# Machinery failure assessment through pattern recognition of energy consumption

Maria Magdalena Kolmer
magdalena.kolmer@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa, Portugal

March 2017

## Abstract

The purpose of the thesis was to develop a fault detection system of a fan based on pattern recognition in energy consumption. The aim of the system is to detect abnormal conditions in ventilation systems in buildings, which result in the increase of the energy costs and the decrease of indoor air quality. The process involved building the laboratory stand required for the data acquisition. Measurements of energy consumption were performed during normal and faulty operating conditions. The examined states include clogged air intake area and fault causing additional friction between a shaft and housing. The analysis of results indicated that energy consumption varies depending on working conditions. The failure assessment was based on damage sensitive features extracted from the dataset and the set of selected features include average load, time of record and standard deviation. The failure assessment was performed with two methods: decision trees and neural networks. Both algorithms were suitable for the given problem and had similar overall classification accuracy. Algorithms were highly efficient in predicting clogged air intake area (more than 80%), but the fault causing extra friction was rarely detected - the malfunction was correctly classified only in 16% of cases. The main advantage of using the decision trees its clear structure which is easy to understand, while the neural network is highly efficient due to the utilisation of nonlinear function and its flexibility in the learning process.

## 1. Introduction

The increase in the world electricity consumption has been observed during the last decades and the demand is predicted to grow continuously over the next years. In the OECD countries, electricity consumption almost tripled over the last 40 years, for an average annual growth rate of 2.5%. Part of the increase is attributable to rising GDP per capita as well as the development of the commercial and public service sector and the multiplication of small appliances in offices and homes [1]. One of the main drivers of the energy consumption increase is high demand of heating, ventilating and air conditioning (HVAC) systems. The impact of HVAC technology on peak demand is even of greater consequence. In result, in the well-developed countries, the peak consumption is observed not only during the winter time but also in the summer due to widely used air conditioning installations. In the context of growing electricity demand, the global energy market faces the challenge of developing technologies, that will increase energy savings in appliances usage. In the terms of HVAC systems, energy savings will be obtained mainly through optimal control and early fault detection.

The purpose of the thesis is to develop a fault detection system of a fan based on pattern recognition in energy consumption. The system aims to detect machinery failures and diagnoses and their causes. A failure is typically an event, that is characterised by the temporary or permanent termination of the machine ability to perform a required function. Malfunctions in ventilating and air conditioning systems may occur in many components [2]. The typical faults include wear of fan bearings, clogged filters or malfunction in temperature, pressure, and humidity controllers [3]. Faults result in inefficient usage of energy, cause loss of quality in a device and shorten the remaining useful life of machine components. If the loss of quality is within an acceptable range of tolerance, the device needs to be post-processed and when it is out of range, the fault may result in the breakdown of the whole installation. Therefore, the early fault detection in machinery is important for the system security and presents an advantageous economic reason [2].

Various methods can be used to identify a process fault. The development of computational intelligence and sensor technology enables the usage of a real-time fault diagnosis system monitoring the operation of components. Such system allows to detect, locate and predict the presence of the defects and it can be achieved through online analysis of the electrical energy intake of machine tools. As defects develop, the machine abnormal behaviours are automatically monitored, logged and analysed to identify the need for the maintenance.

The benefits from the application of early fault detection systems in the air conditioning and ventilation installations will be the following:

- energy savings through early detection of inefficiently operating processes and faster location of faults,
- increased quality of living, e.g. faults can be detected before they affect the indoor air quality,
- reduction of cost as a result of scheduled and planned beforehand maintenance,
- increase in safety as detecting the faults early decreases the risk of personal injuries, as improperly operating processes may cause a health risk [4].

In this work, the measurement of energy consumption is proposed as the base parameter for the faults detection. As a result, the energy audit can be also used as a source of valuable information. Analysis of device energy consumption helps to identify energy wastages, indicate necessity of energy savings or estimate the financial impact of the energy conservation projects.

## 2. Methodology

The development of a complete fault detection system involves several steps: data acquisition and pre-processing, feature extraction and modelling. The sequence of steps, which are undertaken in to develop fault detection algorithm of this thesis, are presented in the diagram below (*Figure 1*).

The first step involves the data acquisition and pre-processing, which is a crucial part during development of model – free algorithm. Recorded data include energy consumption of the fan and ambient air temperature. Energy consumption measurements involve the application of two devices: the single-phase electronic energy meter and the EOT module, which transmit data to the cloud service using Wi-Fi connection. The online platform enables direct and real-time access to energy consumption data. Temperature measurements are performed with the use of EOT Temperature Meter at the sampling rate of 1 min [5].

In the given work the cleaning procedure of data is limited to detecting outliers - values in the data set that deviate significantly from the most of observations. Deviations might be caused by a sensor noise, process disturbances or human-related errors. In the applied procedure outliers are detected with the use of standard deviation method. The method indicates that measurements outside intervals may be considered as outliers. The range of the interval is calculated as the triple value of the standard deviation [6].

Identifying a significant number of indicators from a large amount of signal data is a focal point in the domain of fault diagnosis. It is desirable that features extracted from the sensory signal are sensitive to machinery abnormal behaviour and robust to the varying machinery operating conditions and background noise. Algorithms for extracting features from a large amount of signal data should also be in inexpensive in computations [7].

In the given example of fault diagnosis, the statistical parameters are extracted from the dataset over a certain time period, which is established beforehand. The following parameters are calculated: standard deviation (2.1), average value (2.2), and peak to peak value (2.3).

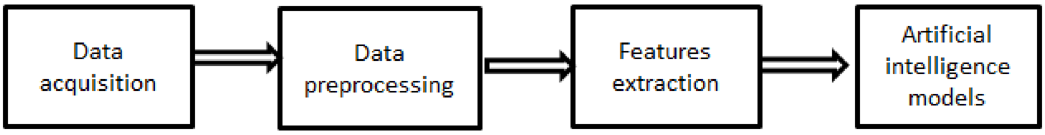

**Figure 1 Sequence of development of a fault detection system**

These are defined as [8]:

$$Sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^2} \qquad (2.1)$$

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \qquad (2.2)$$

$$Pp = \frac{1}{2}\left(max_{i=0}^{n-1} a_i - min_{i=0}^{n-1} a_i\right) \qquad (2.3)$$

The mean value reflects the average energy consumption of the fan over the investigated time period, standard deviation measures the dispersion of the dataset and the peak-to- peak value indicates the absolute amplitude level of the energy consumption fluctuations.

The next step of development of the fault detection system involves application of artificial intelligence methods, which are chosen to classify data. Two methods characterised by different features are selected: decision tree and neural network. The decision tree is a method for approximating discrete-valued target functions, the tree structure is among the most popular of inductive inference algorithms and has been successfully applied to a broad range of tasks. Decision trees classify instances by sorting them from the root to some leaf node, down the tree. Each node in the tree specifies a test of some attribute and each branch descending from the node corresponds to one of the possible values of the attribute. An instance is classified by starting at the tree root, testing the attribute specified by each node and moving down the tree branch corresponding to the output value in the given test [9]. The example of a structure of a decision tree is shown in *Figure 2*.

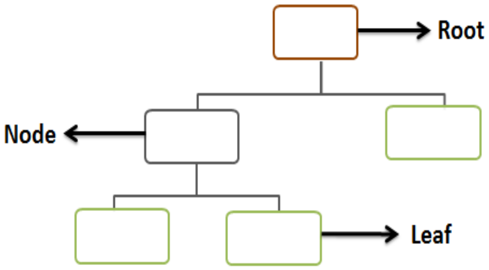

**Figure 2 Decision tree structure**

The decision tree algorithm is an easy to understand structure and has the clear relationship between input and output, thus it is selected as the first model to perform the classification task. After the model evaluation, the set of the input variables resulting in the best model performance is used for the development of the neural network algorithm. The network is a mathematical model that tends to simulate the structure and functionalities of biological neural networks. The basic component of a model is an artificial neuron, which illustrates a mathematical function of three simple sets of rules: multiplication, summation and activation. The obtained sum is projected on an activation function, the so-called transfer function and at the end an artificial neuron passes the information via outputs. The process of information flow in the artificial neuron is depicted in *Figure 3*.

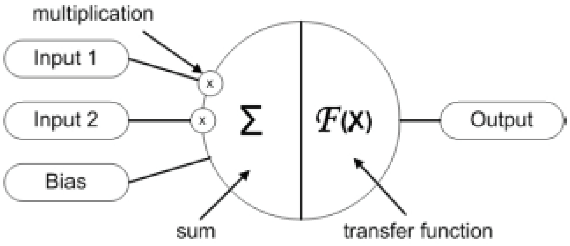

**Figure 3 Artificial neuron: working principle**

The working principle of individual elements of the network is not complicated -neurons read an input, process it, and generate an output. The calculation power of models occurs when basic blocks are interconnected into artificial neural networks, which is capable of solving complex real-life problems by processing information in a non-linear, distributed, parallel and local way. The network was chosen as the second model since it has a black box learning approach. On the other hand, the advantage of this method is the possibility of using nonlinear boundaries, what results in the high efficiency of classification.

## 3. Laboratory facility

Development of a fault detection system of a fan using a history based method and supervised learning algorithm requires large dataset of historical records. To collect data a low scale laboratory stand was designed and built. The installation consists of two fans and variable sensors, which measure, process and send the data to the cloud. The lower fan is subjected to different types of faults (number **9** in the scheme), while the second fan (**8**) is working in the correct manner and it is used as a model device during fans performances comparison. Fans require connection to the grid for the power supply. The energy consumption is measured collectively for both fans by an energy meter (**4**), which displays a result based on current and voltage. Data is sent by a wire to the smart meter EnergyOT (**3**) and transmitted to the cloud service.

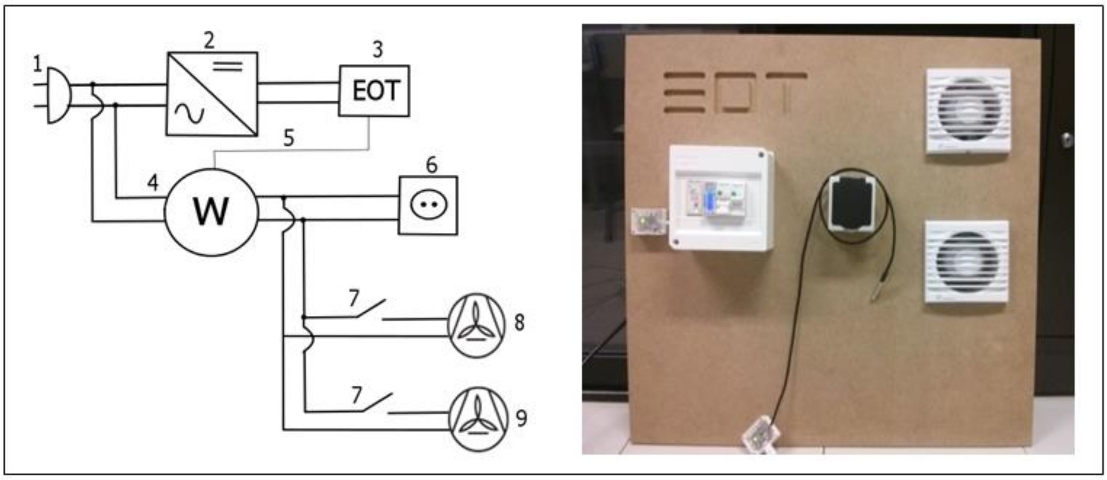

Figure 4 Scheme of the laboratory stand and picture of the ready installation

The results of measurements are available for online monitoring in the website: www.my.eot.pt [5]. The EnergyOT module requires direct current, therefore an AC/DC transformer (**2**) is installed before the device. The whole installation is protected with circuit breakers (**7**), which are also used for switching on fans. Additionally, measurements of temperature are performed with EOT Temperature Meter [5]. shows the scheme of installation and the picture of the ready installation.

To develop an algorithm based on the process history method, measurements must be performed during normal and faulty operating conditions. In the first stage of data acquisition, the performance of a fan without any fault was examined. To register changes in the power consumption, the fan was working continuously for twelve days. The analysis of the energy consumption curve and periodicity analysis indicated, that power requirements changes within the daily cycle: the load decreases during a daytime and the highest value of used power is observed during night hours. In order to investigate reasons for daily fluctuations, additionally, the measurements of air temperatures were performed. The changes in temperature, which affect the air density, are the main factor causing the daily load fluctuation.

The examined state of a fan included common malfunction in the ventilating system, which is clogging of the air intake area. To simulate such operating conditions in the laboratory stand, different areas of intake surface were covered with duck type: a quarter, a half, three-quarters and a whole surface. Another example of the common failure is the wear or misalignment of bearings, which can cause additional friction between a shaft and housing during rotation. The failure was simulated in the laboratory stand by mounting additional material on the fan housing. As a result, every rotation of a shaft led to the interaction between material and rotating blades, causing additional resistance. *Figure 5* presents pictures of different fault cases.

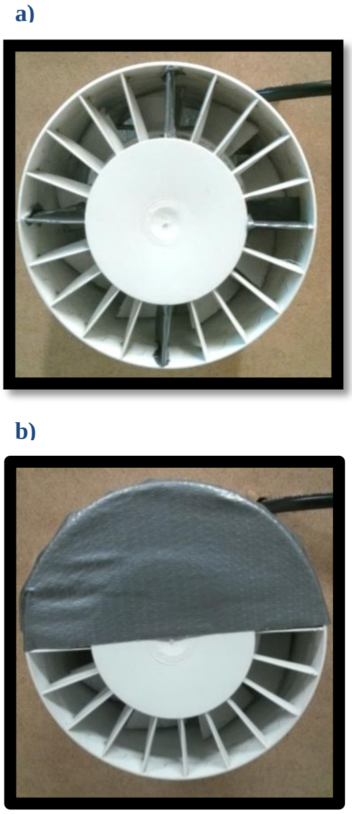

Figure 5 a) Fault in the surface of air intake, half of the area clogged,
b) Fault causing additional friction

In order to visualise the differences between performances of a fan subjected to the various working conditions, the plot (*Figure 6*) gathers all power consumption characteristics.

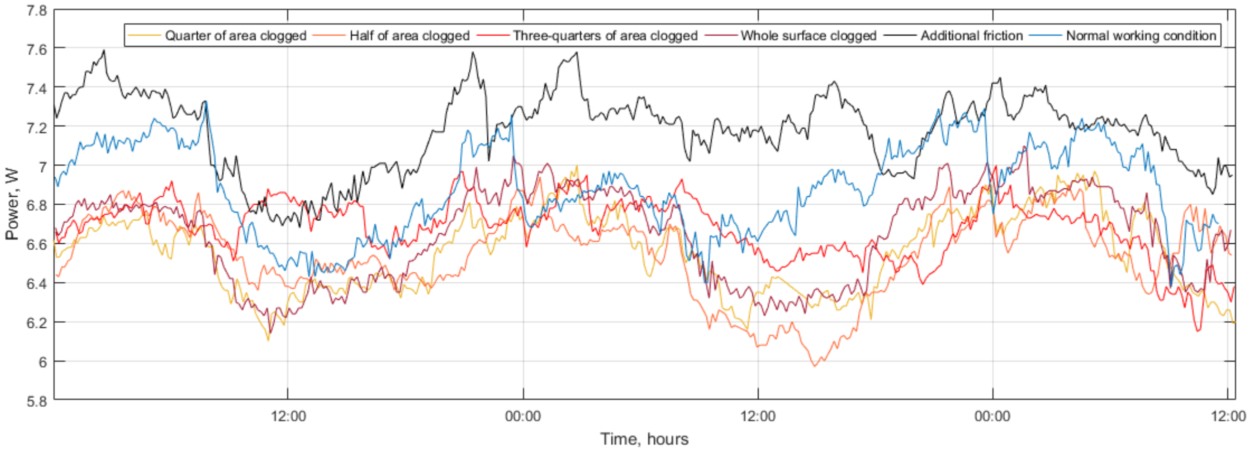

**Figure 6 Power consumption of a fan operating in different conditions**

Based on the graph, we can draw the conclusion that the energy consumption varies depending on the fan working condition. The energy required to run a fan without any failure (indicate by a blue line in the graph) changes during a day. The different conditions of ambient air are said to be the main reason of daily fluctuations. When the motor is forced to overcome additional friction (black line), the required energy increases. However, there are periods throughout the day, where the lines in the graph overlap with each other, meaning that both working states are characterised by the same power requirements. Contrary to the bearing fault, which simulates the additional friction caused by a bearing malfunction, the clogging area of the air intake results in the energy consumption decrease. It can be observed, that the plots of the fan load for different ratios of clogged surface have a similar shape, which makes them difficult to unambiguously identify for any pattern recognition algorithm. Additionally, also in the case of the fault in the air intake surface, there are periods during a day when the power requirement is similar to the normal working conditions.

## 4. Results and discussion

The next step of the fault detection system development includes making an accurate assessment of the malfunction based on extracted features. The algorithm can learn relationships from the data sets and assign a damaged state to a given set of features. The features are formed from measurements, which are sensitive to the malfunction. Once the features have been established, the map between the values and the diagnosis is constructed with the use of two different methods: decision tree and neural network. To develop the most efficient algorithm, several decision trees models with different input variables are examined. After the evaluation of the decision tree performance, the model based on the neural network is developed. The efficiency in the prediction of the fan working state is evaluated on the dataset external to the training dataset.

### 4.1 Decision tree

In order to train the decision tree algorithm, a large dataset with historical records is required. For the given fault detection algorithm, the total number of days, when the measurements were performed is equal to 46, resulting in 7420 measurement points of the fan load. The records were divided into two different sets: 5/6 of total data available was used for the model training and the rest of records were used to evaluate the model performance.

For the purpose of this work, decision tree models were trained with the use of the application Classification Learner included in the software MATLAB R2016a. The application allows to classify data using supervised machine learning techniques and it is a part of the Statistic and Machine Learning toolbox.

In the first step of the decision tree model development, the input values and the complexity of the structure have been defined. For the given case the possible input data include fan load, a period of the day and ambient air temperature. Moreover, the features of power consumption data (standard deviation, mean value, peak to peak value) are calculated over time, which is also set beforehand. The complexity of the decision tree model is measured by the maximum number of splits, which can vary from 4 to 100.

Five models with different structures were developed in order to compare their accuracy in classification. In the model characterised by the best performance, features were calculated every 40 min, which results in the training set consisting of 1237 observation points. The developed algorithm during data classification takes into account three attributes: average load, a period of the day and standard deviation. The peak-to-peak value is not included in a tree structure, what leads to the conclusion, that calculations of this feature are redundant. The model efficiency was evaluated by the software at the level of **67.3 %**, which means that for 833 from 1237 observation points, the class was predicted correctly. The high level of misclassification is observed between classes:” no fault” and “extra friction”. The model performance was also evaluated on the test data set, external to the training data. The algorithm correctly predicted the operating state of a fan in the case of 167 observations from the total of 244 observations. Therefore, the accuracy model is **68.4 %.**

The detailed statistics of the correct predictions are presented in *Figure 7*. It can be observed, that states: normal operating condition and fault in the intake air area are classified with the highest accuracy. Conversely, efficiency in detecting the fault, caused by the additional friction is very poor. The malfunction was correctly classified only in the case of 16% from all observations characterised by this type of the fault.

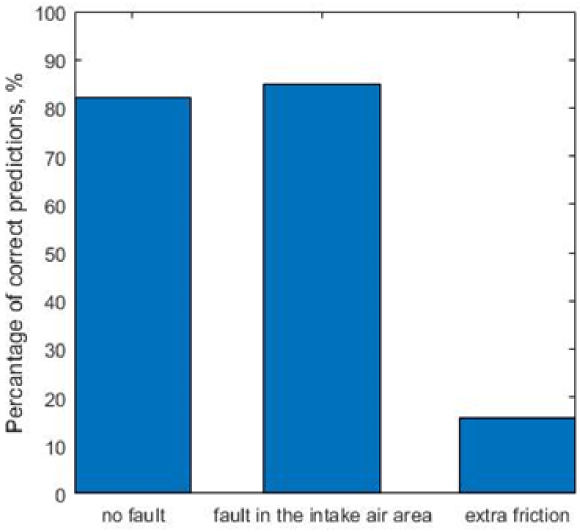

**Figure 7 Statistics of the model correct data classification**

## 4.2 Neural network

Artificial neural network is another example of a computational approach widely applied for the pattern recognition. The algorithm based on the neural network was developed to examine the efficiency in the fan fault detection. The network was trained using the same dataset as the decision tree, using the Neural Network Pattern Recognition toolbox, which is included in the MATLAB software.

The designed network consists of the input layer fed with 4 input variables, 10 hidden layers of neurons and the output layer. The network inputs were chosen based on the model of the decision tree, which was characterised by the best performance. Therefore, the 4 input quantities include average load, a period of the day, standard deviation and peak to peak value. The calculations of features were performed over the time period of 40 min. The result of the output layer of the neural network is a prediction on the fan operating state from the three investigated classes: no fault, fault in the input area and fault causing extra friction.

The activation function used in the network is the non-linear Log – Sigmoid function. The function generates outputs between 0 and 1, which allows to estimate how probable is the prediction of the particular fan working state for each case. The network has the feed-forward architecture; the information flows from the inputs to outputs in only one direction.

After the network training, the efficiency was evaluated at the level of **70%**. Overall 372 observations were classified incorrectly and the highest misclassification was present between the classes: “extra friction” and “no fault”. The neural network accuracy of classification was also evaluated on the data external to the training set. As the output, the neural network determines the probability distribution between classes of the given observation point. For this work, the class with the highest probability was chosen as the network prediction. From the data set consisting of 244 points, 172 events were predicted correctly. The classification accuracy is high and equals to **70.5 %.** The detailed information on the model performance is presented in. *Figure 8*. The graph indicates, that similarly to the decision tree model, the highest classification accuracy is observed for the classes: “no fault” and “fault in the intake air area”. The malfunction causing additional friction was predicted correctly only in 15% of the cases.

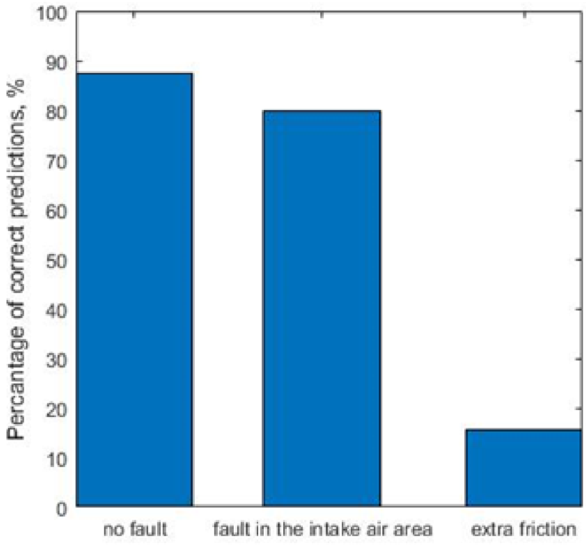

**Figure 8 Statistics of the model correct data classification, neural network**

## 4.3 Discussion of the results

Several decision tree models very analysed to select an algorithm characterised by the best performance. Based on the obtained results for the given dataset, the most efficient model had following characteristics:

- Inputs: power measurements and period of the day,
- Extracted features: average load, average hour and standard deviation,
- Time period over which features are extracted: 40 min,
- The complexity of the tree structure: medium (maximum 20 splits).
- Overall accuracy of data classification: 67.3 % (training set), 68.4 %.( test dataset)

The characteristics of the most efficient decision tree model are selected as the entry parameters for the neural network algorithm. The neural network performance was better evaluated than the decision tree algorithm. However, since the difference in classification accuracy is small (two percentage points only), both methods, decision tree and neural network were considered to be suitable for the given problem. It is important to underline, that the algorithms rarely detected the fault causing extra friction during the fan operation.

The high efficiency of the neural network model is caused by the utilisation of nonlinear functions and adaptive learning. The flexibility in learning is particularly useful when new data is continuously introduced to model. The neural network is capable of reflecting the information of the new instance on the model very efficiently just by changing the weight values. The advantage of the algorithm is also the form of prediction – results are presented in the form of probability distribution. On the other side, in the neural network relationship between input and outputs cannot be visualised and analysed in the way, it is possible with the decision tree algorithm.

The decision tree model is easy to understand. The clear and easy to follow the tree structure enables to analyse the process of data classification and identify errors or wrong assumptions, as it happened for Model V. Decision trees are also fast to be trained and power – efficient. However, their work principle is the simple data division, which can result is the worse classification accuracy. Conversely to the neural networks, decision trees algorithms are not suitable for application which require constant updates of the model. If new data includes some exceptional situation, it will be necessary to train the algorithm one more time.

# 5 Conclusions

The thesis aimed to develop a fault detection system of a fan based on the pattern recognition of its energy consumption. The development of a fault detection system with the use of a history based method and supervised learning algorithm requires large datasets of measurements for both: normal and faulty operating conditions. For this purpose, a laboratory stand was built and that allowed to acquire data of both states. However, in real world application, acquisition of large datasets of measurements carried out during operation with all possible machinery failures is often unrealistic. Moreover, the dataset used for the algorithm training should include records performed during various ambient conditions. The incomplete training dataset can lead to poor efficiency in fault detection and results in the need to perform constant updates. The acquisition of large and complete dataset is a significant drawback of fault detection system based on the process history method.

Another important issue during the system development is the choice of features extracted from the dataset and the algorithm input values. The feature selection is often not a straightforward task. The extracted features should be damage sensitive and robust to the varying machinery operating conditions and background noise. On the other hand, it is recommended to perform computations in an inexpensive way due to the large size of signal data.

The choice of the algorithm inputs should also be done carefully in order to prevent the overfitting problem. The developed model should not be tailored fit to the specific sample, but should reflect the overall population.

The analysis of developed models leads to several conclusions. Both examined algorithms: decision tree and neural network had a similar overall accuracy in the prediction of the operating state, thus both methods are suitable for the given classification problem. The developed algorithms were characterised by the high efficiency in the prediction of clogged air intake area, but the fault causing extra friction was rarely detected. Therefore, not all malfunctions of the fan are possible to diagnose with the use of pattern recognition of the energy consumption. For this type of failures, it is recommended to apply a method, that is more damage sensitive (e.g. vibration analysis).

Both algorithms: decision tree and neural network were trained with the use of Statistics and Machine Learning toolbox and Neural Network Pattern Recognition toolbox. The applications are included in the MATLAB 2016a software. The software is characterised by the user-friendly structure and interface and also by the broad scope of available documentation.

## 5.1 Future work

While carrying out the investigations described in this thesis, it became apparent that there is still lots of work that could be done to develop complete fault detection system. To improve the performance of the already developed algorithms, the training dataset should be extended by a greater amount of measurements carried out in the various conditions. The set of possible faults could also include different types of common fan malfunction, for example, blade failures. Moreover, it is recommended to investigate the possibility of the implementing algorithms in the real-time fault detection system. In that case, the developed models will be connected to the online platform: www.my.eot.pt. The models will continuously process the data sent by the energy meters in order to detect device malfunctions. The work could be also expanded with measurements carried out in the big, industrial ventilation system, where the failure detection is of high importance and results in both, economic and comfort, gains.

# 6. Bibliography

[1] International Energy Agency, How has electricity consumption evolved in OECD countries? *https://www.iea.org/newsroom/graphics/oecd-total-final-electricity-consumption-by-sectoral-share-1971-2013.html*, date of access: 22.10.2016.

[2] S. Emec, J. Krüger, G. Seliger. Online fault-monitoring in machine tools based on energy consumption analysis and non-invasive data acquisition for improved resource-efficiency, *Procedia CIRP 4, 236 – 243,* 2016.

[3] J. Hyvärinen, and S.Kärki. Building optimization and fault diagnosis source book. *IEA Annex 25,* pages:14-18, 1996.

[4] D. Woradechjumroena, H. Lib, T. Promwattanapakeec Fault Detection and Diagnosis Process for Oversizing Design on Multiple Packaged Air-conditioning Units, *Procedia Computer Science 8, 341 – 344,* 2016.

[5] Energy OT, Temperature and optic meter complete manual, *http://eot.pt/en/downloads.html,* date of access: 23.01.2016.

[6] S. Seo, A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets, *BS, Kyunghee University,* 2002.

[7] Q. Sun, P. Chen, D. Zhang, F. Xi Pattern Recognition for Automatic Machinery, fault diagnosis, *Journal of Vibration and Acoustics,* April 2004.

[8] T.W. Rauber, E. Mendel do Nascimento, E. D.Wandekokem and F. M.Varejão. Pattern Recognition based Fault Diagnosis in Industrial Processes: Review and Application, *Universidade Federal do Espírito Santo,* 2010.

[9] T. M. Mitchell, Machine Learning, *McGraw-Hill Science/Engineering/Math*; pages: 52-77, 1997.

[10] A. Krenker, J. Bester, A. Kos, Introduction to the Artificial Neural Network, *Artificial Neural Networks – Methodological Advances and Biomedical Applications,* 2011.