

## SCADA Data-Driven Wind Turbine Main Bearing Fault Prognosis Based on One-Class Support Vector Machines

A. Insuasty<sup>1</sup>, C. Tutivén<sup>2,3,4</sup> and Y. Vidal<sup>4,5</sup>

<sup>1</sup>Electronic Engineering  
Faculty of Engineering  
Universidad de Nariño

San Juan de Pasto, Colombia

Phone number:+0057 317 8346569, e-mail: [andresinsuasty@udenar.edu.co](mailto:andresinsuasty@udenar.edu.co)

<sup>2</sup>Mechatronics Engineering  
Faculty of Mechanical Engineering and Production Science  
Escuela Superior Politécnica del Litoral  
Guayaquil, Ecuador

Phone number:+593 958621189, e-mail:[cjtutive@espol.edu.ec](mailto:cjtutive@espol.edu.ec)

<sup>3</sup>Universidad ECOTEC  
Km. 13.5 Vía a Samborondón  
Guayaquil, Ecuador

<sup>4</sup>Control, Modeling, Identification and Applications  
Department of Mathematics

Escola d'Enginyeria de Barcelona Est  
Universitat Politècnica de Catalunya

Campus Diagonal-Besós, 08019 Barcelona, Spain

Phone number:+0034 934 137309, e-mail: [yolanda.vidal@upc.edu](mailto:yolanda.vidal@upc.edu)

<sup>5</sup> Institute of Mathematics (IMTech)  
Universitat Politècnica de Catalunya  
Pau Gargallo 14, 08028 Barcelona, Spain

**Abstract.** This work proposes a fault prognosis methodology to predict the main bearing fault several months in advance and let turbine operators plan ahead. Reducing downtime is of paramount importance in wind energy industry to address its energy loss impact. The main advantages of the proposed methodology are the following ones. It is an unsupervised approach, thus it does not require faulty data to be trained; ii) it is based only on exogenous data and one representative temperature close to the subsystem to

diagnose, thus avoiding data contamination; iii) it accomplishes the prognosis (various months in advance) of the main bearing fault; and iv) the validity and performance of the established methodology is demonstrated on a real underproduction wind turbine.

### Key words

wind turbine, fault prognosis, main bearing, SCADA data.

## 1. Introduction

Nowadays, the challenge to attain energy while minimizing costs and pollution is the matter in hand, due to climate change and global warming concerns. In particular, renewable energies are a significant opportunity. Among renewable energy sources, the wind energy sector has grown significantly within the last two decades. However, to really unleash the massive potential of wind energy is crucial to reduce its levelized cost of electricity (LCOE), see [1]. In this matter, the reduction of operation and maintenance costs, through the use of condition monitoring strategies, is imperative.

Condition monitoring is the crux of the matter to move from time-based preventive maintenance (which is still the current mainstream practice for wind turbines) to predictive maintenance. A significant amount of research on fault prognosis for wind turbines exists, see [2]. The main subsystems that the majority of the research focuses on includes: blades (e.g., [3], and [4]), gearbox (e.g., [5], and [6]), and bearings (e.g., [7], [8], and [9]). These aforementioned works use the data from specific and costly condition monitoring sensors as they are mainly based on high-frequency vibration analysis, acoustic emission signals, or oil analysis sensors. By contrast, in this work, the proposed predictive maintenance strategy is achieved without needing to invest in extra hardware, but only through the already available SCADA (supervisory control and data acquisition) data.

These SCADA data are highly variable due to the constantly changing environmental conditions, are affected by seasonality, have a low sampling ratio (usually, as in this work, 10 minute average value) compared to the kHz frequency of traditional condition monitoring strategies, are rarely standardized, and the description of work orders is generally not clear. Thus, resulting in challenges for researchers in contextualizing these data to be used for fault prognosis, see [10], [11], and [12]. However, recent research has focused on this approach, and there are some success stories about using solely SCADA data from real wind turbines for condition monitoring. For example, in [13] the diagnose and prediction of wind turbine faults from SCADA data is accomplished using support vector machines, and in [14] a framework for automatically identifying periods of faulty operation using rules applied to the turbine alarm system is presented and applied to perform fault classification.

The previous works use SCADA data and validate their approach on real wind turbines, but all of them require faulty data. In particular, historical SCADA data must be accurately labeled with the periods when turbines were down due to faults, as well as with the reason for the fault. This is time

consuming, error prone, and likely to result in a set of labeled vectors with an unbalanced number of classes. By contrast, in this work, there is no need of historical fault data, thus, the proposed strategy can be applied to any wind farm (even when no faulty data has been yet recorded). In particular, in this work, a one-class support vector machine (SVM) anomaly detection method is proposed. Thus, the model is built by using only normal (healthy) operation data.

Finally, this work deals with wind turbine main bearing faults because, as stated by the European Academy of Wind Energy (EAWWE) [15],

The wind industry has identified the main bearing failures as being a critical issue in terms of increasing wind turbine reliability and availability, as they lead to a major repair with a high replacement cost and long downtime period.

The proposed methodology to prevent this important fault is demonstrated in this work on a real underproduction wind turbine.

## 2. Brief Wind Turbine Description

The wind turbine under study belongs to an onshore wind park located in Poland. It has a power of 2300 kW and a diameter of 101 m. Figure 1 shows its major components. A summary of the wind turbine technical specifications is

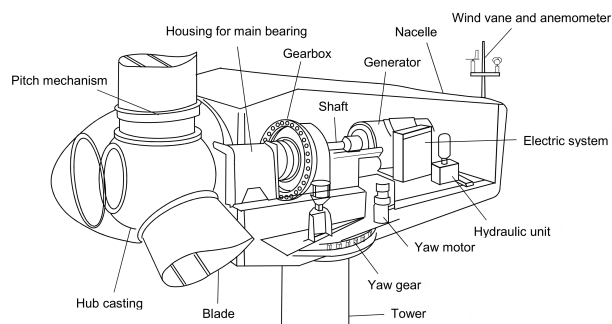


Fig. 1. Main components of the wind turbine [16].

given in Table I.

The wind farm SCADA data contain different variables that can be broadly grouped in environmental variables, electrical variables, component temperature variables, hydraulic variables, and control variables. The mean, maximum, minimum, and standard deviation of the 10-minute averaging period of 1 Hz sampled values are available for all these variables. In this work, the prognostic model is constructed based only on the exogenous variables mean wind speed and ambient temperature together with the mean low speed shaft temperature.

TABLE I  
TECHNICAL SPECIFICATIONS OF THE WIND TURBINE.

Number of blades	3
Nominal power	2300 kW
Rotor diameter	101 m
Wind class	IEC IIb
Swept area	8012 m <sup>2</sup>
Cut-in wind speed	3 m/s
Rated wind speed	12 m/s
Cut-out wind speed	20 m/s

Note that, apart from the SCADA data, there is available information about the maintenance and repair actions. These data provide information on the failures occurred, when did they occur, when the work was carried out, and information about the subsystem that was repaired or replaced. In particular, from this aforementioned information, it is known that the wind turbine under study had a main bearing fault on June 11, 2018. This information is used in this work to test whether the proposed methodology is capable to predict months in advance the appearance of this fault.

### 3. Failures Description

The studied wind turbine uses a double spherical main roller bearing. This part is composed by machine elements that permit rotary motion of shafts [17], such as a cage, the rolling elements (that are in contact under heavy dynamic loads and relatively high speeds), an outer race and an inner race, see Figure 2. Any of these parts is at risk of failure [18].

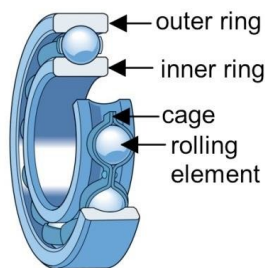


Fig. 2. Main components of the rolling bearing. [19].

Bearing damage and failure can be classified in different ways and use different terminology. Because of that, in 1995, a working group was formed to define a common classification method and terminology for bearing failure types. This group classified the failure into 6 main modes and thereafter into sub-modes [20]. The main modes are: i) fatigue, ii) wear, iii) corrosion, iv) electrical erosion, v)

plastic deformation, and vi) fracture and cracking. There are 14 failure modes in total, see Table II. All the named modes have different causes and behaviors, and could develop/appear progressively (slow degradation) which could allow them to be predicted months in advance. For more detailed information about the mentioned failures modes, see [20].

TABLE II  
SKF BEARING FAILURE CLASSIFICATION ADAPTED FROM ISO 15243:2004.

Failure Mode	Failure Sub-modes
Fatigue	Subsurface initiated fatigue
	Surface initiated fatigue
Wear	Abrasive wear
	Adhesive wear
Corrosion	Moisture corrosion
	Frictional corrosion
	Fretting corrosion
Electrical erosion	Excessive current erosion
	Current leakage erosion
	False brinelling
Plastic deformation	Overload deformation
	Indentations from debris
Fracture and cracking	Forced fracture
	Fatigue fracture
	Thermal cracking

### 4. Methodology

This work is focused on developing a methodology to prevent the main bearing failure in wind turbines. Particularly, it is proposed a anomaly detection method, analyzing the SCADA data using a one-class SVM algorithm. Besides, an indicator is stated based on monthly grouping the anomalies and using a predefined threshold, to minimize the number of false positives alarms.

As previously mentioned, in this work a 2300 kW wind turbine is used. The work orders show that the wind turbine presents the main bearing failure on June 11, 2018. The main challenge of this research is to predict the failure several months in advance. The strategy uses the data between the dates of January 01, 2014 until December 11, 2017 to train the one-class SVM. Then, data from December 11, 2017 until December 12, 2019 is used to test the model (making predictions). Figure 3 details the split into train and test datasets.



Fig. 3. Wind turbine data for training and test.

The proposed fault prognosis strategy follows the next steps. First, knowledge based selection of variables is performed. Second, the data is split, cleaned and pre-processed. Third, a one-class SVM with two inputs is trained for anomaly detection. Finally, an indicator is stated. The next subsections describe in detail the aforementioned procedure.

### A. Variable selection

Variable selection, also called feature selection or attribute selection, is the process of selecting a subset of relevant features to use in the model construction [21]. In this work, variable selection is carried out to construct a model based only on exogenous variables and the temperature measurement closely related to the studied failure.

The SCADA data used in this work were obtained from a 2300 kW under operation wind turbine, where continuous operational data were collected from the beginning of January 2014 to December 12, 2019. This data contains 434 different variables of the monitored wind turbine systems (pitch, main shaft bearing, gearbox, generator, etc.) with a sampling time averaged over 10 minutes, with a total of 312433 samples. In this work, only two environmental SCADA data (mean wind speed and ambient temperature) together with the mean low speed shaft temperature are used. The wind speed is the exogenous measurement that most directly affects the behavior of the wind turbine as it defines the different operating regions [22].

### B. Data cleaning and preprocessing

As stated previously, the dataset is divided into two parts: the first one that presents a normal operating behavior is used as training (from 01/01/2014 to 11/12/2017, with 207192 samples), and the second one that has a failure report is used as test dataset (from 11/12/2017 to 12/12/2019, with 105241 samples). The first step in the data cleaning preprocess is to eliminate missing data in the selected features. Then, to eliminate seasonality in the main bearing temperatures, the mean ambient temperature is subtracted to it. Finally, a Z-score normalization is employed,

$$z = \frac{x - \mu}{\sigma}, \quad (1)$$

where  $z$  is the value normalized,  $x$  is the original value,  $\mu$  is the mean value of  $x$  and  $\sigma$  is the standard deviation.

It is important that the data in the training set must correspond to a normal functioning behavior, a healthy behavior, to train the normality model. Thus, strictly speaking, it is a semisupervised algorithm.

A plot of the training and test sets is given in Figure 4, where the temperature variables and the wind speed variable

are shown. In this Figure also the moving average is plotted for better interpretation, but it is not used by the model.

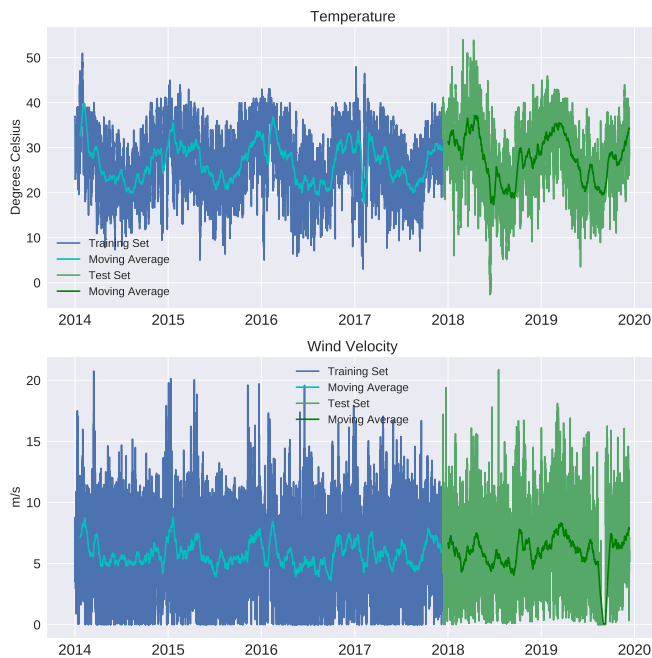


Fig. 4. In blue color the data used for training dataset and in green color the data used for test dataset, those variables with their respective moving average just for a better understanding of visualization.

### C. One-class Support Vector Machine

The SVM algorithm has shown excellent results for condition monitoring systems. However, the vast majority of these studies are based on training the model using both healthy and failure samples, as the basic SVM paradigm suggests. However, for many applications where there is an absence of negative (failure) samples, or is difficult to obtain them, as happens with wind turbine systems, applying this algorithm is nonviable. One alternative is the one-class SVM algorithm. One-class SVM is an extended algorithm of SVM [23], that is trained only on the 'normal' data, in our case the healthy samples. It learns the boundaries of these points and is therefore able to classify any points that lie outside the boundary as anomalies or outliers. Because of that, this algorithm widely used to discover anomalies in an unsupervised learning. In particular, the one-class SVM main objective is to classify the data as follows [24]:

$$f(x) = \begin{cases} 1, & \text{if } x \in S \\ -1, & \text{if } x \notin S \end{cases}$$

where  $S$  is a subset of a high dimensional feature space  $H$ . The main idea is that the algorithm maps all the samples into  $H$  by using an appropriate kernel function [25]. Then, it attempts to find the hyperplane that separates the mapped vectors from the origin with maximum margin. In this case, it has to determine how far from the origin a point can be before it can be classified as an anomaly. In our context, let  $x_1, x_2, \dots, x_l$ , be training examples belonging to the healthy class  $X$ . Note that  $l$  is the number of training samples and  $X$  is a compact subset of  $R^N$ . Then, given  $\Phi : X \rightarrow H$ , a kernel map which transforms the training examples to the feature space, to separate the data set from the origin, the following quadratic programming problem needs to be solved:

$$\min \frac{1}{2} \|w\|^2 + \frac{1}{vl} \sum_{i=1}^l \xi_i - \rho,$$

subject to

$$w^T \cdot \Phi(x_i) \geq \rho - \xi_i \quad i = 1, 2, \dots, l \quad \xi_i \geq 0$$

where  $\xi_i > 0$  are the slack variables,  $v \in (0, 1)$  is a parameter that controls the trade off between maximizing the distance from the origin and containing most of the data in the region created by the hyperplane,  $w^T$  is the vector transpose of  $w$ , and  $\cdot$  stands for the scalar product. Assuming  $w$  and  $\rho$  are the optimized parameters, then the decision function

$$f(x) = \text{sign}((w^T \cdot \Phi(x)) - \rho)$$

will be positive for most examples,  $x_i$ , contained in the training set.

The hyperplane makes a linear division. However, most data sets are not linearly separable. Thus, kernel functions, such as polynomial kernel and radial basis kernel, are widely used in SVMs to map from the original input data sets to a high dimensional feature space to make the data sets linearly separable [26]. In this work the radial basis function (RBF) kernel is used.

#### D. Indicator

Recall that SCADA data is 10 minutes averaged, thus at the end of the day it stores 144 samples of data for each sensor. Assuming a 30-day month, at the end of this period 4320 samples are registered. When testing a wind turbine, all these samples go through the proposed algorithm and they are classified as an anomaly or as a normal sample. Next, for each month, the total number of reported anomalies are counted and when this value exceeds the threshold of 1000 samples, then an alarm is triggered. In a nutshell, the indicator is based on the number of samples reported as anomalies per month.

## 5. Results

Recall that the studied wind turbine has a main bearing failure on 11/06/2018, and our main objective is to be able to predict it several months in advance.

The indicator of anomalies reported for each month in the test dataset is shown in Figure 5. The representation of the moving average shows a clear growing trend, exceeding 1400 samples classified as anomalies in the month of April 2018, thus leading to trigger the alarm two months in advance prior to the fault. On the other hand, after the month of June (after the repair), the report of anomalies drops noticeably going back to values below the threshold.

It is also interesting to note that in the months of August and September 2019 there is an unusual increase in the number of reported anomalies, but without trespassing the threshold. This behavior could be related to an atypical behavior in the wind speed variable, which causes some samples to be classified as anomalies.

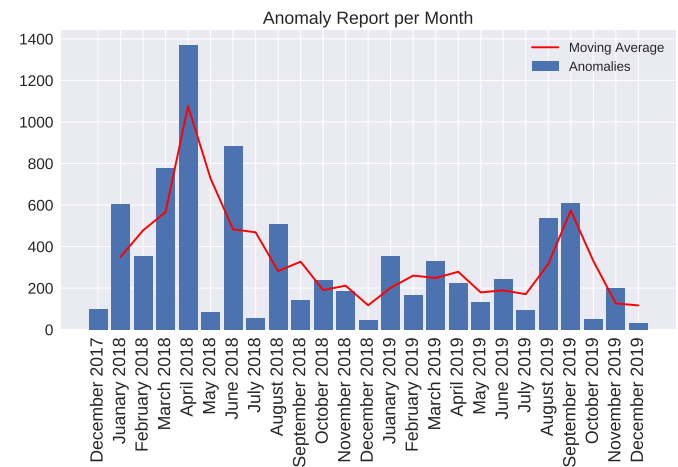


Fig. 5. Number of samples in the test dataset classified as anomalies grouped by month of the year and representation of the two-month moving average of reported anomalies.

## 6. Conclusions

In this work, a fault prognosis methodology to predict the main bearing fault is stated. It is based only on SCADA data and needs only healthy data to be trained. Moreover, the proposed methodology works under the different and varying operating and environmental conditions to which wind turbines are subject to. Finally, the methodology is demonstrated on a real underproduction wind turbine. The results show that the alarm is triggered several months in

advance from the total breakdown, thus giving time to the plant operator to schedule maintenance.

## Acknowledgments

The authors express their gratitude and appreciation to the Smartive company (<http://smartive.eu/>), as this work would not have been possible without their support and ceded wind farm data.

This work was partially funded by the Spanish Agencia Estatal de Investigación (AEI) - Ministerio de Economía, Industria y Competitividad (MINECO), and the Fondo Europeo de Desarrollo Regional (FEDER) through research project DPI2017-82930-C2-1-R; and by the Generalitat de Catalunya through research project 2017 SGR 388.

## References

- [1] W. Shen, X. Chen, J. Qiu, J. A. Hayward, S. Sayeef, P. Osman, K. Meng, Z. Y. Dong, A comprehensive review of variable renewable energy levelized cost of electricity, *Renewable and Sustainable Energy Reviews* 133 (2020) 110301.
- [2] M. L. Hossain, A. Abu-Siada, S. Muyeen, Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review, *Energies* 11 (5) (2018) 1309.
- [3] M. Florian, J. D. Sørensen, Wind turbine blade life-time assessment model for preventive planning of operation and maintenance, *Journal of Marine Science and Engineering* 3 (3) (2015) 1027–1040.
- [4] J. Tang, S. Soua, C. Mares, T.-H. Gan, An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades, *Renewable Energy* 99 (2016) 170–179.
- [5] F. Cheng, L. Qu, W. Qiao, Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis, *IEEE Transactions on Sustainable Energy* 9 (1) (2017) 157–167.
- [6] R. Dupuis, Application of oil debris monitoring for wind turbine gearbox prognostics and health management, in: *Annual Conference of the prognostics and health management society*, 2010, pp. 10–16.
- [7] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, M. Saif, An integrated feature-based failure prognosis method for wind turbine bearings, *IEEE/ASME Transactions on Mechatronics* (2020).
- [8] M. Motahari-Nezhad, S. M. Jafari, Anfis system for prognosis of dynamometer high-speed ball bearing based on frequency domain acoustic emission signals, *Measurement* 166 (2020) 108154.
- [9] F. Elasha, M. Greaves, D. Mba, D. Fang, A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetary gearbox, *Applied Acoustics* 115 (2017) 181–195.
- [10] K. Leahy, C. Gallagher, P. O'Donovan, D. T. O'Sullivan, Issues with data quality for wind turbine condition monitoring and reliability analyses, *Energies* 12 (2) (2019) 201.
- [11] M. Ruiz, L. E. Mujica, S. Alferes, L. Acho, C. Tutiven, Y. Vidal, J. Rodellar, F. Pozo, Wind turbine fault detection and classification by means of image texture analysis, *Mechanical Systems and Signal Processing* 107 (2018) 149–167.
- [12] F. Pozo, Y. Vidal, J. M. Serrahima, On real-time fault detection in wind turbines: Sensor selection algorithm and detection time reduction analysis, *Energies* 9 (7) (2016) 520.
- [13] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, A. M. Agogino, D. T. O'Sullivan, Diagnosing and predicting wind turbine faults from scada data using support vector machines, *International Journal of Prognostics and Health Management* 9 (1) (2018) 1–11.
- [14] K. Leahy, C. Gallagher, P. O'Donovan, K. Bruton, D. T. O'Sullivan, A robust prescriptive framework and performance metric for diagnosing and predicting wind turbine faults based on scada and alarms data with case study, *Energies* 11 (7) (2018) 1738.
- [15] E. Hart, B. Clarke, G. Nicholas, A. Kazemi Amiri, J. Stirling, J. Carroll, R. Dwyer-Joyce, A. McDonald, H. Long, A review of wind turbine main bearings: design, operation, modelling, damage mechanisms and fault detection, *Wind Energy Science* 5 (1) (2020) 105–124.
- [16] Z. Jiang, W. Hu, W. Dong, Z. Gao, Z. Ren, Structural reliability analysis of wind turbines: A review, *Energies* 10 (12) (2017) 2099.
- [17] V. Srinivasan, Analysis of dynamic load characteristics on hydrostatic bearing with variable viscosity and temperature using simulation technique, *Indian Journal of Science and Technology* 6 (6) (2013) 4797–4803.
- [18] M. Hamadache, D. Lee, Wind turbine main bearing fault detection via shaft speed signal analysis under constant load, in: *2016 16th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 2016, pp. 1579–1584.
- [19] A. Nabhan, N. Ghazaly, A. Samy, M. Mousa, Bearing fault detection techniques-a review, *Turkish Journal of Engineering, Sciences and Technology* 3 (2) (2015).
- [20] Bearing damage and failure analysis, [https://www.skf.com/binaries/pub12/Images/0901d1968064c148-Bearing-failures---14219\\_2-EN\\_tcm\\_12-297619.pdf](https://www.skf.com/binaries/pub12/Images/0901d1968064c148-Bearing-failures---14219_2-EN_tcm_12-297619.pdf), accessed: 2021-01-24 (2017).
- [21] M. Kang, J. Tian, Machine learning: Data pre-processing, Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things (2018) 111–130.
- [22] P. Tavner, C. Edwards, A. Brinkman, F. Spinato, Influence of wind speed on wind turbine reliability, *Wind Engineering* 30 (1) (2006) 55–72.
- [23] L. M. Manevitz, M. Yousef, One-class svms for document classification, *Journal of machine Learning research* 2 (Dec) (2001) 139–154.
- [24] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, R. C. Williamson, Estimating the support of a high-dimensional distribution, *Neural computation* 13 (7) (2001) 1443–1471.
- [25] T. Zhang, J. Wang, L. Xu, P. Liu, Fall detection by wearable sensor and one-class svm algorithm, in: *Intelligent computing in signal processing and pattern recognition*, Springer, 2006, pp. 858–863.
- [26] Y. Wang, J. Wong, A. Miner, Anomaly intrusion detection using one class svm, in: *Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop*, 2004., IEEE, 2004, pp. 358–364.