Politecnico di Torino

Master's Degree in Aerospace Engineering

Politecnico
di Torino

Utrecht
University

Master's Degree Thesis

HEALTH MONITORING FOR WIND TURBINES DATASETS PROCESSING AND DEVELOPMENT OF RUL PROGNOSTICS

Supervisors

Prof. Matteo Davide Lorenzo Dalla Vedova
Prof. Mihaela Mitici
Prof. Gaetano Quattrocchi

Candidate

Davide Manna

Academic Year 2022-2023

Contents

1 Introduction

1.1 Why wind energy?

The big environmental crisis, that has been affecting the World, has highlighted the need of looking around for new sustainable energy sources. Greenhouse gas emissions have had an important role for this crisis [1] and for this reason the society is trying to move towards renewable energy and in particular to wind energy [2]. According to the European Green Deal, the main target of the Global Wind Energy Council is:

- to provide the 20% of the electricity demand around the world for the 2030 thanks to wind energy systems
- to provide a completely decarbonised electricity supply where wind energy will be the leading renewable source before the 2050

The focus on the wind energy is related to its advantages. As reported in [3] and [4], the wind energy benefits are :

- **It is an inexhaustible renewable resource**.
  Wind energy is readily available
- **The provision for a clean and sustainable source of energy**.
  Wind energy systems allow to deliver electricity without producing $C0_2$ and particulates and there is only a small amount of Green House Gas emission associated to the manufacture and the transport of turbines and blades.
- **The Location**.
  The wind energy system can be placed almost everywhere (obviously good windy sites are preferred) with a little disturbance to the animal and the general farming activities
- **It helps to diversify the national energy portfolio**.
  In this way, countries can rely on more than one type of energy
- **Wind energy systems have low operating costs**.
  This because there aren't fuel costs associated
- **Reduction of costly transport costs of electricity from far-away power stations.**

On the other hand, the disadvantages are [4]:

- **The intermittency of wind (the wind is unpredictable)**
- **Noise pollution**
- **Initial cost**

According to the Global Wind Energy Council (GWEC), globally, 77.6 GW of new wind power capacity was introduced in 2022, having, in this way, a total installed wind capacity of 906 GW, which means a growth of 9%. From the projections, GWEC expects 680 GW of wind capacity to be added between 2023 and 2027. In particular, 2023 should be the very first year to exceed 100 GW [5].
In the recent years, the wind power industry has spread all around the world given the serious environmental issues created by the extensive use of petroleum as source of energy.

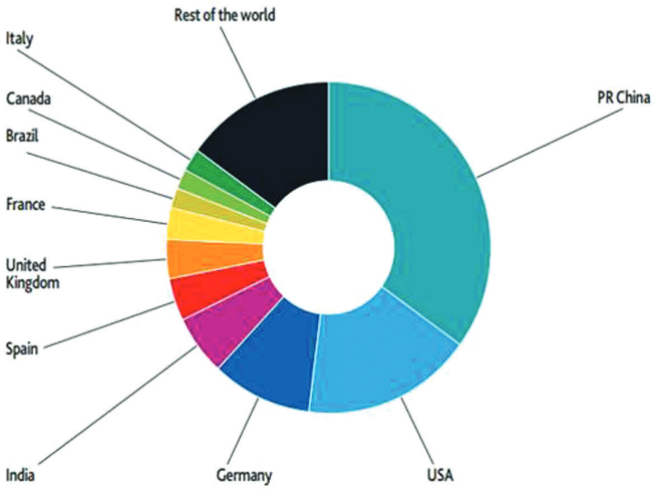


Figure 1: Top ten countries with the highest cumulative capacity [6]

Figure 1 highlights how China is the leader in the current wind energy market.
The projection for the new installation of Regional onshore and offshore wind energy, according to GWEC, is reported in Figure 2.

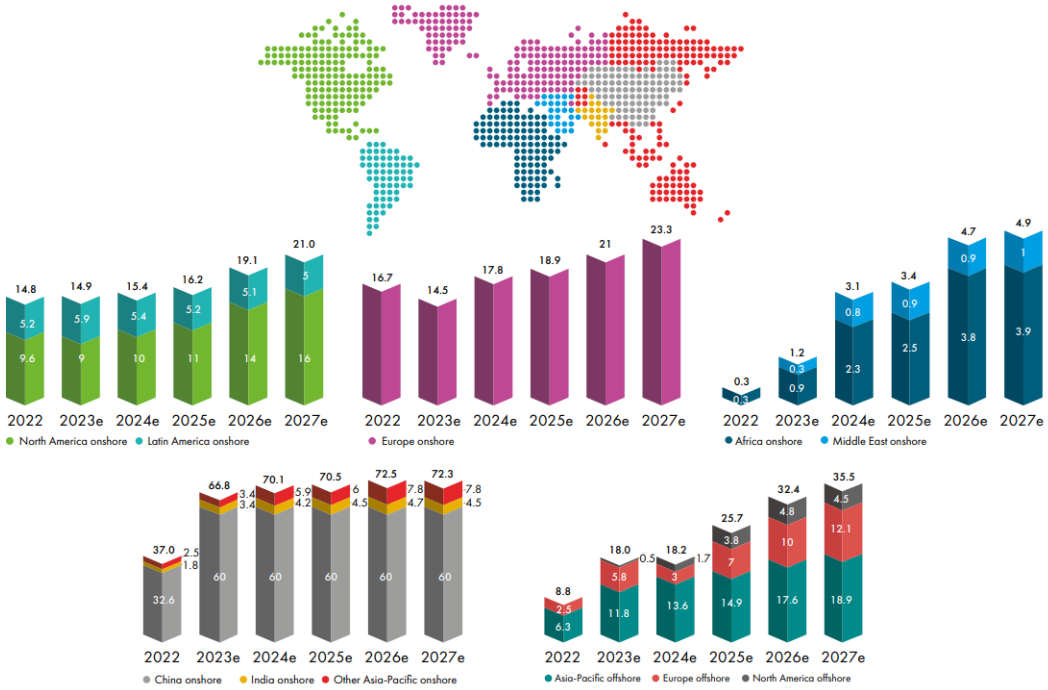


Figure 2: Regional onshore and offshore wind outlook for new installations [5]

In total, 60 GW of onshore wind capacity is expected to be added in the next five years in North America, of which 92% will be built in the US and the rest in Canada [5]. In Europe in 2023, it is expected a decreasing of the onshore wind capacity equals to 13%, due to a slowdown in the Nordic countries [5]. In Africa and Middle East, in total, from the 2023 to 2027, it is expected an addition of 17 GW to the current wind capacity. In more detail, 5.3 GW will come from South Africa, 3.6 GW from Egypt, 2.4 GW from Saudi Arabia and 2.2 GW from Morocco [5]. In the end, in Latin America, GWEC Market Intelligence expects 26.5 GW of onshore wind capacity to be added in this region in the next five years with Brazil, Chile and Colombia contributing 78% of the additions [5].

1.2 Physics beyond wind energy and wind power

Wind energy depends on three different parameters [4]:

- volume of air
- mass of air
- speed of air

It can be defined as the energy content of the air flow due to its motion and this kind of energy is also called kinematic energy:

$$KE = \frac{1}{2}mU^2 \tag{1.2.1}$$

where m is the mass of air, U is the airspeed [4].
Wind power is defined as the rate of kinetic energy flow:

$$P = \frac{1}{2}\rho AU^3 \tag{1.2.2}$$

where A is the cross section area and ρ is the air density. From Eq. 1.2.2, it can be seen a linear relationship between the Wind Power, the air density, a linear relationship between the Wind Power and the cross section area and a non linear cubic relationship between the Wind Power and the airspeed.
The amount of wind energy that can be transferred to the wind turbine blades, in order to generate mechanical energy, is lead by the power coefficient. So, in this sense, it can be said that the power coefficient Cp is a measurement of the efficiency in wind power extraction. The Cp definition is given by

$$Cp = \frac{P_T}{P_{wind}} \tag{1.2.3}$$

where P_T is given by Eq.1.2.2; while P_{wind} is the total power of the wind resource [4].
Strictly related to the Cp is the Beth's Law which states that it is possible to convert a maximum of 59% of the kinetic energy to mechanical energy by using a wind turbine. This is because the wind on the back side of the rotor must reach such high velocity in order to move away and to allow more wind through the rotor plane.

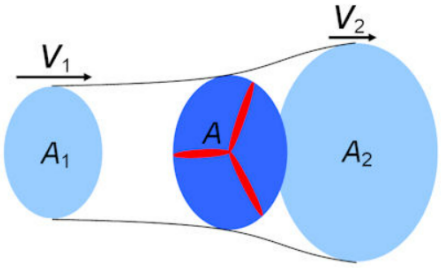


Figure 3: Wind energy scheme [7]

The maximum Cp possible, according to Beth's Law, is $Cp = \frac{16}{17}$ with $v_2 = \frac{1}{3}v_1$ [8]. As it can be possible to see in Fig. 3, the v_1, that is the air speed before the wind turbine, is higher than the airspeed downstream, because the mass flow must be continuous and for the same reason area A2 after the wind turbine must be bigger than the area A1 before [7].

1.3 Wind Turbine

The main goal of Wind Turbine is to convert the kinematic energy of wind into mechanical energy and then, this mechanical energy is transformed into the electrical one [9]. Wind Turbines can be classified into:

- Horizontal axis (HAWT)
  This type of Wind Turbine is the most common and it is characterised by the blades rotating axis parallel to the wind stream. It has the main rotor shaft and the electrical generator at the top of a tower. For this reason, it must be pointed in the wind direction in order to provide high turbine efficiency, high power density, low cut in wind speed and low cost per unit power output [10].

- Vertical axis (VAWT)
  The blades rotating axis is perpendicular to the ground. This kind of WT can accept wind from any direction and there is no need for a yaw control system. The cost is lower than HAWT because all the components (generator, gearbox etc..) can be set up on the ground using a direct drive from the rotor assembly to the ground-based gearbox. The VAWT uses an external energy source to allow the rotation of the blades during the inizialization and consequently it can be pointed independently from the wind direction [10].

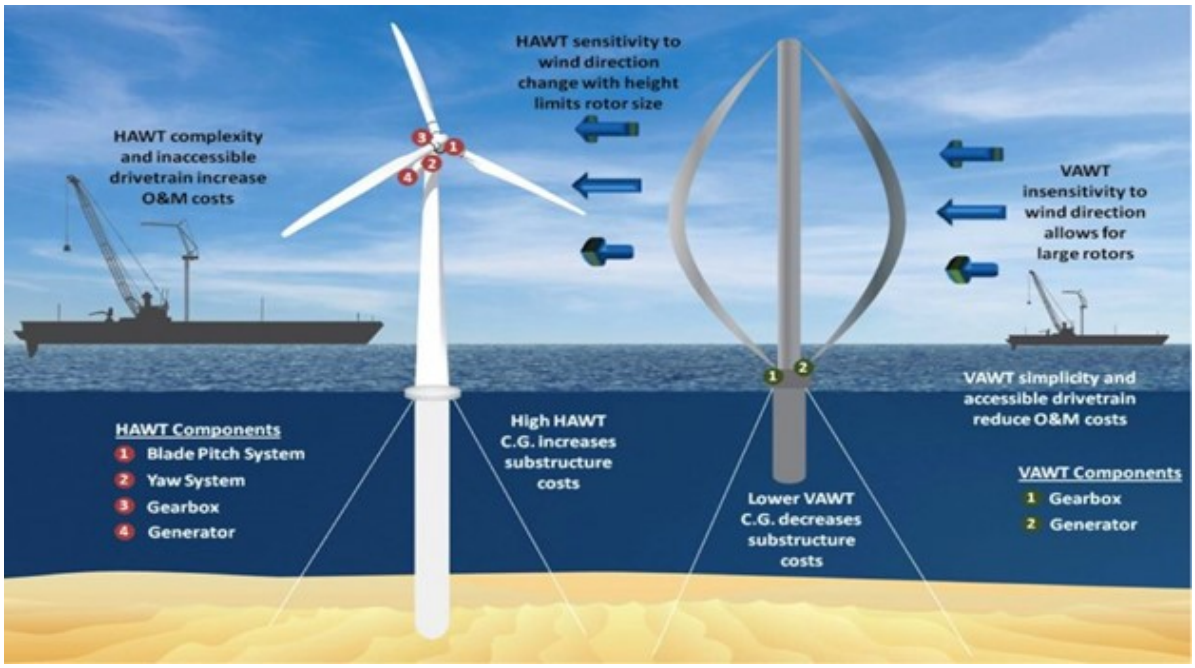


Figure 4: HAWT and VAWT [11]

- Upwind WT
  The wind rotor faces the wind and this allows to avoid the distortion of the flow stream. The majority of the HAWTs are upwind [10].

- Downwind WT
  Wind passes before through the nacelle and the tower and then through the rotor blades. This configuration allows the rotor blades to be more flexible and to reduce the wind resistance. There are great fluctuations in the flow stream and this means that an unstable flow field is generated (causing aerodynamics losses and more fatigue loads on the Wind Turbine that could be damaged) [10].

- Geared Drive
  WT uses a multistage gearbox to increase the generator rotor rotating speed in order to have higher power in output and this gearbox takes the rotational speed from the low speed shaft of the blade rotor and converts it into the fast rotation on the high speed shaft of the generator rotor. This allows to have

WT with smaller size and less weight but, on the other hand, the introduction of the gearbox increases the turbine noise and mechanical losses [10].

- Direct Drive
  The generator shaft is directly connected to the blade rotor and this allows to be more efficient in terms of energy and reliability [10].

According to [12], when the wind speed is high enough to overcome friction in the wind turbine drivetrain, the control systems allow the rotor to rotate in order to generate a very small amount of power. The output power increases rapidly as the wind speed rises [12]. When the output reaches the maximum power the Wind Turbine is designed for, control systems govern the output to the rated power. The wind speed at which the rated power is reached is called Rated Wind Speed and it is usually a strong wind of about 15 m/s[12]. Eventually, if the wind speed increases further, the control system shuts the wind turbine down to prevent damage to the machinery [12].

1.4 Horizontal axis Wind Turbine

The typology of Wind Turbine object of this work is the Horizontal Axis Wind Turbine. More in detail, the HAWT with gearbox is the one most common configuration and for this reason it provides more data in order to develop a data-driven model for prognostics application.
The architecture of this configuration is characterised by a split shaft system, where the main shaft (low-speed shaft) turns slowly with the rotor blades and the torque is transmitted, through the gearbox, to the secondary shaft (high speed shaft) which drives the few pole pair generator [13], as shown in Fig. 5.

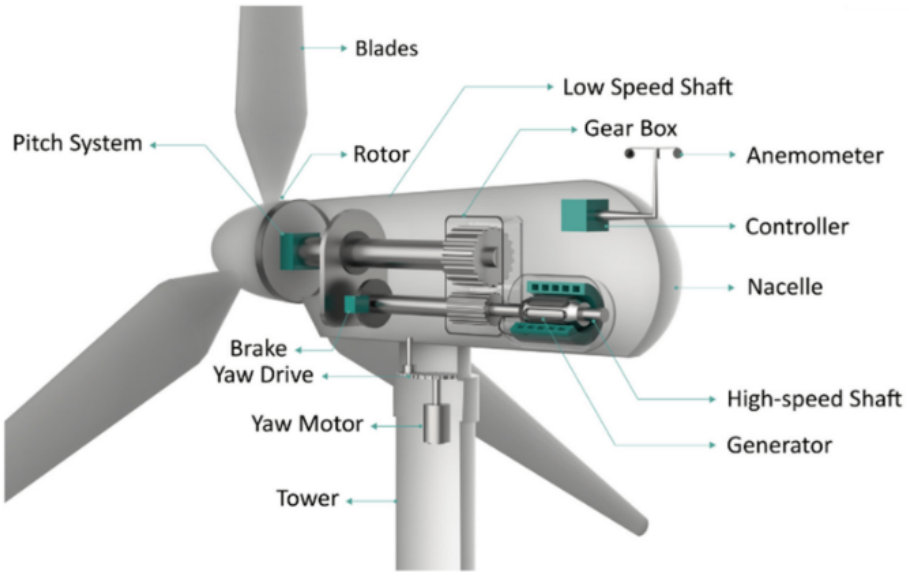


Figure 5: Scheme of HAWT [14]

The gearbox acts as a rotational speed increaser [4] and it is typically lubricated with oil. Due to mechanical losses, the oil could be heated and in this case the cooling becomes mandatory for the Wind Turbine correct functioning.

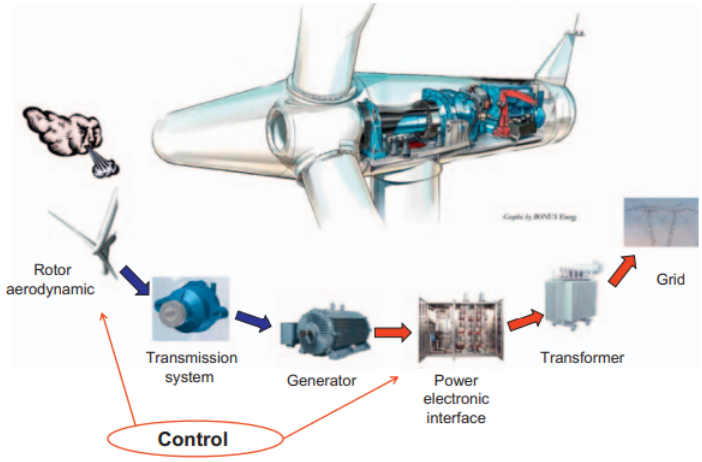


Figure 6: Sub-assemblies of HAWT [4]

The gearbox converts the slow high torque rotation of the aerodynamic rotor (with blades) into the much faster rotation of the secondary shaft (that is the generator shaft) [4]. There are different kind of gearboxes:

- spur and helical
  Spur gearbox has gear teeth parallel to the gear axis and they load bearings radially. Typically, Spur gears are noisier than helical gears because fewer teeth are in contact but the dynamic loads imposed on the gear teeth are greater than helical configuration [15].
  The helical gearbox can be divided into single or doubled helical. In particular, Single helical gears impose both radial and thrust loads on bearings and typically they are quieter than spur gearboxes because there are more teeth in contact. In this sense, helical gear teeth are inclined to the gear axis like a helical screw [15]. In the end, Double helical gears are characterised by all the advantages of single helical gears, and they additionally balance internally generating thrust loads [15].

- Planetary
  Planetary gearboxes allow to have higher power density than parallel axis gears systems, and are able to offer a multitude of gearing options, and a large change in rpm within a small volume. On the contrary, the planetary gearing systems are very complex, vital components are basically inaccessible, and high loads are imposed on the shaft bearings [16].

Wind Turbines are also governed by brake system which controls the power generation. In particular, thanks to the brake large disk on the main shaft, Wind Turbines have the capacity to interrupt the transmission of the torque to the generator [12]. The group of components made by main shaft, mechanical brake and gearbox is called Transmission system.
Another key component for the correct Wind Turbine functioning is the rotor. It is constituted by a hub with blades and typically the most common configuration has 3 blades. The hub is strictly linked to blade control and it is responsible for the connection of the rotor mechanism with the rotor shaft (and consequently with the electrical generator) [17]. The rotor has the function of capturing the power from the wind and to convert it into kinetic mechanical power [4]. The nacelle contains all mechanical organs of the system (gearbox, rotor brake, bearings, the main shaft, the secondary shaft) as well as the generator and control systems [17]. A control system, the pitch mechanism, that is guided by an hydraulic system or by an electrical system, allows to optimize the angle of attack of the blades in order to have higher efficiency for the wind energy conversion and for the power generation stability and it guarantees safe condition in case of high wind speed and/or in case of emergencies [9]. Typically, WTs are also equipped with Yaw system, which uses electric motors and gearboxes, to force the HAWT in the wind direction [17].
One of the most important Wind Turbine component is the generator. In particular, the generator is an electro-mechanical component, which converts the mechanical power into electrical power [4].

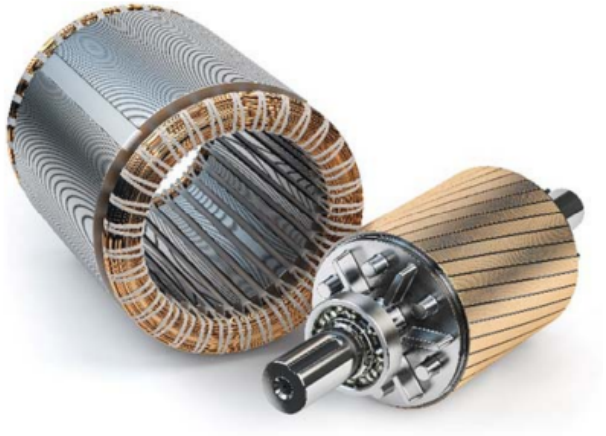


Figure 7: Stator and rotor of a generator [4]

As it can be seen in Fig. 7, the generator is made by a stator, that is a stationary part (it has coils of wire mounted in a certain pattern) and by a rotor that is the rotating part and it is responsible for the generation of the magnetic field of the rotor [4]. The magnetic field passes through the stator windings and a voltage is created in the terminal of the stator [4]. In more detail, if the magnetic field of the stator follows the magnetic field of the rotor, the generator is called synchronous [4].

2 Literature review

In general, the maintenance strategies employed for WT are [18]:

- Corrective, where there is not a condition monitoring approach. As reported in [19], there is a maintenance task only after that a fault has been detected and a refurbishment or replacement of parts could be needed. This would bring to unscheduled downtime of the WT.
- Preventive, which has been considered as scheduled maintenance activities. In fact, repair or replace activities before failures are adopted. Reducing failures implies more frequent maintenance tasks [19].

An evolution of these two previous strategies is the Condition based Maintenance, which is a robust approach where maintenance tasks are performed only in case of a pre-estabilished alarm [18]. It is characterized by continuous monitoring and inspection techniques in order to detect incipient faults and to define the optimum maintenance task before the fault [19], with the aim of avoiding any unscheduled downtime of systems.

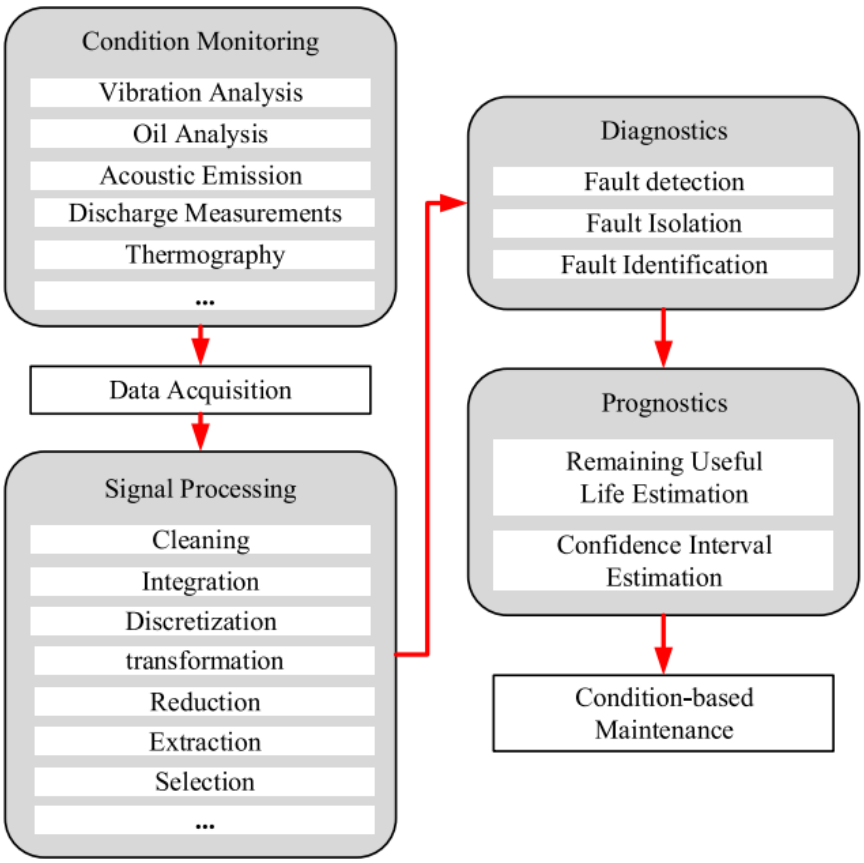


Figure 8: Flow chart PHM of WT [20]

This approach is based on the acquisition, processing, analysis and interpretation of data gathered [19]. When we speak of WTs, we speak of complex machines, assembled thanks to the combination of multiple technologies (Rezamand et al.[20]). Horizontal Axis Wind Turbine is the most promising wind energy technology worldwide, as reported by Astolfi et al.[21]. In general, performances of whatever system degrade with age [21], but being Wind Turbines located in sites with highly variable and harsh meteorological conditions, their degradation rate is much higher. Being exposed to constant changing loads (so highly variable operational conditions) means that WTs are subjected to intense mechanical stress [22]. The most of times, WTs are located in remote sites and for this reason it is not always possible to have access to these places. In this sense, it becomes really important to make predictions and to plan the maintenance activities [23],[24]. The expected life for WT is typically of about 25 years and condition monitoring and maintenance can extend this duration according to the increasing demand of wind energy. The lifetime extension must be applied without compromising safety

standards (in this way it is also possible to increase the return of the initial investment) [1]. It has been esteemed that the operational unavailability of WTs is about 3% of their lifetime. Operation and Maintenance costs can be 10%-20% of the total cost of energy for a wind project and this percentage can also reach 35% for a WT at the end of life, as reported by Tchakoua et al. [22]. For this reason, a predictive maintenance strategy is more suggested in order to reduce systems' unavailability and to reduce maintenance costs. In fact, as reported in [20], the unexpected failure of WT and of its sub-assemblies causes important economic losses. In this sense, a Prognostic and Health Monitoring (PHM) approach becomes fundamental.

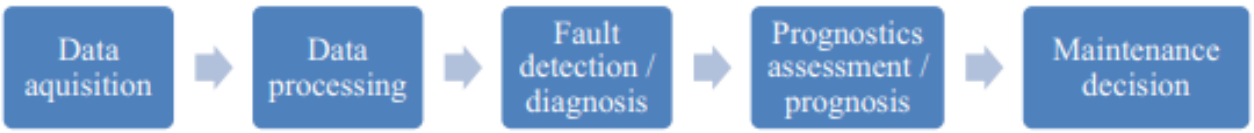


Figure 9: Condition based maintenance [18]

Moving towards predictive maintenance has been possible thanks to the introduction of condition monitoring systems (CMSs) and of Standard Supervisory Control and Data Acquisition (SCADA) system [23]. CMSs allow to monitor several key parameters related to WTs operations (e.g vibration analysis) but they are characterized by high costs. These systems are capable of capturing high frequency dynamics (in contrast with SCADA systems which can't meet the CMSs wide bandwidth) [24],[20]. All large WTs have a SCADA system [23]. As reported in [25], this monitoring system bases its functioning on wireless sensors which communicate with an embedded microprocessor mounted on the devices and the SCADA data are represented on a PC in the form of a database, as showed in Fig.10.

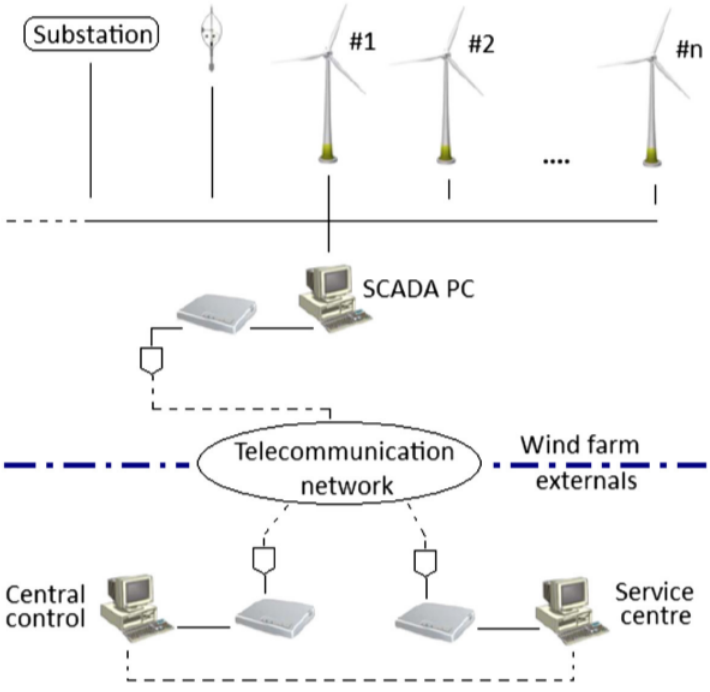


Figure 10: SCADA system scheme [24]

SCADA data are operational data mainly used to monitor WTs performances [24],[23],[25]. They provide a wealth of data, typically, at 10 minute resolution. These data are used for both diagnostics and/or prognostics applications, allowing to always monitor the overall health status of WT and its components [25], without requiring additional sensors [23]. The type of data, representing normal operation and faulty conditions, registered by SCADA system are represented in Table 1.
Using SCADA data, for condition monitoring, allows to inform the operator about the WT conditions, but

Environmental	Electrical characteristic	Part Temperature	Control variables
wind speed	active power output	gearbox bearing	pitch angle
wind direction	power factor	gearbox lubricant oil	yaw angle
ambient temperature	reactive power	generator winding	rotor shaft speed
nacelle temperature	generator voltages	generator bearing	generator speed
	generator phase current	main bearing	fan speed/status
	voltage frequency	rotor shaft	cooling pump status
		generator shaft	number of yaw movements
		generator slip ring	set pitch angle/deviation
		inverter phase	number of starts/stops
		converter cooling water	operational status code
		transformer phase	
		hub controller	
		top controller	
		converter controller	
		grid busbar	

Table 1: SCADA data [23]

as reported by Tavner et al. [24], they can't guarantee a full WT condition monitoring (incapacity of fault isolation and identification) as it is guaranteed by wind turbine condition monitoring systems.
In addition, another downside of the SCADA systems is that could report false alarms led by the wide variety of loads applied.

2.1 Feature extraction and Data Preprocessing

According to [18], SCADA data can contain errors due to malfunctions in the data collection system and to sensor faults. In this sense, the pre-processing phase becomes very important for treating the data which will be used to feed a Machine Learning (ML) model. Pre-processing steps involve data exploration and filtering out errors, using criteria based on the operator's experience [18]. Zhao et al. [26] proposed a data pre-processing procedure for SCADA data made by 4 steps: data cleaning, features selection, features reduction and data balancing. According to [27], in the initial phase (data cleaning), before building a ML model, it is necessary to define the outliers. Puig et al. [28] investigated some techniques in order to identify and to remove outliers. Each SCADA dataset considered in the analysis has been divided into a training dataset and a testing dataset. Puig's analysis took into account also WTs failures events. The techniques investigated have been: Extreme Studentized Deviate (EDS) Filter, Quantile Filter, Hampel Identifier. They showed that the outliers filtering methods can decrease the error on the training dataset but increase the error in the testing data. Yang et al. [29] applied the Chebyshev inequality method in order to remove the outliers in raw SCADA data, because this method does not make any assumptions about the type of distribution of data. Udo et al. [30], considering the "La Haute Wind born" Wind Farm took some key parameters such as the active power, the wind speed, ambient Temperature, generator bearing. Then, for the phase of the cleaning data affected by errors, Udo et al. adopted 4 different criteria depending on the type of error. According to Stetco et al. [27], feature selection is the process of selecting variables related to the output that we want to study and they defined 3 approaches: wrapper methods, embedded methods, Filter methods. The use of these different approaches for selecting variables can depend on the particular case and on the object of the study. In this sense, the feature reduction consists of transforming high-dimensional data into a meaningful representation of reduced dimensionality and there are many techniques that can be implemented, but the most used is the Principal Component Analysis (PCA).

2.2 Fault Detection and Diagnosis

As reported in [31], one of the most important step in Condition monitoring approach, in order to detect and isolate the type of failures, is the Fault Detection and Diagnosis (FDD). We have to make a first distinction between Onshore Wind Farm (WF) and Offshore Wind Farm. Being under different environmental conditions, as we can see in Fig. 11, implies that same components have different failure rates depending on the WT environmental conditions.

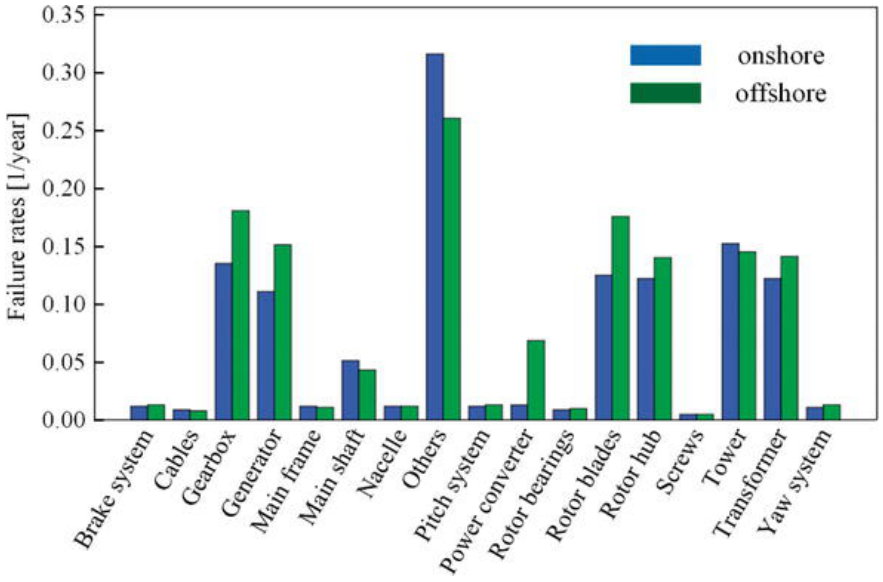


Figure 11: Failure rate of Onshore and Offshore WT [32]

For the onshore WF, the components that have the highest faults rate are: towers, gearboxes, and rotor blades; while for the offshore wind farm gearboxes, rotor blades, generators, and towers are those most affected by anomalies [20]. In this sense, it has been observed, in general, that the most critical components in the WT are: gearbox, main bearing and blades.

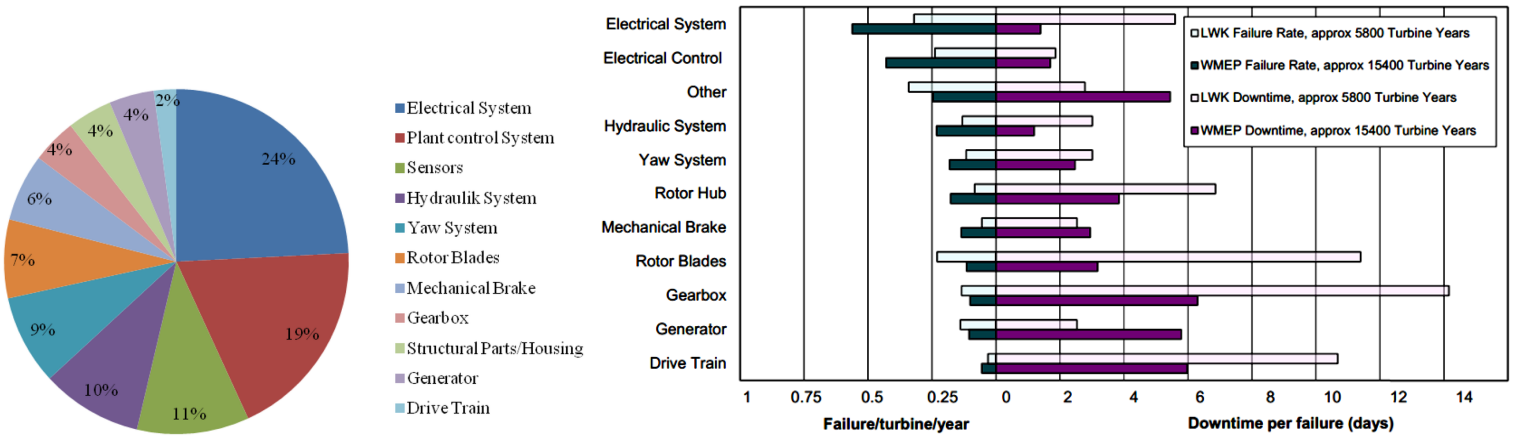


Figure 12: Percentage of failure of WT subsystems [22]

Yang et al. [33] investigated, through the analysis of the SCADA data, observing the correlation between SCADA variables, the possible causes that can bring to failure the main components of the WT, as reported in Fig. 13.

Subassembly	SCADA parameters	Reasoning rules for fault location
Anemometer	1. Wind speed measured from the anemometer installed at the top of WT 2. Wind speeds measured at different heights of the metrological mast	Normal correlation of metrological wind speeds at different heights, and miscorrelation of anemometer data vs. metrological wind speeds indicates a failure with the anemometer.
Yaw system	1. Yaw angle 2. Wind vane data	Yaw angle should be well correlated to wind vane data. Miscorrelation between them indicates the fault. Yaw motor temperature, lubrication grease level and brake pressure are helpful in fault shooting.
Pitch system	1. Pitch angles of blades 2. Wind speed	Miscorrelation of pitch angles of different blades, and miscorrelation of either one of the pitch angles and wind speed indicate the occurrence of fault.
Blades	1. Wind speed 2. Rotor speed 3. Generator power 4. Generator speed 5. Gearbox vibrations 6. Gearbox oil level	Correction between generator speed and generator power is used to confirm the condition of generator; Correlation between rotor speed and generator power, correlation between rotor speed and generator speed, correlation between gearbox vibration and rotor speed, and gearbox oil level against time can be used for checking the health condition of gearbox and coupling. If above correlations are well maintained, then the miscorrelation between wind speed and rotor speed, and miscorrelation between wind speed and the shaft torque derived from the ratio of generator power and generator speed could indicate something is wrong with one or more blades of the WT.
Tower & foundation	1. Wind speed 2. Tower vibration	Miscorrelation of wind speed vs. tower displacement implies the occurrence of tower structural failure or foundation failure.
Main bearing	1. Temperature of main bearing 2. Grease level for lubrication of main bearing 3. Rotor speed 4. Generator power	Miscorrelation between main bearing temperature and rotor speed, and miscorrelation between main bearing temperature and generator power indicates a fault in the bearing. Grease level is helpful for excluding the lubrication problem.
Gearbox	1. Gearbox vibrations 2. Temperature of gearbox oil 3. Rotor speed 4. Oil level in gearbox sump 5. Gearbox oil pressure	Both miscorrelation between rotor speed and gearbox vibration, and miscorrelation between rotor speed and gearbox oil temperature indicate the occurrence of a fault in gearbox. Both oil level and oil pressure will give indication when oil leakage happens.
Generator	1. Generator speed 2. Generator active power 3. Generator reactive power 4. Generator side 3 phase power factor measured at converter	Miscorrelation between generator speed and generator active power will indicate a fault in the generator, and miscorrelation between generator speed and generator reactive power will indicate an electrical fault in the generator. Generator side 3 phase power factor measured at converter is also a good indicator of generator fault.
Converter	1. Generator side temperature 2. Grid side temperature 3. Generator side 3 phase power factor 4. Grid side 3 phase power factor 5. Generator side 3 phase-to-ground voltages 6. Generator side 3 phase currents 7. Grid side 3 phase-to-ground voltages 8. Grid side 3 phase currents	Miscorrelation between the converter temperatures at grid side and generator side will indicate a converter failure. Miscorrelation between the 3 phase power factors at generator side and grid side also clearly indicates a fault occurring in the converter. Analysing the unbalance of 3 phase-to-ground voltages and 3 phase currents at both generator side and grid side is helpful for locating the fault.
Transformer	1. Transformer temperature at turbine side 2. Transformer temperature at grid side 3. Turbine side 3 phase-to-ground voltages 4. Turbine side 3 phase currents 5. Grid side 3 phase-to-ground voltages 6. Grid side 3 phase currents	Miscorrelation of temperatures and power values measured at both sides of transformer could indicate a fault of it. Analysing the unbalance of 3 phase-to-ground voltages and 3 phase currents at both turbine side and grid side will be helpful for locating the fault.

Figure 13: Correlation between SCADA parameters and possible causes of faults [33]

For this reason, with the aim of reducing the failure rate inevitably associated with systems' downtime, it becomes really important speaking of diagnosis, prognostics and RUL prediction. According to Gao et al. [34], the approaches to the faults diagnosis can be divided into 4 macro areas: model-based methods, signal-based methods, data driven methods, and hybrid methods. The hybrid methods are combination of the previous approaches and for this reason they are not considered in this literature review.

2.2.1 Model Based Diagnosis

This approach requires to build a Wind Turbine model which takes as input the same inputs of the real time wind turbine. Then, the goal is to monitor and verify the differences in the outputs of the two wind turbine models [34]. To be able to consider healthy conditions, the residual must ideally be zero or less than a threshold that has been imposed according to a particular criteria considered. The basic concept of the model based approach can be seen in Fig. 14 .

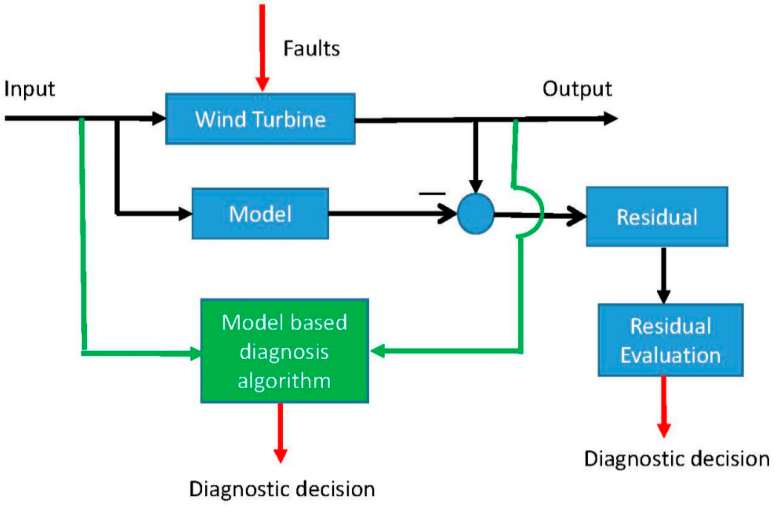


Figure 14: Model based Approach to Diagnosis [34]

An example of model based approach is provided by Garlick et al. [35] who assumed a least squares (LS) method and an Autoregressive with eXogenous input (ARX) model structure, as system identification method applied to raw SCADA data, in order to develop a discrete time dynamic model for the early indication of bearing and gearbox faults. Zhang et al. [36] based their diagnosis of rotor winding failure on a dynamic model sensor method which reported the relationship between the generator temperature, wind speed and ambient Temperature. According to [37], an abductive diagnosis can be considered for WTs and this approach strongly depends on the formalization of the relationship between specific failures and their effects. One of the model based approach most used for fault diagnosis is the observer based [34]. Cross et al. [38] employed the simplified refined instrumental variable (SRIV) algorithm for discrete-time transfer function model identification. This algorithm is based on an iterative procedure in which each step allows a linear-like computation of recursive estimates. This method provided some alerts related to the system even if its Temperature was within the range and in phase of check a blade of a fan was discovered broken. Shao et al. [39] proposed a parameter-varying model, which is a real time updating non linear model, in order to provide fault reconstruction on a 4.8MW wind turbine system. The model is designed offline and performed and regulated automatically online. Sanchez et al. [40] introduced a model based fault detection method which combines the use of analytical redundancy relations (ARRs) and interval observers.
Another model based approach is based on Kalman filter. Dey et al. [41] used a Cascade Kalman filter based technique in order to detect and isolate several faults in components like rotor, converter actuator, drivetrain. Wei et al. [42] introduced a Kalman filter based diagnosis algorithm in order to detect additive and multiplicative sensor faults. In particular, they focused on the detection of blade root moment sensor faults for a three bladed horizontal axis wind turbine. The approach developed by the authors has been also applied in order to detect additive and multiplicative faults using the residual of the Kalman filters. Cho et al. [43] focused on a Kalman filter approach based on residual generation to detect faults generated in blade pitch sensors and actuator in a floating wind turbine.

2.2.2 Signals based Diagnosis

Signal based methods are based on data like electrical signals, vibration, sound signals gathered by sensors installed in wind turbines [34]. This kind of approach needs of processing techniques that allow to extract the critical information from signals. In order to understand the flow chart of signals based diagnosis approach, we can see Fig. 15 .

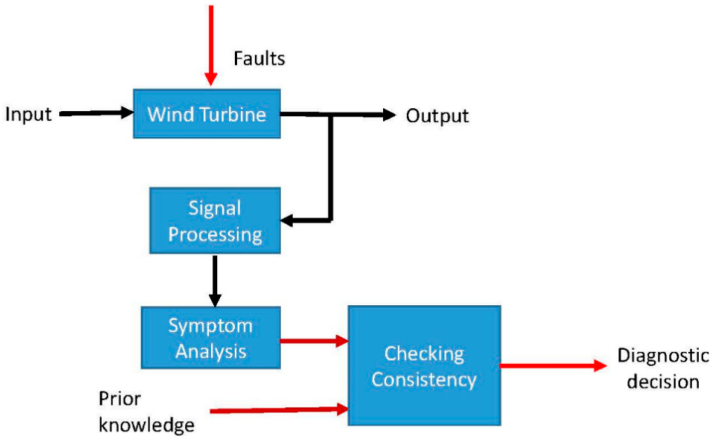


Figure 15: Signals based Approach to Diagnosis [34]

Tamilsevan et al.[44] presented a vibration based two stage fault detection which integrated both analytical diagnosis and graphical diagnosis of a NREL wind Turbine gearbox. Zappalà et al. [45], starting from vibration data of a NREL wind turbine Gearbox Condition Monitoring Round Robin project, developed a gear condition indicator, the sidebands power factor algorithm (SBPF), to detect and diagnose damages on the high speed shift (HSS). The SBPF algorithm allowed to assess the gear fault severity by tracking a progressive tooth damage. Yang et al. [46] investigated the condition monitoring and the faults diagnosis for a wind turbine subjected to varying torque from the wind resource. In this study, the discrete wavelet transforms (DWT) and the continuous wavelet transforms (CWT) have been employed. In particular, the DWT has been used for noise cancellation, while the CWT has been used for the features extraction. Watson et al. [47] proposed a wavelet analysis applied to two doubly fed induction generators (DFIGs) to detect the rotor eccentricity analysing the generator shaft misalignment. Rotor eccentricity is often the result of increased bearing wear and an indication of potential failure.
According to Gao et al. [34], Signal-based monitoring methods don't need to define an explicit mathematical model for wind turbine system. In general, it is implied to monitor and to do diagnosis of rotating components of wind turbines, such as wheels and bearings of gearbox, bearings of generator and main bearing.

2.2.3 Data driven Diagnosis

One approach to anomalies detection, proposed by Weinert et. al [23], is trending. This approach can be considered as a static analysis based and it is really useful in the anomaly detection but it can bring to false alarms [1] because, considering also the high variability in the operational conditions of WT, a change in the value of a SCADA parameter is not an evidence of a fault [23]. For example, as reported in [23], [48], it was applied a PCA trending approach with an auto-associative neural networks. The PCA approach allows to extract the useful and non redundant information registered by the sensors, but it allows only a linear data analysis. For this reason it has been combined with AANN in order to obtain a non linear PCA[48]. Using this approach, as indicated by Weinert et al. [23], it has been possible to detect a failure also when no advance signals alarms were detected. Feng et al.[49], analysing the physical procedure of kinetic energy transmission and dissipation, developed a monitoring algorithm that connected the transmission efficiency with the temperature rise and the rotational speed in order to detect failures related to the gearbox of a 2MW variable speed turbine. They showed that the gearbox efficiency decreased with the rise of temperature of the gearbox. Yang et al. [33], reported that the value of WT SCADA data would have been predictable only once the operational condition is known. Any deviation of the measured value from its predicted value would have indicated a fault. For this reason, it has been developed a trending method that uses bin averaging by wind speed, generator speed or output power to make those predictions [23],[33].
According to [50], data-driven WT fault diagnosis methods are based on data mining technology to obtain useful information to define the normal conditions of WT and to detect the fault. Widely used for WT diagnosis are ML model. We can have different ML Diagnosis approaches:

- **Supervised Learning Diagnosis approach**
  Firstly, we have raw SCADA data gathered by Wind Turbine sensors. Then, we have to divide these data into training, validation and testing set. The pre-processing is performed on the dataset and the normalization is needed after the processing of the missing values to avoid gradient-explosion problems. Secondly, a ML algorithm is chosen in order to train the training set which is used for modeling [50]. The testing set is used to evaluate the model quality and as the final output an accurate fault classification is obtained. Artificial Neural Network (ANN), according to [50], is one of the most popular and widely used supervised learning model. Zhang et al. [51] developed an ANN which can be used to detect the corresponding fault of rear bearing by comparing estimated and actual temperature. For example. Zaher et al. [52] introduced a multilayer neural networks in order to detect anomalies analysing the gearbox temperature of a Scottish wind farm. They presented as the best architecture an ANN with 3 neurons in the hidden layer. Zhang et al. [53] made a comparison between an ANN model and a mathematical model in order to define which was able to detect a faults of rear bearing by comparing estimated and actual temperature. The ANN provided results more accurate than the mathematical model. Brandao et al. [54] proposed a NN approach to detect failures of the gearbox thanks to the deviation of the real Oil Temperature of the component from the Temperature esteemed by the neural network. Wang et. al [55] investigated a deep neural network (DNN) approach in order to monitor and identify WT gearbox failures using SCADA data. In addition they made a comparative analysis between 6 data mining algorithms: k-nearest Neighbors (kNN), Lasso, Ridge, SVM, shallow neural network and DNN. The result was that the DNN algorithm was the most accurate. Bangalore et al. [56] investigated, analysing SCADA data, the performance, in faults detection for a WT gearbox, of a NARX-ANN algorithm which provided accurate results.
  Another Supervised Learning ML methods is Support Vector Machine (SVM). It is a non parametric statistical approach which can be easily used to detect faulty response of WT[34]. Laouti et al. [57] used a SVM approach in order to detect fault in a variable speed horizontal-axis wind turbine composed of three blades and a full converter. The Kernel used for learning all the faults is Gaussian (with different values of variance). This approach was able to detect and isolate faults related to different systems such as sensors of pitch position, sensors of generator, sensors of rotor speed and convertor actuator. Santos et al. [58] studied the performance, to detect mechanical faults (like imbalance and misalignment) in a wind turbine, of two different methods: SVM and ANN. These two approaches were applied at the same vibrations dataset. In their work, the authors used 4 kernels: linear, Gaussian, stump and perceptron.

At the end, the linear SVM showed the highest accuracy. Liu et al. [59] applied a single "1-a-1" support vector machine (SVM) classifier for fault identification, and a multi-SVM, with radial basis function (RBF) as the kernel function, in order to identify the location and fault type of renewable energy power grid. Castellani et al. [60] focused on the diagnosis of electrical faults of generators using SCADA data. In particular, they developed a data driven normal behaviour models, built considering a support vector regression with a Gaussian kernel, that describes the relationship between electrical parameters and operation variables. According to [50], the SVM approach uses inner product kernel function to turn the raw data gathered from the wind turbine to linear data through mapping the raw data to a high dimensional space. Yang et al. [29] proposed a Support Vector Regression (SVR) approach. This machine learning algorithm showed excellent performance in pattern recognition and regression analysis. Using SCADA data, the object of interest of their work was the Wind turbine high Temperature anomalies.
As supervised Learning approach for WT diagnosis, we also consider the Decision Tree (DT) method. Abdallah et al. [61] applied a DT algorithm in order to detect wind turbine faults using telemetry data. The authors trained an esemble Bagged DT classifier on a dataset from an offshore wind farms made by 48 WTs. The DT algorithm has been used to link excessive vibration faults to their possible causes. DT approach is easy to implement but it has limitations in dealing with missing values [50].
Another Supervised Learning approach is the Esemble Learning. This method relies on the concept to adjust and train base learners as esemble members into a strong Learner that should have greater performance [50] in detecting faults. Bootstrap aggregating (also called bagging) algorithm helps to reduce variance and to prevent overfitting [50]. Mansouri et al. [62] applied a Gaussian Process Regression Multi-Class Random Forest GPR-MCRF. The authors extracted the effective features through a Gaussian process and then they applied a multi class random forest classifier in order to diagnosis different type of faults for a Wind turbine. They compared the performance of different approaches like Naive Bayes, kernel PCA-based MCRF, Support Vector Machines, PCA-based MCRF, Decision Tree, K-Nearest Neighbors, and Discriminant Analysis approach. It resulted that the GPR-MCRF was the most accurate in detecting faults. Li et al. [63] used a deep random forest fusion (DRFF) technique in order to improve fault diagnosis performance for gearboxes by using measurements of an acoustic emission (AE) sensor and an accelerometer that are used for monitoring the gearbox condition simultaneously. The RF is introduced to fuse simultaneously acoustic and vibratory features into an integrated aspect. It resulted that the DRFF can improve fault diagnosis capabilities for gearboxes compared with conventional RF. Boosting methods can be applied in order to improve the performance classification. Examples of boosting methods are XGBoost and LightGBM. Zhang et al. [64] applied a random forest (RF) approach in combination with extreme gradient boosting (XGBoost). The RF model is used in order to evaluate and sort the importance of features for all the faults. Then the fault classifier is trained based on XGBoost model with three top ranking features selected in the previous step.
The last Supervised Learning approach that has been considered is the Deep Learning (DL). Zhao et al. [65] proposed a deep learning method based on a deep auto-encoder (DAE) network using Wind Turbines SCADA data. The DAE network model is more accurate in modelling wind turbine component dynamic behaviour by working on a closer level of mimicking the working process of a natural brain. This method can provide the early warning of the faulty component and derive the physical location of the faulty WT component through the residual of the deep autoencoder network mode [50],[65]. Xiao et al. [66] presented convolutional neural network models to wind turbine converter fault detection using convolutional neural network by using wind turbine SCADA system data.

- **Unsupervised Learning Diagnosis Approach**
The unsupervised learning is based on the concept of learning unlabelled data in order to reveal the hidden structure and to define the key features of data, allowing to divide the data into several categories[50]. The representative technique is clustering. Clustering is an evolution of the trending and cluster algorithm are used to make the separation between normal and faulty observations. As the trending method also the clustering approach can provide false alarms and for this reason it is important having a large number of failures data [23]. For example, Kim et al. [48] developed an unsupervised clustering algorithm: a self organizing feature map (SOFM) which forms neurons located on a regular grid, generally in one ore two dimensions. This algorithm can detect regularities and correlations in its input and it will adapt the

responses to the input [48]. The SOFM has an isolation capability that it is able to generate a number of clusters according to the number of the failure modes detected. According to Kim et al.[48], with appropriate algorithms, performance monitoring, thanks to SCADA data, can be employed in order to detect individual component fault, because a reduced efficiency in the performance means that something is not working well and it needs an investigation. Kim et al. [48] proposed a diagnostics technique based on SCADA data using anomaly detection algorithms and clustering techniques. Luo et al. [67] introduced a combination of Ensemble Local Means Decomposition (ELMD) and Fuzzy C-Means clustering (FCM) methods for fault diagnosis of a gearbox of a wind turbine. The FCM algorithm is a clustering algorithm in which each data point can belong to more than one cluster [50]. Lapira et al. [68] investigated 3 methods, using SCADA data gathered from a large scale on shore wind turbine for 27 months. In particular they compared Neural Networks based on residual analysis with two unsupervised methods Self-Organizing Maps (SOMs) and Gaussian Mixture Models (GMMs). GMM resulted more suitable for showing degradation trend. Zhao et al. [69] investigated the application of a SOM neural network in order to detect the faults of a WT using the status data gathered. The authors showed the good robustness and validity of the approach in detecting wind turbine malfunctions.

2.3 Prognostics and RUL prediction

Prognostics is defined as predictive diagnostics, which includes determining the Remaining Useful Life (RUL) or time span of proper operation of the component [70]. The methods for RUL estimation, according to [18], can be divided into 4 groups, as reported in Fig. 16: Knowledge based models, Life Expectancy models, Artificial Neural Network, Physical models

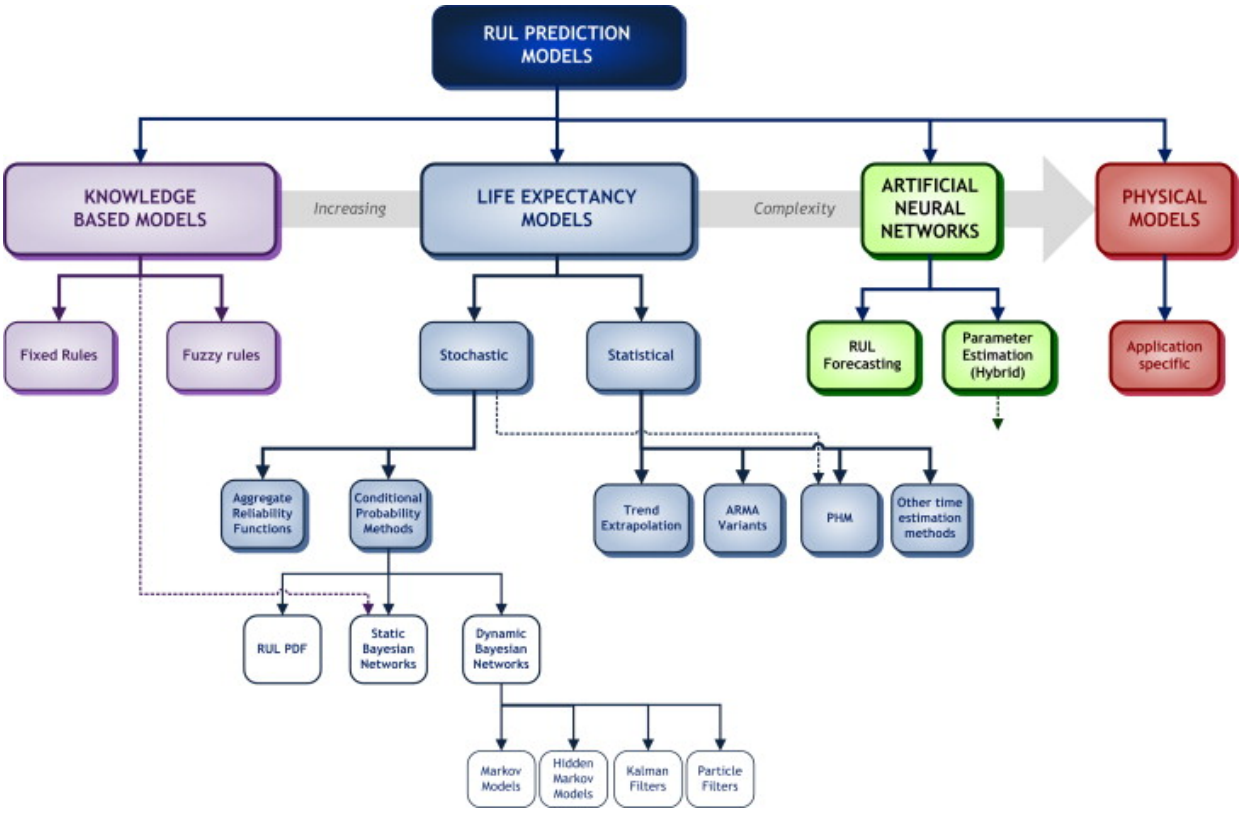


Figure 16: RUL [71]

But these 4 kind of model can be divided into 3 categories: physics-model based, data-driven model and hybrid model. A summary of the advantages and disadvantages of different prognostics approaches is reported in Fig. 17.

Prognostics Approaches	
Advantages	**Disadvantages**
Data-driven approaches	
- Relatively simple and rapid to deploy. - Aids in the assessment of massive amounts of data to acquire a better knowledge of bodily dynamic system behavior.	- There are no physical cause and effect relationships used. - It is difficult to strike the right balance between generalization and learning specific data trends. - Requires large amounts of data.
Physics-based approaches	
- Based on modeled cause-effect correlations, the prediction results are intuitive. - Only calibration may be required for different cases once a model has been built. - Drives sensing requirements.	- Developing models is not trivial. - High-fidelity models can be computationally costly to run, making them unsuitable for real-time applications.
Hybrid approaches	
- No necessity for high-fidelity models or big data. - Maintains the model's intuitiveness while explaining observed data. - Assists in the handling of uncertainty.	- Requires both data and the models. - A faulty model or noisy data can lead each other's approaches to be biased.

Figure 17: Advantages and Disadvantages of Prognostics Approaches[72]

2.3.1 Physics based Prognostics

Physics-based prognostic methods are based on the concept to build mathematical model in order to describe failure modes physics [20]. Breteler et al. [73] introduced a general physics of failure based methodology considering both SCADA and CMS data from operating wind turbines. This method relied on the concept to relate the design load and additional load generator to the consumption and RUL for two cases. In this work, the prognostics phase analysed the degradation pattern and it aimed to define if the pattern found could be representative for future degradation for WTs gearbox. Zhu et al. [74] derived a physical based model in order to define the mathematical relationship between lubrication oil degradation and particle contamination level for a wind turbine. Wang et al. [75] developed a new model based approach of integrated fault diagnosis and prognosis for wind turbine remaining useful life estimation, especially the cases with limited degradation data. Wang et al. [76] introduced probabilistic damage growth model to characterize individual wind turbine performance degradation and failure prognostics. Saidi et al. [77] introduced an integrated prognostic method for WT high speed shaft bearing, which integrated physical degradation model and data driven models. They built a model based on Paris's law and used a Kalman smoother to estimate the RUL. The physics based approach has been applied mainly for the prognostics of WTs structural damages, considering fatigue life models, progressive damage models, deterministic physical law or a stochastic process. The physical based approaches, according to [72], are more accurate than data-driven and/or hybrid methods, but they are very limited because physical based models are built under ideal conditions with numerous assumptions.

2.3.2 Data Driven Prognostics methods

The focus of this literature review is on the data driven approach using SCADA data. SCADA data provides a rich source of continuous time observations that can be used for the monitoring of the whole WT [48]. Performance degradation can highlight problems related to different components. In order to do prognostics, different methods have been developed, using SCADA data. A data driven approach to WT prognosis is AI-Based Prognostic Method such as ANNs, DL, and ANFIS. Artificial Neural Network are mainly employed in case of non linear relationship between observations, using training data. There is one input layer, a variable number of hidden layers and one output layer and each layer is made by a different number of neurons which consist in a non linear transfer function to combine the inputs and an activation function [23]. A key feature of the ANN is the high processing speeds and it is possible to provide correct responses from noisy or partially incomplete data [25]. Matthews et al. [78][79] investigated different AI techniques, including Fuzzy Inference System (FIS), k-means clustering, Self-organizing Map, Artificial Neural Network (ANN), Naïve Bayes, Bayesian Network, Support Vector Machine and Adaptive Neuro-Fuzzy Inference System (ANFIS) considering SCADA data in order to detect the incipient WT pitch system faults. They proved the feasibility of a prognosis based on a-priori knowledge-based ANFIS applied to Wind turbine pitch faults prognosis considering SCADA data. Kusiak et al. [80], in order to study bearing faults, made a comparison between 5 ANN models with different number of neurons and activation functions considered. The best model was ANFIS with 18 neurons and logistic hidden activation. Carrol et al. [81] investigated as machine learning

algorithms: artificial neural network, a support vector machine (SVM) and a logistic regression approach in order to predict the failure and remaining useful life of wind turbine gearboxes. Between the three methods, the artificial neural network provided 72.5% of failures being correctly predicted and least missed failures. Schlechtingen et al. [82] introduced two neural network models: FSRC, Autoregressive. Both models have been implemented in order to predict the generator bearing temperature. The FSRC NN was capable of predicting the temperature analysing the stator and nacelle temperature, shaft speed and the output power. On the other hand, the Autoregressive NN gave more accurate results due to the large heat capacity. Kusiak et al. [83] studied the accuracy in predicting faults and turbine states of five data-mining algorithms: neural network (NN), support vector machine (SVM), random forest algorithm (RFA), boosting tree algorithm (BTA), and general chi-square automatic interaction detector (CHAID) algorithm. In the phase 1 of predicting the best two algorithms were BTA and CHAID. Zhao et al. [26] used an unsupervised learning to cluster the operational state of the generators in order to determine the RUL of the generator. Firstly, a DBSCAN method has been applied for a density-based spatial clustering of applications with the noise and then, in order to measure a wind turbine's performance, it has been proposed the anomaly operation index. Wang et al. [25] characterized ANFIS model defining the advantages of this approach and some application. Hsu et al.[84] analyzed and predicted maintenance needs of wind turbines by using the wind turbine historical data collected in the ChangHua Coastal Industrial Park, Taiwan. The authors, after the phase of identifying four different types of faults thanks to a statistical approach, used two machine learning algorithms, decision tree and random forest classifications, in order to predict wind turbine abnormalities with accuracy rates higher than 92%.
To predict failures related to WTs components, supervised learning approach can be implemented: regression and classification. In particular, with regression models, it is possible to predict a numerical variable (dependent variable like power); while for the classification a categorical variable is predicted [21]. Guo et al. [85] proposed a temperature trend analysis method based on the Nonlinear State Estimate Technique (NSET). NSET is applied in order to define the normal operating model for the wind turbine generator temperature and then at each time step the model is used to predict the generator temperature, using SCADA data.
According to Stetco et al. [27], classification finds a relationship between independent variables typically grouped in vectors and one of several predefined categories identified by labels. Leahy et al. [86], using SCADA data, applied classification techniques in order to make predictions. After under/oversampling, a SVM classifier approach has been used. In particular, the training data fed the SVM algorithm. In addition to this, they used ensemble meta-learners in order to reduce bias and variance in the results. Herp et al. [87] developed a model in order to define a high prediction horizon and to be able to predict the remaining lifetime until the failure, where the T of the main bearing component is modelling applying an ANN.

2.3.3 Life expectancy Prognostics models

These models can be divided into two categories: Stochastic and Statistical.
The first stochastic prognostics based techniques, that have been considered are Bayesian networks which are a type of probabilistic open chain graphical model for estimating probabilities [20]. According to Rezamand et al. [20], Bayesian networks can be applied in order to assess, for different scenarios, the root causes of an event or, in the case of time-series modeling, to determine probabilities associated with a particular future event. The most common Bayesian techniques includes the Markov models, Kalman Filters (KPs) and Particle Filters (PFs). The Markov models allow to estimate probabilities of future failure by determining probabilities associated with each state and probabilities associated with transitioning from one state to another [20]. Lau et al. [88] proposed an estimation of the RUL considering an Hidden Markov Model (HMM) which is a statistical Markov model. Lau et al.[88] developed a Particle filter approach in order to do WT prognostics. The method aims to approximate the relevant distributions with articles and their associated weights. The state of Probability Density Function (PDF) is estimated and it is used to predict the evolution in time of the fault indicator. According to [89], the PF method is suitable for predicting a nonlinear stochastic process with noisy measurements. Cheng et al. [89] developed a new particle filtering (PF) method applied in order to predict the RUL of a bearing in the drivetrain gearbox of a 2.5 MW wind turbine. Fan et al. [90] applied a PF method in order to predict the RUL of a wind turbine gearbox based on information from SCADA system. Some hybrid approaches for the RUL prediction have been introduced using SCADA data. Hsu et al.[84] defined, thanks

to statistical approaches (e.g. Pareto chart and the cause and effect diagram), the frequency of faults that interested the Wind Turbines, after having applied two machine learning algorithms (decision tree and random forest classifications).

It is possible to observe in Fig. 18, a summary of the advantages and disadvantages of the different approaches introduced in the previous sections.

Prognosis technique	Advantages	Limitations
AI-based methods (ANNs/ DL/ DBNs/ RNNs/ CNNs/ ANFIS)	Efficient and practical at modeling multi-dimensional, complex, and non-linear systems	Require a large amount of data as representative of actual data range and its variability for training, computationally intensive
Aggregate reliability functions	Capable of performing at all equipment hierarchy levels, particularly when a few failure modes exist, simple and well understood by reliability engineering community	Require a statistically meaningful sample size of each failure mode for reliable RUL predictions, require statistically independent and identically distributed failures
Bayesian networks	Can easily manage imprecise, noisy, or incomplete datasets	Computational challenging in determining a prior unknown network. Thus, a Bayesian network is only as beneficial as the previous knowledge is reliable
Gaussian Process Regression	Easy coding and implementation	Require a smooth and monotonic data trend
Markov/Semi-Markov	Well organized method, capable of modeling various system designs and failure modes, capable of managing incomplete data sets, providing confidence limits as part of the RUL prediction	Considering a single monotonic and non-temporal failure degradation trend, cannot model previously unanticipated faults and/or root causes
Hidden Markov/Semi-Markov	Capable of modeling varying stages of degradation. Therefore failure pattern does not require to be monotonic	Require a significant volume of data for training, proportional to the number of hidden states
Kalman Filter	Capable of accommodating incomplete and noisy measurements, Being computationally effective	computationally intensive of variants for non-linear systems, easily divergence of some variants
Particle Filter	Well-suited to perform the task of inference (state estimation) in nonlinear dynamic systems with non-Gaussian sources of uncertainty, and has become the de facto state of the art for real-time uncertainty characterization in the implementation of failure prognostic algorithms	Require a significant sample size to prevent degeneracy problem, can be more computationally intensive than basic Kalman filters
Physics-based prognostics	Require fewer data compared to data-driven methods, providing the most accurate estimates of all modeling options if the physics of models remain consistent across the component	Complexity in developing mathematical models, require detailed and complete knowledge of system behavior, being defect-specific

Figure 18: Summary of Prognostics Techniques [20]

3 Open Source Datasets

In this section, we report 15 Open Source datasets. These datasets have been found online on the web and they are related to SCADA data of real world WTs. As reported by [31], data gathered by WT fall under the concept of 'Big Data'. There are several aspects to consider, according to [31], linked to the following datasets, and they are mainly: Volume (a WT generated 60-100 SCADA data and this accounts for 0.2 GB of raw data), Velocity (the frequency at which data is produced and transmitted), Variety (providing a mix of information that can be for example images, videos), Veracity (data should not have missing values and inconsistencies). The quality of data can depend on different factors such as the purpose of the research, the organization that gathers all these data etc. We consider the following datasets: Energias de Portugal (EDP) Dataset [91], Le Haute Wind Born Dataset (LHWB) [92], Vestas V52 Dataset (VV52) [93], Yalova Dataset [94], Sotavento Dataset [95], Eolos Dataset [96], Inland Wind Farm Dataset 1 (IWFD1) [97], Offshore Wind Farm Dataset 1 (OWFD1) [97], Inland Wind Farm Dataset 2 (IWFD2) [98], Offshore Wind Farm Dataset 2 (OWFD2) [98], Beberibe Dataset [99], GRC Dataset[100], GRC2 Dataset [101] Penmanshiel Dataset [102], Kelmarsh Dataset [103]. The characteristics of these 15 Datasets are reported in the following tables. Table 2 gives a general overview of each dataset.
The Wind Farms considered are located in different countries and made by operational Wind Turbines (except for [97], [98] for which this information is not available). It is very important to observe the type of information reported, in order to evaluate the best dataset for PHM application. In this sense, it can be seen as all the datasets report SCADA (like Temperature of components, RPM, Active Power) and meteorological measurements (like wind speed, wind direction, pressure), but only the EDP dataset [91] contains an historical failures logbook. Furthermore, Table 2 also contains additional information for [91], [92], [102] and [103] like technical data (such as hub height, rotor diameter) and like the status (logs) of Wind Turbines for [91], [102], [103].

Datasets	Format	Access	Provider	Information reported	Country	Operative
EDP [91]	.csv	Open	EDP Inovaçao	SCADA, metmast, historical failures logbook, logs	West African Gulf of Guinea	Yes
LHWB [92]	.csv	Open	Engie France	Static, SCADA	France	Yes
VV52 [93]	.xlsx	Open	Dundalk Institute of Technology	SCADA	Ireland	Yes
Yalova [94]	.csv	Open	-	SCADA	Turkey	Yes
Sotavento [95]	graphic data	Open	Sotavento Galicia Foundation	SCADA	Spain	Yes
Eolos [96]	.xlsx	Open	Eolos Wind Research Group	SCADA, metmast	USA	Yes
IWFD1 [97]	.csv	Open	Data Science for Wind Energy book	SCADA, metmast	-	-
IWFD2 [98]	.csv	Open	Data Science for Wind Energy book	SCADA, metmast	-	-
OWFD1 [97]	.csv	Open	Data Science for Wind Energy book	SCADA, metmast	-	-
OWFD2 [98]	.csv	Open	Data Science for Wind Energy book	SCADA, metmast	-	-
Beberibe [99]	.nc	Open	Brazilian Electricity Regulatory Agency	SCADA	Brazil	Yes
GRC [100]	.tdsm	Open	NREL-GRC	SCADA, Mechanical	USA	-
GRC2 [101]	.tdsm	Open	NREL-GRC	SCADA, Mechanical	USA	-
Penmanshiel [102]	.csv	Open	Cubico Sustainable Investments Ltd	Static, SCADA, metmast, logs	UK	Yes
Kelmarsh [103]	.csv	Open	Cubico Sustainable Investments Ltd	Static, SCADA, metmast, logs	UK	Yes

SCADA data: Temperature; RPM Active Power, Reactive Power, Control Variables, etc. **Metmast**: Wind Speed, Wind Direction, Ambient Temperature, Ambient Pressure, Ambient Humidity, etc. **Logs**: Status Remarks like external power, Yaw speed, Hub fan, generator, nacelle Temperature, components activeted (in/off) **Mechanical**: stress, strain, displacement, bending, torsion moment **Static**: Hub Height, Rated Power, Rotor diameter, Manufacturer

Table 2: General Information of Open Source datasets for Wind Turbines

The detailed analysis of each dataset, which can support the choice of the most appropriate one, is conducted in Table 3. It can be seen how the number of Wind Turbines varies widely from 1 for [93], [94], [96], [100] and [101] to a maximum of 34 for [95]. The Wind Turbines considered are all Horizontal Axis, which is the most common configuration in the industry, and onshore except for [91], [97], [98]. Important aspects to consider, in order to find the best dataset for PHM application, are: the time span, the sampling rate and the number of parameters. In particular, having larger time span, smaller sampling rate and higher number of parameters means having more data that can be exploited. In this sense, we can see as the time span varies from 1 hour for [96] to 14 years for [92] and the typical sampling rate is 10 min for all except for [96], which presents a sampling rate of 1 min and except for GRC [100] and GRC2 [101] for which this information is not available. In addition, the number of parameters varies from 7 for [97] to a maximum of 303 for [102] and [103].

Datasets	Number of WT	Type of WT	Capacity (MW)	Time Span	Sampling rate (min)	Number of parameters
EDP [91]	5	Offshore	10	01/2017-12/2017	10	133
LHWB [92]	4	Onshore	17.30	01/2013-12/2016 01/2017-12/2020	10	138
VV52 [93]	1	Onshore	8.5	01/2006-03/2020	10	22
Yalova [94]	1	Onshore	53	01/01/2018-13/12/2018	10	5
Sotavento [95]	34	Onshore	18	10/2022	10	3
Eolos [96]	1	Onshore	2.5	19/08/2012 24/08/2012, 29/12/2013, 08/01/2014	1	40
IWFD1 [97]	4	Onshore	-	08/2010-08/2011 04/2010-04/2011	10	7
IWFD2 [98]	2	Onshore	42	01/2008-10/2011	10	8
OWFD1 [97]	2	Offshore	-	01/2009-12/2009	10	9
OWFD2 [98]	2	Offshore	-	01/2007-12/2010	10	8
Beberibe [99]	32	Onshore	-	09/2013-08/2014	10	43
GRC [100]	1	-	0.75	12/2013-01/2015	-	165
GRC2 [101]	1	-	0.75	09/2016-12/2016	-	190
Penmanshiel [102]	14	Onshore	28.7	07/2016-07/2021	10	303
Kelmarsh [103]	6	Onshore	12	07/2016-07/2021	10	303

Table 3: Detailed Information of Open Source datasets for Wind Turbines

The analysis about which are the components and the parameters for each dataset is reported in Table 6. [91],[92],[93],[102],[103] consider measurements associated to a wide range of components, but the most monitored are gearbox, generator with their sub-assemblies (like bearing, stator) and rotor because are those most critical for Wind Turbines. [94],[95],[96],[97],[98] contain only power production variables, and for this reason they are not useful for the aim of this work.
The parameters shown in Table 4 have been divided into 4 categories: meteo, production, signals and additional.

Datasets	Components	Type of Data	Main parameters
EDP [91]		Meteo	WS(*), WD(*),T$_a(*)$, $P_a(*)$, $H_a(*)$, $Prcp(*)$
		Production	ActP(4), Q(4)
	Gearbox	Monitoring Signals	T_{oilAVG}
	Gearbox bearing		T_{AVG}
	Generator		RPM(*), ActP(4),Q(4),$T_{ringchamber}$
	Generator Bearing		T_{AVG}
	Generator stator		T_{iAVG}
	Transformer		T_{iAVG}
	Grid		T_{AVG}(4),P, Cosphi, f, T_{busbar}, ActP(*), Q(*)
	Rotor		RPM(*)
	Blades		PA(*)
	Nacelle		T_{AVG}, Dir
	Controller		$T_{nacelleAVG}$, T_{hubAVG},$T_{vcpboardAVG}$,$T_{vcssectionAVG}$
	Spinner		T_{AVG}
	Hydraulic group		T_{oilAVG}
		Additional	Nu,Nc,PLAP,PGAP,PGIR,GCRP,RD,AF,AO,ACO,ACG, DAP,PAF
LHWB [92]		Meteo	WS, WD,T_a
		Production	ActP, Q
	Gearbox	Monitoring Signals	T_{oil}, T_{inlet}
	Gearbox bearing		T
	Generator		RPM
	Generator converter		RPM
	Generator bearing		T
	Generator stator		T
	Converter		Tq
	Nacelle		Dir, T, A
	Grid		f
	Rotor		RPM
	Rotor Bearing		T
	Hub		T
	Blades		PA
		Additional	Cosphi, Nu, Tq, S, Va
Yalova [94]		Meteo	WS, WD
		Production	ActP
	-	Monitoring Signals	-
		Additional	PC
Sotavento [95]		Meteo	WS, WD
		Production	En
	-	Monitoring Signals	-
		Additional	-
Eolos [96]		Meteo	WS(11), WD(11), , T_a(6), H(6),P_a
		Production	-
	-	Monitoring Signals	-
		Additional	RP

Table 4: Components and Main Parameters

Datasets	Components	Type of Data	Main parameters
IWFD1-OWFD1[97]		Meteo	WS, WD, D
		Production	y
	-	Monitoring Signals	-
		Additional	TI, WSh
IWFD2-OWFD2[98]		Meteo	WS, WD, D
		Production	Y
	-	Monitoring Signals	-
		Additional	TI, WSh
Beberibe [99]		Meteo	WS(6*), WD(2), T_a, H, P_a, D
		Production	ActP(*),
	Rotor	Monitoring Signals	RPM(*)
		Additional	T_{logger}, O(16)
GRC [100]		Meteo	-
		Production	P(2), Q(2)
	Gearbox	Monitoring Signals	Tq(5), T(19), RPM(5), p_{oil}
		Additional	ϵ(72), F(3), Spt(24), Mb(16), O(17)
GRC2 [101]		Meteo	-
		Production	P(2), Q(2)
	Gearbox	Monitoring Signals	Tq(6), T(25), RPM(7), p_{oil}
		Additional	ϵ(87), F(3), Spt(23), Mb(16), O(18)
VV52 [93]		Meteo	WS(2), WD(2), , T_a
		Production	P(*), Q_{oil}
	Gearbox	Monitoring Signals	T_{oil}
	Gearbox Bearings		T
	Generator		RPM, T
	Generator Bearing		T
	Generator Winding		T_i
	Nacelle		T
	Blades		PA
		Additional	-
Penmanshiel [102]		Meteo	WS(*), WD(*),$T_a(*)$, $P_a(*)$, $H_a(*)$, $Prcp(*)$
		Production	P(*), ActP(*), Q(*)
	Gearbox	Monitoring Signals	T_{oilAVG},RPM(*)
	Gearbox Bearing		T_{AVG}
	Front Bearing		T(*)
	Rear Bearing		T(*)
	Stator		T(*)
	Generator		RPM(*)
	Generator Bearing		$T_{rear}(*)$, $T_{front}(*)$
	Transformer		T(*)
	Gear		$T_{oilInlet}$, $T_{oil}(*)$, $P_{oilInlet}(*)$,$P_{oilPump}(*)$
	Rotor		RPM(*)
	Rotor Bearing		T(*)
	Yaw Bearing		YA(*)
	Blades		PA(3*)
	Nacelle		T, Pos(*)
	Motor axis		T(*)
	Grid		F(*)
	CPU		T(*)
	Hub		T(*)
	Top Box		T(*)
		Additional	Va,EE,EI,LP(20),EB,ET,VP,PPD(11),TPS,ACP(2),Cosphi(*),V(4*),C(7*),CF,DA, TBI(9),PB(5),REI,REE,S(*),EFLH(2),PF,PI, DTA(*),CWCP(*),TA(*),MPC(2)
Kelmarsh [103]		Meteo	WS(*), WD(*),$T_a(*)$, $P_a(*)$, $H_a(*)$, $Prcp(*)$
		Production	P(*), ActP(*), Q(*)
	Gearbox	Monitoring Signals	T_{oilAVG},RPM(*)
	Gearbox Bearing		T_{AVG}
	Front Bearing		T(*)
	Rear Bearing		T(*)
	Stator		T(*)
	Generator		RPM(*)
	Generator Bearing		$T_{rear}(*)$, $T_{front}(*)$
	Transformer		T(*)
	Gear		$T_{oilInlet}$, $T_{oil}(*)$, $P_{oilInlet}(*)$,$P_{oilPump}(*)$
	Rotor		RPM(*)
	Rotor Bearing		T(*)
	Yaw Bearing		YA(*)
	Blades		PA(3*)
	Nacelle		T, Position(*)
	Motor axis		T(*)
	Grid		F(*)
	CPU		T(*)
	Hub		T(*)
	Top Box		T(*)
		Additional	Va,EE,EI,LP(20),EB,ET,VP,PPD(11),TPS,ACP(2),Cosphi(*),V(4*),C(7*),CF,DA, TBI(9),PB(5),REI,REE,S(*),EFLH(2),PF,PI, DTA(*),CWCP(*),TA(*),MPC(2)

The number reported into the round brackets () represents the number of parameters related to the considered quantity.
T= Temperature (°C), **ActP**= Active Power (kW), **Q** = Reactive Power(kVAr), **Va**=Vane Position (°), **S**=Apparent Power (kVA), **WS**= Wind speed(m/s), **WD**= Wind direction (°), **P** = Power (kW), **F**=frequency (Hz), **Cosphi**= Actual phase displacement, **C**=Current (A), **i**=phase1, phase2, phase3, **(*)**=min,max,AVG, **Pa**=pressure of air (Pa), **Ta**= ambient temperature (°C), **H** = humidity (%), **V**=Voltage (V), **EE**= Energy Export (kWh), **EI**=Energy Import (kWh), **LP** =Lost Production (kWh), **EB**=Energy Budget (kWh), **PF**= Production Factor, **TPS**= Turbine Power Setpoint (kW), **ACP**= Available Capacity for Production (kWh) , **PA**= Pitch Angle (°), **ET**= Energy Theoretical (kWh), **VP**= Virtual Production (kW), **PPD**= Potential Power Default (kW), **CF**= Capacity Factor, **DA**= Data Availability, **TBI**= Time Based Information, **PBI**= Production Based Information, **REI**= Reactive Energy Import (kVarh), **REE**= Reactive Energy Export (kVarh), **EFLH**= Equivalent Full Load Hours (s), **PI**= Performance Index, **DFA**= Drive Train Acceleration (mm/ss), **CWCP**= Cable Windings from Calibration Point, **TA**= Tower Acceleration (mm/ss), **MPC**= Metal Particle Count, **Nu**= Grid Voltage (V), **Tq**= Torque (Nm), **A**= Angle (°), **Dir**= Direction (°), **Nc**=Grig Current (A), **PLAP**= Power Latest Average Period (kW), **En**= Energy (kWh), **PGIR**= Possible Grid inductive ReactPwr (kVAr), **GCRP**= Grid Capacitive React Power (kVAr), **AO**=Anemometer offset (m/s), **RD**= Rain Detection, **ACO**= Anemometer correct offset, **ACG**= Anemometer CorrGain, **PAF**=Pressure air frequency (Hz), **DAP**= distance air pressure, **PC**= Power Curve, **RP**=Real Power (kW), **TI**= Turbulance Intensity, **D**= density, **Y**=% relative to rated power, **YA**= Yaw Angles (°), **Prcp**= Precipitation (mm), **AF**=Anemometer frequency (Hz), **PGAP**= Possible Grid Active Power (kW), **Pos**= Position (°), **O**= Others

Table 4: Components and Main Parameters (Continued)

Each dataset, except for [100] and [101], which contains benchmark measurements, reports meteorological information, and in particular, the focus is on the wind speed because strictly connected to the power generation. The production related information are mainly Reactive Power and Active Power which, correlated to the wind speed, allows to plot the power curve of the Wind Turbine to monitor its correct behaviour. As SCADA signals associated to different components, the most reported variables, which can be considered important systems health indicators are: Temperature, Speed of the rotational components (like generator), Control Variables like pitch angle. In Table 4, some additional parameters, which vary according to the provider, have been reported in order to give a complete overview of what each dataset contains.

3.1 Research on Open Source Datasets

In this section, we focus on the research that has been done for the datasets considered in the previous section. In particular, research only related to [91] and [92] has been found and for this reason two tables have been built, each one for each dataset. In Table 5, 9 research papers about EDP dataset [91] are reported. In particular, in this table we have investigated data driven approaches. The components on which the research has focused are almost all the same, mainly gearbox and generator (with their sub-assemblies) because, as explained in the previous section, are those with the highest failure rate. For this reason, they are object of many research in order to develop approaches to improve their reliability. In Table 5, a brief description of what is done in research is reported with a focus on the main aspect of models for faults detection and diagnosis. The range of methods considered in this literature review is quite wide, but we can see that the most employed are supervised learning models. The last column of the table reports the performance evaluation of the model introduced, but not for all the research has been possible to find the evaluation metrics. Many of them, to evaluate the model, consider a graphical representation of the models' output compared with the actual value and any significant deviation testifies an abnormal behaviour that must be investigated in more detail. Table 6 reports 3 research papers related to [92]. The Table structure is the same of Table 5. In fact, we can see the component object of research, a brief description of the work with the models developed and, lastly, a performance evaluation. The performance metrics, introduced in both tables, to evaluate the capability in faults detection of these models, are listed below:

- RMSE= Root Mean Squared Error $=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_p^i - y_{true}^i)^2}$ where y_p= forecasting values, y_{true}=true values
- MAPE= Mean Absolute Percentage Error(%)$=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_p^i - y_{true}^i}{y_{true}^i}\right| * 100\%$
- R^2= Coefficient of Determination $=1 - \frac{\sum_{i=1}^{n}(y_p^i - y_{true}^i)^2}{\sum_{i=1}^{n}(y_{mean}^i - y_{true}^i)^2}$ where y_{mean}= mean of all true values
- MIR= Missed Isolation Rate= $\frac{FNor}{FNor + TIso}$ where FNor= number of faulty variables regarded as normal, TIso= number of faulty variables isolated as faulty
- MSE=Mean Square error=$\frac{1}{n}\sum_{i=1}^{n}(y_p^i - y_a^i)^2$ where y_p= predicted values, y_a= actual value
- MAE= Mean Absolute Error $=\frac{1}{n}\sum_{i=1}^{n}\left|y_p^i - y_{true}^i\right|$;
- SDAPE= Standard Deviation of APE=$\sqrt{\frac{1}{n_i - 1}\sum_{i=1}^{n_i}(APE - MAPE)}$
- Acc= Accuracy=$\frac{TP + TN}{TP + FP + TN + FN}$, where TP= True Positive, FP= False Positive, TN= True Negative, FN= False Negative
- Rec= Recall= $\frac{TP}{TP + FN}$

- Prec= Precision= $\frac{TP}{TP+FP}$

3.1.1 EDP Dataset

In this section, the focus is on the EDP dataset [91], that is the most used and counts 9 research papers. In particular, Data Driven approaches have been investigated in order to perform faults identification (FDI), diagnosis, root causes analysis and RUL prediction.

Paper	Application	Objective	Method	Performance Evaluation
[104]	Transformer, Generator,HG	Constructing a Health Indicator(HI), based on the Temperature (T), and using a predefined threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty (HI>Th)	Hotelling's T^2	-
			TEMI	MAR=0.05,MIR=0.12,FIR=0.02,FAR≤0.01
			PCA-CP	MAR=0.57,MIR=0.71,FIR=0.03,FAR=0.06
			MRBD	MAR=0.78,MIR=0.86,FIR≤0.02,FAR≤0.01
[105]	Transformer	Using ML approach to classify whether a component is faulty or not.	Soft SVM, NSGA-II	Pr=1, Rec=0.83
	Gearbox			Pr=0.75, Rec=0.75
	Generator			Pr=0.86, Rec=0.75
[106]	Wind Turbine	Constructing a Health Indicator (HI), based on a RE, and using a pre-defined threshold (0.5<Th<0.6) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th)	DAE	Graphical representation of the RE normalised
	Generator bearing	Constructing a Health Indicator(HI) based on Temperature (T). Predict the future values (y_p) of HI and compare with actual value (y_{true})	RAM-CNN	RMSE=2.345°C,MAPE=2.519%,R^2=0.961
			CNN	RMSE=4.025°C,MAPE=6.893%,R^2=0.822
			Seq2Seq	RMSE=2.971°C,MAPE=2.841%,R^2=0.956
			LSTM	RMSE=3.706°C,MAPE=6.428%,R^2=0.876
			WFSM	RMSE=6.747 ,MAPE=12.400%,R^2=0.679
	HVT		RAM-CNN	RMSE=1.390 °C,MAPE=1.767%,R^2=0.964
			CNN	RMSE=3.094°C,MAPE=5.874%,R^2=0.889
			Seq2Seq	RMSE=0.956°C,MAPE=1.482%,R^2=0.971
			LSTM	RMSE=2.532 °C,MAPE=3.472 %,R^2=0.927
			WFSM	RMSE=4.312 °C,MAPE=6.809 %,R^2=0.655
[107]	Wind Turbine	Constructing a Health Indicator (HI) based on a RE, interpreted in order to perform a RCA	AE+ARCANA	AE Train: MSE=0.097, AE Val: MSE=0.010, ARCANA importance= 80%
[108]	Gearbox Generator Generator bearing HG Transformer	Comparison of 6 approaches' capability to detect faults. Constructing a Health Indicator(HI) for each model. Predict the future values (y_p) of HI and compare with actual value (y_{true})	NBM	Graph. visualization of prediction error and anomaly detection period
			LoMST-CUSUM	
			WHC-LOF	
			NBM-LI	
			CCA	
			KCPD	-
[109]	Generator bearing	Constructing predictors combined to detect anomalies and failures in advance	CPPS+ MLP	Graph. visualization of anomalous period
[110]	Wind Turbine	Constructing 2 HIs based on Temperature (T), and using these HIs the PP degradation over time is estimated	Distance Index	Graphical evaluation of power production
			Regression Model	
[111]	Gearbox	Constructing a Health Indicator(HI), based on the Cumulative Score, and using a predefined threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty (HI>Th)	CUSUM	Graph. representation of anomaly score and cumulative anomaly score
[112]	Wind Speed Sensor	Using ML algorithms to detect failures and to improve RUL prediction	DT	Acc=96.77%

HVT= High Voltage Transformer, **HG**=Hydraulic Group, **NSGA-II**= Non-dominated Sorting Genetic Algorithm II, **RAM-CNN**= residual attention module convolution neural network, **DAE**= Denoising Autoencoder, **XGB**= Extreme Gradient Boost, **CNN**= Convolution Neural Network, **WFSM**= Without features selection models, **NN**= Neural network, **LSTM**=Long short-term memory, **Seq2Seq**=sequence to sequence, **MAR**= missed alarm rate, **NBM**=Normal Behaviour Models, **LoMST-CUSUM**= Combined Local Minimum Spanning Tree and Cumulative Sum of Multivariate Time Series Data, **WHC-LOF**=Combined Ward Hierarchical Clustering and Novelty Detection with Local Outlier Factor, **NBM-LI**= Normal Behaviour Model with Lagged Inputs, **CCA**= Canonical Correlation Analysis, **KCPD**= Kernel Change-Point Detection, **CUSUM**= Cumulative Sum of Multivariate Time Series Data, **MLP**= Multilayer Perceptron, **TEMI**=Temperature-based Monitoring and Isolation,**PCA-CP**= principal Component analysis and contribution plot, **MRBD**= minimum risk Bayesian decision. **CPPS**= Combined Power Predictive Scores, **RE**= Reconstruction Error, **PP**= Power Production, **AE**= Autoencoder

Table 5: EDP related work

In Table 5, a brief description of the main goal of each study is reported with a focus on the principal aspects of models introduced for the FDI and the diagnosis. Most of the research aims to construct an appropriate health indicator, which is able to well represent the behavior of the system (mainly Temperature or Reconstruction error), that is compared with a threshold in order to diagnose when the faulty condition occurs.
It can be seen in Table 5, how the EDP dataset [91] has been object of different publications where the range of methods considered is really wide. The most monitored components are those most critical for failures which means gearbox, generator and transformer. The last column of the table reports the performance evaluation of the algorithms, but not for all has been possible to find the evaluation metrics.
Many of these works propose a graphical representation to evaluate model performance, where the model output is compared with the actual value to observe possible deviations which testify a faulty behaviour that must be investigated in further detail.
We can see in Table 5, if we want to evaluate the algorithms performance in terms of accuracy for systems behaviour forecasting, as the CNN-RAM and Seq2Seq show the best performances [106]. In terms of Total Savings Prediction, [108] shows that higher cost savings are provided by LoMST-CUSUM and WHC-LOF.

In the end, [104] introduces a novel TEMI approach which has, in terms of alarms and isolation rate, better performance than traditional techniques like PCA-CP and MRBD.
Given the availability for the EDP dataset [91] of information related to the faulty events that affected the Wind Farms, the tasks of Root Cause analysis and RUL prediction have been tackled. For each of these topics, only 1 research publication has been found, i.e. [107] and [112]. For the Root Causes Analysis, [107] introduces a novel approach that relies on an optimization algorithm that aims to find which input features contributed to the anomalies detected by an Autoencoder model. Such model is able to associate the importance of each feature considered for the Root Causes Analysis. In this sense, the feature related to the wind speed sensor, which is the faulty components, has an importance of 80% (much higher than all the other features, which reach the 10% of importance regarding the fault).
Ref. [112] shows how a data-centric approach overperforms, in terms of RUL prediction, a model-centric approach.

3.1.2 LHWB Dataset

Table 6 reports 3 research papers related to the second Open Source dataset LHWB [92].

Paper	Application	Objective	Method	Performance Evaluation
[30]	Generator bearing R80736	Constructing a Health Indicator (HI), represented by the deviations of the forecast values from the real values, and using a pre-defined threshold (Th=+3σ) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th)	XGB	R^2=0.877,RMSE=1.419,MAE=1.000,MAPE=0.025
			LSTM	R^2=0.880,RMSE=1.446,MAE=1.034,MAPE=0.026
			MLR	R^2=0.850,RMSE=1.616,MAE=1.180,MAPE=0.305
	Generator bearing R80790		XGB	R^2=0.815,RMSE=1.713,MAE=1.105,MAPE=0.027
			LSTM	R^2=0.816,RMSE=1.708,MAE=1.051,MAPE=0.026
			MLR	R^2=0.793,RMSE=1.813,MAE=1.248,MAPE=0.031
	Generator bearing R80721		XGB	R^2=0.936,RMSE=1.060,MAE=0.801,MAPE=0.020
			LSTM	R^2=0.921,RMSE=1.172,MAE=0.938,MAPE=0.024
			MLR	R^2=0.897,RMSE=1.357,MAE=1.052,MAPE=0.027
	Generator bearing R80711		XGB	R^2=0.926,RMSE=1.094,MAE=0.797,MAPE=0.020
			LSTM	R^2=0.928,RMSE=1.074,MAE=0.782,MAPE=0.020
			MLR	R^2=0.879,RMSE=1.396,MAE=1.092,MAPE=0.028
	Gearbox bearing R80736		XGB	R^2=0.967,RMSE=0.975,MAE=0.707,MAPE=0.011
			LSTM	R^2=0.950,RMSE=1.204,MAE=0.951,MAPE=0.016
			MLR	R^2=0.954,RMSE=1.163,MAE=0.845,MAPE=0.014
	Gearbox bearing R80790		XGB	R^2=0.956,RMSE=1.086,MAE=0.848,MAPE=0.014
			LSTM	R^2=0.954,RMSE=1.110,MAE=0.833,MAPE=0.014
			MLR	R^2=0.934,RMSE=1.329,MAE=1.012,MAPE=0.017
	Gearbox bearing R80721		XGB	R^2=0.949,RMSE=1.125,MAE=0.775,MAPE=0.013
			LSTM	R^2=0.941,RMSE=1.207,MAE=0.847,MAPE=0.014
			MLR	R^2=0.929,RMSE=1.324,MAE=0.904,MAPE=0.015
	Gearbox bearing R80711		XGB	R^2=0.9569,RMSE=1.092,MAE=0.809,MAPE=0.014
			LSTM	R^2=0.9567,RMSE=1.094,MAE=0.804,MAPE=0.014
			MLR	R^2=0.938,RMSE=1.305,MAE=0.944,MAPE=0.016
[113]	Gearbox	Comparison of 6 different ML models evaluated constructing a Health Indicator y_p, based on the Temperature (T), compared with the actual value y_{true} of T, in order to evaluate the deviations of the forecast T from the real values to understand when an overheating fault occurs	RF	Train:R^2=99.6%, MSE=0.39,MAE=0.44
				Val:R^2=99.1%, MSE=0.42,MAE=0.45
			kNN	Train:R^2=82.05%, MSE=22.54,MAE=2.92
				Val: R^2=84.14%, MSE=19.54,MAE=3.12
			ADA	Train:R^2=98.2%, MSE=2.25,MAE=1.12
				Val:R^2=94.22%, MSE=5.25,MAE=2.12
			MLP	Train:R^2=73.31%, MSE=41.1,MAE=4.96
				Val:R^2=86.42%, MSE=39.23,MAE=7.63
			BA	Train:R^2=98.5%, MSE=0.33,MAE=0.44
				Val:R^2=99.4%, MSE=0.36,MAE=0.47
			ETR	Train:R^2=43.01%, MSE=0.86,MAE=0.79
				Val:R^2=48.2%, MSE=0.93,MAE=0.91
[114]	Wind Turbine	Constructing Health Indicators to define the status and to identify the attributes contributing the abnormal status	GMM, IF	NOA detected: 306

XGB= Extreme Gradient Boost, **LSTM**= Long Short term Memory, **MLR**= Multiple Linear Regression, σ= deviation stardard, **NOA**= Number of Anomalies, **Train**= Training, **Val**= Validation, **WT**= Wind Turbine, **RF**=Random Forest, **IF**= Isolate Forest, **GMM**= Gaussian Mixture Model, **kNN**=k-Nearest Neighbors, **ADA**=Adaboost regressor, **MLP**=Multi Layer Perceptron, **BA**= Bagging regressor, **ETR**= Extra tree regressor

Table 6: La Haute Wind Born Datset [92]

The table structure is the same as Table 5, where it can be seen the component object of research, a brief description of the work with the models developed and a performance evaluation. The approaches developed to detect failures are quite wide. Typically, the goal is to create an health indicator, mainly for generator and gearbox, and comparing it against a threshold value.
Within the ML algorithms introduced, the generator behaviour is very well described considering a LSTM approach; while gearbox is well described considering a bagging model with a MSE=0.33.

The research considering the LHWB [92] is quite limited and it has been object only of diagnosis because it lacks the information related to the faults and for this reason, the root causes analysis and the prognosis haven't been performed.

3.2 Research on Diagnosis and Prognostics for Not Open Source Datasets

In this section, we review research publications for diagnosis and prognosis application, considering Not Open Source Datasets, and discussing the objective, methods and performance. We make a distinction between Faults Detection, Diagnosis and Prognostics and RUL prediction. Firstly, considering the fault detection and the diagnosis, 3 Tables have been considered divided into: Signals Based, Model Based and Data Driven. In each table, it is reported the type of data considered. In this sense, we can see as the most used data for almost all the research are SCADA data, Vibrations and CM data. The components object of the analysis are mainly gearbox and generator. All the works are briefly described in order to understand the flowchart followed by the authors for the aim of their research. Lastly, the method developed and its performance is reported. In Table 7, 4 papers related to the Signals Based Diagnosis are considered and each one, in order to detect an anomaly, analyses the frequency spectrum of the signals gathered by sensors. In Table 8, 9 papers related to the Model Based Diagnosis are reported. The modelling approaches of the system considered are very different and at the end these works focused, mainly, on the prediction of a given quantity that is then compared with a threshold value. The performance metrics, in this table, are quite various and the indices not already explained in the previous section, are reported at the bottom of the table. Table 9 is focused on Data Driven works and in detail, it counts 25 papers. The approaches considered are supervised and unsupervised learning models and for each work it is reported the performance evaluation. For the Prognosis and RUL prediction Table 10, Table 11 and Table 12 are considered in order to report 26 research papers related to: Physics Based (7), Life expectancy(4) and Data Driven Prognostics Approach(15). To evaluate the confidence level of the prognosis, some metrics are introduced: prognosis horizon, $\alpha - \lambda$ performance, relatively accuracy. Many of these approaches rely their capability in predicting RUL or faults on the definition of health indices which are compared with a threshold in order to identify the healthy or faulty condition.

3.2.1 Signals Based Diagnosis

This section shows some works related to the Signal based diagnosis of Wind Turbine.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[44]	Vibrations	Gearbox	Using the frequency spectrum analysis to detect faults	Analytical+ Graphical Analysis	Analysis of severity factors, defect matrix and evaluation of graphical trends
[45]	Vibrations	HSS	Constructing a Health indicator (HI) to evaluate the damages	SBPF	for early stages of tooth wear Average SBPF=100%, for missing tooth SBPF=320%
[46]	ESs	Generator, DrTr	Using a novel technique based on the signals analysis to detect faults	DWT + CWT	Graph. evaluation
[47]	Vibrations + SCADA	Generator	Using characteristic frequencies to detect faults	WA	Graph. evaluation of the faulty conditions

VB2SFD= Vibration based two stage fault detection, **SBPF**= Side-Band Power Factor Algorithm **HSS**= high speed shaft, **DWT**= Discrete Wavelet Transforms , **CWT**= Continuous Wavelet Transforms , **WA**= Wave Analysis, **DrTr**= Drive Train, **ES**= Electrical Signals, **SBPF**=side band power factor=$\frac{SBPF_f - SBPF_h}{SBPf_h}$, **WA**= Wave Analysis

Table 7: Signals Based Fault Detection and Diagnosis

The type data used in the different approaches developed are mainly vibrations and all the techniques are based on signal processing to extract the main information needed. As we can see in Table 7, the evaluation of the performance of the different methods is made by graphs interpretation.

3.2.2 Model Based Diagnosis

In this section, our attention is focused on all those research related to model based diagnosis. In particular, 9 papers have been found.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[35]	Real SCADA	Gen. Bear.	Constructing a Health Indicator y_p based on the Temperature (T) compared with the actual value y_{true} of T.	LS+ARX with 3 param	R^2=0.8018, AIC=-0.6971, Graph. distribution analysis of the model error
[36]	Real SCADA	Gen.	Using faults sensitive features for faults detection	DM + NOFRF	For FD Graph. analysis of indices I_0^{101}, I_0^{102}, I_1^{101} and I_1^{102}; for prognosis AUC=0.9982, TPR=0.8501, FPR=0.0104
[37]	Real SCADA	TS, CV, IV, RT	Using health variables and extensive expert knowledge on specific components failures enables to determine faults root causes	ADM	N.A.
[38]	Real SCADA	Gb Bear, Gen. Wind.	Investigating different techniques for CM systems. Constructing a Health Indicator (HI) and using an adaptive threshold (Th) to establish whether the status of WT is healthy (H≤ Th) or faulty (HI>Th).	LM	Gear: R^2=0.710, Gen: R^2=0.833
				MSDP	Gear: R^2=0.997, Gen: R^2=0.983
				ANN	Gear: R^2=0.992, Gen: R^2=0.977
[39]	Bench. SCADA	Gen., Blades, Act.	Using a parameter varying modelling for state estimation and fault reconstruction.	PVM	Graph. analysis of Temperature and viscous friction parameter
[40]	Bench. CM Data	DrTr, PS, Generator, Tower	Constructing a Health Indicator (HI) and using a threshold (Th) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th).	ARR+IO	Min. Faults: f1=0.997, f2=-0.145, f3=0.9999, f5=1.00035, f9=25
[41]	Simul. CM data	Generator, CV	Comparison of 3 approaches, Constructing a Health Indicator (HI) and using a threshold (Th) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th).	cKalman, DObs,H_∞	Graph. evaluation of the error
[42]	Simul. CM Data	Blades sensor	Constructing a Health Indicator (HI) and significant deviations of the HI predicted from the actual value are used to establish whether the status of WT is healthy or faulty.	KF	Evaluation of the sensitivity of the method considering 0 or 1. 1= test sensitive to the fault, 0= test not sensitive, Graph. visualization of the prediction error
[43]	Real CM data	Blades sensor, Actuator	Constructing a Health Indicator (HI) and using a threshold (Th) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th).	KF+IM	Isolation capacity 99%, if the generated residual exceeds the threshold false alarm is set to 1

DTDM= Discrete Time Dynamic Model, **LS**= Least Square, **AIC**=Akaike's Information Criterion, **ARX**=AutoRegressive with eXogenous input model,**AUC**= Area Under the Receiver operating characteristic curve, $I_0^{101} = \left|G_0^{103}\right| - \left|G_0^{101}\right|$, $I_0^{102} = \left|G_0^{103}\right| - \left|G_0^{102}\right|$, $I_0^{101}(j\omega_c) = \left|G_0^{103}(j\omega_c)\right| - \left|G_0^{101}(j\omega_c)\right|$, $I_0^{102}(j\omega_c) = \left|G_0^{103}(j\omega_c)\right| - \left|G_0^{102}(j\omega_c)\right|$, **f1**= blade root bending moment sensor fault, **f2**= accelerometer faults, **f3**= generator speed sensor faults, **f5**= generator power sensor faults, **f9**= torque offset fault, **FI**= Fuzzy inference, **CV**= Converter, **TS**= Transformer, **IV**= Inverter, **RT**=Rectifler, **DrTr**=Drive Train, **PS**= Pitch System, **Gen. Wind.**= Generator Windings, **Gb**= Gearbox, **Bear**= Bearing, **Gen**= Generator, **CM**= Condition Monitoring, **Bench**= Benchmark, **Simul**= Simulation **DMS**=Dynamic Model Sensor, **T**= Temperature,**WS**= Wind Speed, **Ta**= Ambient Temperature, **SFA**= System Frequency Analysis, **DM**= Dynamic model, **NOFRF**= nonlinear output frequency response functions, **ADM**= Abductive diagnosis model, **LM**= Linear Model, **MSDP**= Multi State dependent parameter models, **PVM**= Parameter-varying model, **ARR**=analytical redundancy, relations, **IO**= internal observer, **KF**= kalman Filter, **NL**= Non Linear, **cKF**= Cascade Kalman Filter, **DObs**= Dedicated Observer-Based Approach, **IM**= Inference Method

Table 8: Model Based Fault Detection and Diagnosis related work

The most of research builds an Health indicator, which is compared with a threshold, to perform the WT diagnosis. The data gathered by sensors are mainly SCADA data and the several approaches introduced are mainly evaluated graphically. In this sense, it is possible to visually define when a possible fault has occurred.

3.2.3 Data Driven Diagnosis

The current research, considering Data-Driven approaches for Fault Detection (FDI) and diagnosis for Not Open Datasets about Wind Turbines, are presented in Table 9. This Table reports 25 papers briefly described in terms of type of data, components, study goal, method and its performance.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[62]	Real CM data	IGBT	Using ML approach to classify different types of faults	GPR + MCRF	Train: Acc=F1=99.98% Test: Acc=F1=99.94%
[58]	Vibrations	MTC	Using ML approach to classify whether the component state is faulty or not	SVM linear kernel ANN	Acc=98.26% Acc=97.47%
[66]	Real SCADA	CV	Constructing Health indicators (HIs) based on fault indicators to generate radar charts whose analysis allows to detect faults	AOC-ResNet50	Acc=98.04%, Pr=98.41%, Rec=97.66%
[86]	Real SCADA	Generator	Using CL approach to predict heating and excitation faults	CW	Pr=0.24, Rec=0.98, F1=0.38
[53]	Real SCADA	Rear Bearing	Constructing a Health Indicator (HI) and significant deviations of the HI predicted (y_p) from the actual value (y_{true}) are used to establish whether the status is healthy or faulty	ANN	RMSE=0.29
[52]	Real SCADA	Gearbox	Constructing a Health Indicator (HI) based on the Temperature (T) and significant deviations of the HI predicted (y_p) from the actual value (y_{true}) are used to establish whether the status is healthy or faulty	MNN	RMSE=1.18
[115]	Real SCADA	Bearing	Constructing a Health Indicator (HI) and significant deviations of the HI predicted (y_p) from the actual value (y_{true}) are used to establish whether the status is healthy or faulty	ANN	RMSE=0.2
[55]	Real SCADA	Gearbox	Constructing a Health Indicator (HI) and using a predefined threshold (Th) to establish whether the status is healthy (Th≤HI) or faulty (Th>HI)	DNN	MAPE=6.01, SDAPE=4.47
[56]	Real SCADA	Gearbox	Constructing a Health Indicator (HI) based on thermal equilibrium and significant deviations of the HI predicted (y_p) from the actual value (y_{true})are used to establish whether the status is healthy or faulty	NARX-ANN + MHD	MAE=0.44, RMSE=0.77
[60]	SCADA	Generator	Constructing a Health Indicator (HI) and significant deviations of the HI predicted (y_p) from the actual value (y_{true})are used to establish whether the status is healthy or faulty	PCA + SVM	P:NMAE=2±0.1, NRMSE=3.1± 0.1, MAPE=5.5± 1.2 V:NMAE=0.63±0.01, NRMSE=0.8±0.01, MAPE=0.04±0.02 I:NMAE=1.9±0.1, NRMSE=2.9±0.2, MAPE=6.3±1
[80]	Real SCADA	Bearing	Constructing a Health Indicator (HI) based on the Temperature (T) and using a pre-defined threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty(HI>Th)	NN	Train: MAE=0.659, MRE=1.39, R^2=0.9926 Test: MAE=0.663, MRE=1.379, R^2=0.9905
[48]	Real SCADA	HSS	Constructing 2 Health Indicators (HIs) and using a threshold (Th) to establish whether the status of WT is healthy (HI≤Th) or faulty (HI>Th). Investigating classification technique for capturing fault signatures	PCA, AANN, SOFM	Graph. evaluation of Q statistic and Hotelling T^2 statistics
[49]	SCADA + CM	Gearbox	Analysing the physical procedure of kinetic energy transmission and dissipation to detect failures		-
[63]	AE, Vibrations	Gearbox	Using ML approach to classify the faults pattern	DRFF	CL rate= 97.68%
[67]	Vibrations	Gearbox	Using ML approach to classify faults samples	ELMD+ FCM	-
[54]	Real SCADA	Gearbox	Constructing a Health Indicator (HI) based on Temperature (T) and significant deviations of the HI predicted (y_p) from the actual value (y_{true})are used to establish whether the status is healthy or faulty	NN	Graph. evaluation of deviation
[59]	Sim. CM data	DrTr	Using ML approach to classify whether the component state is faulty or not	SBS/SFS-SVM	Acc=97.27%
[57]	Real CM data	Sensors, CVA	Constructing a Health Indicator (HI) based on the residuals between the predicted (y_p) and the actual values (y_{true}) and using a pre-defined threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty(HI>Th)	SVM	-
[61]	SCADA	Generator	Combining engineering knowledge, failure modes and domain knowledge to define faults root causes using a ML approach	DT	-
[64]	Bench. data	Sensors, Actuator	Constructing a Health Indicator (HI) and using a pre-defined Threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty(HI>Th)	RF+XGB	hit rate=0.9999, Graph. evaluation of the deviation
[65]	Real SCADA	Gearbox, Generator bearing	Constructing a Health Indicator (HI) based on RE and using an adaptive Threshold (Th) to establish whether the status is healthy (HI≤Th) or faulty(HI>Th)	DAE	Graph. evaluation of the difference between HI and the threshold
[68]	Real SCADA	Gearbox	Constructing a Health Indicator (HI) and significant deviations of the predicted (y_p) from the actual values (y_{true})are used to establish whether status is healthy (HI≤Th) or faulty(HI>Th)	SOM, GMM, NN	Graph. evaluation of deviation
[69]	Vibrations	Bearing	Using the position of the neurons in the output layer, it's possible to detect a fault.	SOM	Graph. trend analysis of faulty and healthy signals
[82]	Real SCADA	Gearboxbearing, Generator stator	Constructing a Health Indicator (HI) based on the Temperature (T) and significant deviations of the predicted (y_p) from the actual values (y_{true})are used to establish whether status is healthy (HI≤Th) or faulty(HI>Th)	FSRC-NN, AUTONN, RM	AUTONN: ALV=50, Graph. of the prediction error

Table 9: Data Driven Fault Detection and Diagnosis work

[83]	Real SCADA	Wind Turbine	Faults Prediction	NN	Acc=97.6%
				SVM	Acc=95.8%
				RF	Acc=99.4%
				BTA	Acc=98.8%
				CHAID	Acc=96%

HSS= High Speed Shaft, **AANN**=auto-associative neural network, **SOFM**= self-organizing feature map, **CM**= Condtion Monitoring, **NARX**= Non linear Auto-Regressive with exogenous input, **DT**= Decision Tree, **SBS**=sequential backward selection,**SFS**=sequential forward selection, **P**= Power, **I**= Current, **V**= Voltage, **CVA**= Converter Actuator, **AE**= Acoustic Emissions, **MNN**= Multilayer neural network, **MHD**= Mahalanobis Distance, **DNN**= Deep Neural Network, **CL**= Classification, **MTC**= Mechanical Transmission Chain, **DrTr**= Drive Train, **FSRC-NN**= Full Signals Reconstruction Neural Network, **AUTONN**= Autoregressive neural network, **ALV**= alarm limit violation, **GPR**= Gaussian process regression-based, **MCRF**= Multiclasse Random Forest, **DRFF**= Deep Random Forest Fusion, **RE**= Reconstruction Error, **RF**= Random Forest, **CW**= Addition of Class Weight, **DAE**= Deep Auto-encoder network, **AOC**=Attention Octave Convolution, **XGB**= Extreme Gradient Boost, **ELMD**= Ensemble Local Means Decomposition, **FCM**= Fuzzy C-means Clustering **GMM**= Gaussian Mixture Model, **CV**= Converter,**BTA**=boosting tree algorithm , **CHAID**=chi-square automatic interaction detector, **RF**= Random Forest, **RM**= Regression model, **SOM**= Self Organizing Map

Table 9: Data Driven Fault Detection and Diagnosis work (Continued)

To perform a diagnosis, most of the techniques are based on the construction of a health indicator, which is then compared with a threshold value. Table 9 reports the techniques used in order to detect failures. For the most part, supervised learning models are used.
It is evident how different Neural Network models, which differ in terms of performance, are widely adopted for Wind Turbines diagnosis. Such ANN models have similar performances for fault detection, while the best results are provided by the Artificial Neural Network in Ref. [115] with a RMSE=0.2. These models are fed by SCADA, Vibrations and CM data gathered by the sensors inside the Wind Turbines.
Observing the second column of Table 9, it can be seen as the most monitored components are gearbox and generator.

3.2.4 Physics based Prognostics

The physics based prognostics related works are showed in Table 10. The main purpose of each work is to predict RUL.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[73]	SCADA, CM	Wind Turbine	RUL prediction	DP	RUL values tabled.
[74]	CM	Gearbox	RUL Prediction	PF	N=1000: RMSE=2.38, σ=0.12
[75]	CM signals	Bearing	RUL prediction with uncertainty quantification	WT+PF	Graph. trend analysis considering α-λ accuracy
[76]	Real CM data	Wind Turbine	RUL prediction with a stochastic decision making framework	SDGM + SBI	Evaluation in terms of total costs savings
[77]	Vibrations	Gen. bearing	RUL Prediction	PL+KF	HI and a threshold are compared
[72]	Vibrations	HSSB, IBGT	RUL Prediction	PL+KF, MMLE, GPR	Graph. Eval. of HI
[116]	Vibrations	Bearing	RUL Prediction with a robust model	IUPF	Graph. RMS, MAD=3.77

DP=Degradation Pattern,**SDGM**=Stochastic Damage Growth model,**PL**=Paris'Law,**PF**=Particle filter,**HSSB**=High Speed Shaft Bearing,**SBI**=Similarity based interpolation,**WT**=Wavelet transforms,**HI**=Health Index,**GPR**=Gaussian Process Regression,**MMLE**=Modified Maximum Likelihood Estimator,σ=standard deviation, **IUPF**= Improved Unscented Particle Filter, **MAD**=mean absolute deviation=$\sum_{i=1}^{N}\left|\frac{Predicted(RUL)-Actual(Rul)}{k}\right|$ where k= number of prognosis step

Table 10: Physics Based Prognostics Related work

We can see how the range of data used for developing the different methods is quite wide and in particular, the authors, for their models, exploited SCADA, Condition monitoring and Vibrations data.

3.2.5 Life Expectancy Prognostics

The current research, considering Life expectancy prognostics approaches, is reported in Table 11.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[88]		Wind Turbine	Comparison of prognostics approaches	HMM, NN, PF	N.A.
[89]	Vibrations	Gearb. bearing	RUL Prediction	EPF+NFI	e1=0.3645,e2=0.6472, Graph. evaluation of fault predictor
[90]	Real SCADA	Gearbox	RUL prediction	PF	N.A.
[117]	Bench. data	Gearbox	RUL prediction.	Generalized Cauchy	RMSE=1.0274, MAPE=0.0532

HMM=Hidden Markov Model,**EPF**=Enhanced Particle Filter,**BI**=Bayesian Inference, **GP**=Gaussian Process,**NFI**=Neuro Fuzzy Inference,**DSTM**=Dependent State Transition Model,**SDGM**=Stochastic Damage-Growth Model,**ACC**=$\left|t_{E_k^n}-t_n\right|$ where $t_{E_k^n}$=real time,tn=predicted time, e1=$\frac{1}{N_sT_m}\sum_{u=1}^{N_s}\sum_{k=1}^{T_m}|z_k-{}^u_k|$, e2=$\frac{1}{N_sT_p}\sum_{u=1}^{N_s}\sum_{T_m+1}^{T_m+T_p}|z_k-{}^u_k|$ where T_m=monitoring period, N_s=Number of experiments,T_p=prediction period,x_k^u=state predicted at the time instant k for the u^{th} experiments;

Table 11: Life Expectancy Prognostics related work

3.2.6 Data-Driven Prognostics

For the Prognosis and RUL prediction, 15 Data-Driven Prognostics papers are considered.
We can see in Table 12, how most of the research about the prognosis for Wind Turbines relies on Data-driven models fed by condition monitoring measurements and in particular by vibrations. The target of each study for the prognosis is to compute the RUL of the most critical components. In this sense, we can see particular attention towards components like gearbox, generator, and bearing. The method developed to predict the RUL are really wide, but the current research considers mainly supervised ML algorithms and the most performed are different Neural Networks models (NN [126], ANN [87] [81], MANN [118], DNN [122], ENN [125]). If we want to evaluate the performance in terms of RMSE, we can see how [125] provides the lowest errors.

Paper	Type of Data	Application	Objective	Method	Performance Evaluation
[78] [79]	SCADA	Pitch System	Using a ML methods to perform a fault prognosis achieving a high prediction horizon	APK-ANFIS	PH=21 days
[118]	Vibration	Gearbox	RUL Prediction	Regression+ MANN	SSE=661.98
[119]	Vibration	Wind Turbine	RUL Prediction	CDM+ triggering algorithm	Case Study 1: PEP=100% Case Study 2: PEP=85%
[120]	Vibration	HSS bearing	RUL Prediction	OCSVM+GLM	RMSE=16.484, MAPE=42.908%
[121]	Vibration	bearing	RUL Prediction	FDMPD+KELM +WAFT	4.68% <E<458.14%
[122]	SCADA	bearing	Prediction performance of WT's risk condition with respect to RFL	NN+DNN	average accuracy>0.8, average sensitivity>0.85%
[123]	Vibration	bearing	RUL prediction	SOM+UKF	E=47.47%%
[124]	SCADA	generator	RUL prediction	ARIMA	MRE=0.27
[125]	Vibration	HSS bearing	RUL prediction	SSF+ENN	Train: RMSE=1.65e^{-5}, $Test: RMSE = 0.0025$
[126]	Vibration	generator	RUL prediction	NN	RUL=26.6 days, prognostics accuracy=12.78%
[87]	SCADA+ status	Bearing	Achieving a high prediction horizon and the time of failure within a specified model accuracy	BI, GP, DSTM, ANN	Defining of a Prediction horizon where ACC=106 hours
[127]	Vibration	Gearbox	RUL prediction	FOA-ELM	RMSE=0.91h, MAE=0.734h, Acc=95.4%
[81]	Real SCADA	Gear bearing	RUL and Faults Predictions	ANN	CorrPred:72.5%
				SVM	CorrPred:60%
				LR	CorrPred:59%
[128]	Vibration	HSS bearing	RUL prediction	ENN	MSE=0.0023

APK-ANFIS= A priori knowledge Adaptive Neuro-Fuzzy Inference System, **ANN**= Artificial Neural Network, **LR**= Linear Regression, **CorrPred**= Correct predictions, **MANN**= Multilayer Artificial Neural Network,**BI**=Bayesian Inference, **GP**=Gaussian Process,**NFI**=Neuro Fuzzy Inference,**DSTM**=Dependent State Transition Model, **CDM**= Cumulative degradation model, **PEP**=Percentage of Effective Predictions, **OCSVM**=One Class Support Vector Machine , **GLM**= Generalized Linear Model, **FDMPD**=Fitting Curve Derivative Method of Maximum Power Spectrum Density, **KELM**= Kernel Extreme Learning Machine, **WAFT**=weight application to failure times, **RFL**= Remaining Functional Life, **ELM**= Extreme Learning Machine, **FOA**= Fruit Fly of Algorithm, **ENN**= Elman Neural Network, **SOM**= Self Organizing Map, **UKF**= Unscented Kalman Filter, **ARIMA**= Autoregressive Integrated Moving Average, **SSF**=Spectral Shape Factor

Table 12: Data Driven Prognosis Related Work

3.3 Application of Open Source Datasets for PHM and Predictive Maintenance Planning

The type of studies that can be conducted considering the 15 Open source datasets, presented in the previous sections, are reported in Table 13. Their suitability is divided into: Diagnosis, Prognostics, Root Causes Analysis (RCA).

Datasets	Suitability for PHM and predictive maintenance planning	Research done
EDP [91]	Diagnosis, Prognostics and RUL prediction, Root Causes Analysis	Diagnosis ([108] [109] [110] [111][104] [105] [106]), Root Causes Analysis ([107]), RUL ([112])
LHWB [92]	Diagnosis	Diagnosis ([30] [114] [113])
VV52 [93]	Diagnosis	-
Penmanshiel [102]	Diagnosis	-
Kelmarsh [103]	Diagnosis	-
Yalova [94]	-	-
Sotavento [95]	-	-
Eolos [96]	Diagnosis	-
IWFD1 [97]	-	-
IWFD2 [98]	-	-
OWFD1 [97]	-	-
OWFD2 [98]	-	-
Beberibe [99]	Diagnosis	-
GRC [100]	Diagnosis	-
GRC2 [101]	Diagnosis	-

Table 13: Possible application of Open Source Datasets for Wind Turbines

Table 13 is built looking inside each dataset, using the information provided by the previous sections. Most of the Open Source datasets available are suitable for health monitoring and diagnosis ([91],[92],[93],[102],[103], [96], [99]), because they provide all the SCADA measurements for the Wind Turbines components. In addition to those information, only the EDP dataset [91] allows to perform prognosis and RUL prediction, given the availability of information about failures/anomalies like date, components interested, Wind Turbine affected, and type of damage. The remaining datasets ([94],[95],[97],[98]) only provide data related to power production and meteorological parameters and given the lack of quantities gathered by components sensor, it is not possible to perform neither a fault detection and/or a diagnosis analysis.

Observing Table 13, as the only open dataset suitable for all kinds of investigation for PHM and predictive maintenance planning is the EDP [91].
The focus of the research done has been monitoring the normal behaviour of the Wind Farms and diagnosing

the anomalies using a threshold. The most of authors used a pre-defined threshold, in general establishing an upper control limit (UCL). In this sense, the definition of a Health Index and of a threshold is very important for an accurate diagnosis with a low margin of error. Then, it can be seen as the challenge for the diagnosis is defining an appropriate threshold that allows having, in output, an accurate evaluation of the normal behaviour and of any significant deviations from it it is caused by faults.
Given the fact that the prognosis for the EDP [91] has not been widely faced by the current research, another challenge is performing the prognosis (computing the RUL of components most affected by failures) to improve the performance of the entire Wind Farm.
In the end, the possibility to also perform a Root Causes Analysis for the EDP [91] introduces another challenge which is to develop a consistent approach able to provide a good understanding of the causes of a defined anomaly.

3.4 EDP preliminary data analysis

The dataset analysed is EDP Dataset [91]. All the data considered for this section have been gathered between 01/01/2017 and 31/12/2017 with a sampling rate of 10 min (Table 3). It has been verified that for the quantities of interest, linked to the correct behaviour of components, like Temperature, RPM, are not present Missing Values. In order to understand the relationship between all the parameters and to understand which variables are strictly connected to each other (to avoid multicollinearity problems for future prognostics applications), the correlation matrix for the EDP dataset [91] has been represented in Fig. 19.

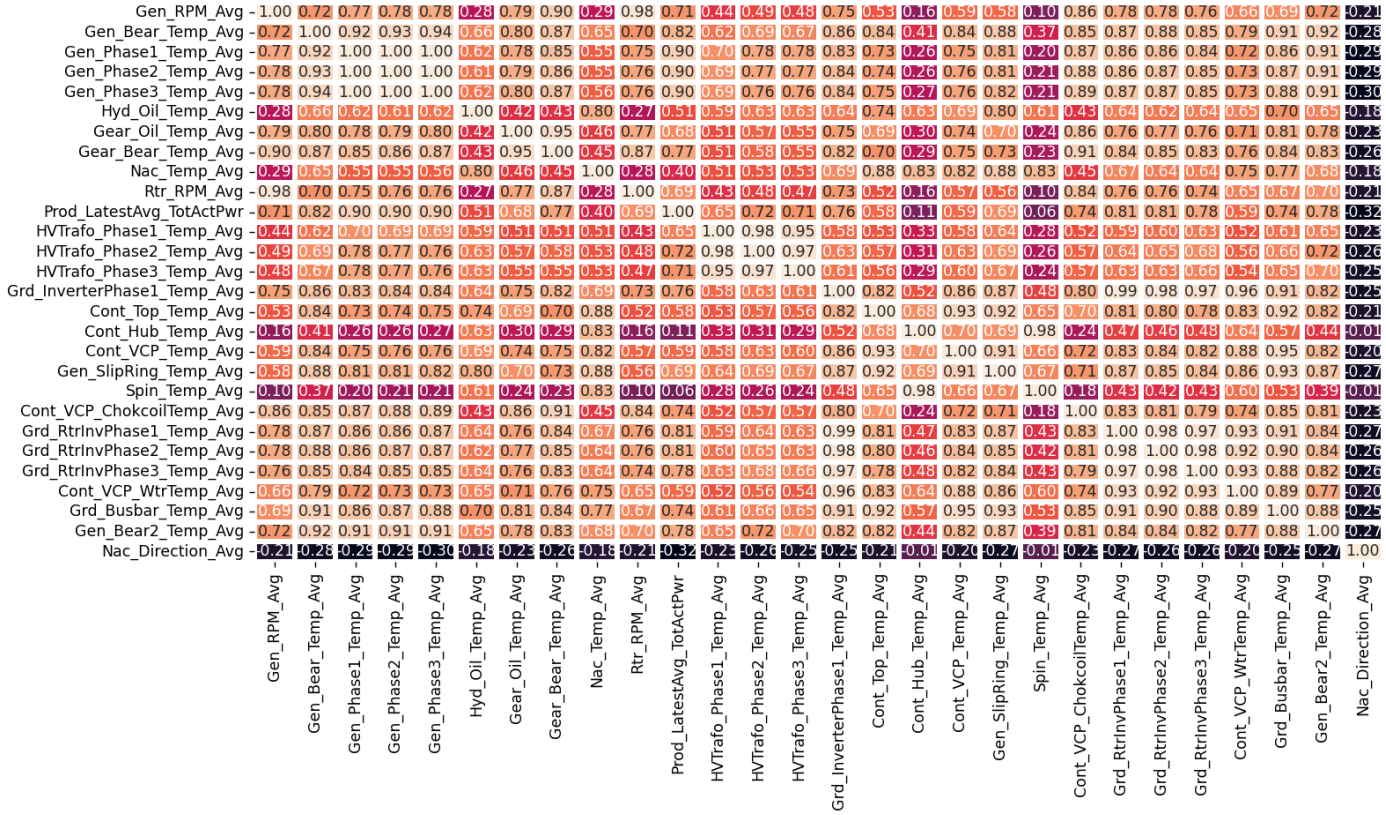


Figure 19: Correlation matrix for EDP Dataset [91]

Fig. 19 shows a strict relationship between the parameters reported on the y axis and the active power production. The same quantities on the y-axis are reported in Fig. 20, in order to visualize how the Temperature of different components increases increasing the power production. In this sense, it is possible to see a sort of linear relationship between the T and the active power.

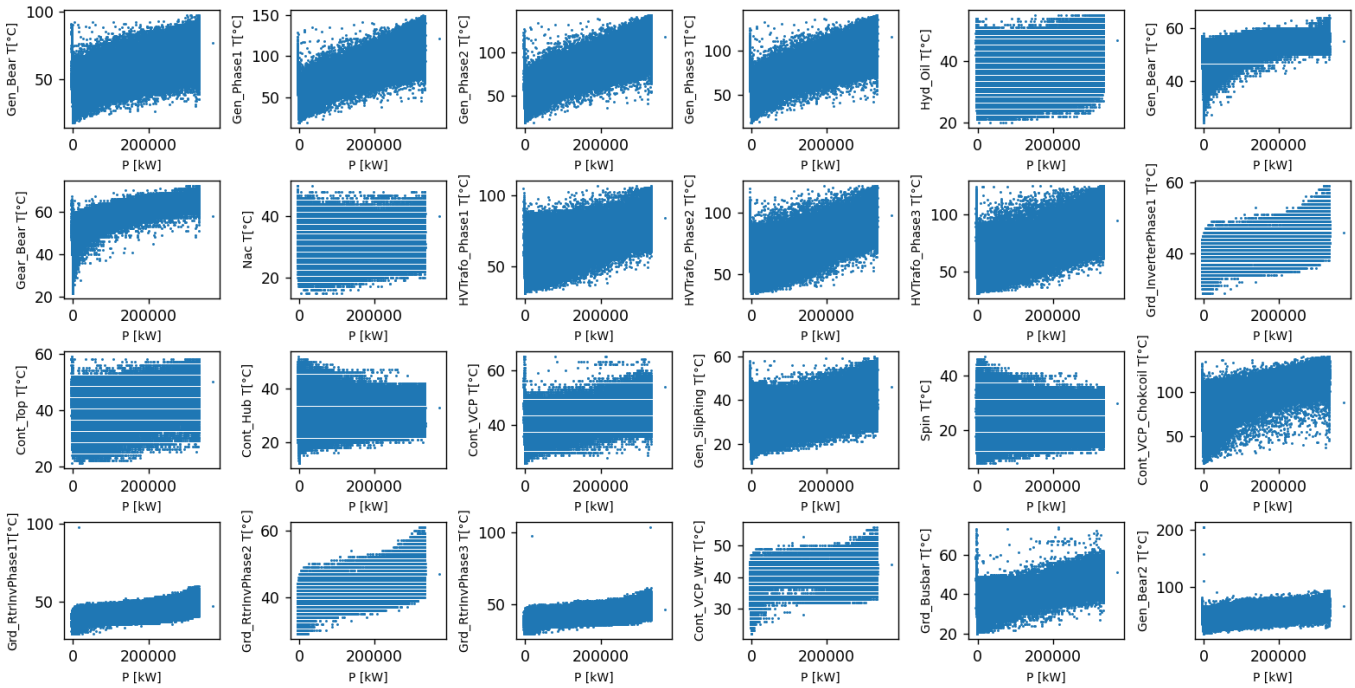


Figure 20: Raw Data for EDP Dataset [91]

A basic analysis of the outliers computed considering the interquartile approach using a box plot representation is reported in Fig. 21. We can see as the number of outliers is quite limited with respect to the number of observations considered which is equal to 209236.

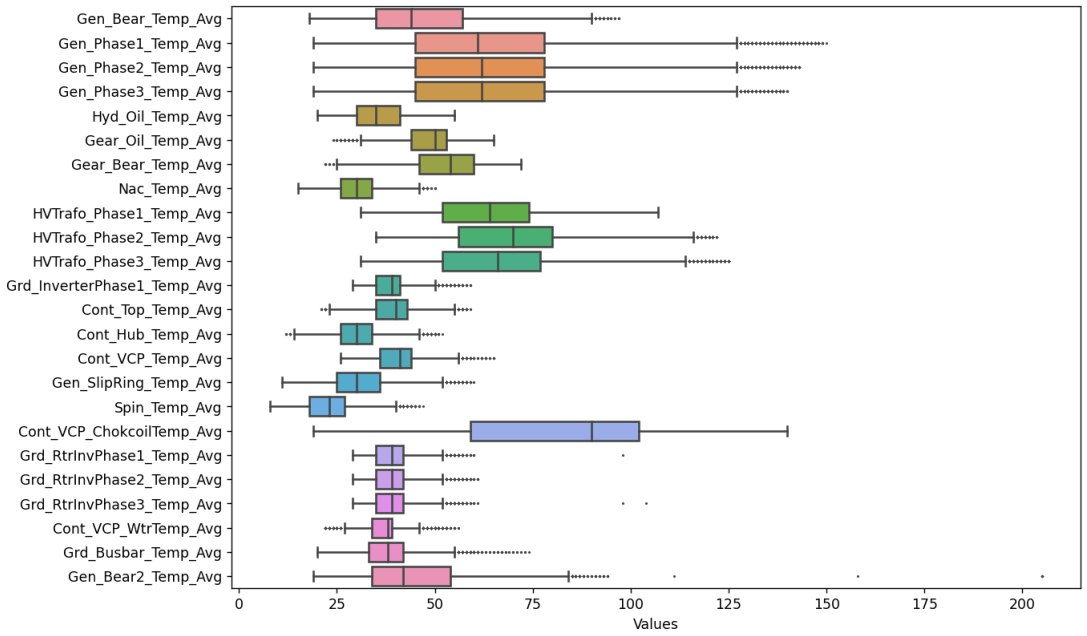


Figure 21: Outliers for EDP Dataset [91]

The attention has been focused on the relationship of various parameters with Active Power Production. Active power is the first basic health indicator for the Wind Turbine, because through the study of the performance and in particular comparing the real power generation with the expected power generation, provided by the manufacturer, it is possible, doing a preliminary analysis on the power curve, to understand if any anomalies occur. On the other hand, the analysis of the wind power curve doesn't help to detect the root of the faults and to understand which are the components involved that show an abnormal behaviour and for this reason it

needs more detailed analysis. The performance of the EDP can be seen in Fig. 22.

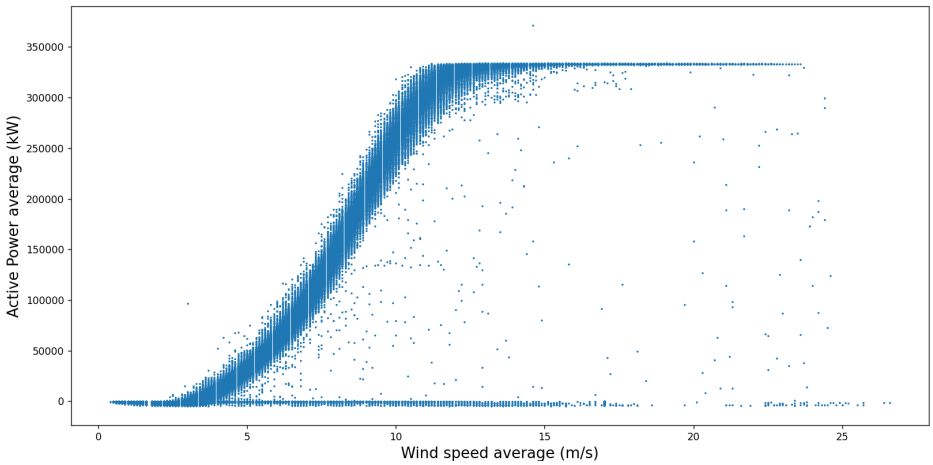


Figure 22: Power Curve for EDP Dataset [91]

From Fig. 22, it can be seen as the power generated varies from 0 MW (when the wind speed is not sufficient to allow power generation) to a maximum value of about 350 MW. The curve is affected by some outliers that should be properly processed for further analysis. Indeed, these outliers in the power generation could significantly affect the accuracy of the result. In this sense, exceeding the wind speed of cut in (v_{cutin}=4 m/s) that indicates the activation of the Wind turbine, we can see many points where the power generated is 0 MW. Still, these points are under power points representing an abnormal wind turbine behavior. The power generated takes a constant value at the cut-out wind speed (v_{cutoff}=25m/s).
In order to complete the preliminary analysis of [91], the information related to the faults, that have affected the correct behaviour of the wind turbines of the wind farms, are shown in Fig. 23, which reports 3 histograms with respectively the number of faults for component, for Wind Turbine and for month.

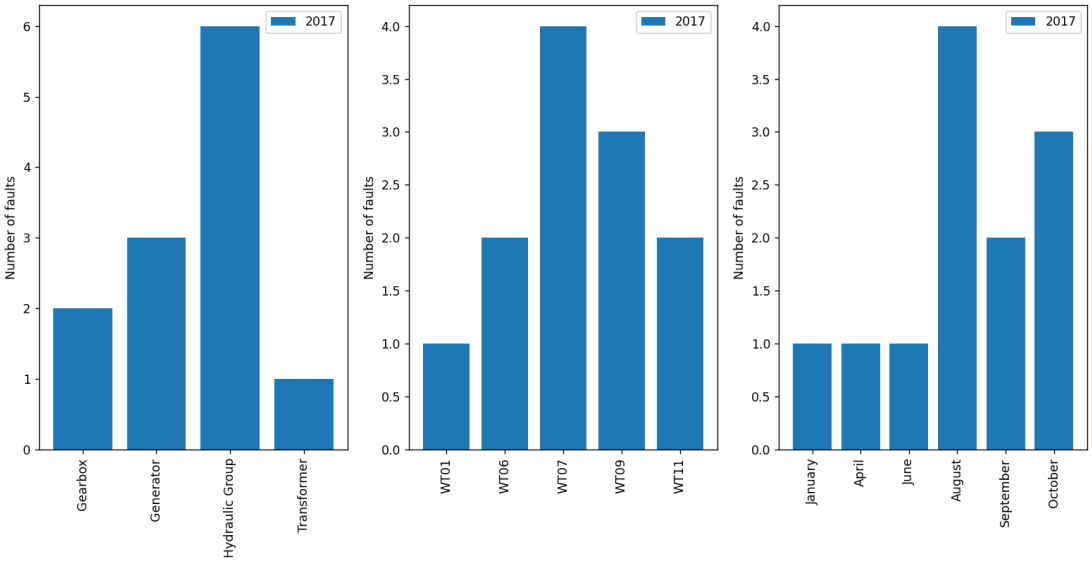


Figure 23: Number of faults

4 EDP Dataset Prognostic

4.1 Theoretical Background

4.1.1 Recurrent Neural Network RNN

A recurrent neural network (RNN), according to [129], is a particular neural network characterised by an internal loop which allows to keep states of earlier input in order to optimize the output. The state of the RNN is reset between processing two different, independent sequences, in order to be able to consider only one sequence a single data point [129]. According to [130], the RNN operates both on an input and internal state space. This state space enables the representation of temporally/sequentially extended dependencies over unspecified (and potentially infinite) intervals. RNNs and their variants have been used in many contexts where the temporal dependency in the data is an important implicit feature in the model design [131].

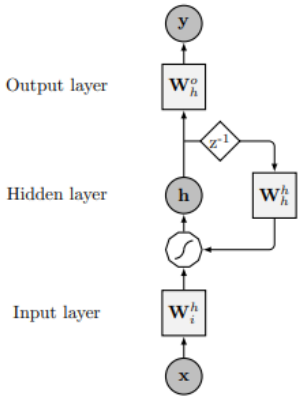


Figure 24: RNN Architecture [131]

Fig. 24 depicts the general RNN architecture. As reported in [131], the circles represent input x, hidden h, and output nodes y, respectively. The solid squares W_i^h, W_h^h and W_o^h are the matrices which represent input, hidden and output weights respectively. Their values are commonly tuned in the training phase through gradient descent. The polygon represents the non-linear transformation performed by neurons and Z^{-1} is the unit delay operator [131]. In particular, the most general form as a RNN can be seen is as a weighted, directed and cyclic graph that contains three different kinds of nodes, namely input, hidden and output nodes [131]. Input nodes do not have incoming connections, output nodes do not have outgoing connections, hidden nodes have both. An edge can connect two different nodes which are at the same or at different time instants.
Bodén [130] introduced the math for a simple RNN model made by 2 layers (one 'hidden' or 'state' layer, and one 'output' layer) nodes excluding the input layer (see Fig. 25).

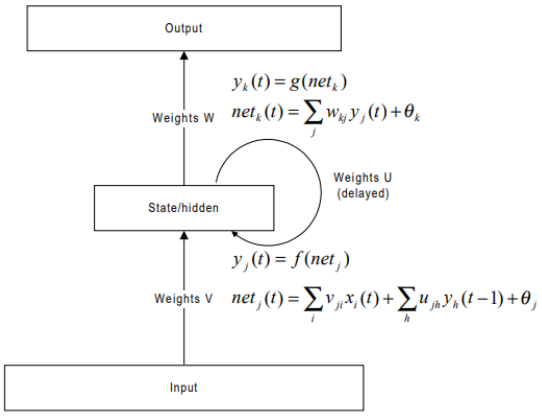


Figure 25: RNN [130]

Studies of RNN model used for prognostic and RUL predictions has been already introduced in the previous sections.

4.1.2 Long-short term memory LSTM

The Long Short-Term Memory (LSTM) architecture was originally proposed by Hochreiter and Schmidhuber [132] and it is widely used because it is able to accurately model both short and long term dependencies in data [131]. LSTM is a idiosyncratic kind of recurrent neural network [133]. LSTM works very well for time series data and it can solve the problem of long-term dependence of the series, avoiding the problems of gradient disappearance and gradient explosion, so it can be used for the analysis of long-term data [134].
The LSTM consists of L LSTM layers which allow to carry information across many timesteps. Each layer, basically, saves information for later, thus preventing older signals from gradually vanishing during processing, according to [129]. The LSTM takes in input data, properly shaped, in order to have shape=[nsamples, timesteps, nfeatures].

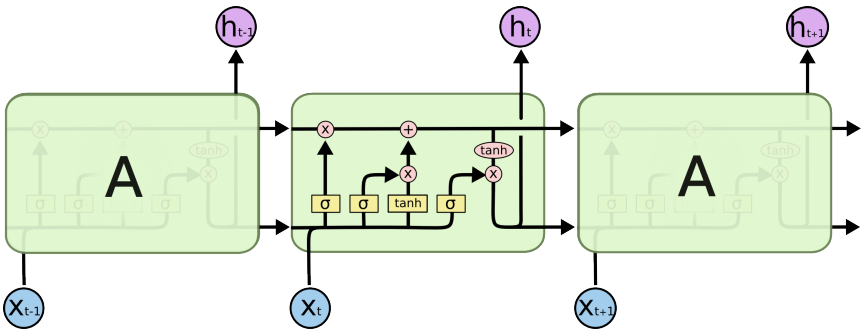


Figure 26: LSTM cell [135]

The LSTM, as showed in Fig. 26, has three gates which are non linear summation units that collect from inside and outside the block, and control the activation of the cell via multiplications [136]. The input and output gates multiply the input and output of the cell while the forget gate multiplies the cell's previous state [136]. According to [137], the LSTM network is based on the idea to generate a constant error path between subsequent time steps. In more detail, the forget gate allows to define which information has to be removed from the cell state, the input gate allows to regulates which new information has to be stored in the cell state, and the output gate controls the output of the LSTM unit defined as the hidden state h(t) [138]. Below the equations for a Vanilla LSTM are reported. Being x_t the input vector at time t [137], we can write:

$$g_t = tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \tag{4.1.1}$$

$$i_t = \sigma(W_{ix}x_t + W_{ix}h_{t-1} + b_i) \tag{4.1.2}$$

$$f_t = \sigma(W_{fx}x_t + W_{fx}h_{t-1} + b_f) \tag{4.1.3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{4.1.4}$$

$$o_t = \sigma(W_{ox}x_t + W_{ox}h_{t-1} + b_o) \tag{4.1.5}$$

$$h_t = o_t \odot tanh(c_t) \tag{4.1.6}$$

where W, b are respectively the weight and the bias, σ represents the sigmoid function, g is the block input, i, f and o are the input, forget and output gates, c represents the memory cell value and h is the block output. In these equation $\odot$ denotes element-wise Hadamard product [137].

4.1.3 Activation Function

There are several activation function, but in our research we have focused on: Rectified Linear Unit (ReLU) and Leaky Rectified Linear Unit (Leaky-ReLu).
The ReLu, according to [139], is a simple function which is the identity function for positive input and zero for negative input and as indicated below:

$$ReLu(x) = max(0, x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{4.1.7}$$

The ReLu activation function, according to [140], allows to solve the problem of "expansion and disappearance" in the sigmoid and tanh functions. The downside of ReLU is the vanishing gradient problem for the negative inputs [139].

The Leaky-ReLu was first proposed by Andrew L. Maas. from the Department of Computer Science at Stanford University and it is an improved version of the ReLu activation function. The way it works has been reported in the following equations:

$$LeakyReLu(x) = \begin{cases} ax & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{4.1.8}$$

where a is the slope value that can be chosen arbitrarily. In particular, when this activation layer takes a negative input, it returns a smaller linear component of the input.
Leaky ReLU can also avoid the dead ReLU problem because it allows a smaller gradient when calculating the derivative, but as for the ReLu also with the LeakyReLu activation function it's not possible to the problem of gradient explosion [141].
In Fig. 27 , we can notice the difference between the 2 activation functions introduced before.

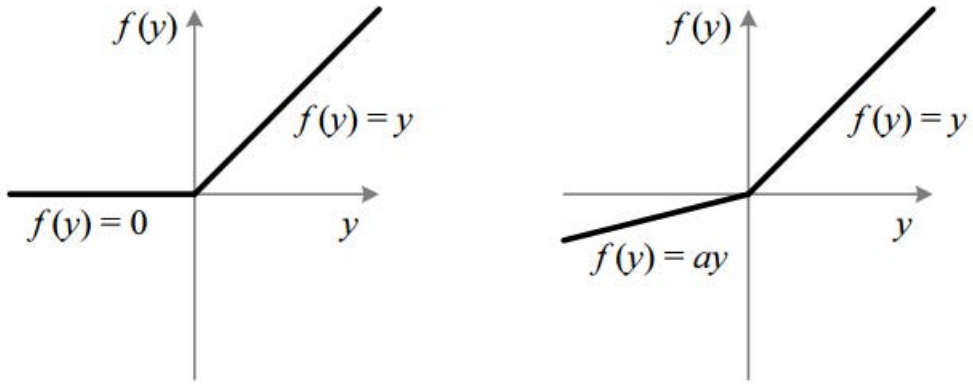
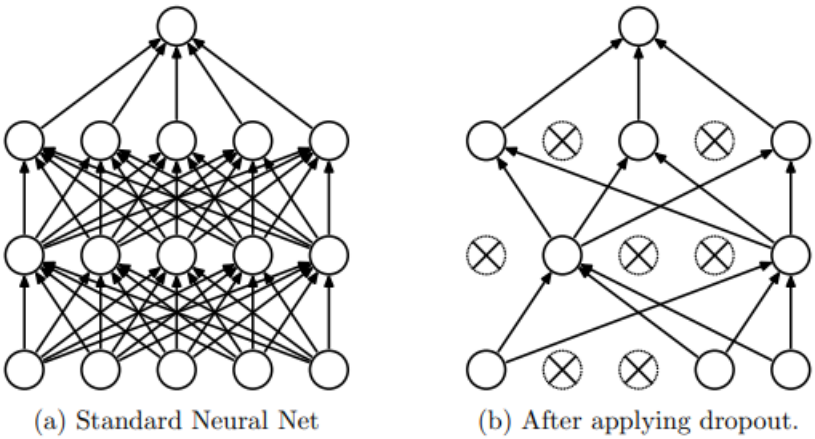


Figure 27: ReLu and LeakyReLu activation function [142]

4.1.4 Dropout Regularization

Large Neural Networks are slow to use, making it difficult to deal with overfitting by combining the predictions of many different large neural nets at test time [143]. As indicated in [144], the overfitting is a problem which occurs when the model performs very well with training data and fails to perform well on test data. In particular, the model learns the noise patterns present in the training data and this means that a large gap between the training and test error is generated [144]. In order to avoid this problem, the dropout regularization has been introduced. In this sense, the dropout is a technique which allows to randomly delete units (along with their connections) from the neural network during training, as showed in Fig. 29.



(a) Standard Neural Net (b) After applying dropout.

Figure 28: Dropout Regularization [143]

According to [145], in a LSTM-Dropout architecture, the hidden state and the output can be computed as:

$$h^{<t>} = \Theta_f \cdot (W_{hh} \odot h^{<t-1>} + W_{zh} \odot z^{<t>} + b_h) \tag{4.1.9}$$

$$H^{<t>} = \Theta_f \cdot (W_{hH} \odot h^{<t>} + b_H) \tag{4.1.10}$$

where, as reported in [145], we have $h^{<t>}$ is the hidden layer, $h^{<t-1>}$ is the previous hidden layer, Θ_f is the activation function and $H^{<t>}$ is the output. W_{hh}, W_{zh} and W_{hH} are, respectively, weights for the connection of the input layer to the hidden layer, the hidden layer to the hidden layer and the hidden layer to the output layer. In the end, the symbol <> indicates a vector and the symbol $\odot$ represents a matrix product.

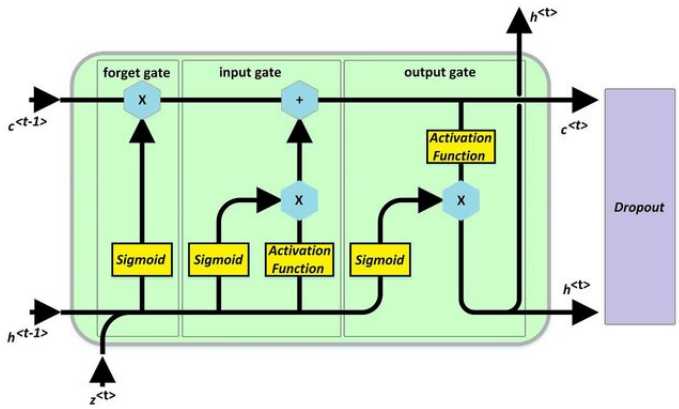


Figure 29: Dropout layer [145]

4.1.5 Monte Carlo dropout

The neural networks are affected by a certain uncertainty and when a model is developed, it is important to evaluate how certain is the output. In this sense, following the categorization of uncertainty indicated by Der Kiureghian & Ditlevsen [146], we can define two different kind of uncertainty:

- Epistemic uncertainty or model uncertainty, correlated to the fact that we do not know which could be the best model for the data
- Aleatoric uncertainty or data uncertainty, where the data does not fully contain the information they should have captured [147]

While the first type of uncertainty is defined as a reducible uncertainty because we can developed the best model implementing hyperparameters tuning techniques; the aleatoric is defined as an irreducible uncertainty because the information that is not contained in the dataset considered can't be recovered
According to [148], despite the fact the Bayesian networks work well in terms of uncertainty quantification, they are characterised by high computation costs. For this reason, we introduce Monte Carlo dropout which is an alternative to the Bayesian inference but more computationally efficient.
Labach et al. [149] explained how to implement Monte Carlo dropout. In particular, firstly a neural network is trained using standard dropout regularization. Then, to perform inference on an input sample, the network is run T times (the number of iterations depends on the dataset) with regular dropout layers, all with the same input but with different randomly generated dropout masks each time [149].

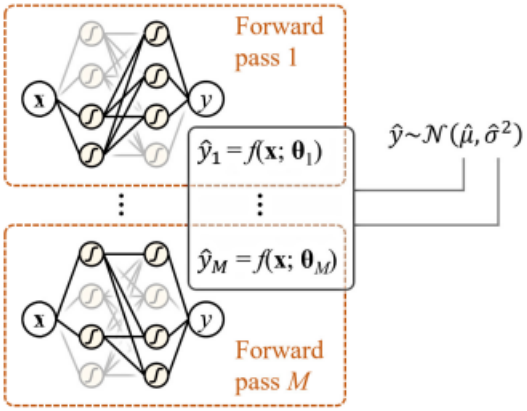


Figure 30: Monte Carlo Dropout [148]

Typically, Monte Carlo dropout is used only during training to prevent overfitting. In this work, we use Monte Carlo dropout during training to prevent overfitting of the model, and during testing to obtain the probability distribution of the RUL, as it will be explained in the next sections.

4.2 Experimental Set-up

4.2.1 Data Description

We consider the EDP Wind Farm dataset available at [91]. The considered Open Source dataset, object of our analysis for prognostics application, has been created for the challenge "Hack the Wind" and it contains all the monitoring parameters of real Wind Turbines. In particular, the EDP datasets contains the monitoring parameters related to a Wind Farm made of 5 Wind Turbines of the same model with a time span of 1 years (from January 2017 to December 2017). In more detail, the data contained in the EDP dataset is reported in Table 16.

Variables		
Generator RPM max rpm	Latest Production Active P Generator0 avg Wh	Grid Rotor Inverter Phase1 T avg °C
Generator RPM min rpm	Latest Production Active P Generator1 avg Wh	Grid Rotor Inverter Phase2 T avg °C
Generator RPM avg rpm	Latest Production Active P Generator2 avg Wh	Grid Rotor Inverter Phase3 T avg °C
Generator RPM std rpm	Latest Production Total Active P avg Wh	Controller Cooling water T avg °C
Generator Bearing T avg °C	Latest Production Reactive P Generator0 avg Wh	Grid Power avg kW
Generator Phase1 T avg °C	Latest Production Reactive P Generator1 avg Wh	Grid Power cosphi
Generator Phase2 T avg °C	Latest Production Reactive P Generator2 avg Wh	Grid Frequency avg Hz
Generator Phase3 T avg °C	Latest Production Total Reactive P avg Wh	Grid Voltage Phase1 avg V
Hydraulic Group oil T avg °C	High volt Transformer phase 1 T avg °C	Grid Voltage Phase2 avg V
Gearbox oil T avg °C	High volt Transformer phase 2 T avg °C	Grid Voltage Phase3 avg V
Gearbox Bearing T avg °C	High volt Transformer phase 3 T avg °C	Grid Current Phase1 avg A
Nacelle T avg °C	Grid Inverter Phase1 T avg °C	Grid Current Phase2 avg A
Rotor RPM max rpm	Controller Top T avg °C	Grid Current Phase3 avg A
Rotor RPM min rpm	Controller Hub T avg °C	Grid Max Power kW
Rotor RPM avg rpm	Controller VCP T avg °C	Nacelle Direction avg °
Ambient Windspeed max m/s	Generator Slip ring T avg °C	Grid Min Power kW
Ambient Windspeed min m/s	Spinner T avg °C	Grid Busbar T avg °C
Ambient Windspeed std m/s	Blades Pitch Angle Min °	Rotor RPM std rpm
Ambient Windspeed avg m/s	Blades Pitch Angle Max °	Ambient Windspeed est avg m/s
Ambient Wind Direction relative avg °	Blades Pitch Angle avg °	Grid Production Power std kW
Ambient Wind Direction absolute avg °	Blades Pitch Angle std °	Grid Production Reactive Power avg kVAr
Ambient T avg °C	Controller VCP Chokcoil T avg °C	Grid Production Reactive Power max kVAr
Grid Production Reactive Power min kVAr	Grid Production Reactive Power std kVAr	Grid Production Possible Power avg kW
Grid Production Possible Power max kW	Grid Production Possible Power min kW	Grid Production Possible Power std kW
Grid Production Possible Inductive Reactive Power avg kVAr	Grid Production Possible Inductive Reactive Power max kVAr	Grid Production Possible Inductive Reactive Power min kVAr
Grid Production Possible Inductive Reactive Power std kVAr	Grid Production Possible Capacitive Reactive Power avg kVAr	Grid Production Possible Capacitive Reactive Power max kVAr
Grid Production Possible Capacitive Reactive Power min kVAr	Grid Production Possible Capacitive Reactive Power std kVAr	Generator Bearing2 T avg °C

Table 16: EDP Data description

We have 81 features related to all the aspects of the functioning of the most important components of the Wind Turbines, production information, environmental condition and turbine condition variables. With a 10-min time of resolution, the 81 features are not all independent because, for some of them the dataset, also contains statistics, such as the average, minimum, maximum and standard in the 10 min period. To discuss method and algorithm implemented in this research, we extract from the 81 parameters, some features that will feed the LSTM model.

4.2.2 Features extraction

In this work, the Wind Turbines have been considered as if they would be single systems and not focusing only on some sub-components. From the 81 initial monitoring data, contained in the dataset, we have selected 31 features which are those directly related to the whole Wind Turbines and to their subsystems.
The parameters chosen for implementing the LSTM model of this work have been reported in Table 17.
As it is reported in Table 17, we have considered all those variables which are strictly correlated to the behaviour of the Wind Turbine and its components. In fact, in this sense, we can observe as the most used feature for almost all the component is the Temperature, because often anomalies in the temperature trend testifies the occurrence of problems or faults. The average, minimum, maximum and standard deviation is always computed each 10 minutes.

Variables
Hydraulic Oil T avg °C
Generator RPM max rpm
Generator RPM min rpm
Generator RPM avg rpm
Generator RPM std rpm
Generator Bearing T avg °C
Generator Bearing2 T avg °C
Generator Phase1 T avg °C
Generator Phase2 T avg °C
Generator Phase3 T avg °C
Gearbox Oil T avg °C
Gearbox Bearing T avg °C
Nacelle T avg °C
Rotor RPM avg °C
Rotor RPM min °C
Rotor RPM max °C
High Volt Transformer Phase1 T avg °C
High Volt Transformer Phase2 T avg °C
High Volt Transformer Phase3 T avg °C
Grid Inverter Phase1 T avg °C
Controller Top T avg °C
Controller Hub T avg °C
Controller VCP T avg °C
Controller VCP Chokcoil T avg °C
Controller VCP Cooling Water T avg °C
Spinner T avg °C
Generator Slip Ring T avg °C
Grid Rotor Inverter Phase1 T avg °C
Grid Rotor Inverter Phase2 T avg °C
Grid Rotor Inverter Phase3 T avg °C
Latest Production Total Active Power Wh

Table 17: Features extracted

4.2.3 Cases of study

For the aim of this work, we have considered 4 different cases of study to test the algorithm implemented. In particular, the different cases have been defined: CASE 1, CASE 2, CASE 3, CASE 4.
Each case consists of a training, validation and testing set. For each Wind Turbine in the training, validation and testing set, the measurements considered are all those directly related to the components and sub-components of a Wind Turbine as explained in the previous section. The goal is to predict the RUL at the moment of failure, i.e, the number of days until the Wind Turbine fails.
Table 18 shows the characteristics of each case.

	CASE 1	CASE 2	CASE 3	CASE 4
training set	WT01,WT07	WT01,WT06	WT06,WT07	WT06,WT07
validation set	WT11	WT11	WT01	WT11
testing set	WT06	WT07	WT11	WT01
length training set	390	453	399	399
length validation set	116	116	222	116
length testing set	231	168	116	222
First Faults	Hydraulic Group	Hydraulic Group	Hydraulic Group	Transformer
Date of first fault	2017-08-19	2017-06-17	2017-04-26	2017-08-11
Lifetime	8 months	6 months	4 months	8 months

Table 18: EDP Cases description

4.2.4 Data pre-processing

Before injecting the features in the LSTM model, several steps of data preparation have been made. As anticipated in the previous section, from the original dataset with 81 features, we have kept only those features which are meaningful for the monitoring of the correct behaviour of Wind Turbines. For this reason, we have extracted 31 variables which will be used as input for our model. For each Wind Turbine, indicated with the nomenclature WT01, WT06, WT07, WT11, we have a record of 1 year with a time sampling of 10 min. The measurements considered, in our model, have been stopped when the very first fault occurs for a given Wind Turbine. Then, in order to avoid any overfitting problems, a time sampling of 24 hours has been adopted.
After this preliminary features engineering phase, the next step has been to normalize our data, in order to eliminate any gradient-exploding problem that could occur in case we directly use the raw data. For each case, the method used for the normalization has been the Min Max normalization [150]:

$$x_i' = \frac{x_i - x_{imin}}{x_{imax} - x_{imax}}$$

where x_i' is the normalised i-th feature, x_i is the raw i-th feature, x_{imin} minimum of the i-th feature, x_{imax} maximum of the i-th feature.
It has been adopted the MinMax method because it allows to maintain the exact scale, despite the fact that it is quite sensitive to the outliers (this is not a problem because we have showed as our dataset is quite robust towards to ourliers).
Then, we have applied a sliding window in order to segment the data as indicated in [150].

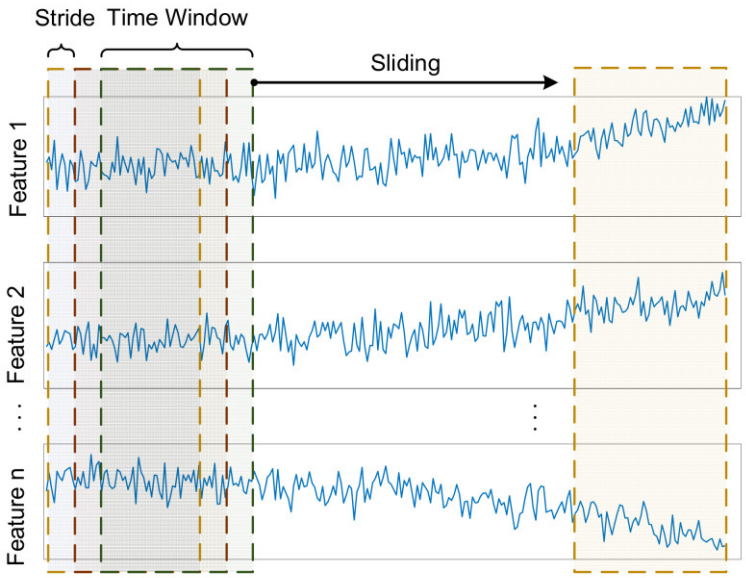


Figure 31: Sliding window processing [151]

In particular, indicating the lifetime of the Wind Turbine until the failure with $\boldsymbol{Ts}$, the length of the window is indicated as $\boldsymbol{w}$ and the stepsize of the stride is indicated as $\boldsymbol{s}$, the real RUL has a length of Ts-w-(s-1) [150]. In this work, we have considered w=3 and s=1.

4.2.5 Proposed Methodology

At time k of a Wind Turbine i, we consider the normalised data x_k^i as input:

$$x_k^i = \{x_{1,k}'^i, x_{2,k}'^i, x_{3,k'}^i, x_{4,k}'^i, ..., x_{m,k}'^i\} \tag{4.2.1}$$

where m is the total number of the considered features and $x'_{m,k}$ is the normalised sensor measurements of day k of Wind Turbine i from feature m.
Considering i Wind Turbines, the real RUL, at the time k, can be indicated as:

$$RUL^a(WT_i) = min\{tc_1^i - k, tc_2^i - k, tc_3^i - k, tc_4^i - k, ..., tc_n^i - k\} \tag{4.2.2}$$

where $c^i = \{c_1^i, c_2^i, c_3^i, c_4^i, ..., c_n^i\}$ is the set of components for WT_i and $\{tc_1^i, tc_2^i, tc_3^i, tc_4^i, ..., tc_n^i\}$ is the set of the failures times of these the components c^i. In our research, we have considered a piece-wise linear RUL target function and in this sense, we have assumed a linear degradation path of Wind Turbines.
The Long Short term memory model consists of L LSTM layers. Each layer contains N neurons and between them, to get better prediction of the Point-RUL has been added LeakyReLu activation layers, because they improve the performance of the model given the high non-linearity of the input data. The last layer of the architecture is a Dense layer which is a simple layer of neurons which takes as input all the inputs of the neurons of the previous layer. For the Dense layer, a ReLu activation function has been selected because it takes the negative input of the previous layers and it returns zero. This has been very useful in order to make predictions when we are close to the fault. After the point-RUL prediction, our attention has been focused on finding which are the most significant measurements implementing a game theoretic approach in order to understand which component, for our case study, has more impact of the RUL of the entire Wind Turbine. In the end, after the phase of features evaluation, all the features have been used to feed an LSTM model characterised by Monte Carlo dropout layers in order to predict the possible RUL distribution each timestep to understand when could be more efficient performing a maintenance task.

4.2.6 Features Importance - SHAP Method

After having implemented a LSTM model which is able to well predict the true RUL, we have investigated which are the features, considered in the training set, that have the highest impact on the output of the model. In this sense, in order to analyse the features interactions and effect on the model's prediction we have determined Shapley additive explanations (SHAP) value of a feature i, which we denote as ϕ_i as follows [152]:

$$\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} |f(S \cup \{i\}) - f(S)| \tag{4.2.3}$$

where F is the set of all features considered for the LSTM model, $S \subseteq F$ is a subset of features obtained from the set F except feature i, and f(S) is the expected prediction output given by the set S of features. The SHAP values show which features have a significant impact on the point-RUL prediction.
We have considered SHAP summary plot, because it allows to analyse and evaluate all these aspects of a feature's importance. Features are sorted for their global impact:

$$\sum_{j=1}^{N} \left| \phi_i^{(j)} \right|$$

where the $\phi_i^{(j)}$ is the SHAP value of the j-th feature.

4.2.7 Monte Carlo dropout

Let consider X as the samples with sensor measurements in the training set of the LSTM, and let Y be the corresponding RUL values, as indicated in [153]. The objective is to predict the distribution of RUL given the training samples X and Y. For a Bayesian Neural Network, we have:

$$p(y|x, X, Y) = \int p(y|x, \omega) p(\omega|X, Y) d\omega \tag{4.2.4}$$

where ω represents all the weights in the neural network. As in [153], $p(y|x,\omega)$ is the probability that the RUL equals y, given test sample x and the weights of the neural network ω. $p(\omega|X,Y)$ is the posterior distribution of the weights, and represents the probability that the weights are ω, given the training samples X and Y.
But as anticipated before, the Bayesian Neural Network has high computational costs. According to [154], the posterior distribution $p(y|X,Y)$, in the variational inference, is approximated with a distribution $q(\omega)^*$. This approximation is needed because we can't evaluate $p(y|X,Y)$ analytically [154]. Firstly, we define a family Q of possible posterior distribution $q(\omega)^*$, as done in [153]. The aim is to find an approximate distribution $q(\omega)^* \in$ Q, which minimises the Kullback–Leibler (KL) divergence, in order to have a distribution which is as close as possible to the posterior distribution obtained from the full Gaussian process [154]. According to [155], we can express:

$$q(\omega)^* = argmin_{q(\omega)\in Q}\{KL(q(\omega|p(\omega|X,Y)))\} \tag{4.2.5}$$

Using $q(\omega)^*$, we approximate the posterior distribution of the RUL of a test sample by:

$$q(y|x) = \int p(y|x,\omega)q(\omega)^* d\omega \tag{4.2.6}$$

where $q(y|x)$ is the approximation of $p(y|x,X,Y)$.
The authors of [154] show how we can approximate the expected value $\hat{y}$ of the RUL of a test sample as:

$$\hat{y} = E_{q(y|x)}(y) = \frac{1}{M}\sum_{1}^{M}\hat{y_j}(x,\omega^j) \tag{4.2.7}$$

where M is the number of forward passes through the neural network, ω^j are the weights of the neural network belonging to the j-th forward pass (i.e., where some neurons are dropped out), and $\hat{y_j}(x,\omega^j)$ is the RUL prediction obtained from the j-th forward pass through the neural network, as indicated in [153].

4.2.8 Hyperparameter tuning

The considered hyperparameters are reported in Table 19. After having implemented different methods for hyperparameter tuning like Grid Search, Random Search and Bayesian Optimization, it has been possible to extract the number of layers which allows to have good performance for all cases study. Then, defined the number L of layers, the other hyperparameters reported in Table 19, has been found, iteratively, considering values within ranges provided by the different hyperparameters methods implemented in the preliminary phase.

Hypeparameters	
Number of layers	4
Neurons layer 1	128
Neurons layer 2	64
Neurons layer 3	64
Neurons layer 4	64
Number of Dropout layers	3
Dropout rate	0.5
Optimizer	Adam
Activation Function	ReLu, LeakyRelu
epochs	40
batch size	32
window length	3

Table 19: Hyperparameters for point-RUL prediction

Then, the LSTM architecture implemented for the Probabilistic RUL is the same of the point-RUL prediction with the only one exception represented by the Monte Carlo Dropout layers and the optimizer (only for CASE 1 the Adam optimizer has been kept). Table 20 reports the hyperparameters considered for the Probabilistic RUL.

Hypeparameters	
Number of layers	4
Neurons layer 1	128
Neurons layer 2	64
Neurons layer 3	64
Neurons layer 4	64
Number of MC Dropout layers	3
Dropout rate	0.5
Optimizer	Adamax
Activation Function	ReLu, LeakyRelu
epochs	40
batch size	32
window length	3

Table 20: Hyperparameters for Probabilistic RUL

4.2.9 Performance metrics

To evaluate the performance of the LSTM model implemented for the Point-RUL prediction, we have considered the Mean Absolute error and the Root Mean Square error.

$$MSE = \sum_{n=1}^{N} \frac{|RUL^a - RUL^p|}{N} \tag{4.2.8}$$

$$RMSE = \sqrt{\sum_{n=1}^{N} \frac{(RUL^a - RUL^p)^2}{N}} \tag{4.2.9}$$

where N is the number of days, RUL^a is the true RUL and RUL^p is the predicted RUL.
To evaluate the reliability and the goodness of our model for the probabilistic RUL, we have considered the reliability curve and the Continuous Ranked Probability Score (CRPS) [156] [157].
According to [157], the CRPS allows to evaluate whether the estimated RUL distribution is centered around the actual RUL of a component and the sharpness of the RUL prognostic (evaluates if the variance is low), even if it is not very used to evaluate the probabilistic RUL prognostic.

$$CRPS = \frac{1}{N} \sum_{i=1}^{N} CRPS_i \tag{4.2.10}$$

$$CRPS_i = \int_{-\infty}^{\infty} (F_{\hat{y}_i}(x) - I\{y_i \leq x\})^2 dx \tag{4.2.11}$$

with $I\{y_i \leq x\} = \begin{cases} 1 & \text{if } y_i \leq x \\ 0 & \text{if } y_i \; x \end{cases}$

According to [157], the CRPS can be considered as a probabilistic generalization of the absolute error. Smaller the CRPS metric is, closer the RUL prediction is to the actual RUL and in an ideal case when a perfect RUL prediction without uncertainty (i.e., a point RUL prediction) is obtained, CRPS equals zero [157]. To conduct a generic, parameter-free reliability analysis of the estimated RUL distribution, the Reliability Score (RS) has been computed [157]. As in [157], we have defined a reliability curve C(α) based on the αCoverage for probabilistic RUL prognostics.
Here, $C(\alpha)_i$=$\{\alpha - Coverage, \alpha \in 0.00, 0.05, ..., 1.00\}$, where $\alpha - Coverage = \frac{1}{N} \sum_1^N I(\alpha)_i$
with $I(\alpha)_i = \begin{cases} 1 & \text{if } y_i \; \in [\hat{y_i}^{0.5-0.5\alpha}, \hat{y_i}^{0.5+0.5\alpha}] \\ 0 & \text{if } y_i \; x \end{cases}$

Then, the reliability score (RS) can be defined as following:

$$RS^{over} = \int_0^1 I\{C(\alpha) \leq \alpha\}(\alpha - C(\alpha))d\alpha \tag{4.2.12}$$

$$RS^{under} = \int_0^1 (1 - I\{C(\alpha) \leq \alpha\})(C(\alpha) - \alpha))d\alpha \tag{4.2.13}$$

$$RS^{total} = RS^{under} + RS^{over} \tag{4.2.14}$$

with $I\{C(\alpha) \leq \alpha\} = \begin{cases} 1 & \text{if } C(\alpha) \leq \alpha \\ 0 & \text{otherwise} \end{cases}$

The RS^{under} quantifies the underestimation and the RS^{under} quantifies the overestimation.
In general, the reliability curve shows visually whether the uncertainty associated with the RUL predictions is overestimated or underestimated.

4.3 Results - Point RUL prediction

In this section, we have reported the results obtained implementing a LSTM model for the point-RUL prediction. The following figures show a representation of the RUL prediction depicted along with the actual target RUL and they allow to understand how good is the RUL prediction. This shows the predictability of the model over the complete life span of the Wind Turbine considered. We can observe how the prediction follows the general direction of the true RUL and it is quite close to the target curve.

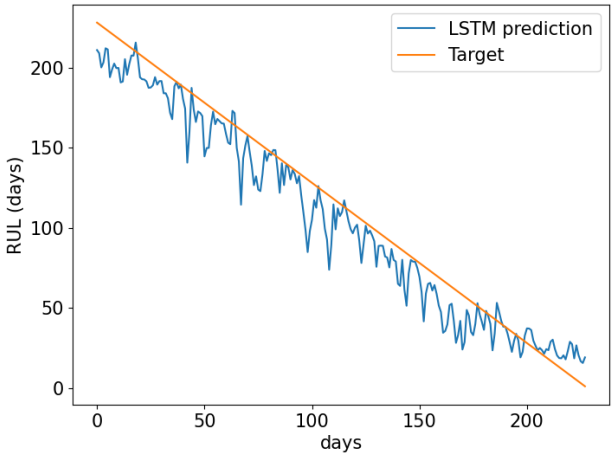


Figure 32: Point-RUL prediction Case 1

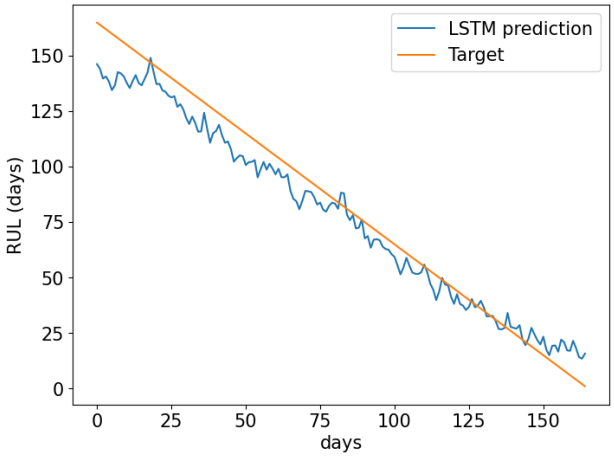


Figure 33: Point-RUL prediction Case 2

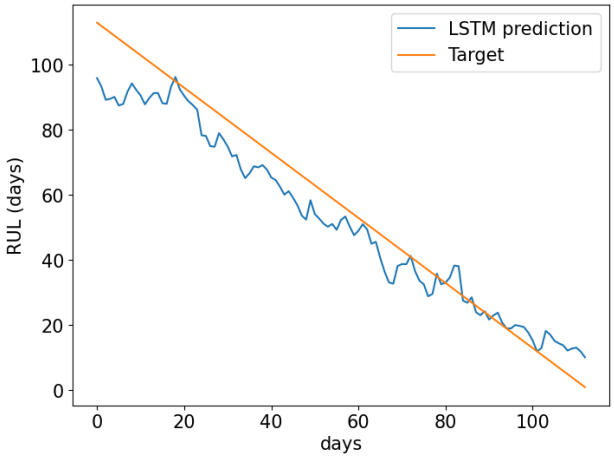


Figure 34: Point-RUL prediction Case 3

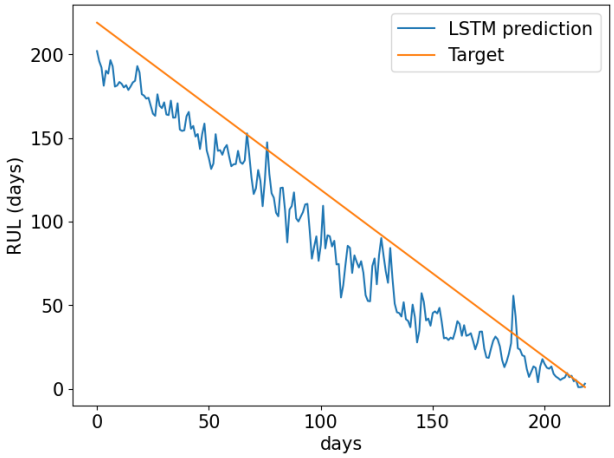


Figure 35: Point-RUL prediction Case 4

The predicted RUL is almost always below the target RUL for the most of the life time considered. This means that we are more conservative for the prediction. Observing the graphs, when we are close, for CASE 1, CASE 2 AND CASE 3, to the fault, we can see that the prediction goes above the target overestimating the RUL. In this sense, it could be more interesting performing an inspection or a maintenance task in the moment of intersection between the predicted curve and the target one to stay safe and not to risk to schedule an inspection when the fault has already occurred. On the other hand, Fig. 35, for CASE 4, shows how the model, differently from the previous cases, is able to predict the RUL when it is close to the fault. In this case, we could have the advantage to have more time to perform a task. However, in order to better understand when it could be more efficient to perform a maintenance task we have to also evaluate the probability distribution function of the RUL that will be introduced in the next section.
Table 21 shows the RUL prediction with the True RUL for each case.

Usage	Case 1		Case 2		Case 3		Case 4	
	RUL^a	RUL^p	RUL^a	RUL^p	RUL^a	RUL^ap	RUL^a	RUL^p
1	228	212.86	165	146.25	113	95.96	219	201.99
25	204	186.13	141	132.03	89	78.4	195	169.24
50	179	167.42	116	104.75	64	58.43	170	142.68
75	154	133.37	91	82.95	39	33.72	145	109.22
100	129	96.79	66	60.59	14	17.74	120	76.55
125	104	89.62	41	35.4	-	-	95	77.99
150	79	75.75	16	19.9	-	-	70	37.73
175	54	37.16	-	-	-	-	45	18.76
200	29	34.45	-	-	-	-	20	17.78
225	4	19.85	-	-	-	-	-	-
227	1	18.79	-	-	-	-	-	-

Table 21: Point-RUL prediction

The following Table shows the errors for all the case study considered.

Point-RUL prediction Errors

	MAE (days)	RMSE (days)
Case 1	12.27	15.49
Case 2	7.97	9.58
Case 3	7.19	8.7
Case 4	21.73	24.28

We can observe how the LSTM model, with the Dropout layers, implemented has quite very good performance for each case investigated. The best result, in terms of MAE and RMSE, is provided by the Case Study 3 (testing=WT11). The Case 4 is the one with the highest errors but it is the one which allows to better predict the RUL when we are close to the fault.

4.3.1 Features Importance - SHAP Method

The following Figures are the SHAP summary plot, one for each of the case study investigated.

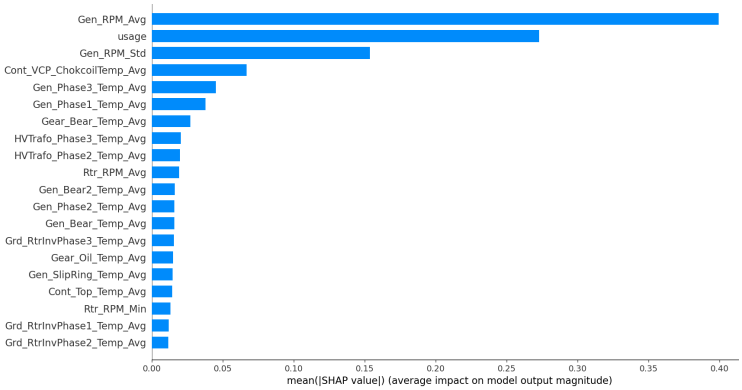


Figure 36: Shap Values Case 1

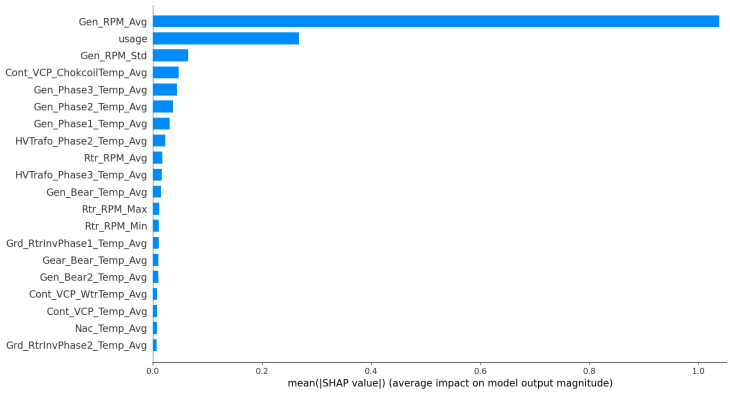


Figure 37: Shap Values Case 2

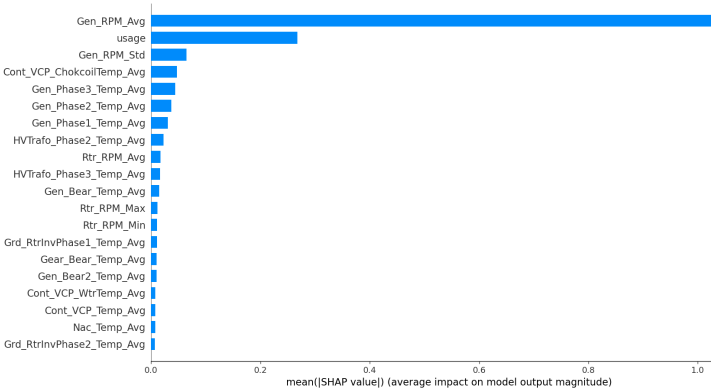


Figure 38: Shap Values Case 3

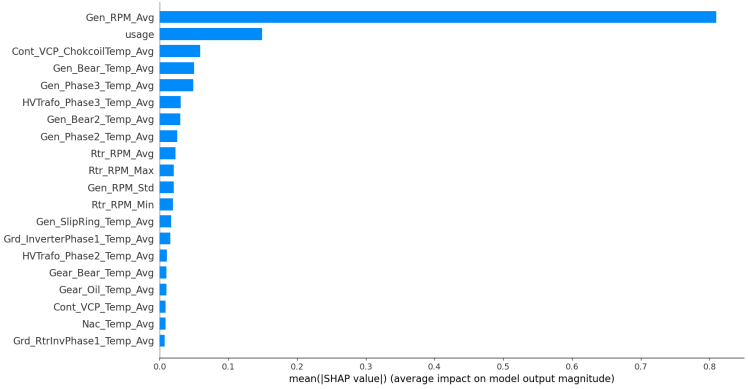


Figure 39: Shap Values Case 4

The features are sorted by the sum of the SHAP values. In this sense, the feature at the top of the graph has the highest impact on the RUL prediction, while the feature at the bottom of the graph has the lowest impact. We can observe, for all the case study, as the driving parameter for the RUL prediction, is the Generator RPM. These results confirm, according to what introduced in the literature review, that the generator is one of the most important component for the correct behaviour of a Wind Turbine.

4.4 Results - Probabilistic RUL prediction

In this section, we present the probabilistic RUL prognostics for Wind Turbines. In order to visualize the trend of our probabilistic RUL prediction, we represent the mean of the probabilistic RUL for each day. In this sense, we can see how the mean of the RUL distribution follows the target RUL. The ideal case would have been that the mean of the prediction would have overlapped the target RUL.

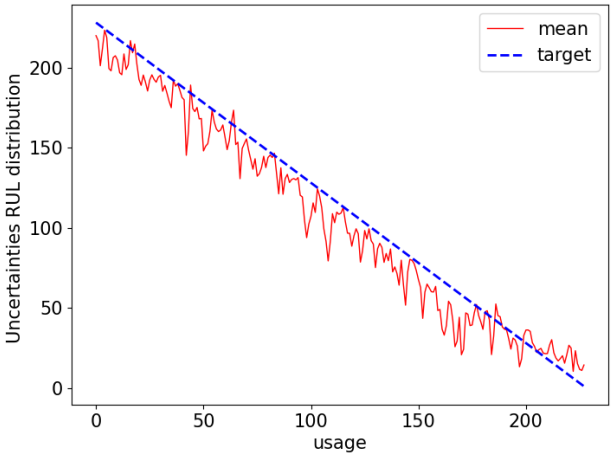


Figure 40: Mean Distribution Case 1

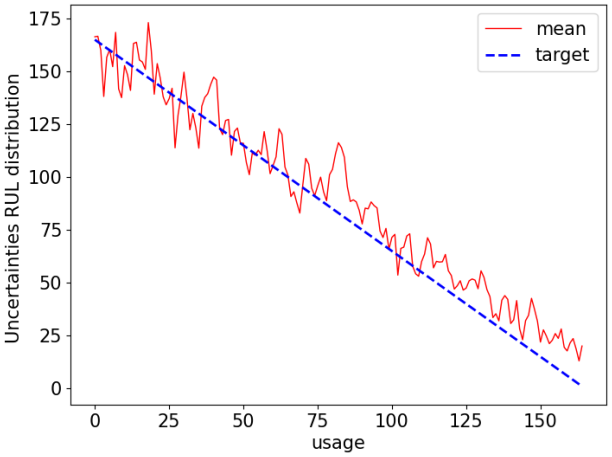


Figure 41: Mean Distribution Case 2

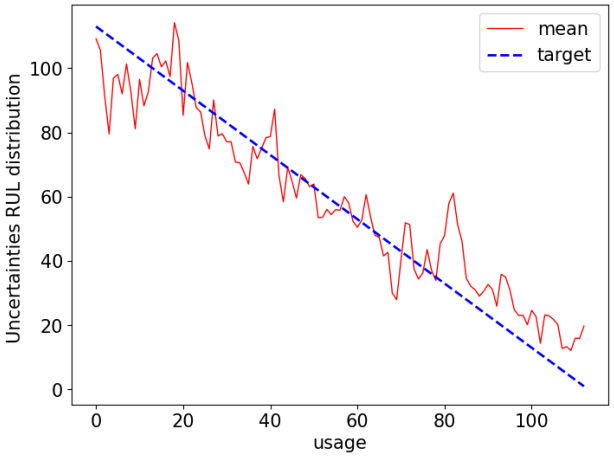


Figure 42: Mean Distribution Case 3

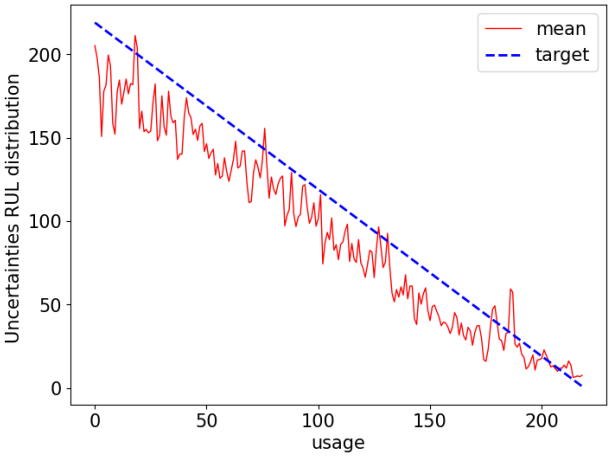


Figure 43: Mean Distribution Case 4

For Case 1 and Case 4, it can be seen as the mean is below the target RUL until we are close to the fault when the mean tends to be slightly above the true RUL value. The mean of the RUL distribution for Case 2 is almost for all the time span considered above the True RUL and this means that we shouldn't be too close to the fault to do a maintenance task because we could risk to already have a malfunction. In this sense, to stay the most conservative we can perform a task between 55-70 days of usage. In the end, Case 3 is characterised by a small testing set and this is the main reason because the mean of the distribution doesn't follow very well the dashed curve. In this case, given the fact that close to the fault the mean of the distribution is above, to be safe we could think to perform an inspection between 65-75 days of Wind Turbine usage. The following table shows the error between the mean of the probabilistic RUL and the target RUL.

Probabilistic RUL Error

	MAE (days)	RMSE (days)
Case 1	11.65	14.48
Case 2	10.03	12.17
Case 3	8.2	10.32
Case 4	19.56	22.95

Instead of predicting only one number for the RUL, however, we predict the PDF of the RUL for all the cases study. The following Figures show the evolution of the estimated RUL distribution over time for the 4 Cases study analysed.

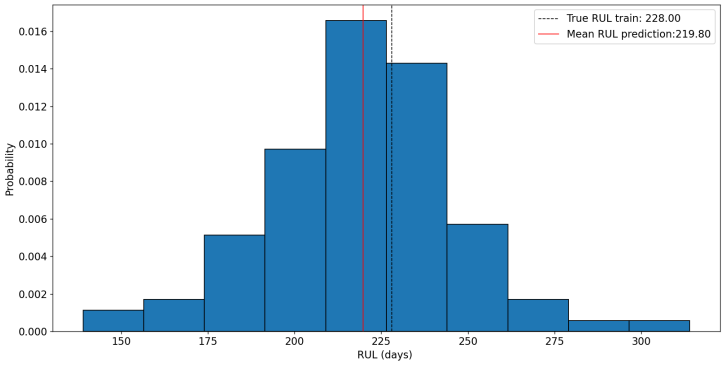


Figure 44: RUL=228 Case 1

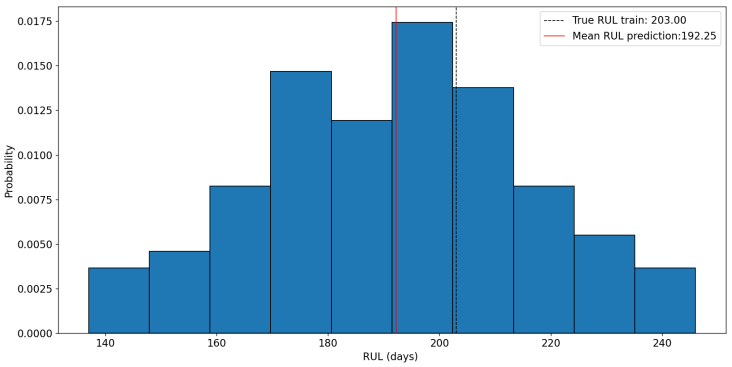


Figure 45: RUL=203 days Case 1

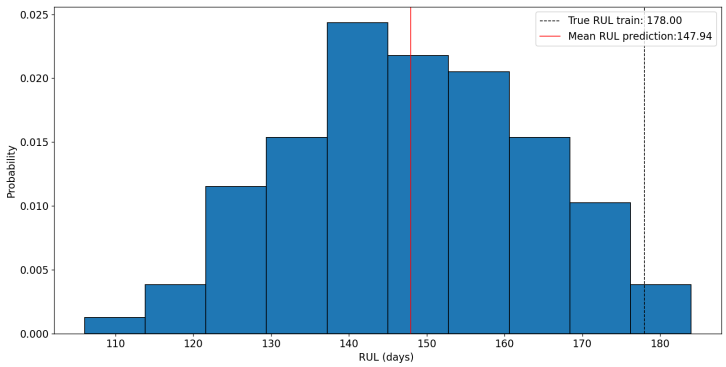


Figure 46: RUL=178 days Case 1

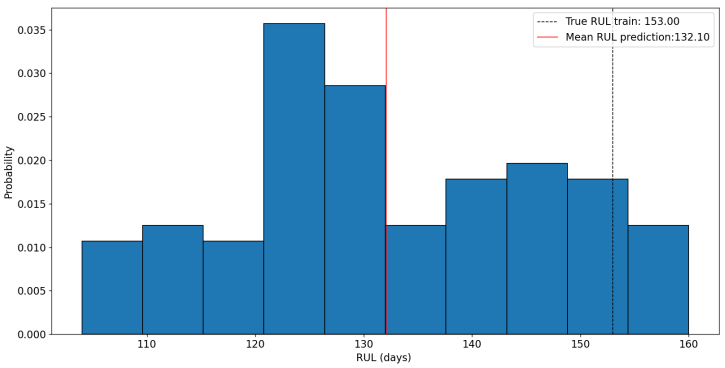


Figure 47: RUL=153 days Case 1

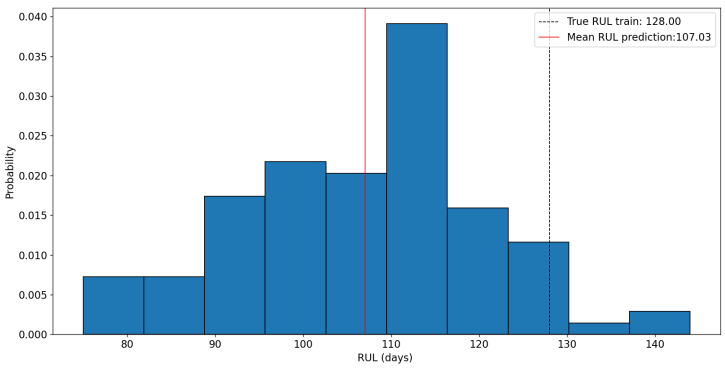


Figure 48: RUL=128 days Case 1

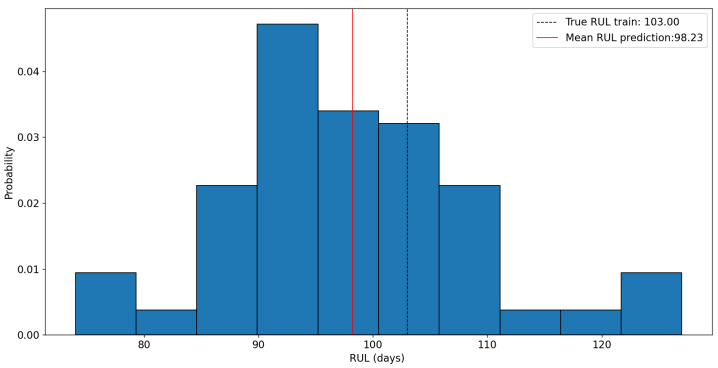


Figure 49: RUL=103 days Case 1

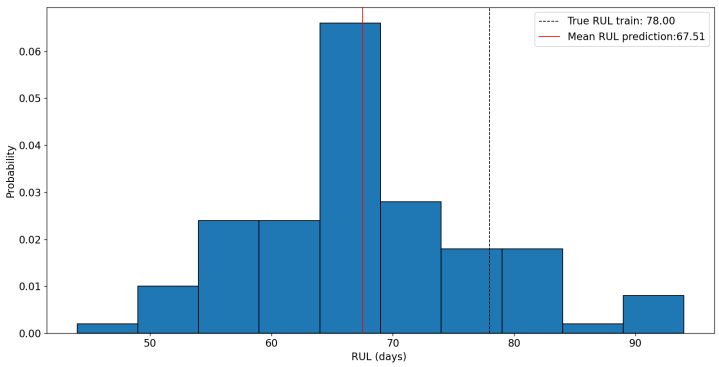


Figure 50: RUL=78 days Case 1

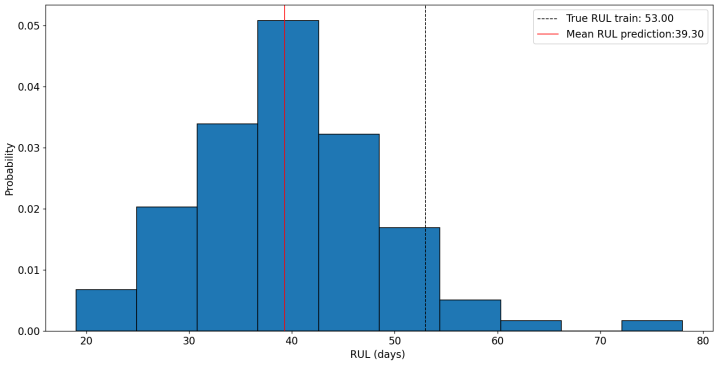


Figure 51: RUL=53 days Case 1

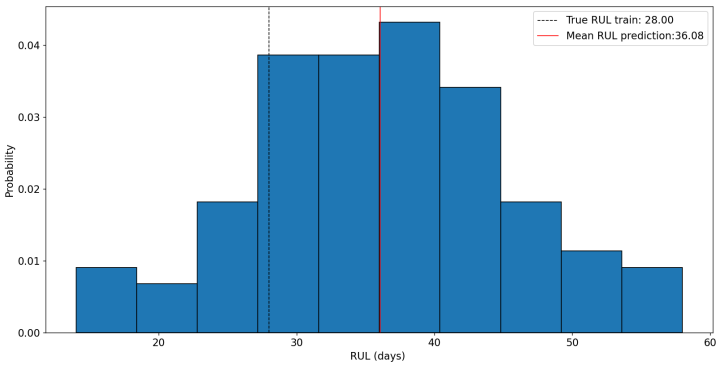


Figure 52: RUL=28 days Case 1

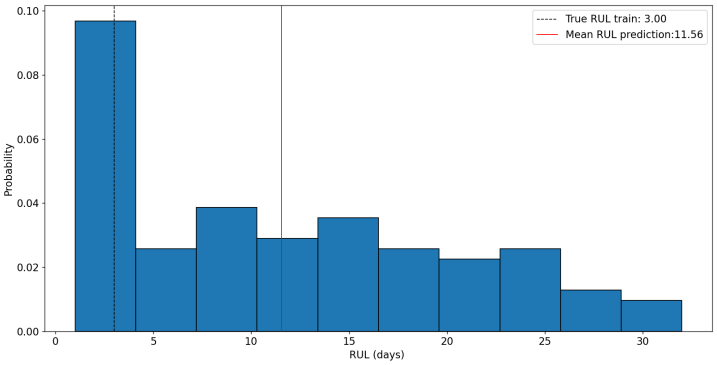


Figure 53: RUL=3 days Case 1

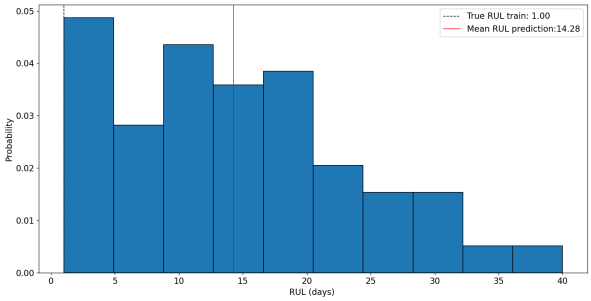


Figure 54: RUL=1 days Case 1

The probabilistic RUL for the second case study, where WT07 is testing, is reported.

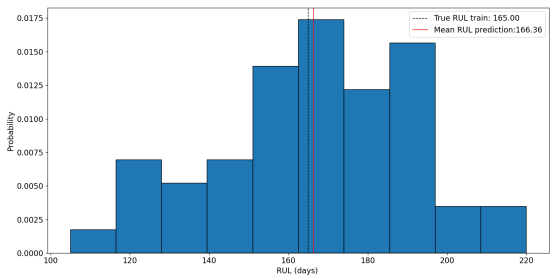


Figure 55: RUL=165 Case 2

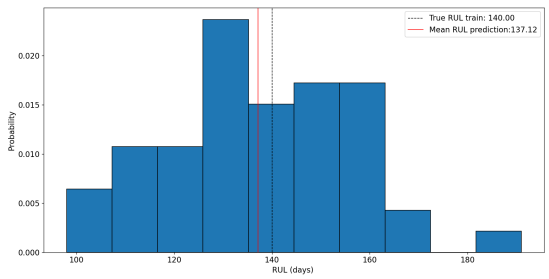


Figure 56: RUL=140 days Case 2

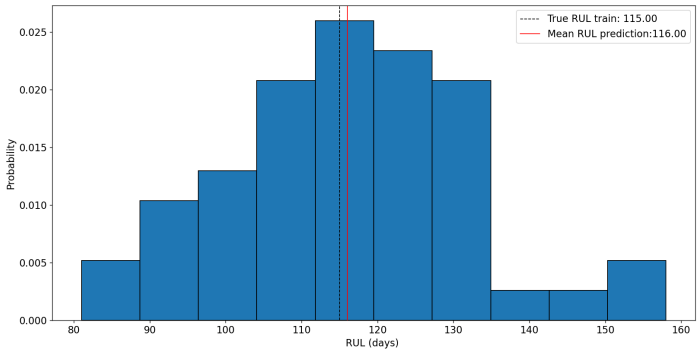


Figure 57: RUL=115 days Case 2

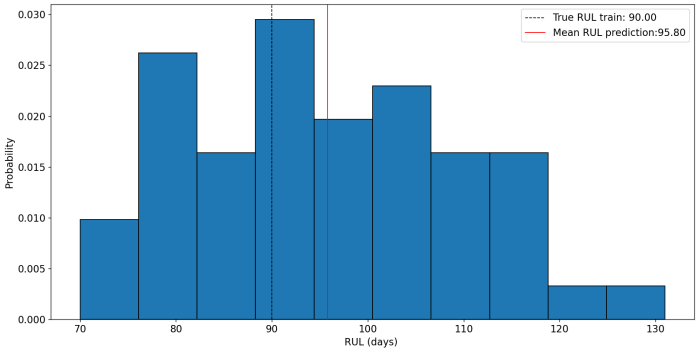


Figure 58: RUL=90 days Case 2

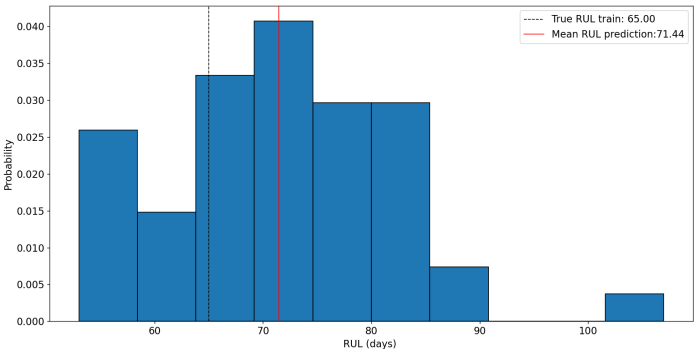


Figure 59: RUL=65 days Case 2

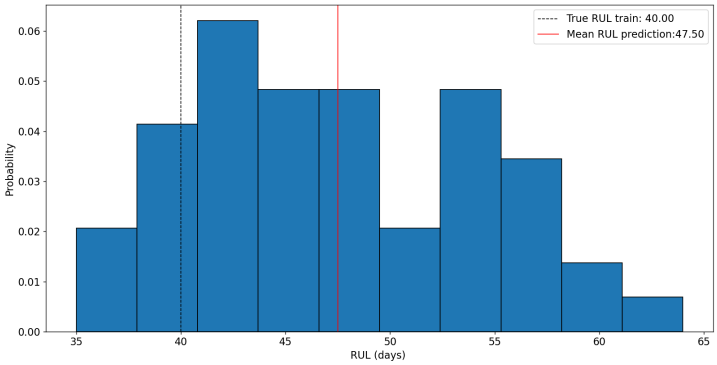


Figure 60: RUL=40 days Case 2

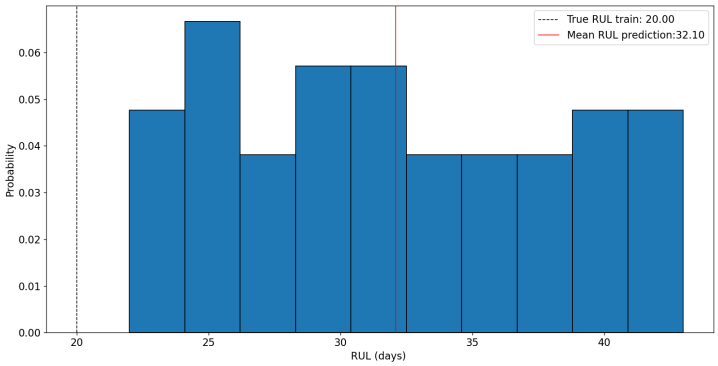


Figure 61: RUL=20 days Case 2

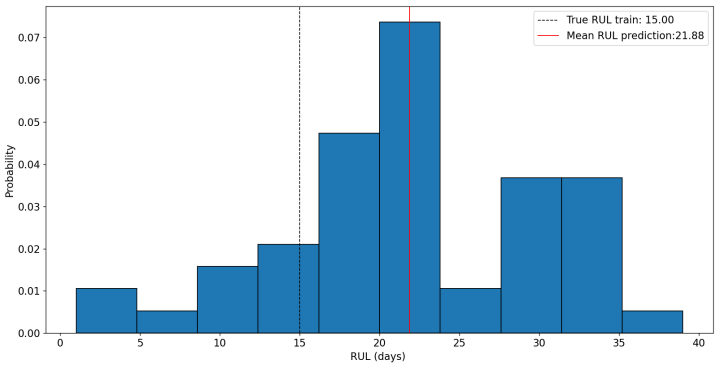


Figure 62: RUL=15 days Case 2

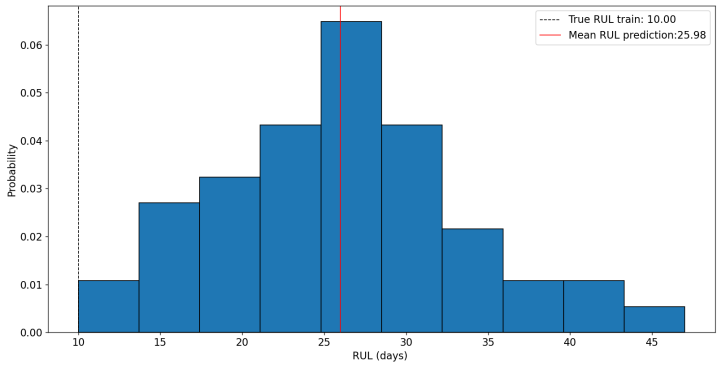


Figure 63: RUL=10 days Case 2

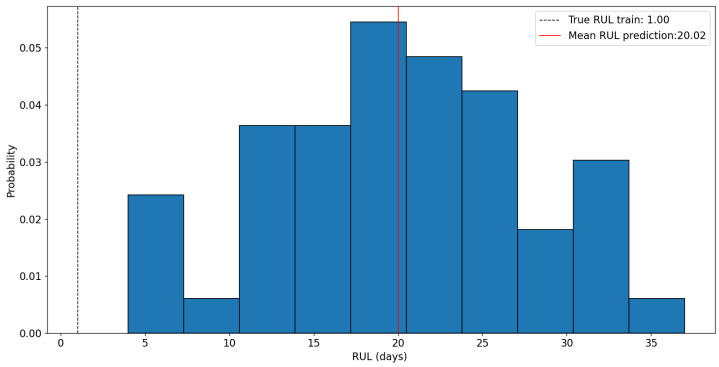


Figure 64: RUL=1 days Case 2

The probabilistic RUL for CASE 3 is represented by the following Figures.

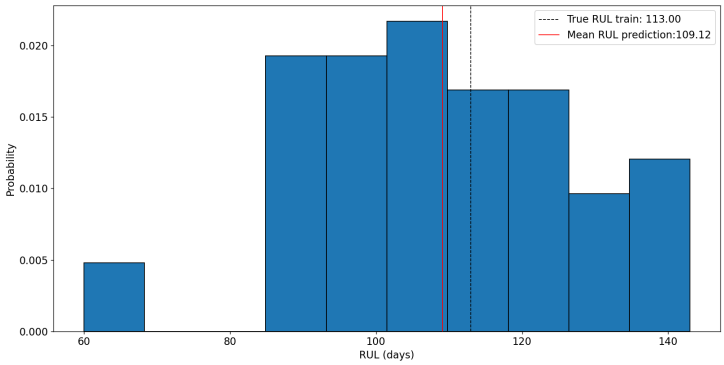


Figure 65: RUL=113 Case 3

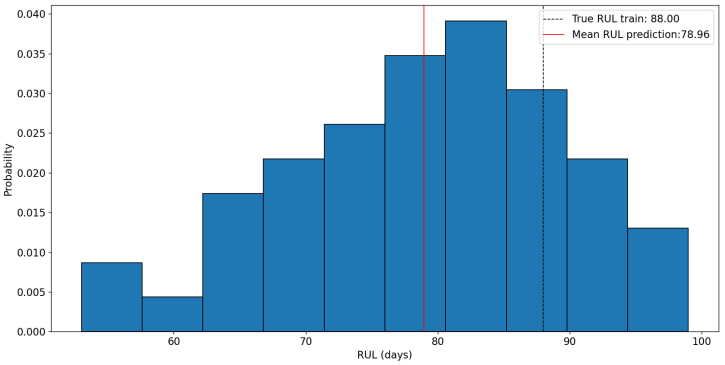


Figure 66: RUL=88 days Case 3

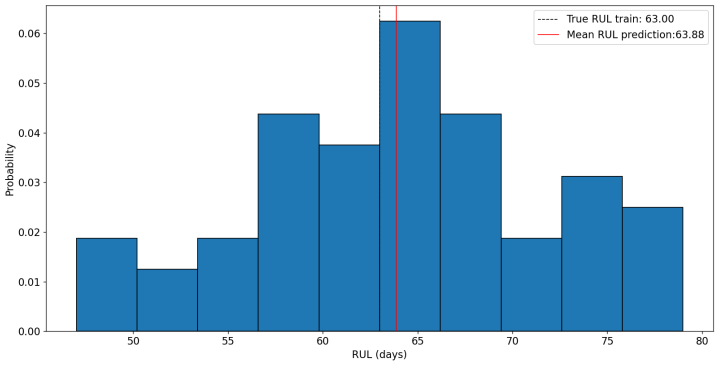


Figure 67: RUL=63 days Case 3

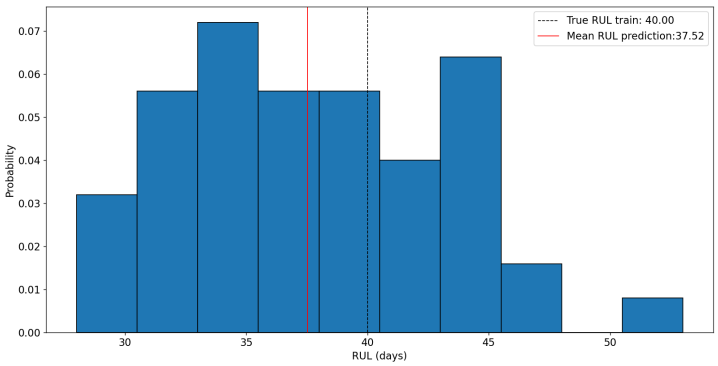


Figure 68: RUL=40 days Case 3

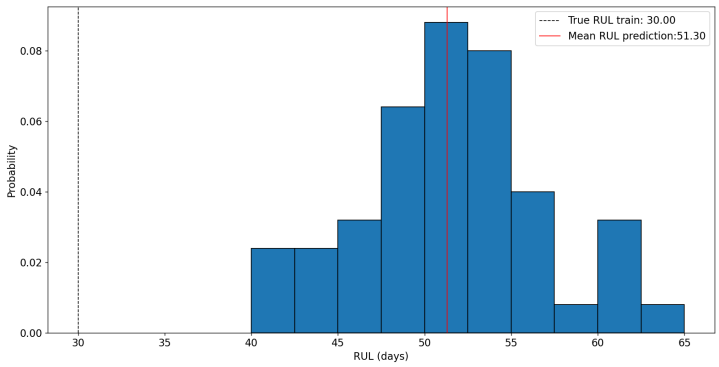


Figure 69: RUL=30 days Case 3

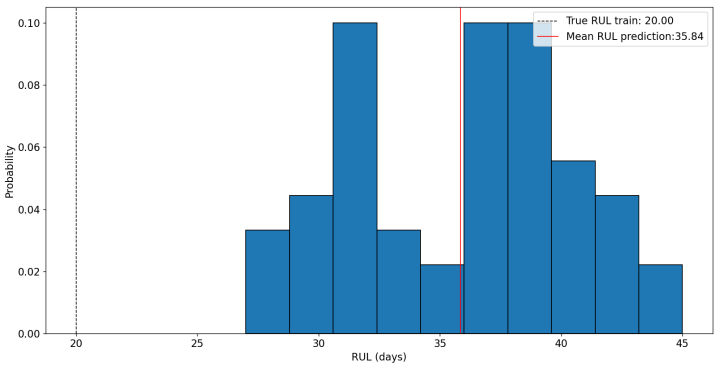


Figure 70: RUL=20 days Case 3

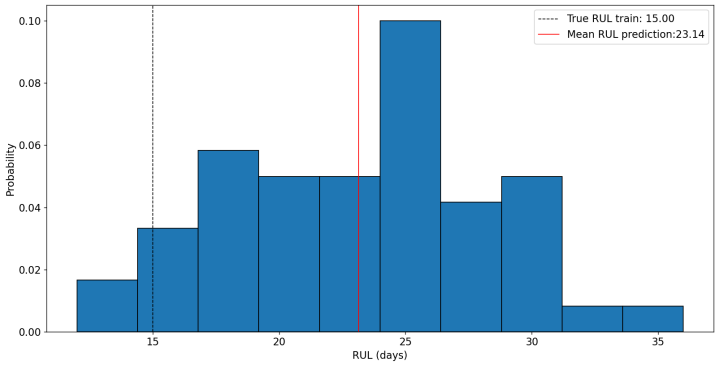


Figure 71: RUL=15 days Case 3

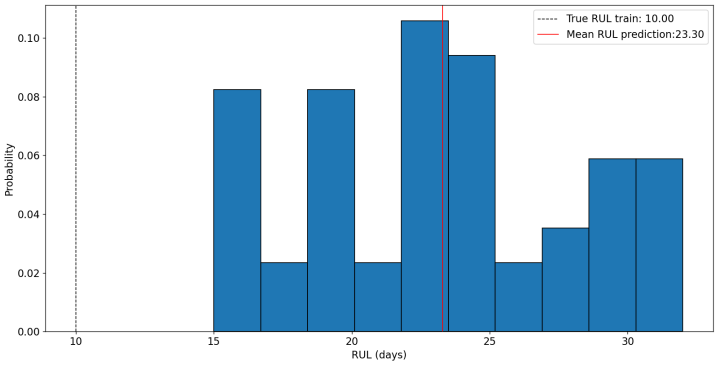


Figure 72: RUL=10 days Case 3

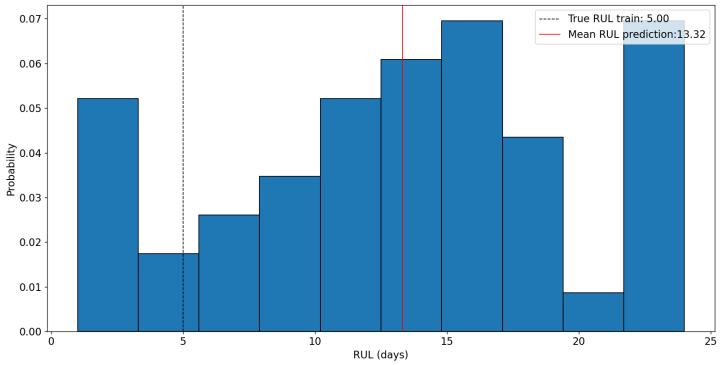


Figure 73: RUL=5 days Case 3

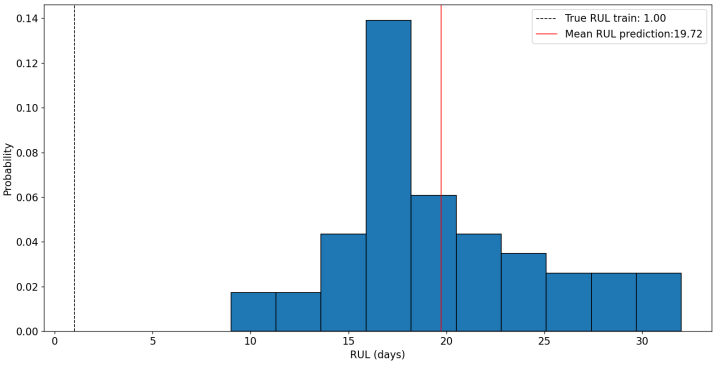


Figure 74: RUL=1 days Case 3

The probabilistic RUL for CASE 4 is represented by the following Figures.

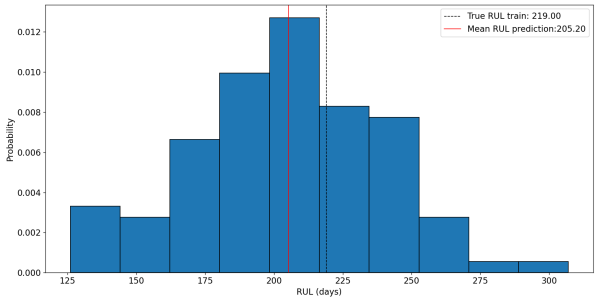


Figure 75: RUL=219 Case 4

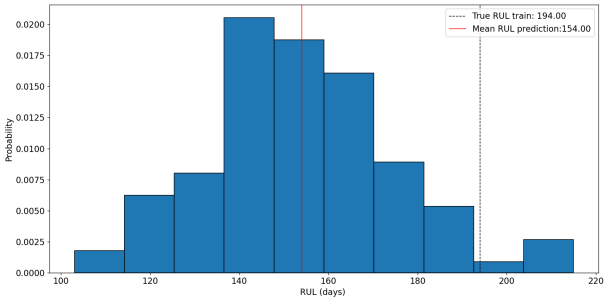


Figure 76: RUL=194 days Case 4

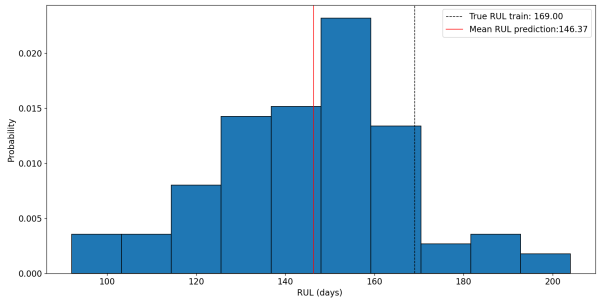


Figure 77: RUL=169 days Case 4

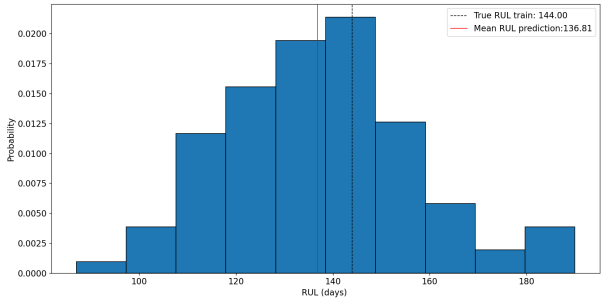


Figure 78: RUL=144 days Case 4

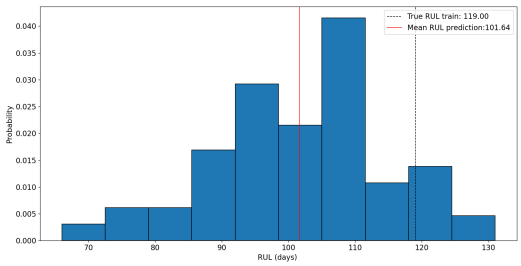


Figure 79: RUL=119 days Case 4

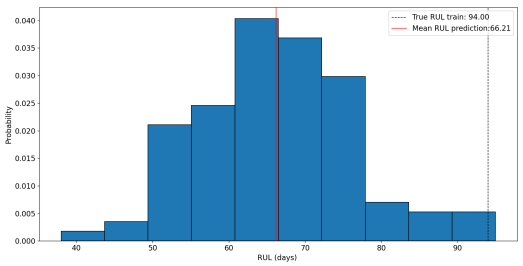


Figure 80: RUL=94 days Case 4

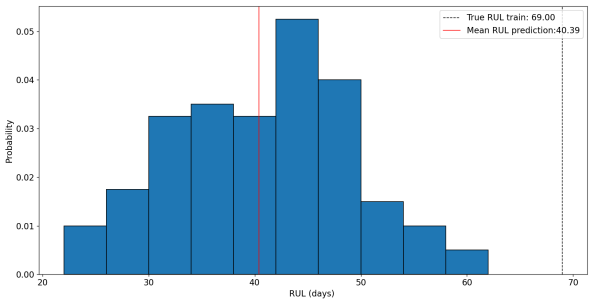


Figure 81: RUL=69 days Case 4

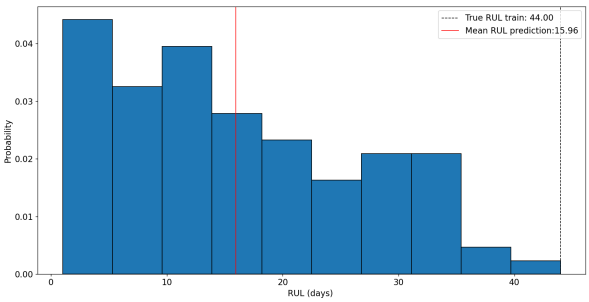


Figure 82: RUL=44 days Case 4

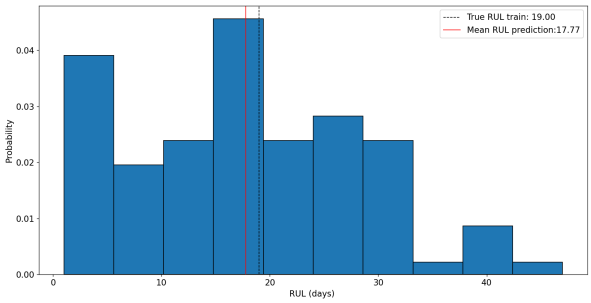


Figure 83: RUL=19 days Case 4

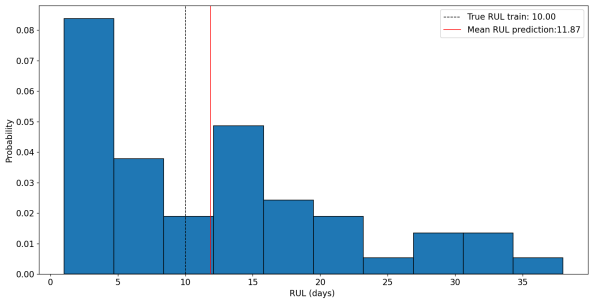


Figure 84: RUL=10 days Case 4

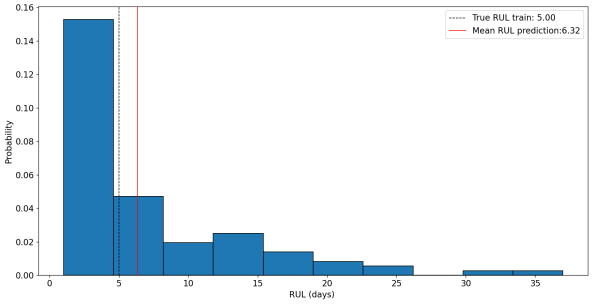


Figure 85: RUL=5 days Case 4

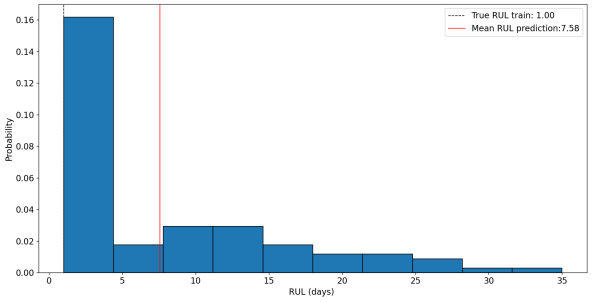


Figure 86: RUL=1 days Case 4

4.4.1 PDF RUL performance

Fig. 87 depicts the reliability curve for all the case study considered. We can observe how the reliability of the prediction made by our model is quite good for all the cases and more in detail we can see how the reliability of the RUL distribution related to the case study interested by an hydraulic group fault (CASE 1, CASE 2, CASE 3) is very similar and close to the ideal curve (red dashed line).

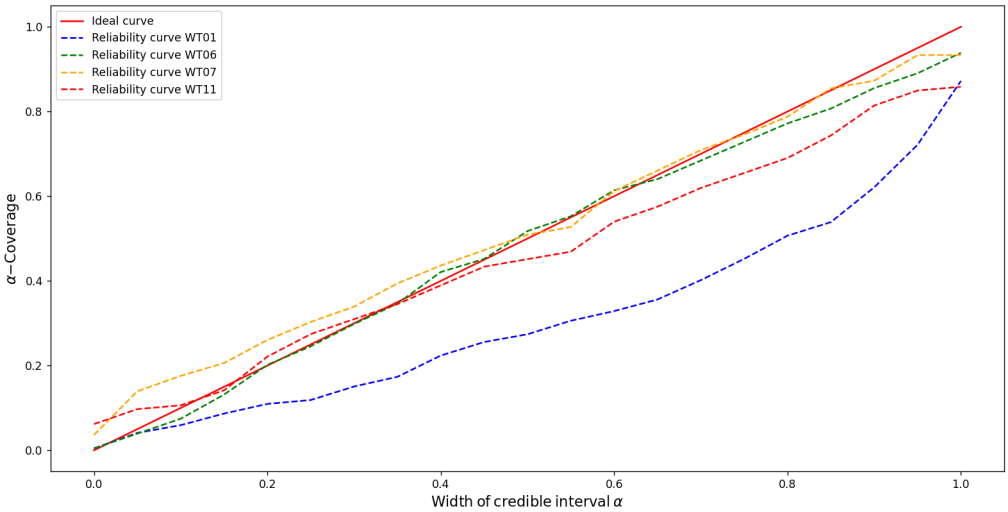


Figure 87: Reliability Curve

In the following Table, there is a summary of the performance for the PDF RUL prediction with a LSTM model.

Probabilistic RUL metrics

	CRPS	RS^{total}	RS^{under}	RS^{over}
Case 1	9.11	0.041	0.0409	0.00049
Case 2	7.08	0.034	0.02	0.014
Case 3	6.26	0.076	0.0725	0.00355
Case 4	15.05	0.22	0.22	$1.04e^{-5}$

From the Table, we can observe how the reliability of our model is very good and in particular this architecture tends to slightly underestimate the uncertainties.

5 Conclusion

Fifteen Open Source Datasets for Wind Turbines have been investigated in terms of features like type of information, number and type of parameters, time span, time sampling and type of components. These characteristics have to be evaluated in order to find the best dataset with enough and appropriate information that can be exploited for PHM and predictive maintenance planning.

The research publications reviewed, for some of the Open Source Datasets and for some Not Open Source Datasets, are divided into Diagnosis and Prognosis related work. The most common aim for the Diagnosis is to construct an appropriate Health Indicator which well represents the system's behavior to compare it with a predefined or adaptive threshold or to evaluate significant deviations from the standard conditions. The models developed, within the ML algorithm, vary in terms of performances evaluated through indices or graphically. In this sense, the choice of an appropriate algorithm depends on the objective of the study and on its performance. These algorithms are fed by condition monitoring measurements, mainly SCADA data, which are easier and cheaper to retrieve because they don't need special sensors, of mechanical, electrical and hydraulic parts. The most monitored components are gearbox, generator and hydraulic group because are those most prone to the failures. This report has also highlighted that the prognosis related work for Wind Turbines are quite limited and are focused on the RUL prediction considering ML algorithms.

Observing the content of each Open Source Dataset, it has been possible to understand the potential application of each of them for PHM and predictive maintenance purposes. All the Open Source Datasets

contain SCADA signals which allow to extract appropriate health indicator for the monitoring of the status of systems; but only the EDP dataset [91] reports the historical failures, which have affected the Wind Turbines. For this reason, only for the EDP [91] it is possible to perform a prognosis investigation, RUL prediction and a Root Causes analysis.

On the other hand, from this report, it can be seen as the availability of Open Source Datasets for Wind Turbines is quite limited and the most of them are not suitable for the prognosis and RUL Prediction.

The LSTM model introduced, in this work, has good performance for each case study considered. Beginning from the point-RUL prediction, we can give a first assessment on when it could be more convenient to perform a maintenance task or inspection.

Considering the SHAP method has been possible to identify which is the component that is the most significant for the Wind Turbines object of this research. In fact, according to the results obtained by the algorithm implemented, the most important component is the generator. In more detail, the driving feature for the RUL prediction has been the Generator RPM Avg.

In the end, implementing the same LSTM architecture with the Monte Carlo dropout, it has been possible to observe the evolution of the RUL prediction and also to evaluate the uncertainty linked to the model in order to take the best decisions on when it is the best period to perform a task.

6 Acknowledgements

At the end of this work, I would like to express my sincere gratitude to Prof. Matteo Davide Lorenzo Dalla Vedova and Prof. Gaetano Quattrocchi for their great availability and support through all these months. Their expertise and encouragement have been fundamental to write this thesis.

I would like, also, to express my deepest gratitude to Prof. Mihaela Mitici for the opportunity she gave me to conduct this research at the Utrecht University. Her guide, her suggestions, her feedback and her support have been invaluable to me and have played a crucial role for this work.

I would express my infinite gratitude to my father, my mother and my "little" brother Daniele Pio for their unconditional love, continuous encouragement and support through this fantastic journey and not only.

List of Figures

List of Tables

References

[1] Manisha Sawant, Sameer Thakare, A. Prabhakara Rao, Andrés E. Feijóo-Lorenzo, and Neeraj Dhanraj Bokde. A review on state-of-the-art reviews in wind-turbine- and wind-farm-related topics. *Energies*, 14(8), 2021.

[2] Moses Apunda and Benard Nyangoye. Challenges and opportunities of wind energy technology. 07 2017.

[3] Ruth Baranowski: NREL. Wind energy benefits. 2015.

[4] Trevor M. Letcher. *Wind Energy Engineering: A Handbook for Onshore and Offshore Wind Turbines*. 2015.

[5] GWEC Report: https://gwec.net/globalwindreport2023/.

[6] Samira Falani, Mario González, Fernanda Barreto, José Toledo, and Ana Torkomian. Trends in the technological development of wind energy generation. *International Journal of Technology Management Sustainable Development*, 19:43–68, 03 2020.

[7] Physics of wind turbines. `https://home.uni-leipzig.de/energy/energy-fundamentals/15.htm`.

[8] Hermann-Josef Wagner. Introduction to wind energy systems. December 2020.

[9] Wei Tong. *Fundamentals of wind energy*, volume 44. 2010.

[10] Wei Tong. *Wind power generation and wind turbine design*. WIT press, 2010.

[11] D. Griffith, Joshua Paquette, Matthew Barone, Andrew Goupee, Matthew Fowler, Diana Bull, and Brian Owens. A study of rotor and platform design trade-offs for large-scale floating vertical axis wind turbines. *Journal of Physics: Conference Series*, 753, 09 2016.

[12] RETScreen International. *WIND ENERGY PROJECT ANALYSIS*. 2004.

[13] Hermann-Josef Wagner. Introduction to wind energy systems. In *EPJ Web of Conferences*, volume 246, page 00004. EDP Sciences, 2020.

[14] Gürdal Ertek and Lakshmi Kailas. Analyzing a decade of wind turbine accident news with topic modeling. *Sustainability*, 13(22), 2021.

[15] Robert Errichello and Jane Muller. Design requirements for wind turbine gearboxes. *NASA STI/Recon Technical Report N*, 09 1994.

[16] Adam Ragheb and Magdi Ragheb. Wind turbine gearbox technologies. In *2010 1st international nuclear & renewable energy conference (INREC)*, pages 1–8. IEEE, 2010.

[17] Ideen Sadrehaghighi. Horizontal axis wind turbines (hawt) with case studies. 02 2022.

[18] Gustavo de Novaes Pires Leite, Alex Maurício Araújo, and Pedro André Carvalho Rosas. Prognostic techniques applied to maintenance of wind turbines: a concise and specific review. *Renewable and Sustainable Energy Reviews*, 81:1917–1925, 2018.

[19] Fausto Pedro García Márquez, Andrew Mark Tobias, Jesús María Pinar Pérez, and Mayorkinos Papaelias. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy*, 46:169–178, 2012.

[20] Milad Rezamand, Mojtaba Kordestani, Rupp Carriveau, David S.-K. Ting, Marcos E. Orchard, and Mehrdad Saif. Critical wind turbine components prognostics: A comprehensive review. *IEEE Transactions on Instrumentation and Measurement*, 69(12):9306–9328, 2020.

[21] Davide Astolfi, Ravi Pandit, Ludovico Terzi, and Andrea Lombardi. Discussion of wind turbine performance based on scada data and multiple test case analysis. *Energies*, 15(15), 2022.

[22] Pierre Tchakoua, René Wamkeue, Mohand Ouhrouche, Fouad Slaoui-Hasnaoui, Tommy Andy Tameghe, and Gabriel Ekemb. Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, 7(4):2595–2630, 2014.

[23] Jannis Tautz-Weinert and Simon Watson. Using scada data for wind turbine condition monitoring - a review. *IET Renewable Power Generation*, 11:382–394, 03 2017.

[24] Wenxian Yang, Peter J. Tavner, Christopher J. Crabtree, Y. Feng, and Y. Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2014.

[25] Ke-Sheng Wang, Vishal Sharma, and Zhenyou Zhang. Scada data based condition monitoring of wind turbines. *Advances in Manufacturing*, 2, 03 2014.

[26] Yingying Zhao, Dongsheng Li, Ao Dong, Dahai Kang, Qin Lv, and Li Shang. Fault prediction and diagnosis of wind turbine generators using scada data. *Energies*, 10(8), 2017.

[27] Adrian Stetco, Fateme Dinmohammadi, Xingyu Zhao, Valentin Robu, David Flynn, Mike Barnes, John Keane, and Goran Nenadic. Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, 133:620–635, 2019.

[28] Pere Marti-Puig, Alejandro Blanco-M, Juan José Cárdenas, Jordi Cusidó, and Jordi Solé-Casals. Effects of the pre-processing algorithms in fault diagnosis of wind turbines. *Environmental Modelling Software*, 110:119–128, 2018. Special Issue on Environmental Data Science and Decision Support: Applications in Climate Change and the Ecological Footprint.

[29] Chunzhen Yang, Jingquan Liu, Yuyun Zeng, and Guangyao Xie. Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model. *Renewable Energy*, 133, 10 2018.

[30] Wisdom Udo and Yar Muhammad. Data-driven predictive maintenance of wind turbine based on scada data. *IEEE Access*, 9:162370–162388, 2021.

[31] Ravi Pandit, Davide Astolfi, Jiarong Hong, David Infield, and Matilde Santos. Scada data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. *Wind Engineering*, 0(0):0309524X221124031, 2022.

[32] Caichao Zhu and Yao Li. Reliability analysis of wind turbines. In Kenneth Eloghene Okedu, editor, *Stability Control and Reliable Performance of Wind Turbines*, chapter 9. IntechOpen, Rijeka, 2018.

[33] Wenxian Yang, Richard Court, and Jiesheng Jiang. Wind turbine condition monitoring by the approach of scada data analysis. *Renewable Energy*, 53:365–376, 2013.

[34] Zhiwei Gao and Xiaoxu Liu. An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. *Processes*, 9(2), 2021.

[35] William Garlick, Roger Dixon, and Simon Watson. A model-based approach to wind turbine condition monitoring using scada data. 01 2009.

[36] Sikai Zhang and Zi-Qiang Lang. Scada-data-based wind turbine fault detection: A dynamic model sensor method. *Control Engineering Practice*, 102:104546, 2020.

[37] Christopher S. Gray, Roxane Koitz, Siegfried Psutka, and Franz Wotawa. An abductive diagnosis and modeling concept for wind power plants. *IFAC-PapersOnLine*, 48(21):404–409, 2015. 9th IFAC Symposium on Fault Detection, Supervision andSafety for Technical Processes SAFEPROCESS 2015.

[38] Philip Cross and Xiandong Ma. Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines. *International Journal of Automation and Computing*, 12:25–34, 02 2015.

[39] Hui Shao, Zhiwei Gao, Xiaoxu Liu, and Krishna Busawon. Parameter-varying modelling and fault reconstruction for wind turbine systems. *Renewable Energy*, 116:145–152, 2018. Real-time monitoring, prognosis and resilient control for wind energy systems.

[40] Hector Sanchez, Teresa Escobet, Vicenç Puig, and Peter Fogh Odgaard. Fault diagnosis of an advanced wind turbine benchmark using interval-based arrs and observers. *IEEE Transactions on Industrial Electronics*, 62(6):3783–3793, 2015.

[41] Satadru Dey, Pierluigi Pisu, and Beshah Ayalew. A comparative study of three fault diagnosis schemes for wind turbines. *IEEE Transactions on Control Systems Technology*, 23(5):1853–1868, 2015.

[42] Xiukun Wei, Michel Verhaegen, and Tim Engelen. Sensor fault detection and isolation for wind turbines based on subspace identification and kalman filter techniques. *International Journal of Adaptive Control and Signal Processing*, 24:687 – 707, 01 2009.

[43] Seongpil Cho, Zhen Gao, and Torgeir Moan. Model-based fault detection, fault isolation and fault-tolerant control of a blade pitch system in floating wind turbines. *Renewable Energy*, 120, 12 2017.

[44] Prasanna Tamilselvan, Pingfeng Wang, Shawn Sheng, and Janet Twomey. A two-stage diagnosis framework for wind turbine gearbox condition monitoring. *International Journal of Prognostics and Health Management*, 4:1–11, 05 2013.

[45] Donatella Zappalá, Peter J. Tavner, Christopher J. Crabtree, and Shuangwen Sheng. Side-band algorithm for automatic wind turbine gearbox fault detection and diagnosis. *IET Renewable Power Generation*, 8(4):380–389, 2014.

[46] Wenxian Yang, P.J. Tavner, and M.R. Wilkinson. Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train. *Renewable Power Generation, IET*, 3:1 – 11, 04 2009.

[47] Simon Jonathan Watson, Beth J. Xiang, Wenxian Yang, Peter J. Tavner, and Christopher J. Crabtree. Condition monitoring of the power output of wind turbine generators using wavelets. *IEEE Transactions on Energy Conversion*, 25(3):715–721, 2010.

[48] Kyusung Kim, Girija Parthasarathy, Onder Uluyol, Wendy Foslien, Shuangwen Sheng, and Paul Fleming. Use of SCADA Data for Failure Detection in Wind Turbines. ASME 2011 5th International Conference on Energy Sustainability, Parts A, B, and C:2071–2079, 08 2011.

[49] Yanhui Feng, Yingning Qiu, Christopher Crabtree, Hui Long, and P.J. Tavner. Use of scada and cms signals for failure detection diagnosis of a wind turbine gearbox. 01 2011.

[50] Mingzhu Tang, Qi Zhao, Huawei Wu, Ziming Wang, Caihua Meng, and Yifan Wang. Review and perspectives of machine learning methods for wind turbine fault diagnosis. *Frontiers in Energy Research*, 9, 2021.

[51] Wang Ke-Sheng Zhang, Zhen-You. Wind turbine fault detection based on scada data analysis using ann. *Advances in Manufacturing*, 23:2195–3597, 2014.

[52] A. Zaher, Stephen McArthur, David Infield, and Y. Patel. Online wind turbine fault detection through automated scada data analysis. *Wind Energy*, 12:574 – 593, 09 2009.

[53] Zhenyou Zhang. Comparison of data-driven and model-based methodologies of wind turbine fault detection with scada data. *European Wind Energy Association Conference and Exhibition 2014, EWEA 2014*, 01 2014.

[54] R.F.M. Brandao, José Carvalho, and Fernando Maciel-Barbosa. Neural networks for condition monitoring of wind turbines. volume 6, pages 1–4, 01 2010.

[55] Long Wang, Zijun Zhang, Huan Long, Jia Xu, and Ruihua Liu. Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics*, PP:1–1, 09 2016.

[56] Pramod Bangalore, Simon Letzgus, Daniel Karlsson, and Michael Patriksson. An artificial neural network based condition monitoring method for wind turbines, with application to the monitoring of the gearbox. *Wind Energy*, 20, 03 2017.

[57] Nassim Laouti, Nida Sheibat-Othman, and Sami Othman. Support vector machines for fault detection in wind turbines. *IFAC Proceedings Volumes*, 44(1):7067–7072, 2011. 18th IFAC World Congress.

[58] Pedro Santos, Luisa Villa Montoya, Anibal Reñones, Andrés Bustillo, and Jesús Maudes. An svm-based solution for fault detection in wind turbines. *Sensors (Basel, Switzerland)*, 15:5627–48, 03 2015.

[59] Keyan Liu, Weijie Dong, Huanyu Dong, Jia Wei, and Shiwu Xiao. A complex fault diagnostic approach of active distribution network based on sbs-sfs optimized multi-svm. *Mathematical Problems in Engineering*, 2020:1–12, 05 2020.

[60] Francesco Castellani, Davide Astolfi, and Francesco Natili. Scada data analysis methods for diagnosis of electrical faults to wind turbine generators. *Applied Sciences*, 11(8), 2021.

[61] Imad Abdallah, V Dertimanis, H Mylonas, Konstantinos Tatsis, Eleni Chatzi, N Dervili, K Worden, and Eoghan Maguire. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. In *Safety and Reliability–Safe Societies in a Changing World*, pages 3053–3061. CRC Press, 2018.

[62] Majdi Mansouri, Radhia Fezai, Mohamed Trabelsi, Hajji Mansour, Hazem Nounou, and Mohamed Nounou. Fault diagnosis of wind energy conversion systems using gaussian process regression-based multi-class random forest. *IFAC-PapersOnLine*, 55(6):127–132, 2022. 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2022.

[63] Chuan Li, René Sánchez, Grover Zurita, Mariela Cerrada, Diego Cabrera, and Rafael Vasquez. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing*, 76, 02 2016.

[64] Dahai Zhang, Liyang Qian, Baijin Mao, Can Huang, Bin Huang, and Yulin Si. A data-driven design for fault detection of wind turbines using random forests and xgboost. *IEEE Access*, 6:21020–21031, 2018.

[65] Hongshan Zhao, Huihai Liu, Wenjing Hu, and Xihui Yan. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, 127:825–834, 2018.

[66] Cheng Xiao, Zuojun Liu, Tieling Zhang, and Xu Zhang. Deep learning method for fault detection of wind turbine converter. *Applied Sciences*, 11(3):1280, 2021.

[67] Xianjin Luo and Xiumei Huang. Fault diagnosis of wind turbine based on elmd and fcm. *The Open Mechanical Engineering Journal*, 8:721–725, 12 2014.

[68] Edzel Lapira, Dustin Brisset, Hossein Davari Ardakani, David Siegel, and Jay Lee. Wind turbine performance assessment using multi-regime modeling approach. *Renewable Energy*, 45:86–95, 2012.

[69] Li Zhao, Zuowei Pan, Changsheng Shao, and Qianzhi Yang. Application of som neural network in fault diagnosis of wind turbine. In *International Conference on Renewable Power Generation (RPG 2015)*, pages 1–4, 2015.

[70] A. Hess, G. Calvello, and P. Frith. Challenges, issues, and lessons learned chasing the "big p". real predictive prognostics. part 1. pages 3610–3619, 2005.

[71] J.Z. Sikorska, M. Hodkiewicz, and L. Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5):1803–1836, 2011.

[72] Lotfi Saidi and Mohamed Benbouzid. Prognostics and health management of renewable energy systems: State of the art review, challenges, and trends. *Electronics*, 10, 11 2021.

[73] Dick Breteler, Christos Kaidis, Tiedo Tinga, and Richard Loendersloot. Physics based methodology for wind turbine failure detection, diagnostics & prognostics. *EWEA 2015 Annual Event*, 25, 2015.

[74] Junda Zhu, Jae Yoon, Bin Qiu, David He, and Eric Bechhoefer. Online condition monitoring and remaining useful life prediction of particle contaminated lubrication oil. 06 2013.

[75] Jinjiang Wang, Yuanyuan Liang, Yinghao Zheng, Robert X. Gao, and Fengli Zhang. An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable Energy*, 145:642–650, 2020.

[76] Pingfeng Wang, Prasanna Tamilselvan, Janet Twomey, and Byeng Dong Youn. Prognosis-informed wind farm operation and maintenance for concurrent economic and environmental benefits. *International Journal of Precision Engineering and Manufacturing*, 14(6):1049–1056, 2013.

[77] Lotfi Saidi, Jaouher Ben Ali, Mohamed Benbouzid, and Eric Bechhofer. An integrated wind turbine failures prognostic approach implementing kalman smoother with confidence bounds. *Applied Acoustics*, 138:199–208, 2018.

[78] Peter Matthews, Bindi Chen, and P.J. Tavner. Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition. *IET Renewable Power Generation*, 9, 07 2015.

[79] Bindi Chen, Peter Matthews, and P.J. Tavner. Wind turbine pitch faults prognosis using a-priori knowledge-based anfis. *Expert Systems with Applications: An International Journal*, 40:6863–6876, 12 2013.

[80] Andrew Kusiak and Anoop Verma. Analyzing bearing faults in wind turbines: A data-mining approach. *Renewable Energy*, 48:110–116, 2012.

[81] James Carroll, Sofia Koukoura, Alasdair McDonald, Anastasis Charalambous, Stephan Weiss, and Stephen McArthur. Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. *Wind Energy*, 22, 11 2018.

[82] Meik Schlechtingen and Ilmar Ferreira Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*, 25(5):1849–1875, 2011.

[83] Andrew Kusiak and Anoop Verma. A data-mining approach to monitoring wind turbines. *IEEE Transactions on Sustainable Energy*, 3(1):150–157, 2012.

[84] Jyh-Yih Hsu, Yi-Fu Wang, Kuan-Cheng Lin, Mu-Yen Chen, and Jenneille Hsu. Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning. *IEEE Access*, PP:1–1, 01 2020.

[85] Peng Guo, David Infield, and Xiyun Yang. Wind turbine generator condition-monitoring using temperature trend analysis. *IEEE Transactions on sustainable energy*, 3(1):124–133, 2011.

[86] Kevin Leahy, R. Hu, Ioannis Konstantakopoulos, Costas Spanos, Alice Agogino, and Dominic O' Sullivan. Diagnosing and predicting wind turbine faults from scada data using support vector machines. *International Journal of Prognostics and Health Management*, 9, 02 2018.

[87] Jürgen Herp, Mohammad H. Ramezani, Martin Bach-Andersen, Niels L. Pedersen, and Esmaeil S. Nadimi. Bayesian state prediction of wind turbine bearing failure. *Renewable Energy*, 116:164–172, 2018. Real-time monitoring, prognosis and resilient control for wind energy systems.

[88] Bill Chun Piu Lau, Eden Wai Man Ma, and Michael Pecht. Review of offshore wind turbine failures and fault prognostic methods. In *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, pages 1–5, 2012.

[89] Fangzhou Cheng, Liyan Qu, Wei Qiao, and Liwei Hao. Enhanced particle filtering for bearing remaining useful life prediction of wind turbine drivetrain gearboxes. *IEEE Transactions on Industrial Electronics*, 66(6):4738–4748, 2019.

[90] Xiaoliang Fan, Xiao Yang, Xinli Li, and Jianming Wang. A particle-filtering approach for remaining useful life estimation of wind turbine gearbox. 01 2015.

[91] EDP Dataset:`https://opendata.edp.com/open-data/en/data.html` (Last Access: 13/04/2023).

[92] La Haute Wind Born Dataset:`https://opendata-renewables.engie.com/` (Last Access: 13/04/2023).

[93] Vestas V52 Dataset:`https://data.mendeley.com/datasets/tm988rs48k` (Last Access: 13/04/2023).

[94] Yalova Dataset:`https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset` (Last Access: 13/04/2023).

[95] Sotavento Dataset:`https://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/` (Last Access: 13/04/2023).

[96] Eolos Dataset:`http://eolos.umn.edu/education`(Last Access: 13/04/2023).

[97] Inland and Offshore wind farm Dataset 1:`https://zenodo.org/record/5516552#.Y3KRTvfMLIV` (Last Access: 13/04/2023).

[98] Inland and Offshore wind farm Dataset 2:`https://zenodo.org/record/5516554#.Y3KRjvfMLIW`.

[99] Beberibe Dataset:`https://zenodo.org/record/1475197#.Y3KRoPfMLIU` (Last Access: 13/04/2023).

[100] GRC Dataset:`https://data.nrel.gov/submissions/45` (Last Access: 13/04/2023).

[101] GRC2 Dataset:`https://data.nrel.gov/submissions/56` (Last Access: 13/04/2023).

[102] Penmanshiel Dataset:`https://zenodo.org/record/5946808#.Y3KP1vfMLIU` (Last Access: 13/04/2023).

[103] Kelmarsh Dataset`https://zenodo.org/record/7212475#.Y3KQCvfMLIU` (Last Access: 13/04/2023).

[104] Xingchen Liu, Juan Du, and Zhi-Sheng Ye. A condition monitoring and fault isolation system for wind turbine based on scada data. *IEEE Transactions on Industrial Informatics*, 18(2):986–995, 2022.

[105] Fernando PG de Sá, Diego N Brandão, Eduardo Ogasawara, Rafaelli de C Coutinho, and Rodrigo F Toso. Wind turbine fault detection: A semi-supervised learning approach with automatic evolutionary feature selection. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 323–328. IEEE, 2020.

[106] Xiongjie Jia, Yang Han, Yanjun Li, Yichen Sang, and Guolei Zhang. Condition monitoring and performance forecasting of wind turbines based on denoising autoencoder and novel convolutional neural networks. *Energy Reports*, 7:6354–6365, 2021.

[107] Cyriana MA Roelofs, Marc-Alexander Lutz, Stefan Faulstich, and Stephan Vogt. Autoencoder-based anomaly root cause analysis for wind turbines. *Energy and AI*, 4:100065, 2021.

[108] Sarah Barber, Luiz Andre Moyses Lima, Yoshiaki Sakagami, Julian Quick, Effi Latiffianti, Yichao Liu, Riccardo Ferrari, Simon Letzgus, Xujie Zhang, and Florian Hammer. Enabling co-innovation for a successful digital transformation in wind energy using a new digital ecosystem and a fault detection case study. *Energies*, 15(15):5638, 2022.

[109] Fabrizio Bonacina, Eric Stefan Miele, and Alessandro Corsini. On the use of artificial intelligence for condition monitoring in horizontal-axis wind turbines. *IOP Conference Series: Earth and Environmental Science*, 1073(1):012005, sep 2022.

[110] Agnieszka Jastrzebska, Alejandro Morales Hernández, Gonzalo Nápoles, Yamisleydi Salgueiro, and Koen Vanhoof. Measuring wind turbine health using fuzzy-concept-based drifting models. *Renewable Energy*, 190:730–740, 2022.

[111] Effi Latiffianti, Shawn Sheng, and Yu Ding. Wind turbine gearbox failure detection through cumulative sum of multivariate time series data. *Frontiers in Energy Research*, 10, 5 2022.

[112] Maryna Garan, Khaoula Tidriri, and Iaroslav Kovalenko. A data-centric machine learning methodology: Application on predictive maintenance of wind turbines. *Energies*, 15(3):826, 2022.

[113] Haroon Rashid and Canras Batunlu. Anomaly detection of wind turbine gearbox based on scada temperature data using machine learning. *renewable energy*, 3:33, 2021.

[114] Hansi Chen, Hongzhan Ma, Xuening Chu, and Deyi Xue. Anomaly detection and critical attributes identification for products with multiple operating conditions based on isolation forest. *Advanced Engineering Informatics*, 46:101139, 2020.

[115] Zhen-You Zhang and Ke-Sheng Wang. Wind turbine fault detection based on scada data analysis using ann. *Advances in Manufacturing*, 2:70–78, 2014.

[116] Wei Teng, Chen Han, Yankang Hu, Xin Cheng, Lei Song, and Yibing Liu. A robust model-based approach for bearing remaining useful life prognosis in wind turbines. *IEEE Access*, 8:47133–47143, 2020.

[117] He Liu, Wanqing Song, Yuhui Niu, and Enrico Zio. A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes. *Mechanical Systems and Signal Processing*, 153:107471, 2021.

[118] Faris Elasha, Suliman Shanbr, Xiaochuan Li, and David Mba. Prognosis of a wind turbine gearbox bearing using supervised machine learning. *Sensors*, 19(14):3092, 2019.

[119] Naipeng Li, Pengcheng Xu, Yaguo Lei, Xiao Cai, and Detong Kong. A self-data-driven method for remaining useful life prediction of wind turbines considering continuously varying speeds. 165:1, 2022.

[120] Lixiao Cao, Zheng Qian, and Yan Pei. Remaining useful life prediction of wind turbine generator bearing based on emd with an indicator. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pages 375–379. IEEE, 2018.

[121] Huimin Zhao, Haodong Liu, Yang Jin, Xiangjun Dang, and Wu Deng. Feature extraction for data-driven remaining useful life prediction of rolling bearings. *IEEE transactions on instrumentation and measurement*, 70:1–10, 2021.

[122] Jianing Man, Zijun Zhang, and Qiang Zhou. Data-driven predictive analytics of unexpected wind turbine shut-downs. *IET Renewable Power Generation*, 12(15):1833–1842, 2018.

[123] Xiaohang Jin, Zijun Que, Yi Sun, Yuanjing Guo, and Wei Qiao. A data-driven approach for bearing fault prognostics. *IEEE Transactions on Industry Applications*, 55(4):3394–3401, 2019.

[124] Yingying Zhao, Dongsheng Li, Ao Dong, Jiajia Lin, Dahai Kang, and Li Shang. Fault prognosis of wind turbine generator using scada data. In *2016 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2016.

[125] Boualem Merainani, Sofiane Laddada, Eric Bechhoefer, Mohamed Abdessamed Ait Chikh, and Djamel Benazzouz. An integrated methodology for estimating the remaining useful life of high-speed wind turbine shaft bearings with limited samples. *Renewable energy*, 182:1141–1151, 2022.

[126] Wei Teng, Xiaolong Zhang, Yibing Liu, Andrew Kusiak, and Zhiyong Ma. Prognosis of the remaining useful life of bearings in a wind turbine gearbox. *Energies*, 10(1):32, 2016.

[127] Yubin Pan, Rongjing Hong, Jie Chen, Jaskaran Singh, and Xiaodong Jia. Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion. *Mechanism and machine theory*, 137:509–526, 2019.

[128] Sharaf Eddine Kramti, Jaouher Ben Ali, Lotfi Saidi, Mounir Sayadi, and Eric Bechhoefer. Direct wind turbine drivetrain prognosis approach using elman neural network. In *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 859–864. IEEE, 2018.

[129] François Chollet. *Deep Learning with Python*. Manning, November 2017.

[130] Mikael Bodén. A guide to recurrent neural networks and backpropagation. 12 2001.

[131] Filippo Maria Bianchi, Enrico Maiorino, Michael Kampffmeyer, Antonello Rizzi, and Robert Jenssen. *Recurrent Neural Network Architectures*, pages 23–29. 11 2017.

[132] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[133] Ioannis E Livieris, Emmanuel Pintelas, and Panagiotis Pintelas. A cnn–lstm model for gold price time-series forecasting. *Neural computing and applications*, 32:17351–17360, 2020.

[134] Beakcheol Jang, Myeonghwi Kim, Gaspard Harerimana, Sang-ug Kang, and Jong Wook Kim. Bi-lstm model to increase accuracy in text classification: Combining word2vec cnn and attention mechanism. *Applied Sciences*, 10(17):5841, 2020.

[135] LSTM cell: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[136] Alex Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Studies in computational intelligence. Springer, Berlin, 2012.

[137] Federico Landi, Lorenzo Baraldi, Marcella Cornia, and Rita Cucchiara. Working memory connections for lstm. *Neural Networks*, 144:334–341, 2021.

[138] Roland Bolboacă and Piroska Haller. Performance analysis of long short-term memory predictive neural networks on time series data. *Mathematics*, 11(6):1432, 2023.

[139] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 2022.

[140] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015.

[141] Yuhan Bai. Relu-function and derived function review. In *SHS Web of Conferences*, volume 144, page 02006. EDP Sciences, 2022.

[142] Activation functions in neural networks. `https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6`.

[143] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[144] Shaeke Salman and Xiuwen Liu. Overfitting mechanism and avoidance in deep neural networks. *arXiv preprint arXiv:1901.06566*, 2019.

[145] H Tilaver, Mustafa Salti, Oktay Aydoğdu, and Evrim Kangal. Deep learning approach to hubble parameter. *Computer Physics Communications*, 261:107809, 04 2021.

[146] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009. Risk Acceptance and Risk Communication.

[147] Ronald Seoh. Qualitative analysis of monte carlo dropout, 2020.

[148] Venkat Nemani, Luca Biggio, Xun Huan, Zhen Hu, Olga Fink, Anh Tran, Yan Wang, Xiaoping Du, Xiaoge Zhang, and Chao Hu. Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial. *arXiv preprint arXiv:2305.04933*, 2023.

[149] Alex Labach, Hojjat Salehinejad, and Shahrokh Valaee. Survey of dropout methods for deep neural networks. *arXiv preprint arXiv:1904.13310*, 2019.

[150] Amgad Muneer, Shakirah Mohd Taib, Sheraz Naseer, Rao Faizan Ali, and Izzatdin Abdul Aziz. Data-driven deep learning-based attention mechanism for remaining useful life prediction: Case study application to turbofan engine analysis. *Electronics*, 10(20):2453, 2021.

[151] Zuo H Jiang H Li P Li X Liu Y, Liu Z. A dlstm-network-based approach for mechanical remaining useful life prediction. *Sensors (Basel)*, 2022.

[152] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[153] Mihaela Mitici, Ingeborg de Pater, Anne Barros, and Zhiguo Zeng. Dynamic predictive maintenance for multiple components using data-driven probabilistic rul prognostics: The case of turbofan engines. *Reliability Engineering & System Safety*, 234:109199, 2023.

[154] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[155] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

[156] Mihaela Mitici, Ingeborg de Pater, Anne Barros, and Zhiguo Zeng. Dynamic predictive maintenance for multiple components using data-driven probabilistic rul prognostics: The case of turbofan engines. *Reliability Engineering System Safety*, 234:109199, 2023.

[157] Ingeborg de Pater and Mihaela Mitici. Novel metrics to evaluate probabilistic remaining useful life prognostics with applications to turbofan engines. *PHM Society European Conference*, 2022.