focusing on single imputation," *Annals of translational medicine*, vol. 4, no. 1, 2016.
[12] M.-J. Lai, "Scattered data interpolation and approximation using bivariate c1 piecewise cubic polynomials," *Computer Aided Geometric Design*, vol. 13, no. 1, pp. 81–88, 1996.
[13] S. Duval and R. Tweedie, "Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis," *Biometrics*, vol. 56, no. 2, pp. 455–463, 2000.
[14] M. Kang and J. Tian, "Machine learning: Data pre-processing," *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*, pp. 111–130, 2018.
[15] F. Xue, W. Zhang, F. Xue, D. Li, S. Xie, and J. Fleischer, "A novel intelligent fault diagnosis method of rolling bearing based on two-stream feature fusion convolutional neural network," *Measurement*, vol. 176, p. 109226, 2021.
[16] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*, vol. 89, 2015, pp. 89–94.