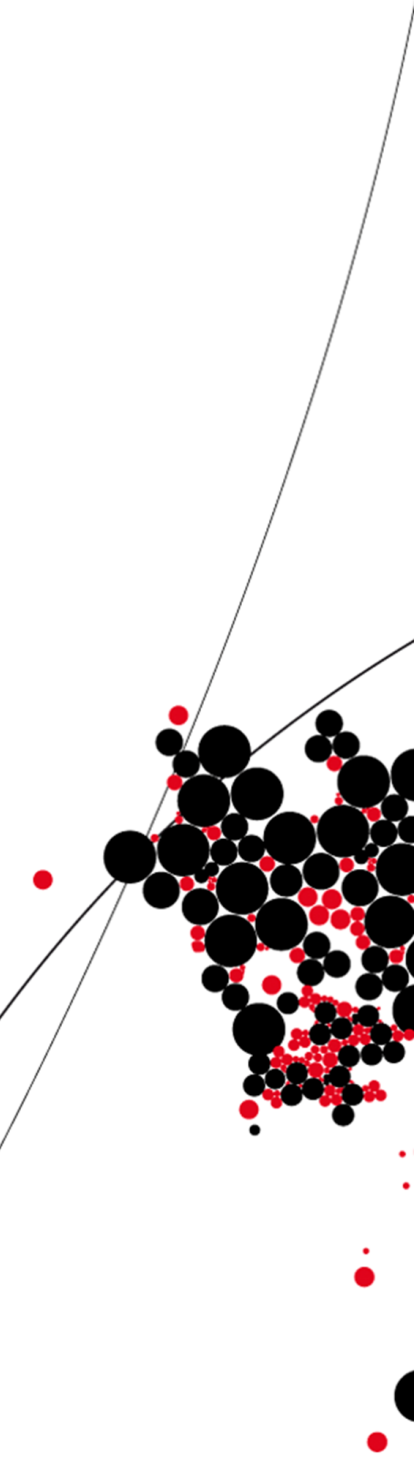
UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Explaining system behaviour in radar systems

Jan Thiemen Postema

Master Thesis
August 2019



Supervisors:
dr. M. Poel
R. aan de Stegge
K. Mussche
dr. E. Mocanu

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Summary

Radar systems are large, complex, and technologically advanced machines that have a very long life-span. This inherently means that there are a lot of parts and subsystems that can break. Thales Nederland develops a whole range of radar systems, including the *SMART-L MM*, one of the world's most advances radar systems, capable of detecting targets at a distance of up to 2.000 kilometers. In order to aid the maintenance and repair of the radar it is equipped with a wide range of sensors, which results in a total of 1.100 sensor signals. The sensor signal are currently processed by two programs, a Built-In Test system, which gives alarms based on a set of rules and an outlier detection algorithm.
In the case of the anomaly detection algorithm the main shortcoming is the lack of explanations. Even though an outlier might be detected, there is still no explanation or label assigned to it. In order to resolve this shortcoming Thales wants to create a system which is capable of recognizing and grouping previously seen behaviour. This results in the following research questions:

1. To what extent can the system state be diagnosed automatically?
    (a) Which techniques are available to diagnose the outliers and which are most suitable given the case described in Section 1.1 and the available data?
    (b) How to assess the quality of the methods used to provide a diagnosis?
    (c) How do the methods selected in RQ 1.a stack up against each other based on the metric found in RQ 1.b and training and diagnostic speed?

Based on an extensive literature review this report proposes to use a clustering algorithm to provide the explanations based on annotations. To find out which algorithm works best, a total of seven combinations are tested. To find out if semi-supervised learning provides a substantial benefit over unsupervised learning for the case of Thales, this report also proposes a novel, semi-supervised constraint-based variant of the Self-Organizing Map (SOM) called the Constraint-Based Semi-Supervised Self-Organizing Map (CB-SSSOM).

The methodology with which these algorithms are tested consists of four steps, (1) pre-processing, (2) dimensionality reduction, (3) clustering and (4) evaluation. This is done on three synthetic data sets and one real data set. The latter is annotated manually by a domain expert to ease the evaluation.

A quick overview of the most important results can be found in Table 1. Most algorithms were tried both with and without dimensionality reduction performed by a Deep Belief Network (DBN).
The conclusion of the report is that unsupervised clustering is most likely not a viable option, although there is still some hope in the form of subspace clustering. However semi-supervised clustering did offer some promising results and could be a viable solution, especially when combined with Active Learning.

Algorithm	**Dimensionality Reduction**	**u/i/s** [1]	**F-score**
k-Means	-	u	0.2460
k-Means	DBN	u	0.4114
c-Means	-	u	0.2460
c-Means	DBN	u	0.2460
SOM	-	u	0.4020
SOM	DBN	u	0.2483
CB-SSSOM	-	i	0.5286

Table 1: Summary of the results obtained on a real data set

[1] u: Unsupervised, i: Semi-Supervised, s: Supervised

Contents

Glossary

AE Auto Encoder

AL Active Learning

ANN Artificial Neural Networks

ANWGG Adaptive Non-parametric Weighted-feature GathGeva

AP Affinity Propagation

BIT Built-In Test

BMU Best Matching Unit

CE Classification Entropy

C-L Cannot-Link

CB-SSSOM Constraint-Based Semi-Supervised Self-Organizing Map

CNN Convolutional Neural Network

D-S Dempster-Shafer

DAE Deep Auto Encoder

DBN Deep Belief Network

DT Decision Tree

EA Evolutionary Algorithm

EM Expectation Maximization

FDD Fault Detection and Diagnosis

FFNN Feed Forward Neural Network

GA Genetic Algorithm

GG Gath-Geva

GK GustafsonKessel

HMM Hidden Markov Model

HUMS Health and Usage Monitoring System

k-NN k-Nearest Neighbour

LDA Linear Discriminant Analysis

LSTM Long Short Term Memory

M-L Must-Link

MLPNN Multi-Layer Perceptron Neural Network

MM Mixture Model

MRW Markov Random Walk

NMI Normalized Mutual Information

NN Neural Network

NWFE Non-parametric Weighted Feature Extraction

OAA One-Against-All

OAO One-Against-One

PC Partition Coefficient

PCA Principal Component Analysis

PSO Particle Swarm Optimization

PUFK Pearson VII Universal Function Kernel

RBF Radial Basis Function

RBM Restricted Bolzman Machine

RF Random Forest

RNN Recurrent Neural Network

SAE Stacked Auto Encoders

SC Subspace Clustering

SDAE Stacked Denoising Auto Encoder

SOM Self-Organizing Map

SVD Singular Value Decomposition

SVM Support Vector Machine

TL Transfer Learning

TSVM Transductive Support Vector Machine

Chapter 1

Introduction

This document describes the results of my master thesis at the University of Twente, which forms the conclusion of the Master Computer Science with a specialization in Data Science. The master thesis is performed at Thales Nederland in Hengelo.
Thales Group S.A. is a multinational company with over 80.000 employees that builds and develops electronic systems for many markets, including aerospace, defence, security and transportation. A large portion of the group's sales are in defence, which makes it the tenth largest defence contractor in the world.
One of the subsidiaries that develops defence systems is Thales Nederland. This company that has its origins in the Hollandse Signaalapparaten B.V. primarily concerns itself with the development of combat management, radar and sensor systems.
One of those radar systems is the *SMART-L MM*. This is one of the largest and most advanced radar systems developed by Thales Nederland and can detect targets at a distance of up to 2.000 kilometers. This project will focus on the *SMART-L MM*.

1.1 Motivation

The *SMART-L MM* radar systems are complex and expensive machines that are crucial to the operations of the defence forces. Therefore unexpected downtime will create serious issues for the operators. This is where the Health and Usage Monitoring System (HUMS) team comes in.
HUMS is responsible for monitoring the complete radar system and giving alarms when a certain part is not functioning as expected. In order to do this a modern radar system has a large number of sensors that monitor many properties including the temperature of the cooling water and the electric current required by the motor that rotates the radar. In the case of the *SMART-L MM* this results in a total of about 1.100 sensor signals. Those signals are currently monitored by two separate

computer programs. The first one is the traditional Built-In Test (BIT) program. This program is based around a predefined set of rules which it uses to fire alarms. So, if a certain sensor reading crosses a threshold value an alarm is fired. The BIT program also tries to find the origin of the problem and, if there are multiple alarms, to group them and find the possible causes. This is partly done based on rules and partly on a model of the radar system. These rules are entered through an elaborate set of Excel sheets and are then parsed into a set of if-else clauses.
The second program that monitors the sensor readings is an outlier detection system. This outlier detection is currently a univariate program that works based on a statistical model of the data. When the likelihood of seeing a certain value falls below a chosen threshold value an alert is send to the user through a monitoring dashboard. This program is slated to be extended by a multivariate outlier detection algorithm that should be capable of better handling the different usage states the radar is in, which have a large influence on its behavior. It should also be able to find correlations between different sensor readings. In the future Thales also wants to include the temporal aspects of the data in the detection algorithm.
The goal of this research is to explain system behaviour in radar systems, however this requires some further specification. A **radar system**, in this case, is a complete radar system such as the *SMART-L MM*, including all of its sub components, such as the cooling system, the rack PCs, send/receive modules, etc. **System behaviour** is defined as the combination of sensor readings and system states at a certain point in time. **System states** describe the current state and activities of the radar. This includes information on whether the radar is rotating or not, whether it is operational and if it is in an eco state. These states have been shown to have a substantial influence on the sensor readings and are therefore an important part of the ”behaviour” of the system. The last part of the title is the term **explanation**, which might be slightly confusing given the growing importance in both literature and practice of explainable AI, which is not what this thesis concerns itself with. In this case it refers to explaining the behaviour of a system based on textual annotations. These annotations will be further described in the next section, however it comes down to assigning labels to periods of time in a data-driven fashion, in other words, classifying the combined sensor reading and system states. During the rest of this report, explaining system behaviour will also be referred to as diagnosing system behaviour.

1.1.1 Annotations

The HUMS team has also been developing an annotation server. This annotation tool is integrated in the monitoring dashboard and can be used by the operators to provide truth values by adding a label to a certain point in time or to an anomaly, or by giving feedback on an existing label, such as the BIT alarms. The user could for example say that an anomaly was indeed a failure. The operator of the radar could also say that the failure was not legitimate and give it a label. There are four types of annotations, which are defined below.

- General; An annotation that spans a certain time period but is not assigned to a specific element or time series
- Time series; An annotation that spans a certain time period and is assigned to a specific data source
- BIT Alarm; An annotation that is assigned to a certain occurrence of a BIT alarm
- Outlier; An annotation that is assigned to a certain outlier

1.1.2 Explanations

One of the things that is currently lacking is an explanation of what is happening on or in the system. When an anomaly occurs it is presented as just that, an abnormal value in a certain time series and when a series of alarms is fired, such as during the startup sequence, all of those alarms are presented without a context. This is a limitation as it is unclear to the operator how he should interpret those notifications. The lack of an explanation also means that it is impossible to filter the outliers or to decide what to do about them, without having in depth knowledge of the radar system. Ideally such an explanation would also include a diagnostic of the underlying cause of the problem, especially when a combination of outliers is detected. For example, when ten temperature readings are reported as outliers because they are too high, the operator does not want to get ten notifications, but just one with the most likely cause, in this case the cooling system.
There are different kinds of behaviour that can occur in the radar system, ”normal” behaviour, when the radar is operating as it should and the resulting data is as expected and ”abnormal” behaviour, when it is not. The latter can be further separated into three variants, abnormal behaviour caused by the operator, abnormal behaviour caused by external factors, such as the temperature, and abnormal behaviour cased by *failures*. Failures in this case are the consequence of *faults*, which arise when

a component does not operate according to its specifications, i.e. a defect. These faults can become gradually worse or it can arise abruptly. The goal of this thesis is to provide a diagnosis for each type of behaviour. For example, when starting up the radar system a certain combination of alarms is raised at the same time. This behaviour is normal, however it would still benefit from an explanation, which in this case can be as simple as ”Startup”. A form of abnormal behaviour caused by external factors is when, due to a low outside temperature, ice has formed on the outside of the radar system, which might interfere with its ability to send and receive. In this case a diagnosis ”Icing on antenna” would be very helpful to the engineers. When the temperature of the radar system suddenly rises a diagnosis could be ”Cooling system offline”. These states might either be available explicitly in the data, such as ”Startup”, which is included as a system state, or be hidden states, which are only available implicitly through other time series. Ideally the resulting explanations can be based on both the explicit and the hidden states, however this thesis will focus primarily on the latter of the two.
Those aforementioned diagnoses, or class labels, are not known apriori and are entered by the operator through the annotation server. The class labels do not have to be related to outliers or alarms though. It could also happen that the operator indicates that he is running an endurance test, in that case future endurance tests should also be recognized. These, user-generated, class labels are used as an explanation of the current system behaviour. The goal of this project is to do a literature review in order to find out if there already exists a method to perform this task that is suitable for the type of data described in the next section. If there is, it will be tested and if there is not, an attempt will be made to create a novel method that is capable of handling the problem.

1.2 Data

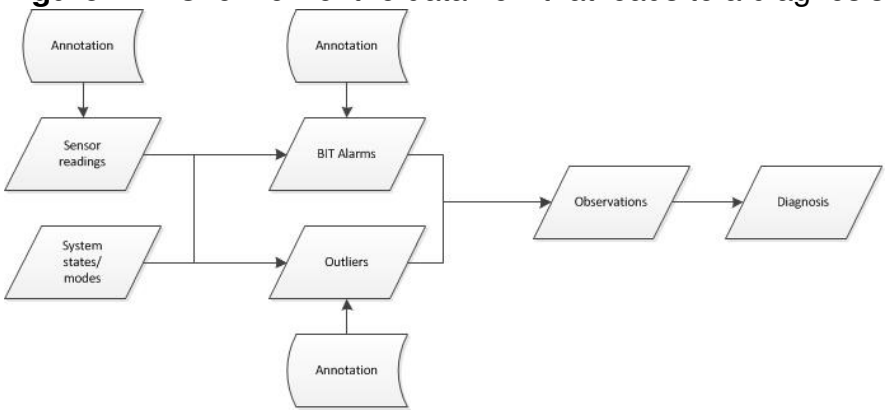
In order to apply machine learning there needs to be enough data. In this subsection the data will be explained and a number of associated challenges will be listed. The available data consists of six parts;

- Sensor readings (continuous data)
- System states (discrete data)
- System modes (discrete data)
- BIT Alarms (discrete data)
- Detected outliers (discrete data)

- Annotations (textual labels)

All of these are collected periodically and are therefore stored as time series. An overview of the data flow and how this leads to a diagnosis that explains the current situation is given in Figure 1.1.

Figure 1.1: Overview of the data flow that leads to a diagnosis



1.2.1 Challenges

There are a number of challenges that result from the data described above. This subsection lists those challenges.
First of all, the dimensionality of the data is quite high, there are about 1.100 sources of continuous data such as sensors and 19.000 discrete data sources. This data is collected over time and is therefore temporal data. The sampling rate differs per sensor, some are collected every 10 seconds and some are only stored when they differ substantially from the previous reading. Given that some outliers, such as those generated during the startup sequence, are more likely to occur than others, there is also a high sample imbalance.
The annotations are currently created by the test engineers, which makes it user generated data, thus it might be ”messy” and sometimes downright wrong. It could for example happen that the maintenance team replaced the wrong part and entered that part as the cause or that they just selected the wrong reason from the list, due to human error. The final challenge is that the annotation server is not yet working. It will take a while for this data to become available. Therefore, should this data be necessary in order to answer the research question, there are a couple of options. The first option is to ask a domain expert to manually label the outliers in (a subset of) the data. The other option is to generate synthetic data with the accompanying labels. However, even when the annotation server is live, there will be a (very) limited amount of annotated data. Even though the amount would increase once the

annotation server is live, they are still outliers and therefore the data is by definition sparse.
Radar systems have a long lifespan of more than 30 years. This means that any fault diagnosis solution should be durable enough to keep functioning throughout the entire lifetime. This long lifetime comes with its own set of unique challenges. For one, it might happen that after 10 years a suppliers decides that a specific part will not be produced anymore, which means that it has to be replaced by another part. This means that the data that is collected is most likely different. The fault diagnosis system could handle such a change in multiple ways, one of which is through an update with a new, pre-trained, model. Another way to handle this issue is having a system that is capable of handling the change and training itself again using operator input. Since this would imply online training it is important to note that any method which uses this solution should be scalable enough to be trained now, on the currently collected data during testing, but also to train itself after 30 years worth of data has been collected.
All of this data will be collected by multiple radar systems, however all of the systems are hand-build, which means that there are slight variations in each machine. Those variations will probably make it difficult to use the trained model of one radar on the other systems. It might be possible to solve this problem through the use of techniques like Transfer Learning, however investigating this is beyond the scope of this project. Given that the market for radar systems is relatively small there will not be a lot of systems sold. Most models are sold somewhere between 10 and 100 times. This means that there will most likely not be enough data to discover the underlying structures in the data.

1.3 Research questions

The main goal of this project is to find a way to provide a diagnosis for the generated outliers automatically based on the data, which should help the operators in maintaining the radar system. This goal has lead to the following research questions:

1. To what extent can the system state be diagnosed automatically?
    (a) Which techniques are available to diagnose the outliers and which are most suitable given the case described in Section 1.1 and the available data?
    (b) How to assess the quality of the methods used to provide a diagnosis?
    (c) How do the methods selected in RQ 1.a stack up against each other based on the metric found in RQ 1.b and training and diagnostic speed?

1.4 Research Method

Before trying to answer the research questions, it is important to decide on a valid research methodology for each of them. The goal of the primary research question is to find a feasible method to perform the task of diagnosing the outliers. This question will be answered through answering the sub questions. To answer RQ 1.a and RQ 1.b a literature review is performed, the result of which can be found in Chapter 3. RQ 1.c combines the two prior sub-questions to compare the found methods and find out which one is most suitable for the case of Thales as described in Section 1.2. Given that there might be (very) few labels available during the course of this project, the methods will also be compared based on a synthetic data set which has a similar dimensionality as the Thales data set, but with artificially introduced outliers and the accompanying labels.

1.5 Report organization

The remainder of this report is organized as follows. In Chapter 3 the literature study that was performed is described. Then, in Chapter 4 a more extensive research methodology is given. The results of the, in the methodology chapter described, experiments are given in Chapter 5, which is followed by the discussion in Chapter 6 and the conclusion in Chapter 7.

Chapter 2

Background

Throughout this report a large number of techniques will be mentioned that the reader is assumed to be familiar with. This chapter serves as a fallback for the cases where this assumption is incorrect. When the reader does have the background knowledge the rest of the report can be followed without reading this chapter.

2.1 Clustering methods

The goal when clustering data points is to group those data points which are more similar to each other than they are to data points in other groups. There are a large number of techniques available that try to accomplish this. This section will discuss those that form an integral part of the report.

2.1.1 K-Means

K-Means might be one of the simplest but also one of the most effective clustering methods. It tries to cluster the data points by assigning them to the nearest cluster center, often based on the Euclidian distance between the data point and the cluster center. The underlying problem is NP-hard, however when applying a heuristic algorithm it is usually possible to quickly converge to locally optimal solution. In this case the Lloyd's algorithm is used to solve the problem. This is an iterative process that uses the following steps to cluster the data;

1. Cluster center initialization
2. Assigning data points to cluster centers
3. Updating the cluster center to be the mean of the assigned data points
4. If the means have not yet converged, return to step two

There are k cluster centers, with k being a predetermined hyper-parameter. The k-Means algorithm has been proven to always converge to a solution. This solution might however be a local optimum [1]. One solution to this problem is to restart the algorithm several times, each with a random initialization. Another, more efficient, solution to mitigate the problem is to use a heuristic in initializing the cluster centers. One such heuristic was proposed by Arthur and Vassilvitskii. They proposed to initialize the cluster centers to be generally distant from each other and proved that this leads to better results than those obtained when the initialization is done at random [2]. They dubbed this initialization strategy *k-means++*. The complete algorithm is described in full detail by Hastie, Tibshirani and Friedman in their 2001 book [3].

2.1.2 Fuzzy c-Means clustering

Each clustering method has its own characteristics and one of those characteristics is whether the clusters are hard or soft. In hard clustering, each label is assigned to one cluster and one cluster only. In soft clustering on the other hand each data point has a degree of membership to a certain cluster. For example, a data point could be 70% likely to be a part of cluster A, 25% likely a member of cluster C and 5% likely a member of cluster B. Fuzzy c-Means clustering is a soft clustering variant of the previously described k-Means technique. A complete description of the algorithm is given by Dunn, who first introduced the algorithm in his 1973 paper [4].

2.1.3 Self-Organising Maps

A Self-Organizing Map (SOM) is a type of artificial neural network that was mainly intended as a tool for dimensionality reduction, however it has also proven itself useful in the field of clustering. A SOM consists of a two dimensional map of ”units”. Each unit is assigned a weight vector in the same dimensionality as that of the original data. The weights are usually initialized at random. Then, through an iterative process, that runs a predetermined amount of times (the number of epochs), the weights of the units are updated to best match those of the original data. The premise here is that this creates a good, two dimensional, representation of the original data. This is done through a concept called the Best Matching Unit (BMU), which is the unit whose weights are closest to those of the selected data point. The weights of all the units are updated to be closer to those of the selected data point. The update is larger when the unit is closer to the BMU. For the complete formula and a more detailed description, please refer to Kohonen’s 1982 paper in which he first introduced the concept [5].

When the training is done, the data points are assigned to a cluster based on their BMU. When data points have the same BMU, they belong to the same cluster. SOMs are a hierarchical clustering method, which in this case means that there is more information available than just which cluster a data point belongs to. For example, if data point A belongs to the unit at $(2, 15)$ and data point B belongs to the unit at $(3, 15)$ then the chances of the two data points are related and should actually be in the same cluster is higher than when they would be at $(1, 3)$ and $(50, 69)$ respectively.

2.1.4 Mixture Models

A Mixture Model (MM) is a linear mixture of multiple probability distributions. The goal is to find the combination of distributions that best describe the data. Each of the distributions describes a cluster and data points are assigned to the distribution that has the highest probability for their value. Even though this sounds good in theory it does require that the parameters of each distribution are estimated. Since it is not clear which of the data points belong to which distribution it is difficult to properly estimate the values. This is where the Expectation Maximization (EM) algorithm comes in. This algorithm, which consists of an expectation and a maximization step, iteratively estimates the parameters to achieve the maximum likelihood. In his 2006 book Bishop gives a more detailed description of MMs [6].

2.1.5 Support Vector Machines

Traditional Support Vector Machine (SVM)s are not a clustering, but a classification tool. They try to calculate the hyper plane that separates the data with the widest margin. This calculation is done using a so-called kernel, of which the two most popular are the Radial Basis Function (RBF) [7], [8] and the Pearson VII Universal Function Kernel (PUFK) [9]. Traditionally, SVMs are a binary classification tool, however it is possible to use them for multi-class classification problems by constructing multistage binary classifiers. This can be done in different manners, the two most popular of which are One-Against-All (OAA) and One-Against-One (OAO) [10]. When the OAA strategy is used, one classifier is trained for each class such that the instances in that class are the positive training samples and all other instances are the negative samples. The sample is then assigned to the class which has the decision function with the highest value. In the OAO strategy a classifier is trained for each combination of two classes. When a new sample comes in it is classified by each classifier and the ”winning” class gets one vote each time. The resulting class will be the one that has the most votes [6]. In the literature the latter

of these two methods proofed most effective [11], [12].
Even though SVMs are traditionally supervised learning algorithms, it is possible to also include unlabelled instances. This is done through a Transductive Support Vector Machine (TSVM). TSVMs calculate the maximum margin solution, while simultaneously finding the most suitable label for the unlabeled instances [13].

2.1.6 Subspace clustering

When working with high-dimensional data there are often subspace in the data that can be identified and utilised to perform clustering. Subspace Clustering (SC) as this is called is a group of clustering techniques. This subset of clustering algorithms tries to cluster the data points based on a limited number of subspaces, so cluster a and b might only exist in the combination of dimension 1 and 2, whereas cluster c only exists in dimensions 3 and 4. In order properly differentiate these clusters they should be looked at in their respective dimensions. More details on this technique and the exact implementation can be found in Parsons' 2004 book [14].

2.1.7 Semi-supervised clustering

In unsupervised learning the algorithms do not use any information about the underlying data to create clusters, whereas supervised learning requires all of the data to be labelled in order to train. A middle ground here is semi-supervised learning, which uses side information about the data to create a more accurate clustering.
Semi-Supervised learning arose in the 60s when the concept of self-learning algorithms was introduced by Scudder. The technique really started to take of in the 70s however when researchers began to incorporate unlabelled data whilst training mixture models [15]. According to Chapelle, Schlkopf, and Zien [15] the term semi-supervised learning itself was first introduced in the context of classification by Merz et al. [16].
There are multiple techniques available to incorporate information into semi-supervised learning algorithms. Two of the most popular are label based and constraint based semi-supervised learning. When semi-supervised learning is done using partial labels it means that there are labels available, however not on the complete dataset. Therefore the algorithm should be able to function with just a subset of the labels. This is the kind of data that is used in algorithms such as TSVMs. In constraint based semi-supervised learning the labels themselves are not available, but rather there are instance level constraints. There are two types of constraints that are often used, Must-Link (M-L) and Cannot-Link (C-L). A M-L between two samples means

that they must be linked in the same cluster, whereas a C-L means that the samples must be in a different cluster. The label based method is more informative, but the constraint based technique is a more generally applicable. It is possible to use labels as the basis in a constraint based setting, but not the other way around [17].

2.2 Dimensionality Reduction

When the data consists of a large number of dimensions it might be required to reduce the dimensionality before using the data for applications such as clustering. This is because relevant patterns are likely to be overshadowed by meaningless data from other dimensions. The general goal of dimensionality reduction is to map the data to a lower dimensional subspace without losing information. There are multiple methods that try to achieve this goal. This sections tries to give some background on three that are important to this report.

2.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a linear transformation algorithm. The goal of the algorithm is to find the directions (lines) of maximum variance. It iteratively starts by finding the one with the maximum variance and is called the first component. It then tries to find another component, with the second highest variance, that has to be mutually orthogonal to the others. The first direction is called the first principal component, the second is called the second principal component, etc. Each component has an eigenvalue associated with it, that indicates the amount of variance that it is responsible for. This means that those components with the lowest eigenvalue are also the ones that could best be thrown away. There are multiple techniques to find those components, however the most popular one is Singular Value Decomposition (SVD). More details on this method can be found in Bishop (2006) [6].

2.2.2 Stacked Auto-Encoder

To understand what a Stacked Auto Encoders (SAE) is, one first has to understand what an Auto Encoder (AE) is. An AE is an unsupervised artificial neural network that tries to find a, lower dimensional, encoding for the data. An AE typically consists of an input layer, a reduction side, that maps the input layer to a hidden layer and a reconstruction side, that tries to reconstruct the original data based on the hidden layer. A SAE or Deep Auto Encoder (DAE) are a deep learning variant of the traditional AE, where multiple AEs are stacked on top of each other to get an even

lower dimensional representation of the data. A more detailed explanation can be found in Aggarwal's 2018 book [18].

2.2.3 Deep Belief Network

A Deep Belief Network (DBN) is in the basis a stack of Restricted Bolzman Machine (RBM)s. A RBM is a shallow generative neural network, which consists of a hidden and a visible layer. The goal of a RBM is to, given a set of outputs, find which inputs produced these outputs. A DBN is a stack of these RBMs where the hidden layer of one RBM is used as the visible layer for the next network. These networks have a large number of applications, however the one that will be used in this report is dimensionality reduction. When a DBN find the input used to create the output it has also found a lower dimensional representation of the output data. After all, the rest of the output can be created by the network and is therefore the same or similar to all other outputs, which means that the input which remains is what makes the sample distinctive. For more information and the formulas, please refer to Aggarwal (2018) [18].

Chapter 3

Literature review

As was established in the Chapter 1, the goal of this report is to find a method to automatically explain the system's behaviour, either through a literature review or by creating a novel method. In this section a literature review will be performed to do so. With the advent of cheaper sensors and storage, the field of monitoring equipment and detecting and classifying faults has become an increasingly popular one. This is often done by inferring the system state from the relevant sensor readings. In the literature this is often called Fault Detection and Diagnosis (FDD). According to the Scopus database a total of 3.397 articles or conference papers have been published on the topic of Fault Diagnosis in 2018 alone. Thales already has a system in place to detect the outliers. However it is not yet possible to diagnose those outliers.

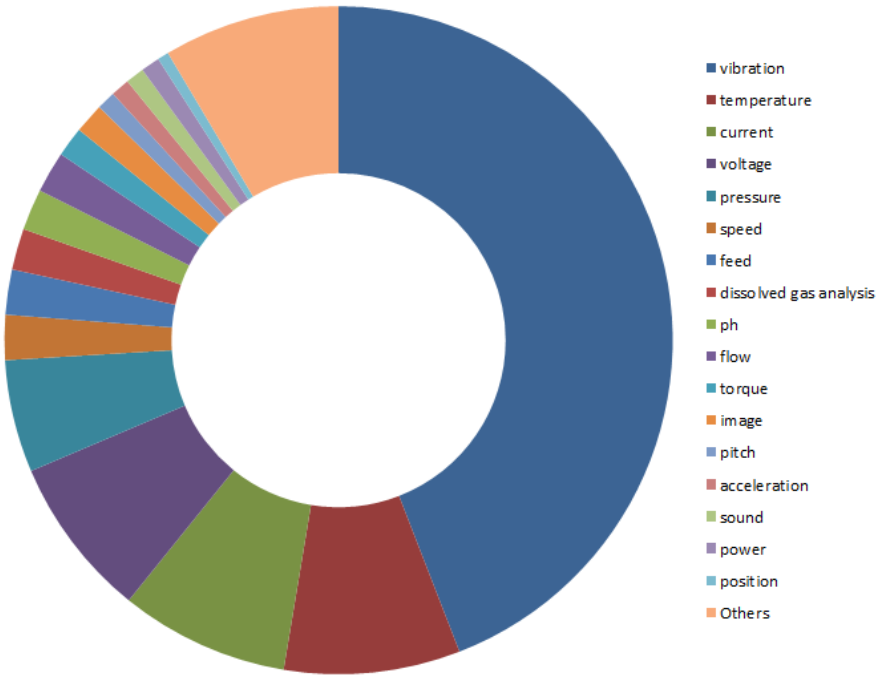
3.1 Fault diagnosis

Isermann identified two methods of performing fault diagnosis [19], data-driven methods and reasoning based methods. Of these two methods the latter mainly relies on a-priori knowledge of the system, whereas the prior is a data driven approach. Given the high complexity of the radar system it is infeasible to encode all the expert knowledge required to diagnose every possible fault into the system. Another complicating factor with the reasoning based approach is that there is most likely not enough expert knowledge to determine every possible fault beforehand. Therefore this research will focus on the data-driven diagnosis method. In order to review the literature available on data-driven fault diagnosis a structured literature review was performed. This search was performed on the Scopus database, with the goal of finding all articles and conference papers related to clustering or classification in the area of fault diagnosis. Since this yielded over 5.500 results, the search was limited to work published after 2016. This resulted in the following search query and a total of 1.280 results at the time of writing (the 13th of March 2019).

```
TITLE-ABS-KEY (fault AND diagnosis) AND (TITLE-ABS-KEY(classification) OR
TITLE-ABS-KEY(clustering)) AND (DOCTYPE(ar OR cp) AND PUBYEAR > 2016)
```

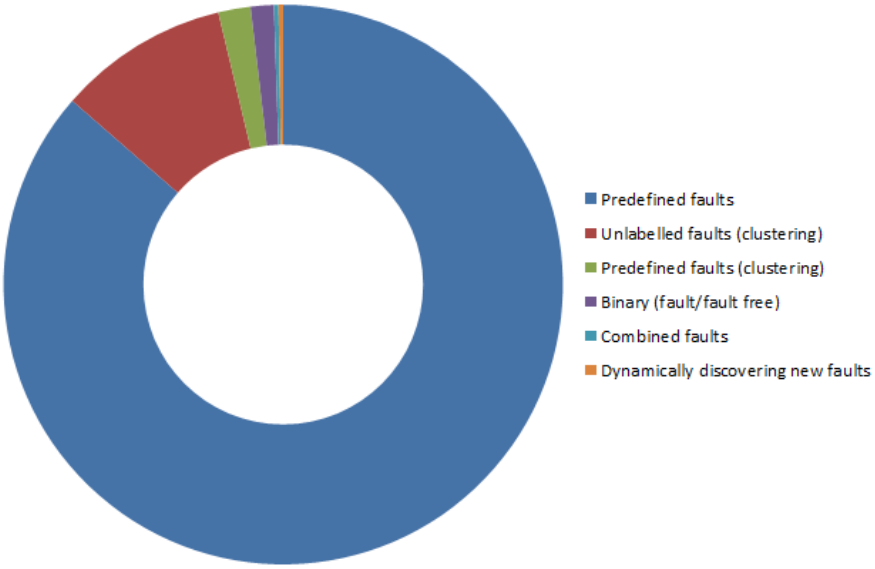
The results of this search were then scanned manually to filter out work that did not concern itself with fault diagnosis in machines or electrical equipment, which brought the number of papers down to 404. Those papers were then categorized based on the techniques used, their application area, the type of data, the method through which the diagnosis was created and whether they are supervised, unsupervised or semi-supervised. The complete categorization can be found in Appendix A. The doughnut charts in Fig. 3.1 and Fig 3.2 are generated based on this categorization. From those charts it becomes apparent that the most popular type of data is vibration data. This was used in almost half of all cases. It is also clear that the most popular type of fault diagnosis is through predefined faults. In those cases there is a set of fault classes and the algorithm assigns one of those to a reading. The second most used method is by clustering the readings and not assigning descriptions at all.

Figure 3.1: Data types used for fault diagnosis



Based on this categorization, four papers will be looked at in-depth to get a deeper insight into the methodologies that were used. Those papers were selected based on how compareable they are to the case of Thales and how much insight they can provide.

Figure 3.2: Types of faults that were diagnosed



3.1.1 Spacecraft

The first selected paper is by Li et al. They tried to diagnose faults in spacecraft based on high-dimensional electric data. Their dataset consisted of 1000 time series, which had 22.800 readings each. Their fault diagnosis system works based on a set of predefined fault modes which are then associated with the data using a classification algorithm. In order to do this the problem was divided into three parts, data cleaning, feature extraction and classification. The first of those three is performed using the wavelet threshold denoising method. This method removes the noise from the signal in an attempt to get a clearer approximation of the underlying signal. Li et al. performed a data-driven experiment to test three feature extraction methods, PCA, SAE and DBN, and four classification methods, Naïve Bayesian Model, k-Nearest Neighbour (k-NN), SVM and Random Forest (RF). When comparing those methods based on their accuracy they found that RF performed best in all situations, irrespective of whether and how dimensionality reduction was applied. However, when applying PCA the accuracy of all classifiers improved, except for RF, which had a worse performance with PCA than without. They also found that the best performing method was a DBN for dimensionality reduction, combined with a RF classifier. This combination had an accuracy of 99.5%. The comparison was done by training the classifiers on a training set, which consisted of 12.800 samples and then testing it on the remaining 10.000 samples. The paper generally does not describe the parameter selection for the models. The only parameter that was described is the number of decision trees in the RF classifier. This value was set to 100 based on a visual inspection of an error rate graph, combined with the training times, which showed that after 100 trees the error rate stayed relatively stable,

whereas the training time did increase substantially [20].

3.1.2 Power systems

Wu et al. designed a fault diagnosis system for power systems where they tried to differentiate between three fault modes. This diagnosis is done using a classification algorithm which sees the fault modes as classes. The test set used by Wu et al. is smaller than the one used by Li et al., this set has nine features and 200 samples. The classification between the three classes was done using a one-vs-all SVM classifier. All of these classifiers are trained twice, once for the voltage data and once for the current data. This leads to two sets of three classifiers. If the classifications of those two are inconsistent with each other the results will then be processed by the fusion step. This fusion step uses the Dempster-Shafer (D-S) evidence theory to decide which of the two classifications to use. However, when applying SVM there are two parameters that need to be determined beforehand, the penalty factor (C) and the kernel parameter (γ). Since there is no way to mathematically find the optimal value for these parameters, there is no straightforward method of finding a good value for them. The way Wu et al. solved this problem was by a grid search of all possible values. However, with two real-valued parameters that are not limited the possible combination are endless. Therefore they used a Genetic Algorithm (GA) to speed the process up and come up with a good value within a reasonable amount of time. A GA is a metaheuristic that draws from the Darwinistic principles of nature to iteratively improve the solution. To prevent overfitting k-fold cross validation was applied in this grid search. The accuracy of the proposed classifier were compared to those of a standard SVM and a SVM that was optimized using Particle Swarm Optimization (PSO). This comparison showed that the GA SVM D-S algorithm performed better than the two opposing approaches [21].

3.1.3 Bearings

Whereas the two previously described papers focused on supervised classification, Zhao and Jia did fault diagnosis using an unsupervised clustering technique. Their application focuses on diagnosing faults in bearings using large amounts of vibration data. They tested their methodology on three cases, one with seven, one with three and one with six fault states. The basis of their method is the Gath-Geva (GG) clustering algorithm. This fuzzy clustering method measures the distance between samples using a fuzzy maximum likelihood estimation. Zhao and Jia identified two

main problems with GG clustering for their application, it counts all samples equally and it has difficulty in selecting the optimal number of clusters. To overcome the former of those issues they incorporated the Non-parametric Weighted Feature Extraction (NWFE) method. This technique assigns weights to the different samples, more effectively use the local information. To determine the number of clusters, the PBMF clustering validity index is used. This index, which looks to balance out intra-class compactness and inter-class separability, provides a score given a number of clusters, the higher this value, the better suited the clustering. To determine the number of cluster K the proposed Adaptive Non-parametric Weighted-feature Gath-Geva (ANWGG) iteratively increases the K by one, until it reaches K_{max} which is set to $\sqrt{N}$, with N being the number of samples. To do dimensionality reduction Zhao and Jia also incorporated a DBN. The proposed setup was tested on three different cases, all of which concerned vibration data from a bearing test setup with a number of labelled faults. The algorithm was compared to GG clustering, Fuzzy c-means clustering and GustafsonKessel (GK) clustering based on four criteria, the PBMF index, the Partition Coefficient (PC), Classification Entropy (CE) and the error rate. The first three are indicators of clustering quality and are calculated solely based on the cluster structure, whereas the error rate is calculated using the labelled samples. The clusters are given labels during the training phase based on the largest membership degree, that is to say, the cluster label is determined by which class is the most common in the cluster. Based on this measure it became clear that the proposed setup performed better than the alternatives, albeit with a slightly higher runtime [22].

3.1.4 General datasets

Hou and Zhang developed a clustering methodology which they proposed as a candidate for fault diagnosis, but did not test in this situation directly. However they took a slightly different approach. Instead of using a density based clustering method, which uses the distance to the cluster center they applied the Dominant Set algorithm, which determines the clustering based on the pairwise similarity between two points. The main benefit of this is that it allows for non circular clusters. Another benefit is that there are no parameters that need to be tuned. The test case that was used to validate the proposed methodology on a variety of public datasets which are published with the goal of providing researchers with a dataset to test their algorithm against. The Dominant Set based algorithm was compared to traditionally popular clustering methodologies such as Affinity Propagation (AP), *k*-means and DBSCAN, based on two measures, the Rand-index and the Normalized Mutual In-

formation (NMI) index, which both compare the results of the clustering to the ground truth to come up with a score of the clustering performance. This test showed mixed signals, the proposed algorithm performed better than the rivals on some datasets but worse on others [23].

3.2 Performance metrics

The metric that is used to measure the performance of a clustering algorithm is an important decision as it defines what is considered ”good”. When class labels are available and the algorithm in question is capable of assigning class labels the most used measure of quality is accuracy, which is defined as the percentage of correctly classified samples. This metric is used among others by the first two papers described above. However, when no ground truth is available or when the algorithm is not capable of assigning class labels, another way of measuring the quality should be found. This has proven to be a challenging task, mainly due to the fact that there is no knowledge of the underlying structure of the data. There are generally speaking three types of criteria by which to judge the performance of clustering algorithms, external-, internal-, and relative criteria [24]. External criteria judge the resulting clusters based on a priori knowledge of the structure of the data. This can manifest itself in different forms, though the most common one is labelled data. Although such criteria would often lead to the most reliable results, it is often impossible to use them outside a controlled test environment. Internal criteria on the other hand only use the data itself to rank the result. The last option compares the resulting clusters to other clusters, generated by the same algorithm. Since there is no knowledge of the underlying distribution of the data and the goal of this study is to compare different algorithms, their performance will be measured based on internal criteria. This type of measurements usually focuses on two main properties, inter-cluster separability and intra-cluster density [25].
Over the years a number of methods have been proposed to measure the performance. Three of those have already been mentioned before, namely the PBMF index, the Partition Coefficient and Classification Entropy. In 2007 Arbelaitz et al. did an extensive comparison study of 18 cluster validation indices [24]. They compared the indices on both a synthetic and a real data set, in different configurations. They found that the three metrics which performed best overall were the Silhouette-, Davies-Bouldin- and Calinski-Harabasz index. The exact definitions of the metrics are given below. In the definitions of the indices the following notation will be used:

X	The data set
N	The number of samples in data set X
C	The set of all clusters
$\overline{c_k}$	The centroid of cluster c_k, defined as the mean vector of the cluster: $\overline{c_k} = \frac{1}{\lvert c_k \rvert} \sum_{x_i \in C_k} x_i$
$\overline{X}$	The centroid of data set X, defined as the mean vector of the data set: $\overline{X} = \frac{1}{N} \sum_{x_i \in X} x_i$
$\mathsf{d_e}(x_i, x_j)$	The euclidean distance between two points x_i and x_j

Silhouette index

The silhouette index measures for each sample how similar it is to its own cluster based on the distance between the sample and the other samples in the same cluster ($\mathsf{a}(x_i, c_k)$) and what the distance is to the closest cluster that it is not part of ($\mathsf{b}(x_i, c_k)$) [26]. The equation below shows the mathematical definition as it was used by Arbelaitz et al. [24].

$$\mathsf{Sil}(C) = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{\mathsf{b}(x_i, c_k) - \mathsf{a}(x_i, c_k}{\max\left(\mathsf{b}(x_i, c_k), \mathsf{a}(x_i, c_k)\right)} \tag{3.1}$$

Where:

$$\mathsf{a}(x_i, c_k) = \frac{1}{\lvert c_k \rvert} \sum_{x_j \in c_k} \mathsf{d_e}(x_i, x_j),$$

$$\mathsf{b}(x_i, c_k) = \min_{c_l \in C \setminus c_k} \left(\frac{1}{\lvert c_l \rvert} \sum_{x_j \in c_l} \mathsf{d_e}(x_i, x_j) \right).$$

Davies-Bouldin index

The Davies-Bouldin index measures the same qualities as the Silhouette index, but does so using the cluster centroids instead. The density is measured based on the distance from the sample to the cluster centroid and the separability is calculated as the distance from the sample to the nearest cluster centroid that it is not part of [27]. This is defined as follows by Arbelaitz et al. [24].

$$\mathsf{DB}(C) = \frac{1}{K} \sum_{c_k \in C} \max_{c_l \in C \setminus c_k} \left(\frac{\mathsf{S}(c_k) + \mathsf{S}(c_l)}{\mathsf{d_e}(\overline{c_k}, \overline{c_l})} \right) \tag{3.2}$$

Where:

$$\mathsf{S}(c_k) = \frac{1}{\lvert c_k \rvert} \sum_{x_i \in c_k} \mathsf{d_e}(x_i, \overline{c_k}).$$

Calinski-Harabasz index

In the Calinski-Harabasz index the separability is not determined based on the individual samples but instead on the distance from the cluster centroid to the global centroid. Just as in the Davies-Bouldin index the density is measured by taking the distance from each sample in the cluster to the cluster centroid [28]. The mathematical definition in Equation 3.3 is courtesy of Arbelaitz et al. [24].

$$\mathsf{CH}(C) = \frac{N-K}{K-1} \frac{\sum_{c_k \in C} |c_k| \mathsf{d_e}(\overline{c_k}, \overline{X})}{\sum_{c_k \in C} \sum_{x_i \in c_k} \mathsf{d_e}(x_i, \overline{c_k})} \tag{3.3}$$

3.2.1 Supervised metrics

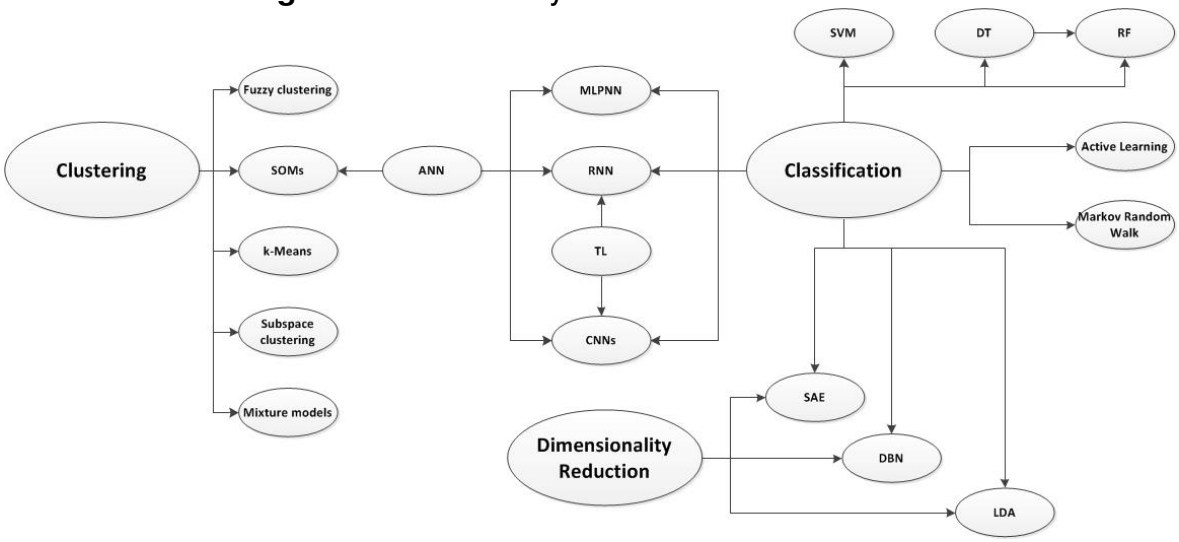
To evaluate the results during training and on the real data set, in other words, when no truth data is available, measures such as the ones above need to be used. However when the actual labels are known it is possible to use other, more informative, measures. There are a number of often used validity measures for supervised classification tasks. These metrics are based on the results from a confusion table. These results include the *True Positives* (TP) (the number of samples that were correctly classified as positive), the *True Negatives* (TN) (the number of samples that were correctly classified as negative), the *False Positives* (FP) (the number of samples that were wrongly classified as positive) and the *False Negatives* (FN) (the number of samples that were wrongly classified as negative).

Sokolova and Lapalma performed a systematic analysis of classification performance measures for, among others, multi-class classification [29]. The measures that are proposed are average accuracy, error rate, precision$_\mu$, recall$_\mu$, F-Score$_\mu$, precision$_M$, recall$_M$ and F-score$_M$. For the precise definition please refer to [29]. The measures they propose for multi-class classification are the same as the ones for binary classification, with the exception that they are averaged for multiple classes. The averaging can be done in two ways; macro averaging (M), where the sum of the measures is averaged, and micro averaging (μ), where the sums of the different parts (TP, TN, FP, FN) are averaged. In both macro- and micro-averaging the TP, TN, FP and FN are calculated in a one-against-all fashion, where a value is true if the class in question is assigned to the sample and false if it isn't.

3.3 Techniques

The goal of this section is to provide an overview of all the methods that have been applied to fault diagnosis over the last two years and to identify other, high potential methods. This is done using the same search query that was used in Section 3.1. A taxonomy of all the methods that are discussed in this section, can be found in Figure 3.3.

Figure 3.3: Taxonomy of the described methods



3.3.1 Fault Diagnosis

This section will list the methods that were found in the fault diagnosis literature. The methods are divided into two groups, classification and clustering.

Classification

One of the most popular methods to perform fault diagnosis is the SVM. In total, 102 of the reviewed papers used some form of SVM. Naturally there are differences between the different implementation such as the use of different kernel functions, e.g. RBF [7], [8] or the PUFK [9]. Another difference between the implementations is how many classes they can differentiate. Traditional SVMs can only differentiate between two classes. When doing multi-class classification instead a choice has to be made between the OAA and the OAO strategy. In the literature the latter of these two methods proofed most effective [11], [12]. Lou et al. used TSVMs to include unlabeled instances as well while doing fault diagnosis [30].
Another widely used supervised classification algorithm is the Decision Tree (DT). A

DT consists of a number of nodes. Each node in a DT splits the dataset based on one attribute. This is done until only the instances of one class remain. In total 18 of the reviewed papers used DTs to perform classification. One of these papers combined DTs with the unsupervised k-means algorithm, which will be described later on, to create a semi-supervised classifier. In their approach two models are trained, one DT based on the labeled data and one k-means based on the unlabelled data. These two models are then combined to classify new instances [31]. A benefit of DTs over most other methods described in this section is it's explainability. Since a DT is only a combination of decisions steps, it is easy to explain how a certain decision was made.

RF is an ensemble learning technique that combines multiple DTs to get a better classification result than with just one DT. In a RF each DT only has access to a random subset of features and training samples, which should prevent issues such as overfitting. RFs are used by 11 of the reviewed papers, however all of these used it in a supervised fashion.

A technique that has seen more and more widespread use in recent years are Artificial Neural Networks (ANN). ANNs are popular because of their widespread applicability, they have been used for a lot of things, including image classification, sales forecasting and fault diagnosis. There are different implementations of this idea, the simplest of which is probably an Feed Forward Neural Network (FFNN) with only an input and an output layer. When more layers are added, this forms a Multi-Layer Perceptron Neural Network (MLPNN). A MLPNN has one or more hidden layers between the input and the output layer where calculations are performed. The training of MLPNNs is usually performed in a two phased fashion, a *forward phase* and a *backward phase*. The forward phase is used to calculate the loss function and the backward phase is used to update the weights. This technique is called *back propagation* [18]. Since 2016 the MLPNN has been applied to fault diagnosis a total of 29 times, sometimes on its own, sometimes in cooperation with other algorithms, such as DT [32] or the Hidden Markov Model (HMM) [33].

Another ANN architecture that is often used for fault diagnosis is a Recurrent Neural Network (RNN). One of the main advantages of RNNs over traditional feed-forward networks is that they can use an internal memory to process sequences of data. This is especially useful in sequences of data, such as text or speech. Another area where this comes in handy is sensor data, which is usually represented as time series data. Basic RNNs have been used four times in the reviewed papers. There are however also other variants of RNNs, such as Long Short Term Memory (LSTM). LSTMs try to solve one of the main problems in RNNs, namely the vanishing gradient problem [18]. This problem can be encountered during the training of RNNs and implies that a certain gradient tends to zero. This type of network has been

used five times in the field of fault diagnosis, of which one time in conjunction with a Convolutional Neural Network (CNN) [34].
The origin of CNNs lies in the field of image recognition, where it has proven to be one of the most successful Neural Network (NN) architectures. Based on this success CNNs have been used in all kinds of fields, including object detection and text processing. Since 2016 CNNs have been applied 28 times to fault diagnosis problems, sometimes in conjunction with other algorithms such as DBN [35] or HMMs [36].
One of the issues with CNNs is that it takes a lot of labelled data to train the network. One solution to this is to use Transfer Learning (TL). TL is a technique that uses information from a previous task to speed up training on this task. It requires another neural network that is trained on different, but similar data. For example if you want to train a NN that identifies cows you could use a NN that classifies cats and dogs as a basis. You then delete the existing loss-layer while keeping the rest of the network and train it again on the, smaller, set of labelled cow images. This technique is not only used for CNNs but also for example for RNNs [18]. It has been applied to fault diagnosis a number of times during the last years, among others by Hasan et al. [37] and Wang and Wu [38]. DBNs have also been applied to fault diagnoses. When applied to classification, which is their main task in the fault diagnosis literature, the DBNs are mainly used in a supervised manner [39]–[41]. There have however also been papers that applied DBNs to fault diagnosis in a semi-supervised or unsupervised manner, although this is mainly in the role of a dimensionality reduction algorithm. Zhao and Jia used a fuzzy clustering method for the semi-supervised clustering and a DBN to perform dimensionality reduction in the context of rotating machinery [22]. The same combination was also applied by Xu et al., though in their case the clustering was performed completely unsupervised and in the context of roller bearings [42].
AEs originate in the field of dimensionality reduction and are an unsupervised neural network that try to find a, lower dimensional, encoding for the data. SAEs or their relative the Stacked Denoising Auto Encoder (SDAE) [43] have been used in the fault diagnosis literature to do both classification, in conjunction with another layer or algorithms such as a SVM [44] or a softmax classifier [45], [46] or semi-supervised clustering, where the data is first clustered unsupervised and the clusters are then improved using a small set of labelled data [47].

Clustering

The SOM is a paradigm that was introduced by Kohonen in 1982 [5]. In contradiction to most of the aforementioned architectures, the SOM does not use error learning

but instead uses competitive learning [18]. Like AEs, SOMs are typically used for dimensionality reduction, but can also be used for clustering. This was done, among others, in the field of fault diagnosis by Blanco et al. [48] and eight others.
Even though neural networks are widely popular nowadays there are also other technologies that offer promising results. One of the most popular of these is k-means clustering and its supervised relative k-NN classification. K-means is used a total of five times to perform fault diagnosis [49], often in conjunction with other methods such as HMMs [50] or SVMs [51].
The k-NN algorithm is about as popular when it comes to fault classification, seeing as it was applied six times since 2016.
Another popular clustering technique is fuzzy clustering. The main difference between fuzzy clustering and traditional or ”hard” clustering methods such as k-means is that fuzzy clustering allows data points to belong to multiple clusters. One of the most popular variants of fuzzy clustering is fuzzy c-means clustering. This technique was first introduced by Dunn in 1973 [4] and is highly similar to the traditional k-means clustering. It was successfully applied to fault diagnosis [52], [53] a total of 17 times in different fields, including roller bearings [54] and wind turbines [55].
When working with high-dimensional data there are often subspace in the data that can be identified and utilised to perform clustering. SC uses these subspaces to identify clusters in the data. SC has been applied to fault diagnosis in bearings by Gao et al. [56].

3.3.2 Classification in high-dimensional sparse data

In the previous section it became apparent that most of the techniques that were recently used in fault diagnosis are based on classification or clustering algorithms. In order to find ways to expand this body of methods a second search was performed. This search focused on the type of data as it was described in Section 1.2, namely sparse, high-dimensional data, with a small and incomplete set of labels, although the labels are not explicitly taken into account in the search query, which looks as follows:

```
(ab:(classification) OR ab:(clustering) OR ab:(categorization) OR ab:(grouping))
AND (ab:(sparse) OR ab:(sporadic) OR ab:(infrequent) OR ab:(scattered)) AND
(ab:(high-dimensionality) OR ab:(high dimensionality))
```

This search query was performed on the University of Twente WorldCat database and resulted in 262 results at the time of writing (the 18th of March 2019). The

results were categorized based on four characteristics, the main application (dimensionality reduction, classification, regression or outlier detection), the techniques that were used, the application field and whether they were supervised, unsupervised or semi-supervised. On the results snowballing was performed, up to three levels deep, with increasing scrutinization. The complete result can be found in Appendix B.

A large portion of the literature focused on dimensionality reduction, however there are a number of interesting algorithms that have not been used for fault recognition yet. Given that the data set has a (very) limited amount of labels, only unsupervised and semi-supervised techniques will be considered.

Linear Discriminant Analysis (LDA) is a technique that is mainly used for supervised dimensionality reduction and classification. Zhoa et al. used this technique in a semi-supervised manner to perform image recognition by using a process called label propagation, where the labels are propagated to the unlabelled instances. Those new labels are called 'soft labels' [57]. To make sure that LDA has not been used in fault diagnosis yet, a search was done. This turned up a total of 82 results where LDA was used for fault diagnosis, including for Re-useable Launch Vehicles [58] and motor bearings [59], meaning that the use of LDA in fault diagnosis is not new.

Another clustering algorithm that came up in the search is the MM. MMs try to separate the clusters based on a probability distribution, for which the parameters are approximated using a technique called EM [60], [61]. Even though it did not appear in the search for papers after 2016, MMs have been used to perform fault diagnosis before, among others Luo and Liang [62] and by Sun, Li and Wen [63].

One way of doing semi-supervised learning is through TL, a technique that was explained in Section 3.1. Self-taught learning is another technique that works in a similar manner. However, self-taught learning does not assume that the additional, unlabelled, samples have the same distribution or class labels [64]. This techniques has not been applied to fault diagnosis yet.

Although they have delivered great results in the past, unsupervised learning methods and techniques like TL might not always be up to the task. In that case a supervised algorithm could be required. However, it is usually still too expensive to label all the data. This is where Active Learning (AL) comes in. AL is a supervised algorithm that selects instances from the pool of unlabelled samples and asks an "oracle", usually a human annotator, to label them. This limits the amount of work that the oracle needs to perform, while still being able to train a supervised classifier [65], [66].

The last method that was not found in the fault diagnosis search was the Markov Random Walk (MRW). This technique is used to perform semi-supervised classification [67]. In this technique the classified instances are used as starting points for

the random walk and the parameters are optimized using the EM algorithm. Like self-taught learning, no papers could be found in the Scopus database where this tactic was applied to the field of fault diagnosis.

3.4 Summary

The literature review on machine learning models in fault diagnosis has resulted in a lot of promising results. A large number of clustering and classification algorithms have already been applied successfully to this field. However there are still a number of shortcomings when it comes to the specific case of Thales as it was described in Chapter 1.
The main problem here lies in the type of data that was used as an input for the diagnosis. In the reviewed literature it often concerned either vibration data or image data. In the case of Thales the data consists of a large number of sensor measurements, which makes for a very different situation. A large number of the reviewed papers also assumed that there were a limited number of fault classes or that all data is labelled. This is not the case here. The majority of the data is unlabelled, which makes is infeasible to apply supervised classifiers. This means that only a few of the proposed algorithms can be applied to Thales' problem as only semi-supervised and unsupervised algorithms qualify. Out of the algorithms described in Section 3.1 the following algorithms are capable of semi-supervised or unsupervised learning: TSVMs, k-means clustering, SOMs, fuzzy c-means clustering and SC. However most of these, with the exception of subspace clustering, will not perform well on high-dimensional data without some sort of preprocessing such as feature transformation or feature extraction. In the literature the process of feature extraction was often done using DBNs or SAEs.
In Section 3.3.2 five methods were identified that have been applied successfully to classification or clustering in high-dimensional data with few or no labels. Two of these, LDA with label propagation and MM have been applied successfully to fault diagnosis, albeit before 2017. In the literature there was no example found where the three remaining methods, self-taught learning, AL and the MRW were applied to fault diagnosis.
An overview of the previously mentioned methods can be found in Table 3.1. This table shows that the most popular method is fuzzy clustering, in one of its forms (e.g. GG clustering, fuzzy c-means clustering). There are three methods which have been applied to fault diagnosis 9 times or more, fuzzy clustering, k-Means clustering and SOMs. All of the methods that were not applied to fault diagnosis yet require additional data. The MRW needs a graph of the system and self-taught learning needs a similar dataset to train on, both of which are not available. Therefore these

methods are deemed infeasible in the situation of Thales during the course of this project, however they would warrant future research. Another method that requires extra data is AL. This method needs an oracle that can provide feedback. Given that the time and availability of the oracles is limited, it would be uncertain if it is possible to implement and test this method within the course of this project. Given this constraint, AL will not be implemented, however it is advised that this method is researched further in the future.

Based on the background set out in the previous section and popularity in the fault diagnosis literature the following methods were selected: fuzzy clustering, k-Means clustering and SOMs.

Another area of research that showed some promising results in the literature but was not very popular was semi-supervised learning. In order to test whether this is indeed of added value, a semi-supervised method will also be tested. In order to easily compare this approach to the unsupervised methods, a semi-supervised variant of the SOM will be used.

The performance of those techniques will be compared on the Thales dataset and an artificial dataset, which will be described later in Section 4.1. Parameter estimation will be performed using a Genetic Algorithm as was proposed by, among others, by Wu et al. [21]. Finally, every method will be tested without and with dimensionality reduction performed by a DBN, which in the research of Li et al. turned out to outperform PCA and SAE [20].

Technique	**Times applied**	**Remarks**
Fuzzy clustering	23	
K-Means clustering	9	
SOM	9	
LDA	3	Needs at least one sample of each class
MM	3	
TSVM	1	Needs at least one sample of each class
Subspace clustering	1	
Self-taught learning	-	Requires a relatively similar dataset
AL	-	Requires interaction with an expert
MRW	-	Needs a graph of the system

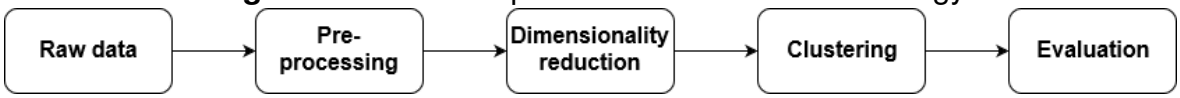
Table 3.1: Unsupervised or semi-supervised fault diagnosis methods

Chapter 4

Methodology

A valid research methodology is the basis of every good piece of research. Based on the conclusions from the previous section, the methodology that was proposed in Section 1.4 can be extended. RQ1.a and RQ1.b have already been answered in the previous section through a literature review. The one remaining question is RQ1.c. This question will be answered in a data-driven fashion, as is common in the related literature described above. To evaluate the three algorithms that were selected, k-Means clustering, fuzzy c-Means clustering and SOMs, they will be tested on multiple data sets. The evaluation is done in three steps, (1) pre-processing, (2) dimensionality reduction, (3) clustering and (4) evaluation. Figure 4.1 provides a visual representation of these steps, which will be described in more detail in the sections below. During steps two and three there are a number of hyper-parameters that need to be determined. The search spaces in which those values can lie are already described in each respective subsection. The manner in which the search is performed is described at the end of this chapter in Section 4.6. Before going into detail on the steps themselves, the data sets will be discussed shortly.

Figure 4.1: Visual representation of the methodology



4.1 Data

As was mentioned before, the different methods will be tested on three synthetic data sets and one real data set. The synthetic data sets are created by domain experts with the goal to be similar to the real data set. In this case synthetic data sets are used because it is difficult and expensive to get a real data set. Using synthetic data makes it possible to quickly test the performance of the algorithms in

a controlled environment. Since it is possible to control the number of dimensions, it is also easier to visualize the data, which helps to get a quicker insight into the data. The real data set covers a period of five months and is annotated manually by a domain expert. This was done through a visual inspection of the time series. Since domain experts are mere humans, it is possible that this annotation is not 100% correct, however it should serve as a good benchmark for the algorithms. More details on these data sets are provided in Appendix C.

4.2 (1) Pre-processing

The data set consists of a large number of sensor reading that all have their own characteristics. If they are compared without any pre-processing, based on the euclidean distance, observations that have a different magnitude will have a disproportionate effect on the results. Therefore some data pre-processing is required. The features will be scaled using a technique called z-normalization, where the mean of the feature is subtracted from the data and it is then divided by its variance. The exact formula is $X_i = \frac{\mu_X - X_i}{\sigma_X}$ with X being a feature vector, i being a sample in that feature vector, μ_X being the mean of the feature and σ_X being the standard deviation. As a further pre-processing step, all alarms will be removed from the data, since according to Thales' domain experts most of the variables are merely an alarm raised when another reading crosses a certain threshold. Thus this data is already derived from other observations and will be disregarded.

4.3 (2) Dimensionality reduction

As was described in Section 3.4, the dimensionality reduction will be performed by a DBN. Each of the clustering methods will be tested both with and without dimensionality reduction, to get a clear view of its influence on the results. When implementing a DBN there are a number of hyper-parameters, most of which are found in most neural networks. These includes the activation function, the learning rate, the number of epochs, the batch size and the structure of the network (i.e. the number of hidden layers and the number of neurons per layer). To determine the size of the search space the recommendations of Bengio for hyper-parameters in neural networks will be used [68]. As activation function only the ReLu function will be used, since this is one of the few activation functions that do not suffer from the vanishing gradient problem [69]. According to Bengio the optimal learning rate for cases where the input is mapped to a range of $(0, 1)$, lies in the range $(10^{-6}, 1]$. Since a change of 0.01 will be relatively higher when it is done on a value of 0.02 than when

it is added to an initial value of 0.1, the values in the range will increase with 10% of their value at each step (i.e. from 0.01 to 0.011 and from 0.1 tot 0.11). Since there is no ”rule of thumb” available for the number of epochs, the search range will be limited at the interval $(5, 250)$, with a uniform distribution over the interval. This range is chosen to get a good overview of all values, while still maintaining a reasonable training time. The batch size is an important hyper parameter to optimize as it determines how many times per epoch the parameters are updated. There are generally speaking three options, a stochastic, or online, training where the batch size is one (i.e. there are no batches), a mini-batch, where the data set is divided into multiple batches which are all larger than one and ”normal” batch training where all samples are entered in one batch. In the case of a large training set, the batch method is infeasible as it will most likely have trouble converging or at least be very slow. In accordance with the recommendation from Bengio the batch size will be searched over the interval $(1, 200)$. The network will be tested with zero to five hidden layers and 25 to 200 neurons per layer.

4.4 (3) Clustering

As was concluded in the previous chapter, a total of three clustering methods will be compared to each other on three synthetic data sets and one real data set. This section describes the algorithms that are compared to each other.

4.4.1 k-Means Clustering

The k-Means clustering algorithm is a popular algorithm to partition a data set into k different clusters. The details of the algorithm were described in 2.1.1. As an initialization schema *k-Means++* will be used. The k-means algorithm is relatively simple, which means that there is only one parameter that needs to be estimated, namely k. This parameter determines the number of clusters that the data set needs to be divided into. Since, given the nature of the problem, there are always a number of labels available, the number of distinct available labels will be used as a lower bound for this parameter. The upper bound of the interval will be given by the lower bound plus 250. This upper bound has a disadvantage though, when the number of samples is relatively low, e.g. 500, and the number of clusters is 250, the clusters contain very few samples. This makes that the F-score gives a high value, however the clustering barely has any meaning. Therefore the upper bound will be set to the minimum of 250 plus the number of distinct classes that are known a priori and the number of samples divided by 5.

4.4.2 Fuzzy Clustering

Fuzzy clustering was implemented using the Fuzzy c-Means algorithm. This technique, which is related to the k-Means algorithm which was described in the previous section, is used because of its widespread usage in the related works. The implementation is based on the one put forward by Bezdek, Ehrlich and Full in their 1982 paper [70]. Since fuzzy clustering is a soft clustering methodology, samples are not automatically assigned one specific cluster, but instead have a membership to each cluster. In order to be able to compare the results from the fuzzy clustering to that of the other clustering methodologies, the data points are later assigned to the cluster with the highest membership factor. In fuzzy c-Means clustering there is also just one hyper parameter that may be tuned, namely the number of cluster centers. Given that this is the same parameter as the one that needs to be tuned in k-Means clustering, the search space will be the same as well.

4.4.3 Self-Organizing Maps

The concept of a SOM was first introduced by Kohonen in 1982 [5]. In this case the SOM is used as an unsupervised clustering algorithm, however it can also be used for dimensionality reduction. SOMs revolve around BMUs. A SOM consists of a map of nodes. In this map the different units are assigned weights, which are then updated based on the samples closest to them (i.e. the samples for which the unit is the best match). The clustering is done based on these BMUs. Each BMU represents a cluster and the samples are assigned to the BMU, and thereby to the cluster, that best matches them. In order to train a SOM there are four parameters that need to be set, the number of epochs, the learning rate and the width and the height of the map. For the number of epochs and the learning rate the same search space is used as for the DBN. The height and width of the network is, among other things, dependent on the number of samples and the number of clusters in the data. The width and height together determine the size of the map ($size = width * height$). The lower bound for the size of the map is the number of distinct clusters. In this case the lower bound is given by the number of distinct, known, class labels (see Section 4.4.1 for more details). The upper bound is given by the number of samples in the entire data set times two.

4.4.4 Constraint-Based Semi-Supervised Self-Organizing Map

As was concluded in Section 3.4, a semi-supervised of variant of the SOM will be tested, since this makes it possible to directly compare it to the results of the unsupervised variant and determine if semi-supervised learning provides a substan-

tial benefit. Some background on semi-supervised clustering can be found in Section 2.1.7. Since the constraint-based variant of semi-supervised clustering is more versatile than its label-based counterpart the learning will be performed using constraints. This offers more possibilities when it comes to gathering user input, such as asking the question ”Do these two samples belong to the same fault?”, which makes it easier for the operator to provide input. In the Scopus database only one paper could be found that worked on incorporating constraints into the SOM. Allahyara, Sadoghi and Haratia did worked on such a variant, however they focused solely on an online variant, whereas the focus of this report is on offline learning [71]. Therefore this section proposes a new extension to the SOM to incorporate constraints while doing offline learning. The details of this implementation are described in Appendix D.

4.5 (4) Evaluation

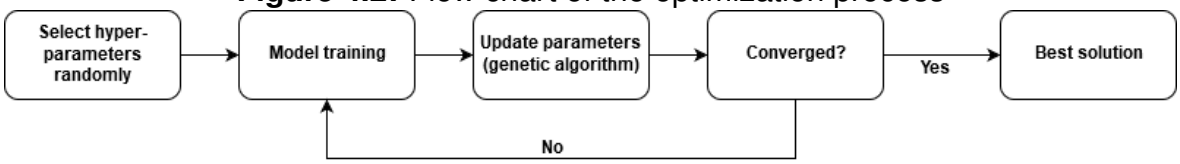
Based on the results of the literature review performed in Section 3.2 two metrics are selected to evaluate the results of the clustering, one that does not require truth data and one that does.
From the definitions given in Section 3.2, it shows that the Davies-Bouldin index and the Calinski-Harabasz index are both calculated based on the cluster centroids, whereas the Silhoutte index is calculated by taking the distance between each sample. One of the main pitfalls of the centroid based methods is that they, due to their nature, favour circular clusters. Since this is an assumption that can not be made at this point, the Silhoutte index will be used when there is no truth data available, even though it is a more computationally intensive measure.
When selecting a supervised quality metric an important point to keep in mind here is that the data is highly imbalanced. When using macro-averaging, all classes are weighted equally, whereas in micro-averaging the classes with more samples are favoured. Given Thales' imbalanced data set, macro-averaging will be used. The imbalanced data set also implies that some measures are more suitable than others. For example, if an event occurs only three times in a data set of 200 samples and the classifier gives none of the samples this label, it will still have an accuracy of 98.5%. Therefore measures such as recall and precision are more suitable. In this case the F-score will be used. This score is the harmonic mean of the recall and the precision and provides a comprehensive representation of performance, without suffering the effects of an unbalanced data set.
Something that has to be kept in mind when using the F-score, or other metrics based on the recall, is that the annotations in the real data set might be inconclusive, as was mentioned in the previous section. In other words, the algorithm might

pick up on an anomaly that the domain expert did not notice.

4.6 Hyper-parameter estimation

Hyper-parameter estimation is one of the most challenging tasks in machine learning, most of all because there is usually no way to mathematically determine the optimal value. Over the years a number of methods have been proposed. These methods include widely used tactics such as grid searches, which enumerate over all possible solutions in order to find the best performing one, and random searches, where possible combinations are selected at random from the total population. The former of these methods works well when the size of the population is limited and training the model is fast, however it is very computationally intensive when the population size or the training time increases. The random search does not have this issue, given that it only tries a subset of all methods, however, given that the hyper parameters are selected randomly from the population it follows logically that there is a substantial chance that the optimal parameters are not selected.
Another way to estimate the hyper parameters is through an Evolutionary Algorithm (EA). Using this method solves the two aforementioned problems to a large extend and has therefore been often used in the literature described in Chapter 3. EAs try to solve black-box optimization problems, of which the hyper-parameter estimation problem is one. The implementation of the EA for hyper-parameter estimation is done using the python package sklearn-deap [72]. A flowchart of the optimization process is given in Fig. 4.2.

Figure 4.2: Flow chart of the optimization process



Chapter 5

Results

This chapter describes the results that were obtained by testing the classification performance of the selected methods on four different data sets. The data sets and their characteristc are described in full detail in Appendix C and are quickly summarized in Table 5.5. A complete overview of all results can be found in Table 5.1 and Table 5.2. The former contains the Silhouette scores and the latter contains the F-scores. In Table 5.3 an overview is given of all the hyper-parameters as they were determined by the genetic algorithm.
Unfortunately, due to time constraints it was not possible to test the Constraint-Based Semi-Supervised Self-Organizing Map (CB-SSSOM) on the synthetic data sets. Therefore these results will only be provided and discussed for the real dataset since this is the most informative and relevant to the case of Thales.

5.1 Synthetic data set

The first of the three synthetic data sets is also the smallest. It is used primarily to test whether or not the algorithms work as expected. Also, if they already perform poorly on a small data set the changes are slim to none that they will perform well on a large, high-dimensional data set. As was mentioned before and is described in detail in Appendix C, the synthetic data set has one type of anomaly, which is visible in most of the time series. To validate the workings of the algorithms, they are first tested on a synthetic data set with just two time series, thereby creating a very basic clustering problem. The results of these tests were generally favourable for all methods. The resulting clusters may be found in Fig. 5.2. Visually speaking DBN-SOM seems to have the best separation between the clusters. All four obvious clusters have their own colour. This is also visible in its silhouette score, which is the highest of all methods. DBN k-Means also has a high silhouette score, however from a visual inspection the separation does not seem substantially better than that

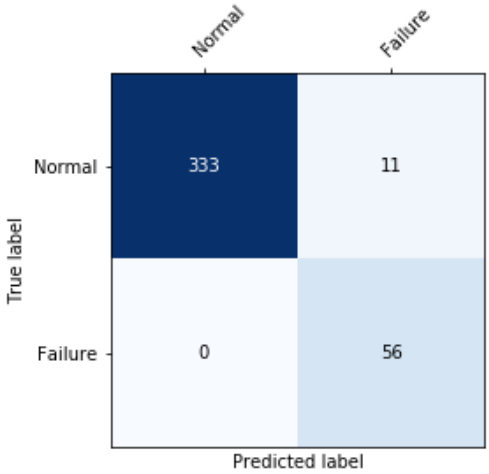
of k-Means, c-Means or SOM. When looking at the confusion matrices in Fig. 5.1 and the F-Scores in Table 5.2 one noticeable thing is that all algorithms have the same F-Score and the same confusion matrices. This probably means that those samples are very hard or impossible to distinguish using any of these techniques. Based on the aforementioned visual inspections of the clustering results the hypothesis is that DBN-SOM will perform best on the other data sets.

When applying the methods to the second synthetic data set, the results start to fluctuate more. The best score is achieved by the combination of DBN and k-Means, however the improvement over normal k-Means is small. In the case of c-Means the score is also improved when the dimensionality is reduced beforehand, however in the case of SOM doing dimensionality reduction prior to clustering has a detrimental effect, which is not in line with the previous results or the proposed hypothesis. This might be explained by the fact that both of these methods attempt to find a lower dimensional subspace to project the data on, which may lead to too much information being discarded. This reduction in dimensionality might also be the reason why the SOM performed worse than the other two algorithms. More detailed results can be obtained when looking at the confusion matrices in Fig. 5.3, which show that the combination of DBN SOM has the tendency to classify all instances as ”Normal”, which given the label propagation method used would imply that the resulting clusters are too large. The same problem, to a lesser extend, is also visible when SOM is used on its own and, to an even lesser extend, when c-Means is combined with DBN. Each method made mistakes on different parts of the data. This implies that the methods complement each other, so a committee of clustering algorithms might be worth looking into. When hypothesizing about the next data set, which is larger only in the number of samples, the same pattern is likely to arise, with DBN k-Means having the highest F-Score.

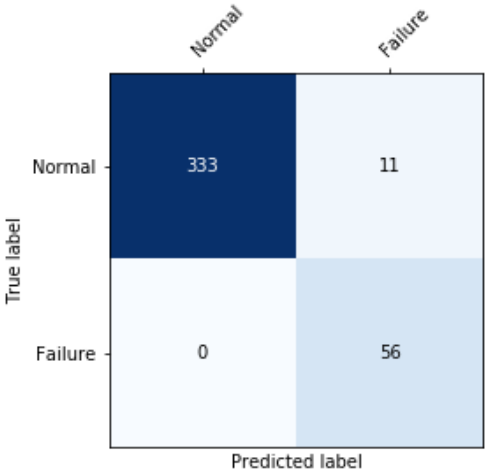
In the first two variants on the data set there were only 400 samples. Given that the real data has substantially more samples it is important that the algorithms are also tried on this kind of data. Therefore the same model is used to generate a new data set, although this time with 10.000 samples. Even though the underlying distribution has not changed, other core characteristics of the data set have. Therefore the hyper-parameters are determined again on this new data set. The results are again very different than in the two previous tries. This means that each method responds differently to an increase in the number of samples. In all of the cases the performance of the methods worsened, however this was most visible in the case of the DBN-SOM combination. The F-Score of this method came down to 0.4763, compared to 0.9472 when the data set only had 400 samples. A similar decrease,

Figure 5.1: Confusion matrices for the different algorithms on synthetic data set 1a. All methods have the same confusion matrix.

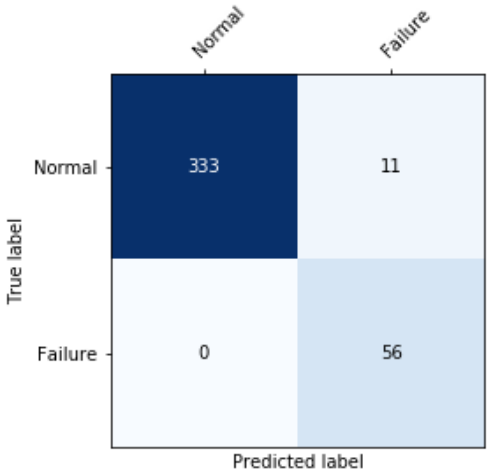
(a) k-Means



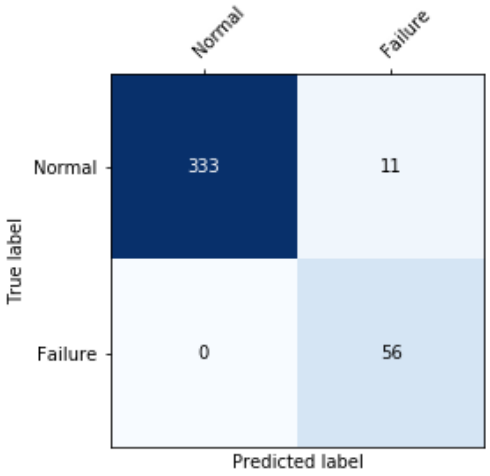
(b) DBN k-Means



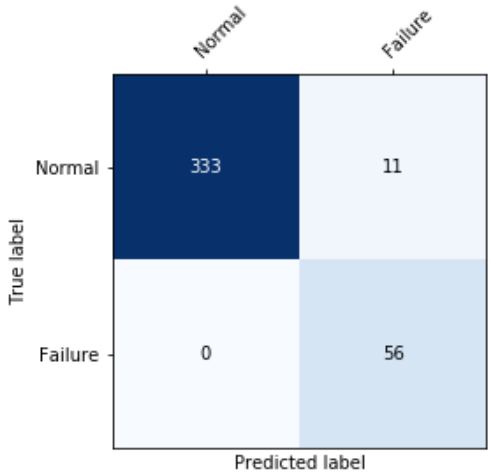
(c) c-Means



(d) DBN c-Means



(e) SOM



(f) DBN SOM

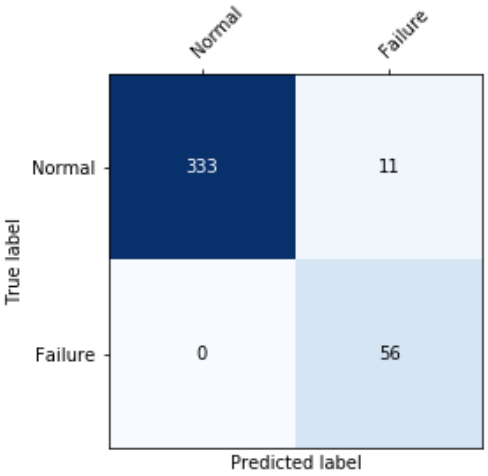
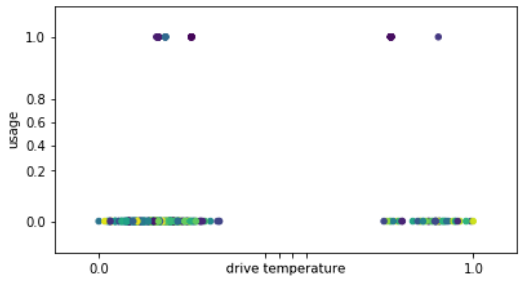
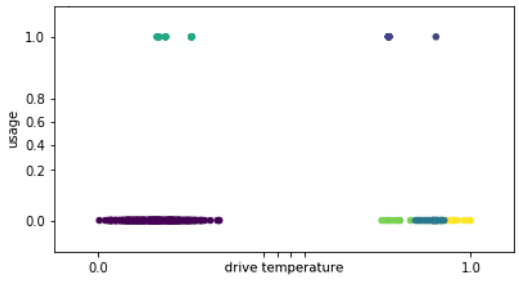


Figure 5.2: Plots of the data points in data set 1a. The data points are colored according to the cluster to which they belong.

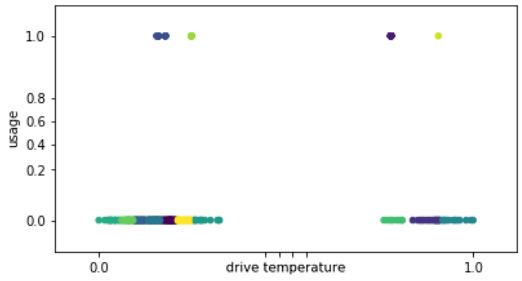
(a) k-Means



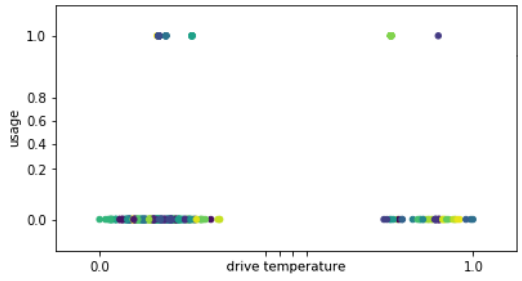
(b) DBN k-Means



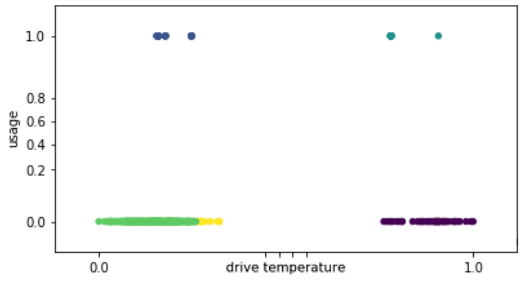
(c) c-Means



(d) DBN c-Means



(e) SOM



(f) DBN SOM

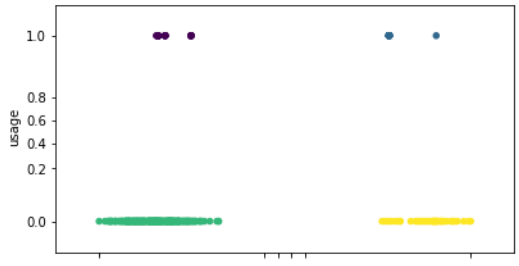
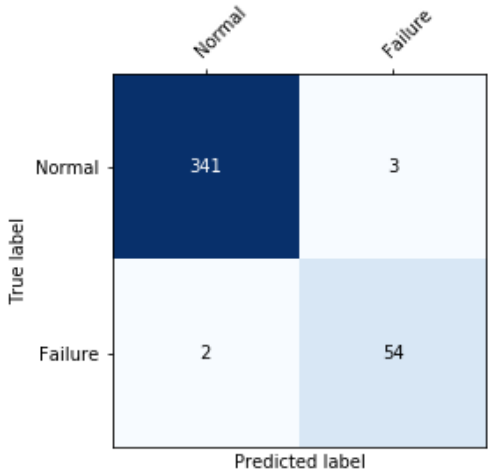
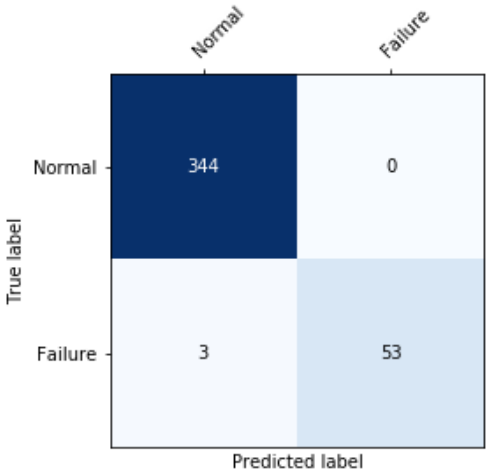


Figure 5.3: Confusion matrices for the different algorithms on synthetic data set 1b

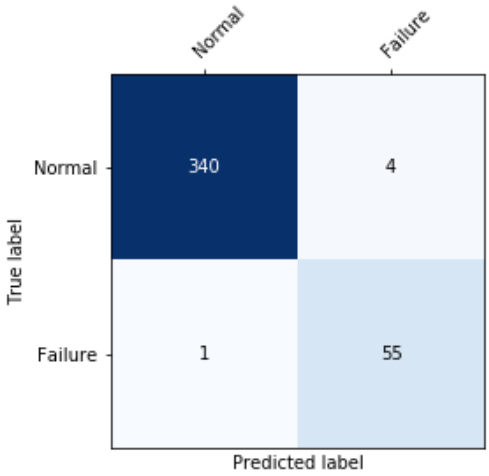
(a) k-Means



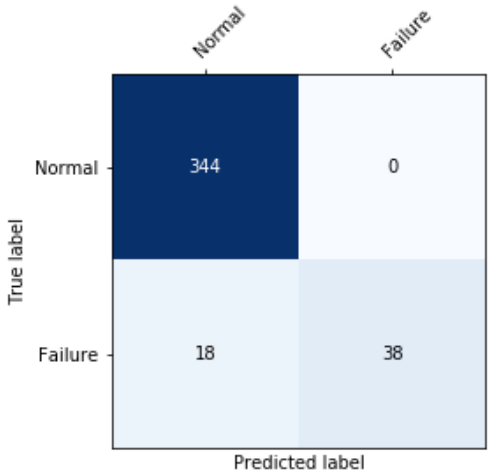
(b) DBN k-Means



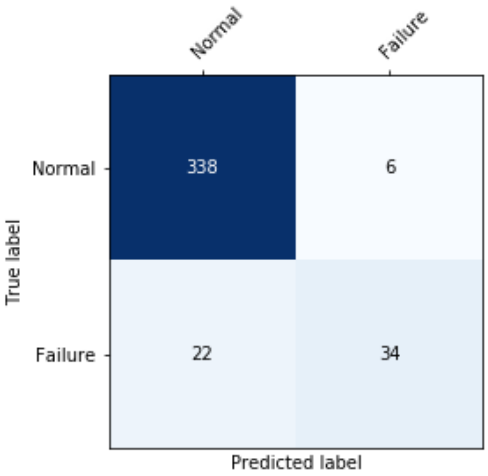
(c) c-Means



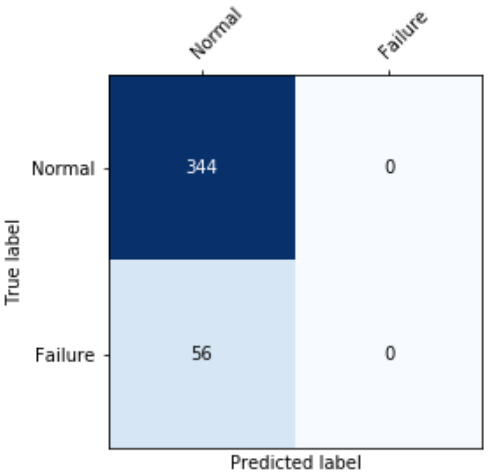
(d) DBN c-Means



(e) SOM



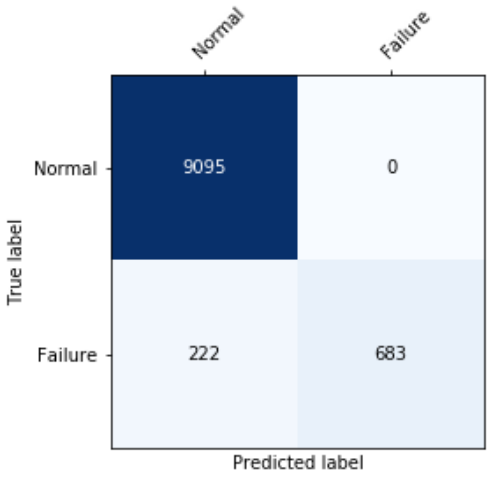
(f) DBN SOM



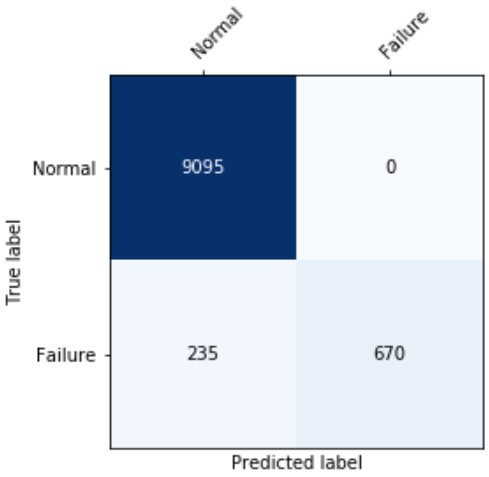
albeit slightly less severe, is visible when SOM is used stand-alone. This decrease is substantial and may be explained by a number of reasons. The first one would be that, due to the larger number of samples, the weights are not updated often enough. Since the weights are initialized randomly it could happen that a unit is placed in the middle of two clusters and serves as the BMU for both. If this happens it means that the weights of the unit are updated by samples from both clusters, meaning it will most likely also stay the BMU for both. This random initialization may also have a negative influence on the performance. That is to say that even though a certain configuration performs well during the hyper-parameter estimation, this doesn't automatically mean that it will perform well the next time. Both of these problems may also occur when the SOM is used together with the DBN. k-Means and c-Means both showed a slight decrease in performance, with a minimal difference between the two methods. DBN c-Means however saw a more substantial performance decrease. This might again be explained by the same problem that was mentioned before, namely that the cluster centers have the tendency to all stay in the middle, which might be further reinforced when there are more samples involved and fewer dimensions to distinguish between them.

Figure 5.4: Confusion matrices for the different algorithms on synthetic data set 1c

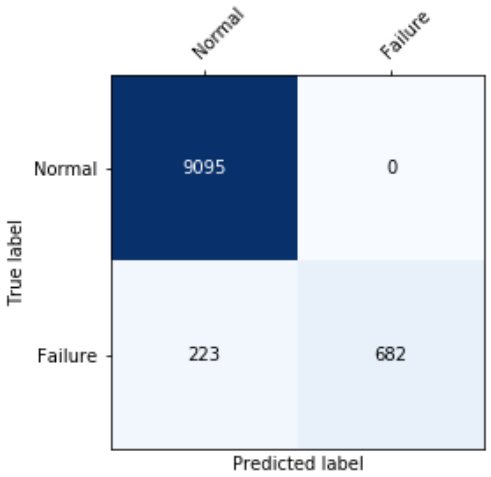
(a) k-Means



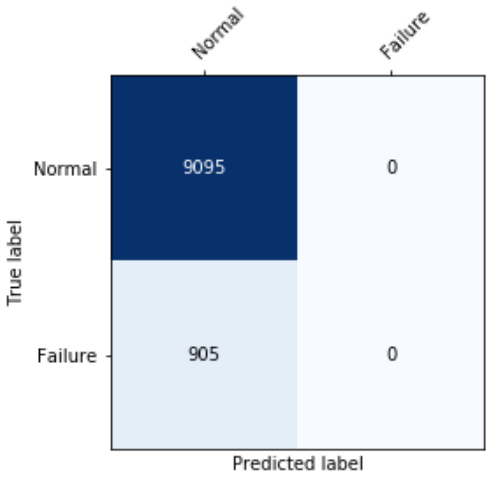
(b) DBN k-Means



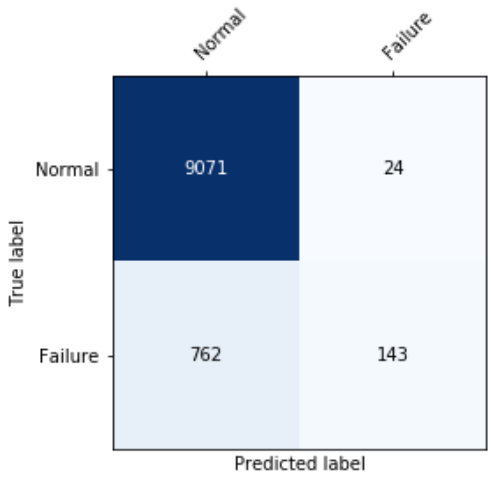
(c) c-Means



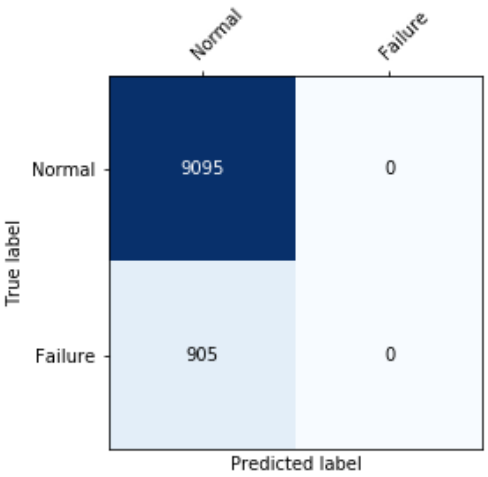
(d) DBN c-Means



(e) SOM



(f) DBN SOM



5.2 Real data set

Even though the results on the synthetic data set are very interesting, it is even more interesting how these algorithms perform when they are tested on the real data set. On this data set, all algorithms performed considerably worse than they did on the synthetic data set, which is likely a result of the faults being less pronounced and the fact that the dimensionality is higher and that there are more types of sensors. When looking at the Silhouette index the results, which can be found in Table 5.2 alongside those of the other data sets, they seem very promising and are generally higher than those on the synthetic data sets, which implies that there is a good separation between the clusters as well as a good cluster density.
This is however a pattern that does not continue when looking at the F-Score. On the synthetic data sets the Silhouette index usually under performed compared to the F-Score, but on the real data set the two have swapped places. In all cases except for the DBN k-Means combination and SOM, the F-Score is quite low with 0.2460. This can be explained when looking at the confusion matrices in Figure 5.6. Here it becomes apparent that all clusters are labelled ”Normal”. The most likely cause of this is that the clusters are too large and will therefore always include a majority of points labelled as ”Normal”.
This problem is to a lesser extend also visible in the case of DBN k-Means. In this case the hyper-parameter estimation came up with a higher number of clusters, namely 29, which led to part of the samples being labelled as fault ”M10”. The two less frequent faults, ”M9” and ”M11” were however still not discovered. However when dimensionality reduction was applied before running the c-Means clustering algorithm the number of clusters did increase, but the F-Score did not and all samples were still labelled as ”Normal”. SOM also performed better than the other methods. The net was relatively large at a width of 4096 and a height of 128 units. This means that the total size is 524.288 units and thereby 524.288 potential clusters.
The problems that DBN k-Means and SOM, as well as the other methods, encountered could in part be due to the fact that the faults are not very pronounced. The distance between a sample which is labelled as ”M11” and the closest other sample with the ”M11” label is on average 1.1012. The distance between a ”M11” sample and the closest sample with the label ”Normal” is on average 1.0830. Therefore it may be concluded that it would be very difficult to achieve a good result here. Possible solutions for this problem are discussed in Chapter 6.
When looking at the results of the CB-SSSOM a couple of thing need to be kept in mind. Since this method is semi-supervised, in contrast to the other methods that were discussed, it will need more information. To provide this the labels provided by the domain expert were used. However since this is a constraint based and not a

label based method, those labels are first converted to constraints. The algorithm is tested in different setups, each with a different amount of labelled samples. The settings as well as the results can be found at the bottom of Table 5.2. Fig. 5.7 shows the confusion matrix for each of the settings. If the algorithm got one percent labelled samples this means that from each class, one percent of the labels was provided, with a minimum of two labelled samples per class.
From the table it is clear that with one percent of the labels available the best F-score is achieved. The score then goes down, but increases again once a quarter of all labels are included. A possible explanation for this phenomenon might be that the algorithm is already able to extract the necessary structure from the one percent, but that, due to having just one BMU for a whole Must-Link group the algorithm has trouble to correctly update these weights, while this problem disappears again once the groups become large enough to enforce a certain structure.
Overall the results of the CB-SSSOM were substantially better than those of the unsupervised algorithms. This was especially visible in the case of fault ”M9”, which was not found by any of the unsupervised algorithms. The F-scores were also better than those of the unsupervised algorithms.
Due to time constraints it was impossible to re-estimate the hyper-parameters, therefore the hyper-parameters that were found for the normal SOM in Chapter 5 were used.

Algorithm	**Metric**	*Synth. 1a*	*Synth. 1b*	*Synth. 1c*	*Real*
k-Means	Silhouette	0.6254	0.5309	0.3965	0.5879
DBN k-Means	Silhouette	0.9245	0.5376	0.3262	0.5061
c-Means	Silhouette	0.5973	0.4656	0.3596	0.2540
DBN c-Means	Silhouette	0.9221	0.4364	0.2316	0.5501
SOM	Silhouette	0.5626	0.2571	0.2610	0.4683
DBN SOM	Silhouette	0.9667	0.5230	0.5198	0.4999

Table 5.1: An overview of all Silhouette scores

Algorithm	**Metric**	*Synth. 1a*	*Synth. 1b*	*Synth. 1c*	*Real*
k-Means	F-score	0.9472	0.9792	0.9241	0.2460
DBN k-Means	F-score	0.9472	0.9795	0.9190	0.4114
c-Means	F-score	0.9472	0.8403	0.9237	0.2460
DBN c-Means	F-score	0.9472	0.8885	0.4763	0.2460
SOM	F-score	0.9472	0.7916	0.6126	0.4020
DBN SOM	F-score	0.9472	0.4624	0.4763	0.2483
CB-SSSOM (1%)	F-score	-	-	-	0.5286
CB-SSSOM (5%)	F-score	-	-	-	0.4383
CB-SSSOM (10%)	F-score	-	-	-	0.3633
CB-SSSOM (25%)	F-score	-	-	-	0.5266

Table 5.2: An overview of all F-scores

Figure 5.5: Visual overview of the F-scores on the real data set

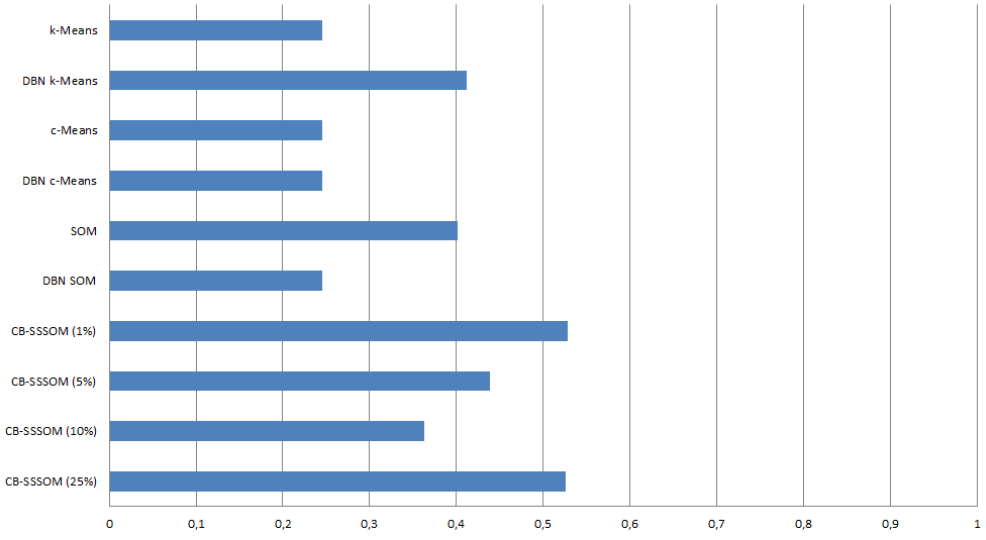
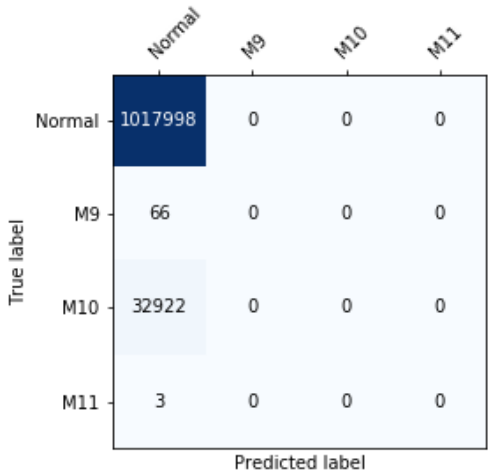
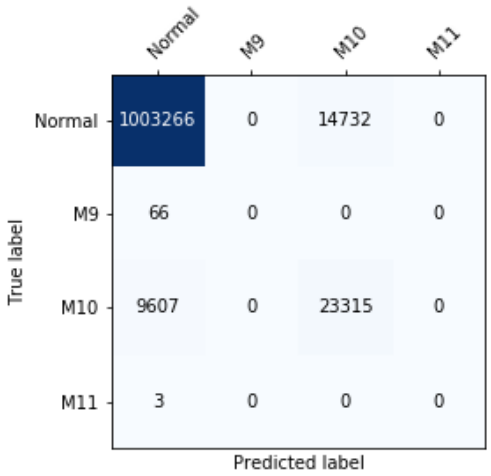


Figure 5.6: Confusion matrices for the different algorithms on the real data set. Many methods classify all data points as ”Normal”, which is why three of the confusion matrices are the same.

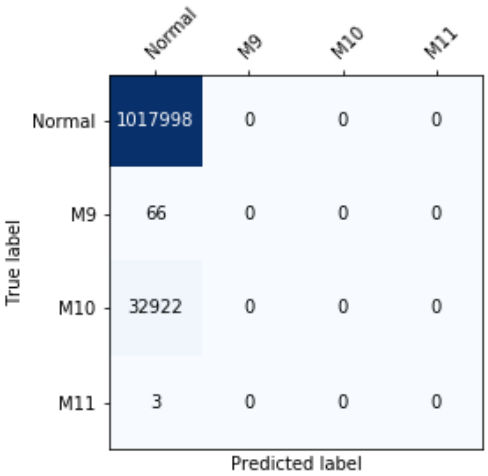
(a) k-Means



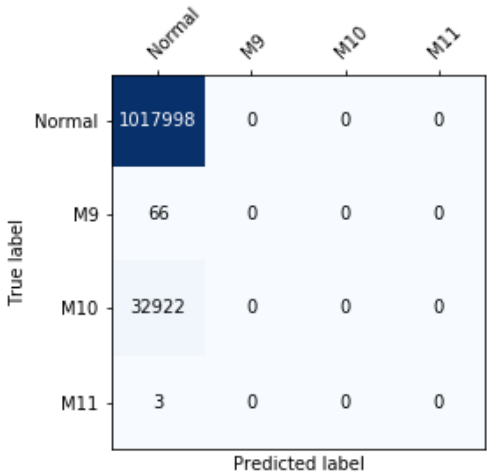
(b) DBN k-Means



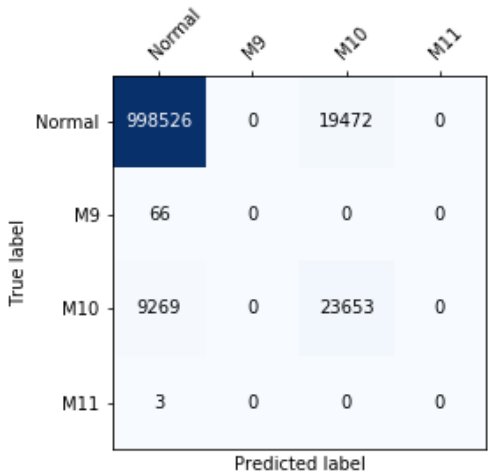
(c) c-Means



(d) DBN c-Means



(e) SOM



(f) DBN SOM

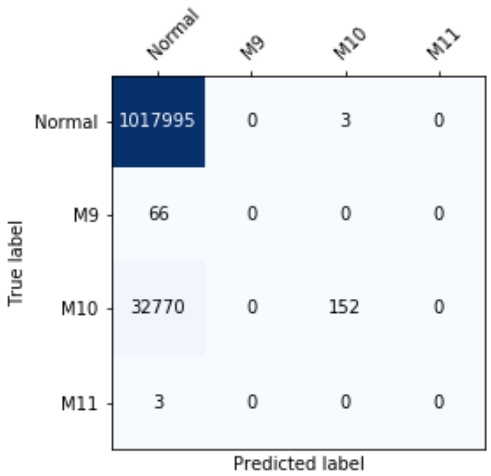
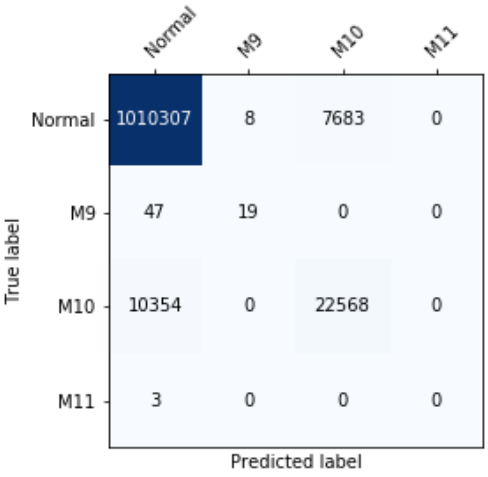
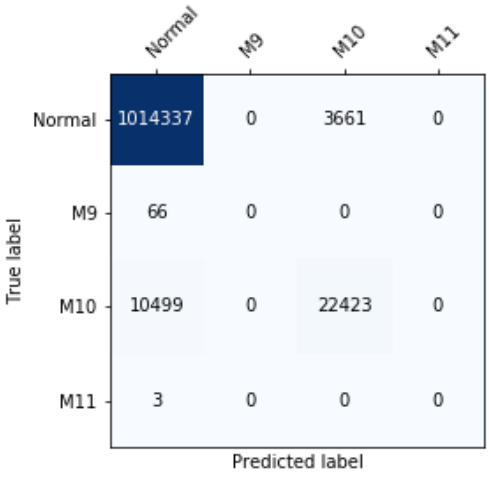


Figure 5.7: Confusion matrices for the CB-SSSOM for the different included percentages.

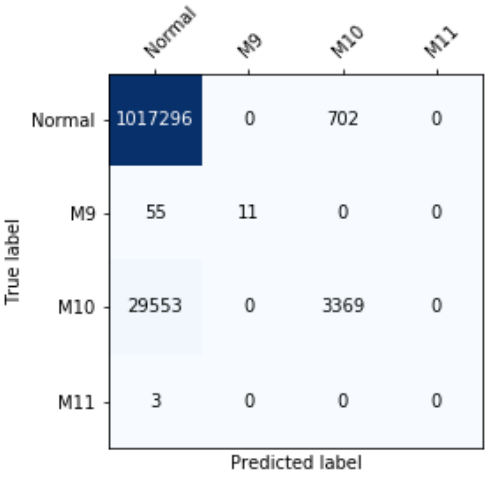
(a) 1 percent



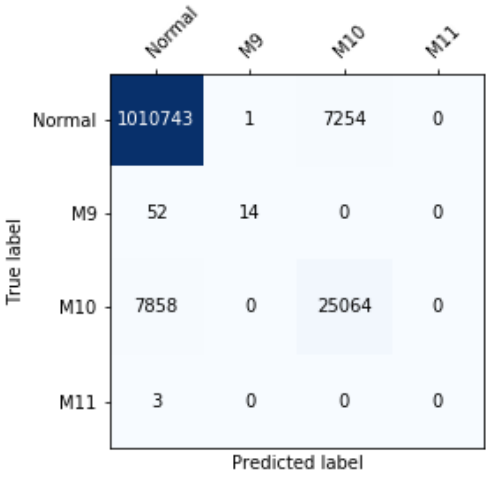
(b) 5 percent



(c) 10 percent



(d) 25 percent



Algorithm	**Parameter**	*Synth. 1a*	*Synth. 1b*	*Synth. 1c*	*Real*
k-Means	k	27	6	29	7
DBN k-Means	k	42	6	48	29
	Epochs	45	137	60	110
	Hidden layers	[103, 103, 103]	[39, 39, 39, 39]	[103, 103, 103]	[102, 102, 102]
	Batch size	29	23	32	54
	Learning rate	0.0766	0.3521	0.0927	0.0394
c-Means	c	35	6	21	5
DBN c-Means	c	29	6	29	26
	Epochs	215	202	187	53
	Hidden layers	[172, 172]	[52, 52, 52]	[172, 172]	[145, 145, 145, 145]
	Batch size	38	110	75	185
	Learning rate	0.0766	0.0058	0.0004	0.0179
SOM	Width	1	1	6	4096
	Height	339	728	8071	128
	Epochs	190	75	145	68
	Learning rate	0.000006	0.000097	0.001051	0.0222
DBN SOM	Width	10	10	10	2048
	Height	41	75	506	8
	Epochs	72	141	156	202
	Learning rate	0.00033	0.00065	0.00033	0.0013
	Epochs (DBN)	214	238	229	186
	Hidden layers	[21, 21, 21, 21]	[173, 173, 173, 173]	[38, 38, 38, 38]	[171, 171, 171]
	Batch size	5	191	23	66
	Learning rate (DBN)	0.0058	0.000002	0.0044	0.0017
CB-SSSOM [1]	Width	-	-	-	4096
	Height	-	-	-	128
	Epochs	-	-	-	68
	Learning rate	-	-	-	0.0222

Table 5.3: Results of the hyper-parameter search for each of data sets

[1] Due to time constraints the hyper-paramters have not be re-estimated, but instead the values found for the SOM were used

5.3 Runtime

The previous two sections described the results in terms of the Silhouette index and the F-Score. However when building a product, the runtime is also a very important factor to consider. To compare the costs of the algorithms in terms of time, the runtimes of the algorithms are compared on synthetic data set 1b. Another factor that has to be kept in mind here is that when the hyper-parameters need to be estimated periodically, then the number of possible combinations and the number of tries needed to converge to a solution are also an important factor. Both the runtimes and these combinations are summed up in Table 5.4.
In terms of combinations, the methods that employ dimensionality reduction have by far the most possible combinations, with the DBN SOM combination leading the pack, with almost 1.5 million possible combinations. In the case of k-Means and c-Means only the number of clusters needs to be determined, which in this case means that there were only 94 possible combinations and that the search was done after 45 and 47 tries respectively. In those cases the genetic algorithm did not perform a very important job, even though the number of tried possible combination was cut in half, compared to a full grid search. However the real benefit here is in the more complex cases, such as the DBN based methods and, to a lesser extend, SOM. In the case of DBN SOM only 284 tries were required before a satisfactory answer was found in the search space of almost 1.5 million options. A similar pattern could be seen at the DBN k-Means, DBN c-Means and SOM.
When looking at the runtimes of the algorithms the DBN-SOM combination takes up most of the time with a wide margin. In that case the time is measured in minutes instead of in seconds as it is with the other methods. One iteration of DBN-SOM takes on average 1.2 minutes on this data set. The simplest of the methods, k-Means and c-Means both take up about 0.5 seconds on average. The time that it takes for DBN-SOM to come up with a solution is in line with the other results, since the other methods also take up 20 to 30 seconds longer per try when a DBN is used. This is also in line with the general complexity of the methods. k-Means and c-Means are often considered as some of the simplest and most straightforward methods to perform clustering and, although they are computationally complex, there has been a lot of research over the last decades into performance enhancing heuristics. SOM on the other hand is a more complex algorithm, as is DBN. Therefore it is evident that training the model also takes up more time.

Algorithm	*Possible combinations*	*Tried combinations*	*Average time per try*
k-Means	94	45	0.5s
DBN k-Means	1.159.526.792	282	17.2s
c-Means	94	47	0.5s
DBN c-Means	1.159.526.792	224	33.5s
SOM	355.612.500	186	23.5s
DBN SOM	1.498.469.552	284	1.2m

Table 5.4: The average search time and number of tried combination for synthetic data set 1b.

	Synthetic 1a	*Synthetic 1b*	*Synthetic 1c*	*Real*
Dimensionality	2	37	37	63
Continuous variables	1	35	35	32
Discrete variables	1	2	2	31
# of samples	400	400	10.000	1.051.871

Table 5.5: Technical characteristics of the data sets

Chapter 6

Discussion

The presented results in Chapter 5 have made it clear that the presented techniques are not yet ready to be used in a production system. The achieved results on the real data set were disappointing. The aim of this chapter is to present some implications of the results, followed by a number of issues with the proposed solutions, including some avenues of future research.

6.1 Implications

When discussing the results it is also important to look at the implication of the results on a real-life scenario. In other words, what would happen when the system was put in use right now. In order to answer this question only the best case scenario, in this case the CB-SSSOM with one percent of the labels available, will be considered. When determining what the real world implications would be there are two primary points to consider, what are the benefits and what are the disadvantages? When looking at the confusion matrix in Fig. 5.7a it becomes clear that in 1.032.894 of the 1.050.989 cases the system made a correct classification, in other words in 98.28% of the cases the classification given was right. When looking at the disadvantages there are two main problems. On the one hand there are false alarms, where a ”Normal” sample is classified as a fault and there are false negatives, when faults are classified as ”Normal”. The former happened in 10.404 or 0.99% of the cases and the latter in 7.691 or 0.73% of the cases. This might not seem like a lot, however when looking at the absolute numbers it means that for almost 29 hours in the five month period there would be false alarms and for more than 21 hours there would be false negatives. In other words, this would not be a viable solution yet when looking at the practical implications.

6.2 Issues

When discussing the issues the elephant in the room ought to be the results themselves. The results were not good enough to produce a production ready system. This is likely a result of the high dimensionality. Since most of the faults are only visible in a couple of time series they are easily overshadowed by other time series with more fluctuations, when looking at the complete system. The proposed solution for this issue was to employ dimensionality reduction however it turned out that this did not perform satisfactorily. Therefore other techniques should be investigated. One possible solutions with high potential that were already discussed passingly in Chapter 3 is subspace clustering, which looks at subsets of the dimensions to find clusters. For a more detailed explanation please refer to Section 2.1.6. Given that this strikes at the core of the problem mentioned before this is a very interesting avenue of future research. Another potential solution that was already tested and deserves further investigation is semi-supervised clustering. The previous chapter showed that even with only 1% of the labels it already performed substantially better than the unsupervised methods. This might even be further improved when the labels are not picked at random as was done in this case, but are picked to be the most informative. This could be achieved by employing Active Learning where the operator is asked to provide labels for points that are considered difficult to cluster. Both subspace clustering and semi-supervised clustering are not methods themselves but rather groups of techniques. However, even though both of these have a high potential of solving the problem mentioned earlier it is also possible that that these do not pan out and it should be concluded that clustering is not a valid solution direction for the case of Thales, however this should not be done unless all potential solutions have been exhausted.

In case a solution is found that improves the results there are still some other issues that should be dealt with. One of those issues is the explainability of a classification. When a new sample is classified, the engineer needs to be able to understand why the classification is given, especially when he does not agree with it. Being able to explain a decision would help in building trust in the system and to further improve it. One way to make the results more explainable is to display the other samples in the cluster, to the users. This might however not be fully satisfying, which makes this issue important to solve before putting the system in production.

Another potential issue is the hyper-parameter search. In some cases, such as in k-Means clustering, the number of clusters needs to be determined beforehand. However as new data comes in, new clusters arise. For example when a previously unseen problem is encountered. This means that the hyper-parameters need to be retrained periodically, which in turn means that it should be feasible to do a hyper-

parameter search on-board. There is however a difference between the different methods. For example, some methods will need a search every time a new cluster arises, but other methods only need to be retrained when a previously unseen time series is included. There are a couple of methods to mitigate this issue, for example limiting the search space around the previously found values or only trying a couple of options, in the expected direction. For example in the case of k-Means it can be assumed that over time the number of clusters will grow, therefore when the hyper-parameters are re-estimated on a weekly basis, it will most likely suffice to try the the values in the range $[k, k + 10]$, with k being the previous number of clusters. However this is only an estimation and such a tactic will require further investigation.
The third problem is one that arises when an engineer does not agree with the given classification. This is a situation that is probably going to occur, because no system is impeccable, and that will have to be dealt with satisfactorily, otherwise the engineers will lose faith in the system. Therefore some way has to be found to incorporate user feedback into the classification, a feature which the given solutions lacks. Some interesting avenues here are Active Learning and Reinforcement Learning and the advice is to further investigate this issue and these possible solutions.
When looking at the methodology there are a couple of open issues that remain. The first of those is the evaluation. In order to compare the different methods a straightforward evaluation strategy was set up based on the Silhouette score and the F-Score. However this required that each of the methods outputted exactly one cluster label for each data point. In other words, all other information, such as the membership probability of the fuzzy c-Means clustering and the topological information which is available in the SOM was discarded. That information might be very useful, however this is currently not investigated. In the future an investigation into how this kind of information can be integrated into the solution would be beneficial.
Another issue with this approach is that, even though soft clustering is used, each data point is still assigned to only one cluster. If the second cluster was also taken into account, the score of c-Means might have been substantially higher.
Something else that could be seen as a drawback of the used evaluation technique is that the number of clusters are not considered. Due to the label propagation technique, the F-Score is generally higher when the clusters are smaller. However having more distinct clusters means asking more questions to the user to find out which label belongs to a certain cluster. This might be prevented by training only on using the Silhouette score as a metric, but the fact that this is not considered explicitly could be seen as a drawback.
When it comes to the hyper-parameter estimation there is also an issue that needs to be considered, namely the problem of local optima. As could be seen in the pre-

vious chapter, the GA only tested a fraction of the possible solutions for the more advanced methods. This means that it is highly likely that the algorithm finished at a local optimum and that the score could be further improved by testing more options. This is something that needs to be kept in mind and that might need to be further investigated.

Another integral part of the methodology is the use of synthetic data. This is not without its dangers and limitations. First and foremost it is difficult to generate high quality data that follows the same distribution as the real data. Thus, if this can not be guaranteed, it can also not be guaranteed that the results on the synthetic data are similar to those on the real data. In order to mitigate this problem, both synthetic and real data sets have been used.

Finally, one of the issues is that over time a lot of data is being collected. Radar systems will typically run for extended periods of time, decades even. During this period data is collected, the question however is, which data should be used to train on? It might not be possible, nor desirable, to train on all the data for a number of reasons. The main one being time, if the model is retrained periodically it will take a long time if there is 20 years worth of data. However, if the system is only trained on the last months, it will not recognise issues that occurred before this period. Something that further complicates this situation is that parts of the radar may be replaced by other types with a similar function during its lifetime, because the original part is no longer manufactured. This has an effect on the (types of) data that is generated by this part. Therefore the selection of training data is a delicate process that needs to be executed carefully and that requires some research to be done beforehand. A possible solution is to only consider the last year in its totality and to only include the abnormal situations during the period before that. However there might be other, more advanced, forgetting strategies available in the literature. This will require some additional research.

Chapter 7

Conclusion

Throughout this report the goal has been to answer the three research questions that were asked in Section 1.3. RQ1.a was answered through an extensive literature review and resulted in a list of algorithms that had in the past been used to solve similar problems. This list was then narrowed down to three methods, k-Means clustering, fuzzy c-Means clustering and SOMs.
The second research question, RQ1.b was also answered through a literature review. Based on this literature review two metrics were chosen to evaluate the quality of the resulting diagnoses, the Silhouette score and the F-Score. The former does not require labels, but the latter does.
The three aforementioned algorithms have been tested on three synthetic and one real data set. The results for each method differed between each of the data sets and on the synthetic data sets there was no clear winner, although in all three cases SOM performed worst, both in terms of Silhouette score and F-Score.
On the real data set none of the methods was able to correctly cluster all the faults. The most likely reason for this is that, due to the high number of dimensions and most faults only being visible in a couple of dimensions, most faults were overlooked.
When answering RQ1.c the selected methods should be compared between themselves. As was mentioned before, no clear winner can be selected when looking at all the data sets at the same time, which is likely a result of the different characteristics of the data sets and the algorithms. In other words, one algorithm might deal better with a lot of data, whereas another algorithm is better capable of handling a large number of dimensions. Therefore when formulating an answer to RQ1.c only the real data set is considered, since it is a better reflection of the problem that was posed in the introduction than the synthetic data sets. In the end, when considering the F-Score and the results as they were displayed in the confusion matrices the CB-SSSOM was the best performing method with a substantial margin. This is also to be expected given that the method has more information to work with. When looking only at the unsupervised methods DBN k-Means was the best performing algorithm,

closely followed by the SOM. However the score was still not satisfactory. Therefore it may be concluded that, although this is certainly an interesting avenue for further work, the current solution is not yet ready to solve the problem. In the discussion a number of points of improvement with regards to the methodology were pointed out and some avenues for future work were presented.

Thus it may be concluded that the major contribution of this work is that an extensive literature review was performed and that a number of possible solutions were implemented and tested on both synthetic and real data. In addition to this a constraint-based semi-supervised variant of the SOM was developed. A methodology was also developed, which may be used in the future by Thales to test other methods while still being able to compare them directly to results obtained earlier. Finally a number of suggestions were done for future work that could be used by Thales to shape the development of a production ready system.

Bibliography

[1] S. Z. Selim and M. A. Ismail, “K-means-type algorithms: A generalized convergence theorem and characterization of local optimality,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 1, pp. 81–87, 1984.

[2] D. Arthur and S. Vassilvitskii, “K-means++: the advantages of careful seeding,” in *In Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007.

[3] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.

[4] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973. [Online]. Available: https://doi.org/10.1080/01969727308546046

[5] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, Jan 1982. [Online]. Available: https://doi.org/10.1007/BF00337288

[6] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[7] J. Martnez-Morales, E. Palacios-Hernndez, and D. Campos-Delgado, “Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions,” *Electrical Engineering*, vol. 100, no. 1, pp. 59–73, 2018, cited By 3.

[8] M.-Y. Cho and H. Thom, “Fault diagnosis for distribution networks using enhanced support vector machine classifier with classical multidimensional scaling,” *Journal of Electrical Systems*, vol. 13, no. 3, pp. 415–428, 2017, cited By 0.

[9] K. Rama Krishna and K. Ramachandran, “Machinery bearing fault diagnosis using variational mode decomposition and support vector machine as a classifier,” vol. 310, no. 1, 2018, cited By 1.

[10] B.-S. Yang, T. Han, and W.-W. Hwang, "Fault diagnosis of rotating machinery based on multi-class support vector machines," *Journal of Mechanical Science and Technology*, vol. 19, no. 3, pp. 846–859, Mar 2005. [Online]. Available: https://doi.org/10.1007/BF02916133

[11] R. Ziani, A. Felkaoui, and R. Zegadi, "Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fishers criterion," *Journal of Intelligent Manufacturing*, vol. 28, no. 2, pp. 405–417, 2017, cited By 13.

[12] Q. Miao, X. Zhang, Z. Liu, and H. Zhang, "Condition multi-classification and evaluation of system degradation process using an improved support vector machine," *Microelectronics Reliability*, vol. 75, pp. 223–232, 2017, cited By 7.

[13] V. Sindhwani and S. S. Keerthi, "Large scale semi-supervised linear svms," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '06. New York, NY, USA: ACM, 2006, pp. 477–484. [Online]. Available: http://doi.acm.org/10.1145/1148170.1148253

[14] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: A review," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 90–105, Jun. 2004. [Online]. Available: http://doi.acm.org.ezproxy2.utwente.nl/10.1145/1007730.1007731

[15] O. Chapelle, B. Schölkopf, A. Zien *et al.*, "Semi-supervised learning, vol. 2," *Cambridge: MIT Press. Cortes, C., & Mohri, M.(2014). Domain adaptation and sample bias correction theory and algorithm for regression. Theoretical Computer Science*, vol. 519, p. 103126, 2006.

[16] C. J. Merz, D. C. St. Clair, and W. E. Bond, "Semi-supervised adaptive resonance theory (smart2)," in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, vol. 3, June 1992, pp. 851–856 vol.3.

[17] I. Givoni and B. Frey, "Semi-supervised affinity propagation with instance-level constraints," in *Artificial Intelligence and Statistics*, 2009, pp. 161–168.

[18] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, 1st ed. Berlin, Germany: Springer, Cham, 9 2018.

[19] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods an introduction," *Control Engineering Practice*, vol. 5, no. 5, pp. 639 – 652, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967066197000464

[20] K. Li, N. Yu, P. Li, S. Song, Y. Wu, Y. Li, and M. Liu, "Multi-label spacecraft electrical signal classification method based on dbn and random forest," *PLoS ONE*, vol. 12, no. 5, 2017, cited By 1.

[21] X. Wu, D. Wang, W. Cao, and M. Ding, "A GA-SVM and D-S evidence theory based method for shunt-fault diagnosis in power systems," *2017 Asian Control Conference, ASCC 2017*, vol. 2018-January, no. 2, pp. 2704–2709, 2018.

[22] X. Zhao and M. Jia, "A novel deep fuzzy clustering neural network model and its application in rolling bearing fault recognition," *Measurement Science and Technology*, vol. 29, no. 12, p. 125005, oct 2018. [Online]. Available: https://doi.org/10.1088%2F1361-6501%2Faae27a

[23] J. Hou and A. Zhang, "Enhanced Dominant Sets Clustering by Cluster Expansion," *IEEE Access*, vol. 6, pp. 8916–8924, 2018.

[24] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Prez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243 – 256, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031200338X

[25] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Cluster validity methods: Part i," *SIGMOD Rec.*, vol. 31, no. 2, pp. 40–45, Jun. 2002. [Online]. Available: http://doi.acm.org.ezproxy2.utwente.nl/10.1145/565117.565124

[26] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53 – 65, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377042787901257

[27] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, Feb. 1979. [Online]. Available: http://dx.doi.org.ezproxy2.utwente.nl/10.1109/TPAMI.1979.4766909

[28] T. Caliski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101

[29] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427 – 437, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457309000259

[30] J. Luo, H. Xu, Z. Su, H. Xiao, K. Zheng, and Y. Zhang, "Fault diagnosis based on orthogonal semi-supervised lltsa for feature extraction and transductive svm for fault identification," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3499–3511, 2018, cited By 1.

[31] T. Abdelgayed, W. Morsi, and T. Sidhu, "Fault detection and classification based on co-training of semisupervised machine learning," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1595–1605, 2017, cited By 6.

[32] F. Tassine and B. Ismail, "Hybrid classifier for fault detection and isolation in wind turbine based on data-driven," 2017, cited By 0.

[33] W.-H. Zhu, J.-Y. Huang, S.-X. Feng, J.-J. Wei, and H.-X. Chen, "Application of feature fusion based on dhmm method and bp neural network algorithm in fault diagnosis of gearbox," 2017, cited By 0.

[34] H. Pan, X. He, S. Tang, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional cnn and lstm," *Strojniski Vestnik/Journal of Mechanical Engineering*, vol. 64, no. 7-8, pp. 443–452, 2018, cited By 0.

[35] Q. Niu, "Discussion on fault diagnosis of and solution seeking for rolling bearing based on deep learning," *Academic Journal of Manufacturing Engineering*, vol. 16, no. 1, pp. 58–64, 2018, cited By 0.

[36] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden markov models for rolling element bearing fault identification," *Knowledge-Based Systems*, vol. 144, pp. 65–76, 2018, cited By 7.

[37] M. J. Hasan, M. M. Islam, and J.-M. Kim, "Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions," *Measurement*, vol. 138, pp. 620 – 631, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263224119301939

[38] K. Wang and B. Wu, "Power equipment fault diagnosis model based on deep transfer learning with balanced distribution adaptation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11323 LNAI, pp. 178–188, 2018, cited By 0.

[39] J. Tao, Y. Liu, D. Yang, and G. Bin, "Rolling bearing fault diagnosis based on bacterial foraging algorithm and deep belief network," *Zhendong yu Chongji/Journal of Vibration and Shock*, vol. 36, no. 23, pp. 68–74, 2017, cited By 0.

[40] Z. Zhang and J. Zhao, "A deep belief network based fault diagnosis model for complex chemical processes," *Computers and Chemical Engineering*, vol. 107, pp. 395–407, 2017, cited By 23.

[41] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. Gao, "A deep learning approach for fault diagnosis of induction motors in manufacturing," *Chinese Journal of Mechanical Engineering (English Edition)*, vol. 30, no. 6, pp. 1347–1356, 2017, cited By 7.

[42] F. Xu, Y. Fang, D. Wang, J. Liang, and K. Tsui, "Combining dbn and fcm for fault diagnosis of roller element bearings without using data labels," *Shock and Vibration*, vol. 2018, 2018, cited By 0.

[43] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1096–1103. [Online]. Available: http://doi.acm.org/10.1145/1390156.1390294

[44] D.-T. Hoang and H.-J. Kang, "A bearing fault diagnosis method based on autoencoder and particle swarm optimization support vector machine," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10954 LNCS, pp. 298–308, 2018, cited By 0.

[45] B. Xin, T. Wang, and T. Tang, "A deep learning and softmax regression fault diagnosis method for multi-level converter," vol. 2017-January, 2017, pp. 292–297, cited By 0.

[46] Q. Xing-Yu, Z. Peng, F. Dong-Dong, and X. Chengcheng, "Autoencoder-based fault diagnosis for grinding system," 2017, pp. 3867–3872, cited By 0.

[47] M. Xia, T. Li, L. Liu, L. Xu, and C. de Silva, "Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder," *IET Science, Measurement and Technology*, vol. 11, no. 6, pp. 687–695, 2017, cited By 13.

[48] M. Blanco, K. Gibert, P. Marti-Puig, J. Cusid, and J. Sol-Casals, "Identifying health status of wind turbines by using self organizing maps and interpretation-oriented post-processing tools," *Energies*, vol. 11, no. 4, 2018, cited By 3.

[49] M. Khanlari and M. Ehsanian, “An improved kfcm clustering method used for multiple fault diagnosis of analog circuits,” *Circuits, Systems, and Signal Processing*, vol. 36, no. 9, pp. 3491–3513, 2017, cited By 2.

[50] J. Liu, Q. Li, W. Chen, and T. Cao, “A discrete hidden markov model fault diagnosis strategy based on k-means clustering dedicated to pem fuel cell systems of tramways,” *International Journal of Hydrogen Energy*, vol. 43, no. 27, pp. 12 428 – 12 441, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360319918313521

[51] H. Rostami, J. Blue, and C. Yugma, “Equipment condition diagnosis and fault fingerprint extraction in semiconductor manufacturing,” 2017, pp. 534–539, cited By 0.

[52] A. Rodrguez Ramos, O. Llanes-Santiago, J. Bernal de Lzaro, C. Cruz Corona, A. Silva Neto, and J. Verdegay Galdeano, “A novel fault diagnosis scheme applying fuzzy clustering algorithms,” *Applied Soft Computing Journal*, vol. 58, pp. 605–619, 2017, cited By 1.

[53] R. Zeng, S. Zhang, R. Zeng, H. Shen, and L. Zhang, “A method of fault detection on diesel engine based on emd-fractal dimension and fuzzy c-mean clustering algorithm,” 2017, pp. 7679–7683, cited By 0.

[54] F. Xu, Y. J. Fang, Z. Wu, and J. Q. Liang, “A method based on multiscale base-scale entropy and random forests for roller bearings faults diagnosis,” *Journal of Vibroengineering*, vol. 20, no. 1, pp. 175–188, 2018.

[55] C. Liu and X. Yan, “Fault diagnosis of a wind turbine rolling bearing based on variational mode decomposition and condition identification,” *Dongli Gongcheng Xuebao/Journal of Chinese Society of Power Engineering*, vol. 37, no. 4, pp. 273–278 and 334, 2017, cited By 1.

[56] J. Gao, M. Kang, J. Tian, L. Wu, and M. Pecht, “Unsupervised locality-preserving robust latent low-rank recovery-based subspace clustering for fault diagnosis,” *IEEE Access*, vol. 6, pp. 52 345–52 354, 2018, cited By 0.

[57] M. Zhao, Z. Zhang, T. W. Chow, and B. Li, “Soft label based linear discriminant analysis for image recognition and retrieval,” *Computer Vision and Image Understanding*, vol. 121, pp. 86 – 99, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314214000150

[58] A. Krishnamurthy, M. Belur, and D. Chakraborty, “Comparison of various linear discriminant analysis techniques for fault diagnosis of re-usable launch vehicle,” 2011, pp. 3050–3055, cited By 1.

[59] X. Jin, M. Zhao, T. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014, cited By 170.

[60] K. Yeung, M. Medvedovic, and R. Bumgarner, "Bayesian mixture model based clustering of replicated microarray data," *Bioinformatics*, vol. 20, no. 8, pp. 1222–1232, 02 2004. [Online]. Available: https://dx.doi.org/10.1093/bioinformatics/bth068

[61] F. Liang, "Use of svd-based probit transformation in clustering gene expression profiles," *Computational Statistics & Data Analysis*, vol. 51, no. 12, pp. 6355 – 6366, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016794730700059X

[62] M.-H. Luo and P. Liang, "Turbine faults diagnosis based on gaussian mixture models," *Hedongli Gongcheng/Nuclear Power Engineering*, vol. 30, no. 6, pp. 86–90, 2009, cited By 0.

[63] J. Sun, Y. Li, and C. Wen, "Fault diagnosis and detection based on combination with gaussian mixture models and variable reconstruction," 2009, pp. 227–230, cited By 0.

[64] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07. New York, NY, USA: ACM, 2007, pp. 759–766. [Online]. Available: http://doi.acm.org.ezproxy2.utwente.nl/10.1145/1273496.1273592

[65] N. Abe, B. Zadrozny, and J. Langford, "Outlier detection by active learning," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '06. New York, NY, USA: ACM, 2006, pp. 504–509. [Online]. Available: http://doi.acm.org/10.1145/1150402.1150459

[66] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[67] M. Szummer and T. Jaakkola, "Partially labeled classification with markov random walks," 2002, cited By 185.

[68] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *CoRR*, vol. abs/1206.5533, 2012. [Online]. Available: http://arxiv.org/abs/1206.5533

[69] M. M. Lau and K. H. Lim, "Investigation of activation functions in deep belief network," in *2017 2nd International Conference on Control and Robotics Engineering (ICCRE)*, April 2017, pp. 201–206.

[70] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191 – 203, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0098300484900207

[71] A. Allahyar, H. S. Yazdi, and A. Harati, "Constrained semi-supervised growing self-organizing map," *Neurocomputing*, vol. 147, pp. 456 – 471, 2015, advances in Self-Organizing Maps Subtitle of the special issue: Selected Papers from the Workshop on Self-Organizing Maps 2012 (WSOM 2012). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092523121400811X

[72] rsteca, "sklearn-deap," 2018. [Online]. Available: https://github.com/rsteca/sklearn-deap

[73] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, 2019. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.09.050

[74] Z. Yuan, L. Zhang, L. Duan, and T. Li, "Intelligent Fault Diagnosis of Rolling Element Bearings Based on HHT and CNN," *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 292–296, 2019.

[75] C. Li, L. Ledo, and M. Delgado, "GKFP: A New Fuzzy Clustering Method Applied to Bearings Diagnosis," *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, no. 1, pp. 1295–1300, 2019.

[76] Y. Liu, Z. Jiang, C. Hu, G. Zhao, and Y. Gao, "A New Approach for Analog Circuit Fault Diagnosis Based on Extreme Learning Machine," *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 196–200, 2019.

[77] B. P. Duong and J.-M. Kim, "Fault diagnosis of multiple combined defects in bearings using a stacked denoising autoencoder," in *Advances in Computer Communication and Computational Sciences*, S. K. Bhatia, S. Tiwari, K. K. Mishra, and M. C. Trivedi, Eds. Singapore: Springer Singapore, 2019, pp. 83–93.

[78] X. Xue, C. Li, S. Cao, J. Sun, and L. Liu, “Fault diagnosis of rolling element bearings with a two-step scheme based on permutation entropy and random forests,” *Entropy*, vol. 21, no. 1, 2019.

[79] A. Rodríguez-Ramos, A. J. da Silva Neto, and O. Llanes-Santiago, “An approach to fault diagnosis with online detection of novel faults using fuzzy clustering tools,” *Expert Systems with Applications*, vol. 113, pp. 200–212, 2018.

[80] H. Feng, Y. Wang, J. Zhu, and D. Li, “Multi-Kernel Learning Based Autonomous Fault Diagnosis for Centrifugal Pumps,” *ICCAIS 2018 - 7th International Conference on Control, Automation and Information Sciences*, no. 11671400, pp. 548–553, 2018.

[81] S. Cao, L. Wen, X. Li, and L. Gao, “Application of generative adversarial networks for intelligent fault diagnosis,” vol. 2018-August, 2018, pp. 711–715, cited By 0.

[82] Z. Y. Wang, C. Lu, and B. Zhou, “Fault diagnosis for rotary machinery with selective ensemble neural networks,” *Mechanical Systems and Signal Processing*, vol. 113, pp. 112–130, 2018. [Online]. Available: https://doi.org/10.1016/j.ymssp.2017.03.051

[83] D. Binu and B. S. Karivappa, “Support Vector Neural Network and Principal Component Analysis for Fault Diagnosis of Analog Circuits,” *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018*, no. Icoei, pp. 1152–1157, 2018.

[84] A. Sharma, R. Jigyasu, L. Mathew, and S. Chatterji, “Bearing Fault Diagnosis Using Weighted K-Nearest Neighbor,” *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018*, no. Icoei, pp. 1132–1137, 2018.

[85] Z. Ren, J. Chen, L. Ye, C. Wang, Y. Liu, and W. Zhou, “Application of RBF Neural Network Optimized Based on K-Means Cluster Algorithm in Fault Diagnosis,” *ICEMS 2018 - 2018 21st International Conference on Electrical Machines and Systems*, pp. 2492–2496, 2018.

[86] M. Lv, X. Su, and S. Liu, “Application of SVM Optimized by IPSO in Rolling Bearing Fault Diagnosis,” *MATEC Web of Conferences*, vol. 227, p. 02007, 2018.

[87] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, “Unsupervised fault diagnosis of rolling bearings using a deep neural network based on

generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412–424, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.07.034

[88] X. Zhao and M. Jia, "Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis," *Neurocomputing*, vol. 315, pp. 447–464, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.07.038

[89] G. Cai, C. Yang, Y. Pan, and J. Lv, "EMD and GNN-AdaBoost fault diagnosis for urban rail train rolling bearings," *Discrete & Continuous Dynamical Systems - S*, vol. 0, no. 0, pp. 1471–1487, 2018.

[90] C. Wu, P. Jiang, C. Ding, F. Feng, and T. Chen, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network," *Computers in Industry*, vol. 108, pp. 53–61, 2019. [Online]. Available: https://doi.org/10.1016/j.compind.2018.12.001

[91] D. Zhao, T. Wang, and F. Chu, "Deep convolutional neural network based planet bearing fault classification," *Computers in Industry*, vol. 107, pp. 59–66, 2019. [Online]. Available: https://doi.org/10.1016/j.compind.2019.02.001

[92] Y. Qin, X. Wang, and J. Zou, "The Optimized Deep Belief Networks with Improved Logistic Sigmoid Units and Their Application in Fault Diagnosis for Planetary Gearboxes of Wind Turbines," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3814–3824, 2019.

[93] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019. [Online]. Available: https://doi.org/10.1016/j.renene.2018.10.031

[94] X. J. Luo, K. F. Fong, Y. J. Sun, and M. K. Leung, "Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system," *Energy and Buildings*, vol. 186, pp. 17–36, 2019. [Online]. Available: https://doi.org/10.1016/j.enbuild.2019.01.006

[95] S. De and B. Chakraborty, "Case Based Reasoning (CBR) Methodology for Car Fault Diagnosis System (CFDS) Using Decision Tree and Jaccard Similarity Method," *2018 3rd International Conference for Convergence in Technology, I2CT 2018*, pp. 1–6, 2018.

[96] W. F. Wang, X. H. Qiu, C. S. Chen, B. Lin, and H. M. Zhang, "Application Research on Long Short-Term Memory Network in Fault Diagnosis," *Proceedings - International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 360–365, 2018.

[97] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.05.002

[98] J. Yu, B. Ding, and Y. He, "Rolling bearing fault diagnosis based on mean multigranulation decision-theoretic rough set and non-naive Bayesian classifier," *Journal of Mechanical Science and Technology*, vol. 32, no. 11, pp. 5201–5211, 2018.

[99] Y. Vidal, F. Pozo, and C. Tutivén, "Wind Turbine Multi-Fault Detection and Classification Based on SCADA Data," *Energies*, vol. 11, no. 11, p. 3018, 2018.

[100] Y. K. Gu, X. Q. Zhou, D. P. Yu, and Y. J. Shen, "Fault diagnosis method of rolling bearing using principal component analysis and support vector machine," *Journal of Mechanical Science and Technology*, vol. 32, no. 11, pp. 5079–5088, 2018.

[101] S. Wan, L. Chen, L. Dou, and J. Zhou, "Mechanical Fault Diagnosis of HVCBs Based on Multi-Feature Entropy Fusion and Hybrid Classifier," *Entropy*, vol. 20, no. 11, p. 847, 2018.

[102] B. Wu and X. Li, "Fault diagnosis method based on kernel fuzzy C-means clustering with gravitational search algorithm," *Proceedings of 2018 IEEE 7th Data Driven Control and Learning Systems Conference, DDCLS 2018*, vol. 2, no. 3, pp. 235–239, 2018.

[103] S. A. Taqvi, L. D. Tufa, H. Zabiri, A. S. Maulud, and F. Uddin, "Multiple Fault Diagnosis in Distillation Column Using Multikernel Support Vector Machine," *Industrial and Engineering Chemistry Research*, vol. 57, no. 43, pp. 14 689–14 706, 2018.

[104] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Differential evolution optimization for resilient stacked sparse autoencoder and its applications on bearing fault diagnosis," *Measurement Science and Technology*, vol. 29, no. 12, 2018.

[105] R. R. Kumar, G. Cirrincione, M. Cirrincione, M. Andriollo, and A. Tortella, "Accurate Fault Diagnosis and Classification Scheme Based on Non-Parametric, Statistical-Frequency Features and Neural Networks," *Proceedings - 2018 23rd International Conference on Electrical Machines, ICEM 2018*, pp. 1747–1753, 2018.

[106] E. F. Swana and W. Doorsamy, "Fault Diagnosis on a Wound Rotor Induction Generator Using Probabilistic Intelligence," *Proceedings - 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I and CPS Europe 2018*, pp. 1–5, 2018.

[107] Z. Shang, X. Liao, R. Geng, M. Gao, and X. Liu, "Fault diagnosis method of rolling bearing based on deep belief network," *Journal of Mechanical Science and Technology*, vol. 32, no. 11, pp. 5139–5145, 2018.

[108] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.05.021

[109] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.05.024

[110] H. Tang, K. Zhang, D. Guo, L. Jia, H. Qiao, and Y. Tian, "Stacked denoising autoencoder based fault diagnosis for rotating motor," vol. 2018-July, 2018, pp. 5757–5762, cited By 0.

[111] Y. Wang, T. Sun, B. Cai, L. Lu, X. Jiang, and X. Hu, "Sensor Fault Diagnosis of Control System Based on CMKPCA and BP Neural Network," *Chinese Control Conference, CCC*, vol. 2018-July, pp. 6061–6066, 2018.

[112] H. Li, F. Zhou, B. Xu, Y. Liu, and B. Yan, "Fault diagnosis of rolling bearing based on variable mode decomposition and multi-class relevance vector machine," *Chinese Control Conference, CCC*, vol. 2018-July, pp. 5954–5960, 2018.

[113] Y. Xie and T. Zhang, "Imbalanced Learning for Fault Diagnosis Problem of Rotating Machinery Based on Generative Adversarial Networks," *Chinese Control Conference, CCC*, vol. 2018-July, pp. 6017–6022, 2018.

[114] H. Li, "Bi-spectrum analysis based bearing fault diagnosis," *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2010*, vol. 6, pp. 2599–2603, 2010.

[115] D. K. Appana, A. Prosvirin, and J. M. Kim, "Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks," *Soft Computing*, vol. 22, no. 20, pp. 6719–6729, 2018. [Online]. Available: https://doi.org/10.1007/s00500-018-3256-0

[116] V. Tran, F. AlThobiani, T. Tinga, A. Ball, and G. Niu, "Single and combined fault diagnosis of reciprocating compressor valves using a hybrid deep belief network," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 232, no. 20, pp. 3767–3780, 2018, cited By 0.

[117] Q. Chen, Y. Hu, J. Xia, Z. Chen, and H. W. Tseng, "Data fusion of wireless sensor network for prognosis and diagnosis of mechanical systems," *Proceedings of the 2017 IEEE International Conference on Information, Communication and Engineering: Information and Innovation for Modern Technology, ICICE 2017*, pp. 331–334, 2018.

[118] X. J. Wan, L. Liu, Z. Xu, Z. Xu, Q. Li, and F. Xu, "Fault diagnosis of rolling bearing based on optimized soft competitive learning Fuzzy ART and similarity evaluation technique," *Advanced Engineering Informatics*, vol. 38, no. May, pp. 91–100, 2018. [Online]. Available: https://doi.org/10.1016/j.aei.2018.06.006

[119] R. Benkercha and S. Moulahoum, "Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system," *Solar Energy*, vol. 173, no. August, pp. 610–634, 2018. [Online]. Available: https://doi.org/10.1016/j.solener.2018.07.089

[120] S. R. Madeti and S. N. Singh, "Modeling of PV system based on experimental data for fault detection using kNN method," *Solar Energy*, vol. 173, no. June, pp. 139–151, 2018. [Online]. Available: https://doi.org/10.1016/j.solener.2018.07.038

[121] A. Khiam, M. Ouassaid, and N. Ngote, "A Hybrid TSA-Neural Network Approach for Induction Motor Faults Diagnosis," *Proceedings of 2017 International Renewable and Sustainable Energy Conference, IRSEC 2017*, pp. 1–6, 2018.

[122] J. Xiao and R. Zhuo, "Research on Motor Rolling Bearing Fault Classification Method Based on CEEMDAN and GWO-SVM," *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018*, no. Imcec, pp. 1123–1129, 2018.

[123] Y. Benmahamed, Y. Kemari, M. Teguar, and A. Boubakeur, "Diagnosis of Power Transformer Oil Using KNN and Nave Bayes Classifiers," *2018 IEEE 2nd International Conference on Dielectrics (ICD)*, no. 3, pp. 1–4, 2018.

[124] Y. Liu, F. Zhou, C. Zeng, J. Ye, D. Hu, L. Li, B. Tan, and X. Chen, "An Automatic Pattern Recognition Methods for Power Network Fault Diagnosis," *International Conference on Innovative Smart Grid Technologies, ISGT Asia 2018*, pp. 396–400, 2018.

[125] J. A. Carino, M. Delgado-Prieto, J. A. Iglesias, A. Sanchis, D. Zurita, M. Millan, J. A. Ortega Redondo, and R. Romero-Troncoso, "Fault Detection and Identification Methodology under an Incremental Learning Framework Applied to Industrial Machinery," *IEEE Access*, vol. 6, no. i, pp. 49 755–49 766, 2018.

[126] J. H. Zhong, P. K. Wong, and Z. X. Yang, "Fault diagnosis of rotating machinery based on multiple probabilistic classifiers," *Mechanical Systems and Signal Processing*, vol. 108, pp. 99–114, 2018. [Online]. Available: https://doi.org/10.1016/j.ymssp.2018.02.009

[127] X. Zhu, J. Zheng, H. Pan, J. Bao, and Y. Zhang, "Time-shift multiscale fuzzy entropy and laplacian support vector machine based rolling bearing fault diagnosis," *Entropy*, vol. 20, no. 8, 2018.

[128] S. Y. Li and L. Xue, "Motor's Early Fault Diagnosis Based on Support Vector Machine," *IOP Conference Series: Materials Science and Engineering*, vol. 382, no. 3, 2018.

[129] F. Wang, B. Dun, G. Deng, H. Li, and Q. Han, "A deep neural network based on kernel function and auto-encoder for bearing fault diagnosis," *I2MTC 2018 - 2018 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Proceedings*, pp. 1–6, 2018.

[130] T. Tang, L. Bo, X. Liu, B. Sun, and D. Wei, "Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification," *Journal of Sound and Vibration*, vol. 425, pp. 137–148, 2018. [Online]. Available: https://doi.org/10.1016/j.jsv.2018.03.032

[131] Z. Zhang, X. Ren, and H. Lv, "Fault diagnosis with feature representation based on stacked sparse auto encoder," *Proceedings - 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2018*, no. 1, pp. 776–781, 2018.

[132] Z. Zhang, Q. Yang, and D. An, "An improved K-means algorithm for reciprocating compressor fault diagnosis," *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pp. 276–281, 2018.

[133] F. Zhou, S. Yang, C. Wen, and H. P. Ju, "Improved DAE and application in fault diagnosis," *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pp. 1801–1806, 2018.

[134] X. Gao, H. Wang, H. Gao, X. Wang, and Z. Xu, "Fault diagnosis of batch process based on denoising sparse auto encoder," *Proceedings - 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2018*, pp. 764–769, 2018.

[135] P. Wen, T. Wang, B. Xin, T. Tang, and Y. Wang, "Blade imbalanced fault diagnosis for marine current turbine based on sparse autoencoder and softmax regression," *Proceedings - 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2018*, pp. 246–251, 2018.

[136] S. Liang and F. Wang, "Fault diagnosis based on pole learning and MCS-LSSVM," *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pp. 2527–2532, 2018.

[137] L. Wen, X. Li, L. Gao, and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[138] C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse Deep Stacking Network for Fault Diagnosis of Motor," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3261–3270, 2018.

[139] X. Chen, Z. Wang, Z. Zhang, L. Jia, and Y. Qin, "A semi-supervised approach to bearing fault diagnosis under variable conditions towards imbalanced unlabeled data," *Sensors (Switzerland)*, vol. 18, no. 7, 2018.

[140] M. Fei, L. Ning, M. Huiyu, P. Yi, S. Haoyuan, and Z. Jianyong, "On-line fault diagnosis model for locomotive traction inverter based on wavelet transform and support vector machine," *Microelectronics Reliability*, vol. 88-90, no. May, pp. 1274–1280, 2018. [Online]. Available: https://doi.org/10.1016/j.microrel.2018.06.069

[141] Y. Xin, S. Li, C. Cheng, and J. Wang, "An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis," *Journal of Vibroengineering*, vol. 20, no. 6, pp. 2321–2335, 2018.

[142] J. Zhang, J. Jiang, and Y. Han, "Semisupervised Regression With Optimized Rank for Matrix Data Classification," 2018.

[143] C. Y. Chang, J. Y. Hong, and W. C. Chang, "Preventive Maintenance of Motors and Automatic Classification of Defects using Artificial Intelligence," *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, MESA 2018*, pp. 1–6, 2018.

[144] Q. Xing-Yu, Z. Peng, X. Chengcheng, and F. Dong-Dong, "RNN-based Method for Fault Diagnosis of Grinding System," *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, CYBER 2017*, pp. 673–678, 2018.

[145] G. Karatzinis, Y. S. Boutalis, and Y. L. Karnavas, "Motor Fault Detection and Diagnosis Using Fuzzy Cognitive Networks with Functional Weights," *MED 2018 - 26th Mediterranean Conference on Control and Automation*, pp. 709–714, 2018.

[146] C. H. Fontes and H. Budman, "Evaluation of a Hybrid Clustering Approach for a Benchmark Industrial System," *Industrial and Engineering Chemistry Research*, vol. 57, no. 32, pp. 11 039–11 049, 2018.

[147] G. Li and Y. Hu, "Improved sensor fault detection, diagnosis and estimation for screw chillers using density-based clustering and principal component analysis," *Energy and Buildings*, vol. 173, pp. 502–515, 2018. [Online]. Available: https://doi.org/10.1016/j.enbuild.2018.05.025

[148] H. Qiao, T. Wang, P. Wang, S. Qiao, and L. Zhang, "A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series," *Sensors (Switzerland)*, vol. 18, no. 9, 2018.

[149] E. Li, L. Wang, B. Song, and S. Jian, "Improved Fuzzy C-Means Clustering for Transformer Fault Diagnosis Using Dissolved Gas Analysis Data," *Energies*, vol. 11, no. 9, p. 2344, 2018.

[150] X. Shi, W. Li, Q. Gao, and H. Guo, "Research on fault classification of wind turbine based on IMF kurtosis and PSO-SOM-LVQ," *Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2017*, vol. 2018-January, pp. 191–196, 2018.

[151] H. A. B. Illias, L. J. Awalin, S. S. Gururajapathy, H. Mokhlis, and A. H. Abu Bakar, "Fault location in an unbalanced distribution system using support vector classification and regression analysis," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 13, no. 2, pp. 237–245, 2017.

[152] K. Yan, Z. Ji, L. Ma, D. Xie, Y. Dai, and W. Shen, "Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis," *International Journal of Refrigeration*, vol. 86, pp. 401–409, 2017. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2017.11.003

[153] Z. Li, H. Fang, M. Huang, Y. Wei, and L. Zhang, "Data-driven bearing fault identification using improved hidden Markov model and self-organizing map," *Computers and Industrial Engineering*, vol. 116, no. December 2017, pp. 37–46, 2018. [Online]. Available: https://doi.org/10.1016/j.cie.2017.12.002

[154] V. Pashazadeh, F. R. Salmasi, and B. N. Araabi, "Data driven sensor and actuator fault detection and isolation in wind turbine using classifier fusion," *Renewable Energy*, vol. 116, pp. 99–106, 2018. [Online]. Available: https://doi.org/10.1016/j.renene.2017.03.051

[155] H. Zhao, S. Sun, and B. Jin, "Sequential Fault Diagnosis Based on LSTM Neural Network," *IEEE Access*, vol. 6, pp. 12 929–12 939, 2018.

[156] N. A. Mohsun, "Broken rotor bar fault classification for induction motor based on support vector machine-SVM," *Proceedings - 2017 International Conference on Engineering and MIS, ICEMIS 2017*, vol. 2018-January, no. 1, pp. 1–6, 2018.

[157] J. Zhao, P. Zhao, S. Liu, Y. Liu, Q. Wang, and N. Jiao, "A fault diganosis method based on bispectrum-LPP and PNN," *Proceedings - 4th International Conference on Dependable Systems and Their Applications, DSA 2017*, vol. 2018-January, pp. 98–102, 2018.

[158] W. Yu and C. Zhao, "A multi-model exponential discriminant analysis algorithm for online probabilistic diagnosis of time-varying faults," *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017*, vol. 2018-January, no. Cdc, pp. 5751–5756, 2018.

[159] W. Grega and A. Tutaj, "A Co-design method for networked feedback control," *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, no. 1, pp. 1–8, 2018.

[160] J. Li, Y. Zhu, Y. Xu, and C. Guo, "A transformer fault diagnosis method based on sub-clustering reduction and multiclass multi-kernel support vector machine," vol. 2018-January, 2018, pp. 1–6, cited By 1.

[161] Y. Wang, "Mechanical Fault Diagnosis Based on Fuzzy Clustering Algorithm," *International Journal of Mechatronics and Applied Mechanics*, vol. 1, no. 3, 2018.

[162] S. Atanassov, Krassimir T Kacprzyk, Janusz Krawczak, Maciej Szmidt, Eulalia Zadrożny, *Advances in Fuzzy Logic and Technology 2017*, 2018, vol. 642. [Online]. Available: http://link.springer.com/10.1007/978-3-319-66824-6

[163] C. Li, M. Cerrada, D. Cabrera, R. Sanchez, F. Pacheco, G. Ulutagay, and J. Valente De Oliveira, "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3565–3580, 2018, cited By 10.

[164] Y. Yang, P. Fu, and Y. He, "Bearing Fault Automatic Classification Based on Deep Learning," *IEEE Access*, vol. 6, pp. 71 540–71 554, 2018.

[165] B. Zheng, H.-Z. Huang, W. Guo, Y.-F. Li, and J. Mi, "Fault diagnosis method based on supervised particle swarm optimization classification algorithm," *Intelligent Data Analysis*, vol. 22, no. 1, pp. 191–210, 2018, cited By 5.

[166] J. Meng, L. Zhao, F. Shen, and R. Yan, "Gear fault diagnosis based on recurrence network," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3651–3660, 2018, cited By 1.

[167] Q. Fu, B. Jing, P. He, S. Si, and Y. Wang, "Fault Feature Selection and Diagnosis of Rolling Bearings Based on EEMD and Optimized Elman-AdaBoost Algorithm," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5024–5034, 2018.

[168] S. Zhou, S. Qian, W. Chang, Y. Xiao, and Y. Cheng, "A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier," *Sensors (Switzerland)*, vol. 18, no. 6, 2018.

[169] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, "On-line failure diagnosis of grid-connected photovoltaic systems based on fuzzy logic," *Proceedings - 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG 2018*, pp. 1–6, 2018.

[170] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018. [Online]. Available: https://doi.org/10.1016/j.isatra.2018.04.005

[171] B. Wang, J. Pan, Y. Zi, J. Chen, and Z. Zhou, "LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy Mechanical Data for Fault Classification," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4973–4982, 2017.

[172] Z. Zilong and Q. Wei, "Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification," *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control*, pp. 1–6, 2018.

[173] S. Guo, T. Yang, W. Gao, and C. Zhang, "A novel fault diagnosis method for rotating machinery based on a convolutional neural network," *Sensors (Switzerland)*, vol. 18, no. 5, 2018.

[174] L. Jing, L. Yankai, C. Shanshan, W. Zhijie, and J. Haipeng, "Application of deep learning in the hydraulic equipment fault diagnosis," *Proceedings - 2018 IEEE International Conference on Smart Manufacturing, Industrial and Logistics Engineering, SMILE 2018*, vol. 2018-January, pp. 12–16, 2018.

[175] S. Mjahed, S. El Hadaj, K. Bouzaachane, and S. Raghay, "Engine Fault Signals Diagnosis using Genetic Algorithm and K-means based Clustering," pp. 1–6, 2018.

[176] M. L. Fadda and A. Moussaoui, "Hybrid SOMPCA method for modeling bearing faults detection and diagnosis," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 40, no. 5, pp. 1–8, 2018. [Online]. Available: https://doi.org/10.1007/s40430-018-1184-7

[177] S. Zgarni, H. Keskes, and A. Braham, "Nested SVDD in DAG SVM for induction motor condition monitoring," *Engineering Applications of Artificial Intelligence*, vol. 71, no. February, pp. 210–215, 2018. [Online]. Available: https://doi.org/10.1016/j.engappai.2018.02.019

[178] M. Zhang, W. Cheng, and Y. Wang, "Multiple-Fault Classification for Hot-Mix Asphalt Production by Machine Learning," *Journal of Construction Engineering and Management*, vol. 144, no. 5, p. 04018024, 2018.

[179] A. Picot, J. Riviere, and P. Maussion, "Bearing fault diagnosis based on the analysis of recursive PCA projections," *Proceedings of the IEEE International Conference on Industrial Technology*, vol. 2018-February, pp. 2093–2098, 2018.

[180] S. T. Gul, M. Imran, and A. Q. Khan, "An online incremental support vector machine for fault diagnosis using vibration signature analysis," *Proceedings of the IEEE International Conference on Industrial Technology*, vol. 2018-February, pp. 1467–1472, 2018.

[181] S. E. Pandarakone, M. Masuko, Y. Mizuno, and H. Nakamura, "Fault classification of outer-race bearing damage in low-voltage induction motor with aid of fourier analysis and SVM," *Proceedings of the IEEE International Conference on Industrial Technology*, vol. 2018-February, pp. 407–412, 2018.

[182] Q. Han, F. Wang, Z. Xue, W. Su, H. Li, and C. Liu, "Combined Failure Diagnosis of Slewing Bearings Based on MCKD-CEEMD-ApEn," *Shock and Vibration*, vol. 2018, pp. 1–13, 2018.

[183] A. Ragab, M. El-Koujok, B. Poulin, M. Amazouz, and S. Yacout, "Fault diagnosis in industrial chemical processes using interpretable patterns based on logical analysis of data," *Expert Systems with Applications*, vol. 95, pp. 368–383, 2018, cited By 4.

[184] B. Duong and J.-M. Kim, "Non-mutually exclusive deep neural network classifier for combined modes of bearing fault diagnosis," *Sensors (Switzerland)*, vol. 18, no. 4, 2018, cited By 3.

[185] J. Ji, J. Qu, Y. Chai, Y. Zhou, Q. Tang, and H. Ren, "An algorithm for sensor fault diagnosis with eemd-svm," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 6, pp. 1746–1756, 2018, cited By 0.

[186] H. Yuan, X. Hou, and H. Wang, "Multi-fault diagnosis for rolling bearing based on double parallel extreme learning machine & kurtosis spectral entropy," vol. 2018-March, 2018, pp. 758–763, cited By 0.

[187] Z. Shi, F. Chen, and L. Cao, "Fault diagnosis of steam turbine rotor based on permutation entropy and ifoa-rvm," *Zhendong yu Chongji/Journal of Vibration and Shock*, vol. 37, no. 5, pp. 79–84 and 113, 2018, cited By 0.

[188] W. Shangguan, J. Zhang, J. Feng, and B. Cai, "Fault diagnosis method of the on-board equipment of train control system based on rough set theory," vol. 2018-March, 2018, pp. 1–6, cited By 0.

[189] S. Marmouch, T. Aroui, and Y. Koubaa, "Induction machine faults diagnosis by statistical neural networks with selection variables based on principal component analysis," vol. 2018-January, 2018, pp. 99–103, cited By 0.

[190] Y. Liu, J. Zhang, K. Qin, and Y. Xu, "Diesel engine fault diagnosis using intrinsic time-scale decomposition and multistage adaboost relevance vector machine," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 232, no. 5, pp. 881–894, 2018, cited By 3.

[191] A. Prosvirin, J. Kim, and J.-M. Kim, "Bearing fault diagnosis based on convolutional neural networks with kurtogram representation of acoustic emission signals," *Lecture Notes in Electrical Engineering*, vol. 474, pp. 21–26, 2018, cited By 2.

[192] M. Cerrada, R.-V. Snchez, and D. Cabrera, "A semi-supervised approach based on evolving clusters for discovering unknown abnormal condition patterns in gearboxes," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3581–3593, 2018, cited By 1.

[193] M. Sohaib and J.-M. Kim, "Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network," *Shock and Vibration*, vol. 2018, 2018, cited By 3.

[194] P. Panigrahy and P. Chattopadhyay, "Improved classification by non iterative and ensemble classifiers in motor fault diagnosis," *Advances in Electrical and Computer Engineering*, vol. 18, no. 1, pp. 95–104, 2018, cited By 0.

[195] B. Jiao and Z. Zhou, "The fault diagnosis of wind turbine bearing based on isapso-svm," vol. 1, 2018, pp. 77–81, cited By 0.

[196] Y. Chen, Y. Chen, and Q. Dai, "Novel gear fault diagnosis approach using native bayes uncertain classification based on pso algorithm," *Journal of Vibroengineering*, vol. 20, no. 3, pp. 1370–1381, 2018, cited By 0.

[197] M. Zair, C. Rahmoune, and D. Benazzouz, "Multi-fault diagnosis of rolling bearing using fuzzy entropy of empirical mode decomposition, principal component analysis, and som neural network," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2018, cited By 0.

[198] X. Tang, Y. Chai, Y. Mao, and J. Ji, "Fault diagnosis of rolling bearing based on wavelet packet and extreme learning machine," *Lecture Notes in Electrical Engineering*, vol. 459, pp. 131–140, 2018, cited By 0.

[199] S. Ravikumar, S. Kanagasabapathy, V. Muralidharan, R. Srijith, and M. Bimalkumar, "Fault diagnosis of self-aligning troughing rollers in a belt conveyor system using an artificial neural network and naive bayes algorithm," 2018, pp. 401–408, cited By 0.

[200] X. Wang, J. Zhao, S. Wang, and Z. Ma, "Fault diagnosis of electromechanical actuator based on principal component analysis and support vector machine," vol. 2018, no. CP743, 2018, pp. 1467–1471, cited By 0.

[201] A. Moosavian, S. Jafari, M. Khazaee, and H. Ahmadi, "A comparison between ann, svm and least squares svm: Application in multi-fault diagnosis of rolling element bearing," *International Journal of Acoustics and Vibrations*, vol. 23, pp. 432–440, 2018, cited By 0.

[202] F. Dong, X. Yu, E. Ding, S. Wu, C. Fan, and Y. Huang, "Rolling bearing fault diagnosis using modified neighborhood preserving embedding and maximal overlap discrete wavelet packet transform with sensitive features selection," *Shock and Vibration*, vol. 2018, 2018, cited By 1.

[203] Y. Li, M. Han, B. Han, X. Le, and S. Kanae, "Fault diagnosis method of diesel engine based on improved structure preserving and k-nn algorithm," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10878 LNCS, pp. 656–664, 2018, cited By 0.

[204] Z. Chen, W. Li, and K. Gryllias, "Gearbox fault diagnosis based on convolutional neural networks," 2018, pp. 941–953, cited By 0.

[205] X. Bai, J. Tan, and X. Wang, "Fault diagnosis and knowledge extraction using fast logical analysis of data with multiple rules discovery ability," *IFIP Advances in Information and Communication Technology*, vol. 539, pp. 412–421, 2018, cited By 0.

[206] J. Yu, M. Bai, G. Wang, and X. Shi, "Fault diagnosis of planetary gearbox with incomplete information using assignment reduction and flexible naive bayesian classifier," *Journal of Mechanical Science and Technology*, vol. 32, no. 1, pp. 37–47, 2018, cited By 4.

[207] T. Narendiranath Babu, A. Aravind, A. Rakesh, M. Jahzan, and D. Rama Prabha, "Application of emd ann and dnn for self-aligning bearing fault diagnosis," *Archives of Acoustics*, vol. 43, no. 2, pp. 163–175, 2018, cited By 0.

[208] R.-V. Snchez, P. Lucero, R. Vsquez, M. Cerrada, J.-C. Macancela, and D. Cabrera, "Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and knn," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3463–3473, 2018, cited By 3.

[209] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69 907–69 917, 2018, cited By 0.

[210] M. Pea, M. Cerrada, X. Alvarez, D. Jadn, P. Lucero, B. Milton, R. Guamn, and R.-V. Snchez, "Feature engineering based on anova, cluster validity assessment and knn for fault diagnosis in bearings," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3451–3462, 2018, cited By 1.

[211] W. Song, J. Xiang, and Y. Zhong, "A simulation model based fault diagnosis method for bearings," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3857–3867, 2018, cited By 1.

[212] K. Li, R. Ran, S. Song, J. Wang, and L. Wang, "Spacecraft electrical signal classification method of reliability test based on random forest," *Lecture Notes in Electrical Engineering*, vol. 456, pp. 457–465, 2018, cited By 0.

[213] K. Moloi, J. Jordaan, and Y. Hamam, "Support vector machine based method for high impedance fault diagnosis in power distribution networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11314 LNCS, pp. 9–16, 2018, cited By 0.

[214] Z. Tong, W. Li, B. Zhang, and M. Zhang, "Bearing fault diagnosis based on domain adaptation using transferable features under different working conditions," *Shock and Vibration*, vol. 2018, 2018, cited By 0.

[215] T. Sang and C. Yang, "Application of improved rough set reduction algorithm in online power distribution fault diagnosis," *Paper Asia*, vol. 1, no. Compendium8, pp. 114–117, 2018, cited By 0.

[216] T. Benkedjouh, T. Chettibi, Y. Saadouni, and M. Afroun, "Gearbox fault diagnosis based on mel-frequency cepstral coefficients and support vector machine," *IFIP Advances in Information and Communication Technology*, vol. 522, pp. 220–231, 2018, cited By 1.

[217] K. Leahy, R. Hu, I. Konstantakopoulos, C. Spanos, A. Agogino, and D. OSullivan, "Diagnosing and predicting wind turbine faults from scada data using support vector machines," *International Journal of Prognostics and Health Management*, vol. 9, no. 1, 2018, cited By 8.

[218] J. Ma, J. Wu, and X. Wang, "Fault diagnosis method based on wavelet packet-energy entropy and fuzzy kernel extreme learning machine," *Advances in Mechanical Engineering*, vol. 10, no. 1, 2018, cited By 4.

[219] X. Qu and S. Patnaik, "Resource optimisation and fault detection algorithms for cloud computing platforms based on svm and resource reserve strategy,"

International Journal of Applied Decision Sciences, vol. 11, no. 3, pp. 223–237, 2018, cited By 0.

[220] T. Praveenkumar, B. Sabhrish, M. Saimurugan, and K. Ramachandran, "Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox," *Measurement: Journal of the International Measurement Confederation*, vol. 114, pp. 233–242, 2018, cited By 7.

[221] I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervilis, K. Worden, and E. Maguire, "Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data," 2018, pp. 3053–3062, cited By 2.

[222] M. Gong, H. Song, J. Tan, Y. Xie, and J. Song, "Fault diagnosis of motor based on mutative scale back propagation net evolving fuzzy petri nets," vol. 2017-January, 2017, pp. 3826–3829, cited By 1.

[223] C. Baigen, Y. Jiaming, S. Weif, and C. Bin, "Research on fault recognition method of on-board equipment based on bp neural network optimized by bayesian regularized," vol. 2017-January, 2017, pp. 6835–6840, cited By 0.

[224] S. Zolfaghari, S. Noor, M. Mehrjou, M. Marhaban, and N. Mariun, "Broken rotor bar fault detection and classification using wavelet packet signature analysis based on fourier transform and multi-layer perceptron neural network," *Applied Sciences (Switzerland)*, vol. 8, no. 1, 2017, cited By 5.

[225] L. Kai, L. Bo, M. Tao, Y. Xuefeng, and W. Guangming, "Unsupervised learning - a novel clustering method for rolling bearing faults identification," vol. 280, no. 1, 2017, cited By 0.

[226] R. Swain and P. Khilar, "Soft fault diagnosis in wireless sensor networks using pso based classification," vol. 2017-December, 2017, pp. 2456–2461, cited By 3.

[227] C. Teh, A. Aziz, I. Elamvazuthi, and Z. Man, "Classification of bearing faults using extreme learning machine algorithm," vol. 2017-December, 2017, pp. 1–6, cited By 0.

[228] N. Aditiya, M. Dharmawan, Z. Darojah, and D. Raden Sanggar, "Fault diagnosis system of rotating machines using hidden markov model (hmm)," vol. 2017-January, 2017, pp. 177–181, cited By 1.

[229] W. Chine and A. Mellit, "Ann-based fault diagnosis technique for photovoltaic stings," vol. 2017-January, 2017, pp. 1–4, cited By 0.

[230] A. Belaout, F. Krim, A. Sahli, and A. Mellit, "Multi-class neuro-fuzzy classifier for photovoltaic array faults diagnosis," vol. 2017-January, 2017, pp. 1–4, cited By 1.

[231] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors (Switzerland)*, vol. 17, no. 12, 2017, cited By 17.

[232] J. Xiong, S. Tian, C. Yang, B. Yu, and S. Chen, "Analog fault feature extraction and classification based on lmd and lvq neural network," vol. 2017-January, 2017, pp. 1–6, cited By 1.

[233] F. Pacheco, M. Cerrada, and R.-V. Sanchez, "Framework for discovering unknown abnormal condition patterns in gearboxes using a semi-supervised approach," vol. 2017-December, 2017, pp. 63–68, cited By 0.

[234] Y. Liao, L. Zhang, and W. Li, "Regrouping particle swarm optimization-based neural network for bearing fault diagnosis," vol. 2017-December, 2017, pp. 628–631, cited By 0.

[235] J. Meng, L. Zhao, and R. Yan, "Gear fault diagnosis based on recurrence network," vol. 2017-December, 2017, pp. 514–517, cited By 0.

[236] G. Tang, G. Li, and H. Wang, "Sparse component analysis based on support vector machine for fault diagnosis of roller bearings," vol. 2017-December, 2017, pp. 415–420, cited By 1.

[237] S. He, L. Xiao, Y. Wang, X. Liu, C. Yang, J. Lu, W. Gui, and Y. Sun, "A novel fault diagnosis method based on optimal relevance vector machine," *Neurocomputing*, vol. 267, pp. 651–663, 2017, cited By 7.

[238] J. Cui, G. Shi, and C. Gong, "A fast classification method of faults in power electronic circuits based on support vector machines," *Metrology and Measurement Systems*, vol. 24, no. 4, pp. 701–720, 2017, cited By 1.

[239] Y.-P. Zhao, F.-Q. Song, Y.-T. Pan, and B. Li, "Retargeting extreme learning machines for classification and their applications to fault diagnosis of aircraft engine," *Aerospace Science and Technology*, vol. 71, pp. 603–618, 2017, cited By 5.

[240] D. Dey, B. Chatterjee, S. Dalai, S. Munshi, and S. Chakravorti, "A deep learning framework using convolution neural network for classification of impulse fault patterns in transformers with increased accuracy," *IEEE Transactions on*

Dielectrics and Electrical Insulation, vol. 24, no. 6, pp. 3894–3897, 2017, cited By 3.

[241] J. Rapur and R. Tiwari, "Experimental time-domain vibration- based fault diagnosis of centrifugal pumps using support vector machine," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 3, no. 4, 2017, cited By 7.

[242] A. Glowacz, W. Glowacz, Z. Glowacz, J. Kozik, M. Gutten, D. Korenciak, Z. Khan, M. Irfan, and E. Carletti, "Fault diagnosis of three phase induction motor using current signal, msaf-ratio15 and selected classifiers," *Archives of Metallurgy and Materials*, vol. 62, no. 4, pp. 2413–2419, 2017, cited By 9.

[243] P. Wang, "The application of radial basis function (rbf) neural network for mechanical fault diagnosis of gearbox," vol. 269, no. 1, 2017, cited By 2.

[244] C. Yao, B. Lai, D. Chen, F. Sun, and S. Lyu, "Fault diagnosis method based on med-vmd and optimized svm for rolling bearings," *Zhongguo Jixie Gongcheng/China Mechanical Engineering*, vol. 28, no. 24, pp. 3001–3012, 2017, cited By 0.

[245] R. Chen, X. Yang, L. Yang, J. Wang, X. Xu, and S. Chen, "Fault severity diagnosis method for rolling bearings based on a stacked sparse denoising auto-encoder," *Zhendong yu Chongji/Journal of Vibration and Shock*, vol. 36, no. 21, pp. 125–131 and 137, 2017, cited By 1.

[246] M. Zhao, B. Li, J. Qi, and Y. Ding, "Semi-supervised classification for rolling fault diagnosis via robust sparse and low-rank model," 2017, pp. 1062–1067, cited By 1.

[247] G. Li, C. Yu, H. Fan, Y. Liu, and Y. Song, "Deep fault diagnosis of power transformer based on multilevel decision fusion model," *Dianli Zidonghua Shebei/Electric Power Automation Equipment*, vol. 37, no. 11, pp. 138–144, 2017, cited By 1.

[248] X. Ma, X. Zhang, and Z. Yang, "Fault diagnosis and classification of mine motor based on rs and svm," *Studies in Computational Intelligence*, vol. 672, pp. 17–30, 2017, cited By 0.

[249] L. Umasankar and N. Kalaiarasi, "Multi wavelet transform with radial basis function neural network for internal faults diagnosis in power transformer," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 11, pp. 5525–5538, 2017, cited By 0.

[250] J. Oliveira, K. Pontes, I. Sartori, and M. Embiruu, "Fault detection and diagnosis in dynamic systems using weightless neural networks," *Expert Systems with Applications*, vol. 84, pp. 200–219, 2017, cited By 11.

[251] W. Jiang and L. Yang, "A research on simultaneous fault diagnosis based on paired-rvm," 2017, pp. 483–488, cited By 0.

[252] Y.-T. Hu, F.-H. Qu, and C.-J. Wen, "Bearing fault diagnosis based on multi-scale possibilistic clustering algorithm," 2017, pp. 354–357, cited By 0.

[253] Q. Feng, H. Lian, and J. Zhu, "Multi-level genetic programming for fault data clustering," 2017, cited By 1.

[254] Y. Xin, S. Li, and J. Wang, "A novel roller bearing fault diagnosis method based on the wavelet extreme learning machine," 2017, cited By 0.

[255] S. Wu, X. Chen, Z. Zhao, and R. Liu, "Data-driven discriminative k-svd for bearing fault diagnosis," 2017, cited By 1.

[256] S. Dong, Z. Zhang, G. Wen, and G. Wen, "Design and application of unsupervised convolutional neural networks integrated with deep belief networks for mechanical fault diagnosis," 2017, cited By 0.

[257] H. Malik and S. Mishra, "Application of fuzzy q learning (fql) technique to wind turbine imbalance fault identification using generator current signals," 2017, cited By 0.

[258] S. Laamami, M. Benhamed, and L. Sbita, "Artificial neural network-based fault detection and classification for photovoltaic system," 2017, cited By 2.

[259] O. Duque-Perez, C. Del Pozo-Gallego, D. Morinigo-Sotelo, and W. Fontes Godoy, "Bearing fault diagnosis based on lasso regularization method," vol. 2017-January, 2017, pp. 331–337, cited By 1.

[260] K. Elangovan, Y. Tamilselvam, R. Mohan, M. Iwase, T. Nemoto, and K. Wood, "Fault diagnosis of a reconfigurable crawling-rolling robot based on support vector machines," *Applied Sciences (Switzerland)*, vol. 7, no. 10, 2017, cited By 5.

[261] J. Senanayaka, H. Van Khang, and K. Robbersmyr, "Towards online bearing fault detection using envelope analysis of vibration signal and decision tree classification algorithm," 2017, cited By 2.

[262] F. Wang, W. Zhang, and Y. Ding, “Bearing fault diagnosis based on intrinsic time-scale decomposition and extreme learning machine,” vol. 14, 2017, pp. 97–101, cited By 0.

[263] L. Chao, C. Lu, and J. Ma, “An approach to fault diagnosis for gearbox based on reconstructed energy and support vector machine,” vol. 14, 2017, pp. 136–140, cited By 0.

[264] K. Dek and I. Kocsis, “Support vector machine with wavelet decomposition method for fault diagnosis of tapered roller bearings by modelling manufacturing defects,” *Periodica Polytechnica Mechanical Engineering*, vol. 61, no. 4, pp. 276–281, 2017, cited By 0.

[265] L. Wang, J. Shen, and X. Mei, “Cost sensitive multi-class fuzzy decision-theoretic rough set based fault diagnosis,” 2017, pp. 6957–6961, cited By 4.

[266] X. Xu, C. Wen, and Z. Zhou, “Fault classification method based on random projections and svm,” 2017, pp. 7242–7246, cited By 0.

[267] X. Zhang, D. Jiang, Q. Long, and T. Han, “Rotating machinery fault diagnosis for imbalanced data based on decision tree and fast clustering algorithm,” *Journal of Vibroengineering*, vol. 19, no. 6, pp. 4247–4259, 2017, cited By 2.

[268] M. Islam, G. Lee, and S. Hettiwatte, “A nearest neighbour clustering approach for incipient fault diagnosis of power transformers,” *Electrical Engineering*, vol. 99, no. 3, pp. 1109–1119, 2017, cited By 7.

[269] H. Fu, Z. Qi, and R. Ren, “A new fault diagnosis method for transformer based on improved rvm,” *Liaoning Gongcheng Jishu Daxue Xuebao (Ziran Kexue Ban)/Journal of Liaoning Technical University (Natural Science Edition)*, vol. 36, no. 9, pp. 964–970, 2017, cited By 0.

[270] G. Jiang, H. He, P. Xie, and Y. Tang, “Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017, cited By 20.

[271] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, “Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests,” *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017, cited By 17.

[272] R. Taleb, L. Seddiki, K. Guelton, and H. Akdag, “Merging fuzzy observer-based state estimation and database classification for fault detection and diagnosis of an actuated seat,” 2017, cited By 0.

[273] T. Dos Santos, F. Ferreira, J. Pires, and C. Damasio, "Stator winding short-circuit fault diagnosis in induction motors using random forest," 2017, cited By 1.

[274] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017, cited By 28.

[275] M. Cho and T. Hoang, "A differential particle swarm optimizationbased support vector machine classifier for fault diagnosis in power distribution systems," *Advances in Electrical and Computer Engineering*, vol. 17, no. 3, pp. 51–60, 2017, cited By 0.

[276] V. Vakharia, V. Gupta, and P. Kankar, "Efficient fault diagnosis of ball bearing using relieff and random forest classifier," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 39, no. 8, pp. 2969–2982, 2017, cited By 5.

[277] D.-R. Huang, C.-S. Chen, G.-X. Sun, L. Zhao, and B. Mi, "Linear discriminant analysis and back propagation neural network cooperative diagnosis method for multiple faults of complex equipment bearings," *Binggong Xuebao/Acta Armamentarii*, vol. 38, no. 8, pp. 1649–1657, 2017, cited By 7.

[278] Q. Jiang, Q. Zhu, B. Wang, and L. Guo, "Nonlinear machine fault detection by semi-supervised laplacian eigenmaps," *Journal of Mechanical Science and Technology*, vol. 31, no. 8, pp. 3697–3703, 2017, cited By 1.

[279] Y. Qi, E. Zhang, S. Gao, L. Wang, and X. Gao, "Wind turbine rolling bearings fault diagnosis based on eemd-keca," *Taiyangneng Xuebao/Acta Energiae Solaris Sinica*, vol. 38, no. 7, pp. 1943–1951, 2017, cited By 0.

[280] P. Yang, M. Liu, Q. Peng, Y. Xing, and B. Li, "A improved way for fault diagnose based on eemd and svm," 2017, pp. 331–335, cited By 0.

[281] G. Chen, F. Liu, and W. Huang, "Sparse discriminant manifold projections for bearing fault diagnosis," *Journal of Sound and Vibration*, vol. 399, pp. 330–344, 2017, cited By 7.

[282] D. Khan and S. Saha, "Neural network based cycloconverter fault detection using wavelet decomposition," 2017, pp. 352–356, cited By 0.

[283] J. He, S. Yang, and C. Gan, "Unsupervised fault diagnosis of a gear transmission chain using a deep belief network," *Sensors (Switzerland)*, vol. 17, no. 7, 2017, cited By 7.

[284] F. Boem, R. Reci, A. Cenedese, and T. Parisini, "Distributed clustering-based sensor fault diagnosis for hvac systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 4197–4202, 2017, cited By 1.

[285] Y.-B. Jing, C.-W. Liu, F.-R. Bi, X.-Y. Bi, X. Wang, and K. Shao, "Diesel engine valve clearance fault diagnosis based on features extraction techniques and fastica-svm," *Chinese Journal of Mechanical Engineering (English Edition)*, vol. 30, no. 4, pp. 991–1007, 2017, cited By 3.

[286] K. Vernekar, H. Kumar, and K. Gangadharan, "Engine gearbox fault diagnosis using empirical mode decomposition method and nave bayes algorithm," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 42, no. 7, pp. 1143–1153, 2017, cited By 1.

[287] R. Lizarraga-Morales, C. Rodriguez-Donate, E. Cabal-Yepez, M. Lopez-Ramirez, L. Ledesma-Carrillo, and E. Ferrucho-Alvarez, "Novel fpga-based methodology for early broken rotor bar detection and classification through homogeneity estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1760–1769, 2017, cited By 12.

[288] N. Barbosa, L. Trav-Massuys, and V. Grisales, "Diagnosability improvement of dynamic clustering through automatic learning of discrete event models," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1037–1042, 2017, cited By 0.

[289] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017, cited By 42.

[290] H.-C. Yan, J.-H. Zhou, and C. Pang, "Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 723–733, 2017, cited By 12.

[291] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017, cited By 30.

[292] Z. Huang, Z. Wang, and H. Zhang, "Multilevel feature moving average ratio method for fault diagnosis of the microgrid inverter switch," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 177–185, 2017, cited By 5.

[293] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *International Journal of Advanced Manufacturing Technology*, vol. 89, no. 9-12, pp. 3487–3494, 2017, cited By 10.

[294] T. Han, D. Jiang, X. Zhang, and Y. Sun, "Intelligent diagnosis method for rotating machinery using dictionary learning and singular value decomposition," *Sensors (Switzerland)*, vol. 17, no. 4, 2017, cited By 11.

[295] L. Ying, L. Zhengda, X. Minghua, Z. Zhen, and Y. Lifen, "Neural network based preprocessing and dictionary methods for switched current circuit fault diagnosis," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 4, pp. 1914–1924, 2017, cited By 0.

[296] S. Chen, "A kind of semi-supervised classifying method research for power transformer fault diagnosis," 2017, pp. 1013–1016, cited By 1.

[297] F. Zhou, Y. Gao, and C. Wen, "A Novel Multimode Fault Classification Method Based on Deep Learning," *Journal of Control Science and Engineering*, 2017.

[298] R. Zhang, Z. Peng, L. Wu, B. Yao, and Y. Guan, "Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence," *Sensors (Switzerland)*, vol. 17, no. 3, 2017, cited By 21.

[299] C.-W. Liu and Z.-W. Shang, "Fault diagnosis of rolling machinery based on scattering transform and smooth one dimensional manifold embedding," vol. 2, 2017, pp. 990–995, cited By 2.

[300] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers," *Journal of Electrical Engineering and Technology*, vol. 12, no. 2, pp. 830–839, 2017, cited By 6.

[301] J. Jiang, H. Wang, Y. Ke, and W. Xiang, "Fault diagnosis based on ltsa and k-nearest neighbor classifier," *Zhendong yu Chongji/Journal of Vibration and Shock*, vol. 36, no. 11, pp. 134–139, 2017, cited By 1.

[302] Y. Zhang, R. Wang, P. Chen, L. Jiao, J. Gu, J. Xie, Y. Cui, F. Liu, and S. Yang, "Random subspace based ensemble sparse representation," *Pattern Recognition*, 2017.

[303] J. Senanayaka, S. Kandukuri, H. Khang, and K. Robbersmyr, “Early detection and classification of bearing faults using support vector machine algorithm,” 2017, pp. 250–255, cited By 10.

[304] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani, and A. Deliou, “A new time-frequency method for identification and classification of ball bearing faults,” *Journal of Sound and Vibration*, vol. 397, pp. 241–265, 2017, cited By 12.

[305] B. Qin, G. Sun, L. Zhang, J. Wang, and J. Hu, “Fault features extraction and identification based rolling bearing fault diagnosis,” vol. 842, no. 1, 2017, cited By 1.

[306] K. Yu, T. Lin, and J. Tan, “A bearing fault diagnosis technique based on singular values of eemd spatial condition matrix and gath-geva clustering,” *Applied Acoustics*, vol. 121, pp. 33–45, 2017, cited By 17.

[307] C. Cai, X. Weng, and C. Zhang, “A novel approach for marine diesel engine fault diagnosis,” *Cluster Computing*, vol. 20, no. 2, pp. 1691–1702, 2017, cited By 2.

[308] S. Dong, X. Xu, and J. Luo, “Mechanical fault diagnosis method based on lmd shannon entropy and improved fuzzy c-means clustering,” *International Journal of Acoustics and Vibrations*, vol. 22, no. 2, pp. 211–217, 2017, cited By 1.

[309] M. Golmoradi, E. Ebrahimi, and M. Javidan, “Compressor fault diagnosis based on svm and ga,” vol. 12, 2017, pp. 49–53, cited By 0.

[310] ——, “Fault diagnosis of compressor based on decision tree and fuzzy inference system,” vol. 12, 2017, pp. 54–60, cited By 0.

[311] W. Luo, M. Peng, X. Wan, S. Wang, and X. Liao, “Transformer fault diagnosis based on chemical reaction optimization algorithm and relevance vector machine,” vol. 108, 2017, cited By 0.

[312] L. Li and Z. Wu, “An on-line fault diagnosis method for gas engine using ap clustering,” 2017, pp. 149–155, cited By 0.

[313] K. Lee, S. Cheon, and C. Kim, “A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135–142, 2017, cited By 16.

[314] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057–3065, 2017, cited By 32.

[315] S.-W. Fei, "Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and svm classifier," *Arabian Journal for Science and Engineering*, vol. 42, no. 5, pp. 1967–1975, 2017, cited By 11.

[316] A. Patil, J. Gaikwad, and J. Kulkarni, "Bearing fault diagnosis using discrete wavelet transform and artificial neural network," 2017, pp. 399–405, cited By 4.

[317] Y. Sun, "Optimized svm based on improved foa and its application in fault daignosis," *Jixie Qiangdu/Journal of Mechanical Strength*, vol. 39, no. 2, pp. 285–290, 2017, cited By 0.

[318] G. Mani and N. Sivaraman, "Integrating fuzzy based fault diagnosis with constrained model predictive control for industrial applications," *Journal of Electrical Engineering and Technology*, vol. 12, no. 2, pp. 886–889, 2017, cited By 0.

[319] J. Tao, Y. Liu, Z. Fu, D. Yang, and F. Tang, "Fault damage degrees diagnosis for rolling bearing based on teager energy operator and deep belief network," *Zhongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Central South University (Science and Technology)*, vol. 48, no. 1, pp. 61–68, 2017, cited By 0.

[320] Y. Sun, N. Pan, and Z. Yi, "Rolling bearing faults detection based on improved sparse component analysis," 2017, pp. 772–775, cited By 2.

[321] G. Yanqing, L. Lu, and F. Yongling, "Fault diagnosis of sensing elements in solenoid valve neutral function identification system based on svm," 2017, pp. 2045–2050, cited By 0.

[322] B. She, F. Tian, J. Tang, and K. Li, "Fault diagnosis of rolling bearing based on adaptive ltsa algorithm," *Huazhong Keji Daxue Xuebao (Ziran Kexue Ban)/Journal of Huazhong University of Science and Technology (Natural Science Edition)*, vol. 45, no. 1, pp. 91–96, 2017, cited By 0.

[323] P. Ma, H. Zhang, and W. Fan, "Fault diagnosis of rolling bearings based on local and global preserving embedding algorithm," *Jixie Gongcheng Xuebao/Journal of Mechanical Engineering*, vol. 53, no. 2, pp. 20–25, 2017, cited By 5.

[324] Y. Zhao, F. Di Maio, E. Zio, Q. Zhang, C.-L. Dong, and J.-Y. Zhang, "Optimization of a dynamic uncertain causality graph for fault diagnosis in nuclear power plant," *Nuclear Science and Techniques*, vol. 28, no. 3, 2017, cited By 1.

[325] A. Singh and A. Parey, "Gearbox fault diagnosis under fluctuating load conditions with independent angular re-sampling technique, continuous wavelet transform and multilayer perceptron neural network," *IET Science, Measurement and Technology*, vol. 11, no. 2, pp. 220–225, 2017, cited By 7.

[326] A. Joshuva. and V. Sugumaran., "A data driven approach for condition monitoring of wind turbine blade using vibration signals through best-first tree algorithm and functional trees algorithm: A comparative study," *ISA Transactions*, vol. 67, pp. 160–172, 2017, cited By 12.

[327] S. Tyagi and S. Panigrahi, "A dwt and svm based method for rolling element bearing fault diagnosis and its comparison with artificial neural networks," *Journal of Applied and Computational Mechanics*, vol. 3, no. 1, pp. 80–91, 2017, cited By 4.

[328] Y. Jiarula, J. Gao, Z. Gao, J. Hongquan, and W. Rongxi, "Fault mode prediction based on decision tree," 2017, pp. 1729–1733, cited By 2.

[329] Y. Liu, B. He, F. Liu, Y. Zhao, and J. Fang, "Comprehensive recognition of rolling bearing fault pattern and fault degrees based on two-layer similarity in phase space," *Zhendong yu Chongji/Journal of Vibration and Shock*, vol. 36, no. 4, pp. 178–184 and 191, 2017, cited By 2.

[330] Z. Zheng, S. Morando, M.-C. Pera, D. Hissel, L. Larger, R. Martinenghi, and A. Baylon Fuentes, "Brain-inspired computational paradigm dedicated to fault diagnosis of pem fuel cell stack," *International Journal of Hydrogen Energy*, vol. 42, no. 8, pp. 5410–5425, 2017, cited By 8.

[331] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors (Switzerland)*, vol. 17, no. 2, 2017, cited By 50.

[332] B. Saha, B. Patel, and P. Bera, "Dwt and bpnn based fault detection, classification and estimation of location of hvac transmission line," 2017, pp. 174–178, cited By 1.

[333] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mechanical Systems and Signal Processing*, vol. 85, pp. 746–759, 2017, cited By 39.

[334] H. Zhang, Y. Qi, L. Wang, X. Gao, and X. Wang, "Fault detection and diagnosis of chemical process using enhanced keca," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, pp. 61–69, 2017, cited By 8.

[335] M. Asr, M. Ettefagh, R. Hassannejad, and S. Razavi, "Diagnosis of combined faults in rotary machinery by non-naive bayesian approach," *Mechanical Systems and Signal Processing*, vol. 85, pp. 56–70, 2017, cited By 24.

[336] A. Aggarwal, H. Malik, and R. Sharma, "Feature extraction using emd and classification through probabilistic neural network for fault diagnosis of transmission line," 2017, cited By 2.

[337] F. Zhang, T. Zhang, and H. Yu, "A novel rolling bearing fault diagnosis method," 2017, pp. 1148–1152, cited By 0.

[338] Y. Cao, J. Zhang, Y. Li, and L. Zhang, "Aero-engine fault diagnosis based on fuzzy rough set and svm," *Zhendong Ceshi Yu Zhenduan/Journal of Vibration, Measurement and Diagnosis*, vol. 37, no. 1, pp. 169–173, 2017, cited By 1.

[339] M. Islam, J. Kim, S. Khan, and J.-M. Kim, "Reliable bearing fault diagnosis using bayesian inference-based multi-class support vector machines," *Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. EL89–EL95, 2017, cited By 10.

[340] M. Una, Y. Sahin, M. Onat, M. Demetgul, and H. Kucuk, "Fault diagnosis of rolling bearings using data mining techniques and boosting," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 139, no. 2, 2017, cited By 6.

[341] H. Meng, H. Xu, and Q. Tan, "Fault diagnosis based on dynamic svm," 2017, pp. 966–970, cited By 2.

[342] S. Qian, X. Yang, J. Huang, and H. Zhang, "Application of new training method combined with feedforward artificial neural network for rolling bearing fault diagnosis," 2017, cited By 1.

[343] T. Liu, "Fault diagnosis of gearbox by selective ensemble learning based on artificial immune algorithm," 2017, pp. 460–464, cited By 0.

[344] P. Mehta, J. Gaikwad, and J. Kulkarni, "Application of multi-scale fuzzy entropy for roller bearing fault detection and fault classification based on vpmcd," 2017, pp. 256–261, cited By 2.

[345] A. Belaout, F. Krim, and A. Mellit, "Neuro-fuzzy classifier for fault detection and classification in photovoltaic module," 2017, pp. 144–149, cited By 8.

[346] Y. Yan, "Research on electrical equipment's fault diagnosis based on the improved support vector machine and fuzzy clustering," *Chemical Engineering Transactions*, vol. 59, pp. 865–870, 2017, cited By 0.

[347] K. Li, Y. Wu, S. Song, Y. Sun, J. Wang, and Y. Li, "A novel method for spacecraft electrical fault detection based on fcm clustering and wpsvm classification with pca feature extraction," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 231, no. 1, pp. 98–108, 2017, cited By 3.

[348] X. Zhang, D. Jiang, T. Han, N. Wang, W. Yang, and Y. Yang, "Rotating machinery fault diagnosis for imbalanced data based on fast clustering algorithm and support vector machine," *Journal of Sensors*, vol. 2017, 2017, cited By 6.

[349] A. Sun and Y. Che, "A novel compound data classification method and its application in fault diagnosis of rolling bearings," *International Journal of Intelligent Computing and Cybernetics*, vol. 10, no. 1, pp. 80–90, 2017, cited By 1.

[350] X. Yao, S. Li, and J. Hu, "Improving rolling bearing fault diagnosis by ds evidence theory based fusion model," *Journal of Sensors*, vol. 2017, 2017, cited By 7.

[351] H. Malik and R. Sharma, "Emd and ann based intelligent fault diagnosis model for transmission line," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 4, pp. 3043–3050, 2017, cited By 17.

[352] W. Ngui, M. Leong, M. Shapiai, and M. Lim, "Blade fault diagnosis using artificial neural network," *International Journal of Applied Engineering Research*, vol. 12, no. 4, pp. 519–526, 2017, cited By 1.

[353] X. Zhou, "Fault diagnosis of garment and textile spinning frame based on support vector machine," *Academic Journal of Manufacturing Engineering*, vol. 15, no. 4, pp. 61–66, 2017, cited By 0.

[354] J. Shuai, C. Shen, and Z. Zhu, "Adaptive morphological feature extraction and support vector regressive classification for bearing fault diagnosis," *International Journal of Rotating Machinery*, vol. 2017, 2017, cited By 3.

[355] N. Md Nor, M. Hussain, and C. Che Hassan, "Fault diagnosis based on multiscale classification using kernel fisher discriminant analysis and gaussian mixture model and k-nearest neighbor method," *Jurnal Teknologi*, vol. 79, no. 5-3, pp. 89–96, 2017, cited By 1.

[356] T. Lin and K. Yu, “A study of bearing fault diagnosis using deep belief networks,” 2017, cited By 0.

[357] D. Appana, W. Ahmad, and J.-M. Kim, “Speed invariant bearing fault characterization using convolutional neural networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10607 LNAI, pp. 189–198, 2017, cited By 4.

[358] M. Shiblee, S. Yadav, and B. Chandra, “Fault diagnosis of internal combustion engine using empirical mode decomposition and artificial neural networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10363 LNAI, pp. 188–199, 2017, cited By 0.

[359] Q. Zhou, T. Xiong, M. Wang, C. Xiang, and Q. Xu, “Diagnosis and early warning ofwind turbine faults based on cluster analysis theory and modified anfis,” *Energies*, vol. 10, no. 7, 2017, cited By 3.

[360] W. Pan and H. He, “An ordinal rough set model based on fuzzy covering for fault level identification,” *Journal of Intelligent and Fuzzy Systems*, vol. 33, no. 5, pp. 2979–2985, 2017, cited By 0.

[361] D.-T. Hoang and H.-J. Kang, “Convolutional neural network based bearing fault diagnosis,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10362 LNCS, pp. 105–111, 2017, cited By 0.

[362] D. Appana, M. Islam, and J.-M. Kim, “Reliable fault diagnosis of bearings using distance and density similarity on an enhanced k-nn,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10142 LNAI, pp. 193–203, 2017, cited By 5.

[363] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, “Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and i-v characteristics,” *Applied Energy*, vol. 204, pp. 912–931, 2017, cited By 29.

[364] H. Tanatavikorn and Y. Yamashita, “Batch process monitoring based on fuzzy segmentation of multivariate time-series,” *Journal of Chemical Engineering of Japan*, vol. 50, no. 1, pp. 53–63, 2017, cited By 1.

[365] C. Chen, F. Shen, and R. Yan, "Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis," *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, vol. 38, no. 1, pp. 33–40, 2017, cited By 9.

[366] R. Sharma, V. Sugumaran, H. Kumar, and M. Amarnath, "Condition monitoring of roller bearing by k-star classifier and k-nearest neighborhood classifier using sound signal," *SDHM Structural Durability and Health Monitoring*, vol. 12, no. 1, pp. 1–16, 2017, cited By 1.

[367] A. De Araujo Cruz, R. Gomes, F. Belo, and A. Lima Filho, "A hybrid system based on fuzzy logic to failure diagnosis in induction motors," *IEEE Latin America Transactions*, vol. 15, no. 8, pp. 1482–1489, 2017, cited By 2.

[368] N. Gangadhar, K. Vernekar, H. Kumar, and S. Narendranath, "Fault diagnosis of single point cutting tool through discrete wavelet features of vibration signals using decision tree technique and multilayer perceptron," *Journal of Vibrational Engineering and Technologies*, vol. 5, no. 1, pp. 35–44, 2017, cited By 0.

[369] Y. Xiao, N. Kang, Y. Hong, and G. Zhang, "Misalignment fault diagnosis of dfwt based on iemd energy entropy and pso-svm," *Entropy*, vol. 19, no. 1, 2017, cited By 14.

[370] W. Zhang, G. Peng, and C. Li, "Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal," *Smart Innovation, Systems and Technologies*, vol. 64, pp. 77–84, 2017, cited By 0.

[371] C. Wu, T. Chen, and R. Jiang, "Bearing fault diagnosis via kernel matrix construction based support vector machine," *Journal of Vibroengineering*, vol. 19, no. 5, pp. 3445–3461, 2017, cited By 2.

[372] J. Ma, J. Wu, and X. Wang, "Fault diagnosis of rolling bearing based on tkeo-elm," *UPB Scientific Bulletin, Series D: Mechanical Engineering*, vol. 79, no. 4, pp. 215–226, 2017, cited By 0.

[373] S. Wen, J.-S. Wang, and J. Gao, "Fault diagnosis strategy of polymerization kettle equipment based on support vector machine and cuckoo search algorithm," *Engineering Letters*, vol. 25, no. 4, pp. 474–482, 2017, cited By 0.

[374] R. Tiwari, D. Bordoloi, S. Bansal, and S. Sahu, "Multi-class fault diagnosis in gears using machine learning algorithms based on time domain data," *International Journal of COMADEM*, vol. 20, no. 1, pp. 3–13, 2017, cited By 0.

[375] S. Pranesh, S. Abraham, V. Sugumaran, and M. Amarnath, "Fault diagnosis of helical gearbox using acoustic signal and wavelets," vol. 197, no. 1, 2017, cited By 0.

[376] S. Jan, Y.-D. Lee, J. Shin, and I. Koo, "Sensor fault classification based on support vector machine and statistical time-domain features," *IEEE Access*, vol. 5, pp. 8682–8690, 2017, cited By 19.

[377] Y. Chang, F. Jiang, Z. Zhu, and W. Li, "Fault diagnosis of rotating machinery based on time-frequency decomposition and envelope spectrum analysis," *Journal of Vibroengineering*, vol. 19, no. 2, pp. 943–954, 2017, cited By 1.

[378] D. Yang, H. Mu, Z. Xu, Z. Wang, C. Yi, and C. Liu, "Based on soft competition art neural network ensemble and its application to the fault diagnosis of bearing," *Mathematical Problems in Engineering*, vol. 2017, 2017, cited By 2.

[379] H. Wang, L. Li, X. Gong, and W. Du, "Blind source separation of rolling element bearing single channel compound fault based on shift invariant sparse coding," *Journal of Vibroengineering*, vol. 19, no. 3, pp. 1809–1822, 2017, cited By 2.

[380] K. Jafarian and M. Mobin, "Misfire fault detection in the internal combustion engine using the artificial neural networks (anns)," 2017, pp. 926–932, cited By 0.

[381] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14 347–14 357, 2017, cited By 13.

[382] L. Xie and J. Wei, "Fault diagnosis of power quality and disturbance classification based on kpca - fda method," vol. 2017-December, 2017, cited By 0.

[383] L.-S. Xia, Y.-Y. Yang, and H.-J. Fang, "Fault diagnosis performance improvement for chemical process based on easyensemble method," *Kongzhi Lilun Yu Yingyong/Control Theory and Applications*, vol. 34, no. 1, pp. 49–53, 2017, cited By 1.

[384] J. Ma, J. Wu, and X. Wang, "Fault diagnosis method of check valve based on multikernel cost-sensitive extreme learning machine," *Complexity*, vol. 2017, 2017, cited By 1.

[385] J. Kowalski, B. Krawczyk, and M. Woniak, "Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble," *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 134–141, 2017, cited By 7.

[386] Z. Wang, Z. Wang, S. He, X. Gu, and Z. Yan, "Fault detection and diagnosis of chillers using bayesian network merged distance rejection and multi-source non-sensor information," *Applied Energy*, vol. 188, pp. 200–214, 2017, cited By 15.

[387] X. Qin, Q. Li, X. Dong, and S. Lv, "The fault diagnosis of rolling bearing based on ensemble empirical mode decomposition and random forest," *Shock and Vibration*, vol. 2017, 2017, cited By 7.

[388] W. You, C. Shen, X. Guo, X. Jiang, J. Shi, and Z. Zhu, "A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery," *Advances in Mechanical Engineering*, vol. 9, no. 6, 2017, cited By 2.

[389] Q. Tong, J. Cao, B. Han, X. Zhang, Z. Nie, J. Wang, Y. Lin, and W. Zhang, "A fault diagnosis approach for rolling element bearings based on rsgwpt-lcd bilayer screening and extreme learning machine," *IEEE Access*, vol. 5, pp. 5515–5530, 2017, cited By 8.

[390] G. Song and C. He, "Research on feature extraction of mechanical fault based on orthogonal local fisher discriminant analysis," *Academic Journal of Manufacturing Engineering*, vol. 15, no. 2, pp. 81–86, 2017, cited By 0.

[391] W. Abed, S. Sharma, R. Sutton, and A. Khan, "An unmanned marine vehicle thruster fault diagnosis scheme based on ofnda," *Journal of Marine Engineering and Technology*, vol. 16, no. 1, pp. 37–44, 2017, cited By 0.

[392] D. MacK, G. Biswas, X. Koutsoukos, and D. Mylaraswamy, "Learning bayesian network structures to augment aircraft diagnostic reference models," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 358–369, 2017, cited By 9.

[393] Y. Yang and D. Jiang, "Casing vibration fault diagnosis based on variational mode decomposition, local linear embedding, and support vector machine," *Shock and Vibration*, vol. 2017, 2017, cited By 5.

[394] N. Zhao, H. Zheng, L. Yang, and Z. Wang, "A fault diagnosis approach for rolling element bearing based on stransform and artificial neural network," vol. 6, 2017, cited By 2.

[395] Y. Xie and T. Zhang, "Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition," *Shock and Vibration*, vol. 2017, 2017, cited By 4.

[396] H. Yuan, J. Chen, and G. Dong, "Bearing fault diagnosis based on improved locality-constrained linear coding and adaptive pso-optimized svm," *Mathematical Problems in Engineering*, vol. 2017, 2017, cited By 2.

[397] W. Deng, R. Yao, M. Sun, H. Zhao, Y. Luo, and C. Dong, "Study on a novel fault diagnosis method based on integrating emd, fuzzy entropy, improved pso and svm," *Journal of Vibroengineering*, vol. 19, no. 4, pp. 2562–2577, 2017, cited By 1.

[398] C.-Z. Hu, Q. Yang, M.-Y. Huang, and W.-J. Yan, "Sparse component analysis-based underdetermined blind source separation for bearing fault feature extraction in wind turbine gearbox," *IET Renewable Power Generation*, vol. 11, no. 3, pp. 330–337, 2017, cited By 11.

[399] H. Cui, Y. Qiao, Y. Yin, and M. Hong, "An investigation on early bearing fault diagnosis based on wavelet transform and sparse component analysis," *Structural Health Monitoring*, vol. 16, no. 1, pp. 39–49, 2017, cited By 5.

[400] S. Altaf, M. Soomro, and M. Mehmood, "Fault diagnosis and detection in industrial motor network environment using knowledge-level modelling technique," *Modelling and Simulation in Engineering*, vol. 2017, 2017, cited By 0.

[401] J. W. Liu, L. P. Cui, and X. L. Luo, "MCR SVM classifier with group sparsity," *Optik*, 2016.

[402] N. A. T. Nguyen, H. J. Yang, and S. Kim, "Hidden discriminative features extraction for supervised high-order time series modeling," *Computers in Biology and Medicine*, 2016.

[403] F. Liang, "Use of SVD-based probit transformation in clustering gene expression profiles," *Computational Statistics and Data Analysis*, 2007.

[404] Z. Wang, L.-M. Liu, N.-Y. Deng, L. Bai, Y.-H. Shao, and C.-N. Li, "MBLDA: A novel multiple between-class linear discriminant analysis," *Information Sciences*, 2016.

[405] C. H. Park and H. Park, "A comparison of generalized linear discriminant analysis algorithms," *Pattern Recognition*, 2008.

[406] A. J. Ferreira and M. A. Figueiredo, “An unsupervised approach to feature discretization and selection,” in *Pattern Recognition*, 2012.

[407] H. Zhang, Y. Zhang, and T. S. Huang, “Simultaneous discriminative projection and dictionary learning for sparse representation based classification,” *Pattern Recognition*, 2013.

[408] X. Zhu, Z. Huang, Y. Yang, H. Tao Shen, C. Xu, and J. Luo, “Self-taught dimensionality reduction on the high-dimensional small-sized data,” *Pattern Recognition*, 2013.

[409] B. Raducanu and F. Dornaika, “Embedding new observations via sparse-coding for non-linear manifold learning,” *Pattern Recognition*, 2014.

[410] J. Wen, Z. Lai, Y. Zhan, and J. Cui, “The L2,1-norm-based unsupervised optimal feature selection with applications to action recognition,” *Pattern Recognition*, 2016.

[411] X. Hu, Z. Yang, and L. Jing, “An incremental dimensionality reduction method on discriminant information for pattern classification,” *Pattern Recognition Letters*, 2009.

[412] C. J. Veenman and A. Bolck, “A sparse nearest mean classifier for high dimensional multi-class problems,” *Pattern Recognition Letters*, 2011.

[413] C. Y. Lu and D. S. Huang, “Optimized projections for sparse representation based classification,” *Neurocomputing*, 2013.

[414] Y. Cui, Z. Jin, and J. Jiang, “A novel supervised feature extraction and classification fusion algorithm for land cover recognition of the off-land scenario,” *Neurocomputing*, 2014.

[415] S. Du, Y. Ma, S. Li, and Y. Ma, “Robust unsupervised feature selection via matrix factorization,” *Neurocomputing*, 2017.

[416] Y. Cui, J. Jiang, Z. Lai, W. K. Wong, and Z. Hu, “An integrated optimisation algorithm for feature extraction, dictionary learning and classification,” *Neurocomputing*, 2018.

[417] S. Suresh, S. Saraswathi, and N. Sundararajan, “Performance enhancement of extreme learning machine for multi-category sparse data classification problems,” *Engineering Applications of Artificial Intelligence*, 2010.

[418] V. Pappu, O. P. Panagopoulos, P. Xanthopoulos, and P. M. Pardalos, "Sparse Proximal Support Vector Machines for feature selection in high dimensional datasets," *Expert Systems with Applications*, 2015.

[419] M. Zhao, Z. Zhang, T. W. Chow, and B. Li, "Soft label based Linear Discriminant Analysis for image recognition and retrieval," *Computer Vision and Image Understanding*, 2014.

[420] X. Cai, G. Wen, J. Wei, J. Li, and Z. Yu, "Perceptual relativity-based semi-supervised dimensionality reduction algorithm," *Applied Soft Computing Journal*, 2014.

[421] K. Mallick and S. Bhattacharyya, "Uncorrelated Local Maximum Margin Criterion: An Efficient Dimensionality Reduction Method for Text Classification," *Procedia Technology*, 2012.

[422] R. A. Weekley, R. K. Goodrich, and L. B. Cornman, "An algorithm for classification and outlier detection of time-series data," *Journal of Atmospheric and Oceanic Technology*, 2010.

[423] R. Lotlikar and R. Kothari, "Bayes-optimality motivated linear and multilayered perceptron-based dimensionality reduction," *IEEE Transactions on Neural Networks*, 2000.

[424] I. Dhillon and Y. Guan, "Information theoretic clustering of sparse cooccurrence data," 2004.

[425] J. Yang, D. Zhang, J.-Y. Yang, and B. Niu, "Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics."

[426] Wen-Hui Yang, Dao-Qing Dai, and Hong Yan, "Feature Extraction and Uncorrelated Discriminant Analysis for High-Dimensional Data," *IEEE Transactions on Knowledge and Data Engineering*, 2008.

[427] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Transactions on Audio, Speech and Language Processing*, 2010.

[428] J. Kim and C. Scott, "L Kernel classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[429] J. Luo, L. Jiao, F. Liu, S. Yang, and W. Ma, "A Pareto-Based Sparse Subspace Learning Framework," 2018.

[430] A. Majumdar and R. K. Ward, "Robust classifiers for data reduced via random projections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010.

[431] G. San Martin, E. López Droguett, V. Meruane, and M. das Chagas Moura, "Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis," 2018.

[432] Y. Hu, K. Zhao, and H. Lian, "Bayesian Additive Machine: classification with a semiparametric discriminant function," *Journal of Statistical Computation and Simulation*, 2016.

[433] H. Raeisi Shahraki, S. Pourahmad, and N. Zare, "K Important Neighbors: A Novel Approach to Binary Classification in High Dimensional Data ," *BioMed Research International*, 2017.

[434] M. A. Borgi, T. P. Nguyen, D. Labate, and C. B. Amar, "Statistical binary patterns and post-competitive representation for pattern recognition," *International Journal of Machine Learning and Cybernetics*, 2018.

[435] D. Komura, H. Nakamura, S. Tsutsumi, H. Aburatani, and S. Ihara, "Multidimensional support vector machines for visualization of gene expression data," *Bioinformatics*, 2005.

[436] A. R. Linero and Y. Yang, "Bayesian regression tree ensembles that adapt to smoothness and sparsity," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 2018.

[437] Y. Li and A. Ngom, "Nonnegative least-squares methods for the classification of high-dimensional biological data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2013.

[438] S. Simmuteit, F. M. Schleif, T. Villmann, and B. Hammer, "Evolving trees for the retrieval of mass spectrometry-based bacteria fingerprints," *Knowledge and Information Systems*, 2010.

[439] T. S. Tian, R. R. Wilcox, and G. M. James, "Data reduction in classification: A simulated annealing based projection method," *Statistical Analysis and Data Mining*, 2010.

[440] F. Yao, J. Coquery, and K. A. Lê Cao, "Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets," *BMC Bioinformatics*, 2012.

[441] Y. Zhan, J. Yin, and X. Liu, "Nonlinear discriminant clustering based on spectral regularization," *Neural Computing and Applications*, 2013.

[442] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognition*, 2008.

[443] W. Zhang, Z. Lin, and X. Tang, "Tensor linear Laplacian discrimination (TLLD) for feature extraction," *Pattern Recognition*, 2009.

[444] S. Maldonado and R. Weber, "A wrapper method for feature selection using Support Vector Machines," *Information Sciences*, 2009.

[445] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Extracting the optimal dimensionality for local tensor discriminant analysis," *Pattern Recognition*, 2009.

[446] M. Zhao, Z. Zhang, and T. W. Chow, "Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction," *Pattern Recognition*, 2012.

[447] X. Zhu, Z. Huang, H. Tao Shen, J. Cheng, and C. Xu, "Dimensionality reduction by Mixed Kernel Canonical Correlation Analysis," *Pattern Recognition*, 2012.

[448] H. Xia, J. Zhuang, and D. Yu, "Novel soft subspace clustering with multi-objective evolutionary approach for high-dimensional data," *Pattern Recognition*, 2013.

[449] "faultdiagnosis."

[450] M. B. Blaschko, J. A. Shelton, A. Bartels, C. H. Lampert, and A. Gretton, "Semi-supervised kernel canonical correlation analysis with application to human fMRI," *Pattern Recognition Letters*, 2011.

[451] W. Sun, J. Chen, and J. Li, "Decision tree and PCA-based fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, 2007.

[452] V. Meruane and A. Ortiz-Bernardin, "Structural damage assessment using linear approximation with maximum entropy and transmissibility data," *Mechanical Systems and Signal Processing*, 2015.

[453] Z. Zhang, M. Zhao, and T. W. Chow, "Marginal semi-supervised sub-manifold projections with informative constraints for dimensionality reduction and recognition," *Neural Networks*, 2012.

[454] C. L. Huang and J. F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing Journal*, 2008.

[455] H. Lee, R. Raina, A. Teichman, and A. Y. Ng, "Exponential Family Sparse Coding with Applications to Self-taught Learning," Tech. Rep.

[456] X. He, D. Cai, and P. Niyogi, "Laplacian Score for Feature Selection," Tech. Rep.

[457] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and Robust Feature Selection via Joint 2,1-Norms Minimization," Tech. Rep.

[458] J. Mao and A. K. Jain, "Artificial Neural Networks for Feature Extraction and Multivariate Data Projection," Tech. Rep. 2, 1995.

[459] L. Shuang and L. Meng, "Bearing fault diagnosis based on PCA and SVM," in *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, ICMA 2007*, 2007.

[460] S. J. Pan and Q. Yang, "A survey on transfer learning," 2010.

[461] F. Nie, D. Xu, I. W. H. Tsang, and C. Zhang, "Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction," *IEEE Transactions on Image Processing*, 2010.

[462] B. Bagheri, H. Ahmadi, and R. Labbafi, "Application of data mining and feature extraction on intelligent fault diagnosis by artificial neural network and k-nearest neighbor," in *19th International Conference on Electrical Machines, ICEM 2010*, 2010.

[463] Y. Li and A. Ngom, "Non-negative matrix and tensor factorization based classification of clinical microarray gene expression data," in *Proceedings - 2010 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2010*, 2010.

[464] H. Liu, L. Li, and J. Ma, "Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals," *Shock and Vibration*, 2016.

[465] B. Xue, M. Zhang, and W. N. Browne, "New Fitness Functions in Binary Particle Swarm Optimisation for Feature Selection," Tech. Rep.

[466] X. Liu, T. Guo, L. He, and X. Yang, "A Low-Rank Approximation-Based Transductive Support Tensor Machine for Semisupervised Classification," *IEEE Transactions on Image Processing*, 2015.

[467] J. Tian, M. Li, F. Chen, and N. Feng, "Learning Subspace-Based RBFNN Using Coevolutionary Algorithm for Complex Classification Tasks," *IEEE Transactions on Neural Networks and Learning Systems*, 2016.

[468] M. Medvedovic, K. Y. Yeung, and R. E. Bumgarner, "Bayesian mixture model based clustering of replicated microarray data," *Bioinformatics*, 2004.

[469] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1, pp. 389–422, Jan 2002. [Online]. Available: https://doi.org/10.1023/A:1012487302797

[470] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions," Tech. Rep.

[471] H. Cevikalp, F. Jurie, B. Triggs, and R. Polikar, "Margin-based discriminant dimensionality reduction for visual recognition," in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[472] L. Zhu, L. Cao, and J. Yang, "Multiobjective evolutionary algorithm-based soft subspace clustering," in *2012 IEEE Congress on Evolutionary Computation, CEC 2012*, 2012.

[473] M. Zhao, Z. Zhang, and T. W. Chow, "On the theoretical and computational analysis between SDA and lap-LDA," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2012.

[474] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught Learning: Transfer Learning from Unlabeled Data," Tech. Rep.

[475] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006. [Online]. Available: http://science.sciencemag.org/content/313/5786/504

[476] H. T. Shen, X. Zhou, and A. Zhou, "An adaptive and dynamic dimensionality reduction method for high-dimensional indexing," *VLDB Journal*, 2007.

[477] M. Sugiyama, T. Idé, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Machine Learning*, 2010.

[478] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Mining and Knowledge Discovery*, 2011.

[479] X. L. Zhang, W. Chen, B. J. Wang, and X. F. Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization," *Neurocomputing*, 2015.

[480] X. Zhang, B. Wang, and X. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowledge-Based Systems*, 2015.

[481] C. Lu, Z. Y. Wang, W. L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, 2017.

[482] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, 2016.

[483] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, 2016.

[484] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Systems with Applications*, 2006.

[485] Z. Li, H. Fang, and M. Huang, "Diversified learning for continuous hidden Markov models with application to fault diagnosis," *Expert Systems with Applications*, 2015.

[486] M. Szummer and T. Jaakkola, "Partially labeled classification with Markov random walks," Tech. Rep.

[487] W. Yan, "Application of Random Forest to Aircraft Engine Fault Diagnosis," 2007.

[488] Y. Yang and W. Tang, "Study of remote bearing fault diagnosis based on BP Neural Network combination," in *Proceedings - 2011 7th International Conference on Natural Computation, ICNC 2011*, 2011.

[489] F. Zakaria, D. Johari, and I. Musirin, "Optimized artificial neural network for the detection of incipient faults in power transformer," in *2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*, March 2014, pp. 635–640.

[490] B. S. Yang, X. Di, and T. Han, "Random forests classifier for machine fault diagnosis," *Journal of Mechanical Science and Technology*, 2008.

[491] F. Comitani, "fcomitani/simpsom: v1.3.4," Apr. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2621560

[492] albertbup, "A python implementation of deep belief networks built upon numpy and tensorflow with scikit-learn compatibility," 2017. [Online]. Available: https://github.com/albertbup/deep-belief-network

[493] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[494] J. Warner, "scikit-fuzzy," https://github.com/scikit-fuzzy/scikit-fuzzy, 2019.

[495] T. E. Oliphant, *A guide to NumPy*, 2006, vol. 1.

[496] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–, [Online; accessed 2019-06-25]. [Online]. Available: http://www.scipy.org/

[497] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[498] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, S. van der Walt and J. Millman, Eds., 2010, pp. 51 – 56.

Appendix A

Fault Diagnosis

This table provides an overview of the relevant fault diagnosis literature, published since 2016. In this overview supervised papers are indicated as s, unsupervised as u and semi-supervised as i.

Technique(s)	s/u/i	Field	Type	Types of faults	Ref.
CNN	s	Bearings	Vibration signal	Predefined faults	[73]
Hilbert-Huang Transform (HTT), CNN	s	Bearings	Vibration signal	Predefined faults	[74]
Fuzzy clustering	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[75]
Extreme Learning Machine (ELM)	u	Analog circuits	Voltages	Predefined faults	[76]
SDA	s	Bearings	Vibration signal	Combined faults	[77]
Entropy, RF	s	Bearings	Vibration signal	Predefined faults	[78]
Fuzzy clustering	u	Actuators	Temperature, pressure, flow, density	Dynamically discovering new faults	[79]
Multi-Kernel Learning (MKL)	s	Pumps	Pressure, temperature, flow, current, voltage, vibration	Predefined faults	[80]

Generative Adversarial Networks (GAN)	u	Bearings	Vibration signal	Predefined faults	[81]
selective ensemble learning, PSO	s	Rotary machinery	Vibration signal	Predefined faults	[82]
Support Vector Nerual Network (SVNN)	s	Analog circuits	Voltages	Predefined faults	[83]
k-NN	u	Motors	Vibration signal	Predefined faults	[84]
Radial Based Function Neural Network (RBFNN)	u	synchronous condenser	Voltages	Predefined faults	[85]
SVM, PSO	s	Bearings	Vibration signal	Predefined faults	[86]
GAN	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[87]
k-NN	s	Bearings	Vibration signal	Predefined faults	[88]
empirical mode decomposition (EMD), genetic neural network adaptive boosting (GNN-AdaBoost)	s	Bearings	Vibration signal	Predefined faults	[89]
CNN	s	Rotary machinery	Vibration signal	Predefined faults	[90]
CNN	s	Gearboxes	Vibration signal	Predefined faults	[91]
DBN	s	Turbine	Vibration signal	Predefined faults	[92]
Transfer Learning (TL), CNN	s	Bearings	Vibration signal	Predefined faults	[37]
LSTM	s	Turbine	Displacement, acceleration, wind speed, rotor speed, torque, wind power	Predefined faults	[93]
k-NN	u	HVAC	Temperature	Predefined faults	[94]

Decision tree	s	Motors	Categorical data	Predefined faults	[95]
LSTM	s	Gearboxes	Vibration signal	Predefined faults	[96]
SVM	s	Bearings	Vibration signal	Predefined faults	[97]
non-naive Bayesian classifier (NNBC)	s	Bearings	Vibration signal	Predefined faults	[98]
SVM, PCA	s	Turbine	Power output, speed, torque, pitch angles,acceleration	Predefined faults	[99]
SVM, PCA	s	Bearings	Vibration signal	Predefined faults	[100]
Support Vector Data Description	i	Circuit breaker	Voltages	Predefined faults (clustering)	[101]
Fuzzy clustering	u	Three-tank system	Pump output	Unlabelled faults (clustering)	[102]
SVM	s	Distillation	Feed composition, temperature	Predefined faults	[103]
DBN, Fuzzy clustering	i	Rotary machinery	Vibration signal	Unlabelled faults (clustering)	[22]
DE, Stacked Sparse Autoencoders	s	Bearings	Vibration signal	Predefined faults	[104]
PCA, Multilayer Perceptron NN	s	Motors	Voltages	Predefined faults	[105]
Nave Bayes	s	Motors	Voltages	Predefined faults	[106]
DBN	s	Bearings	Vibration signal	Predefined faults	[107]
Deep CNN	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[108]
GAN, SDA	s	Gearboxes	Vibration signal	Predefined faults	[109]
SDA	i	Rotary machinery	Vibration signal	Unlabelled faults (clustering)	[110]
BPNN	s	Sensors	Level, flow	Predefined faults	[111]

Relevance Vector Machine (RVM), PSO	s	Bearings	Vibration signal	Predefined faults	[112]
GAN	s	Rotary machinery	Vibration signal	Predefined faults	[113]
SVM	s	Bearings	Vibration signal	Predefined faults	[114]
CNN	s	Bearings	Vibration signal	Predefined faults	[115]
DBN	S	Compressor	Pressure, current, vibration	Predefined faults	[116]
BPNN	s	Mechanical faults	Temperature, vibration	Predefined faults	[117]
Fuzzy clustering, PSO	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[118]
Decision tree	s	Photovoltaic system	Current, voltage, temperature	Predefined faults	[119]
k-NN	s	Photovoltaic system	Current, voltage	Predefined faults	[120]
BPNN	s	Rotor bars	Current	Predefined faults	[121]
SVM	s	Bearings	Vibration signal	Predefined faults	[122]
k-NN, Nave bayes	s	Transformers	Dissolved Gas Analysis	Predefined faults	[123]
SVM	s	Power network	Current	Predefined faults	[124]
Subspace clustering	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[56]
SVM	s	Rotary machinery	Torque	Predefined faults (clustering)	[125]
Bayesian Extreme Learning	s	Rotary machinery	Air ratio, sound, voltage	Predefined faults	[126]
SVM	s	Bearings	Vibration signal	Predefined faults	[127]

SVM	s	Motors	Current	Predefined faults	[128]
Stacked AutoEncoder	s	Bearings	Vibration signal	Predefined faults	[129]
Radial Basis Function Neural Network (RBF), Generalized Regression Neural Network (GRNN)	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[130]
Stacked Sparse AutoEncoder (SSAE)	s	Chemical processes	Temperature, pressure	Predefined faults	[131]
k-Means	u	Compressor	Pressure, current, vibration	Unlabelled faults (clustering)	[132]
SDA	s	Bearings	Vibration signal	Predefined faults	[133]
SDA, BPNN	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[134]
Sparse AutoEncode (SAE)	s	Turbine	Images	Predefined faults	[135]
SVM	s	Inverted pendulum	Pole analysis	Predefined faults	[136]
HMM, k-Means	s	Fuel cell	Pressure, temperature, voltage	Predefined faults	[50]
CNN	s	Manufacturing	Images	Predefined faults	[137]
Sparse Deep Stacking Network (SDSN)	s	Motors	Vibration signal	Predefined faults	[138]
k-NN	i	Bearings	Vibration signal	Predefined faults	[139]
SVM	s	Inverter	Voltages	Predefined faults	[140]
DNN	s	Rotary machinery	Vibration signal	Predefined faults	[141]

hierarchical dimension reduction (HDR) classifier	s	Transformers	Dissolved Gas Analysis	Predefined faults	[142]
SVM	s	Motors	Vibration signal	Predefined faults	[143]
RNN	s	Grinding	Temperature, pressure, sound, vibration	Binary (fault/fault free)	[144]
Fuzzy Cognitive Map	s	Motors	Vibration, current	Predefined faults	[145]
PCA, Fuzzy clustering	u	Manufacturing	PH, temperature, feed, pressure	Predefined faults	[146]
PCA	u	HVAC	Pressure, temperature, flow rate	Binary (fault/fault free)	[147]
LSTM	s	Bearings	Vibration signal	Predefined faults	[148]
Fuzzy c-means clustering	u	Transformers	Dissolved Gas Analysis	Unlabelled faults (clustering)	[149]
Dominant Set Clustering	u	General	Images	Unlabelled faults (clustering)	[23]
GA, SVM	s	Power systems	Current, voltage	Predefined faults	[21]
PSO, SOM, Learning Vector Quantization Neural Network	s	Turbine	Vibration signal	Predefined faults	[150]
SVM	s	Power network	Voltage, current	Predefined faults	[151]
RF, SVM	s	Bearings	Vibration signal	Predefined faults	[54]
SVM	s	HVAC	Temperature, pressure, flow rate	Predefined faults	[152]
HMM, SOM	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[153]
Fushion (MLP-NN, RBF-NN, Decision trees, k-NN)	s	Turbine	Speed, torque, power, accelleration	Predefined faults	[154]

LSTM	s	General	PH, temperature, feed, pressure	Predefined faults	[155]
SVM	s	Rotor bars	Vibration signal	Predefined faults	[156]
Locality Preserving Projection	s	General	Vibration signal	Predefined faults	[157]
Exponential Discriminant Analysis (EDA)	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[158]
GMM	i/u	General	Current, acceleration, torque, speed	Unlabelled faults (clustering)	[159]
SVM	s	Transformers	Dissolved Gas Analysis	Predefined faults	[160]
Fuzzy kernel clustering	u	Rotary machinery	Vibration signal	Predefined faults	[161]
Fuzzy c-means clustering	u	General	Images	Unlabelled faults (clustering)	[162]
Fuzzy c-means clustering	u	Bearings	Vibration signal	Predefined faults	[163]
DNN	s	Bearings	Vibration signal	Predefined faults	[164]
DBN, PCA, Fuzzy clustering	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[42]
PSO	s	Engines	Vibration signal	Predefined faults	[165]
RNN	s	Gearboxes	Vibration signal	Predefined faults	[166]
Elman-NN, AdaBoost	s	Bearings	Vibration signal	Predefined faults	[167]
SVM	s	Bearings	Vibration signal	Predefined faults	[168]
Fuzzy inference	s	Photovoltaic system	Temperature, voltage, current, wind speed	Predefined faults	[169]
RNN	s	Bearings	Vibration signal	Predefined faults	[170]
Wavelet transform, CNN	s	Bearings	Vibration signal	Predefined faults	[171]
CNN	s	Bearings	Vibration signal	Predefined faults	[172]

CNN	s	Rotary machinery	Vibration signal	Predefined faults	[173]
SAE, Softmax classifier	s	Hydraulic equipment	Temperature, pressure, shock, revolution	Predefined faults	[174]
GA, k-means	u	Engines	Vibration signal	Predefined faults (clustering)	[175]
SOM, PCA	u	Bearings	Vibration signal	Predefined faults (clustering)	[176]
SVDD, SVM	s	Rotor bars	Current	Predefined faults	[177]
PCA, SVM, PSO	s	Manufacturing	% of aggregates passing the sieve	Predefined faults	[178]
PCA	u	Bearings	Vibration signal	Binary (fault/fault free)	[179]
SVM	s	Bearings	Vibration signal	Predefined faults	[180]
SVM	s	Bearings	Vibration signal	Predefined faults	[181]
SVM	s	Bearings	Vibration signal	Predefined faults	[182]
Logical Analysis of Data (LAD)	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[183]
SDA	s	Bearings	Vibration signal	Predefined faults	[184]
SVM	s	Sensors	Temperature oxygen	Predefined faults	[185]
SOM	u	Turbine	Temperature, speed, power	Predefined faults (clustering)	[48]
Double Paralell feedforward Extreme Learning Machine (DP-ELM)	s	Bearings	Vibration signal	Predefined faults	[186]
Relevance Vector Machine (RVM)	s	Turbine	Vibration signal	Predefined faults	[187]
CNN, HMM	s	Bearings	Vibration signal	Predefined faults	[36]
Rough set theory Neural Network	s	Gearboxes	Vibration signal	Predefined faults	[188]

RBFNN, Probablistic Neural Network (PNN)	s	Motors	Vibration signal	Predefined faults	[189]
SVM	s	Bearings	Vibration signal	Predefined faults	[9]
AdaBoost, RVM	s	Engines	Vibration signal	Predefined faults	[190]
SVM	s	Motors	Vibration, current	Predefined faults	[7]
CNN, LSTM	s	Bearings	Vibration signal	Predefined faults	[34]
CNN	s	Bearings	Vibration signal	Predefined faults	[191]
k-Means, 1-NN	i	Gearboxes	Vibration signal	Predefined faults	[192]
SDA	s	Rotary machinery	Vibration signal	Predefined faults	[193]
RF	s	Motors	Vibration signal	Predefined faults	[194]
SVM	s	Turbine	Vibration signal	Predefined faults	[195]
Native Bayes, PSO	s	Gearboxes	Vibration signal	Predefined faults	[196]
SAE, CNN, DBN	s	Bearings	Vibration signal	Predefined faults	[35]
SAE, PSO, SVM	s	Bearings	Vibration signal	Predefined faults	[44]
Fuzzy entropy, SOM, PCA	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[197]
ELM	s	Bearings	Vibration signal	Predefined faults	[198]
ANN, Nave Bayes	s	Manufacturing	Vibration signal	Predefined faults	[199]
PCA, SVM	s	Actuators	Torque, voltage, current	Predefined faults	[200]
ANN, SVM, LS-SVM	s	Bearings	Vibration signal	Predefined faults	[201]
Neighborhood Preserving Embedding	s	Bearings	Vibration signal	Predefined faults	[202]
k-NN	s	Engines	Vibration signal	Predefined faults	[203]

Transductive Support Vector Machine (TSVM)	i	Gearboxes	Vibration signal	Predefined faults	[30]
SVM	s	Bearings	Vibration signal	Predefined faults	[142]
CNN	s	Gearboxes	Vibration signal	Predefined faults	[204]
LAD	s	General	PH, temperature, feed, pressure	Predefined faults	[205]
Nave bayes	s	Gearboxes	Vibration signal	Predefined faults	[206]
ANN, DNN	s	Gearboxes	Vibration signal	Predefined faults	[207]
RF, k-NN	s	Rotary machinery	Vibration signal	Predefined faults	[208]
TL	i	Rotary machinery	Vibration signal	Predefined faults	[209]
k-NN	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[210]
SVM	s	Bearings	Vibration signal	Predefined faults	[211]
RF	s	Spacecraft	Pitch, position, electronical data, etc. (1000 features)	Predefined faults	[212]
SVM	s	Power network	Current	Predefined faults	[213]
k-NN	s	Bearings	Vibration signal	Predefined faults	[214]
Rought set	u	Power network	Current, voltage, switch	Unlabelled faults (clustering)	[215]
SVM	s	Gearboxes	Acceleration, torque, speed, temperature, vibration, sound	Predefined faults	[216]
SVM	s	Turbine	Temperature, pitch, rotor	Predefined faults	[217]
Fuzzy kernel, ELM	s	Bearings	Vibration signal	Predefined faults	[218]

PCA, SVM	s	Cloud computing	Images	Predefined faults	[219]
Decision trees	s	Gearboxes	Vibration signal	Predefined faults	[220]
Deep transfer learning	i	Power systems	Dissolved Gas Analysis, temperature, moisture, density	Predefined faults	[38]
Decision trees	s	Turbine	Vibration signal	Predefined faults	[221]
Fuzzy Petri Net	s	Motors	Vibration signal	Predefined faults	[222]
BPNN	s	Trains	Vibration signal	Predefined faults	[223]
MLPNN	s	Rotor bars	Vibration signal	Predefined faults	[224]
Locality preserving clustering	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[225]
PSO	s	Cell network	Temperature	Predefined faults	[226]
ELM	s	Bearings	Vibration signal	Predefined faults	[227]
HMM	s	Rotary machinery	Vibration signal	Predefined faults (clustering)	[228]
DBN	s	Bearings	Vibration signal	Predefined faults	[39]
ANN	s	Photovoltaic system	Current, voltage	Predefined faults	[229]
Fuzzy classifier	s	Photovoltaic system	Current, voltage	Predefined faults	[230]
SAE	s	Bearings	Vibration signal	Predefined faults	[231]
Linear Vector Quantization (LVQ) NN	s	Analog circuits	Voltage	Predefined faults	[232]
1-nn, k-means	i	Gearboxes	Vibration signal	Predefined faults	[233]
PSO	s	Bearings	Vibration signal	Predefined faults	[234]
RNN	s	Gearboxes	Vibration signal	Predefined faults	[235]

SVM	s	Bearings	Vibration signal	Predefined faults	[236]
RVM	s	General	Feed, flow, pressure, temperature	Predefined faults	[237]
DBN	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[40]
dag-SVM	s	Analog circuits	Current, voltage	Predefined faults	[238]
ELM	s	Aircraft	Speed, temperature, pressure, fuel flow	Predefined faults	[239]
CNN	s	Transformers	Current	Predefined faults	[240]
SVM	s	Pumps	Multiple (pressure, temperature, flow, current, voltage, vibration)	Predefined faults	[241]
Bayes classifier, LDA, k-NN	s	Motors	Current	Predefined faults	[242]
RBFNN	s	Gearboxes	Vibration signal	Predefined faults	[243]
SVM	s	Bearings	Vibration signal	Predefined faults	[244]
Stacked Sparse Denoising AutoEncoders (SSDA)	s	Bearings	Vibration signal	Predefined faults	[245]
Manifold learning	i	Bearings	Vibration signal	Predefined faults	[246]
Deep confidence nets	s	Transformers	Images	Predefined faults	[247]
DBN	s	Motors	Vibration signal	Predefined faults	[41]
Rough set, SVM	s	Motors	Current	Predefined faults	[248]
RBFNN	s	Transformers	Current, voltage	Predefined faults	[249]
Weightless Neural Networks (WNN)	s	General	Vibration signal	Predefined faults	[250]
Paired RVM	s	General	Vibration signal	Predefined faults	[251]

Multi-scale pssiblistic clustering	u	Bearings	Vibration signal	Predefined faults (clustering)	[252]
Genetic Programming	u	General	Artificial data	Unlabelled faults (clustering)	[253]
DHMM, BPNN	s	Gearboxes	Vibration signal	Predefined faults	[33]
ELM	s	Bearings	Vibration signal	Predefined faults	[254]
K-SVD	s	Bearings	Vibration signal	Predefined faults	[255]
CNN, DBN	s	Mechanical faults	Vibration signal	Predefined faults	[256]
Fuzzy Q-Learning (reinforcement learning)	i	Turbine	Current, voltage	Predefined faults	[257]
ANN	s	Photovoltaic system	Current, voltage	Predefined faults	[258]
Logistic regression classifier (LASSO)	s	Bearings	Vibration signal	Predefined faults	[259]
SAE, softmax classifier	s	General	Voltage	Predefined faults	[45]
SVM	s	Crawling-rolling robot	gyro, accelerometer, magnetometer	Binary (fault/fault free)	[260]
Decision Tree	s	Bearings	Vibration signal	Predefined faults	[261]
ELM	s	Bearings	Vibration signal	Predefined faults	[262]
SVM	s	Gearboxes	Vibration signal	Predefined faults	[263]
Bayes, BPNN, Decision Trees	s	Turbine	Speed, torque, pitch	Predefined faults	[32]
SVM	s	Bearings	Vibration signal	Predefined faults	[264]
ELM	s	Bearings	Vibration signal	Predefined faults	[33]
Rough set, fuzzy classification	s	General	Dissolved Gas Analysis	Predefined faults	[265]
Random projection, SVM	s	General	Vibration signal	Predefined faults	[266]

Fuzzy c-means clustering	u	General	Images	Unlabelled faults (clustering)	[52]
k-means	u	Analog circuits	Voltage	Predefined faults	[49]
Decision Tree	s	Rotary machinery	Vibration signal	Predefined faults	[267]
k-NN	s	Transformers	Dissolved Gas Analysis	Predefined faults	[268]
RVM	s	Transformers	Dissolved Gas Analysis	Predefined faults	[269]
SDA	s	Turbine	Vibration signal	Predefined faults	[270]
RF	s	Bearings	Vibration signal	Predefined faults	[271]
fuzzy Takagi-Sugeno (T-S) state observer	s	Actuators	Current, angle	Predefined faults	[272]
RF	s	Motors	Current, voltage	Predefined faults	[273]
Deep CNN (ConvNet)	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[274]
PSO, SVM	s	Power network	Current, voltage	Predefined faults	[275]
RF	s	Bearings	Vibration signal	Predefined faults	[276]
LDA, BPNN	s	Bearings	Vibration signal	Predefined faults	[277]
SVM	s	Bearings	Vibration signal	Predefined faults	[12]
LS-SVM, LaPlacian eigenmaps	i	General	Images, vibration signal	Predefined faults	[278]
Kernel Entropy Component Analysis (KECA)	s	Turbine	Vibration signal	Predefined faults	[279]
Decision Tree, k-means	i	General	Current, voltage	Predefined faults	[31]
SAE, softmax classifier	s	Grinding	temperature, pressure, sound, vibration	Predefined faults	[46]
Fuzzy c-means clustering	u	Engines	Vibration signal	Unlabelled faults (clustering)	[53]
SVM, GA	s	Circuit breaker	Vibration signal	Predefined faults	[280]

Manifold clustering	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[281]
ANN, PNN	s	Motors	Voltage	Predefined faults	[282]
DBN	s	Gearboxes	Vibration signal	Predefined faults	[283]
Distributed clustering	u	HVAC	Temperature	Unlabelled faults (clustering)	[284]
ICA, SVM	s	Engines	Vibration signal	Predefined faults	[285]
Nave bayes	s	Gearboxes	Vibration signal	Predefined faults	[286]
field programmable gate array (FPGA)	s	Rotor bars	Vibration signal	Predefined faults	[287]
Density based clustering	u	Discrete events	Discrete events	Predefined faults	[288]
SAE, DBN	s	Bearings	Vibration signal	Predefined faults	[289]
GMM	i	Rotary machinery	Vibration signal	Unlabelled faults (clustering)	[290]
Hierarchical CNN	s	Bearings	Vibration signal	Predefined faults	[291]
ANN	s	Inverter	Current, voltage	Predefined faults	[292]
Decision tree, Fuzzy classifier	s	Gearboxes	Vibration signal	Predefined faults	[293]
Dictionary learning, Singular Value Decomposition, PCA, k-NN	s	Rotary machinery	Vibration signal	Predefined faults	[294]
BPNN	s	Analog circuits	Current	Predefined faults	[295]
Fuzzy Nearest Neighborhood Label Propogation	i	Transformers	Dissolved Gas Analysis	Predefined faults	[296]
SVM	s	Bearings	Vibration signal	Predefined faults	[297]
DNN	s	Temporal data	Vibration signal	Predefined faults	[298]
Manifold embedding	s	Bearings	Vibration signal	Predefined faults	[299]
SVM	s	Transformers	Dissolved Gas Analysis	Predefined faults	[300]

local tangent space alignment(LTSA) , k-NN	u	Bearings	Vibration signal	Predefined faults	[301]
PCA, HMM	s	Bearings	Vibration signal	Predefined faults	[302]
SVM	s	Bearings	Vibration signal	Predefined faults	[303]
Adaptive Neuro-Fuzzy Inference System	s	Bearings	Vibration signal	Predefined faults	[304]
ELM	s	Bearings	Vibration signal	Predefined faults	[305]
Gath-Geva Clustering, SVD, EEMD	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[306]
SVM	s	Engines	Vibration signal	Predefined faults	[307]
Fuzzy c-means clustering	u	Mechanical faults	Vibration signal	Unlabelled faults (clustering)	[308]
SVM, GA	s	Compressor	Vibration signal	Predefined faults	[309]
Decision tree, fuzzy inference	s	Compressor	Vibration signal	Predefined faults	[310]
RCM	s	Transformers	Dissolved Gas Analysis	Predefined faults	[311]
Affinity Propagation (AP) clustering, PCA	u	Engines	Rotation speed	Unlabelled faults (clustering)	[312]
CNN	s	Manufacturing	Temperature, pressure, voltage, currents	Predefined faults	[313]
large memory storage retrieval (LAMSTAR) neural network	s	Bearings	Vibration signal	Predefined faults	[314]
SVD, SVM	s	Bearings	Vibration signal	Predefined faults	[315]
DBN, RF	s	Spacecraft	Pitch, position, electronical data, etc. (1000 features)	Predefined faults	[20]
ANN	s	Bearings	Vibration signal	Predefined faults	[316]

Variational Mode Decomposition, Fuzzy c-means clustering	u	Turbine	Vibration signal	Predefined faults	[55]
SVM, fruit Fly Optimization Algorithm (FOA)	s	Bearings	Vibration signal	Predefined faults	[317]
Fuzzy decision making	s	Actuators	Temperature	Predefined faults	[318]
SVM, PSO	s	Bearings	Vibration signal	Predefined faults	[11]
SOM, SVM, PCA, k-means	u	Manufacturing	Discrete events	Binary (fault/fault free)	[51]
DBN	s	Bearings	Vibration signal	Predefined faults	[319]
Fuzzy c-means clustering, sparse component analysis	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[320]
SVM	s	Solenoid valve	Pressure, leakage	Predefined faults	[321]
local tangent space alignment(LTSA) , k-NN	s	Bearings	Vibration signal	Predefined faults	[322]
k-NN	s	Bearings	Vibration signal	Predefined faults	[323]
Dynamic uncertain causality graph (DUCG), Fuzzy Decision Tree	s	Nuclear Power Plants	Pressure, flow, temperature	Predefined faults	[324]
MLPNN	s	Gearboxes	Vibration signal	Predefined faults	[325]
Best fit tree, functional trees	s	Turbine	Vibration signal	Predefined faults	[326]
SVM	s	Bearings	Vibration signal	Predefined faults	[327]
Decision tree	s	Chemical processes	Alarms	Predefined faults	[328]
Normalized cross correlation function, phase space reconstruction (PSR)	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[329]

Resevoir Computing (RC)	s	Fuel cell	temperature, voltage, current, flow, humidity	Predefined faults	[330]
Deep CNN	s	Bearings	Vibration signal	Predefined faults	[331]
BPNN	s	HVAC	Current, voltage	Predefined faults	[332]
Fuzzy entropy, SVM	s	Bearings	Vibration signal	Predefined faults	[333]
Kernel Entropy Component Analysis	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[334]
Non-Nave bayes, EMD	s	Rotary machinery	Vibration signal	Predefined faults	[335]
PNN, IMF	s	Transmission line	Voltage	Predefined faults	[336]
PCA, PSO, SVM	s	Bearings	Vibration signal	Predefined faults	[337]
CNN	s	Bearings	Vibration signal	Predefined faults	[302]
Fuzzy c-means clustering, SVM, rough set	s	Engines	Vibration signal	Predefined faults	[338]
Bayesian inference, SVM	s	Bearings	Vibration signal	Predefined faults	[339]
Decision trees, SVM	s	Bearings	Vibration signal	Predefined faults	[340]
SVM	s	General	Speed, acceleration	Predefined faults	[341]
Feed Forward Neural Network	s	Bearings	Vibration signal	Predefined faults	[342]
Artificial Immune Algorithm, ensemble learning	s	Gearboxes	Vibration signal	Predefined faults	[343]
Fuzzy entropy, Variable predictive model-based class discrimination (VPMCD)	s	Bearings	Vibration signal	Predefined faults	[344]

Neuro-fuzzy classifier	s	Photovoltaic system	Voltage, current, temperature, solar irradiance	Predefined faults	[345]
SVM, Fuzzy clustering	u	Electrical equipment	Voltage, current	Unlabelled faults (clustering)	[346]
Fuzzy c-means clustering, SVM, PCA	u/s	Spacecraft	Temperature, voltage, current, pressure, flow	Unlabelled faults (clustering)	[347]
Fast Clustering Algorithm (FCA), SVM, Variational Mode Decomposition (VMD), PCA	u	Rotary machinery	Vibration signal	Unlabelled faults (clustering)	[348]
SVM	s	Bearings	Vibration signal	Predefined faults	[349]
learning vector quantization (LVQ) neural network, Decision tree	s	Bearings	Vibration signal	Predefined faults	[350]
EMD, ANN	s	Gearboxes	Voltage, current	Predefined faults	[351]
ANN	s	Turbine	Vibration signal	Predefined faults	[352]
SVM	s	Textile spinning	Vibration signal	Predefined faults	[353]
Support Vector Regressive Classification	s	Bearings	Vibration signal	Predefined faults	[354]
Kernel Fisher Discriminant Analysis, GMM, k-NN	u	Chemical processes	PH, temperature, feed, pressure	Unlabelled faults (clustering)	[355]
DBN	s	Bearings	Vibration signal	Predefined faults	[356]
CNN	s	Bearings	Vibration signal	Predefined faults	[357]
EMD, ANN	s	Engines	Vibration signal	Predefined faults	[358]
Adaptive Neuro-fuzzy Inference System (ANFIS),	s	Turbine	Temperature	Predefined faults	[359]

Rough set, fuzzy covering	s	Gearboxes	Vibration signal	Predefined faults	[360]
CNN	s	Bearings	Vibration signal	Predefined faults	[361]
k-NN	s	Bearings	Vibration signal	Predefined faults	[362]
ELM	s	Photovoltaic system	Power output, voltage, current	Predefined faults	[363]
PCA, GG clusering	u	Temporal data	Temperature, pH, dissolved oxygen	Predefined faults	[364]
Least Squares (LS)-SVM, TL	i	Bearings	Vibration signal	Predefined faults	[365]
k-star classifier, k-nn	s	Bearings	Vibration signal	Predefined faults	[366]
Fuzzy logic	s	Motors	Vibration signal	Predefined faults	[367]
Decision tree, MLPNN	s	Manufacturing	Vibration signal	Predefined faults	[368]
PSO, SVM	s	Turbine	Vibration signal	Predefined faults	[369]
CNN	s	Bearings	Vibration signal	Predefined faults	[370]
SVM	s	Bearings	Vibration signal	Predefined faults	[371]
SDA	i	Bearings	Vibration signal	Predefined faults	[47]
Teager-Kaiser Energy Operator, ELM	s	Bearings	Vibration signal	Predefined faults	[372]
SVM, Cuckoo search algorihm	s	PVC production	Speed, pressure, temperature, current	Predefined faults	[373]
SVM	s	Gearboxes	Vibration signal	Predefined faults	[374]
Decision tree, k-star classifier	s	Gearboxes	Sound	Predefined faults	[375]
SVM	s	Sensors	Temperature	Predefined faults	[376]
EMD, IMF, SVM	s	Rotary machinery	Vibration signal	Predefined faults	[377]
ART-NN, Fuzzy competitive learning	s	Bearings	Vibration signal	Predefined faults	[378]

Sparse coding	u	Bearings	Vibration signal	Unlabelled faults (clustering)	[379]
ANN	s	Misfire	Vibration signal	Predefined faults	[380]
TL, ANN	i	Bearings	Vibration signal	Predefined faults	[381]
KPCA, Fisher Discriminant Analysis (FDA)	s	Power systems	Voltage	Predefined faults	[382]
PCA, SVM	s	Chemical processes	PH, temperature, feed, pressure	Predefined faults	[383]
ELM	s	Compressor	Vibration signal	Predefined faults	[384]
one-vs-one ELM	s	Engines	Temperature, exhaust gas	Predefined faults	[385]
Bayesian networks	s	HVAC	Temperature	Predefined faults	[386]
EMD, RF	s	Bearings	Vibration signal	Predefined faults	[387]
CNN, SVR	s	Rotary machinery	Vibration signal	Predefined faults	[388]
ELM	s	Bearings	Vibration signal	Predefined faults	[389]
Orhogonal LDA	s	Mechanical faults	Vibration signal	Predefined faults	[390]
orthogonal fuzzy neighbourhood discriminative analysis, RNN	s	Unmanned Marine Vehicles	Current, vibration	Predefined faults	[391]
Nave bayes, expert system	s	Aircraft	Pressure, temperature, vibration, etc.	Predefined faults	[392]
VMD, Local Linear Embedding, SVM	s	Bearings	Vibration signal	Predefined faults	[393]
RBFNN	s	Bearings	Vibration signal	Predefined faults	[394]
SVM	s	Radial distribution feeder	Voltage, current	Predefined faults	[8]

CNN, EMD	s	Rotary machinery	Vibration signal	Predefined faults	[395]
PSO, SVM	s	Bearings	Vibration signal	Predefined faults	[396]
EMD, Fuzzy entropy, PSO, SVM	s	Bearings	Vibration signal	Predefined faults	[397]
EMD, SVD, Fuzzy c-means clustering	u	Turbine	Vibration signal	Predefined faults	[398]
Sparse Component Analysis	s	Bearings	Vibration signal	Predefined faults	[399]
ANN (FFNN)	s	Rotor bars	Vibration signal	Predefined faults	[400]

Table A.1: Overview of the relevant fault diagnosis literature since 2016

Appendix B

Classification

This table provides an overview of the relevant classification in high-dimensional data literature. In this overview supervised papers are indicated as s, unsupervised as u and semi-supervised as i. They are further categorized based on their type, here DR stands for Dimensionality Reduction and C stands for Classification.

Type	Technique	s/u/i	Application	Ref.
DR/C	SVM	s	Biomedical	[401]
DR	Tensor decomposition	s	EEG	[402]
DR/C	Mixture-gaussian, SVD	s/u	Genes	[403]
DR	LDA	s	General	[404]
DR	LDA	s	Face recognition, text classification	[405]
DR	Linde-Buzo-Gray	u	General	[406]
DR/C	Dictionairy learning	s	General	[407]
DR/C	External data, k-Means	s/u	General	[408]
DR	Manifold learning	u	General	[409]
DR/C	L2,1-norm-based sparse representation model	u	Video recognition	[410]
DR	Random subspaces	s	Images	[302]
DR	LDA, Multidimensional scaling (MDS)	s	Undersampled	[411]
C	Regularized classification method	s	Biomedical	[412]
DR	SRC	s	General	[413]
DR	SR, NPE	s	General	[414]
DR/C	$L_{2,1}$ norm, matrix factorization	s	General	[415]
DR/C	NPE, SRC	s/u	Images	[416]
C	ELM	s	Biomedical	[417]
DR/C	Proximal Support Vector Machines (PSVMs)	s	Biomedical	[418]

DR/C	LDA	i	General	[419]
DR/C	Relative transformation, Neighborhood graph	i	Genes	[420]
DR	Maximum Margin Criterion (MMC), local manifold	s	General	[421]
OD	Density clustering	s	Time series	[422]
DR	Multilayer perceptron	s	Images	[423]
C	k-NN, mutual information	u	General	[424]
DR	Multimanifold learning	u	General	[425]
DR	LDA, MMC, BGA	s	Genes	[426]
DR	PCA	u	Time series	[427]
C	L_2 norm	u	General	[428]
DR/C	Evolutionary Computing (EC), sparse subspace classification	u	Face recognition	[429]
C	Compressed Sensing, Compressed classification, Nearest Neighbour	u	Face recognition	[430]
C	$L_{2,1}$ regression matrix	i	General	[142]
DR/C	VAE, CNNs, Feed forward NNs	u	Fault diagnosis	[431]
C	Bayesian Inference	s	General	[432]
DR/C	k-NN, smoothly clipped absolute deviation (SCAD) logistic regression	s	General	[433]
DR/C	high-order statistical moments (SM), Virtual Collaborative Projection (VCP)	s	Face recognition	[434]
C	SVM	s	General	[435]
C	Bayesian additive regression trees	s	General	[436]
DR/C	Bayesian, kernel dictionairy learning	s	Genes	[437]
C	Evolving tree	s	Biomedical	[438]
C	Simmulated Annealing	s	General	[439]
DR	ICA, PCA	u	Genes	[440]
DR/C	Spectral regularization	u	General	[441]
DR	LDA, MMC, Least-squares classifier	i	Images	[442]
DR	LDA, LLD	i	Time series	[443]
DR/C	Feature selection, SVM	s	General	[444]
DR	LDA	s	General	[445]
DR	Trace Ratio LDA (TR-LDA)	i	General	[446]
DR	PCA, CCA	s/u/i	General	[447]
C	Soft subspace clustering	u	Fault diagnosis	[448]

C	Stacked Denoising Autoencoders	s	Fault diagnosis	[]
C	Kernel Canonical Correlation Analysis	i	Biomedical	[450]
DR/C	PCA, C4.5 Decision trees BPNN	s	Fault diagnosis	[451]
C	Maximum entropy	s	Fault diagnosis	[452]
DR/C	Manifold learning, Pairwise constrained	i	Images	[453]
DR/C	PSO, SVM	s	General	[454]
DR/C	Sparse coding, self-taught learning	i	Images	[455]
DR	Locality Preserving Projection	u	General	[456]
DR	$L_{2,1}$-norm regularization	s	Genes	[457]
DR	Neural Networks	s	General	[458]
DR/C	DNN, Softmax classifier	s	Fault diagnosis	[297]
DR/C	PCA, SVM	s	Fault diagnosis	[459]
DR/C	Tranfer learning	i	General	[460]
DR	Manifold learning	i	General	[461]
DR/C	ANN, k-NN, Improved Distance Evaluation (IDE)	s	Gearboxes	[462]
DR	Non-negative Matrix Factorization (NMF)	u	Genes	[463]
DR/C	Short-time Fourier Transform, Stacked Sparse AutoEncoder	s	Fault diagnosis	[464]
DR/C	Binary Particle Swarm Optimization (BPSO), k-NN	u	General	[465]
C	Transductive Support Tensor Machine (TSTM)	i	General	[466]
DR/C	RBNN, CEA	s	General	[467]
C	Bayesian Mixture Model, Gibbs sampler, Reverse Annealing	u	Genes	[468]
DR/C	SVM, Recursive Feature Elimination (RFE)	s	Genes	[469]
C	Gaussian Fields, Manifold learning	i	Images	[470]
DR/C	Lower supspace projection	s	Face recognition	[471]
DR/C	Subspace clustering	u	General	[472]
DR/C	LDA, Least Squares (LS)	i	General	[473]
C	Self-taught learning	i	Images	[474]
DR	Deep auto-encoders	u	Images	[475]

DR	Subspace clustering, PCA	u	Indexing	[476]
DR	Local Fisher discriminant, PCA	i	Text classification	[477]
DR	Manifold learning, Elastic Net	s	Face recognition	[478]
DR/C	SVM, ACO	s	Fault diagnosis	[479]
C	Icremental SVM	s	Fault diagnosis	[480]
C	SDA	s	Fault diagnosis	[481]
C	DNN	s	Fault diagnosis	[482]
C	Deep Belief Network (DBNs)	s	Fault diagnosis	[483]
DR	GA, SVM	s	General	[484]
DR/C	Continuous Hidden Markov Models (cHMMs), SOMs	s	Fault diagnosis	[485]
C	Markov Random Walk	i	Text classification	[486]
C	Random Forest	s	Fault diagnosis	[487]
C	Expert system, BPNN	s	Fault diagnosis	[488]
C	EP, ANN	s	Fault diagnosis	[489]
C	RF, GA	s	Fault diagnosis	[490]

Appendix C

Data

As was mentioned before the data will be tested on three data sets. Two of those are generated synthetically and one is a real data set, augmented with anomalies. This section provides an overview of these data sets, how they were generated and what their properties are. For the sake of confidentiality the axis have been anonymized when the figures concern real world data.

	Synthetic 1a	*Synthetic 1b*	*Synthetic 1c*	*Real*
Dimensionality	2	37	37	63
Continuous variables	1	35	35	32
Discrete variables	1	2	2	31
# of samples	400	400	10.000	1.051.871

Table C.1: Technical characteristics of the data sets

C.1 Synthetic data

In an effort to generate realistic synthetic data a data generation tool has been build. The data generation assumes that there are a number of relations between states of the radar, that may or may not be observed, and the observations done by the sensors. Therefore a number of states will be generated. The state values are predefined, for example the state ”technical_state::prs” can have three values: ”ON”, ”OFF” and ”RESET”. Each of these states have a predefined transition probability to each other state, thereby creating a Markov Chain. The transition probabilities are selected to be visually similar to the states found in the real data set.
The sensor readings are then generated based on these states. There are multiple options as to what might happen when a state changes. One option is a level shift, where the signal continues in a similar fashion as before, but on a different level. This change can occur instantaneous, such as in Fig. C.1 or more gradually such as

in Fig. C.2.

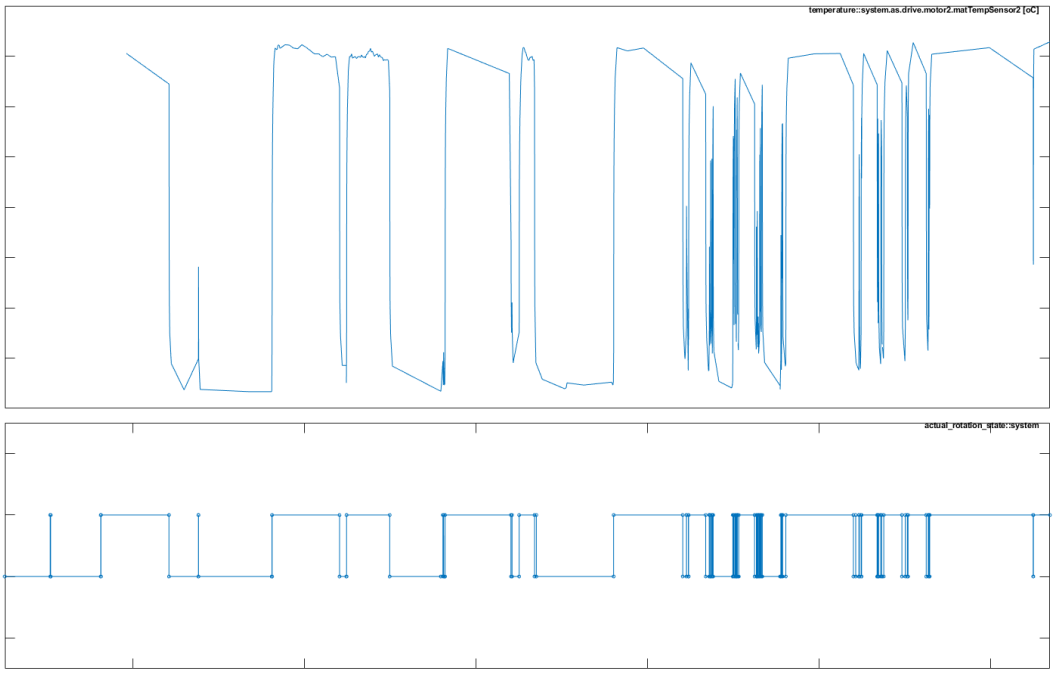


Figure C.1: Instantaneous level shift in the temperature of the motor

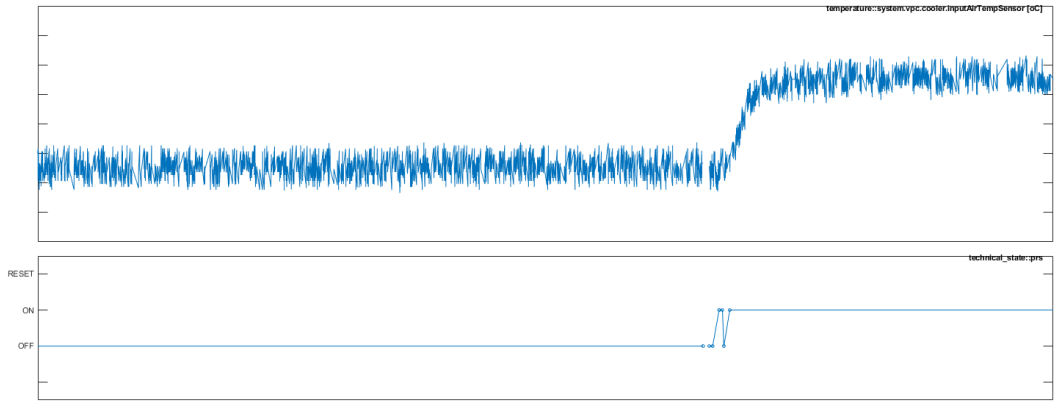


Figure C.2: Gradual level shift in the input air temperature of the cooling system

Another situation that occurs when a state changes is a short peak. For example, when the system goes into operational mode, a sudden peak in the load is expected. This is also visible in the data, when the operational mode changes the load has a short peak. This is shown in Fig. C.3.

There is also a situation possible where the amplitude of the observations changes when a state changes. For example, when the radar starts rotating, the motor current goes from flat to a moving value. This example is illustrated in Fig. C.4.

It is also possible that an observation has no relation at all with states of the radar, but is still included. This is the case for for example the outside humidity and

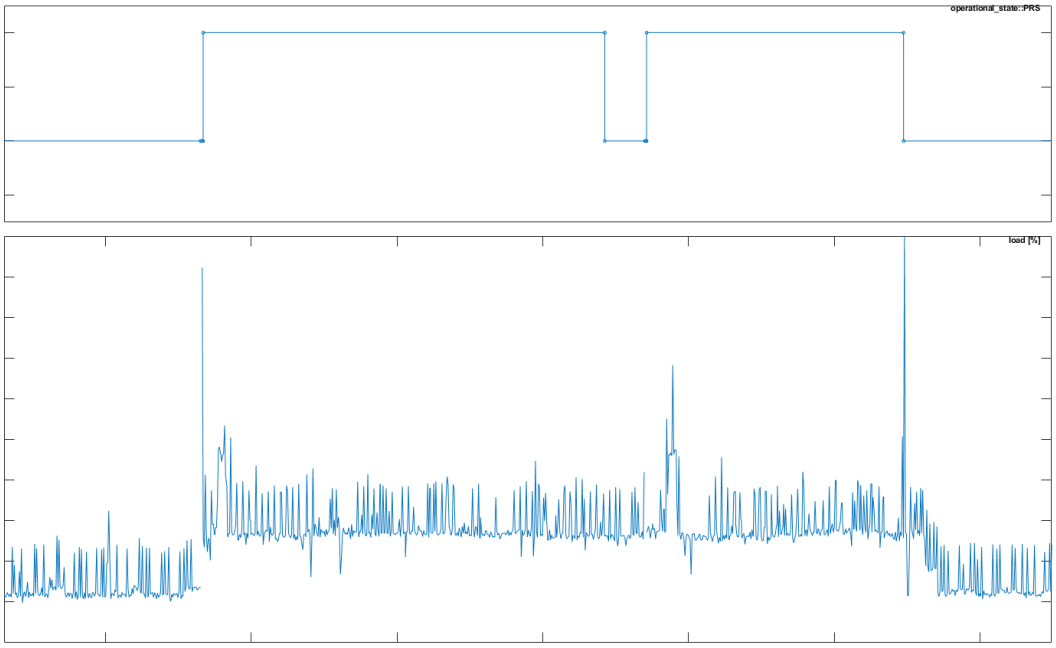


Figure C.3: Short load peak when the system goes into operation

temperature. Those are also included in the data set. The generation of the data is done in such a way that the resulting time series are visually similar to those in the real data set. The result of this comparison can be found in Fig. C.5.

The two synthetic data sets both have a different dimensionality as was already displayed in Table C.1. The first data set has a dimensionality which is about 5% of the original. That way it is possible to first test the proposed methods on a lower dimensional data set to figure out what the impact is. The second set is equal in dimensionality to the original, to provide an approximation that is as close as possible. The ratio of the different variable types (temperature, voltage, load, etc.) is also kept as close to the original as possible.

There are two types of states in the synthetic data, latent and observed. Both of these state types can influence the observed continuous variables described before, however only the latter can be used for training the model. These hidden states are also used for introducing anomalies into the data. The hidden states have, in addition to their usual states, an error state and a transition probability to that error state. When the variable is in an error state the observed variables which are derived from the state will produce an anomalous reading. When the state variable returns to the error state the observed variable will again produce a, similar, anomalous reading.

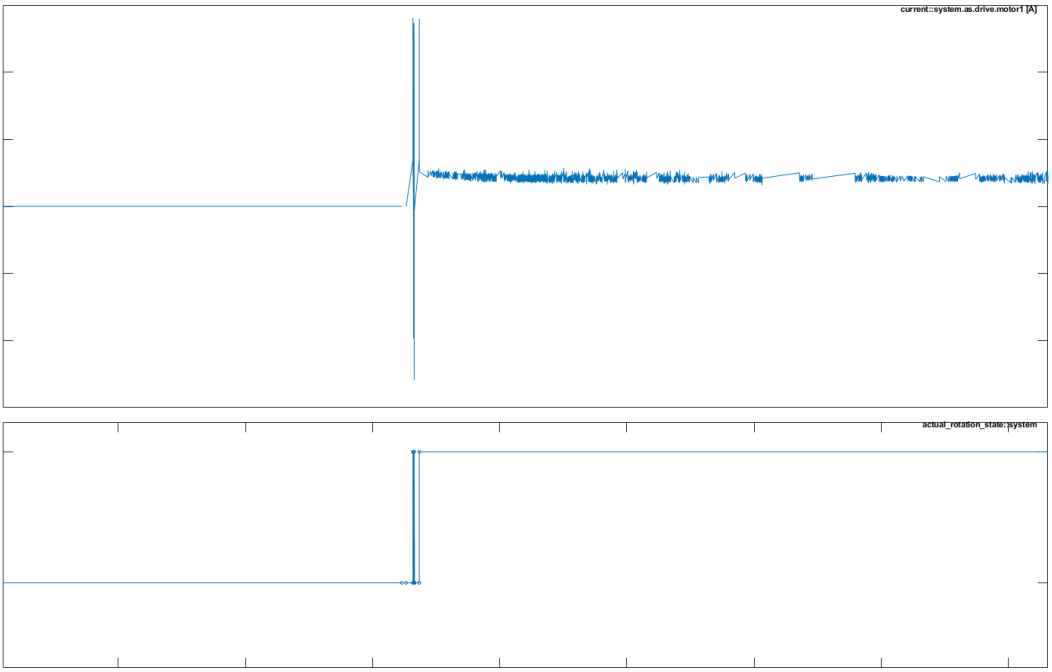
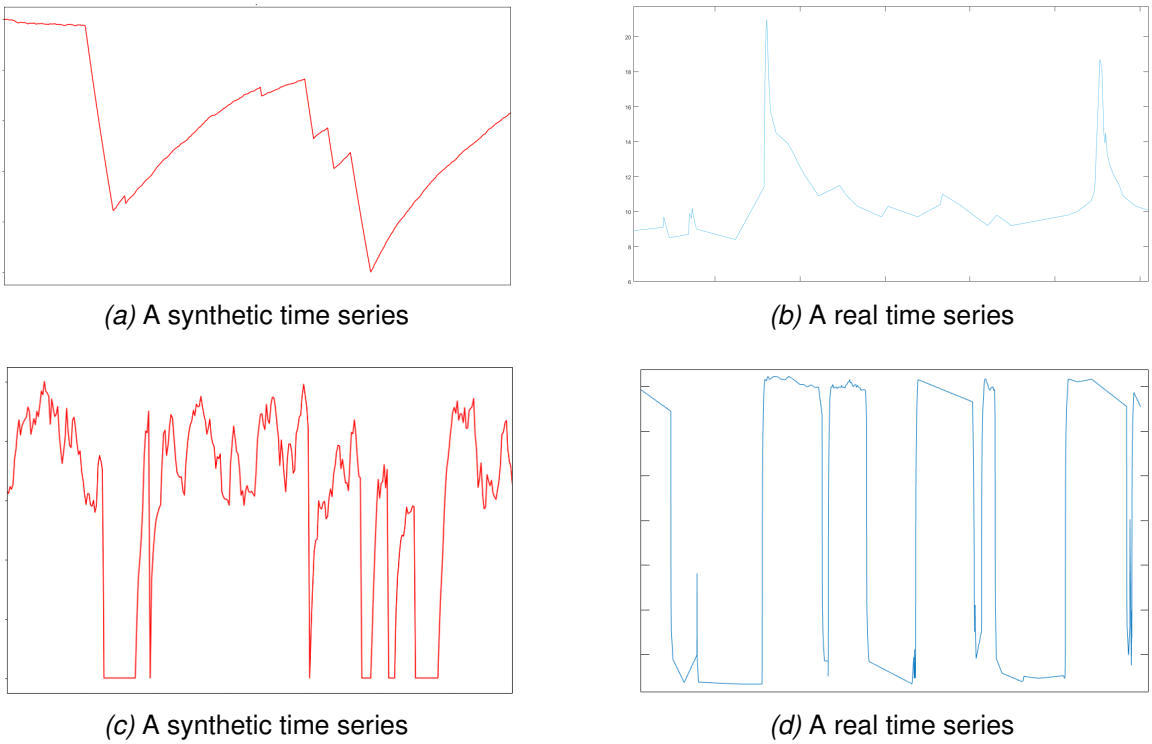


Figure C.4: Amplitude change when the radar starts rotating

C.2 Real data

To test the solutions on a real world dataset the data for the period from 2019-01-01 until 2019-05-29 is used. This data is resampled to include a reading every 10 seconds. This makes that the data set consists of a total of 1.050.989 samples, spread out over a period of almost five months. It has been limited to 63 time series, of which 32 continuous and 31 discrete variables. In this data set every anomaly has been annotated by a domain expert. The points which are not accompanied by an annotation will be considered to be normal behaviour. The time series which are included are picked based on their correspondence to these annotated faults.
In total there were three faults found by the domain expert. These faults were labelled as ”M9”, ”M10” and ”M11”. This naming is in line with other, internal, documentation. Fault ”M9” is a sudden drop in the temperature of the cooling liquid, while the two known states (the system state and the drive state), do not change. An example of this can be found in Figure C.6. Fault ”M10” occurs when there is suddenly a higher average load than would be expected given the technical states and the radar mode. An example of this can be found in Figure C.7. The final fault, ”M11” is a sudden drop in the pressure and flow rate of the cooling liquid, even though the system and drive state do not change. This is shown in Figure C.8.



(a) A synthetic time series

(b) A real time series

(c) A synthetic time series

(d) A real time series

Figure C.5: A visual comparison between a synthetic and a real time series

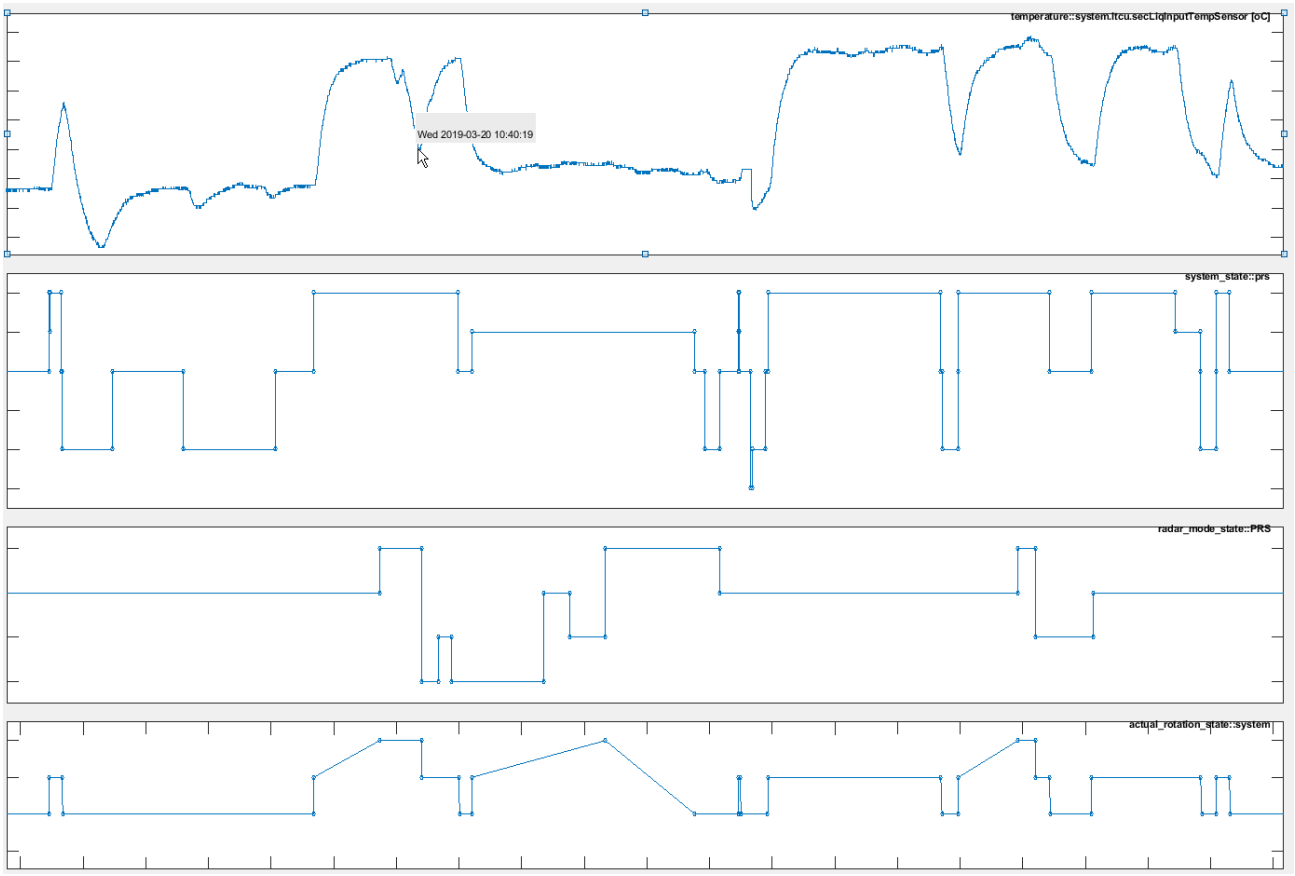


Figure C.6: A sudden drop in the temperature of the cooling liquid.

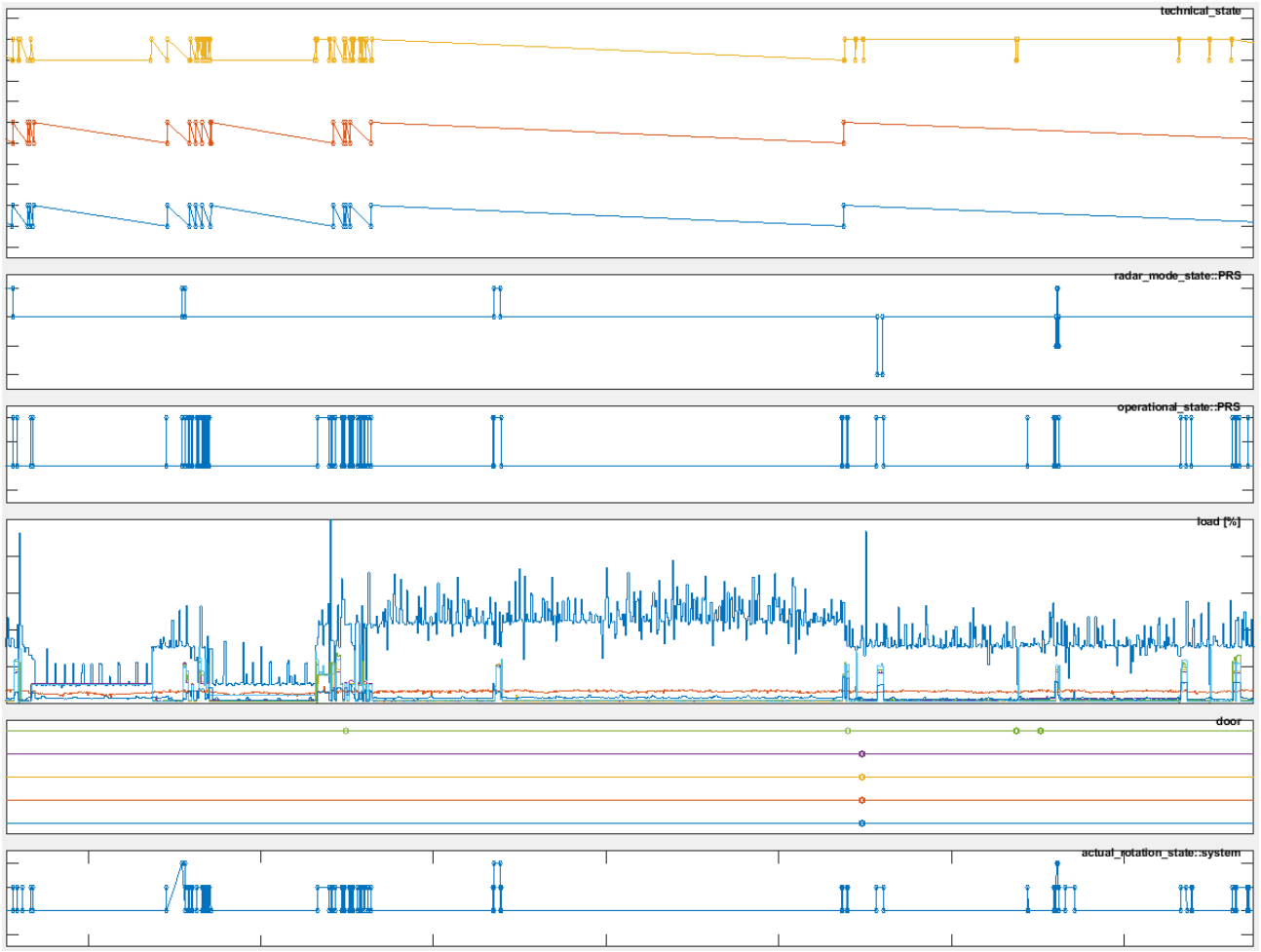


Figure C.7: A sudden higher average load than would be expected.

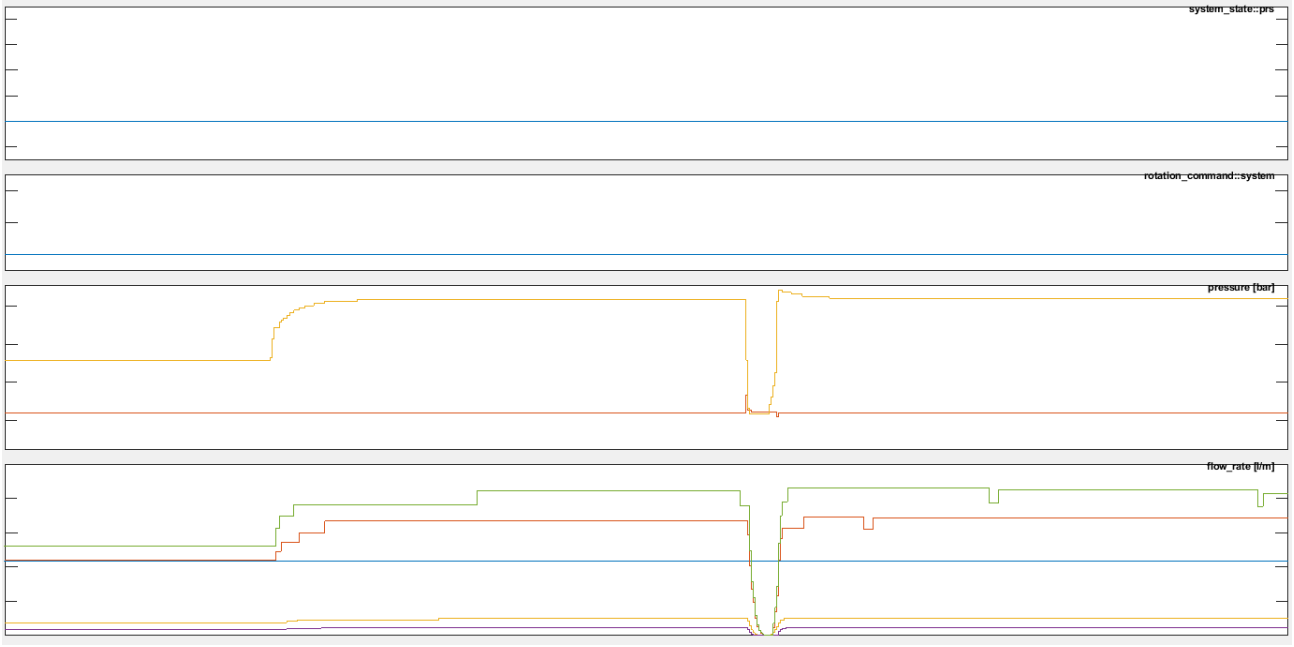


Figure C.8: A sudden drop in the pressure and flow rate of the cooling liquid.

Appendix D

Constraint-Based Semi-Supervised Self-Organizing Map

The goal here is to extend the SOM to incorporate M-Ls and C-Ls. This appendix describes the entire implementation in full detail. In the literature there is a distinction between soft constraints (where solutions that violate the constraints are allowed, though not desired) and hard constraints (where solutions where the constraints are violated are not considered) [15]. In this case the constraints are considered hard. During the rest of this section the following notation will be used:

X	The data set
N	The number of samples in data set X
D	The number of dimensions in data set X
CL	The set of all C-Ls
CL^i	The list of samples with which sample i may not be linked
ML^i	The list of samples with which sample i must linked
W^u	The weights of unit u in all D dimensions
U	The map of units
$U_{2,6}$	The unit at coordinate $(2, 6)$ in the map
BMU_1	The BMU for sample 1

Due to the properties of M-Ls groups may be created. In this implementation the BMU for each of the samples in the group is calculated. Then the modus of these BMUs is selected to be the BMU of the entire group.
When it comes to the C-Ls the BMUs may not be the same. In order to achieve this the BMUs of each of the C-Ls with a lower index than the sample in question are taken out of consideration. So when $CL = [[1, 2]]$, $BMU_1 = U_{4,2}$ then $BMU_2 \neq U_{4,2}$.

D.1 Pseudo code

In order to provide a complete and holistic view of the implementation this section will attempt to describe the algorithm in full detail in psuedo code. The code below is meant to be easy to read, not to be efficient. For that purpose non-essential details have been omitted. The pseudo code as well as the actual implementation are based on the SimpSOM package by F. Comitani [491].

procedure TRAIN($width$, $height$, X, CL, ML, $lrate_{start}$, $epochs$)
  $U \leftarrow$ `create_map`$(width, height)$
  $lrate \leftarrow lrate_{start}$
  $\sigma_{start} \leftarrow$ `max`$(width, height)/2$
  $\tau \leftarrow epochs/$`log`(σ_{start})
  for i `in range`$(epochs)$ **do**
    $\sigma \leftarrow \sigma_{start} - e^{-i/\tau}$
    $lrate \leftarrow lrate_{start} - e^{-i/epochs}$
    $x \leftarrow randint(0, N)$
    $BMU \leftarrow$ `find_bmu_constrained`(x, X, CL, ML, U)
    `update_weights`$(x, X, BMU, \sigma, lrate)$
  return U

procedure CREATE_MAP($width, height$)
  Create a map U of $width \times heights$ units and return it.

procedure FIND_BMU_CONSTRAINED(x, X, CL, ML, U)
  $units \leftarrow []$
  for ml `in` ML^x **do**
    $units.append($`find_bmu_cl`$(ml, X, CL, ML, U))$
  $units.append($`find_bmu_cl`$(x, X, CL, ML, U))$ **return** `get_modus`$(units)$

procedure FIND_BMU_CL(x, X, CL, ML, U)
  $exclude \leftarrow []$
  for cl `in` CL^x **do**
    if $cl < x$ **then**
      $exclude.append($`find_bmu_constrained`$(cl, X, Cl, ML, U))$
  return `find_bmu`$(x, X, U, exclude)$

procedure FIND_BMU(x, X, U, $exclude$)
  Find the unit that is closest to sample x, while excluding the units in $exclude$.

procedure UPDATE_WEIGHTS(x, X, BMU, σ, $lrate$)
  for u `in` U **do**
    $dist \leftarrow$ `get_distance`(u, BMU)
    $upd \leftarrow e^{-dist \times dist/(2 \times \sigma^2)}$
    $W^u \leftarrow W^u - (W^u - X^x) \times upd \times lrate$

procedure GET_DISTANCE(ux, BMU)

Get the distance between unit u and BMU (the euclidean distance between their weights)

procedure CLUSTER(X, CL, ML, U)

$clusters \leftarrow []$

for x `in` X **do**

$clusters.append($`find_bmu_constrained`$(x, X, CL, ML, U))$

return $clusters$

D.2 Assumptions

A couple of assumptions have been made while extending the SOM, the first of which is that there is a solution. Using hard constraints means that it is possible that there is no viable solution. The simplest example of this is that sample 1 and 2 have both a M-L and a C-L or that there is a M-L between $[1, 2]$ and between $[2, 3]$ but also a C-L between $[1, 3]$. The other assumption is that the preprocessing of the constraints has already been performed. That is to say that when there is a M-L between $[1, 2], [2, 3]$ there should also be a M-L between $[1, 3]$ in the list. The same should be true for the C-Ls (e.g. when $ML = [[1, 2], [2, 3]]$ and $CL = [[3, 4]]$ the complete list of C-Ls should be as follows: $CL = [[1, 4], [2, 4], [3, 4]]$).

Appendix E

Packages

Since reinventing the wheel is most of the time a rather useless activity, most of the algorithms have been implementing using a publicly available package. In Table E.1 an overview is given of the different packages used.

Algorithm	**Package**	**Author**	**Version**	**Citation**
DBN	deep-belief-network	A. Dub	1.0.3	[492]
k-Means	Scikit-Learn	D. Cournapeau	0.19.1	[493]
c-Means	Scikit-Fuzzy	J.D. Warner	0.4.1	[494]
SOM	SimpSOM	F. Comitani	1.3.4	[491]
General	NumPy		1.13.3	[495]
	SciPy		0.19.1	[496]
	Tensorflow		1.4.0	[497]
	Pandas		0.20.3	[498]
	Python		3.6	

Table E.1: An overview of the used packages