Contents

[Notation](#_Toc360648423)

[1 Introduction](#_Toc360648424)

[1.1 Theory of condition monitoring](#_Toc360648425)

[1.2 Overview of this thesis](#_Toc360648426)

[2 Feature Extraction](#_Toc360648427)

[2.1 Mel Frequency Cepstral Coefficients](#_Toc360648428)

[2.2 Multi Fractal Dimensions](#_Toc360648429)

[2.3 Kurtosis](#_Toc360648430)

[3 Semi-supervised classification](#_Toc360648431)

[3.1 Support Vector Data Description](#_Toc360648432)

[3.2 K-Center](#_Toc360648433)

[3.3 K-Means Data Description](#_Toc360648434)

[3.4 Nearest Neighbor Data Description](#_Toc360648435)

[3.5 Parzen Window Data Description](#_Toc360648436)

[3.6 Self Organizing Maps](#_Toc360648437)

[3.7 Random forest](#_Toc360648438)

[3.7.1 Variable importance](#_Toc360648439)

[3.7.2 Proximity](#_Toc360648440)

[3.7.3 Class prototypes](#_Toc360648441)

[3.7.4 Outlier measure](#_Toc360648442)

[3.7.5 Unsupervised training](#_Toc360648443)

[3.7.6 Semi-supervised classification](#_Toc360648444)

[4 Experiments](#_Toc360648445)

[4.1 Roller Bearing Datasets](#_Toc360648446)

[4.2 Feature Extraction](#_Toc360648447)

[4.2.1 Mel Frequency Cepstral Coefficients](#_Toc360648448)

[4.2.2 Higuchi Fractal Dimensions](#_Toc360648449)

[4.3 Random Forest Data Description](#_Toc360648450)

[4.4 Classification and Evaluation](#_Toc360648451)

[4.4.1 Performance evaluation](#_Toc360648452)

[4.4.2 Feature reduction experiment](#_Toc360648453)

[5 Conclusions](#_Toc360648454)

[6 References](#_Toc360648455)

***Abstract***

*Roller bearing* datahas already been used in many publications, to evaluate and proof *condition monitoring* concepts (Marwala, 2012), (Nelvamondo, Marwala, & Mahola, 2006), (Li, Chow, Tipsuwan, & Hung, 2000). Most of these papers used both normal and outlier training data to evaluate the efficiency of *supervised* classifiers such as *neural networks* or *support vector* *machines*. Goal of this thesis is the design and evaluation of an exemplary *semi-supervised* condition monitoring approach, using the roller bearing data from (Case Western University) as benchmark.

My contribution involves

* Research on *condition monitoring* and a concise presentation of a general condition monitoring approach (1)
* Research into *feature extraction* methods and selection of suitable methods with respect to the benchmark data set (2,4.2)
* Implementation of a *Higuchi fractal dimension (HFD) feature* extraction method (4.2.2)
* Research on *one-class-classification* and selection of suitable classifiers in relation with the benchmark data set and the available features (3,4)
* Design and implementation of a new *one-class-classifier* based on *random forest*
* Implementation and evaluation of a complete *semi-supervised condition monitoring* approach, including *feature extraction*, *feature selection*, classification and tests (4).

The first part of this thesis introduces several techniques, starting with *feature extraction* methods in chapter 2. Chapter 3 provides a general approach for the construction of *one-class-classifiers* and continues with an introduction of various *semi-supervised* classifiers which can be found in literature. The final part of this chapter discusses several aspects of the *random forest* classifier (3.7), of which a *semi-supervised* version was developed as part of this thesis (3.7.6), (4.3).

The techniques introduced in the first part, are evaluated in the second part of this thesis (4), which starts with a description of the *roller bearing* data set in 4.1. The *feature extraction* results are briefly discussed in 4.2 and 4.3 provides a proof of concept of the *semi-supervised random forest* classifier introduced in 3.7.6. Two experiments where designed to evaluate the complete *outlier detection* scenario, of which the first one (4.4.1) compares classifier performances, while the second analyzes the applicability of *random forest* for feature reduction purposes.