

Article

Anomaly Detection of Wind Turbine Gearbox based on SCADA Temperature Data using Machine Learning

Haroon Rashid ^{1,*}  and Canras Batunlu ²

¹ Sustainable Environment and Energy Systems, Middle East Technical University, Kalkanli Guzelyurt, 99738, Turkey; haroon@metu.edu.tr

² Department of Electrical Engineering, Middle East Technical University, Kalkanli Guzelyurt, 99738, Turkey; cbatunlu@metu.edu.tr

* Correspondence: haroon@metu.edu.tr

Abstract: Wind energy is becoming an essential source of power for countries which have the aim to reduce greenhouse gases emission and mitigate the effects of global warming. The Wind Turbines (WTs) installed around the globe is increasing significantly every year. The dramatic increase in wind power has encountered quite a few challenges, among which the major issues are availability and reliability. The unexpected failure in WTs Gearbox (GB) ultimately increases the Operation and Maintenance (O&M) cost. The identification of faults in the earlier stages before it turns to catastrophic damage to other components of WT is curcial. This research deals with the prediction of WT failures by using a Supervisory Control and Data Acquisition (SCADA) system. The main aim is to forecast the temperature of the WTs GB to predict impending overheating of the GB at an early stage. The earlier prediction will help to optimize the maintenance period and to save maintenance expenses and, even more important, to generate warnings in due time to avoid major damages or even technical disasters. In the proposed method we compared six different machine learning (ML) models based on error and accuracy of prediction. The bagging regressor is the best ML model, which results in the mean square error of 0.33 and R^2 of 99.8 on training data. The bagging regressor is then used to predict the fault in the WT GB, which detected the anomalous behavior of WT GB 59 days earlier than the actual failure. This model also detects the extremely unusual behavior of the GB 9 days earlier than a complete failure.

Keywords: Wind turbine; Renewable energy; Wind energy; Machine learning; Prediction algorithm; Gearbox

1. Introduction

1.1. Background

Renewable energy use is on the rise around the world. Renewable energy is the energy that is obtained from the inexhaustible natural source. This alternative source of power includes wind, solar, geothermal, tidal, hydroelectric, and biofuels [1]. The rapid growth in population and economic expansion imposes higher renewable energy demand [2]. The higher demand for renewable energy leads the researcher, economist, and environmentalists to contribute to this field. On the other hand, global warming is a crucial challenge for the world. The greenhouse gas emission is rapidly increasing due to the drastic increase in population and industrialization. Many countries around the globe are in the transition of renewable energy from fossil fuels in order to combat climate change. According to the International Renewable Energy Agency (IRENA), in the year 2016, solar and wind energy received 94 percent of the total investment in renewable energy [3].

1.2. Wind Energy

Wind energy is the conversion of the movement in the air to electrical energy with the help of the WTs. The wind moves around the earth due to the pressure and temperature difference. The wind energy is produced either in offshore (at sea) or onshore (on sea). Wind power has been expanding by leaps and bounds in the last , 20 years over the world [4]. According to the Global Wind Energy Council (GWEC) global wind statistics 2019 the installed wind power is around 651 GW by the end of 2019 [5]. The global cumulative wind power installing increased exponentially from the year 2001 to 2019 as shown in Figure 1.

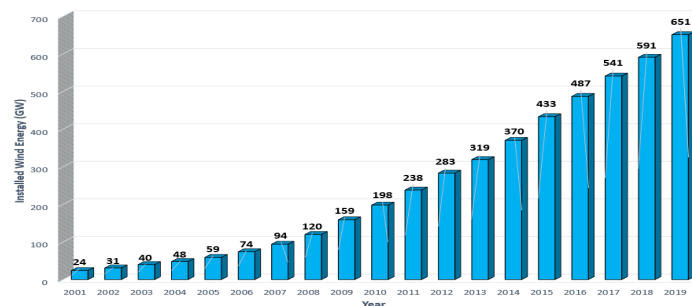


Figure 1. Historical development in wind energy installation [This figure was made in Excel using data from reference [4]

Wind Turbine (WT) is a sophisticated electrical-mechanical system. As compared to traditional turbines that are used in power plants WTs are more prone to failure due to harsh operating conditions. The faults in WTs include a long term or temporary faults depend on severity. The GB expected lifetime is usually 20 years but due to harsh conditions, it fails earlier and downtime is higher as compared to other subsystems. The earlier failure in GB cost wind energy high maintenance cost.

1.3. Challenges

The number of WTs installed all over the world is increasing day by day and the numbers are expected to be 2,02,1000 by 2050¹ [6]. The rapid expansion of wind energy has been encountering many challenges among these; the major challenges are the high maintenance cost and poor reliability. The reliability of WT is susceptible to the failure of its subsystems and degradation.

The undesirable GB reliability is a great challenge to the utilization of wind power more efficiently. The unplanned operation and maintenance have a great impact on the cost of wind power. Table 1 shows the replacement cost of different subsystems in WT in Demark [6]. The GB cost of replacement is more than twice of that of the blade or generator.

Table 1: Subsystem replacement cost of a wind turbine [6]

Name of Subsystem	Cost (DKK)
Gearbox	3,937,000
Generator	3,937,000
Blade	1,968,000

1.4. Literature Review

Reliability is paramount in smooth production of wind energy. This development in WTs results in heavier and larger structures, which also affect the failure rate of WTs.

¹ These numbers are inferred based on the predicted global installation capacity of wind power and the average rated power per WT

These failure rates vary for the same size WT between the onshore and offshore farms [7].

Figure 2 shows the failure rates of different subassemblies and its downtime after failure. The results in Figure 2 show that the lower the subassembly's reliability, the longer is the downtime of the corresponding sub-assemblies [8].

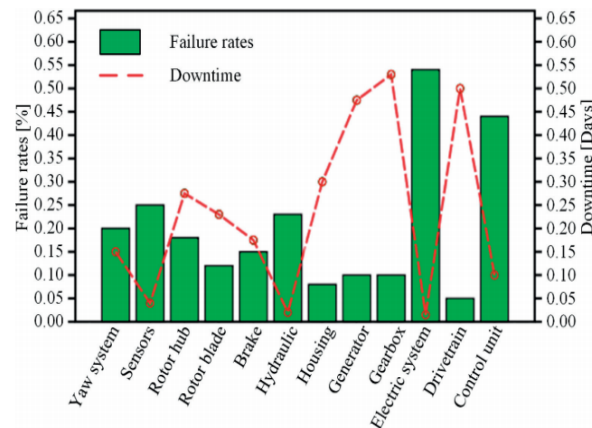


Figure 2. Wind turbine different sub-assemblies downtime and failure rates [8]

WTs GBs are designed according to standards of International Electrotechnical Commission (IEC) for twenty years of a lifetime, but most of the WT GBs do not meet this design lifetime. The wind industries are very concerned with the reliability of GB because of the share almost 13% of total WT cost, and the replacement cost is very high as compared to other components. The authors in [9] pointed out that in terms of WT availability, GB alone shares one-third of the total loss.

The operating conditions of WT are frequently changing in order to meet the grid condition which effects the WT GB reliability. The GBs in WTs suffer severe transient conditions during the low and high ride through, start, and shutdown [10]. The wind industry nowadays uses various condition monitoring (CM) techniques. These techniques include oil monitoring [11,12] acoustic monitoring [13,14] thermal monitoring [15,16], eddy current monitoring [17], and SCADA data analysis [18,19]. Among these mentioned CM techniques electrical parameter, temperature and vibration monitoring is a commonly used method for the health monitoring of WTs.

The artificial intelligence-based CM has recently emerged for fault diagnosis in WTs. Researchers mostly used Artificial Neural Network (ANN) as an effective tool in complicated WTs fault pattern recognition. The authors in [20] used Back Propagation (BP) ANN to detect four different faults in WTs GB. The results of this paper are satisfying but completely ignore the drawbacks of BP neural networks. References [21,22] used data mining and fuzzy algorithms approach to detect the fault in WTs GB and blade. The SCADA based data mining and multi-agent techniques are also widely used techniques. In these techniques the SCADA data is used to predict the health status of WTs using different data mining algorithms and agents [23,24]. These algorithms are boosting wrapper with genetic search, tree algorithm and wrapper with best-first search.

The regression-based model predicts the dependent variable or output based on the inputs when we assumed the machine or component working in its reasonable condition. Figure 3 [25] show the distribution of failure rates in three stages in which represents the Weibull distribution. The normal behavior modeling based on regression can be defined in various complexity levels. For example, the modeling of the output power of WT using the wind speeds. In this case, the whole WT is considered a black box. So this prediction depends on the wind speed, which may not be the same in another location. The authors in [26] made a comparison of two unsupervised methods Gaussian Mixture Model and Self Organizing Maps with a regression-based model. The results of this

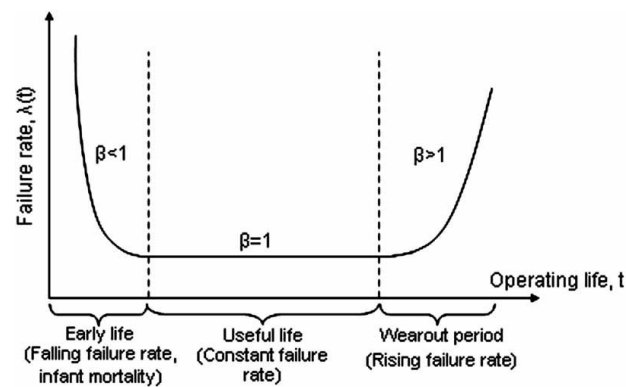


Figure 3. Machinery failure rate with time [25]

study revealed that the GMMs model is more suitable for prediction. Laouti et al [27] used SVMs for fault detection in the WT blades using a radial basis kernel. They trained their model using the data acquired from different sensors on WTs. The authors in [28] done a comparison of autoregressive NN, regression, and NNs to detect the WT bearing temperature using the SCADA data. The results of this comparison study show that the NN approach is best among these three. Wang et al. [29] designed a regression-based deep neural network using SCADA data. They evaluate the accuracy of the model using the standard deviation of the absolute error (SDAPE) and mean absolute percentage error (MAPE). They used the data from 6 different wind farms, and the deep NNs gave a MAPE between 2.9 and 14.25. Orzco et al. [30] used an extensive SCADA data of 614 WTs from 6 different wind farms to build a regression model. The model learns the relation between independent variables such as power output and outdoor temperature and dependent variable the temperature of the components. The multiple design type of regression models like the random forest, linear regression, and they evaluate these models using root mean squared error.

Verma et al. [31] used the data collected from SCADA to detect the failure in the generator brush using classification. They used different techniques for parameter selection like the wrapper method, boosting tree, and chi-square filter. The result of this study shows that the classification accuracy increases from 82.1% to 97.1% using the boosting model. They predicated the brush failure 12 hours ahead.

Abdullah et al.[32] used 64 channel wind farm data collected from 48 WTs over 12 months to detect the abnormalities in different operations. They used decision trees to detect faults. The authors selected the features manually and did not use any dimension reduction techniques. They used bootstrap sampling to train the CART tree. The results of this study show that the proposed method is easy to interpret the faults in wind farms.

1.5. Novelty of this research

The aim of this research was to explore and investigate the application of machine learning (ML) models for WT gearbox condition monitoring based on SCADA data. While doing so, different ML models were taken into account to select the best model based on performance. The particular focus while selecting these models was to choose the model with minimum error and few false alarms. The main objective of this study is to apply advance fault detection technologies. The fault in WT GB will be detected based on the novel ML algorithms using the GB temperature data collected from SCADA.

This paper consists of 4 sections. Section 1 briefly describes the background, motivation, challenges and briefly summarizes the research published for the fault detection of WT GB from the existing literature. Section 2 describes the methodology and all the approaches used to predict the fault in the WT GB. Section 3 summarizes the results obtain from the analyze of SCADA data for fault prediction. Section 4 contains the contribution and conclusion of this work. This chapter further highlights the future work suggestion.

2. Methodology

This section discusses methodology utilized in this study. The complete flow diagram of the methodology of this work is shown in Figure 4.

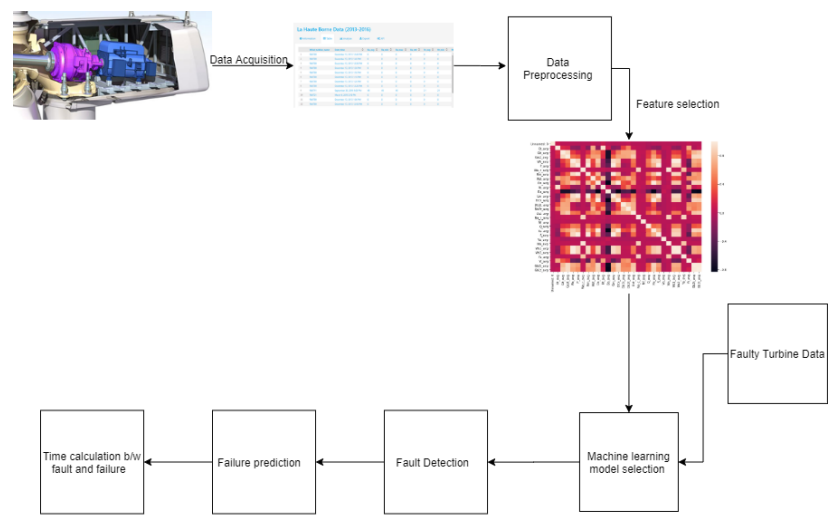


Figure 4. Methodology of this study

Firstly the SCADA data is collected from wind farm then the data is preprocessed before further analysis. The useful features are extracted for analysis before the ML model selection. A comparison of different ML models is done to select the best ML model for fault prediction using the collected historical SCADA data. After the selection of the best model, the faulty turbine data is given as input to detect the fault in the WTs. Once the fault is detected the failure is predicted and the time between the detection and actual failure is calculated.

2.1. Wind turbine SCADA data collection

Supervisory Control and Data Acquisition (SCADA) system is a basic tool in the WT to control and monitor different parameters. It collects all the information about WT in a single point of control. The SCADA system usually collects the data from the wind farms in 10 minutes interval in order to reduce the bandwidth of data [18]. All most all the modern wind farms have the SCADA systems which were used to record different important parameters.

The WT SCADA data for this study was collected from the onshore wind farm La Haute Borne-Vaudeville-le-Haut in France. This wind farm consists of four WTs [33]. The Senvion company manufactures all 4 WTs. The technology of these WTs is highly reliable and well equipped with DC pitch system and doubly-fed induction generator frequency converter. The nominal output power of each turbine is 2 MW. The technical details of this turbine are shown in Table 2.

Table 2: Technical details of wind turbines

Nominal power	2050 kW
Cut-in wind speed	3.5 m/s
Cut-out wind speed	25 m/s
Nominal wind speed	14.5 m/s
Operating temperature range	-20C to +35C
Rotor diameter	82 m
Rotor area	5281 m ²
Rotor blade length	40 m
Nominal frequency	50/60 Hz
Hub height	80 m

The collected parameters are classified into performance parameter (rotor speed, power, generator speed, etc.), controllable parameters (torque and pitch angle, etc.) and non-controllable parameters (temperature and wind speed, etc.). The SCADA system record the average, standard deviation maximum and minimum values of all parameters. The historical SCADA data from the year 2014-2018 was collected consist of 34 parameters which were recorded on 10 minutes interval.

2.2. Data preprocessing

Data pre-processing is crucial in ML because "Better data beats fancier algorithms". The SCADA data used for this study have extensive amount of WT parameters. The dataset contain different outliers for example, mismatch in time and date stamp, negative power values, abnormal wind speed, pitch angle mismatch and out of range values. Data processing is a technique which mostly uses in ML to convert the data into desired and meaningful.

2.2.1. SCADA data rescaling

The collected SCADA has different parameters of varying scales. The different scaled data is very difficult to visualization and this degrades the ML predictive performance as well as slow down the prediction. The data is rescaled using Python to the same scale. The parameters are rescaled in the range of 0 to 1 and this is called normalization. The data is rescaled using the MinMaxScaler class from scikit-learn. The rescaled value is calculated using the following equation

$$W' = \frac{w - \min(w)}{\max(w) - \min(w)} \quad (1)$$

Where W' is rescaled value and w is the original value, $\max(w)$ is the maximum value in features and $\min(w)$ is the minimum value in features.

2.2.2. Missing SCADA Values

The collected SCADA WT data contain some missing values. The reason behind this is the sensor failure, corrupt data, and incomplete extraction. Null values are one of the significant challenges in data science while predicting the output. In this research work, the Python SciKit Learn libraries used to remove the missing value from the collected data. The rows with null values are removed as our dataset has enough parameters for analysis. Deleting the null values will reduce the bias from our prediction. The removal of missing data will increase the accuracy of the model.

2.3. Feature Selection

Feature selection is the method to reduce the number of input parameters while developing a ML predictive model. The feature selection process is critical because it will improve the efficiency of the model. In this study, we use the Recursive feature elimination (RFE) method to choose the most desirable input to predict the output of GB temperature. RFE aims to eliminate the attributes that are not dependent on each other, and are linear correlation between them [34]. RFE is not a deterministic process to specify precisely which features should be acquired for the learning process. In other words, each time we run this method, different outcomes may be observed.

The flow chart of RFE method is illustrated in Figure 5. Firstly the features are selected after that based on the training results the importance of every feature is determined. The features with less contribution are removed and then again the model accuracy is calculated. These steps are repeated until we get the best model. This will continue until the maximum number of folds is reached.

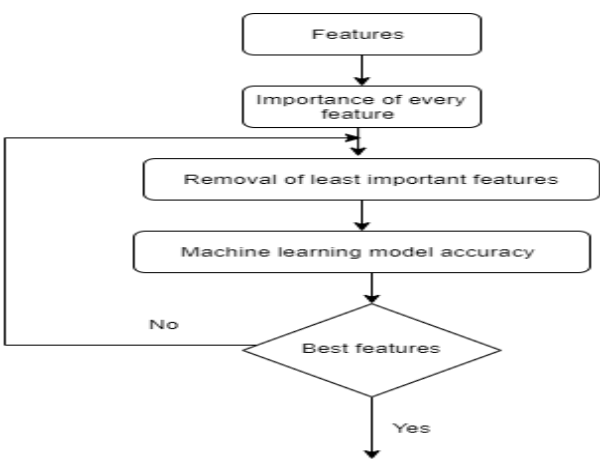


Figure 5. Recursive Feature Elimination flow chart

2.4. The proposed method for gearbox fault detection

The main theme behind this research work is to predict the GB fault using the GB temperature data. The SCADA data is collected from 4 different WTs in the same wind farm. After the preprocessing of data, we got 169322 samples. The data is divided into training and test data. The workflow of fault prediction in the GB is shown in Figure 6.

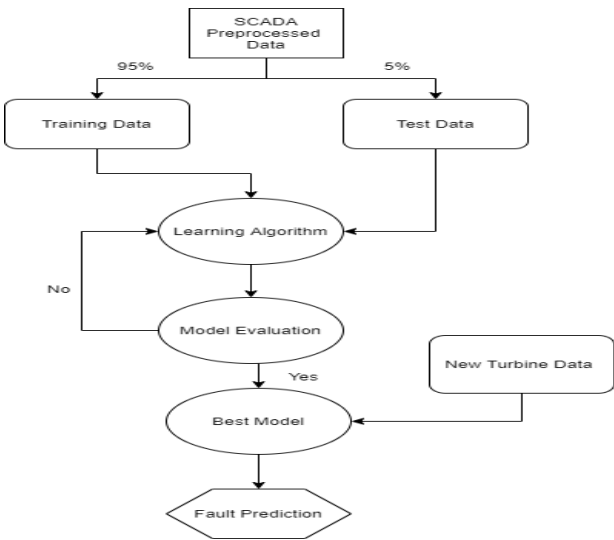


Figure 6. Gearbox fault prediction work flow

2.4.1. Machine learning model selection

The ML model selection is one of the toughness jobs in predicting GB temperature. In this study, we compare the different ML models on the basis of the performance as shown in Figure 7. The main points we consider during our predictive model selection are:

1. The selected model should better predict the future values (unseen values) of WT GB temperature.
2. The selected model should be a minimum error while predicting the future value of GB temperature.
3. The selected model should take minimum computation time while predicting future value.

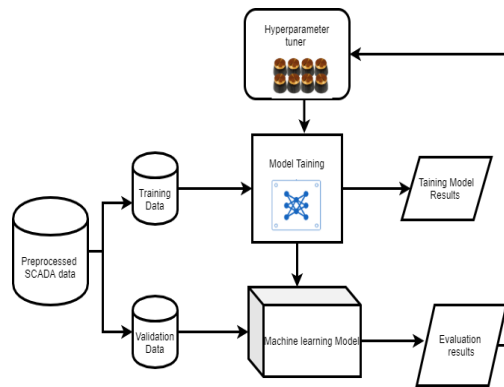


Figure 7. Machine learning model selection

2.4.2. Training model

The ML training model is shown in Figure 8. Firstly we feed the input parameters, which consist of 26 variables. These variable includes wind speed, rotor speed, power, generator speed, etc. as shown in Figure 8. The WT GB is predicted based on these input variables.

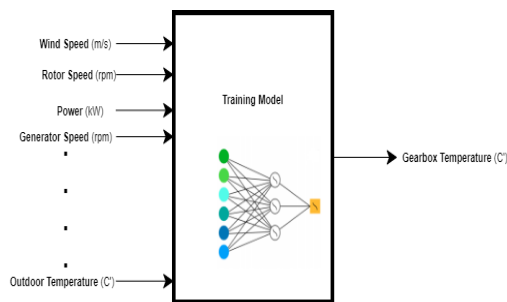


Figure 8. Machine learning model selection

2.4.3. Machine learning Algorithms

In this research work, we used six different ML algorithms. The algorithms are compared for the SCADA data based on the predicted error. The ML models used in this study are the following:

Random Forest:

The random forest algorithm is a supervised learning algorithm that is made up of different decision trees. The accuracy of the model will vary on the number of trees [35].

The random forest algorithm works by constructing different trees, and the predicted output is the mean of all decision trees. The main advantage of random forest over other model is that it prevents overfitting problem in ML because it randomly selects a sample from the data set and generate an output.

Linear Regression:

The linear regression is a supervised ML algorithm that predicts the targeted output based on the input variables. This model is mostly used to find the linear relation between the inputs and output variables. Mathematically speaking the concept behind linear regression is the straight-line equation :

$$y = mx + c \quad (2)$$

Where y is the targeted output , x is the input parameters , m is the slope of linear line and c is y intercept

K-Nearest Neighbor:

The K-Nearest Neighbor is a simple ML algorithm which predicts the targeted output based on the distance functions. The distance function for continuous value are :

$$Minkowski = \left[\sum_{i=1}^k (|x_i - y_i|)^q \right]^{\frac{1}{q}} \quad (3)$$

$$Euclidean = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (4)$$

$$Manhattan = \sum_{i=1}^k |x_i - y_i| \quad (5)$$

Multilayer perceptron : The multilayer perceptron is a neural network algorithm, which consists of many sample layers. The layers are the input layer, hidden layer, and output layer [36]. This algorithm mostly use in finding the non-linear relationship between the inputs and target output [37]. The input layer consists of the input parameter like wind speed, rotor speed, and outdoor temperature, etc. The hidden layer consists of different neurons that help in training the neural network. The output layer consists of the WT GB temperature.

In this research for the training of a multilayer perceptron error backpropagation algorithm is used. In this algorithm, the output and actual output is compared until we got the least error [38]. To transfer the input k_j into the output o_j the log sigmoid function is used which is represented as follow:

$$f(k_j) = \frac{1}{(1 + \exp(-k_j))} \quad (6)$$

Adaptive Boosting algorithm: Adaptive Boosting is a meta ML algorithm that is combined with different types of other algorithms to boost performance. In this algorithm, the data output of other weaker algorithms is merged to one weighted sum, which is then boosted output. The AdaBoost (Adaptive Boosting) algorithm is very sensitive to outliers and the data with noise [39]. The advantage of this algorithm over others is that it is less susceptible to overfitting problems. The AdaBoost algorithm, unlike other NN and SVMs, select only the features that improve the accuracy of the model [40].

Bagging algorithm : The bagging algorithm in ML is an ensemble algorithm which is proposed by Breiman [41]. The bagging algorithm uses the technique known as bootstrap which is a statistical method to predict the output base on the average estimation of input sample data [42]. The flow diagram of this regressor is shown in Figure 9. The main advantage of this algorithm over other discuss algorithms it reduces the variance which helps to minimize the overfitting problem .

2.4.4. Machine learning model comparison and evaluation

In this work to compare and validate the above mentioned machine-learning algorithms following metrics are used:

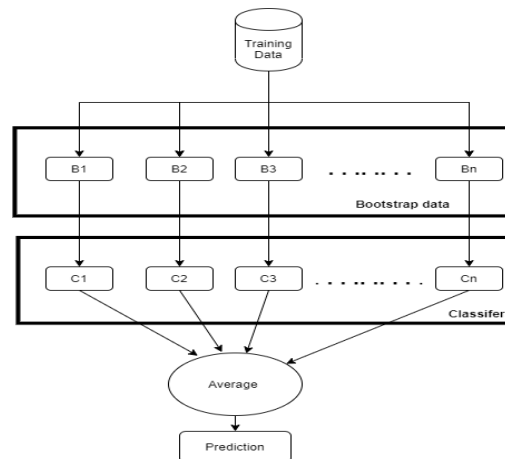


Figure 9. Bagging algorithm flow chart

Coefficient of determination:

The coefficient of determination (R^2) metric measures how good the algorithm to the regression of the equation by calculating the relationship between the input parameters and targeted variables. This metric is also called fitting degree of regression as well. Mathematically this is expressed as:

$$R^2 = 1 - \frac{\sum_{m=1}^K (Gbt_p(m) - Gbt_a(m))^2}{\sum_{m=1}^K Gbt_a(m) - Gbt'_a(m))^2} \quad (7)$$

where Gbt_p is gearbox predicted temperature, Gbt_a is gearbox actual temperature, K show the number of datapoints which is $K=1,2,3,\dots,m$ and Gbt'_a is the mean actual gearbox temperature.

The maximum value of R^2 is 1. If the value is closer to 1 the model is good and fitting to the regression line to the targeted variable [43]. If the value is less (less than 0.5) the model is not good for this data.

Mean Square Error :

The mean square error (MSE) estimated the average square difference between the actual value and predicted value. The MSE measures the quality of the regression ML algorithm. The value closer to zero is better. If the predicted GB temperature is $Gbt_p(m)$ and actual GB temperature is $Gbt_a(m)$ and total data samples are K mathematically MSE is expressed as:

$$MSE = \frac{1}{K} \sum_{m=1}^K ((Gbt_a(m) - Gbt_p(m))^2) \quad (8)$$

Mean Absolute Error :

The mean square error which denoted as MAE. The MAE calculate the average between the absolute actual value and the absolute predicted value. Mathematically this error is represented as :

$$MAE = \frac{1}{K} \sum_{m=1}^K |Gbt_p(m) - Gbt_a(m)| \quad (9)$$

Where $Gbt_p(m)$ is the predicted gearbox temperature, $Gbt_a(m)$ is actual gearbox temperature and K is total data samples.

3. Results and Discussion

The first objective is to deal with the SCADA data of WT and to remove the unnecessary information which is not necessary for health assessment of GB of the WT. The

second objective is to the selection of ML model and train model using the healthy WT GB temperature data. The most accurate model is selected based on the results. The third one is to early predict the GB temperature of the in the unhealthy data of the second WT.

3.0.1. Removal of missing values

The SCADA dataset contain lot of missing value which are not recorded. We first plot the missing value as show in Figure 10. After complete analysis we remove that row which contain more than one missing values. We did not replace the not a number (NaN) values with any statistical approach like mean, mode etc because it can effected negatively on our ML model

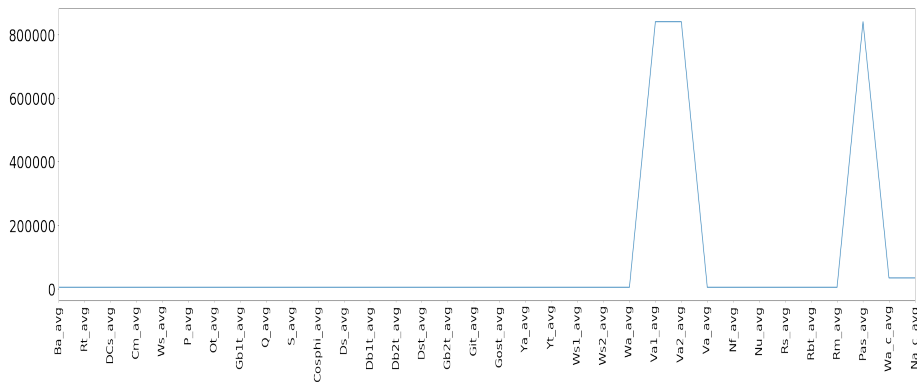


Figure 10. Missing values in SCADA data

3.1. Feature selection

The feature selection is done using the correlation statistics as shown in Figure 11. The highly related input variables to the targeted output (Gearbox output temperature) are selected.The value of correlation is represented in the range of 0 to 1. The input variables with a value of less than 0.5 are removed and greater than 0.5 is selected.

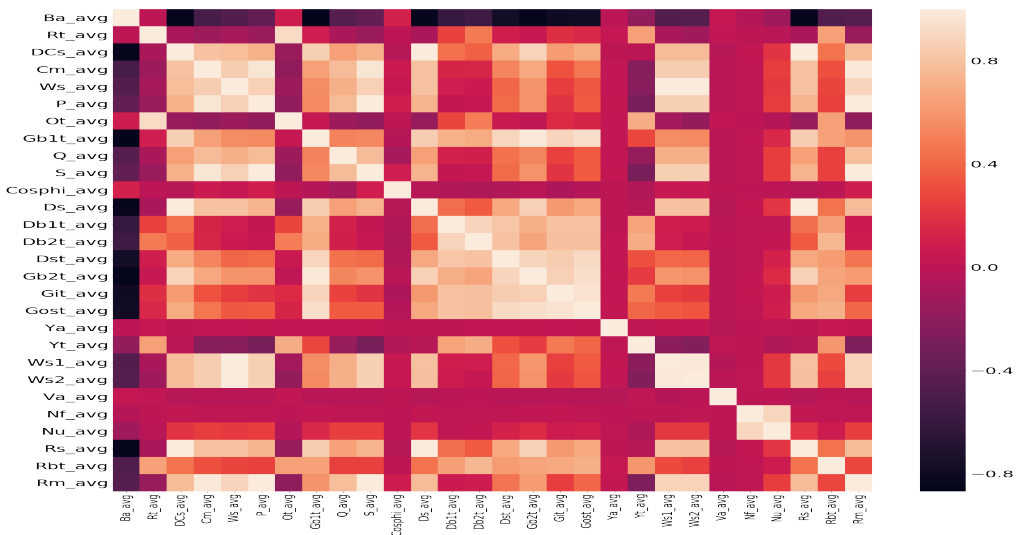


Figure 11. Pearson correlation heatmap

3.2. Machine learning model training

After the SCADA data preprocessing and feature selection the training of the ML model is done using six different regressors which are further described in the following sections.

3.2.1. Extra tree regressor

The first regressor we used to train the WT SCADA data is an extra-tree regressor. The MSE, MAE and R^2 reported were 0.86, 0.79 and 43.01% respectively. The time taken by this regressor was 55.95 seconds. As the dataset we used in this study is big which contain almost 200000 samples which is hard to visualize for better visualization Figure 12 show the prediction of GB temperature for 3 days.

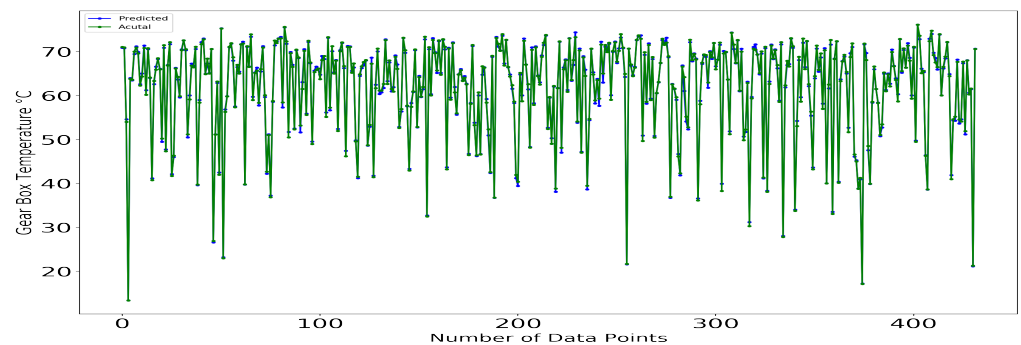


Figure 12. Wind turbine gearbox temperature prediction using extra tree regressor for 3 days

3.2.2. Random forest regressor

The second algorithm we try for our collected SCADA data was a random forest regressor which uses the average value of sub samples as explained in methodology. The MSE, MAE and R^2 reported were 0.39, 0.44 and 99.6% respectively. The time taken by this regressor was 42.95 seconds. Figure 13 represents the prediction with 3 days of data only. Compare to the previous model the prediction accuracy reported for this model is high.

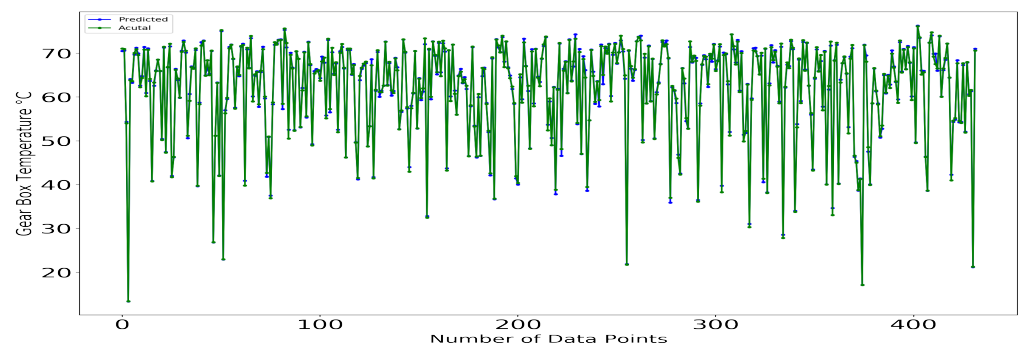


Figure 13. Wind turbine gearbox temperature prediction using random forest regressor for 3 days

3.2.3. KNeighbors regressor

KNeighbors regressor predicts the targeted output by interpolation with nearest neighbors in the training data which is explained in section 2. The MSE, MAE and R^2 reported were 22.54, 2.92 and 82.05% respectively. The time taken by this regressor was 19.93 seconds. Compare to the previous models the prediction accuracy of this model is less and the error is high. The time taken by this model is less than the previous. Figure 14 represents the better visualization of the prediction with 3 days of data only.

3.2.4. AdaBoost regressor

The AdaBoost regressor is a meta-predictive algorithm which predicts the targeted output using decision tree regression as explained in methodology. The MSE, MAE and R^2 reported were 2.25, 1.12 and 98.2% respectively. The time taken by this regressor was 47.95 seconds. Figure 15 represents the better visualization of the prediction with 3 days of data only.

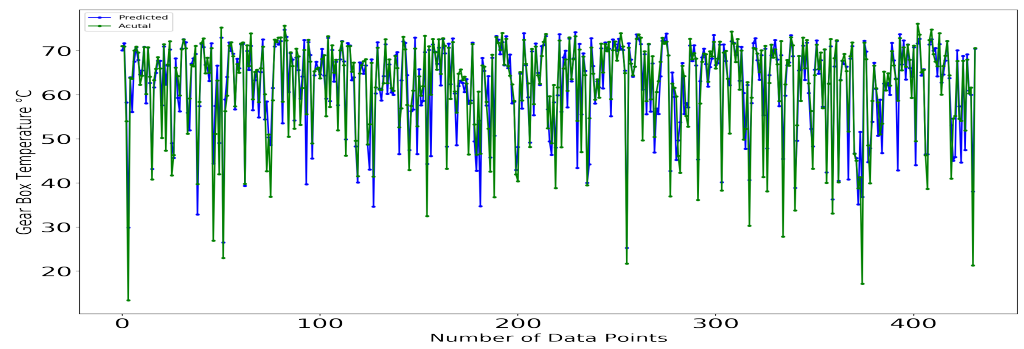


Figure 14. Wind turbine gearbox temperature prediction using KNeighbors regressor for 3 days

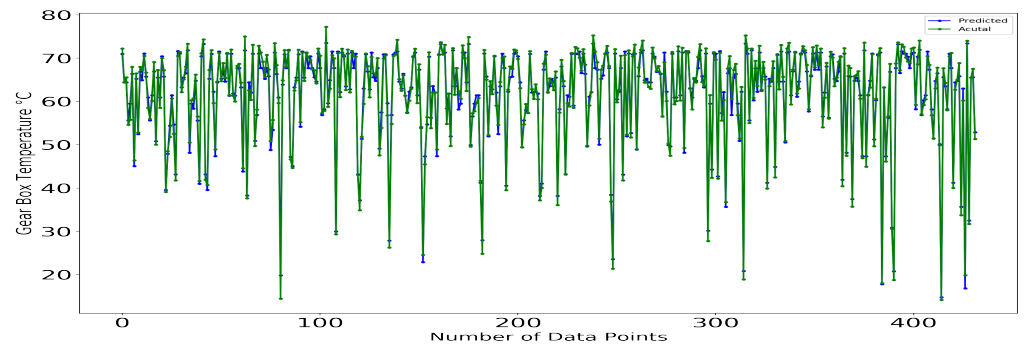


Figure 15. Wind turbine gearbox temperature prediction using adaBoost regressor for 3 days

3.2.5. Multi-layer perceptron regressor.

The multi-layer perceptron (MLP) regressor optimizes the squared loss using gradient descent as explained in section 2. The MSE, MAE and R^2 reported were 41.1, 4.96 and 73.31% respectively. The time taken by this regressor was 49.95 seconds. The output is further visualized in Figure 16 for 3 days. The prediction error for this model is very high as compared to previous models.

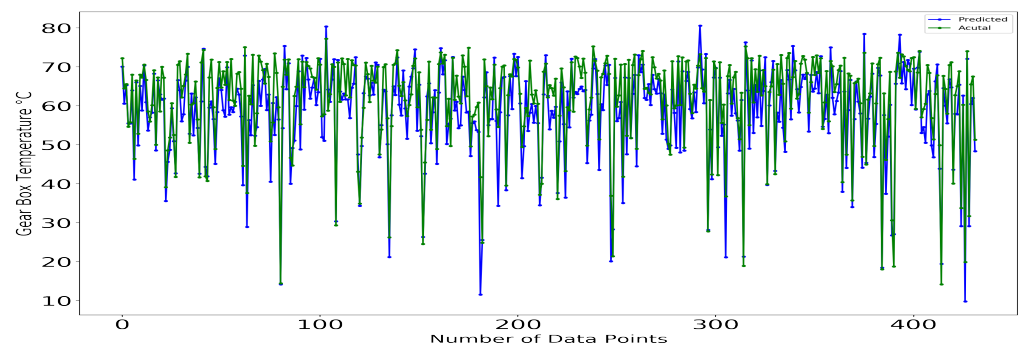


Figure 16. Wind turbine gearbox temperature prediction using multi-layer perceptron regressor for 3 days

3.2.6. Bagging regressor

The bagging regressor is an ensemble that predicts the output GB temperature based on aggregate the average of individuals as explained in the previous section. The MSE, MAE and R^2 reported were 0.33, 0.44 and 98.5 respectively. The time taken by this regressor was 31.77 seconds. Compare to the previous models the prediction accuracy of these models. Figure 17 represents the better visualization of the prediction with 3 days of data only.

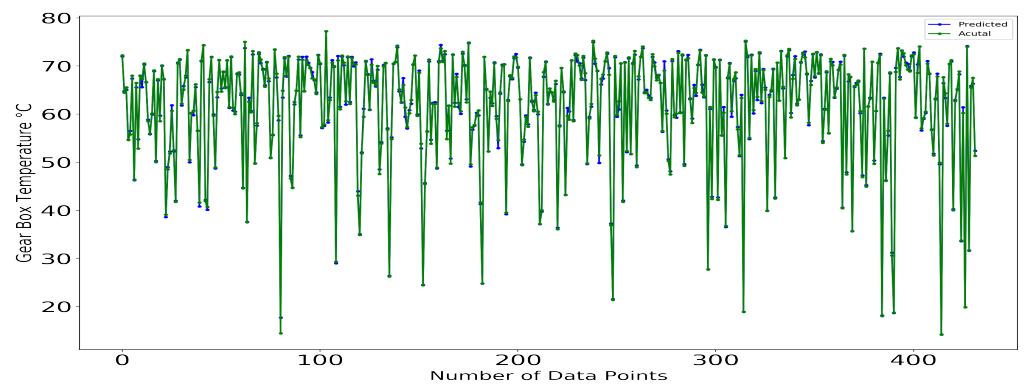


Figure 17. Wind turbine gearbox temperature prediction using bagging regressor for 3 days

3.3. Machine learning model comparison and evaluation

The above-discussed ML model is further compared on the basis of errors and accuracy. The models are further evaluated using backward and recursive feature elimination techniques, these results are tabulated in Table 3. After running a backward elimination algorithm, we remove [Ws avg, Nac avg, Ws1 avg]. As a result, the accuracy slightly improves in some cases and recursive feature elimination technique algorithm, [Gost avg, P avg, Rm avg, Ds avg, 'Cm avg', 'DCs avg', 'Gb1t avg'] are the features that we decide to remove, based on the observed outcomes in each fold.

The results is further evaluated without removing any feature from the dataset. Table 3 indicates that bagging regressor is better than other models because it offers less error and the accuracy of prediction is high as compared to other models. Considering these factors we select the bagging regressor as the most suitable ML model for the output gearbox temperature prediction.

Table 3: Comparison of different feature elimination techniques on ML models

Model	Backward Elimination				Recursive Feature Elimination				Without data manipulation			
	MSE	MAE	R2	Time(s)	MSE	MAE	R2	Time(s)	MSE	MAE	R2	Time(s)
Bagging	0.36	0.44	99.7	81.77	0.65	0.62	99.4	41.77	0.33	0.44	99.8	31.77
Random fores	0.38	0.44	99.7	35.95	0.7	0.64	99.3	35.95	0.39	0.44	99.6	42.95
Adaboost regressor	2.45	1.26	98.07	43.95	1.95	1.1	98.4	33.95	2.25	1.12	98.2	47.95
K-Neighbors	6.14	1.86	95.01	48.95	3.14	1.16	97.05	14.95	22.54	2.92	82.05	19.95
Extra tree regressor	0.38	0.4	38.01	40.95	0.76	0.69	66.01	48.95	0.86	0.79	43.01	55.95
MLP regressor	46.61	5.01	64.31	15.95	1.161	0.96	98.31	19.95	41.1	4.96	73.31	49.95

3.3.1. Validation

The results obtained from the ML models are further validated by dividing the dataset in train, test and validate data as explained in methodology. The data is split in 80,10,10 percent respectively. The results obtained for the validation are shown in Table 4.

Table 4: Validation results using holdout method

Model	MSE	MAE	R ² (%)	Time (sec)
Bagging	0.36	0.47	99.4	34.87
Random forest	0.42	0.45	99.1	45.95
Adaboost regressor	5.25	2.12	94.22	45.15
K-Neighbors	19.54	3.12	84.14	29.35
Extra tree regressor	0.93	0.91	48.21	66.34
MLP regressor	39.23	7.63	86.42	35.53

3.4. Fault detection of wind turbine gearbox using best model

Once the normal behavior of the WT GB is trained by using the best-performed training algorithm, the fault in the GB is now detected by comparing the actual and predicted temperature. Figure 18 shows the change in temperature of GB for 3 month period which eventually fails after 76 days. Figure 19 shows the difference between

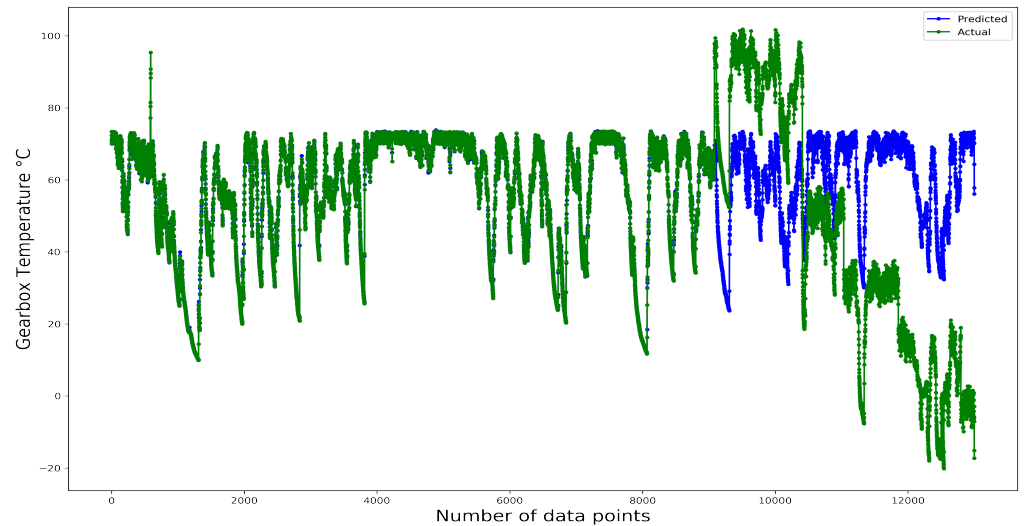


Figure 18. Fault detection of WT gearbox for 3 months

predicted and actual GB temperatures in these 3 months. The first noticeable change from the predicted model occurs at point 1 label in Figure 19, shows that the GB performance has deviated from normal behavior. The frequency of this deviation is increased further in point 2 and the temperature difference is 15°C and lasted for a long time. Point 3 in Fig 19 shows that the turbine is completely failed at point 3 due to the overheating of the GB. This shows that using this model we detect the anomalous behavior of WT GB at point 1 which can be used as an early warning point. Point 2 in this method can be used as an acute warning point (alarm) as the anomalous behavior frequency is increased. When we calculate the difference in time between our warning at point 1 and the complete failure of the WT at point 3 it is 59 days which means we detect this fault 59 days earlier and we can fix this fault to avoid the failure at point 3. The extremely unusual behavior of the GB is model at point 2 is 9 days earlier than actual failure, we can stop the WT at point 2 and do the minor repair at this point to avoid the complete failure in point 3.

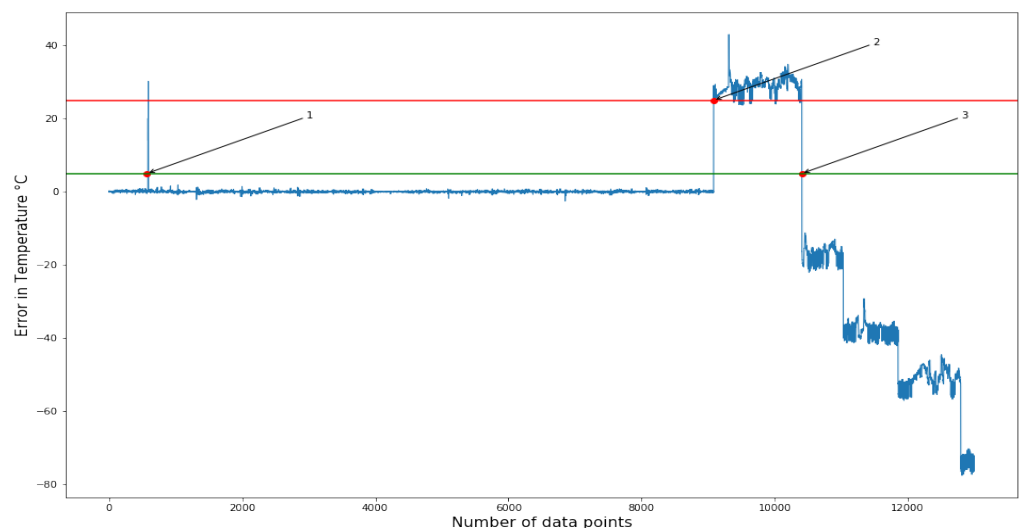


Figure 19. Difference between actual and prediction gearbox temperature for 3 months

4. Conclusion and Future Work

The wind industry in past few years experiences significant growth and will continue to play a significant role in providing energy to households and industries. However, the wind industry faces higher challenges in terms of reliability and availability. Unexpected failure in wind turbine causes unscheduled maintenance, which increases operational and maintenance costs. These unexpected failures can be detected by efficient failure detection methods, which can help in improving WT reliability, availability, power performance, and reduction in operational and maintenance costs.

The collected SCADA dataset for this research contains different outliers so the SCADA data is pre-processed using Python. The features were selected from the dataset to reduce the number of input parameters while developing a ML predictive model. The feature selection process is critical because it will improve the efficiency of the model as well as reduce the computation time of the model.

Six different ML algorithms were trained and compared based on mean square error and coefficient of determination. Considering these evaluation metrics, we select the bagging regressor as the most suitable ML model for the output GB temperature prediction and we used to predict the failure in our faulty WT. The bagging method is further validated based on the validation data. By implementing the proposed method to predict the fault in the faulty we detect the anomalous behavior of WT GB 59 days earlier which considered as an early warning point. The extremely unusual behavior of the GB is model 9 days earlier than actual failure, we can stop the WT 9 days earlier and do the minor repair at this point to avoid the complete failure.

This research work can be further extended to other subsystems in the WT such as power converters, tower, main bearing, blade. Experimental verification of the proposed health monitoring of WT GB should be conducted.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SCADA	Supervisory Control and Data Acquisition
WTs	Wind Turbines
GB	Gearbox
O&M	Operation and Maintenance
ML	Machine Learning

1. Cresswell, L.; Twigg, R.; Buchdahl, J. Energy fact sheet series for key stage 4 and A level. Technical report, Technical report, Atmosphere Climate and Environment Information Programme ..., 2002.
2. Pîrjan, A.; Căruțașu, G.; Petroșanu, D.M. Designing, developing, and implementing a forecasting method for the produced and consumed electricity in the case of small wind farms situated on quite complex hilly terrain. *Energies* **2018**, *11*, 2623.
3. McCrone, A.; Moslener, U.; d'Estais, F.; Usher, E.; Grüning, C. *Global trends in renewable energy investment 2017*; Bloomberg New Energy Finance, 2017.
4. GWEC, G. Global wind statistics 2019, 2020.
5. Rashid, H.; Haider, W.; Batunlu, C. Forecasting of Wind Turbine Output Power Using Machine learning. 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), 2020, pp. 396–399. doi:10.1109/ACIT49673.2020.9208852.
6. Wang, C. Health Monitoring and Fault Diagnostics of Wind Turbines. PhD thesis, Aalborg Universitetsforlag, 2016.
7. Dinmohammadi, F.; Shafiee, M. A fuzzy-FMEA risk assessment approach for offshore wind turbines. *International Journal of Prognostics and Health Management* **2013**, *4*, 59–68.

8. Zhu, C.; Li, Y. Reliability analysis of wind turbines. *Stability Control and Reliable Performance of Wind Turbines* **2018**, p. 169.
9. Gray, C.S.; Watson, S.J. Physics of failure approach to wind turbine condition based maintenance. *Wind Energy* **2010**, *13*, 395–405.
10. Errichello, R.; Muller, J. Gearbox Reliability Collaborative Gearbox 1 Failure Analysis Report: December 2010-January 2011. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2012.
11. Zhu, J.; Yoon, J.M.; He, D.; Bechhoefer, E. Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines. *Wind Energy* **2015**, *18*, 1131–1149.
12. Wang, C.; Liu, X.; Chen, Z. Probe improvement of inductive sensor for online health monitoring of mechanical transmission systems. *IEEE Transactions on Magnetics* **2015**, *51*, 1–4.
13. Sørensen, B.F.; Lading, L.; Sendrup, P.; McGugan, M.; Debel, C.P.; Kristensen, O.J.; Larsen, G.C.; Hansen, A.M.; Rheinländer, J.; Rusborg, J.; others. Fundamentals for remote structural health monitoring of wind turbine blades-a preproject **2002**.
14. Bouno, T.; Yuji, T.; Hamada, T.; Hideaki, T. Failure forecast diagnosis of small wind turbine using acoustic emission sensor. *KIEE International Transaction on Electrical Machinery and Energy Conversion Systems* **2005**, *5*, 78–83.
15. Beattie, A.; Rumsey, M. Non-destructive evaluation of wind turbine blades using an infrared camera. 37th Aerospace Sciences Meeting and Exhibit, 1998, p. 46.
16. Guo, P.; Infield, D.; Yang, X. Wind turbine generator condition-monitoring using temperature trend analysis. *IEEE Transactions on sustainable energy* **2011**, *3*, 124–133.
17. Gros, X.E. An eddy current approach to the detection of damage caused by low-energy impacts on carbon fibre reinforced materials. *Materials & Design* **1995**, *16*, 167–173.
18. Zaher, A.; McArthur, S.; Infield, D.; Patel, Y. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* **2009**, *12*, 574–593.
19. Long, H.; Wang, L.; Zhang, Z.; Song, Z.; Xu, J. Data-driven wind turbine power generation performance monitoring. *IEEE Transactions on Industrial Electronics* **2015**, *62*, 6627–6635.
20. Yang, S.; Li, W.; Wang, C. The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network. 2008 International Conference on Condition Monitoring and Diagnosis. IEEE, 2008, pp. 1327–1330.
21. Li, Y.; Li, G.; Yan, J. Fault diagnosis of wind turbine blades based on fuzzy theory. 2011 International Conference on Control, Automation and Systems Engineering (CASE). IEEE, 2011, pp. 1–3.
22. Chen, Z.; Lian, X.; Yu, H.; Bao, Z. Algorithm of data mining and its application in fault diagnosis for wind turbine. 2009 Second International Symposium on Knowledge Acquisition and Modeling. IEEE, 2009, Vol. 2, pp. 240–243.
23. An, X.; Jiang, D.; Li, S. Application of back propagation neural network to fault diagnosis of direct-drive wind turbine. 2010 World Non-Grid-Connected Wind Power and Energy Conference. IEEE, 2010, pp. 1–5.
24. Ju, L.; Song, D.; Shi, B.; Zhao, Q. Fault predictive diagnosis of wind turbine based on LM arithmetic of artificial neural network theory. 2011 Seventh International Conference on Natural Computation. IEEE, 2011, Vol. 1, pp. 575–579.
25. Faulstich, S.; Hahn, B.; Tavner, P.J. Wind turbine downtime and its importance for offshore deployment. *Wind energy* **2011**, *14*, 327–337.
26. Lapira, E.; Brisset, D.; Ardakani, H.D.; Siegel, D.; Lee, J. Wind turbine performance assessment using multi-regime modeling approach. *Renewable Energy* **2012**, *45*, 86–95.
27. Laouti, N.; Sheibat-Othman, N.; Othman, S. Support vector machines for fault detection in wind turbines. *IFAC Proceedings Volumes* **2011**, *44*, 7067–7072.
28. Schlechtingen, M.; Santos, I.F. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical systems and signal processing* **2011**, *25*, 1849–1875.
29. Wang, L.; Zhang, Z.; Long, H.; Xu, J.; Liu, R. Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics* **2016**, *13*, 1360–1368.
30. Orozco, R.; Sheng, S.; Phillips, C. Diagnostic Models for Wind Turbine Gearbox Components Using SCADA Time Series Data. 2018 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2018, pp. 1–9.

31. Verma, A.; Kusiak, A. Fault monitoring of wind turbine generator brushes: a data-mining approach. *Journal of Solar Energy Engineering* **2012**, *134*.
32. Abdallah, I.; Dertimanis, V.; Mylonas, H.; Tatsis, K.; Chatzi, E.; Dervilis, N.; Worden, K.; Maguire, E. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. *Safety and Reliability—Safe Societies in a Changing World* **2018**, pp. 3053–3061.
33. Menezes, D.; Mendes, M.; Almeida, J.A.; Farinha, T. Wind farm and resource datasets: A comprehensive survey and overview. *Energies* **2020**, *13*, 4702.
34. Wu, Y.; Zhang, A. Feature selection for classifying high-dimensional numerical data. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, Vol. 2, pp. II–II.
35. Lin, W.; Wu, Z.; Lin, L.; Wen, A.; Li, J. An ensemble random forest algorithm for insurance big data analysis. *Ieee Access* **2017**, *5*, 16568–16575.
36. Hinton, G.E. Learning translation invariant recognition in a massively parallel networks. International Conference on Parallel Architectures and Languages Europe. Springer, 1987, pp. 1–13.
37. Rumelhart, D.E. Parallel distributed processing: Explorations in the microstructure of cognition. *Learning internal representations by error propagation* **1986**, *1*, 318–362.
38. Razavi, B.S.; others. Predicting the trend of land use changes using artificial neural network and markov chain model (case study: Kermanshah City). *Research Journal of Environmental and Earth Sciences* **2014**, *6*, 215–226.
39. Sun, Y.; Liu, Z.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE Transactions on Aerospace and Electronic Systems* **2007**, *43*, 112–125.
40. Chu, F.; Zaniolo, C. Fast and light boosting for adaptive mining of data streams. Pacific-Asia conference on knowledge discovery and data mining. Springer, 2004, pp. 282–292.
41. Breiman, L. Bagging predictors. *Machine learning* **1996**, *24*, 123–140.
42. Chernick, M.R.; LaBudde, R.A. *An introduction to bootstrap methods with applications to R*; John Wiley & Sons, 2014.
43. Rashid, H.; Khalaji, E.; Rasheed, J.; Batunlu, C. Fault Prediction of Wind Turbine Gearbox Based on SCADA Data and Machine Learning. 2020 10th International Conference on Advanced Computer Information Technologies (ACIT). IEEE, 2020, pp. 391–395.