Application of Wavelets-based SVM Classification for Automated Fault Diagnosis and Prognosis of Mechanical Systems

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Maschinenbau und Verfahrenstechnik der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

von

Mahmud-Sami Saadawia, M.Sc.
aus
Bengasi, Libyen

Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker
Univ.-Prof. Dr.-Ing. Torsten Jeinsch
Tag der mündlichen Prüfung: 19. August 2015

An meine kleine wunderbare Familie;

meine Frau Fatema,

und meine Kinder Khairia und Anas

Acknowledgement

This work presents the main achievements and results obtained during my research at the Chair of Dynamics and Control (SRS) at the University of Duisburg-Essen, under the direction and supervision of Univ.-Prof. Dr.-Ing. Dirk Söffker. It is my great pleasure to express my thanks and gratitude to him for his efficient guidance and valuable support which helped me to improve myself scientifically and personally. It is also my great pleasure to express my thanks and gratitude to all my dear colleagues during my time in the chair SRS; the PhD students, as well as the technical, administrative, and academic staff of the chair, who gave me the opportunity to live in a wonderful multicultural atmosphere, rich of encouragement, assistance, and support.

I would like to thank Univ.-Prof. Dr.-Ing. Torsten Jeinsch from the University of Rostock for taking the place of my second supervisor, and for his support to improve the value of the thesis with his insightful knowledge in the area. I would like also to thank the members of the examination board; Prof. Johannes Gottschling, the chairman of the board, and Prof. Bettar O. el Moctar.

In addition, thanks to my students who supported my work which led to this thesis, and thanks to my colleagues who participated in the publications mentioned in this thesis.

Last but not least, my special thanks to my wife and children, who gave me more value and power to proceed.

Mahmud-Sami Saadawia
Duisburg, May 2016

Abstract

In this thesis, the application of pattern recognition techniques is considered for different kinds of fault diagnosis and prognosis problems and applications. The investigated applications represent real industrial applications, in which different measurement characteristics (such as cyclic, impulsive, and periodic signals), different recognition objective characteristics (such as accumulative and one-time events), different operational conditions and parameters of the machine, and different faults and detection system requirements (such as wear, crack, and object detection; system state and remaining life time) are challenging the existence of modular pattern recognition procedures and techniques. Different approaches are investigated and applied such as Support Vector Machine (SVM), Discrete Wavelet Transform (DWT), Wavelet Packet Transform (WPT), and Continuous Wavelet Transform (CWT), and many concepts and solutions are proposed and verified, in order to achieve a reliable condition monitoring system, which supports the maintenance planning of the machine and adds value to the production quality and cost.

In the first investigated application in this thesis, an approach for developing a fault diagnosis and prognosis system is presented. The system is used as a prewarning module to detect the necessity for replacing wear parts of production machines and to evaluate the remaining life time of the supervised part. The sensor signals encountered for processing are nondeterministic with cyclic nature related to the operation cycle of the machine.

In the second investigated application, the goal is to monitor a production process for online detection of a target object with the lowest possible false alarm rate. The signals encountered in the system of this work are characterized with nonstationary impulsive one-time events representing the goal object. Another characteristic of the sensor cluster signals is the partly simultaneous stimulation of events which requires the use of suitable decision fusion techniques.

In the last investigated application, two main approaches used for crack detection and prediction in rotating machinery; model based and signal based, are investigated, in order to achieve a prewarning technique for rotor cracks to be applied for online monitoring in turbo-machinery. The signals encountered are periodic vibration signals with accumulative impact of the fault incident.

Open questions arise in the issues of state evaluation, severity estimation, and remaining life time prediction, based on specific sensor data with particular application-oriented characteristics. This work deals with these open questions, in order to achieve a reliable condition monitoring system. As a general conclusion of the work, it can be stated that Wavelets and SVM are reliable tools for feature extraction and classification in the field of condition monitoring, and the feature space of SVM is useful for remaining life prediction. However; specific application oriented solutions and tricks are necessary, considering the diversity of fault diagnosis and prognosis problems and difficulties.

Kurzfassung

In dieser Arbeit werden Techniken der Mustererkennung auf verschiedene Problemstellungen der Fehlerdiagnose und -prognose angewendet. Die untersuchten Anwendungen stellen reale industrielle Anwendungen dar, bei denen verschiedene Messeigenschaften (wie zyklische, impulsive, und periodische Signale), verschiedene Charakteristik der Erkennungsobjektiven (wie kumulativ und einmalige Ereignisse), verschiedene Betriebsbedingungen und -parameter der Maschine, und verschiedene Fehler und Erkennungssystemanforderungen (wie Verschleiss, Riss, und Objekterkennung; Systemzustand und Restlebensdauer) die modulare Mustererkennungsverfahren und -techniken erfordern. Verschiedene Ansätze werden untersucht und angewendet, wie Support Vector Machine (SVM), Continuous Wavelet-Transform (CWT), Wavelet Packet Transform (WPT) und Diskrete Wavelet-Transform (DWT), und viele Konzepte und Lösungen werden vorgeschlagen und überprüft, um ein zuverlässiges Zustandsüberwachungssystem zu erreichen, dass die Instandhaltungsplanung der Maschine unterstützt und die Produktionsqualität und Produktionskosten verbessert.

In der ersten untersuchten Anwendung in dieser Arbeit wird ein Ansatz für die Entwicklung eines Fehlerdiagnose- und -prognosesystems vorgestellt. Das System wird als Vorwarnmodul verwendet, um die Notwendigkeit für das Ersetzen von Verschleissteilen von Produktionsmaschinen zu erkennen und die Restlebensdauer des überwachten Teils zu bewerten.

In der zweiten untersuchten Anwendung wird ein Produktionsverfahren überwacht. Ziel ist die Erkennung eines Objektes mit einer möglichst geringen Fehlalarmrate. Die Signale beinhalten nichtstationäre, impulsartige bzw. einmalige Ereignisse. Ein weiteres Merkmal der Sensorcluster-Signale ist die nicht gleichzeitige Erzeugung von Ereignissen, die die Verwendung von geeigneten Entscheidungsfusionstechniken erfordert.

In der letzten untersuchten Anwendung, werden modell- und signalbasierte Verfahren für die Risserkennung und Prognose in rotierenden Maschinen untersucht, um eine Vorwarnung für Rotor-Risse zu erreichen für Online-Überwachung in Turbomaschinen. Die angetroffenen Signale sind periodische Schwingungssignale mit kumulativen Auswirkungen der Fehlerereignisse.

Offene Fragen stellen sich bei den Themen Zustandsbewertung, Fehlerschweregrad und Restlebensdauer, basierend auf spezifischen Sensordaten mit besonderen anwendungsorientierten Eigenschaften. Diese Arbeit befasst sich mit diesen offenen Fragen, um ein zuverlässiges Zustandsüberwachungssystem zu erreichen. Es kann festgestellt werden, dass Wavelets und SVM sehr nützliche Werkzeuge für die Merkmalsextraktion und Klassifikation im Bereich der Zustandsüberwachung sind. Der Merkmalsraum von SVM ist nützlich für die Bewertung der verbleibenden Lebensdauer. Allerdings zeigt sich ebenfalls, dass angesichts der Herausforderungen anwendungsorientierte Lösungen gefunden werden müssen.

Contents

List of Figures

List of Tables

Nomenclature

Denotations

Symbol	Meaning
$x(t)$	Time series measurement signal
$\psi_n(t)$	Template function
$< x, \psi_n >$	Inner product od the two functions
$\psi_n^*(t)$	Complex conjugate of the function $\psi_n(t)$
$g(t)$	Window function of STFT
$\psi(t)$	Wavelets mother function
s	Wavelet scaling parameter
τ	Wavelet shifting parameter
W	Weight matrix of ANN
C	Penalty parameter of SVM
W	Coefficient vector of SVM separating hyperplane
ζ_i	Slack variables
b	Bias term of SVM separating hyperplane
ϕ	SVM mapping function
α	Lagrange multipliers
$K(x_i, x)$	Kernel function of SVM
V	Vibration velocity
A	Vibration acceleration
P	System hydraulic pressure
D	Displacement of the monitored part
H_j	Shannon entropy
E_j	Total energy
CI	Change Index
Ax	Approximation x of DWT
Dx	Detail x of DWT
T_1	Experimental threshold
$f(t)$	Steering function
M	Mass matrix
D	Damping matrix
G	Gyroscopic matrix
K	Stiffness matrix
Ω	Rotational speed
N	Input matrix of the nonlinearities

Important abbreviations

Abbrev.	Meaning
STFT	Short Time Fourier Transform
GA	Genetic Algorithms
PCA	Principal Component Analysis
FFT	Fast Fourier Transform
ANN	Artificial Neural Networks
MLP	Multilayer Perceptron
k-NN	k-Nearest-Neighbor
SVM	Support Vector Machine
RBF	Radial Basis Function
CBM	Condition-Based Maintenance
SCADA	Supervisory control and data acquisition system
FDI	Fault Detection and Isolation
KPCA	Kernel Principle Component Analysis
PSO	Particle Swarm Optimization
SOM	Self-Organizing Maps
DS	Dempster-Shafer method
WPT	Wavelet Packet Decomposition
DWT	Discrete Wavelet Transform
WVD	Wigner-Ville Distribution
SVD	Singular Value Decomposition
CWT	Continuous Wavelet Transform
CMS	Condition-Monitoring Systems
SHM	Structural Health Monitoring
FEM	Finite Element Method
DOF	Degrees of freedom
FDA	Fisher discriminant analysis
MD	Mahalanobis distance
CV	Cross validation
RMS	Root Mean Square

1 Introduction

Condition monitoring and supervision systems are essential for the automation of industrial processes. Accordingly, abnormal conditions need to be identified for fault detection and diagnostics purposes. The task of diagnosis systems is to establish relevant statements about faults as early as possible. However, in order to maximize the benefits of the operational life of a machine, maintenance procedures should be delayed until the early inception of the fault responsible for the system failure, before the occurrence of the failure. This period of time between the early inception of a machine fault and the system failure represents the racetrack for fault diagnosis techniques and methods. Consequently, more understanding and awareness of fault inception and progression is necessary for more reliable diagnostics.

In practice machines and mechanical systems do not break down or fail without any kind of warning, which is indicated, as an example, by increase in vibration level, temperature, and/or noise. In many complex mechanical systems, the symptoms occur in high dimensional level of measures, which are not directly connected to the sensors. In such cases, usual methods of fault diagnosis, such as thresholds and knowledge-based systems, are not reliable enough to represent the system state. Moreover, model-based monitoring systems are often not suitable for complex systems because of the necessity for complex modeling of the process with detailed information about process parameters and system state changes [WCDB09].

The use of pattern recognition and machine learning approaches has been providing comparatively good performance in several applications such as face recognition, text categorization, machine vision, and many others. The application of pattern recognition on fault diagnosis and condition monitoring problems has emerged as significant potential for researchers [RNWV10], therefore, many achievements have been done, and many challenges have been intensively investigated.

1.1 Motivation

In this work, the application of pattern recognition techniques is considered for different kinds of fault diagnosis and prognosis problems. The investigated problems represent real industrial applications, in which different measurement characteristics, different recognition objective characteristics, different operational conditions of the machine, and different detection system requirements are challenging the existence of modular pattern recognition procedures and techniques. The diversity of the investigated problems helps to aware of difficulties and obstacles encountered in real life industry.

In the first application to be introduced (Chapter 3), a production machine related supervision task is investigated over a long duration. It is required to develop a fault

diagnosis and prognosis system to support condition-based maintenance of wear parts in the machine. Accordingly, a prewarning module is necessary to detect the necessity for replacing wear parts. Moreover, an indication of the state evaluation and remaining life time prediction of the supervised parts is necessary. Wear parts failure should be detected on a real time basis, before scuffing or seizing lead to serious failure of the machine. The sensor signals encountered for processing are nondeterministic with cyclic nature related to the operation cycle of the machine. Operation cycles have changing lengths and characteristics. Real industrial process and sensor signals are considered in this work. However, in principle the developed approaches could include arbitrary data acquisition techniques.

In the second application (Chapter 4), different kind of incidents considered for recognition should be investigated. The individual events with no consequent and accumulating effects to be investigated in this work represent many possible process and fault related incidents. The signals encountered in the system are characterized with nonstatinary impulsive one-time events representing the goal object. Another characteristic of the sensor cluster signals is the partly simultaneous stimulation of events which prevents the advantage of using exact simultaneous events for fusing the data in the features level, which requires the use of suitable decision fusion techniques. Real industrial process and sensor signals are considered in this work. However, in principle the developed approaches could include arbitrary data acquisition techniques.

In the last application (Chapter 5), two main approaches used for crack detection and prediction in rotating machinery; model-based and signal-based, are investigated. It is required to achieve a prewarning technique for rotor cracks to be applied for online monitoring in turbo-machinery. Several strength and weak points are discussed and compared, in order to achieve a comparative overview of the available techniques. The concepts of fault severity estimation and remaining life time prediction are considered in the application. The signals encountered are periodic vibration signals with accumulative impact of the fault incident. A realistic simulation example is considered in this work. However, in principle the developed approaches can include arbitrary data acquisition techniques.

Briefly summarized, it can be stated that open questions arise in the issues of state evaluation, severity estimation, and remaining life time prediction, based on specific sensor data with particular application-oriented characteristics. This work deals with these open questions, in order to achieve a reliable condition monitoring system, which adds value to the performance of the machine.

1.2 Organization of thesis

A brief introduction and the motivation of the work are presented in this chapter. A brief overview about the theoretical background used in the thesis is presented

in Chapter 2, covering pattern recognition systems and condition monitoring, and their current state and future prospects. The following three chapters; Chapter 3, 4, and 5, cover the core of the thesis, which is the application of pattern recognition techniques for condition monitoring, including new developed concepts and solutions for application-oriented problems and difficulties. Three applications are presented; wear plates (Chapter 3), object detection (Chapter 4), and rotor crack detection (Chapter 5). Each application includes objectives and scope of work, and state of the art of the application. Each application comes with specific scientific challenges to be dealt with. It is necessary to mention that, the state of the art required for this work is distributed across the chapters, according to the structure of the thesis. The current state of the main methods is presented in Section 2.6, whereas the application state of the art which includes particular application-oriented requirements is linked with the application chapters in Sections 3.1.2, 4.1.2, and 5.1.2 respectively. Finally, in Chapter 6, the thesis is summarized including the different scientific contributions of the work, with some guidelines which might be helpful for future work.

2 Background: Pattern recognition and condition monitoring

A brief overview of the theoretical background used in the thesis is presented in this chapter, covering pattern recognition systems and condition monitoring. Moreover, the current state of methods and the future prospects are introduced.

2.1 Introduction

The recent developments in the field of pattern recognition and machine learning have been motivated by the huge advances in the field of computers and digitalization, and the requirement for reliable automation techniques and efficient handling of the huge amount of data being generated [HKK06]. Consequently, several applications have been introduced such as face recognition, text categorization, machine vision, and many others. This led to the requirements for solid theoretical foundation and improvements of the field of pattern recognition, in order to comply with the achievements in the computer integrated systems.

The field of fault diagnosis and prognosis and condition monitoring appears as a prospective research field for pattern recognition [RNWV10], therefore, many ambitions have been intensively investigated.

2.2 Pattern recognition

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes [TK08]. Depending on the application, these objects, which are characterized by patterns, could be images, signal waveforms, or any type of measurements.

2.2.1 Modules of pattern recognition systems

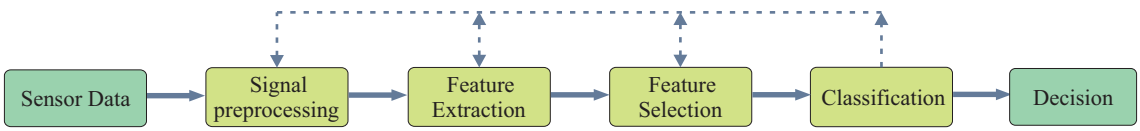


Figure 2.1: Modules of pattern recognition systems

Usually, pattern recognition systems comprise several stages of data handling, in order to achieve highest possible inference of patterns. The overall performance of

the pattern recognition system depends on the performance of the individual stages and the interaction between stages. A typical structure of pattern recognition systems is illustrated in Fig. 2.1. According to the literature [TK08, DHS00, Bis06], the modules mentioned in the structure are discussed in the following:

Signal preprocessing

In practical applications, the sensor data are usually undergone preprocessing procedures, in order to transform the measurements into simpler dataset to apply subsequent operations of pattern recognition. These procedures include idle operation signals removal, scaling and normalization of signals to reduce input variability, segmentation and rearrangement of input signals, and/or removal of outliers. The process should avoid losing relevant information, and should be applied on training and test datasets.

Feature extraction

A feature is a transformed set of system measurements representing a distinctive indicator of the underlying categorical condition of the system, which is referred to as system state. Features identify indicative characteristics of system state, which known as patterns. The process of feature extraction aims at generating of system state indicators or properties, which are referred to as features. Accordingly, suitable transformations are applied to the datasets, in order to extract the indicators and rearrange them into a recognizable structure suitable for classification. Irrelevant and redundant information should be excluded to avoid deterioration of the accuracy. Usual methods for feature extraction include time-frequency transformations such as short time Fourier transform (STFT) and wavelets.

Feature selection

The task of feature selection is to choose the most relevant and indicative features, excluding redundant and irrelevant features, thus a minimum sufficient feature subset is selected, in order to improve the generalization and reduce the computational effort. Feature selection techniques include statistical hypothesis testing, genetic algorithms (GA), and decision trees, in addition to dimensionality reduction techniques such as correlation analysis and principal component analysis (PCA).

Classification

The process of classifying the relevant features into a set of system states or characters represents an essential task in pattern recognition systems. A decision about the state of the system is the outcome of the classification process. The input vectors of features are assigned by the classifier to the relevant classes of system states or

characters, according to the corresponding patterns which are distinguished by the classifier. Accordingly, a model is used to predict the class of unknown input vectors. The model could be presented as IF-THEN rules, decision tree, or a mathematical equation. Many classification methods can be used, such as neural networks, fuzzy classifiers, decision trees, bayesian classifiers, and support vector machine.

However, the previously mentioned modules of pattern recognition systems are usually interrelated. According to [TK08], the sequence of the modules illustrated in Fig. 2.1 is not fixed. It is necessary sometimes to reform the earlier modules according to the results of the last ones. This is illustrated in Fig. 2.1 by the dotted line connecting the modules.

2.2.2 Scope of application

Several applications of pattern recognition have been explored. Consequently, a diversity of signal/data characteristics and natures, as well as difficulties and obstructions, which are related to specific processes, have been researched and investigated. In the following some of the main topics of applications are introduced.

Biometrics

Biometrics aims to achieve reliable and robust identification of humans from their personal traits [Sal10]. The main application of biometrics comprises security and authentication purposes, although identification and tracking the users of smart applications for customized services represent the significantly growing field of applications. Frequently considered approaches mentioned by the authors include recognition of fingerprint [MMJP09], face [KS09], iris [SAHO14], palmprint [FMD14] and voice. Other possible biometrics might include recognition of gait, ear image, retina, DNA, and even odor [RLBSA$^+$13] and behaviors.

A survey of machine learning methods used for biometrics applications is presented in [Sal10]. Relevant research issues and challenges are also identified by the author. Three areas of interest are emphasized; offline methods for biometric template construction and recognition, information fusion of multiple biometrics for robust results, and methods for dealing with temporal information. Some mentioned challenges include processing and handling huge data acquired in real time applications and the protection of stored biometric information for privacy.

Medical applications

Pattern recognition techniques have been given a central role in medical diagnosis. As an example, the pattern recognition techniques have been successfully applied for breast cancer detection [MB04], prostate Cancer Classification [TBR09], lung sound analysis [PSA13], and diagnosis of autoimmune diseases [FPSV14].

In the field of medical imaging, which has the most significant role in diagnosis, several pattern recognition techniques are utilized and adopted, as mentioned in [MB04, SHJ09, ALPM+13]. The methods of wavelets, neural and neuro-fuzzy systems, genetic algorithms, fuzzy-based systems, and decision trees are more emphasized. The methods are used for medical image segmentation, medical image registration, decision support systems and other machine learning tasks, in order to extract clinically relevant information required for diagnosis.

Audio, image, and video analysis

The constantly increasing progress in the digital revolution leads to a huge potential of acquisition, storage, and processing of audio, image, and video materials. These materials require effective handling and management in order to avoid information overload. Two major challenges have been mentioned by [CV08]: data analysis, which aims to extract the data content, i.e. any information that constitutes an asset for potential users, and data processing, which requires representing the data in a form that enables the transmission through physical networks as well as wireless channels.

Several techniques, as well as applications have been mentioned in the literature [CV08, MLS07, NL11]. The RBF neural network has been successfully used to extract prosodic features for audio emotion recognition [OSAC14]. For speech and movie events recognition, audio-visual combination approach has shown improvements in noisy environment [DZG+14, NDSL13]. The features extracted from images, as mentioned in [AKJ02], could be color-based; which is largely independent of view and resolution, shape-based; depending on invariant geometric shape structures, or texture-based; which utilizes the visual characteristics of homogeneous regions.

Fault diagnosis and prognosis

For more understanding of faults inception and progress, it is necessary to adopt the intelligent techniques of pattern recognition and machine learning, which are assisted by the progress of sensor technology and industrial computers. However, this led to a problem of huge amount of raw measurements requiring transformation and handling.

Pattern recognition and machine learning techniques have been successfully applied for many fault diagnosis problems including rotating machines [WPJ13, KLB+13], bearings [BBST13, ZJZL14], and chemical process industries [ZS11, NJR14]. A general strategy for employing machine learning techniques in fault diagnosis is mentioned in [RNWV10]:

1. Get as many raw measurements from as many sensors of a process as possible.

2. Submit the raw measurements to as many feature extraction techniques as possible.
3. Reduce the high dimensional data to only a few final features which simultaneously are the most discriminative descriptors of the process.
4. Induce a classifier for fault diagnosis.

The following literature present a comprehensive review of techniques and applications of pattern recognition in fault diagnosis and prognosis [VLR+06, WG10, FGRX07].

2.3 Feature extraction

The main reason for combining feature extraction with classification is to make the classification of different classes or states easier.

Signals obtained during system monitoring usually consist of three major components: a periodic component resulting from interaction between elements of the cycling dynamics in a mechanical system; a transient component caused by nonstationary events such as crack initiation; and broadband background noise. In applications involving vibrational signals, these signal components are typically associated with a large variety of related frequencies. Thus, feature extraction should be based on a suitable domain-specific transform module such as STFT or wavelet transformation. The goal of such signal processing algorithms is to transform a time-domain signal into a suitable domain to extract those characteristics that are embedded in the time series which cannot be directly observed in the original form [BD12, HYLW10, VLR+06]. Mathematically, this can be achieved by representing a signal $x(t)$ as a series of parameters and inner product coefficients according to comparison of the signal to a set of known template functions $\{\psi_n(t)\}_{n\in Z}$ as

$$C_n = < x, \psi_n > = \int x(t)\psi_n^*(t)dt, \tag{2.1}$$

where $\psi_n^*(t)$ indicates the complex conjugate of the function $\psi_n(t)$, and the inner product C_n represents the similarity between the signal $x(t)$ and the template function $\psi_n(t)$. The more similar $x(t)$ is to $\psi_n(t)$, the larger the inner product C_n will be [BWR13, GY11, HVG11].

The template function differs according to the method used to perform the transformation. To illustrate the differences in the usual methods discussed in the following, an example is presented in Fig. 2.2. A nonstationary signal consists of four groups

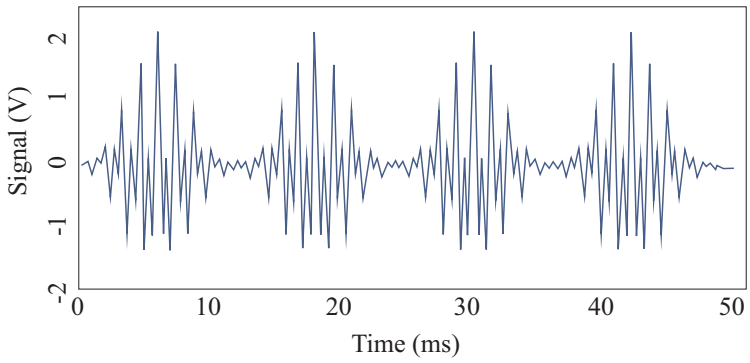


Figure 2.2: A nonstationary signal $x(t)$ (refer to [GY11])

of impulsive signal series, each containing two transient elements of different center frequencies at 1500 and 650 Hz, respectively. The four groups are separated from one another by a 12 ms time interval. Within each group, the two transient elements are time-overlapped. The sampling frequency used to capture the signal is 10 kHz.

2.3.1 Fourier transform

In Fourier transforms, the similarity of a time series signal $x(t)$ to a series of sine and cosine template functions is evaluated. Hence, any periodic signal can be presented as a weighted sum of a series of sine and cosine functions. Accordingly, the signal is transformed from the time domain into the frequency domain in order to reveal its frequency composition.

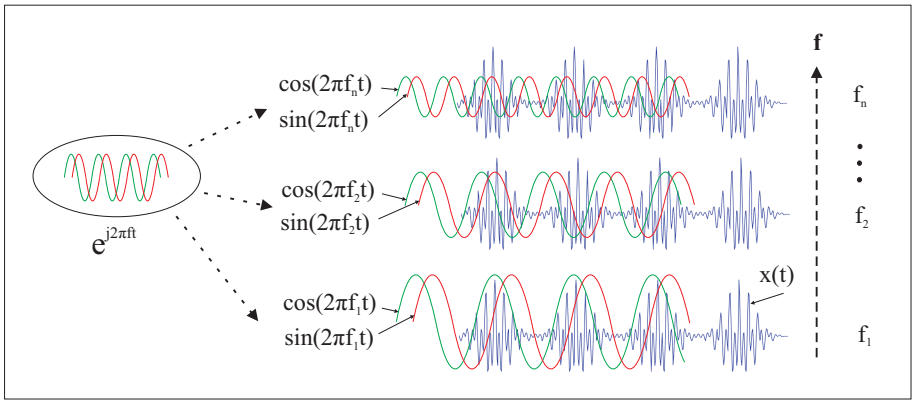


Figure 2.3: Illustration of the Fourier transform of a continuous signal $x(t)$ (acc. to [GY11])

The Fourier transform implementation of a signal $x(t)$ is illustrated in Fig. 2.3, and can be expressed as the inner product function

$$X(f) = < x, e^{i2\pi ft} > = \int_{\infty}^{\infty} x(t)e^{i2\pi ft}dt. \tag{2.2}$$

The Fast Fourier transform (FFT) is an improvement of the Fourier transform obtained by recursively break down a discrete Fourier transform of a large data sample into a series of smaller transforms of smaller samples, thus intensively decrease the computational effort. Correspondingly, the sampling frequency should be at least twice as large as the maximum frequency component contained in the signal, in order to avoid aliasing, which assumes, however, that the maximum frequency component has to be known in advance. In addition, the Fourier transform represents only the average frequency information for the entire period of the signal analyzed and not the variation of its content over time.

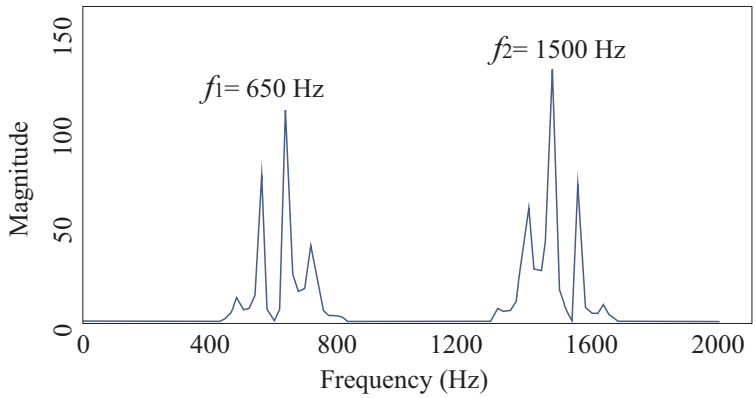


Figure 2.4: Fourier transform results of the signal $x(t)$ (refer to [GY11])

The Fourier transformation of the example signal $x(t)$ illustrated in Fig. 2.2 is shown in Fig. 2.4 as two major frequency peaks at 650 and 1500 Hz, respectively. The resulting transform does not indicate whether the two frequency components are present continuously or at certain intervals. For this reason the application of Fourier transform is limited to stationary signals and it is not suited for analyzing nonstationary signals which is the more encountered type of signals in condition monitoring.

2.3.2 Short Time Fourier Transform (STFT)

Compared to the Fourier transform, the Short Time Fourier Transform (STFT) introduces a measure of similarity between a time series signal $x(t)$ and a time-shifted and frequency-modulated window function. To overcome the limitations of Fourier transforms, STFT uses a window function $g(t)$ centered at τ, that slides

over the signal $x(t)$ along the time axis to perform a localized window-based Fourier transform as

$$STFT(\tau, f) = < x, g_{\tau,f} > = \int x(t) g^*_{\tau,f}(t) dt = \int x(t) g(t-\tau) e^{j2\pi ft} dt. \quad (2.3)$$

The segment of the signal $x(t)$ within the window function is assumed to be approximately stationary. Consequently, STFT transforms a time series signal into a two-dimensional time-frequency representation (Fig. 2.5).

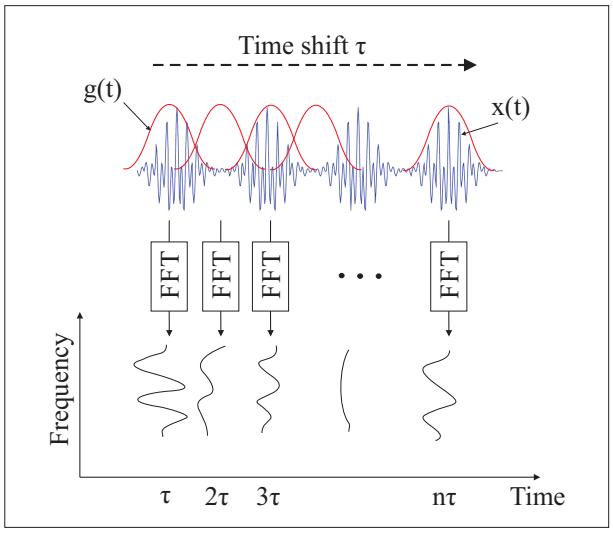
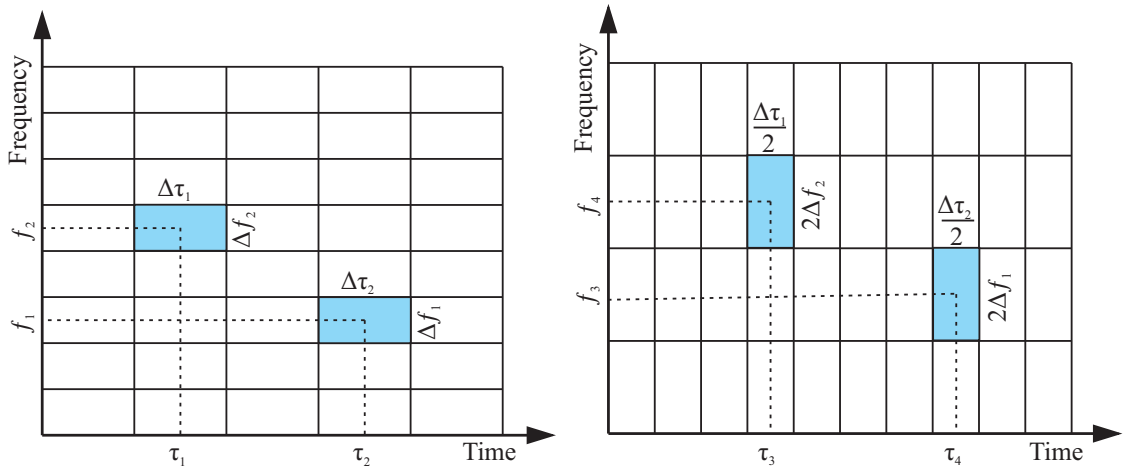


Figure 2.5: Illustration of STFT on the test signal $x(t)$ (acc. to [GY11])

One significant limitation of the STFT is the inevitable conflict between the time resolution and frequency resolution. Specifically, the product of the time and frequency resolution is lower bounded by

$$\Delta\tau \cdot \Delta f \geq \frac{1}{4\pi}, \quad (2.4)$$

where $\Delta\tau$ and Δf represent time and frequency resolution, respectively. Consequently, arbitrary selection of time and frequency resolution concurrently is out of the question. As the time and frequency resolutions of a window function depend on the parameter τ only, once the window function is chosen, the time and frequency resolutions over the entire time-frequency plane are fixed. This is illustrated in Fig. 2.6 by two alternative combinations in which the products of the time and frequency resolutions of the window function (i.e., the area defined by the product of $\Delta\tau \cdot \Delta f$) are the same, regardless of the actual window size employed (τ or $\tau/2$).



(a) Window size τ

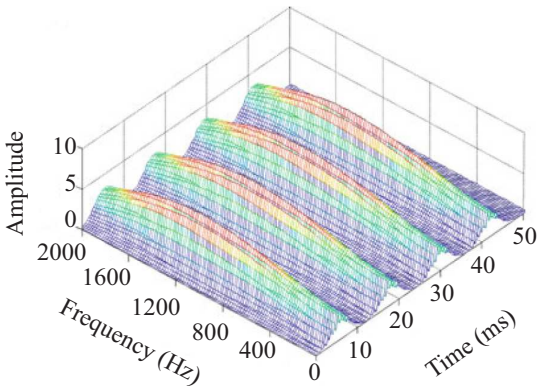
(b) window size $\tau/2$

Figure 2.6: Time frequency resolutions associated with the STFT technique (acc. to [GY11])

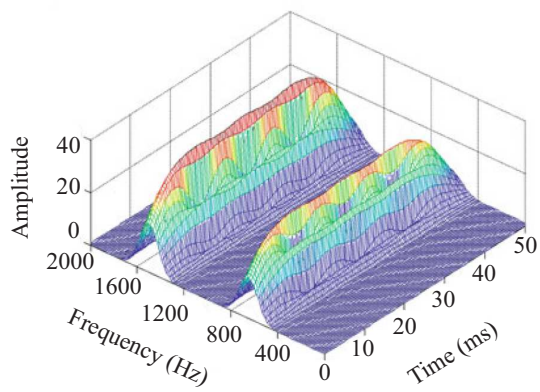
To illustrate the effect of window size on the example signal $x(t)$ introduced in Fig. 2.2, the STFT with Gaussian window is performed on the signal selecting three different window sizes; 1.6 ms, 6.4 ms, and 25.6 ms. The three resulting time-frequency representations are presented in Fig. 2.7. For the 1.6 ms window size, the four pulse trains of the signal were recognized in time, whereas the frequency resolution was insufficient to recognize the two frequency components. Conversely, the 25.6 ms window size results in good frequency resolution at the expense of time resolution. The best results were obtained by the 6.4 ms window size, where the time and frequency components are recognized. However, this required previous knowledge about the specific frequency content of the signal, which is not usually the case in real situations.

2.3.3 Wavelets

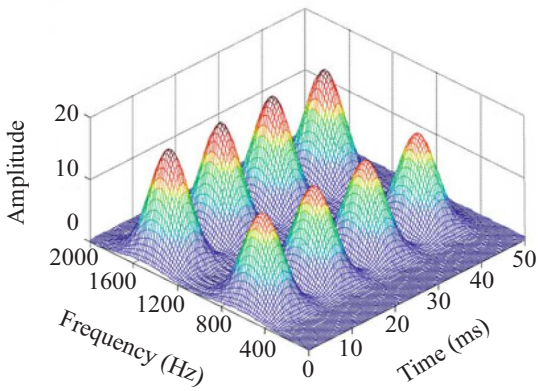
According to [GY11], the name wavelet was introduced by Jean Morlet in the mid-1970s, while working on sonic logging signals for an oil company. Morlet developed the waveforms of varying widths which are based on Haar basis functions. The progress in the field that has led to the prosperity of the wavelets was introduced by Stephane Mallat and Yves Meyer as they introduced the theory of orthogonal wavelets and multiresolution analysis, which made the chance available for strong mathematical construction of the wavelets provided by subsequent researchers such as Ingrid Daubechies.



(a) Window size 1.6 ms



(b) Window size 25.6 ms



(c) Window size 6.4 ms

Figure 2.7: Results of the STFT of the signal $x(t)$ (refer to [GY11])

The wavelet transform, in contrast to STFT, allows for variable window sizes in analyzing different frequency components within the signal. This allows good frequency resolution at low frequencies and good time resolution at high frequencies. Signals are compared to a set of template functions obtained from scaling (stretching or squeezing) and time shifting of a mother wavelet (base) function $\psi(t)$. These scaled and shifted functions represent localized frequencies of varying durations of a sound signal or image details, as an example.

The wavelet transform of a signal $x(t)$ can be expressed as the inner product of

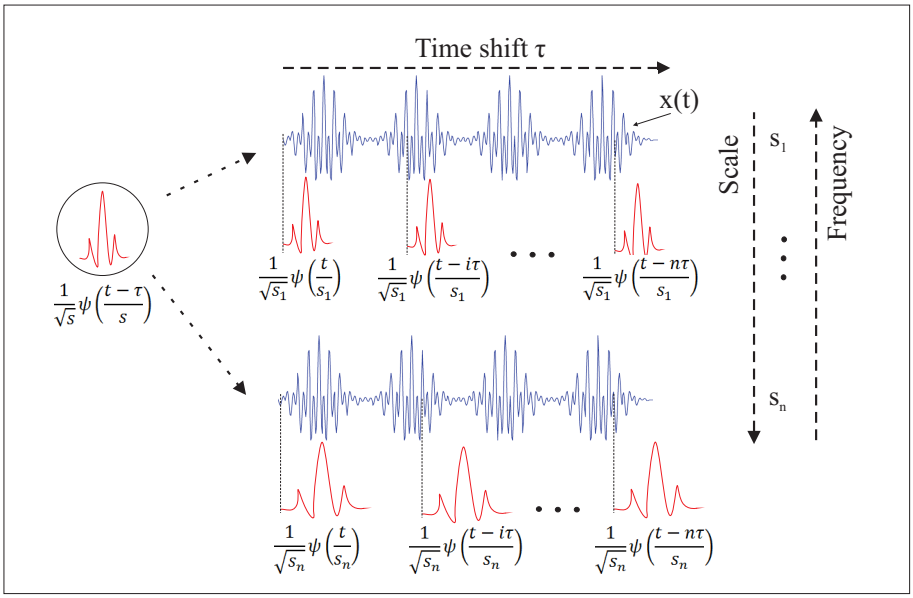


Figure 2.8: Illustration of wavelet transform (acc. to [GY11])

$$wt(s,\tau) = < x, \psi_{s,\tau} > = \frac{1}{\sqrt{s}} \int_{\infty}^{\infty} x(t)\psi^*(\frac{t-\tau}{s})dt. \tag{2.5}$$

where the symbol s represents the scaling parameter, which determines the time and frequency resolutions of the scaled base wavelet $\psi(t-\tau/s)$. The specific values of the scaling parameter s are inversely proportional to the frequency and frequency resolution, and directly proportional to the time resolution. The symbol τ represents the shifting parameter, which translates the scaled wavelet along the time axis. General description of the wavelet transform process is illustrated in Fig. 2.8.

To explain the advantage of the variable window sizes in wavelets, Fig. 2.9 presents variations of the time and frequency resolutions of Morlet mother function at two locations on the time-frequency plane. From the figure, changing the scale from s_1 at the location $(\tau_1, \eta/s_1)$ to $s_2 = 2s_1$ at $(\tau_2, \eta/s_2)$, decreases the time resolution by half (width of time window is doubled), while doubling the frequency resolution (width of frequency window is reduced to half). Thus, the wavelet transform is capable of extracting intrinsic components within a time series over its entire spectrum, through variations of the scale s and time shifts (by τ) of the base wavelet function. This is achieved by utilizing the small scales for decomposing high frequency components and the large scales for low frequency components. This is also coinciding with the result of the wavelet transform applied to the example signal introduced in Fig. 2.2 using Morlet base function. The result is presented in Fig. 2.10, in which all the transient components of the signal are recognized in time and frequency (scale). Here, the selection of window size, which is necessary for STFT, is not required.

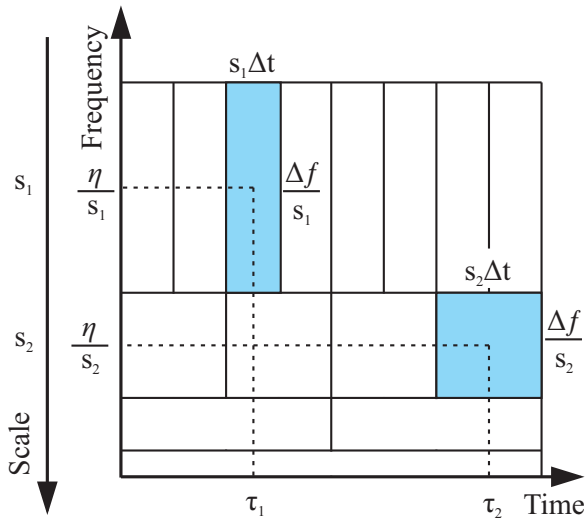


Figure 2.9: Time and frequency resolutions of the wavelet transform (acc. to [GY11])

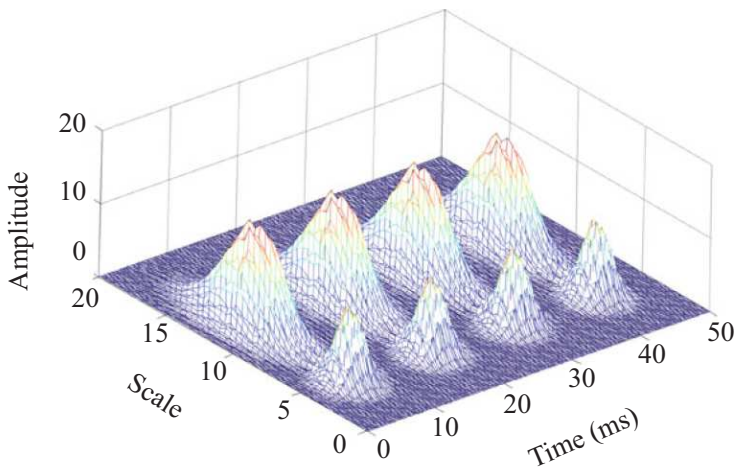


Figure 2.10: Wavelet transform of the signal $x(t)$ (refer to [GY11])

2.4 Classification

The essential task in fault detection and isolation is to recognize and identify faults. Accordingly, the characterizing features are classified into classes of faults. A classifier aims to assign each input vector of characterizing features to one of the finite numbers of classes. The classifier provides a decision about the state of the system, according to a predefined set of classes. Several classification methods have been used, such as neural networks, fuzzy classifiers, decision trees, bayesian classifiers, and support vector machine. Some of the most dominant methods are introduced in the following sections.

2.4.1 Artificial neural networks (ANN)

Artificial neural networks (ANN) are general mathematical computing paradigms that model the operations of biological neural systems [HH02]. The method was developed in the 1940s and 1950s based on the concept presented by McCulloch and Pitts [MP43], and the first generation of neural networks proposed by Rosenblatt [Ros58], which is known as perceptron.

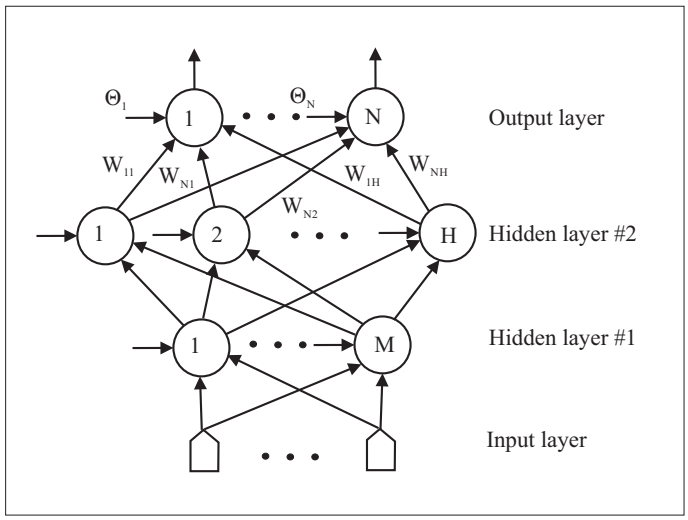


Figure 2.11: A three-layer multilayer perceptron configuration (acc. to [HH02])

An artificial neural network (Fig. 2.11) represents a network with a finite number of layers consisting of sufficient number of simple artificial nodes (neurons) with different types of connections between layers. The main characteristics of neural networks are that they have the ability to approximate complex nonlinear input-output functions, by using sets of numerical parameters (weights) that are tuned by learning [BBK10].

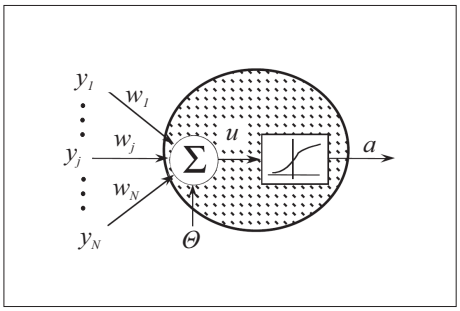


Figure 2.12: McCulloch and Pitts neuron model (acc. to [HH02])

In the diversity of neural networks, the common elementary block is the neuron. The most widely used neuron model is based on the work of McCulloch and Pitts

(Fig. 2.12) and consists of two parts: the net function and the activation function [HH02]. The net function determines how the network inputs (y_j) are combined inside the neuron. The most commonly used net function, although several net functions are proposed, is the weighted linear combination

$$u = \sum_{j=1}^{N} w_j y_j + \theta, \tag{2.6}$$

where w_j are parameters known as synaptic weights. The quantity θ is called the bias (or threshold) and is used to model the threshold. The output of the neuron (a) is related to the net function output (u) via a linear or nonlinear transformation called the activation function ($a = f(u)$). In various neural network models, several activation functions have been proposed. The most commonly used activation functions are the sigmoid, the hyperbolic tangent, the linear, and the Gaussian radial basis functions [HH02].

The multilayer perceptron (MLP) is the most well-known neural network paradigm. A multilayer perceptron (MLP) neural network model consists of a feed-forward, layered network of McCulloch and Pitts neurons (Fig. 2.11), with a nonlinear activation function that is often continuously differentiable, such as the sigmoid function and the hyperbolic tangent function [HH02]. The MLP provides a nonlinear mapping between inputs and outputs of the network. A typical three layer MLP configuration is illustrated in Fig. 2.11. The circles represent the neurons which are organized in layers labeled as the hidden layers #1 and #2, and the output layer, whereas no neurons are usually engaged in the input layer. The name hidden layer refers to the fact that the output of these neurons will be fed into upper layer neurons and, therefore, is hidden from the user who only observes the output of neurons at the output layer [HH02].

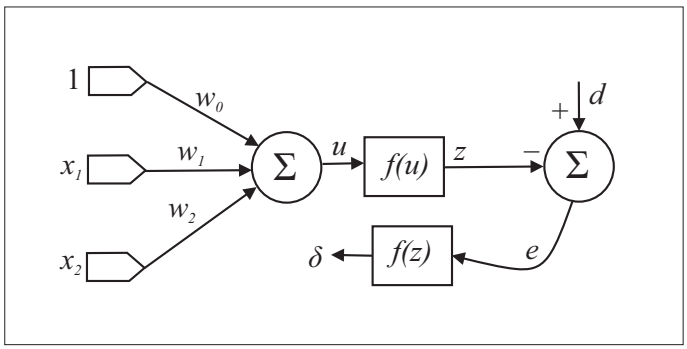


Figure 2.13: Example of MLP back-propagation training, single neuron case (acc. to [HH02])

In order to implement the MLP, the weight matrix, which contains the weight parameters, should be selected for each layer. This is usually done by the method of error back-propagation. An example for the single neuron case is illustrated in Fig. 2.13. The output z is compared with a desired target value d, and the difference (error) $e = d - z$ is back propagated in order to adjust the weights in an iterative formulation [HH02],

$$W(t+1) = W(t) + \Delta W(t), \tag{2.7}$$

where $\Delta W(t)$ is the correction made to the current weights $W(t)$. Here an initial guess for the weight matrix W is necessary. Several algorithms are used to estimate $\Delta W(t)$. The most used algorithms are the steepest descend gradient method and Newton's method. Iteration is continued as far as the error is greater than a negligible value.

2.4.2 k-Nearest-Neighbor (k-NN) classification

The k-Nearest-Neighbor method (k-NN) has been widely used in the area of classification and pattern recognition. According to [Han05],the first description of the method goes back to the early 1950s, and the increased computing power in the 1960s helped to make it popular, although the formulation of the nearest neighbor decision rules for pattern classification could have very old roots in the history [Pel14].

As described in [Bis06], the formulation of k-NN method considers a sphere centered on the test point x to be classified. The radius of the sphere grows until it contains precisely k data points irrespective of class labels. Suppose this sphere contains K_k points from class C_k. Then the posterior probability of class membership is obtained by

$$p(C_k \mid x) = K_k/k. \tag{2.8}$$

The posterior probability indicates the uncertainty of assigning a test point to particular class. The test point x is assigned to the class which has the largest posterior probability, corresponding to the largest value of K_k/k.

The k-NN method requires the entire training data set to be stored, and no model construction is done before the test measurement to be classified is given to the algorithm, which is a style of learning referred to as lazy learning [PDD02].

In spite of the simple architecture of the k-NN classifier; that it requires only a measure of closeness, as the number of training data set becomes larger, the computational effort becomes more expensive. Therefore, as it is stated in [Abe10], methods for speeding up the classification are required. Moreover, and according to [HTF08, BGRS99], the k-NN classifiers are generally affected by the curse of dimensionality, which is an expression indicating severe difficulties that can arise in high dimensional measurements.

As it is stated in [Han05], The appropriate value of k is experimentally determined starting with $k = 1$. By increasing the size of the training data set, the appropriate value of k will be increased accordingly, so that larger portion of the stored samples can be incorporated for classification decisions.

Several distance measurement functions are used in k-NN. The general formulation of Minkowski metric and the specific forms of Euclidean and Manhattan distances are often utilized in the implementation of k-NN algorithms [DHS00]. However, the choice of the distance measurement function in k-NN could be critical. According to [DHS00], a nearest-neighbor classifier would have different results depending upon different scaling processes. Moreover, and as it is stated in [Han05], nearest-neighbor classifiers can suffer from poor accuracy when given noisy or irrelevant attributes.

2.4.3 Classification with Support Vector Machine (SVM)

Since early applications in fault diagnosis [ROG99, SGBD00], SVM has yielded better results than other techniques such as neural networks, decision trees, and model-based reasoning approaches [SEMJ+11, SL12, Zan12]. The method introduced by Cortes and Vapnik [CV95] is based on statistical learning theory and is considered one of the best techniques for pattern recognition. Support vector machine (SVM) implementations have been demonstrated in a wide range of applications, including economics [HL13], text mining [YWY+13], medicine and biology [LT12, OPYM+12], remote sensing [MIO11], image segmentation [YWW11], in addition to machine fault diagnosis and condition monitoring [BD12, WY07, XWYG12].

The SVM was used for classification because of its good generalization ability and its robustness to outliers, owing to the concept of large margin classification [GE03]. Unlike typical classification methods, the learning process of SVM involves using information on the separating margin, which leads to improved separability between classes. The training of SVM aims to maximize the margin (Fig. 2.14), and thus the generalization ability is improved, which is essential in classification, especially under conditions such as scarce training data. Moreover, and according to [Abe10, AC09, Bur98], the solution of SVM does not suffer from the problem of local minima. This is in contrast to other methods such as neural networks. The SVM training also appears to be easier and requires less parameter tuning. The more transparent

geometric interpretation of the separating hyperplane in the SVM feature space provides easier validation of the results than neural networks can do.

For signal fusion tasks, the SVM feature space is used as a tool to realize a complementary transformed combination of individual signals providing better insight into the problem than individual signals. Another advantage of SVM is its robustness to outliers. Proper setting of the penalty parameter C, which controls the misclassification error, can compensate the effect of noise and outliers. It is sometimes necessary in neural networks, as stated in [Abe10], to eliminate outliers before training.

The importance of the SVM robustness to outliers is more emphasized by high-dimensional data sets with large number of features. As a result of the increasing number of features, the performance of traditional classification methods such as neural networks often decreases, which is referred to as the curse of dimensionality. To deal with this problem, dimensionality reduction and feature subset selection techniques are often applied in the earlier stages of classification. In case of SVM, according to [Kec05], the learning complexity is independent of the dimensionality of the input space. Therefore, dimensionality reduction methods do not significantly increase SVM accuracy. As stated in [Kec05], the generalization ability of SVM is not affected even in very high-dimensional spaces.

According to [Abe10, Kec05], the learning task in SVM involves finding the unknown nonlinear dependency mapping between the input vector x of training, which is high-dimensional, and the output scalar y. In general, no information about the underlying joint probability function is available. The available information is implicitly included as features in the training data set used. The solution for SVM problems correspond to minimizing the cost function

$$J = \frac{1}{2}W^T W + C\sum_{i=1}^{n}\zeta_i \tag{2.9}$$

with respect to

$$y_i(W^T\phi(x_i) + b) \geq 1 - \zeta_i \tag{2.10}$$

and

$$\zeta_i \geq 0, \tag{2.11}$$

where W is the coefficient vector for the separating hyperplane, C is a penalty parameter, ζ_i a slack variable associated with x_i, n is the number of data points, b is a scalar representing the bias term of the separating hyperplane, and ϕ is a mapping function.

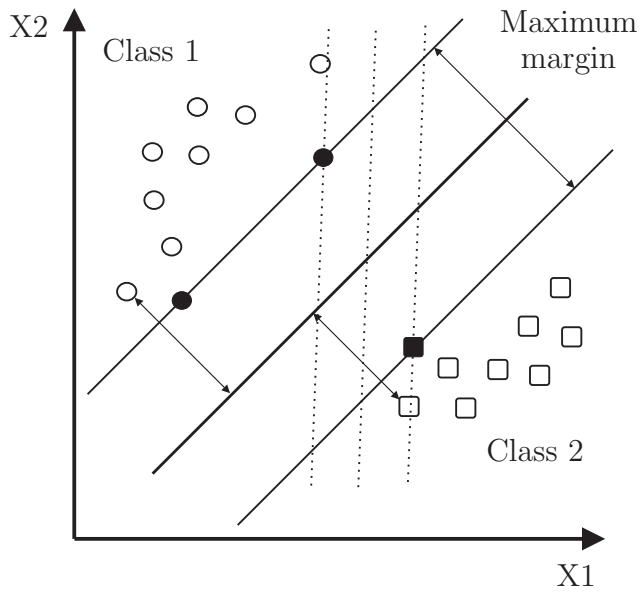


Figure 2.14: Feature space of a support vector machine

Using the techniques of Lagrange multipliers (α), and considering the dual problem, the equations Eq. 2.9 to Eq. 2.11 can be transformed into the following optimization form.

Maximize

$$G_2 = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\phi(x_i) \cdot \phi(x_j)), \tag{2.12}$$

with respect to

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \tag{2.13}$$

and

$$\alpha_i \geq 0, \tag{2.14}$$

which is a convex quadratic optimization problem [KH11]. This type of problems is easier to solve and more robust against the issue of local minima, which is a drawback of the neural network classifiers.

The solution of the transformed optimization problem returns values for α_i and ζ_i, and subsequently the coefficient vector of the separating hyperplane W and its bias term b, thus the separating hyperplane is identified within the feature space, which is the purpose of the training process.

In order to classify the subsequent test samples, the task is to check on which side of the hyperplane an individual sample x is situated. Therefore a decision function for the test data set is calculated as,

$$D = W^T \phi(x) + b = \sum_{i=1}^{n} \alpha_i y_i (\phi(x_i) \cdot \phi(x)) + b. \tag{2.15}$$

More specifically, an individual test sample x is classified, for example, as class 1 ($y = 1$) if

$$W^T \phi(x) + b = \sum_{i=1}^{n} \alpha_i y_i (\phi(x_i) \cdot \phi(x)) + b > 0, \tag{2.16}$$

and x is classified as -1 ($y = -1$) otherwise. The class name is assigned previously before the stage of training. Usual designation in case of binary classification are class 1 and class -1.

This implies that training samples that are not support vectors (i.e. $\alpha_i = 0$) do not influence the classification of new test samples. In fact, it is only necessary to compute the dot product between a test sample x and every support vector. This results in a significant saving of computational time as only the support vectors from the training set are involved in the computation [Iva07].

In case of training data which are not linearly separable, the linear hyperplane is not enough and the use of suitable mapping functions (ϕ) is necessary, in order to map input data from the input space to a higher-dimensional feature space, in which these classes become linearly separable.

As stated in [KH11], the dot product $(\phi(x_i) \cdot \phi(x))$ in Eq. 2.15 can be replaced, under certain conditions, with a kernel function $K(x_i, x)$ that is easy to compute, and the resulting classifier becomes a form of a general kernel classifier as

$$D = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b. \tag{2.17}$$

In practice, once the kernel function K is specified, the mapping and the transformed space are induced by the choice of K so that the training and classification rule can be implemented directly [KH11]. The selection and development of kernels for specific applications are crucial for classification [Abe10]. According to [Abe10, Iva07, HCL10], the most used SVM kernels are mentioned in the following.

Linear kernel

The linear kernel in the form ($K(x_i, x_j) = x_i \cdot x_j$) is used for linear classifiers, in case of linearly separable training set. It is also used to test the nonlinearity of training set and improvement potential with nonlinear kernels. Moreover, it is preferable in case of very large number of features.

Polynomial kernel

The polynomial kernel in the form ($K(x_i, x_j) = (1 + x_i \cdot x_j)^d$) is a simple and efficient method for modeling nonlinear relationships in some applications. The usual polynomial degree is $d = 2$. As the degree increases, the classification surface becomes more complex and overfitting becomes more likely to happen.

Radial Basis Function (RBF) kernel

The Radial Basis Function (RBF) kernel ($K(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2)$) is suitable as first choice in many nonlinear problems. The RBF kernel is simpler than the polynomial kernel as it has less hyperparameters, which influences the complexity of model selection. In addition, the linear kernel (basic kernel) is a special case of RBF kernel. Moreover, the RBF kernel has less numerical difficulties and good performance.

Other kernels

Many other kernels are applied for SVM such as neural (sigmoid) kernel, spline kernel, and Mahalanobis kernel. Many kernels have been developed for specific applications such as image processing and speech recognition, where inputs have variable lenghts.

2.5 Condition monitoring

Condition-Based Maintenance (CBM) activities require routine monitoring and supervision of the critical modules and parts in the equipment. Condition monitoring is the assessment of the condition of equipment, which is performed offline and during operation of the equipment.

2.5.1 Condition monitoring and data acquisition

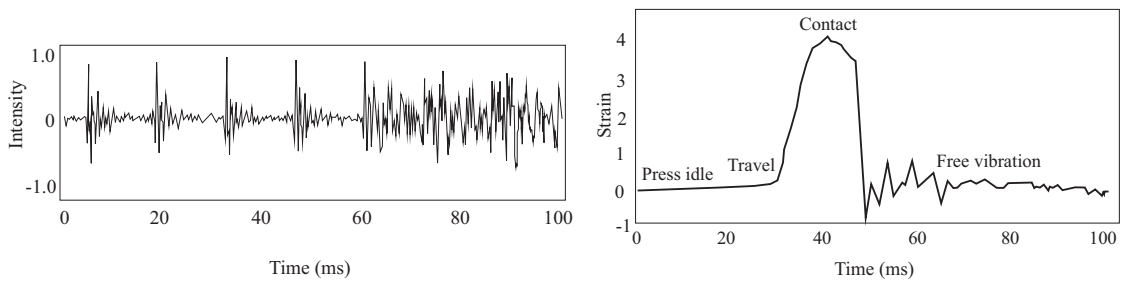
According to [Sta09], condition monitoring is classified into the following two categories: periodic monitoring and continuous monitoring. Periodic monitoring is carried out according to a regular course of procedure on a periodic fixed-time interval basis. This usually includes the regular measurements of vibration and noise as well as lubricant and debris analysis. Correspondingly, the application of continuous monitoring requires online sensors and onboard or mounted instrumentations, which

are sometimes connected to a supervisory control and data acquisition (SCADA) systems.

The outcomes and inferences of the condition monitoring carried out either periodically or continually are the basis for scheduled and urgent maintenance as well as general overhauls for plant and equipment.

2.5.2 Signals in manufacturing

The term signal refers to a physical quantity that carries certain type of information. Signals are classified in general as being either deterministic; which can be defined explicitly by mathematical functions, or nondeterministic; which are random in nature and described in statistical terms [GY11]. Deterministic signals could be periodic or transient, while nondeterministic signals comprise stationary and nonstationary behavior. The vibration caused by imbalance in a rolling bearing is an example of the deterministic signals, whereas the acoustic emission signals generated during a machining process are nondeterministic. Actually, whether a measured signal is deterministic or nondeterministic depends on its reproducibility. A signal that can be generated repeatedly with identical results is considered to be deterministic; otherwise it is nondeterministic [GY11]. However, measured signals in real world applications may contain several of the components mentioned above. Some sample industrial signals are illustrated in Fig. 2.15.



(a) Vibration signal of bearing failure

(b) Stamping process signal

Figure 2.15: Sample industrial signals (refer to [GY11])

From machine elements perspective, measured signals in manufacturing consist of three major components [GY11]:

- A periodic component resulting from the cyclic interactions between the interfacing elements of the machine; the vibrations caused by the interaction between the rolling elements and the raceway as an example.

- A transient component caused by "one-time" events. An example is the inception of a crack in rotor or the sudden breakage of a gear tooth.
- Broadband background noise.

In order to apply a reliable condition monitoring system, the acquired signals should be characterized to extract the relevant information using indicators of machine defects or indicators of product quality. These indicators, or rather these signal components are usually short in duration and weak in amplitude, and moreover, buried under strong background noises.

2.5.3 Fault detection, isolation, and prognosis

As stated in [BKL+06], a fault in a mechanical system is an abnormal condition characterized by a deviation of the system structure or the system parameters from the nominal situation. On the other hand, a failure represents the inability of the system to accomplish its function. A component fault could lead to quality degradation of the product. In addition, faults are usually accumulative which leads eventually, if not detected and repaired, to the failure of the system. Assuming that the performance of a machine (and accordingly the system state) is described by the variables (features) y_1 and y_2, an example of the different regions of the machine performance are illustrated in Fig. 2.16.

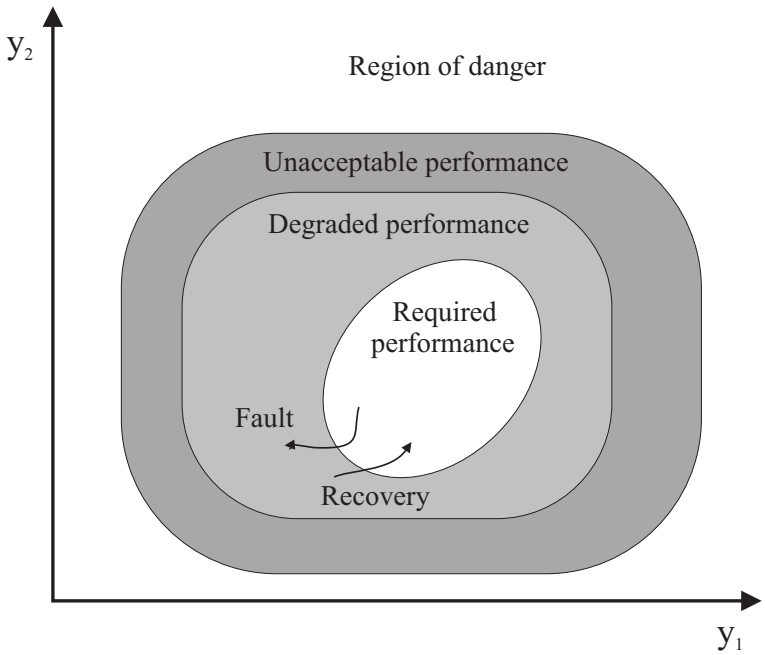


Figure 2.16: Regions of required and degraded performance (acc. to [BKL+06])

Fault detection and isolation (FDI) systems aim to identify machine faults at incipient stages to give time for corrective procedures, and to provide strict control of the product quality. Fault prognosis, on the other hand, deals with the remaining service life of a component based on the current situation. In general, the tasks of fault detection, isolation, and prognosis systems are carried out in different levels [Ryt93, Sin09]:

- Detection: The system gives a qualitative indication that fault might be present in the structure.
- Localization: The system gives information about the probable location of the fault.
- Identification: The system gives information about the nature of the fault.
- Severity quantification: The system gives information about the size and significance of the fault.
- Prognosis: The prediction of the remaining service life of the structure.

However, for efficient use of recent developments in intelligent systems and pattern recognition, a comprehensive change of fault diagnosis and maintenance philosophy is necessary, which is more oriented according to "Not on failure, nor per schedule" principles.

2.6 Current state and future prospects

In addition to exploring more applications, the current research on wavelets and support vector machines has many trends, which gives an impression that there is still more potential for improvements.

In the field of wavelet transform, a constructive work has been done for wavelet customization [CR00]. According to [GY11, SY06, GSK+06], a wavelet that is adapted to a specific signal, or rather to a specific application is designed, in order to have a higher degree of matching with the signal, and accordingly more effective extraction of patterns. Moreover, significant effort has been done to adapt base wavelet function to better analyze signals of finite length or duration, instead of infinite or periodic signals. This led to the second generation of wavelets and wavelet lifting scheme. According to [Swe98], this technique replaces the process of translation and dilation of a fixed function, and can transform signals with a finite length without extension of the signal to infinite duration, which is more useful with irregular signal samplings. Moreover, it is faster to calculate and easier to implement [GY11, URB97].

The support vector machine accuracy, speed, multiclass optimization, interpretation, incremental training, and ability to take high demand and complex structured inputs have been studied since more than a decade. The combination of SVM with other machine learning techniques has been also emphasized. A combination method of genetic algorithms (GA) with SVM is proposed and successfully used in [IT13] to select a representative subset of the DNA sequence variation (SNP), which is characterized by the restriction of genotyping millions of SNPs. Moreover, the method uses particle swarm optimization (PSO) algorithm to optimize the parameters of SVM. A proposed method of Kernel Principle Component Analysis (KPCA) to combine with multiclass SVM is also introduced by [JLZD11] and successfully applied for state recognition of aero-engine rotor during maintenance testing. The nonlinear principal components of the features are taken as eigenvectors of multi-class SVM for training and testing. A cross optimization for SVM parameters is used, and it is stated that there is no practical theoretical guidance to set these parameters to allow for better performance of SVM. A method of compressed classification learning with uniformly ergodic Markov Chain samples is adopted by [CDZT14] for SVM algorithm. The accuracy of the SVM classifier in the compressed domain is found close to that of the best classifier in the data domain, in order to avoid the curse of dimensionality in the learning process. The combination of Self-Organizing Maps (SOM) with least square support vector machine (LSSVM) is proposed in [ISS11] for time series forecasting. The results show that the proposed method of combination performs better than single LSSVM. A modified method of Particle Swarm Optimization (PSO) is proposed in [Wu11] to optimize the parameters of Wavelet ν-SVM. A simulation results show that the method is effective in dealing with many dimensions, nonlinearity and finite samples. For the purpose of dynamic fusion of classifiers, the method of Dempster-Shafer (DS) is successfully used by [BADB14] in combination with SVM classifiers to combine Inertial Navigation System (INS) and Global Positioning System (GPS) for more reliable Land Vehicle Navigation (LVN). The trained SVM model corrects the INS error in case of GPS outages thereby improving the positioning accuracy. Better results than the existing methods of Artificial Neural Network (ANN) and the Random Forest Regression (RFR) are presented.

More recently, the concept of deep learning [BCV13, ARK10, Ben09] has been given more attention, after impressive achievements of application on neural networks and deep belief nets [HOT06]. Deep learning is an efficient way of training neural networks and represented by architectures composed of multiple levels of non-linear transformations and abstractions, such as with many hidden layers in neural networks. The results have been trying to put the method of neural networks back in the lead after being no longer fashionable. However, comparable works to adopt SVM for deep learning are in progress [Tan13, KKL13]. It seems that the competition between ANN and SVM will remain open for longer time.

3 Wavelet-SVM system for evaluation of wear states and remaining life time

In this chapter, a production machine related supervision task is investigated over a long duration. An approach for developing a fault diagnosis and prognosis system to support condition-based maintenance of wear parts is presented. The system is used as a prewarning module to detect the necessity for replacing wear parts of production machines and to evaluate the remaining life time of the supervised part. Wear parts failure should be detected before scuffing or seizing lead to serious failure of the machine. The sensor signals encountered for processing are nondeterministic with cyclic nature related to the operation cycle of the machine. A real industrial process generating real sensor signals are considered in this chapter. However, in principle the developed approaches can include arbitrary data acquisition techniques.

During the progress of this work, some scientific papers were published based on the studies and results of the work. This chapter is prepared to a large extent based on the previous publications [SS12, SLS11, SS10a, SS10b, SS09].

3.1 Introduction

Usually, the failure of machines and mechanical systems is indicated by system or process specific features, for example, by an increased vibration level, increased hydraulic pressure, decreased displacement, or combinations of all these phenomena. On the other hand, the complexity and the high dimensionality of measured signals require reliable, efficient, fast, and less demanding methods, which can be easily validated, to detect and to distinguish faults by measured symptoms. The use of signal-based approaches in machine learning has been providing comparative performance without the need for complex modeling task necessary for model-based approaches.

The cyclic nature of operation can be found in many production machines and systems. Such kind of operation is usually related to the periodic nature of data documented by corresponding data acquisition systems. For reliable diagnosis and prognosis systems, the operation cycles are considered as units. Here the time series data within one cycle operation are considered as features or source of features for classification. In many cases the length of the cycles is variable, which imposes difficulties in constructing the input matrix of the classifier algorithms such as Support Vector Machine (SVM). Related SVM-approaches require constant dimensions of the input feature matrix. This problem is solved usually by zero padding, entropy, and energy measures, which lead to either deterioration of the generalization and robustness of the solution or a loss of system information.

3.1.1 Scope of work and study objectives

The objective of this work is to develop a fault diagnosis and prognosis system to support condition-based maintenance of wear parts in the production machine. The system comprises a prewarning module to detect the necessity for replacing wear parts, based on an indication of the state and predicted remaining life time of the supervised parts. Wear parts failure is detected on a real time basis, before scuffing or seizing lead to serious failure of the machine.

The sensor data considered for this work are cyclic in nature according to the operation cycle of the production machine. The process cycle results from actions of moving cylinders with different characteristics. Accordingly, the cycles comprise segments related to the individual process stages. Comparable segments in the different cycles are not necessarily equivalent, as they change in load and duration, in addition to the system and structure disturbances. Such kind of inevitable changes and nondeterministic nature make it necessary to adopt a detection system, which is invariant in nature.

In order to design a fault detection system for the current problem, different approaches have been investigated and applied based on SVM and wavelet packet decomposition in combination with energy and entropy measures. A new approach for feature-based resampling is presented based on multi-level wavelet decomposition. A related change index is proposed based on the decision value of the SVM to evaluate the remaining life time.

3.1.2 Application state of the art

Wear represents a crucial factor in determining the lifetime of machine elements. It is a complex dynamical process, incorporating surface and material properties, operating conditions, stresses, and geometry. Measurement and monitoring of wear indicators are essential in many industrial applications, which represent a considerable challenge because of the dynamic and complex nature of wear. Susceptible machine elements include, as examples, bearings, presses, railways, cylinder-piston mechanism, cam-follower mechanism, and machining tools.

In addition to the direct offline inspection of the machine element, the most commonly used techniques for wear and deterioration measurement include oil and wear debris analysis, performance analysis, and vibration analysis [Sta05]. The techniques of oil and debris analysis are usually carried out remotely in certified laboratory environment [Sta05], which negatively influences the need for immediate decisions about system state. Efforts have been made to automate wear debris analysis by classification [JMZ11, PK98]. Some other techniques for wear and deterioration measurement include acoustic emission, strain measurement, electrical effects, and thermography, as well as the relationship between performance parameters such as power and speed

[MTPP12, HHC+09, BLS+13]. Techniques based on vibration analysis have been successfully utilized as they can be applied remotely on real time basis. A technique based on the analysis of the structure of vibration signals using Singular Spectrum Analysis (SSA) and cluster analysis is proposed and successfully applied by Alonso and Salgado [AS08] for tool wear detection. The acquired tool vibration signals are decomposed by SSA into an additive set of time series, whereas cluster analysis is used to group the SSA decomposition to obtain several independent components in the frequency domain to a feed forward back-propagation (FFBP) neural network to determine the tool flank wear. Ebersbach et.al. [EPK06] introduced a combination of vibration analysis and wear debris analysis for the purpose of spur gearbox fault diagnosis under constant and cyclic overload conditions. The correlation between wear debris and vibration analysis techniques in an integrated condition monitoring maintenance program, using some specific numerical descriptors is discussed and evaluated.

Many other techniques have been utilized for classification and extraction of features from sensor signals. The Bayesian inference technique is used in [ACSK11] to update the distribution of wear coefficients, which incorporates in situ measurement data to obtain the posterior distribution. The Markov Chain Monte Carlo technique is employed to generate samples from the given distribution. A method based on artificial neural networks is successfully applied in [KTFME10] to model wear damage caused by dry fretting and to describe the dynamical frictional behavior of the interface. The influence of various mechanical parameters on the response of the network is investigated. The methods of STFT and the generalized Wigner-Ville Distribution (WVD) are used in [Sto10] for time-frequency analysis to detect wear of a valve plate in multi-piston axial pump. It is stated that quantitative assessment of time-frequency distributions and root mean square of the STFT spectrum can be useful for tracking the wear of the plate. A new approach based on wavelet packet decomposition (WPD) and empirical mode decomposition (EMD) is proposed by [BGLD12]. Signals with fault frequency feature are decomposed using WPD method for de-noising, then the intrinsic mode functions denoting the features of corresponding frequency bandwidth are obtained by EMD method. A two-stage procedure for the gear dynamic monitoring in rotating machinery is proposed in [SPH12]. Both stages are based on the combined use of Discrete Wavelet Transform (DWT) and neural networks, in order to provide information referred to the gear status (fault or normal condition) and to estimate the mesh stiffness per shaft revolution in case that any abnormality is detected.

Furthermore, the classification of variable-length quasi-periodic signals has been emphasized by several researchers, who investigated signal types such as electrocardiogram (ECG) signals, sound, and rolling element bearing vibration signals. A proposed technique for the analysis of quasi-periodical signals based on Fuzzy Finite States Machines (FFSM) is introduced in [BTvdH08]. The method extends the fuzzy logic modelling techniques by defining a linguistic model of a signal which

is associated with a new kind of membership function. A proposed method based on singular value decomposition (SVD) and polynomial classifiers is introduced in [AAN08] to extract fetal ECG from a single lead abdominal ECG signal. The method of SVD is used to extract an estimate of the maternal component from the composite abdominal signal by exploiting its quasi-periodic nature. The proposed method is validated on both real and synthetic data. The method of envelope analysis among other methods used for diagnostic analysis of acceleration signals from rolling element bearings, as stochastic signals, is investigated in [RA11]. The advantages of recent digital implementations of the method are explained, despite the fact that this technique has been used for 40 years in analogue form. A method using functional self-organizing maps (SOM) approach and k-means clustering is proposed in [DDDLV07] to discriminate between normal shaped heartbeats and abnormal ones. The high dimensional data vectors are transformed by the method in discretized functions, and the analysis is then performed on the projections of the data on a chosen subspace rather than working on the raw high-dimensional vectors. A review of the emerging role of the wavelet transforms in the recognition of electrocardiogram (ECG) signals is introduced in [Add05]. The signals are characterized by the time varying morphology subject to physiological conditions which prevent reliable detection of signal features.

The application of distributions of failure rates and reliability-based projections to the remaining useful life have been investigated and usefully utilized [Sta05, GMR00, LTLQ10]. However; in order to achieve more reliable indication of machine element quality and life, several methods have been investigated. The method of ANN is used in [RHH+13] to analyze the reliability and useful life of a steam turbine blade. A correlation between some system parameters such as damping, frequency ratio, dynamic stresses, and the cycle of life as output is established. A method based on locality preserving projections (LPP) is proposed in [Yu11] for bearing performance degradation. The method is able to discover local structure of the data manifold, which enables finding more meaningful low-dimensional information hidden in the high-dimensional observations. A method of long-term prediction is proposed in [LDN+07], based on comparing signatures from any two degradation processes using measures of similarity that form a match matrix (MM). Similarities with historical records are used to generate possible future distributions of the indicative features, and accordingly the probabilities of failure over time. A performance degradation assessment method for bearings is proposed in [ZZZ13]. The method of rough support vector data description (RSVDD) is based on the one-class SVM classifier and the rough set notion. The presented results show that the method is robust against outliers and overfitting.

3.2 Description of target system and solution concept

A general description of the production machine and the operation sequence is necessary to understand the system and wear plates measurements. The sensor signals as system input are introduced and explained. After that, the solution concept is briefly introduced.

3.2.1 Description of the machine

The press RAS III (Fig. 3.1) from the company Metso Lindemann is a serie of hydraulic triple-acting advanced scrap baling presses, used in scrap yards, car body plants, aluminum processing plants and non-ferrous metal remelting works. They compact melted iron, sheet metal and wire scrap into square edged bales; a preferred feed material for steel mills, foundries, and remelting plants.

Figure 3.1: Metso Lindemann RAS III press [Met00]

The machine RAS III (Fig. 3.2) is a hydraulic press with three hydraulic cylinders perpendicular to each other with peak hydraulic pressure of 350 bars. The special washboard profile of the baler wear plates (Fig. 3.3) should help to prevent the "plate jam" caused by thin pieces of metal and off cuts and the resulting expensive operational interruptions. Extra-long parallel guides reduce the friction and smooth running strokes of the compression units. However, worn-out wear plates are subjected to extensive deterioration, scuffing, and seizing.

The compression chamber is filled with scrap through the opening (Fig. 3.2) by crane and weighing system or controlled dosing mechanism. With the outlet closed,

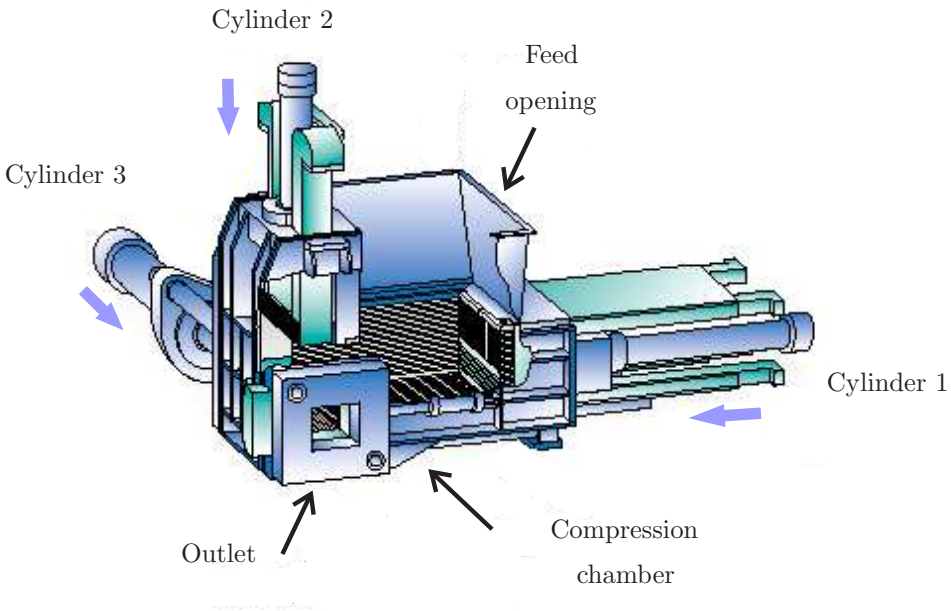


Figure 3.2: RAS III description [Met00]

Figure 3.3: Compression chamber and wear plates [Met00]

the cylinder 1 forwards pushing the scrap in the compression chamber, then cylinder 2 moves forward for more shaped material. After that, the forward movement of cylinder 3 with maximum pressure of 350 bars gives the final form of the material. Once all cylinders are placed in compression, the pressed material is given its final cubic form. The cylinders are released so that the outlet door of the material can be open. After that, the cylinder 2 is pulled back and cylinder 3 pushes the material cube out of the door to be taken by a conveyor for transportation. The cylinder 3 then is taken back behind the door to be closed immediately. At last, cylinders 1 and 3 are taken back to their initial position.

3.2.2 Description of input data

The time series vectors of the sensor data considered in this work (Fig. 3.4) comprise the process cycles in a periodic nature of the signals. A complete cycle of the considered production machine comprises the process which is represented by actions of moving cylinders with different characteristics (Fig. 3.5).

The sensor data include vibration velocity (V) and vibration acceleration (A) signals in addition to other system specific measurements. Vibration measurements are widely used in monitoring to support machinery maintenance decisions. The vibration velocity signal is a crucial variable to measure medium frequencies (until about 1 kHz) [Bru82], where the failure induced results from fatigue and wear out of surfaces. On the other hand, vibration acceleration signals are more relevant in

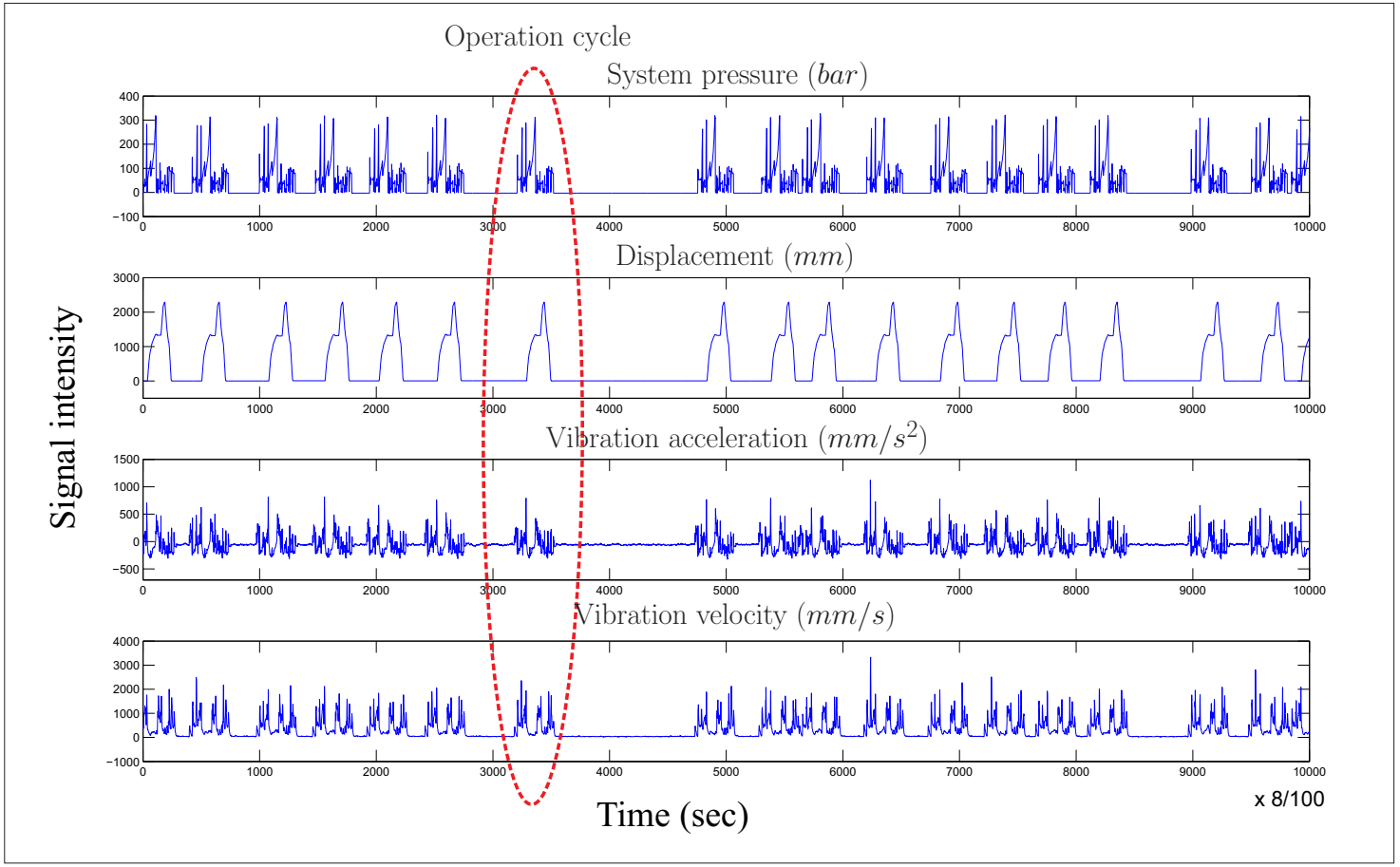


Figure 3.4: Raw data sample

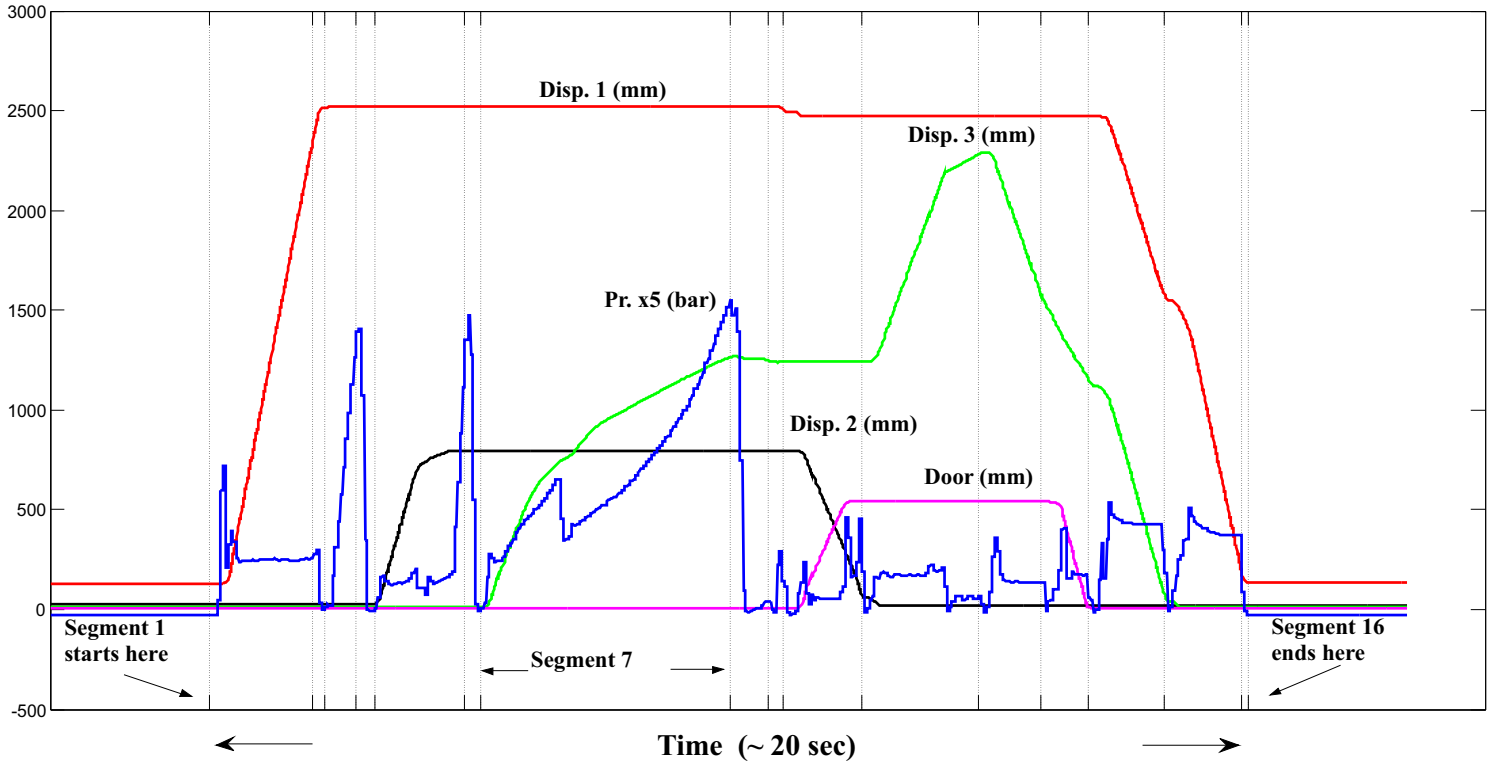


Figure 3.5: Machine operation cycle (example)

extracting transient, process-related, or impulse-like incidents, where the localized high frequencies are dominating for short times [Bru82]. Two other measurements are included in the example's sensor data; the system hydraulic pressure (P) and the piston displacement of the monitored parts (D). According to the concepts of friction, which take into account force and displacement, combining system hydraulic pressure and piston displacement provides implicit information about the friction and hence the tribological condition of the parts surfaces.

3.2.3 Concept of solution

The cyclic nature of the sensor data defines the classification unit which is treated as source of features. Accordingly, the operation cycles recognized in the time series signals (Fig. 3.4) should be transformed into data points within a multidimensional feature space representing the objective states required for classification. For this purpose, the targeted operation cycles are extracted and transformed in order to separate the indicators of the system state. The transformed data points within the state zones in the feature space give indications about fault intensity, which could be the basis for reliable quantification measures required for remaining life indicators.

The variable production parameters and process dynamics, and the consequent variable lengths of cycles and cycle segments (Fig. 3.8) make it necessary to adopt an invariant system, which can represent the system state independent of operation variations and noise.

3.3 Extraction of features

In order to apply a successful classification process, the data must be prepared by careful transformation to extract the classification indicators. Unimportant and redundant information should be excluded to avoid deterioration of the accuracy. Accordingly, the classification indicators containing the useful information should be presented in a recognizable structure suitable for classification. For the purpose of fault detection, the task of classification is to classify features into a number of classes.

3.3.1 Indicators of classification

For classification purposes, the time series vectors of the sensor data (Fig. 3.4) do not conform to the cyclic nature of the considered machine. The classification indicators are the indicators of the system states to be classified, and these indicators might represent the process cycle as well as parts of it. As an example, Fig. 3.6 shows the vibration velocity of the operation cycles before and after a machine part

replacement. It can be seen, considering the signal intensity, that more classification indicators are located at the beginning of the cycles. In the case that classification is applied directly to the time series vectors, the other parts of the cycle will deteriorate the efficiency of the classification (Fig. 3.7). This is because no considerable difference between these points can be observed according to the different machine states.

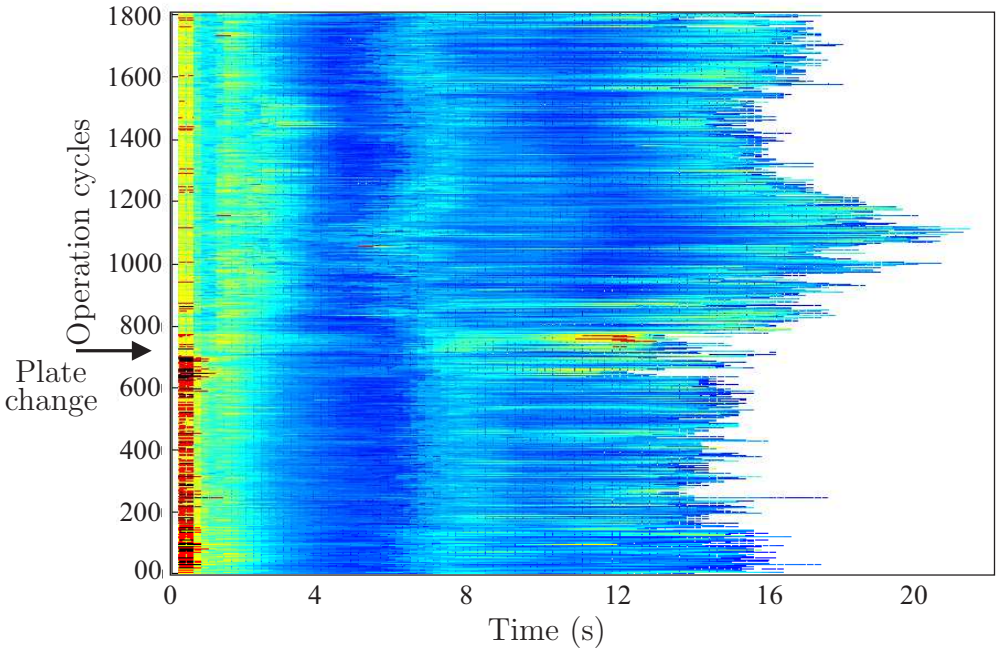


Figure 3.6: Local changes in the cycles (example)

After elimination of the redundant signals (idle operation of the machine), the signals should be structured so that transformed observations are as indicative as possible of the system states in the classification. In the presented example, extraction of the operation cycles of the machine is a preliminary step for further extraction of features. Furthermore, it is important for the purpose of study to divide the process cycle into data segments. This is because of the nature of the parameters of the process. The complete operation cycle of the machine comprises the process. The operation cycle comprises actions of moving cylinders, and these actions have different importance, different shape, and accordingly, different characteristics of related real signals. Additionally, these actions have to be treated independently. Here in the example, the machine cycle is divided into 16 data segments (Fig. 3.5), by aid of the control signals of the machine in addition to comparisons of signal values. Each data segment has similar characteristics in all cycles. All these data segments of the machine process would give information about the behavior of the machine; however, some data segments are more informative and have more classification

indicators than others.

To recognize the segments which have more classification indicators, a feature space representing the two states of the machine is constructed using a training data set in the form of time series. In the feature space the points with the highest separability of the system states (the separable points in Fig. 3.7) are detected (here manually) and re-allocated to the time series vector to specify the best candidate segments for classification. Data segment 7 and data segments 12-14 are found to be the best candidates to be considered. The data segment 7 is the power stage where the machine is subjected to the highest stresses. It is the main stage which is responsible for deterioration because of the direct effect of the load on the wear rate of materials. On the other hand, data segments 12-14 comprise pulling back the cylinder without any load and this would allow for materials contact without disturbance of load changes in material and quantity. Both data segments, 7 and 12-14, where taken into consideration during the design of the system to select the most reliable data segment in diagnosis. The values of vibration velocity, vibration acceleration, the system hydraulic pressure, and displacement, along the data segments 7 and 12-14 respectively where taken as attributes in the classification process.

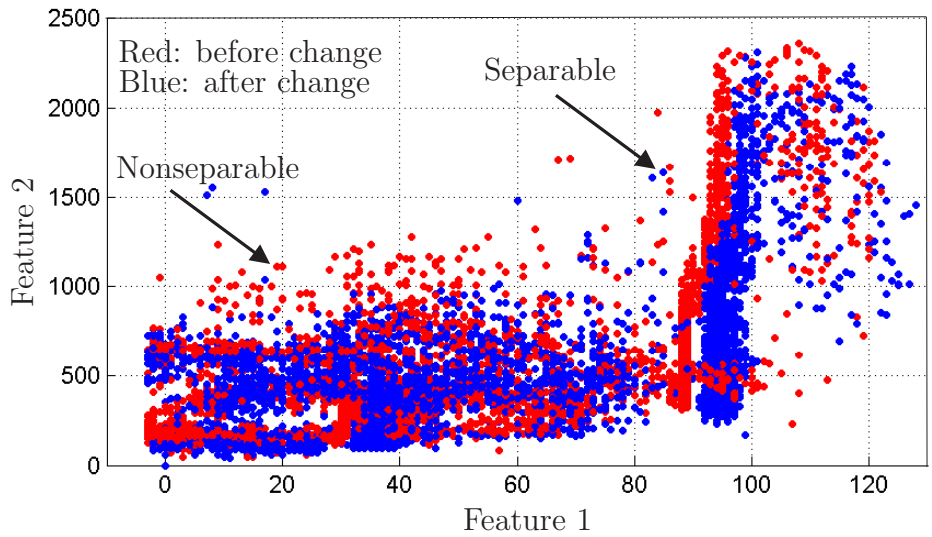


Figure 3.7: Local classification indicators (example)

3.3.2 Direct application

For construction of the feature matrix as classifier input, cycles with constant lengths are required. Padding with zeros is an alternative method providing that changes in length are limited which is not the usual case in reality. Because of the reliability of the solution deteriorated by zero padding, in the presented example the zero

padding is applied only for comparison purposes of the time series signals and their combinations. In order to consider cycles with equal lengths, a process of homogenous resampling is applied to the cycles. In this case the most usual length of the cycles is used as target length. The problem of this approach is that the lengths of the cycle segments are not coinciding with the length of the cycles (Fig. 3.8). It is usual to have longer cycles with shorter individual segments.

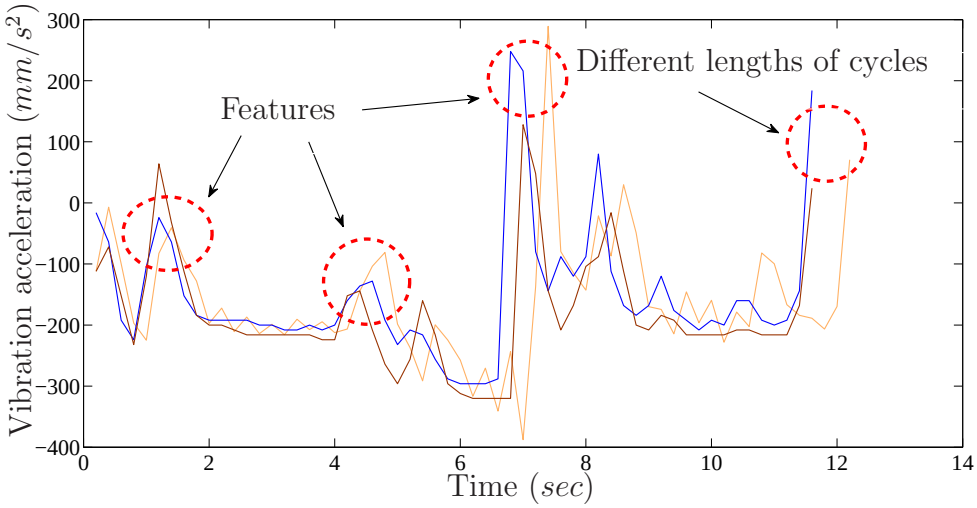


Figure 3.8: Cycles with different lengths (example)

3.3.3 Feature extraction by WPT Entropy and WPT Energy

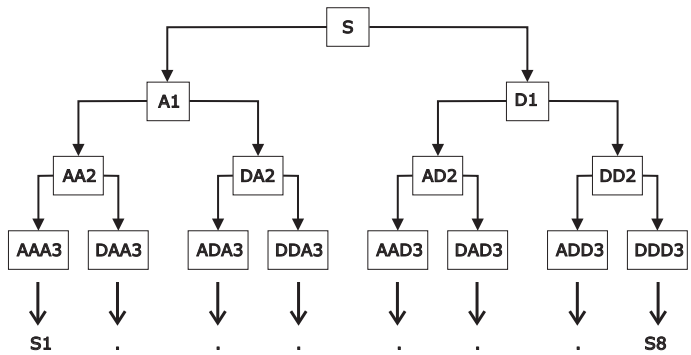


Figure 3.9: The decomposition of WPT

According to Wavelet Packet Transform (WPT) (Fig. 3.9), signals are decomposed into different frequency bands. A group of sub-signals is generated (AAA_3 to DDD_3) ranging from the low frequencies to the high frequency band. The different combinations of A and D indicate the position of nodes. Each node represents a certain

degree of signal characteristics. In the presented example of the cyclic data, the extracted cycle segments representing a cycle are decomposed by WPT into different frequency bands using 6 depth Daubechies db4 wavelet. The wavelet packet coefficients are reconstructed to a group of signals in every frequency range. The decompositions $S1$ to $S8$ indicate the reconstructed signals which represent the components of the original signal S.

The Shannon entropy (H_j) and the total energy (E_j) for each band signal Si (i = 1 to 8) are calculated (Fig. 3.10) according to,

$$\begin{aligned} H_j &= -\sum |S_j|^2 \log |S_j|^2 \\ &= -\sum_{k=1}^{n} |x_{jk}|^2 \log |x_{jk}|^2 \end{aligned} \tag{3.1}$$

$$E_j = \int |S_j|^2 dt = \sum_{k=1}^{n} |x_{jk}|^2. \tag{3.2}$$

Changing the state of a machine has usually an influence on the signals on some frequency bands which is indicated by the measurable changed complexity and energy of the frequency bands. The wavelet packet entropy or energy values on different frequency bands construct the feature vector, which reflects the information distribution of signals in frequency bands and used as the input vector to the SVM. Since one sub-signal corresponds to one entropy value or energy value, the advantage of combining WPT and entropy or energy measures is to have the length of the feature vector decoupled from the length of the original cycle. This means that different lengths of input cycles give a constant structure of the feature vector. On the other hand, since one sub-signal corresponds to one entropy or energy value, the feature vector contains only the information on frequency bands, i.e. these two methods only focus on describing the change in frequency domain. For example, it could happen that two signals do not have the same form but the same entropy or energy values. In order to solve this problem, a method named "feature-based resampling" is proposed in this work to overcome the deterioration in accuracy in case of homogeneous resampling. The method finds a set of segment dividers to apply resampling individually on the segments.

3.3.4 Feature-based resampling method

In the case that two signals do not have the same form but the same entropy or energy values, difficulties in generalizing the solution with respect to the reliability of such high level of signal abstraction occurs.

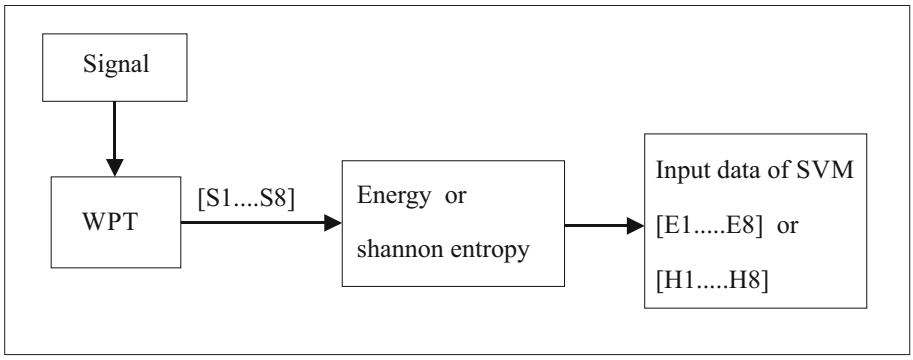


Figure 3.10: The WPT energy and entropy

In Fig. 3.8 an example of the measured data set of vibration acceleration is presented. It shows a general condition that the real measured signals, which have similar characteristics and type features, are usually in different lengths. This means different lengths of feature vectors leading to difficulties in constructing the input matrix of the SVM algorithm. In the direct application of extracted segments and the WPT, this problem is solved by the zero padding and the entropy and energy measures, which leads either to deterioration of the generalization and robustness of the solution or loss of system information required for the classification.

In order to build the feature vector of the system, the direct application of signal characteristics (signal peaks) leads to difficulties because of the high level of noise and disturbance. Therefore, a resampling based on more reliable signal features (nodes) is necessary to normalize the lengths of the cycle segments.

In the case of WPT and Discrete Wavelet Transform (DWT), the length of the decomposed signals is related to the length of the original signal. As shown in Fig. 3.11, by using a 3 depth DWT, the original signal will be decomposed into the high frequency part D_1 and the low frequency part A_1 and then the approximation part A_1 is further decomposed into A_2 and D_2, and so on. By reconstruction of the approximation A_3 and detail parts D_3, D_2 and D_1, a group of sub-signals (S_1...S_4) on different frequency bands is generated. As the noise and disturbance are random, selective reconstruction of sub-signals in DWT can be applied to denoise the signal and detect the reliable features for resampling. Another possibility is to detect these features in the reconstructed signal from the selected components of the DWT and divide them into segments (6 segments in Fig. 3.11) and considered as the keys for the process of resampling (Fig. 3.11). Here three level Daubechies db4 wavelet is used in the DWT decomposition. In this way the time scale of the signal is changed and the corresponding information related to the time is disturbed. To solve this problem, the original positions $P = [p_1, ...p_n]$ of the discrete points in the reconstructed signal are stored in a matrix added to the value matrix $W = [w_1, ...w_n]$ to generate the feature vector T of the SVM classifier as

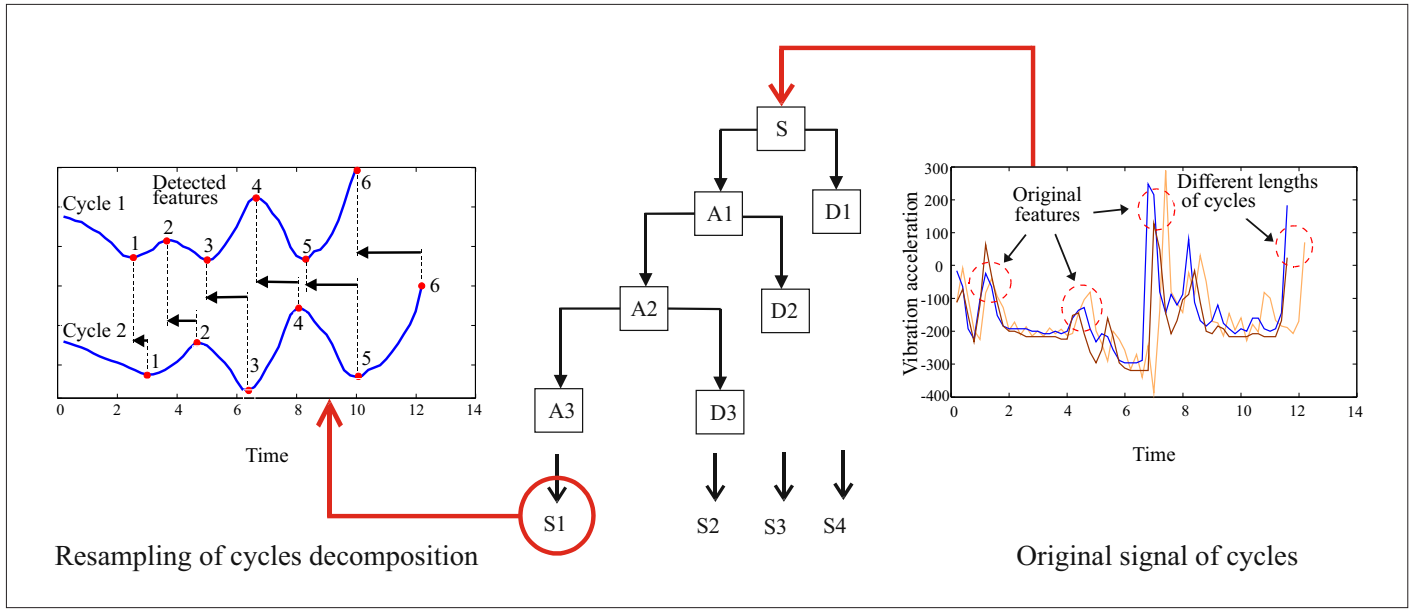


Figure 3.11: Feature-based resampling method

$$T = [W, P] = [w_1, w_1.....w_n, p_1, p_2.....p_n]. \tag{3.3}$$

3.4 Classification

A support vector machine classifier is applied to classify the states of the presented experimental example. The learning task of SVM (refer to Section 2.4.3) is to find the dependency mapping between the feature matrix containing the extracted relevant features and the system state vector using the concept of maximum margin for better generalization. A decision function is used to classify the unknown data points according to the position and distance from the separating hyperplane.

In SVM problems, model selection comprises selection of kernel, kernel parameters, and penalty parameter C. If the number of features is large, as in the current case, one may not need to map the data to a higher dimensional space, due to the nonlinear mapping which does not improve the performance. As observed in the current problem, using the linear kernel is good enough compared to other kernels used, and one only searches for the parameter C. A cross validation is done to estimate the best values of the penalty parameter C. The goal is to identify good value to C so that the unknown data (testing data) can be classified accurately, and not only the training data which comprises overfitting, therefore a 10-fold cross validation was done to predict the value of C. Additionally, a simple scaling is done to the data to avoid numerical difficulties during calculation. Scaling is applied to help preventing the domination of greater numeric ranges attributes on the smaller numeric ranges ones.

Two classes are defined to train the SVM classifier; the first one is the state *before* wear part change (old part), and the second is the state *after* wear part change (new part). A training data of 200 cycles, 100 cycles each class, were taken randomly from 4 places of the data. A linear kernel is considered because of the high number of attributes. The test set has a size 15874 cycles. The algorithm of Libsvm [CL11] is used for training and classification.

3.5 Results and discussion

The results of the different approaches considered for solution are presented and discussed in the following sections.

3.5.1 Direct application and homogenous resampling

In this section, the results are presented for the method of direct application using zero padding and the method of homogenous resampling, which is a general resampling of cycles without considering the segments. Two groups of data are investigated: data segment 7 and data segment 12-14, with fifteen combinations of the four sets of measurement data; vibration acceleration, system pressure, vibration velocity, and displacement. Each test set has a size of 15874 cycles. Using the algorithms of Chang and Lin [CL11], the task was to solve the classification problem with a linear classifier, two classes, 200 cycles training data, 15874 cycles test data, and number of attributes 100, 200, 300, 400, 600, and 800 attributes.

Results of possible sensor combinations are summarized in Table 3.1 and Table 3.2 for data segment 7 and data segment 12-14 respectively. The accuracy in the tables represents the percentage of the correctly classified test measurements (cycles) in the test set of 15874 cycles. In general, data segment 12-14 has better accuracy with maximum accuracy of 97.76 percent than data segment 7 with maximum accuracy 60.71 percent. This is because data segment 12-14 comprises a sensor signal without disturbance from material type and size to be compressed.

Table 3.1: Results for data segment 7

Signal combination	Attributes	S.V.	Accuracy (Perc.)
A	100	20	56.94
V	100	11	60.71
P	100	14	56.15
D	100	9	54.03
A-V	200	19	60.46
A-P	200	14	56.49
A-D	200	10	56.06
P-V	200	9	58.79
V-D	200	9	57.76
P-D	200	9	54.47
A-P-V	300	16	58.59
A-V-D	300	9	56.87
A-P-D	300	10	55.98
P-V-D	300	9	57.71
A-PV-D	400	9	56.84

In order to understand the effect of combining individual signals, the combination accuracy plot is proposed to show the possible combinations and their accuracies.

Table 3.2: Results for data segment 12-14

Signal combination	Attributes	S.V.	Accuracy (Perc.)
A	200	22	94.54
V	200	21	96.96
P	200	70	91.56
D	200	63	93.66
A-V	400	25	97.36
A-P	400	24	95.53
A-D	400	21	96.06
P-V	400	21	97.14
V-D	400	22	97.34
P-D	400	52	95.41
A-P-V	600	26	97.44
A-V-D	600	25	97.67
A-P-D	600	26	96.23
P-V-D	600	22	97.44
A-PV-D	800	25	97.76

The figures, Fig. 3.12 and Fig. 3.13 illustrate the combination accuracies of segment 7 and segment 12-14 respectively. These plots help to recognize the effect of combining sensor information, whether positive or negative, on the accuracy of the classification system. They can also be used to follow a certain sensor and to realize the efficiency of adding different sensors to it. Studying the two figures, Fig. 3.12 and Fig. 3.13, the following observations can be realized: The best single signal accuracy in both groups is the results from the vibration velocity. A possible reason is the previously mentioned effect of wear frequency and the sensitivity of vibration velocity measurements to wear problems. The second best single signal accuracy is the vibration acceleration.

It can also be seen from the two figures that the accuracy is negatively affected by combining sensors in segment 7, while it is positively influenced by combining sensors in segment 12-14. The reason is the high level of noise in segment 7. Adding sensor information can lead to accumulating noise which would deteriorate the accuracy of classification.

The best accuracy in both segments is the combination of all signals in data segment 12-14 with an accuracy of 97.76 percent. In addition to the avoidance of material-related disturbance, the following aspects should be noted:

- The higher dimensional feature space allows better description of the data.

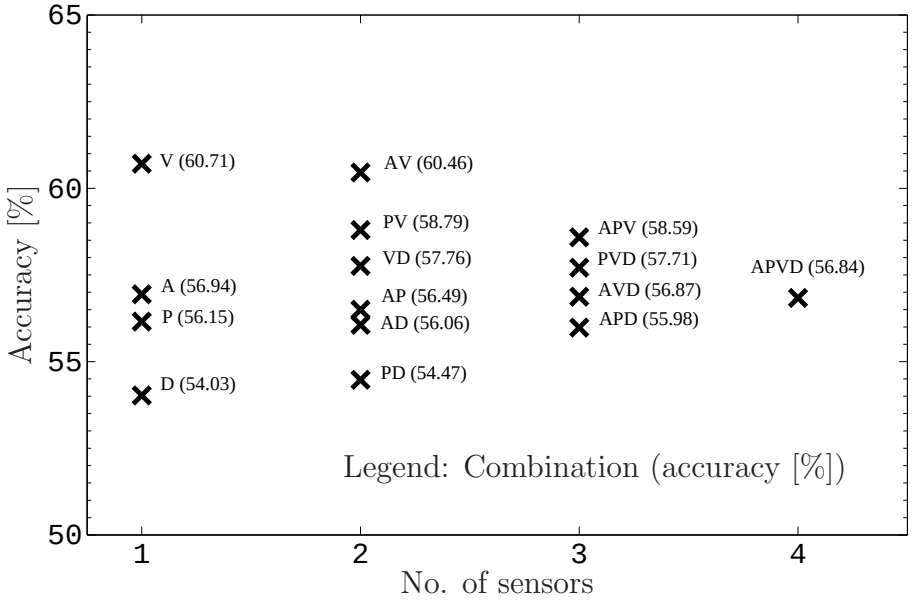


Figure 3.12: Combination accuracies for data segment 7 [SS10b]

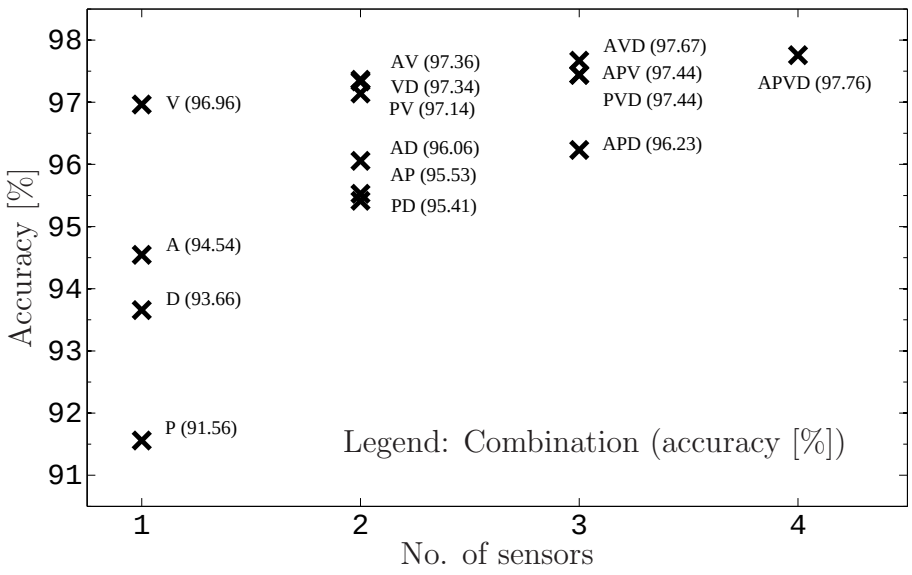


Figure 3.13: Combination accuracies for data segments 12-14 [SS10b]

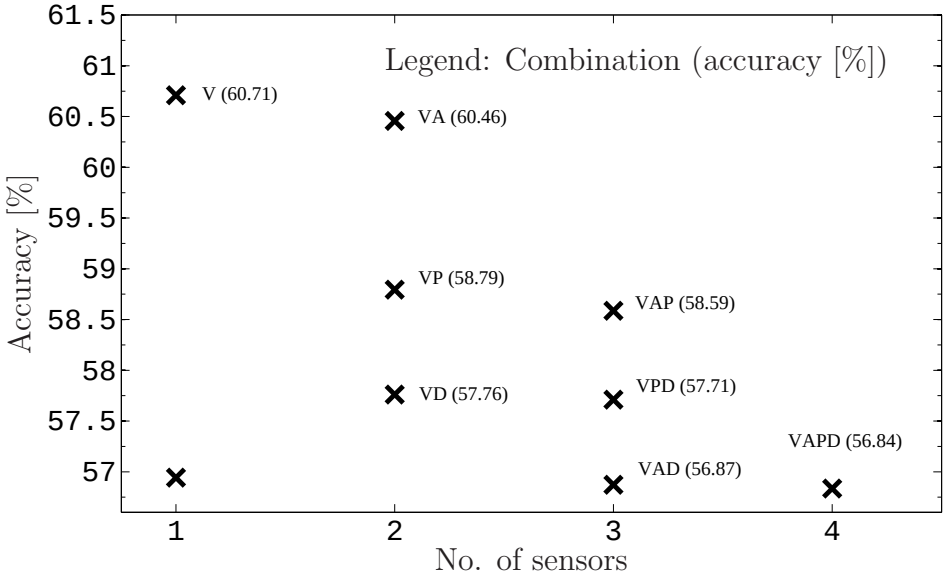


Figure 3.14: Combination accuracies for vibration velocity in data segment 7 [SS10b]

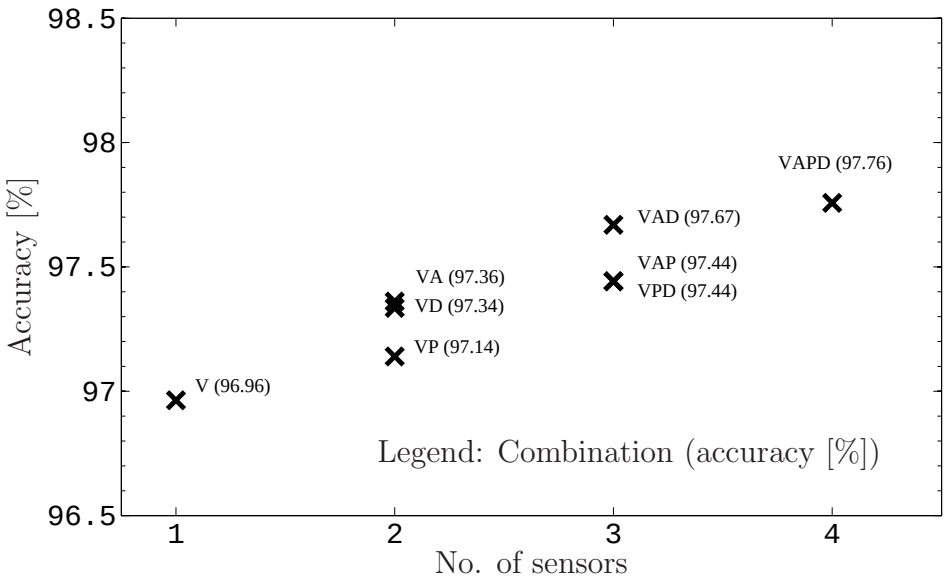


Figure 3.15: Combination accuracies for vibration velocity in data segments 12-14 [SS10b]

- The combination of signals can give better information than individual signals.
- All sensor signals (and accordingly all the related information) are integrated in the combination of best accuracy.

This is not the case with the results of segment 7 where the disturbance confused the results of the signal combinations.

To study the effect of combining a specified sensor with other sensors the vibration velocity is taken as an example to be followed in both segments. Accuracies of vibration velocity combinations are shown in Fig. 3.14 and Fig. 3.15 for segment 7 and segment 12-14 respectively. In addition to the observations and arguments mentioned with Fig. 3.12 and Fig. 3.13 , it can be seen that combining a specified signal with a first signal of better accuracy than a second one gives better accuracy than combining the specified signal with the second signal.

3.5.2 Remaining life prediction

An example of the decision value function of the SVM for the sensor combinations (APVD, direct application) is shown in Fig. 3.16. The figure presents the classification results of the testing data, which include wear plate replacement (indicated in the figure). The decision boarder (decision value 0.0) separates two classes (Before and After plate change) and the operation cycles are classified accordingly, based on the information provided by the extracted features. The decision boarder represents the separating hyperplane of SVM. Consequently, the decision value is a measure of how far a point (cycle) is from the separating hyperplane. It is an indication of how much a cycle's information contributes to a class. In addition, the running-in period of the new material is recognizable in the figure.

A longer test dataset of 241034 cycles (average period of 1.5 year of machine operation) is considered in Fig. 3.17; here the smoothed function is presented. The resulting decision value function and its smoothing are shown in Fig. 3.17. Smoothing of the function shows the increasing tendency for the average signal over time. Briefly described, the two classes become closer over time. This gives an indication about the remaining life time of the part at any time point of operation. Accordingly, a related change index (CI) is proposed based on the decision value of the SVM to evaluate the remaining life time. The CI shows a tendency to change over time coinciding with the deterioration of the part and the remaining life time. The CI in Fig. 3.18 is a measure function of the distance between the current position of the decision value of a measurement and the maximum allowed position of the decision value before changing of the wear part is necessary. The CI is not only an indication of how "worn" a part is, but it can give indication of the remaining life time of the part, which adds a value to the maintenance planning of the machine.

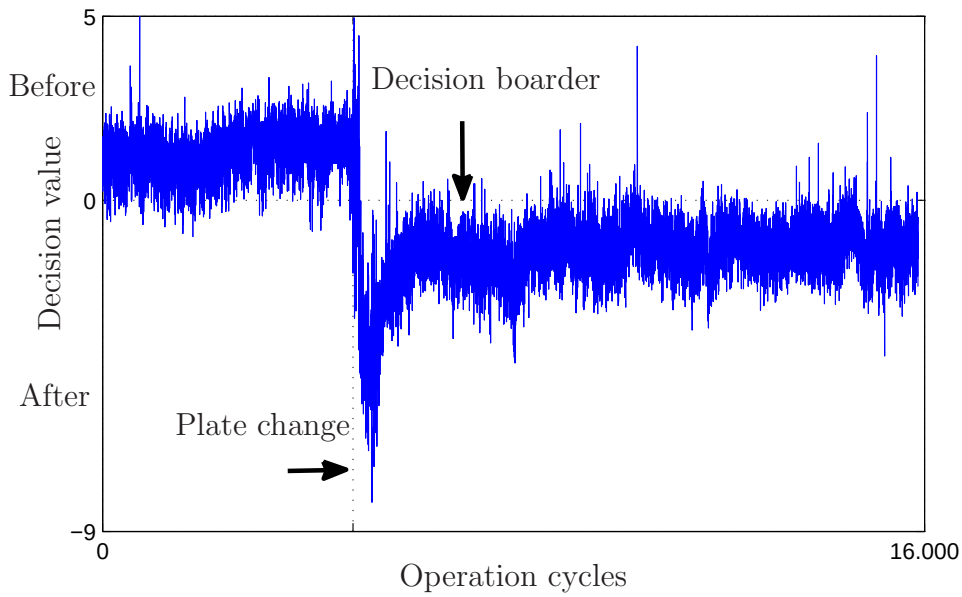


Figure 3.16: Decision value of the signal combination for data segments 12-14 [SS10b]

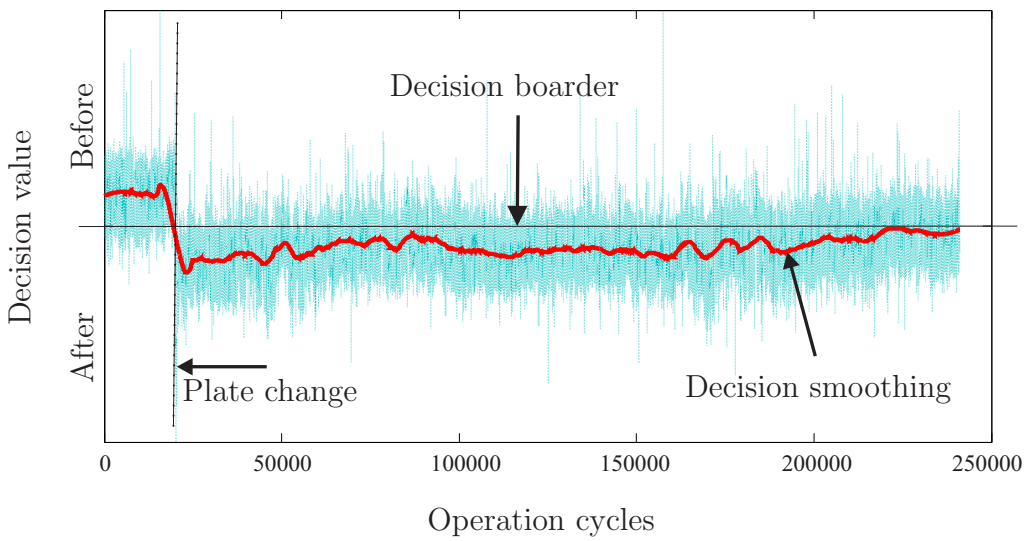


Figure 3.17: Decision value of the extended test data and its smoothing

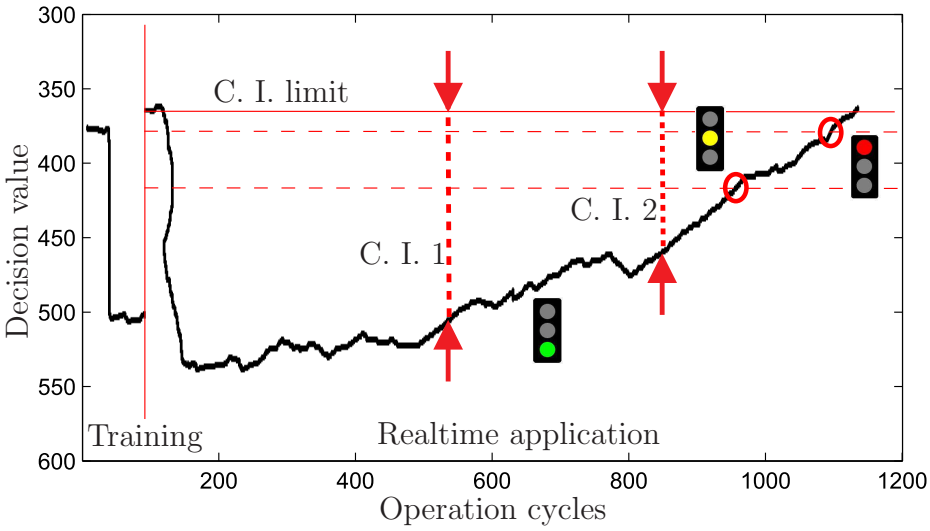


Figure 3.18: The CI based on the decision value [SS10b]

The traffic light indicator in the figure (Fig. 3.18) is introduced to provide a quick indication about the general state of the machine part. This quick indication is corresponding to the state ranges in the decision value which are determined by decision thresholds set by training data. Decision thresholds and the Change Index (CI) are used in parallel for better indication of critical situations, as illustrated in Fig. 3.19 as an industrial implementation arrangement.

3.5.3 WPT and feature-based resampling

The resulting classification accuracies for the four machine parameters (A, P, V, and D) and their possible combinations, using the different approaches discussed above are presented in Table 3.3. The classification accuracy presented in the table represents the percentage of the correctly classified test measurements (cycles) in the test set of 15874 cycles. As mentioned before; the results of the direct application are added only for comparison because of the lack in the solution flexibility which results from the zero padding. In general, it can be seen that the results of the feature-based resampling are better than those of entropy and energy WPT. Additionally, the combination of signals improves the accuracy and the highest combination accuracy (P-D: 98.56%) is achieved by using the feature-based resampling approach.

For the results of feature-based resampling method, it is necessary to select the best combination of the DWT decomposition sub-signals, which gives the best accuracy. The resulting classification accuracies for the possible combinations of the DWT sub-signals (i.e. Approximations (Ax) and Details (Dx)) extracted from the four machine signals (A, P, V, and D) are presented in Table 3.4. The results are based

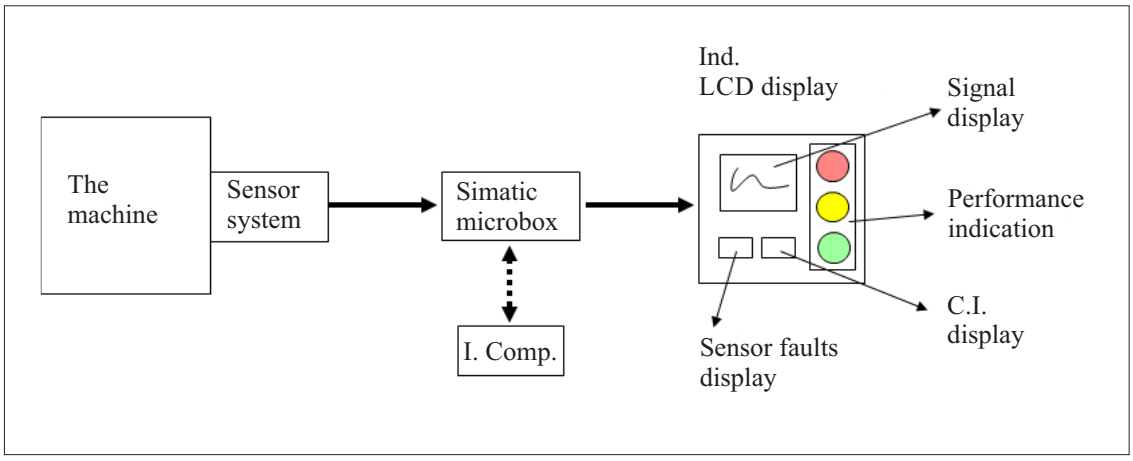


Figure 3.19: Industrial implementation arrangement

Table 3.3: Classification results of the different methods

Signal	Direct app.	WPT-Entropy	WPT-Eneregy	Feature-based R.
A	94.54	94.42	95.75	94.35
P	91.56	83.89	89.05	97.50
V	96.96	94.67	94.65	96.79
D	93.66	88.78	92.89	95.17
A-P	95.53	92.68	93.64	94.67
A-V	97.36	93.17	90.15	96.42
A-D	96.06	92.26	94.48	94.85
P-V	97.14	93.23	93.32	96.80
P-D	95.41	89.51	91.68	98.56
V-D	97.34	92.78	95.17	97.11
A-P-V	97.44	93.44	90.15	96.87
A-P-D	96.23	92.18	92.96	94.95
A-V-D	97.67	93.02	94.29	97.00
P-V-D	97.44	92.82	95.24	97.02
A-P-V-D	97.76	94.53	95.48	97.32

on the feature-based resampling approach with and without consideration of the original positions after resampling. Original positions are considered by connecting the two matrices of the resampled cycles and the original positions of the key features together. It should also be mentioned that the connected position matrix in Table 3.3 is optimized by a weighting factor in the feature matrix in order to improve the accuracy of the solution. No weighting factor was applied to the results in Table 3.4.

It can be seen from Table 3.4, that the accuracy of classification is improved by

considering the position information. It can also be seen that the optimal level for the classification (DWT combination) is different from machine signal to another. A combination of four sub-signals (A3, D3, D2, and D1) gives the best accuracy of the vibration acceleration (A), whereas only one sub-signal (A3) gives the best accuracy of the displacement (D). Moreover, the accuracy of a machine signal is influenced by the selection of particular signal approximations (Ax), as well as signal details (Dx). This is illustrated by comparing the different sub-signal combinations of a particular machine signal. This means that the extracted segment nodes, which not necessarily depend on some underlying system dynamics, are not only based on approximation indicatives. They are also based on details as well.

Table 3.4: Classification results of DWT components

Signal	Combination	Original position not considered	Original position considered
A	A3	83.77	83.85
	A3+D3	87.63	87.75
	A3+D3+D2	90.31	90.39
	A3+D3+D2+D1	90.67	90.70
P	A3	92.33	93.15
	A3+D3	94.10	94.21
	A3+D3+D2	93.48	93.81
	A3+D3+D1	92.93	93.31
V	A3	94.00	94.01
	A3+D3	94.25	94.27
	A3+D3+D2	94.37	94.38
	A3+D3+D2+D1	94.39	94.48
D	A3	93.97	95.17
	A3+D3	92.72	93.03
	A3+D2	91.57	91.78
	A3+D1	91.95	92.20

3.6 Conclusion

An approach for developing a fault diagnosis and prognosis system to support condition-based maintenance of wear parts is presented. The system is used as a prewarning module to detect the necessity for replacing wear parts of production machines and to evaluate the remaining life time of the supervised part. The proposed system uses SVM classification as a signal-based diagnosis technique and

as a feature fusion tool. Alternative combinations of fusing sensors are taken into consideration to define a complementary sensor array for better accuracy. Relevant features are extracted using wavelet transform techniques with the aid of energy and entropy measures. Measurements were processed to select the most reliable combination of characteristics to depend on. A new approach for feature-based resampling is presented based on multi-level wavelet decomposition. A related change index is proposed based on the decision value of the SVM to evaluate the remaining life time.

By applying the proposed method of feature-based resampling a considerable improvement in the accuracy and robustness of the solution is achieved.

4 Wavelet-SVM system for object detection

In this chapter different kind of incidents considered for recognition are investigated. The individual events with no consequent and accumulating effects investigated in this work represent many possible process and fault related incidents. The signals encountered are characterized with impulsive one-time events representing the goal object. Another characteristic of the sensor signals is the partly simultaneous stimulation of events which prevents the advantage of using exact simultaneous events for fusing the data in the features level, which requires the use of suitable decision fusion techniques. A real industrial process generating real sensor signals are considered in this chapter. However, in principle the developed approaches can include arbitrary data acquisition techniques.

During the progress of this work, some scientific papers were published based on the studies and results of the work. This chapter is prepared to a large extent based on these publications [ASSS14, ASSSS12, ASSS11, ASSSS10].

4.1 Introduction

The automation of the industrial processes requires suitable supporting systems which include relevant process monitoring systems. Individual process-related variables are usually considered and thresholds are defined and used to distinguish regular and irregular operations. Accordingly, the regular and irregular conditions of the mechanical system are specified and defined. Such methods are not generally applicable for highly complex systems because the multidimensionality and interrelations involved cannot be handled by low-dimensional approaches that use classical thresholds. Model-based monitoring systems are often not suitable for complex systems because precise models of the mechanical system considered are required for reliable monitoring [AS11]. Model-based methods usually require complex modeling of the process with detailed process parameters and additional information on changes in the system states.

Signal-based diagnostic methods are based on an analysis of measured (physical) signals. They are useful when the measured variables contain direct or implicit information about possible faulty behavior. Signal-based diagnostic methods are easy to use and are widely adopted to extract relevant process characteristics from analyzed sensor data in combination with further knowledge. Feature extraction can be performed in either the time or frequency domain of the signal. The extracted features should be able to represent the regular state of the system, as well as non-regular behaviors; in other words, they should indicate changes in system states. Thus, signal-based methods when combined with machine learning techniques can be used to distinguish system states. Depending on the machine and

process complexity, suitable sensors have to be used to define suitable mappings between machine operating states and sensor data.

4.1.1 Scope of work and study objectives

In this work, a new monitoring approach is developed. The developed approach is inspired by the monitoring task of production processes that suffer from various drawbacks of approaches applied before [ASSS11, Nie09]. In contrast to the input signals of the previous chapter, the signals encountered in the system of this work are characterized with nonstatinary impulsive one-time events representing the goal object. Another characteristic of the sensor cluster signals is the partly simultaneous stimulation of events which prevents the advantage of using exact simultaneous events for fusing the data in the features level, which requires the use of suitable decision fusion techniques. Real industrial process and sensor signals are considered in this work. However, in principle the developed approaches can include arbitrary data acquisition techniques.

The goal of the project is to detect the presence of target objects within the material transported during a production process [ASSSS12]. For such specific applications, the presence of target objects in transported material (overburden) has to be detected (target object present yes/no) to avoid resulting disturbances and failures during the continuous transportation process. The goals were to reduce the rate of false detection and obtain reliable decisions on the presence of target objects. The monitoring system uses acceleration sensors and is used as a sensor-cluster. The approach developed must be practical and suitable for real-time use for standard industrial hardware.

For the purpose of comparison, this work investigates two new multisensor data fusion algorithms for object detection in monitoring of the industrial process. Two approaches were proposed [ASSS14]. The first uses a short-time Fourier transform (STFT) as a prefilter to extract relevant features from the acceleration signals. The features extracted from different sensor channels are first classified using support vector machine (SVM)-based filters. A novel decision fusion process to combine individual decisions was developed. The second approach uses a continuous wavelet transform (CWT) as a prefilter to extract relevant features from the acceleration signals. The features extracted from different sensor signals are subjected to further prefiltering processes before SVM-based classification. The individual decision functions are then combined in a decision fusion module. The classification system was trained and validated using real industrial data. The two approaches were tested using the same data and their performance and modeling complexity are compared. The developed approaches show strong improvements in detection and false alarm rates.

4.1.2 Application state of the art

A few studies have considered the development of monitoring systems for similar production processes. Petrich and Köhler [PK05] developed a monitoring system based on georadar to detect target objects in overburden before excavation. A detection rate of less than 60% was achieved. Their system was not able to distinguish non-critical objects such as frozen overburden from the target objects of interest. Such objects lead to false alarms, disturbing the production process. Their system could also not identify the position of target objects accurately. Petrich and Köhler also installed a radiometric measuring system above a transport belt to detect target objects in the material flow. A detection rate of less than 70% was achieved. Changes in the petrography and elemental composition of the overburden, other objects such as clay chunks, and small objects led to false alarms.

Nieß developed an automatic monitoring system to detect target objects in a material flow using acceleration sensors [Nie09]. Several acceleration sensors were mounted in the area of impact along the production line. The amplitude of the acceleration signals and vibration durations were considered to determine the presence of target objects. A detection rate of approximately 75% was achieved. However, the production process was often disturbed by false alarms for this system [SMS07].

4.2 Description of target system and solution concept

Bucket-wheel excavators are used by RWE Power AG Company for lignite mining in Garzweiler, Hambach and Inden opencast mines (Fig. 4.1). These excavators are utilized to extract the coals as well as to remove the overburden that located above the coal layers.

4.2.1 Description of production process

The excavated materials by the paddle wheel is continuously discharged and transported onto a conveyor belt to another place. The excavators often excavate stones of various sizes as well as other objects, e.g. metallic objects that are often included in the materials flow. The detected outsized objects and stones are removed from the conveyor of the bucket wheel and collected for removal by heavy trucks. Undetected large stones disturb the work of the machine and cause damages (such as longitudinal cracks or punctures of the straps of transport system and damage of the supporting structure) or shutdown of the production system leading to reduction of the production capacity, and increase in the operating costs. The goal objects to be detected (stones that cause disturbance) are set to be recognized with regard to their geometric dimensions. The stones with size of greater than 600 mm are recognized for detection as goal objects (Fig. 4.2).

Figure 4.1: Bucket-wheel excavator

4.2.2 Measurements and input data

In Fig. 4.3, the location of different sensors along the transport belt is shown. Because of the nature of the production process, the efficient use of the measured data encounters some difficulties and obstructions requiring more creative solutions for the developed pattern recognition system. Due to the geometrical distribution of the used sensors, an inevitable and varying time shift between the stimulations of the individual sensors exist. This prevents the advantage of using exact simultaneous events for fusing the data in the features level. Additionally, the operation conditions of the machine and the connections of the machine structures produce impacts and vibrations stimulating the sensors in an event-similar way, which disturb the procedure of recognizing the goal objects. Another difficulty arises from the process of transporting the overburden and the possible accompanying goal objects. The goal objects (stones to be detected) are transported buried to some extent in the overburden. The intensity of the event stimulated by the goal object is connected with the degree of damping by the overburden. It could happen that a small stone induce strong stimulation whereas a big stone induce a strongly damped stimulation. This behavior requires a recognition system which is scale invariant in order to have efficient performance of the system.

4.2.3 Concept of solution

Owing to the complexity and variety of detection schemes in production processes, here is assumed that the task cannot be solved satisfactorily using just one sensor technique. Moreover, due to the complexity of the transportation process, no single

Figure 4.2: Target object on the transport belt [SAS12]

sensor technique can achieve the task directly. Thus, it is concluded that only indirect measurements or suitable combinations of individual measurements of relevant events (physical effects) of transported objects can be performed. The individual signals measured have to be preprocessed, evaluated, and combined to develop a reliable monitoring system to determine the presence of target objects.

One of the physical effects considered is force, which induces impact responses during transportation of target objects. These responses should be measured by sensors (five acceleration sensors here) in the impact area along the transportation line. These signals were investigated to build a detection module (an acceleration module here).

The inevitable time shift between the object impact stimulations of the individual sensors varies making the fusion of the process information difficult. Therefore, signal preprocessing, feature extraction, and classification for the individual acceleration sensors were used. Then a decision fusion process based on specific decision criteria was applied to combine the preliminary individual decisions of different classifiers (target object present yes/no) (Fig. 4.4).

Two different detection approaches for the acceleration module were investigated. The first uses a short-time Fourier transform (STFT) as a prefilter and a support vector machine (SVM) as a classifier. The second approach uses a continuous wavelet transform (CWT) as a prefilter and SVM as a classifier. The decision fusion process in both approaches is realized using different criteria. Both approaches are described and discussed in the following sections.

4.3 Approach I: Detection system based on STFT and SVM

In the STFT-SVM approach, STFT is used to extract relevant information from the signals obtained from different acceleration sensors. The SVM is then used to clas-

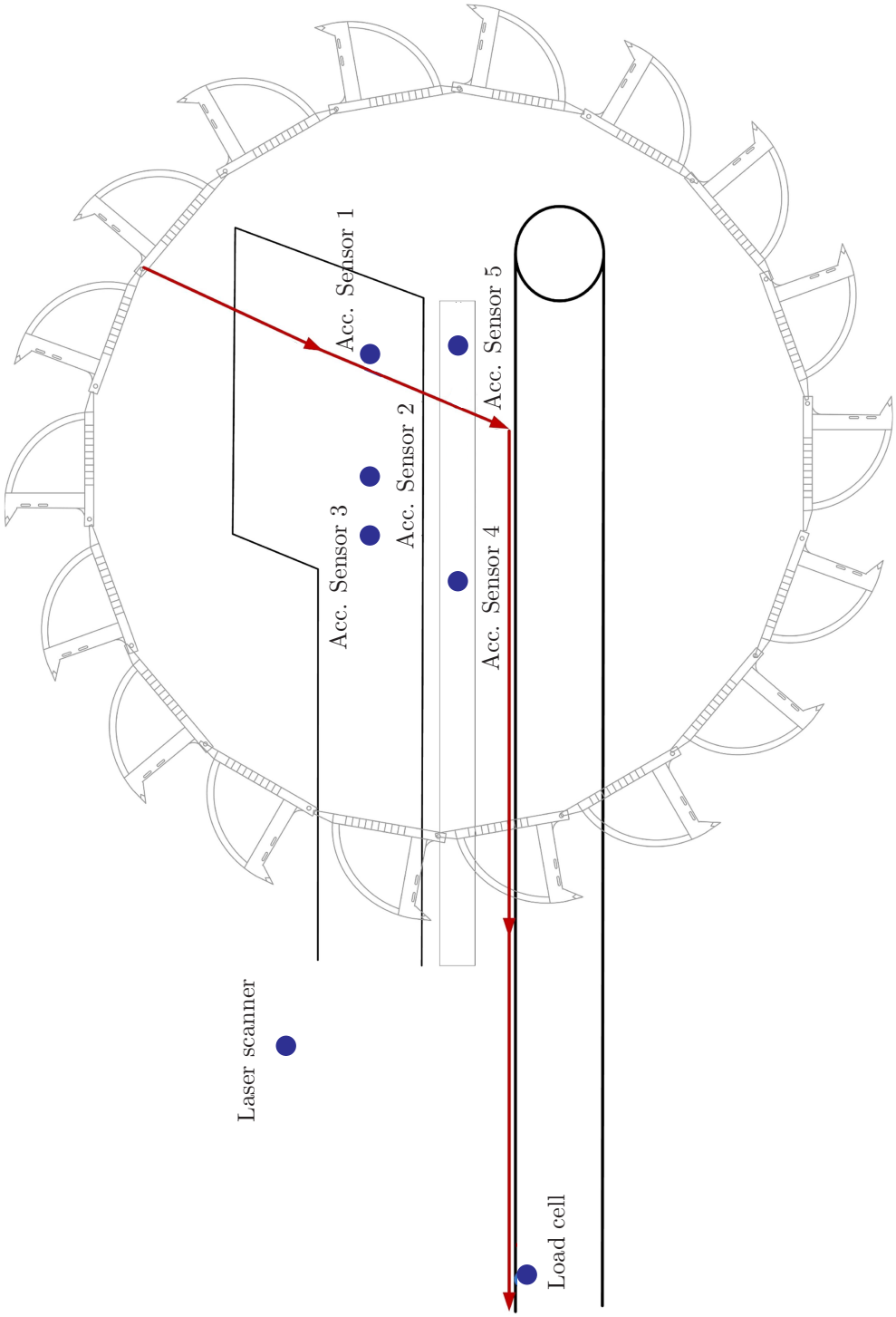


Figure 4.3: Position of different sensors along the transport belt [SAS12]

sify the extracted features. In addition, a specific fusion process based on SVM and experimentally-based decision rules is developed and applied to combine the prelimi-

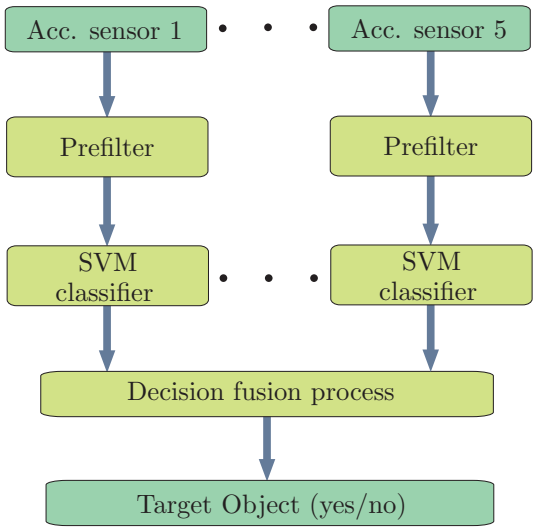


Figure 4.4: Detection approach based on STFT and SVM [ASSS14]

nary decisions of the individual sensors. The signals are individually prefiltered with STFT (Fig. 4.4 and Fig. 4.5). This prefiltering process is used to extract relevant features of the acceleration signals (Section 4.3.1). A set of supervised classification filters, denoted SVM I, is developed to classify the features extracted from each sensor signal. An adjustable decision fusion process is developed to combine the preliminary individual decisions of the different classifiers (Section 4.3.3). Feature extraction, classification, and decision fusion processes are described in detail below.

4.3.1 STFT-based feature extraction

The STFT extracts relevant information on system states. It serves to classify the information related to a single information path based on previously observed phenomena for a sensor signal. Fig. 4.6(a) (raw signal) shows a raw sample acceleration signal during the production process. At 12.5 s, a target object was manually classified. The impact of this object results in strong acceleration signal amplitudes. The other peaks in the signal are caused by other unknown events and are not object-induced signal changes. From the raw acceleration signals in the time domain it is very difficult to distinguish target object's effect from other unknown events. The different events can be classified on the basis of features extracted using prefilters (STFT). As shown by the spectrogram in Fig. 4.6(a), the target object causes strong excitation of low frequencies. By contrast, higher frequencies are excited more by unclassified or unknown events. This effect is used to distinguish events due to target objects from those related to other events.

4.3.2 Classification process

Three classification modules are included in the detection system (Fig. 3). The SVM-based algorithms are used to detect the system states. The Libsvm algorithm [CL11] is used to realize the SVM classifiers.

4.3.2.1 The Module SVM I

Due to the inevitable time delay between the excitation of the sensors which is constantly varying as a result of the structural dynamical behavior between the impact's and the sensor's locations. This affects the feature vectors. A classifier based on the SVM algorithm (SVM I) is developed for each individual acceleration signal. Data clearly indicating the presence of target objects are used to train SVM I. This should limit the false alarm rate, which is directly affected by the strength and intensity of the indicators used for training. Weak indicators lead to a higher rate of false alarms and vice versa. The decision functions of the individual classifiers generated by SVM I are input as preliminary decisions into subsequent stages to confirm the assumed system state.

4.3.2.2 The Module SVM II

When SVM I does not provide sufficient information, the decision on the presence of target objects is uncertain. In such a case, and in the cases mentioned in the following sections, a more local and precise investigation is necessary. Data with weak indications of the presence of target objects are used to train module SVM II; the limited area of application of the signal allows more flexible detection criteria.

The SVM II is trained with data consisting of two states. State 1, "target object present", is represented by training data with weak indications of the presence of target objects. State 2, "uncertain", is represented by training data with all other indications except state 1. The output statement of SVM II is either "target object present" or "uncertain".

4.3.2.3 The Module SVM III

A classification process based on SVM III is performed in cases for which SVM II provides an uncertain output statement. The SVM III provides further data classification for uncertain output statements. The SVM III is trained using data with clear indication of no target objects (state 1) and data with uncertain indications (state 2). The output statement of SVM III is either "target object not present" or "uncertain".

The SVM II and SVM III classifiers are used for more accurate trained classification locally in cases where further assessment of unclear decisions is required.

4.3.3 Adjustable decision fusion process

A new decision fusion process was developed to combine individual preliminary decisions and generate a final decision on the system state. The decision method is based on knowledge derived from analysis of experimental data from different acceleration sensors. It is designed to obtain the highest possible detection rate for the lowest possible false alarm rate. Therefore, tuning parameters are used to systematically adjust the fusion process.

The decision fusion filter consists of two rule-based filter levels and classification levels SVM II and SVM III (Fig. 4.5).

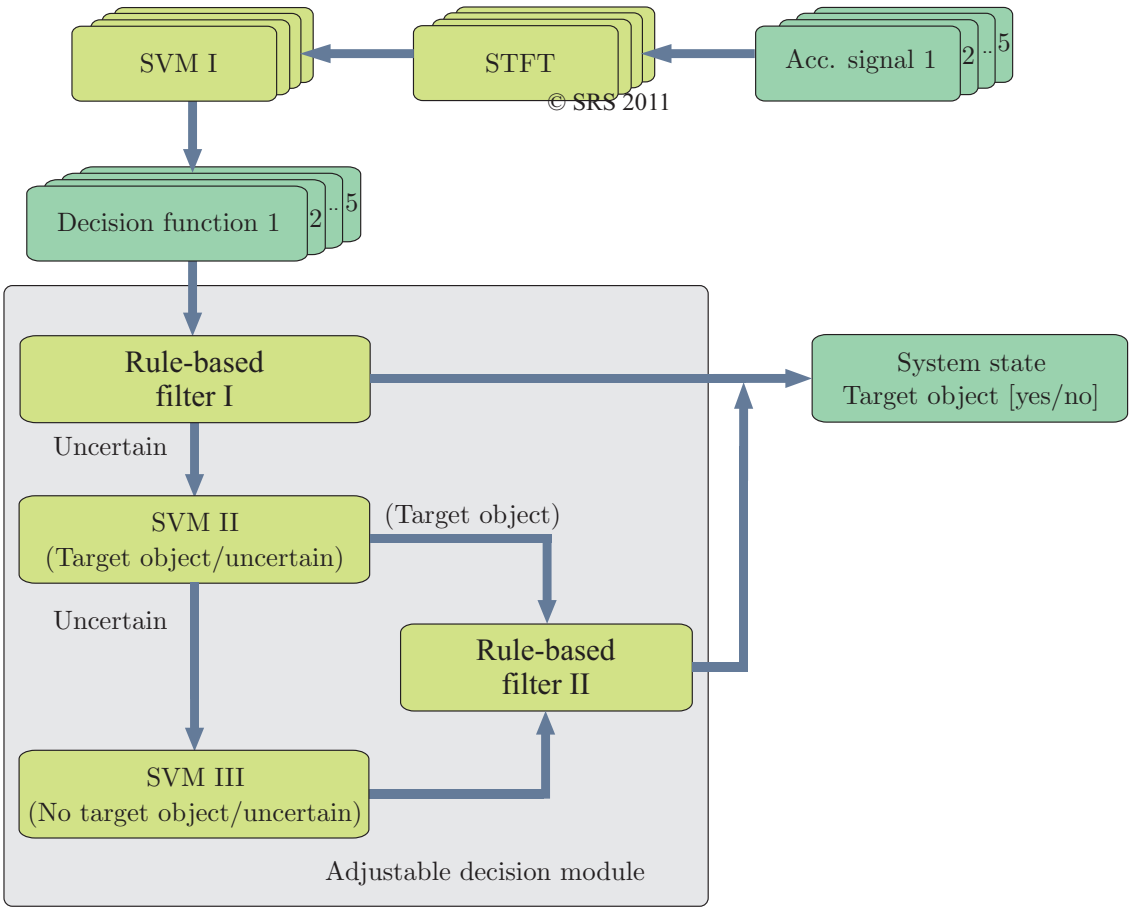


Figure 4.5: adjustable decision fusion process [ASSS14]

Classification level SVM I provides preliminary decision functions for the individual classifiers (Section 4.3.2.1). Depending on the individual preliminary decisions, the final decision of the acceleration module could be met by either rule-based filter I or by rule-based filter II.

The rule-based filter I consists of predefined rules that govern the final decision of the acceleration module. When it is impossible to achieve a reliable final decision from the acceleration module based on current values of the individual preliminary decision functions, specific rules to trigger further classification levels (SVM II and SVM III) are included in rule-based filter I. The rules for rule-based filter I are as follows:

- *Rule I*: At least two simultaneous positive individual decisions lead to a final decision of "positive: target object present".
- *Rule II*: Weak individual positive decisions with a decision value less than the experimentally defined threshold value T_1 are ignored, leading to a final decision of "negative: no target object present".

  Experimental evidence shows that weak individual positive decisions (with decision values less than T_1) are generally caused by events denoted as no target objects. Therefore this kind of decision is ignored and not considered as an indicator of a target object.
- *Rule III*: Individual positive decision with a decision value greater than the experimental threshold T_2 ($T_2 > T_1$) leads to a final decision of "positive: target object present".
- *Rule IV*: Individual positive decision with a decision value greater than T_1 and less than T_2 triggers further classification levels.

As mentioned before, classification levels SVM II and SVM III are activated if rule IV is fulfilled. When an individual positive decision has a decision value greater than T_1 and less than T_2, the other four acceleration signals that provide negative decision values (parallel to the single individual positive decision) are subjected to more accurate evaluation using either SVM II alone or both SVM II and SVM III. Further classification is performed locally on neighboring areas of the corresponding acceleration signals to confirm or disprove the correctness of the single individual positive decision.

The acceleration signals are first evaluated using SVM II. The SVM II output statement is either "target object present" or "uncertain". Thus, SVM II confirms the presence of target objects. The SVM III evaluation is performed for acceleration signals that yield an uncertain output statement from SVM II. These acceleration signals are evaluated for the presence of events denoted as "no object". The output statement of SVM III is either "target object not present" or "uncertain". Thus, SVM III confirms the absence of the target object.

The single individual preliminary decision from SVM I and the other four decisions provided by SVM II, SVM II and SVM III, or SVM III are combined using rule-based filter II. The rules for rule-based filter II are as follows:

- *Rule I*: If the number of output statements "target object present" is greater than the number of "uncertain" statement provided by SVM II and SVM III, the final decision is "positive: target object present" (majority rule).
- *Rule II*: An individual output statement "target object not present" from SVM III leads to a final decision of "negative: no target object".

The decision generated by the fusion module is self-adaptable and depends on re-evaluation of individual partial decisions. Owing to the complexity of the system to be monitored, it is difficult to provide these benefits using classical fusion techniques. The main reason for using the proposed method for fusion is the inevitable and varying time shift between the object impact stimulation of the individual sensors, and the fact that the effect of noise and disturbance signals to the system and sensors (both individual and group sensors) is inevitable. Therefore, the importance of individual decisions is retained by combining and comparing them with other decisions that do not necessarily coincide in time. This data handling requires a floating decision window, which increases the computational load.

The improvement in quality resulting from the developed fusion technique over single-stage SVM classifiers demonstrates the validity of the approach. This multi-stage technique with additional stages focuses mainly on classes that cannot be identified with suitable reliability, and therefore have to be considered separately and in detail.

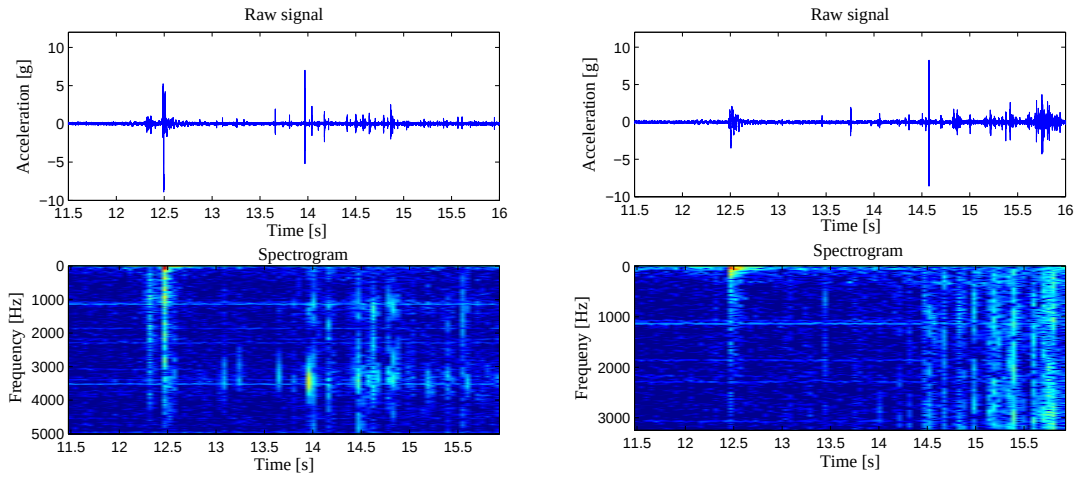
It should be noted that the floating window and decision re-evaluation require buffer savings for the range of data considered. These buffer savings should cover the processing window and transformed data for the different channels and their decisions, as well as decisions generated by the fusion module. These buffer savings increase the memory requirements of the system and the time required for the final decision owing to the inefficient buffering time.

In spite of the complexity of the detection system, the main requirement is that the final decision should at least be faster than subsequent events to provide enough time to isolate target objects that are detected. This requirement is fulfilled according to the implementation results presented below.

4.3.4 First industrial implementation and results

The STFT algorithm is used as a prefilter to extract relevant features from acceleration signals. The spectrograms generated (Fig. 4.6(a) and 4.6(b)) show the features extracted (frequencies in the range up to 3250 Hz) as functions of time. The impulse intensities for each feature are represented by a suitable color map. The feature vector is based on 511 features, one for each acceleration signal.

The spectrogram in Fig. 4.6(a) reveals that frequencies between 1 and 200 Hz are dominant at 12.5 s (red) owing to the impact of a target object. The specific behavior depends on the structural dynamic characteristics of the contact surface on which the sensor is mounted and on the impact position. Higher frequencies at 12.5 s show less energy (light blue) than frequencies below 200 Hz.



(a) Acceleration signal for sensor 1

(b) Acceleration signal for sensor 2

Figure 4.6: Acceleration signal for sensors 1 and 2 [ASSS14]

The amplitude behavior of the acceleration signal at 13.95 s (Fig. 4.6(a)) in the raw time-domain signal is in principle similar to that at 12.5 s; however the range of frequencies excited in the STFT features is different (2560–4200 Hz). Thus, it is experimentally observed that the machine structure reacts differently to target object events and allows a statement about the presence of objects. Therefore, target objects and other events can be detected, classified, and separated. If the resonance properties of the structure at the collision point, taking into consideration the sensor position, are known and are considered to allow reliable distinction between relevant frequency ranges and related impact power, the distinction is considered reliable. In reality, this distinction has been observed as very robust for different sensor locations.

In the raw signal in Fig. 4.6(b), the event observed at 12.5 s and the disturbance effects between 15.6 and 16 s show similar amplitudes and behavior in the time domain. It is expected that the excited structural dynamics at 12.5 s responds differently (in the frequency domain 0–200 Hz (Fig. 4.6(b))).

The approach was tested using an experimental set of real industrial data. The results for preliminary application to the system are summarized in Table 4.1. The

best individual detection accuracy is 58.3% (classifiers 1 and 4). Classifier 5 leads to the lowest accuracy and false alarm rates, although it has the smallest number of support vectors, indicating comparatively low levels of noise. This result demonstrates a typical compromise: an increase in the detection rate leads to an increase in the false alarm rate (Table 4.1).

Table 4.1: Classification results for the STFT-SVM approach

Training data					
Sensor/classifier number	1	2	3	4	5
Target objects	17	19	17	17	16
No. of support vectors	193	210	177	158	79
SVM kernel			Linear		
Individual results for the test data					
Target objects			36		
Objects detected	21	20	19	21	16
Accuracy [%]	58.3	55.5	52.8	58.3	44.4
False alarms	18	14	7	4	2
Fusion results for the test data					
Objects detected			27		
Accuracy [%]			75		
False alarms			7		
False alarms/number of objects [%]			19.4		

The accuracy of the system based on fused decisions for the acceleration sensor network is 75%, which represents an improvement of at least 16.7% over the individual accuracy rates. This improvement in accuracy indicates that the individual sensors have different views, depending on their mounting position and their relation to the materials transported, including target objects. This means that each individual classifier can detect target objects that possibly went undetected by other classifiers. The rate of false alarms can be compromised, however, because the false alarms for individual sensor paths are not necessarily identical. The fusion approach not only improves the detection rate, but also leads to a strong reduction in the number of false alarms (Table 4.1). During development of the detection system, and considering the requirements of the mechanical system, a compromise between accuracy and rate of false alarms must be achieved.

4.4 Approach II: Detection system based on CWT and SVM

The CWT-SVM approach uses CWT to extract relevant information from acceleration signals for the different sensor channels. The SVM is used to classify the

features extracted. A fusion process is applied to combine the individual decisions of the different SVM classifiers (Fig. 4.4).

The individual sensor signals are prefiltered separately using CWT (Fig. 4.4). The features extracted for individual sensors are subjected to multistage filtering. The feature extraction and decision fusion processes are described below.

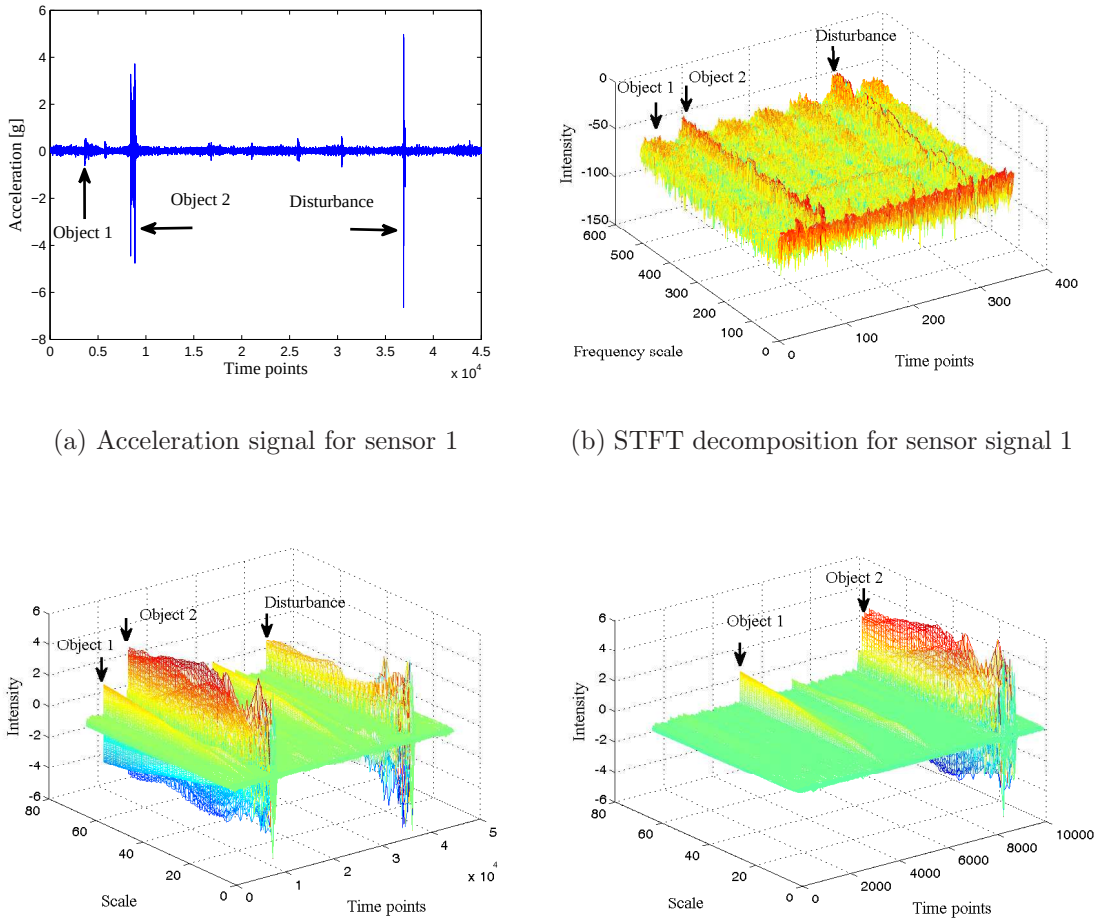
4.4.1 CWT-based feature extraction

As an example, Fig. 4.7(a) shows the acceleration signal for sensor 1. To illustrate the solution concept and for the purpose of comparison [ASSS11], the signal was filtered using STFT (Fig. 4.7(b)) and CWT (Fig. 4.7(c)). The signal has two events, denoted as objects 1 and 2 at time points 4000 and 8500, respectively, that were manually classified as target objects. These events appear at time points 30 and 70 in the STFT extracted feature space (Fig. 4.7(b)) and at time points 4000 and 8500 in the CWT extracted feature space (Fig. 4.7(c)). A third event resulting from an unknown disturbance is evident at time point 37000 in Fig. 4.7(a). Since this event was not classified manually, it cannot denote a target object. It is also evident in the STFT and CWT extracted feature spaces [ASSSS12].

The object at time point 8500 and the event at time point 37000 can be clearly recognized in the STFT and CWT results. In the STFT results the two events seem to be similar, whereas in the CWT results they appear to be different. In the case of CWT, the higher scales (low frequencies) for the second object are excited more strongly than the lower ones, while the lower scales (higher frequencies) are more strongly excited in the case of the disturbing event.

Unlike the case in which a target object is present, the higher scales (lower frequencies) of the disturbance are related to lower energy than the lower scales (higher frequencies). The reason is that the range of the activated frequency and accordingly the center of frequency, which coincides with the energy peak are different for the disturbance and target object. This advantage of the CWT approach is used as a base rule for further filtering steps.

To illustrate this, consider the first target object in the sample data (object 1 at time 4000) that cannot be clearly distinguished from the time series signal (Fig. 4.7(a)). The object is difficult to detect because of its impact on the mechanical structure, which is obviously dampened by the accompanying materials (overburden). In the STFT results (Fig. 4.7(b)), the presence of the object is characterized by weak excitation of low frequencies. In the CWT results (Fig. 4.7(c)), the object can be better recognized and characterized by a longer band of high scales (low frequencies) of a specific shape. The effects of objects 1 and 2 are evident in Fig. 4.7(c) and are magnified in Fig. 4.7(d).



(a) Acceleration signal for sensor 1

(b) STFT decomposition for sensor signal 1

(c) CWT decomposition for sensor signal 1

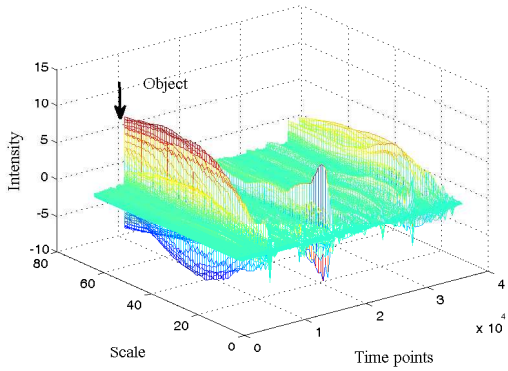
(d) First part of the CWT decomposition

Figure 4.7: Comparison of STFT and CWT [ASSS14]

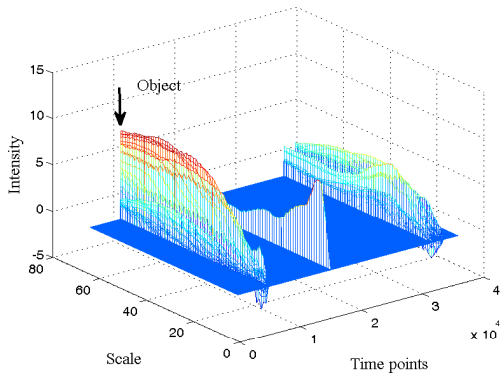
Several noise and disturbance sources are involved in this complex and unstable production process. These lead to difficulties in recognizing target objects. The disturbance can be stationary background noise or non-stationary noise, with large or rapid spectral changes over time, and can therefore resemble events resulting from target objects.

Fig. 4.8(a) shows an acceleration signal resulting from CWT as a function of time and frequency. The signal includes different events. The marked event is the only one that needs to be detected. All other events are caused by different noise events.

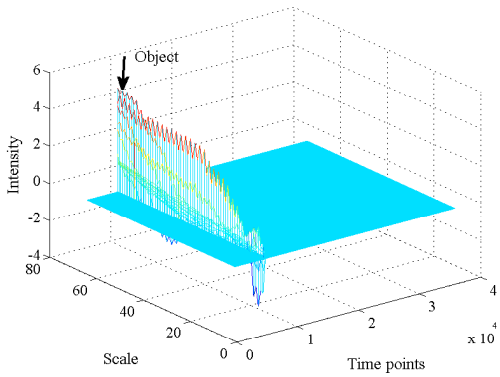
For efficient learning and reliable classification, further filtering is applied to reduce the data complexity. This filtering eliminates known noisy events, as described below.



(a) CWT decomposition of an acceleration signal



(b) CWT decomposition of the acceleration signal after applying first prefiltering rule



(c) CWT decomposition of the acceleration signal after applying second prefiltering rule

Figure 4.8: Prefiltering results of prefilter I and II [ASSS14]

4.4.1.1 Prefilter I: Prefiltering the background noise

An acceleration signal contains permanent background noise in both the time and frequency domains (Fig. 4.8(a)). The presence of events in the training data set complicates SVM training.This complexity results due to difficulty in labeling training data because of the interpenetration of the different events. Inaccurate labeling of the training data affects the classifier performance. The aim of prefilter I is to eliminate stationary background noise from the extracted features to avoid this problem and simplify SVM classifier training. This involves eliminating events in the acceleration signals for time points at which the maximum intensity of low-frequency excitation is less than an experimental threshold according to the prefiltering rule.

If this condition is fulfilled, the intensity for all frequencies at such time points is set to zero (Fig. 4.8(b)).

4.4.1.2 Prefilter II: Prefiltering specific known noise behavior

The distinguishing characteristics are used to filter process-related non-stationary disturbances. The main characteristics differ for detection cases and for disturbances related to gradients and event forms. The gradient from the low-frequency to the high-frequency domain is estimated at each time point and compared to an experimental threshold. The frequency intensity is set to zero at those time points for which the gradient does not exhibit the same behavior as target events. In addition, the relative signal intensity for the low scales is checked and compared to the behavior of the target events. Experimental evidence has shown that certain behaviors for the starting scales can be known as they do not belong to target objects but to specific noise events. This characteristic was confirmed using manual classification of the target objects to be removed.

It should be noted that event strength is not necessarily a reliable characteristic for distinguishing between events to be detected and disturbances owing to the nature of the production process. Some target objects cause weaker effects than those caused by disturbances. Therefore, scale-invariant recognition based on the event form is more reliable and effective.

An example data set filtered using prefilter II is presented in Fig. 4.8(c). The second condition checks whether the gradient at each time point exhibits the behavior of the target events.

4.4.2 Classification process

The Libsvm algorithm [CL11] was used for numerical realization of the SVM classifiers. An experimental data set was prepared using wavelet-based prefilters to build the training data set. The SVM classifier model was then built based on this set and used for classification of the five different acceleration signals.

4.4.3 Decision fusion process

The decision fusion process based on experimental data combines individual preliminary decisions to reach a final decision. The decision fusion filter is a rule-based filter. The rules are as follows:

- *Rule I*: At least two simultaneously positive individual decisions lead to a final decision of "positive: target object present".

- *Rule II*: Several strong positive decisions from at least one of the five classifiers within a specific floating decision window lead to a final decision of "positive: target object present". Several relatively strong positive decisions imply that the number of positive peaks within the floating decision window is greater than an experimental threshold. In addition, the decision value should be greater than an experimental threshold.

4.4.4 First industrial implementation and results

Results for CWT-based detection are summarized in Table 4.2. The best individual accuracy is 58.3% (classifiers 1 and 2), but this corresponds to a higher rate of false alarms. It should be noted that increases in the training accuracy for individual classifiers will lead to improvements in the detection accuracy; the rate of the false alarms would also increase accordingly. During development and depending on the system requirements, a compromise between detection accuracy and false alarms must be achieved.

Table 4.2: Classification results for the CWT-SVM approach

Training data					
Sensor/classifier number	1	2	3	4	5
Target objects			16		
No. of support vectors			1197		
SVM kernel			RBF		
Individual results for test data					
Target objects			36		
Objects detected	21	21	20	18	13
Accuracy [%]	58.3	58.3	55.5	50.0	36.1
False alarms	7	11	12	6	1
Fusion results for test data					
Objects detected			23		
Accuracy [%]			63.9		
False alarms			4		
False alarms/number of objects [%]			11		

The individual and fused results reveal that classifier fusion leads to a reduction in the false alarm rate by approximately 89%. The final detection rate is approximately 6% better than for the best individual classifier.

4.5 Discussion and comparison of approaches

The efficiency of any monitoring system is generally evaluated according to the detection accuracy and the false alarm rate. The challenge for any detection approach is to achieve the highest possible detection rate with the lowest possible false alarm rate. The complexity of the monitoring system should also be considered in evaluations. Complicated systems involve more unexpected defects and flaws than simple ones because of the unpredictable nature of disturbances which is not rigorously perceptible. In addition, implementation of more complicated systems usually requires more effort to realize and can involve greater difficulties in real-time applications.

According to the results in Tables 4.1 and 4.2, which use the same test data set to realize the testing phase, the STFT-SVM approach detects at least 10% more target objects compared to the CWT-SVM approach. The detection rate is 75% for the STFT-SVM approach and only 64% for the CWT-SVM approach. This does not necessarily mean that the STFT-SVM approach is better than CWT-SVM, because the highest number of false alarms for STFT-SVM is approximately double that for CWT-SVM. It should be noted that the sensitivity of the training process affects the rate of detection and the rate of false alarms in various ways. Increasing the sensitivity of the model can increase the rate of detection, as well as the number of false alarms. In fact, a trade-off exists in the relationship between detection and false alarms. This means that both approaches have the same improvement potential with respect to detection and false alarm rates.

Other important issues regarding realization and implementation are as follows. In the first approach (STFT and SVM) a number of 7 SVM classification algorithms (5 of them in parallel) are developed in order to realize the proposed fusion modules, whereas in the second one (CWT and SVM) only one SVM classification algorithm is developed and used for the five individual preliminary classifiers in the system. Moreover, the two sets of multiple rules required for STFT-SVM and the backward evaluation scheme are more complicated than the simple combination rules for the CWT-SVM approach. These are strong factors in judging whether the CWT-SVM approach is more efficient and reliable for real-time applications. The STFT-SVM approach is indeed difficult to design and develop and much processing effort is required for implementation. In spite of the comparable results for detection and false alarm rates, the CWT-SVM approach is more convenient to realize and implement and more appropriate for real-time applications.

For the two approaches, it is evident that the system was successful in isolating the required objects from noise events. After processing and transformation, many noise events acquire different shapes and become visually different from the target object; however, some noise events do not. Such noise events might be detected simultaneously by many sensors, which can lead to false alarms. Such noise events

cannot be avoided. They are usually caused by objects that are similar to the targets but smaller in size.

Since the object size does not necessarily coincide with the event intensity, it is difficult to differentiate these smaller objects by simple adjustment of the system sensitivity, which could lead to deterioration in the performance and detection rate.

It should be noted that the accuracy achieved is not always that targeted for the intended monitoring system. These accuracies represent individual accuracies for the acceleration sensors among other detection modules based on other data acquisition techniques. All these modules are fused together to realize the monitoring system. In this contribution, the principles behind are developed and compared to previous results for similar applications.

In the following, some additional results justify the fusion of all five sensor channels instead of using more simple combinations of two decisions, for instance.

The classifiers based on sensors 1 and 2 were considered and the STFT-SVM approach was applied. In contrast to the 27 objects detected after fusion of five sensors, only 16 target objects were detected as the best achievable result; however, the number of false alarms decreased from seven to six. After fusion of sensors 3 and 4, the detection rate was 11 objects out of 36, with two false alarms. To compare the sensors 1 and 2 with the sensors 3 and 4, the individual numbers of false alarms are added for sensors 1 and 2 as 32 false alarms, and for sensors 3 and 4 as 11 false alarms. The ratios of the fused number of false alarms (6 for sensors 1 and 2, and 2 for sensors 3 and 4) to the sum of the individual number of false alarms (32 for sensors 1 and 2, and 11 for sensors 3 and 4) are almost similar (approximately 0.18).

For the CWT-SVM approach, sensors 1 and 2, and sensors 3 and 4 were fused. The results were as follows. Only 19 objects were detected by fusing sensors 1 and 2 and the number of false alarms was reduced to 1. For fusion of sensors 3 and 4, 18 objects out of 36 were detected, with no false alarms. The ratios of the fused number of false alarms (1 for sensors 1 and 2, and 0 for sensors 3 and 4) to the sum of the individual number of false alarms (18 for sensors 1 and 2, and 18 for sensors 3 and 4) are much smaller than that obtained through STFT-SVM approach. These results clearly indicate the better fusibility of the individual sensors and the lower redundancy.

In general, the individual sensors deliver a high number of false alarms for both approaches, so detection is unreliable. By combining two individual sensors, better reliability can be achieved with fewer false alarms. The detection reliability can be improved by decreasing the number of false alarms and increasing the rate of detection. This is achieved with a fusion process for the five sensors and related filters, in order to achieve deep insight into the problem diversities and distinctions. Although fusion leads to a slight increase in the number of false alarms, this deterioration is

negligible compared to the overall improvement in system performance. The deterioration can be explained by redundancy and the accumulation of disturbances when many sensors are considered together. The improvement potential of fusion lies in the quality of the sensor data obtained and the feature extraction process used.

4.6 Conclusion

Two different approaches for feature-based multisensor fusion were developed for a monitoring system. The goal was to monitor a production process for online detection of a target object with the lowest possible false alarm rate. The application is a stationary process with a continuous stream of sensor signals. The two approaches were compared in terms of performance and complexity.

The STFT-SVM approach uses STFT for feature extraction. The individual feature vectors for different acceleration sensors are classified using SVM to provide individual preliminary decisions about the presence of a target object. These decisions are fused using a novel fusion module to provide a reliable decision about the system state.

The CWT-SVM approach uses CWT as a prefilter to extract features, which are subjected to further prefiltering processes before SVM classification.

The results for both approaches reveal improved detection rates and fewer false alarms. Approach I leads to a better improvement in detection rate compared to individual preliminary decisions. The number of false alarms is also lower. Approach II leads to an even better improvement in the false alarm rate than approach I.

5 Wavelet-SVM system for crack detection

In this chapter, two main approaches used for crack detection and prediction in rotating machinery; model based and signal based, are investigated, in order to achieve a prewarning technique for rotor cracks to be applied for online monitoring in turbo-machinery. Several strength and weak points are discussed and compared, in order to achieve a comparative overview of the available techniques. The concepts fault severity estimation and remaining life time prediction are considered in the application. The signals encountered are periodic vibration signals with accumulative impact of the fault incident. Realistic simulation example is considered in this chapter. However, in principle the developed approaches can include arbitrary data acquisition techniques.

During the progress of this work, some scientific papers were published based on the studies and results of the work. Others are still under processing for publication. This chapter is prepared to a large extent based on these publications [SWWS16, SSW13, SSW12].

5.1 Introduction

Diagnostics, Fault Detection and Isolation (FDI) approaches of rotating machines are crucial factors to ensure cost effective and reliable industrial processes. Advanced approaches featuring monitoring and control of availability aspects of rotating machines like Condition-Monitoring Systems (CMS) or Structural Health Monitoring (SHM) systems require a reliable diagnostics and prognostics system. In industrial practice nowadays these systems include signal-based approaches in combination with machine learning methods. Model-based approaches using suitable system description are also discussed mainly in academic research as a possible alternative to the more application-oriented former signal-based approaches. Modern approaches are based on feature extraction and recognition algorithms, along with mathematical modeling and simulations, in order to detect and/or avoid faults able to breakdown machines and systems by affecting the functionality. In the consequence the task of related diagnostics and prognostics approaches is to establish relevant statements as early as possible.

Cracks in rotating machinery are the most critical and fundamental damages in the related industry, often caused by fatigue stress. Dynamically they often lead to vibration effects similar to those of asymmetry and misalignment, accompanied with changes in the vibration properties. These changing dynamical properties are used as indicators for changed mechanical parameters of the rotor, including deterioration effects. The knowledge about the related causal chain initialized by shaft cracks in effecting the system dynamics of the rotating system can be used by model-based as well as signal-based approaches.

It should be noted that both approaches (model-based, signal-based) use knowledge, but in a different way.

Signal-based approaches are using output signals (denoted as S_o). Fault detection modules relate the raw or filtered signals (denoted as S_f) to reference or threshold values stating changes with respect to the normal (regular, healthy) conditions, called fault detection. Additional knowledge and modules are necessary to relate these signals (S_o, but mainly S_f) to those related to distinguished one compared on a third level (feature space). Other classification approaches are used (possibly in combination with suitable chosen filters) to distinguish states/faults as classes. Relating to this, also model information/knowledge is used implicitly, but also these approaches are denoted as signal-based (or data-driven) approaches.

Model-based approaches uses beside output signals S_o, input signals S_i and a model to be built (parameter identification) or to be assumed (observer). The comparison results directly from this level on the base of comparison (of parameters: identification or the outline establishment of residuum (observer)). Filtering of residuum allows the distinction of errors location within a so called parity space. These models are assumed as given/known directly, so these approaches are denoted as model-based.

In this contribution two different related approaches are explained in detail in order to evaluate their potential and applicability. The differences and the possibilities are illustrated with respect to the development of a reliable crack detection approach to be applied for online monitoring in turbo-machinery.

The work is partly based on a previously published model-based approach [SBM93], firstly published in 1993. Here the approach is briefly introduced and applied to a new example using realistic model data. The signal-based alternative used here has been previously published in complete different context [SS11]. Both approaches are well developed within the last years [SS11], or decades [SBM93]. In addition to a previous publication [SSW13], here the two approaches are applied to a very realistic simulation example for the first time and both are compared in detail using various studies. An addition to previous publications the approaches are given with more details.

5.1.1 Cracks in rotors

Rotating shafts are considered among the most important and most critical machine elements in the industrial processes and machines such as turbines, compressors, and pumps. The rotors of such machines are usually subjected to extreme working conditions of loading and temperature variations. Accordingly the consequent failures can lead to enormous damages, economical losses, and injuries. A profound and actual overview is given in [Ish08], here Ishida introduced case histories of accidents and cracks found in industrial machines.

Different kinds of faults and flaws have been recorded in the rotating machinery such as unbalance, misalignment, rub and looseness [WPJ13]. The cracks in rotors have long been considered as factors limiting the safe and reliable operation of the rotating machinery. A crack may be developed from some surface or internal imperfections and propagate without much apparent warning. In ductile steels used for rotors, cracks are influenced by many factors such as the rapid fluctuation of the bending stresses, the presence of stress raisers and possible design or manufacturing flaws, and the variations in temperature and environment [SKKQ04].

Methods of crack formation and propagation vary from high and low cycle fatigue to temperature and environment effects. A typical event of cracking in ductile steels can be divided into three stages: crack initiation; in which tiny discontinuities are initiated in the uncracked parent material, crack propagation; in which the discontinuity grows in size as a result of the cyclic stresses induced in the material, and failure; which occurs when the material that has not been affected by the crack cannot withstand the applied loads [SKKQ04, BPT10].

The transverse breathing crack is the most critical type of cracks in rotating shafts. Here the cross section is reduced weakening the rotor under certain conditions. In case of breathing, the opening and closing the rotating crack is due to the rotation of the shaft. The crack moves from the upper position in which the static bending moment forces the crack to be closed, to the opposite position in which the crack is forced to be open [BPT10].

The influence of the crack existence in the rotor structure is related to the change of the local stiffness of the crack. A crack introduces local flexibilities and time dependent changes in the stiffness of the structure due to strain energy concentration, hence reducing the natural frequencies of the original uncracked rotor [BPT10, GA08].

5.1.2 Crack detection of elastic rotors: A brief review

Several crack models are developed and used in the 70ties and 80ties, mostly used to be integrated within simulation studies.

Several techniques have been used to monitor cracks in rotors such as vibration sensors, ultrasonic measurements, tribological analysis, and recently the acoustic emission techniques [BPT10, MR06]. However, the vibration-based techniques have been used widely as tools for fault diagnosis in the rotating machinery [GC05, VG12]. Vibration-based systems directly measure the rotor response forced by rotor flaws. A general review of the vibration-based condition monitoring for structures is given by Carden and Fanning [CF04]. They presented several approaches adopted in the literature for time, frequency, and modal domains, such as the natural frequency-based methods which are modal methods dependent on frequency shifts and the

relation between stiffness changes and natural frequency changes. According to Carden and Fanning, the reliability of such methods is limited to single or few damage locations and/or small laboratory structures. They presented and discussed other approaches based on mode shape, modal strain energy, dynamically measured flexibility matrix, residual force vector, wavelet transform, neural networks, genetic algorithms, and statistical pattern recognition. They stressed that there is a lack in research to deal with synchronous faults and the remaining service life. The vibrational behavior of cracked rotors is also studied by Silani et al. [SZRT13]. They used a finite element approach and short time Fourier transform (STFT) to investigate the detection of small cracks. They presented that though the transient response does not change sensibly in the presence of very small cracks, the STFT of the response behavior can clearly identify cracks. In the work of Sawicki et al. [SSL09], the method of Multiresolution Wavelet Analysis has been applied on the vibration signal of a rotating system with and without external force excitation, in order to distinguish the existence of a transverse crack. They found that the differences are more pronounced in the presence of external force excitation. They presented the RMS amplitude values of the vibration signal in different frequency bands as a simple quantification method for fault severity. Several other vibration-based techniques are introduced in the literature [GP07, YMM03].

The objectives of any crack monitoring system comprise crack detection, localization, severity quantification, and remaining service life prediction. The crack detection and localization have been much more emphasized in the literature than the other two objectives.

An early concept of model-based crack detection based on the theory of disturbance observer is introduced by Söffker et al. [SBM93]. Here, based on the nominal behavior of the system, the fault effects caused by the crack are interpreted as unknown external effects acting to the nominal behavior of the rotor. Measurements of displacements and/or velocities of the vibrating dynamic system are necessary, together with further information such as the mechanical model of the rotor and the characteristics of the typical behavior of the crack. Another model-based method is proposed by Sekhar [Sek03] for the on-line identification of cracks in a rotor while it is passing through its flexural critical speed. The fault-induced change of the rotor system is taken into account by equivalent loads in the mathematical model. For crack modeling, the flexibility matrices of the cracked section are utilized according to Papadopoulos and Dimarogonas [PD87]. The equivalent loads are virtual forces and moments acting on the linear undamaged system to generate a dynamic behavior identical to the measured one of the damaged system. The rotor has been modelled using FEM, and the crack has been identified for its depth and location on the shaft for different rotor accelerations. The CWT has been used to extract the sub-harmonic features of crack from the time response. The work results show accurate estimation of crack location, however; the error of crack depth estimation increase with the decrease in measured vibration data (DOF). On the contrary, results of

Xiang et al. [XZCH08] show better accuracy of crack depth estimation than that of location estimation. The model-based detection system proposed by Xiang et al. is based on the combination of wavelet-based elements and genetic algorithm. Genetic algorithm is applied to eliminate the errors of frequencies between numerical simulation and experimental measurement. The wavelet transform is used also in the work of Nagaraju et al. [NNRMR09]. They studied the transient analysis of rotor system with transverse breathing crack for flexural vibrations. To extract the hidden features of the crack, the time signal obtained from the transient analysis is transformed to 3D-CWT plots in which the time-frequency components are clearly represented. A new wavelet plot called cross wavelet transform (XWT) has been applied to the time signals to obtain the phase angles. The XWT gives the phase angles of different frequency components along with the sub-critical peaks in a single plot. The inverse problem of crack detection has been carried out using Artificial Neural Network (ANN).

Some other reference papers for model-based systems are recommended [ELF98, SA08].

The signal-based systems for monitoring rotating machines have been used for long time. The techniques used are gradually updated according to data acquisition and signal processing techniques used. The advances in machine learning and feature extraction techniques have induced new methods and techniques to be adopted in the field of rotating machinery fault detection.

The work introduced by Tao et al. [TLLW13] comprises a detection method based on Fisher discriminant analysis (FDA) as linear dimensionality reduction technique and Mahalanobis distance (MD) for performance assessment. Data samples are projected into a new low dimensional space in which MD between the new measurement data and normal population is calculated for performance assessment. As a conclusion, the transformation of MD into the feature space and the determination of an adaptive threshold for fault detection is still a challenge. An energy-based approach to defect diagnosis in rotary machines is introduced by Yan and Gao [YG09]. The method uses continuous wavelet transform CWT and is based on the analysis of the energy content associated with the signal to determine the best suited base wavelet and decomposition scale for analyzing the signal.

Some other reference papers for signal-based systems are recommended [WP08, YMM02].

For deep understanding about the cracked rotors techniques and recent advances in general and for the two main categories model- and feature-based, some recommended general review references are helpful [WPJ13, SKKQ04, KR09, BPT10].

5.2 Rotor system

Modeling of elastic rotors has been developed over decades to a high degree of sophistication for simulation, fault detection, and isolation purposes. The developed models are validated by comparing numerical results with the natural frequencies, mode shapes and critical speeds acquired from vibration measurements of the rotors. However, sufficient performance of a rotor model should usually consider the dynamics of complex framework foundation which is directly connected to the rotor dynamics. This adds more difficulties and complexity to the process of rotor modeling and restricts the use of models.

The incorporation of breathing behavior into the dynamics of the rotor represents a significant increase in complexity as a result of the nonlinear characteristics of the real transverse fatigue cracks.

During the rotation of the rotor, and mainly as a result of gravity, a portion of the rotor cross section remains under compression and a portion under tension. In case of cracked rotor, the crack section facing tensile stress opens, while the crack section facing compression closes. Therefore, the crack section opens and closes according to the angular position of rotation. In the completely closed position, the rotor behaves as it was uncracked.

Crack breathing is accompanied by periodical changes in the stiffness of the rotor. At certain angular position, when the crack is completely closed, the rotor has almost the stiffness of faultless rotor. Similarly, at certain angular position, when the crack is completely open, a significant decrease in local stiffness exists, however; the reduction in global stiffness of the rotor depends on the depth and location of the crack.

Many approaches have been developed for modeling cracks in rotors, and the subsequent reduction in stiffness. The developed models should accurately represent the rotor system, however; the required accuracy depends on the intended use of the model. On the other hand, the complexity of the breathing crack mechanism leads to necessary approximations and assumptions which are required in order to simplify the process of modeling and use of model.

The earliest, most simple, but often used model, is the hinge model published in 1976 by Gasch [Gas93]. The steering function in the model is a step function approximation of the crack in which the crack is represented as either entirely open or closed. It is assumed in the model that vibrational amplitudes are small compared with the static weight deflection, and the axial and torsional vibrations are ignored [Gas08]. In Fig. 5.1 the "breathing" of the crack under the weight influence when the shaft is slowly turned, and the model steering function $f(t)$ are shown.

A smooth transition between the open and closed positions was introduced by the sinusoidally varying model presented by Mayes [MD84]. The model uses a steering

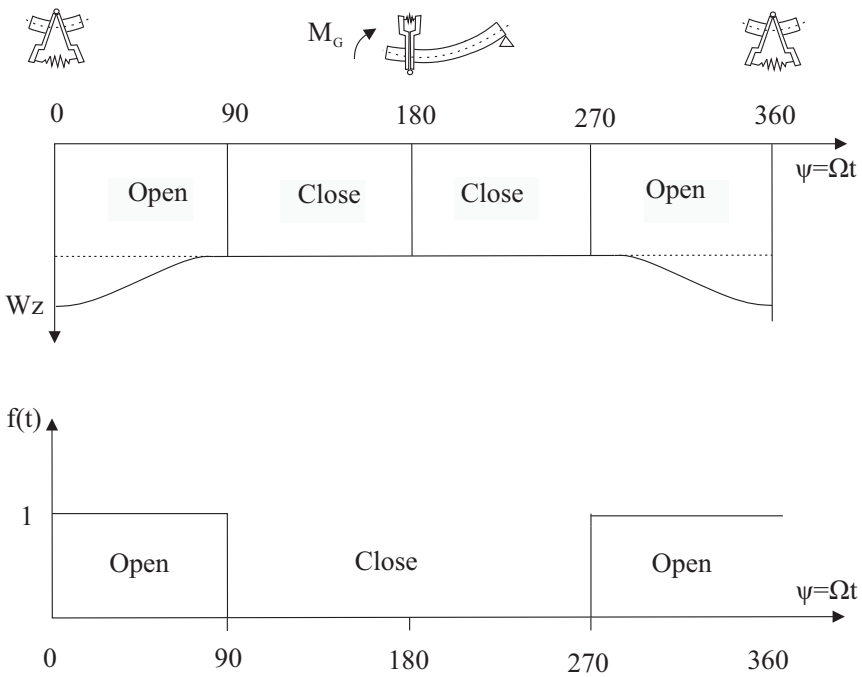


Figure 5.1: Crack breathing, static deflection, and corresponding steering function (ref. to [Gas08])

function in the form $f(t) = (1+cos\Omega t)/2$ as a better approximation of the breathing behavior. The use of Mayes model is more significant for deeper cracks, although the rotor stiffness is not directly connected to the depth of the crack.

Considering the crack related vibrations, more complicated breathing model was introduced by Jun et al. [JEEL92]. The presented model expressed the equation of motion with the response-dependent stiffness in a simple rotor. The model used the fracture mechanics to estimate the cross coupling stiffness, as well as the direct stiffness. The crack openness was determined through the response solved by the governing equation, and the response-dependent stiffness was calculated by numerically integrating over only the open crack area [JG08].

In order to implement more realistic rotor models, the finite element method has been used in connection with the previously mentioned models by many researchers.

The previously mentioned models were applied and compared by Penny and Friswell [PF03] in order to investigate the influence of the crack model on the response of a general rotor model. According to their results, the three crack models examined had relatively small effect on the predicted steady state $1X$ response but they did have some influence on the predicted whirl orbit and the steady state $2X$ response. However, in any crack identification scheme, these differences are not likely to have a significant effect and simple models are more readily used [LF06].

The example system used in this contribution uses an application-oriented modeled rotor based on a finite element model of a length of 4.2 m and a radius of 0.14 m. The rotor is supported by 2 bearings, which are modeled with a simple bearing supporting damping (total damping coefficient d = 9e3 Ns/m and stiffness k = 3e7 N/m). While the 4 sensors at the bearing positions allow measurement of the rotor's displacement; the displacements and velocities of the beam nodes in the x and y plane are measured directly during operation. This measurement scenario is the typical one in practice and different to those used in the previous publications [SBM93]. The discretized model of the rotor (Fig. 5.2), has 8 nodes, 7 elements and 32 elastic degrees of freedom (each node has 4 degrees of freedom, translation and rotation in the x and y planes). The vector equation of motion arises to

$$\begin{aligned} M\ddot{q}(t) + (D + \Omega G)\ddot{q}(t) + Kq(t) =& Fw, \\ y =& C_r q, \end{aligned} \tag{5.1}$$

where M represents the mass matrix, D the damping matrix (including bearing damping), G the gyroscopic matrix, K the stiffness matrix (including bearing stiffness), f the input matrix, q the displacement vector, y the measured nodes, C_r the output matrix (corresponding to sensor nodes), Ω denotes the rotational speed, and w the input vector (unbalance forces and crack forces).

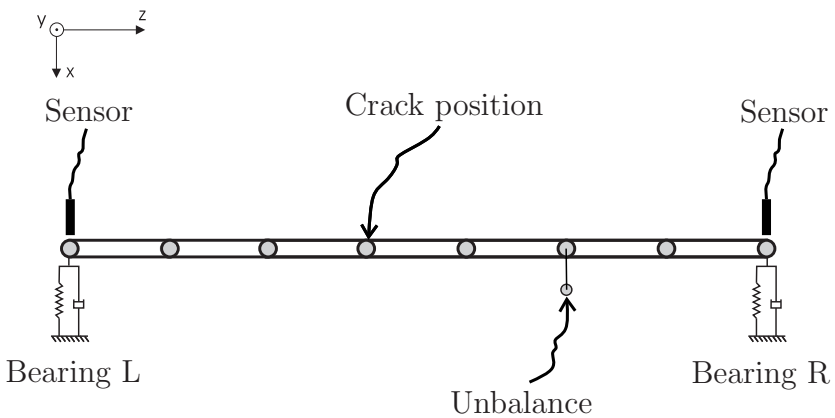


Figure 5.2: Discretized rotor model [SSW13]

The modeling of the shaft cross-crack is realized by integrating the effects of curvature dependent changes applied to the location of the crack. Additionally, crack-specific parameters [SBM93, MD84] have to be adapted. This procedure is used here similar to those introduced in [SBM93]. The integration of the crack effects results in an addition to the rotor dynamics description to additional strongly non-linear effects. The crack-related flexibilities are primarily influenced by the curvature-induced opening of the crack at the crack position (here assumed as known). The

integration into a FEM-based description is necessary for applications because for signal-based approaches output data have to be generated and for model-based approaches input-output information beside the model has to be used. Here the unknown external effects are interpreted as effects on the right side of the dynamical equation. In the case of simulations these external effects have to be taken into account for the simulation as well known effects.

The state space model is used for both simulations as well as for the observer design. It is known that the eigenvalues of a rotor system are speed-dependent. The eigenvalues are calculated at the rotational speed of 9000 rpm. The first four forward modes are illustrated in Fig. 5.3. The amplitudes for each mode are normalized, so that the maximal amplitudes for each mode are taken as unity. The two rigid body modes and the first two bending modes can be clearly identified.

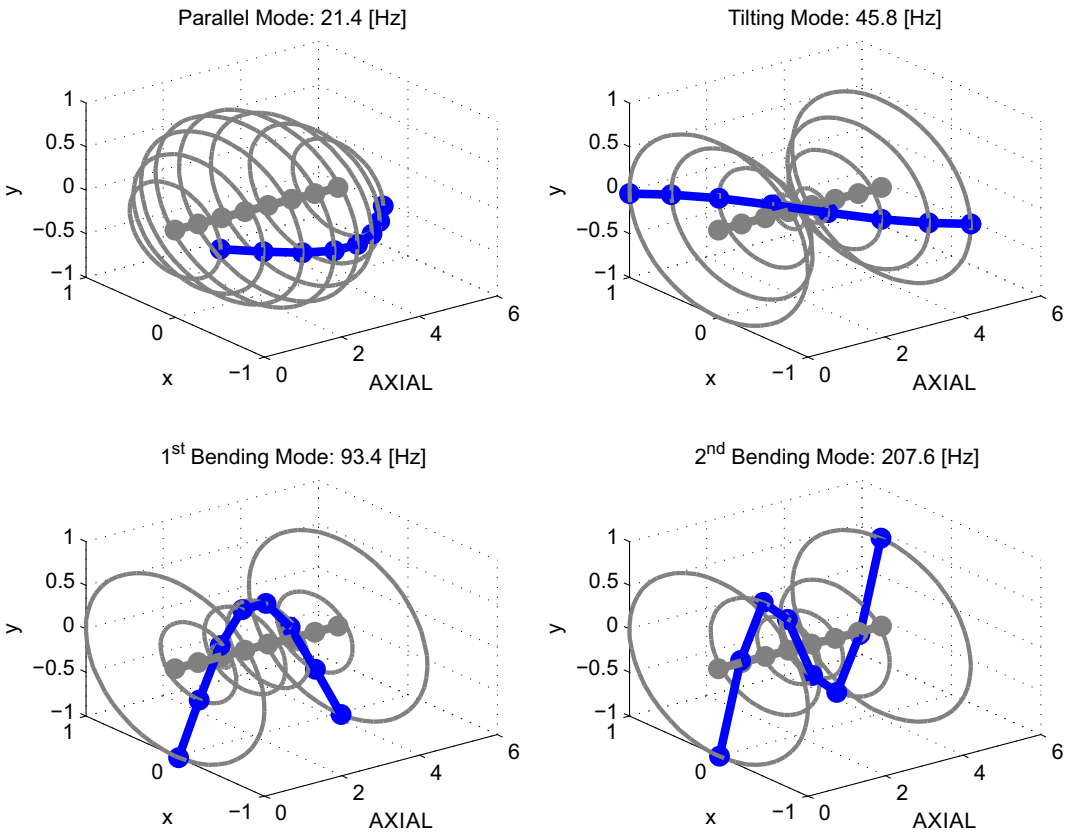


Figure 5.3:]
First 4 rotor eigenmodes [speed of 9000 rpm] [SSW13]

To show the effect of the crack in the shaft dynamics, simulations are performed with and without crack. The system is excited by an unbalance $U = 0.12\ Kg\,m$ acting to node 6 of the rotor; the weight force is taken into account in the form of static forces in the y-direction. Using the location of the crack at the node 4, the

rotor crack is modeled using a relative compliance of $hr = 0,001$. In Fig. 5.4 an example of the system to be considered is shown, it can be stated that the crack affects only a small noticeable change in the oscillation amplitude, but not in phase.

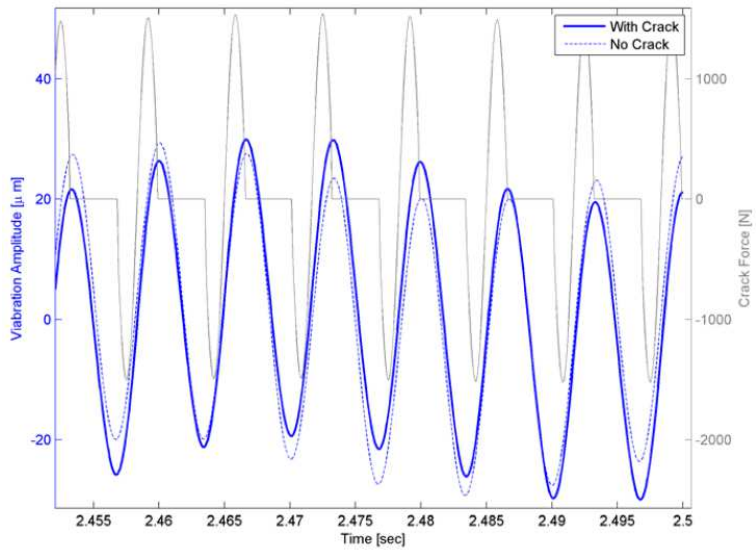


Figure 5.4: Vibration amplitude (blue) and crack force (gray) at Lx [SSW13]

The technical challenge for monitoring is similar for all procedures. The change of the rotor dynamics due to the occurrence of the shaft cross crack through the available (indirect) measurements of the bearing shaft movements has to be identified with respect to fault detection. The signal characteristics or in the sense of a diagnostic task, the identification of the mapping between the detected change and the related causal cause, should be used in the way that the fault should be assigned to the geometrical position of the irregularity.

5.3 PI-Observer (model-based) approach

As mentioned in the section of brief review (section 5.1.2), crack detection based on model-based approaches has been studied in the last years by many groups.

An early concept of crack detection based on the theory of disturbance observer, later denoted as PI-Observer [SYM95], is introduced by Söffker et al. [SBM93], later optimized for practical use and successfully used for several years [SYM95, LS12]. Here, based on the nominal behavior of the system, the fault effects caused by the crack are interpreted as unknown external effects acting to the nominal behavior of the rotor. For this task, measurements of displacements and/or velocities of the vibrating dynamic system are necessary, together with further information such as

the mechanical model of the rotor and the characteristics of the typical behavior of the crack.

The system equation of motion (equation 5.1) mentioned in section 5.2 can be extended to more general description of state-space model as

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix} &= \begin{bmatrix} A & NH \\ 0 & F \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}. \end{aligned} \tag{5.2}$$

Here x denotes the 2n-dimentional state vector (consisting of displacement and velocity variables), A is the 2nx2n system matrix, B represents the input matrix, C the output matrix, y the vector of measurements, and u is the 2n-dimentional vector of known control inputs and/or excitation functions. It is assumed that the system parameters A, B, C, and N, as well as the input and output time signals u and y, are known. The input matrix of the nonlinearities N couples the fictitious approximation Hv of the nonlinearities and unknown inputs n to the states where they appear. The signal characteristics of these inputs are approximated by a linear dynamical system with the system matrix F. The task is to reconstruct the unknown nonlinearities (here the external disturbance forces of the crack) by applying Proportional-Integral Observer (PIO) [SYM95].

From the structure of the PI observer illustrated in Fig. 5.5, the dynamics of the PI-observed system are described by

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{v}} \end{bmatrix} = \underbrace{\begin{bmatrix} A - L_1 C & L_3 \\ -L_2 C & 0 \end{bmatrix}}_{A_e} \begin{bmatrix} \hat{x} \\ \hat{v} \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u + \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} y. \tag{5.3}$$

For the estimation of crack disturbances, it is necessary to emphasize the estimation residual as

$$\begin{bmatrix} \dot{e} \\ \dot{\hat{f}} \end{bmatrix} = A_e \begin{bmatrix} e \\ \hat{f} \end{bmatrix} - \begin{bmatrix} N \\ 0 \end{bmatrix} f \tag{5.4}$$

where the estimation error $e(t)$ is introduced as

$$e(t) = \hat{x}(t) - x(t). \tag{5.5}$$

The analysis of the estimation residual can be used for the detection of faults and related localization. It is shown in [SYM95, LS12] that suitable observer design in combination with large gains lead to acceptable estimation of crack faults.

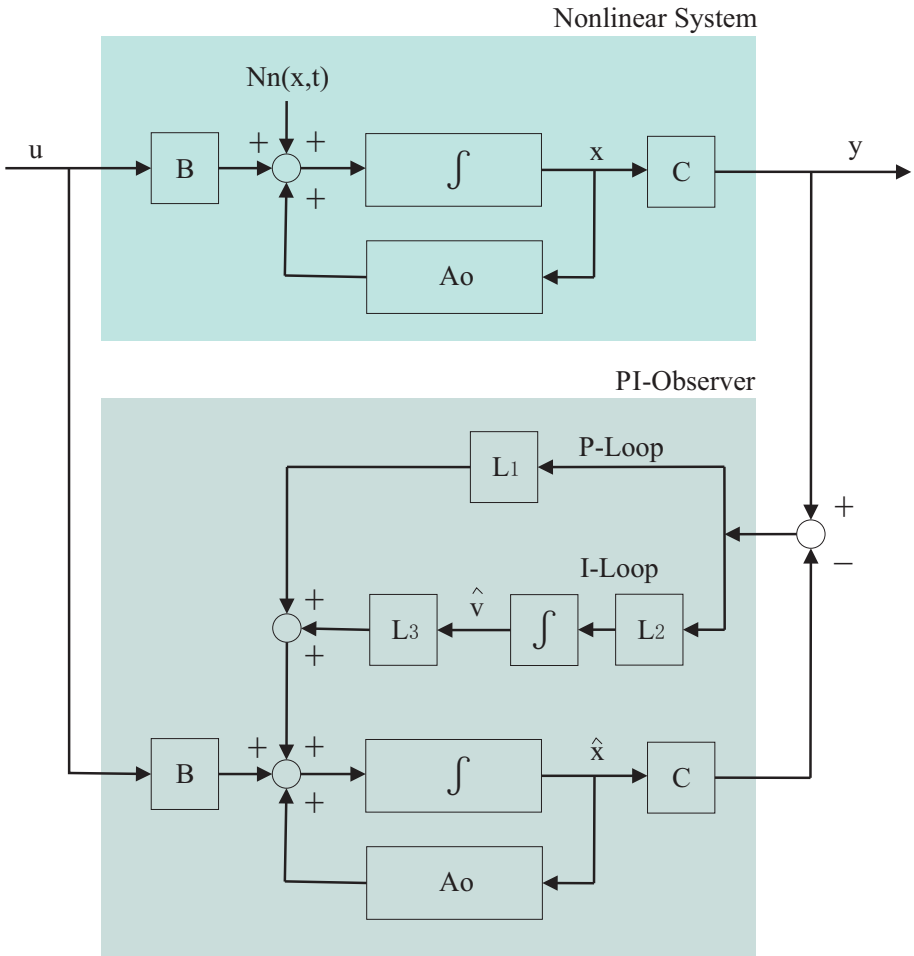


Figure 5.5: Structure of PI-Observer [SYM95]

5.3.1 Implementation, simulation, results, and discussion

The simultaneously realized estimation of the fictitious crack force as well as the related displacement at the location of the crack, the diagnostic indicator 'relative stiffness loss' can be determined as a causal indicator [SBM93] showing the dynamical behavior effect of the crack at the location of the crack. Clearly, the rotation-induced 'breathing' of the crack can be shown (Fig. 5.8). The observer-based results, based on the PI-observer method, are shown in Figs. 5.6 - 5.9. The variables can sometimes be estimated very accurately, as shown in Fig. 5.6. In Figs. 5.6 and 5.7 the time behavior of the estimation of the node displacement has been used to reconstruct the crack-induced effect from the vibrational behavior. Applying simulated noise (to simulate a real application example), a partly strong influence on the reconstructed curves is observed (Fig. 5.9).

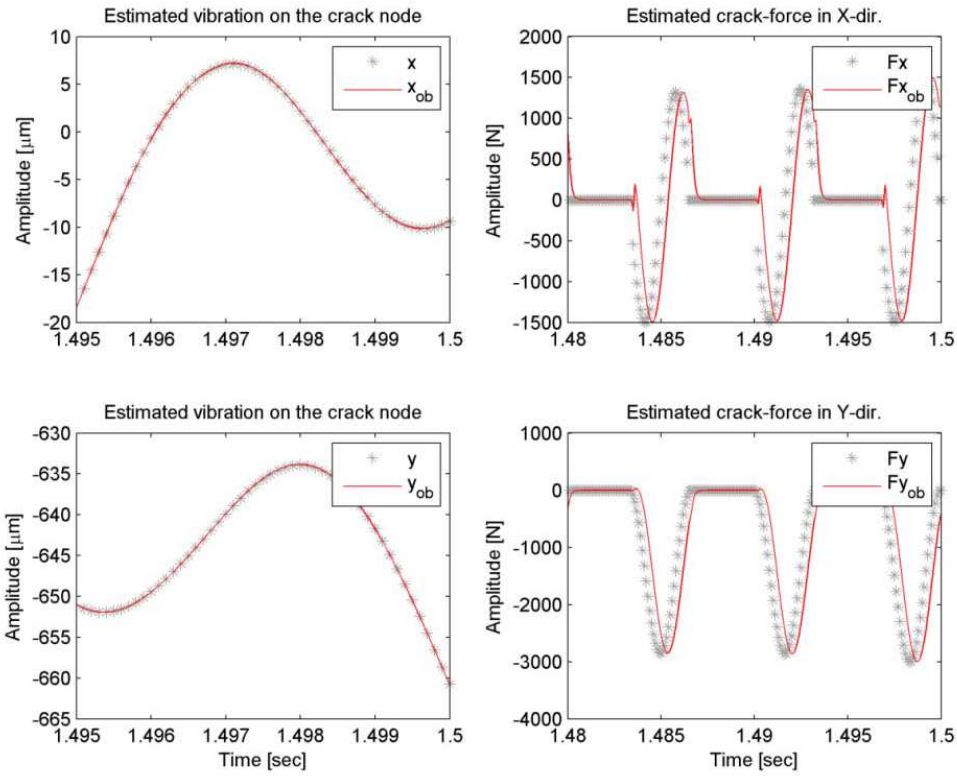


Figure 5.6: Comparison of simulated and estimated vibration (left) as well as crack force (right) [SSW13]

5.4 Data-driven (signal-based) approach

Being easy to apply (no model is needed), signal-based detection concepts represent an usual technique for vibration monitoring systems in practice. The main advantage of this widely used field of techniques is the easy applicability. The main disadvantage is that the conclusion from measurable changes to the physical reason (diagnostic statements) assumes detailed and specific knowledge or assumptions about the physical behavior of the fault. On the other hand, modern machine learning methods are used to generate features representing different states of the rotor related to the existence of faults/changes etc. like due to the dynamical effects resulting from 'breathing' cracks.

The existence and growth of cracks and faults in vibrating systems like turbomachines, can be implicitly observed by monitoring of features generated from measurements of the system. In spite of the easy use of the measurements, the detection of specific physical effects behind the change of signal properties is usually difficult to detect or classify directly, especially in the early stages of damages. This results

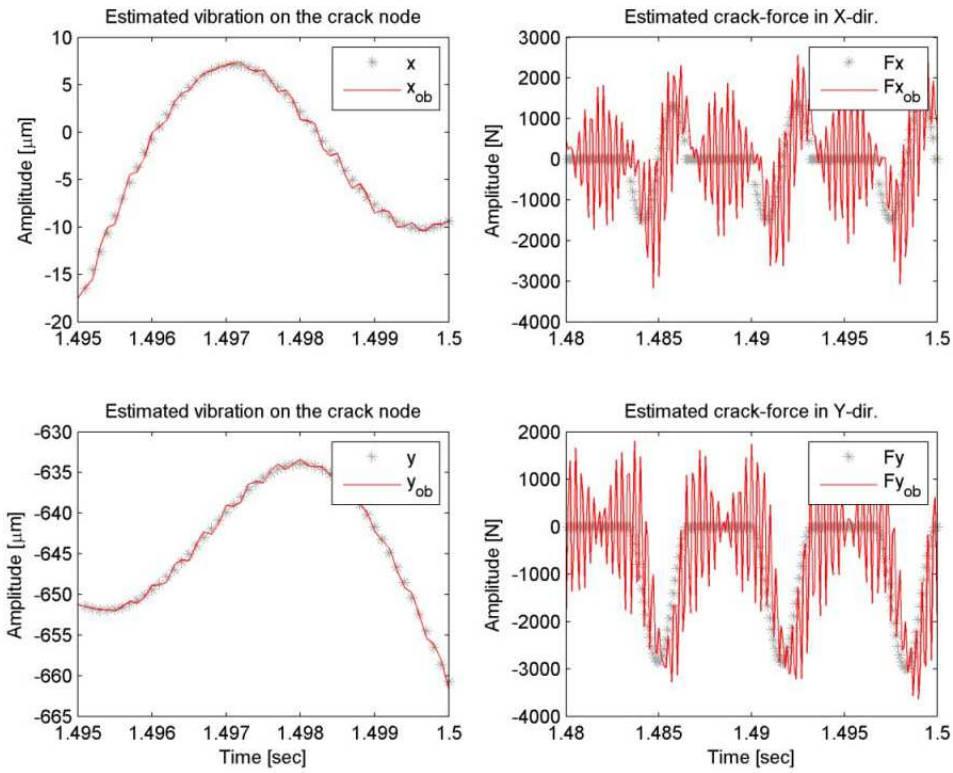


Figure 5.7: Comparison of simulated and estimated vibration (left) as well as crack force (right) with sensor noise [SSW13]

mainly from the weak effects and also from existing disturbances or other effects affecting the measurements as well as the vibrations. However; reliable measurements supported by appropriate information extraction techniques can also in case of the above mentioned effects produce recognizable features and patterns which enable reliable allocation of the physical causes, indicating the existence and size of cracks even in the early stages.

In particular, the diagnosis of failures appears as a complex task. Recent developments, however, permit the use of filtering techniques in combination with methods that do not have the limitations of classical threshold-based methods. Suitable filtering techniques for fault detection are used such as FFT, Cepstrum, STFT or wavelets, which produce sufficiently complex features to define complex characteristics of the vibrational state [GY11]. With the help of suitable pattern recognition and classification methods, the generated complex features can be learned to classify patterns in the application, i.e. assign the learned patterns. As classification methods, the known methods of Neural Networks (NN), support vector machine (SVM), and the fuzzy-based methods can be used. These methods as supervised

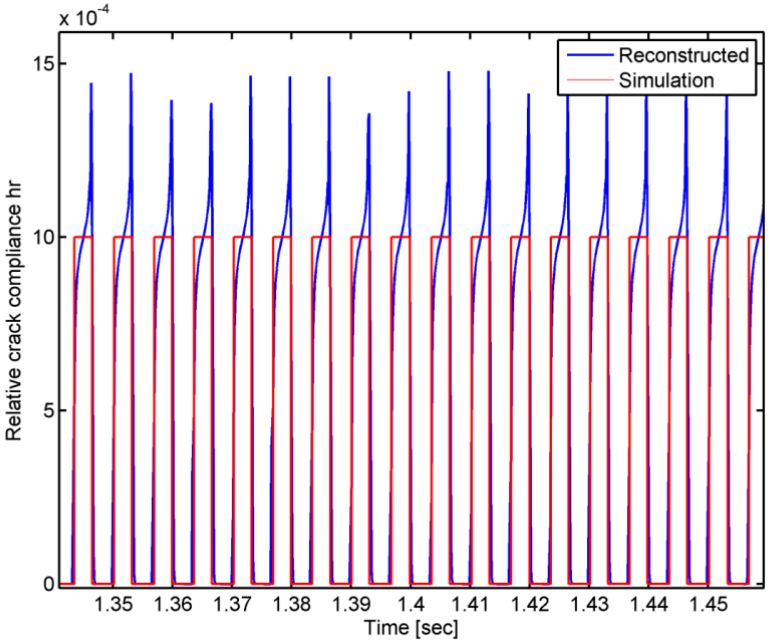


Figure 5.8: Reconstructed relative compliance [SSW13, SBM93]

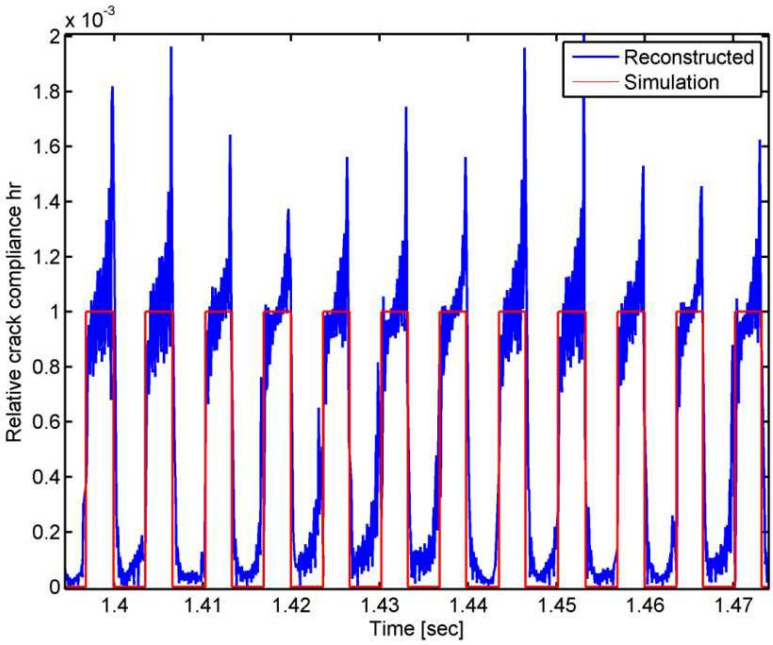


Figure 5.9: Reconstructed relative compliance considering additive sensor noise [SSW13, SBM93]

learning methods use a problem-specific data sets and form method-specific patterns that can differentiate specified faults and machine health states of interest. These techniques are in general easy to apply without the need for complex modeling task necessary for model-based approaches. The main disadvantage is that the conclusion from measurable changes to the physical reason (diagnostic statements) assumes knowledge or assumptions about the physical behavior of the fault.

Suitable machine learning methods are used to generate features representing related different states of the rotor connected to the existence of faults/changes etc. As introduced, one important dynamical effect results from 'breathing' cracks, which is an important but unusual fault. The feature extraction stage is realized to extract suitable features and to exclude the redundant ones. Additionally the suitable features are transformed into a representative form to help to make the recognition process easier. The transformed features are undergone the classification to detect specified machine health states of interest.

In wavelet analysis [GY11], signals are compared with a set of template functions obtained from scaling and shifting of a mother wavelet function. Wavelet-based approaches are widely used in classification and recognition tasks as feature extraction tools. The performance of the wavelets is proved to be more flexible than other usual approaches as they lead to time-frequency analysis with adaptive und suitable time and frequency resolution concurrently, and therefore with perfect reconstruction characteristics.

Vibration signals of the rotating machinery are usually a mixture of periodic and transient components buried in broadband background noise. For applications implying noise to be removed from a signal, a reliable alternative is the Discrete Wavelet Transform (DWT) which is obtained by a process of a dyadic parameter discretization of the Continuous Wavelet Transform (CWT) leading to more efficient computational effort as well as to a suitable size of generated parameters. These advantages make the DWT more appropriate for real-time applications in comparison to other approaches.

In general the system including crack to be monitored should be monitored, so it becomes necessary to observe and isolate growing cracks. The used routines should work robustly independent from changing operating conditions. Operating conditions may change, also damping effects. In many applications, parameters collected from the starting up of the rotor are used. It is also plausible that the stationary signals would provide reliable source of system state information by excluding the disturbing transient events.

The task of the diagnostic system includes the generation and related processing of a suitable feature set which is representative to the different machine states, and the reliable classification of the classes within the feature set. The combination of reliable feature extraction and classification procedures adds enhancement to the individual capabilities of the two modules (Fig. 5.10).

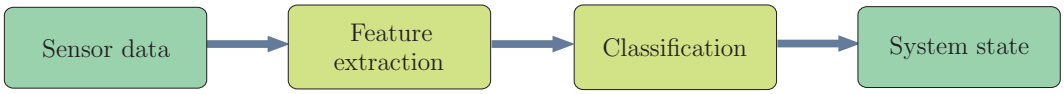


Figure 5.10: Classification modules

The application of SVM classification requires the selection of suitable kernel function and the parameters which adjust the function of the classifier. The Radial Basis Function (RBF) kernel is a usual first choice for diversity of applications for many reasons [HCL10]. Using only one parameter (γ) to be selected, the kernel can handle nonlinear cases more effectively, with fewer numerical difficulties. Additionally, the penalty parameter (C) of SVM should be assigned. A perfect separability of the training data is not necessary as it could be reason for over-fitting which is an indication to deteriorated generalization of the model. In order to avoid over-fitting, the method of cross validation (CV) is used. The training set is divided into k equal parts. One part is chosen for testing and the rest for training the classifier. This is done for all the parts and gives an average indication of the classification accuracy for different values of C and γ. A multistage grid search (Fig. 5.11) is then helpful in order to find the required best accuracy.

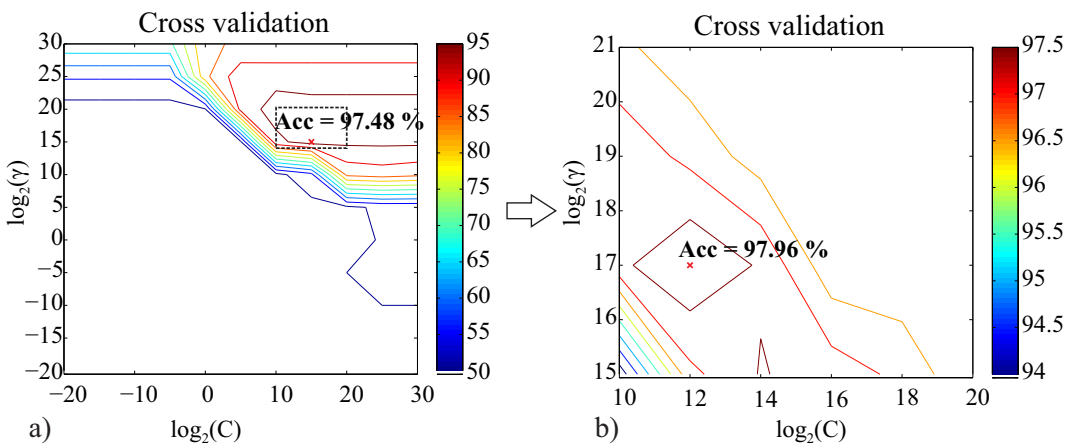


Figure 5.11: Cross validation and parameter combination for a) coarse grid, and b) fine grid

An illustration of the classification module is presented in Fig. 5.12. The parameters C_{opt} and γ_{opt} represent the optimal configuration of the classifier parameters and selected for the classification of the test data. In the case of multi-class SVM, the method used must be considered together with parameter selection strategy. As an example, there are two options to implement the "one-against-one" method considering parameter selection: First, for any two classes of data, the parameter selection is conducted to have the best (C, γ). The second option is that for each (C, γ), cross-validation in combination with the "one-against-one" method is used for

estimating the performance of the model. A sequence of pre-selected (C, γ) is tried to select the best model [CLS05]. Considering the overall accuracy, one parameter set for each individual decision function may lead to over-fitting.

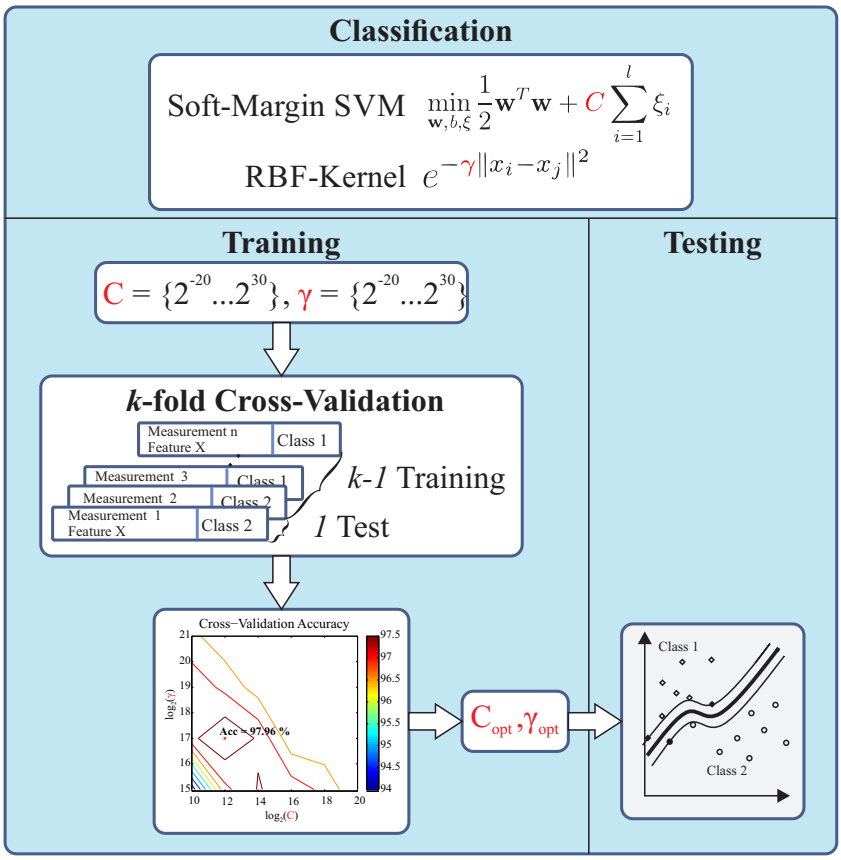


Figure 5.12: Overview of the classification module [SWWS16]

The generation of features using wavelets introduces diversity of choices according to the wavelet method and mother wavelet selection, and the level of decomposition, in addition to the way in which the generated subcomponents are handled. An illustration of the different approaches for features generation using wavelets is presented in Fig. 5.13.

In order to improve the classification process, the features set has to be manipulated for the purposes of

- generating of further features,
- improvement of the existing features, and
- the combination of features.

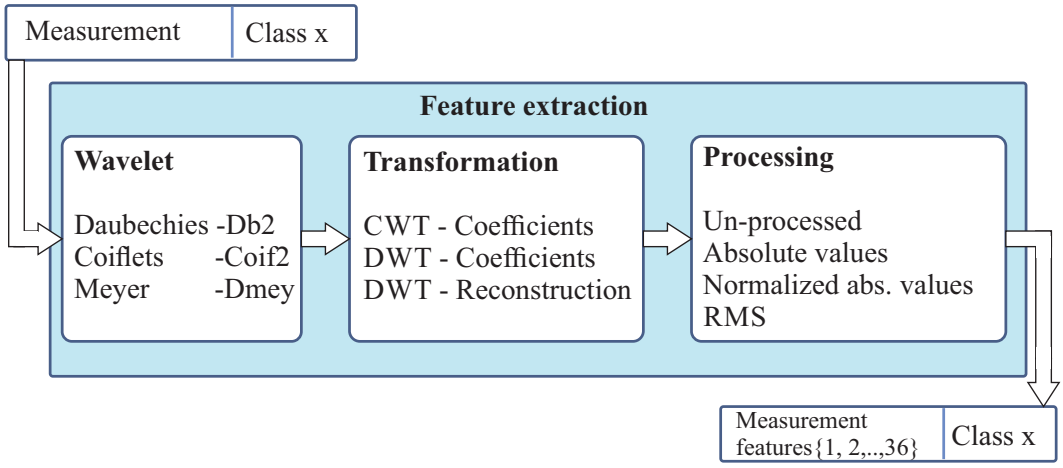


Figure 5.13: Different alternative feature generation procedures [SWWS16]

The information content of the detail and approximation levels of the DWT represents a possible indicator of the system states and accordingly a source of new features. In order to improve the existing features, some features are subjected to elimination because of the low content of classification indicators. The elimination of the coefficients with low classification abilities enables more focus on the data segments with higher CV accuracy. The proposed procedure for this concept is:

1. The best CV for all coefficients is determined.
2. One coefficient is eliminated and the CV is determined again.
3. In case that the CV accuracy after elimination is at least the same as before, the CV is assigned as the new best CV and the coefficient is permanently eliminated. Otherwise, the coefficient remains in the feature set.
4. Steps 2 and 3 are implemented for all the coefficients.

In order to combine features to generate new ones, some statistical measures are used. The Root Mean Square (RMS), as an example, is used to quantify the wavelet coefficients generated within the analyzing window, as a measure of the magnitude of varying quantities.

5.4.1 Implementation and discussion

The sensor data provided from the considered rotor consist of four time series of vibration acceleration signals (Lx, Ly, Rx, and Ry) taken at the two bearing ends (L and R) of the rotor in two independent coordinates, horizontal (x) and vertical

(y), taken from a stationary region. A time window of 1 second with a sample rate of 10 kHz is applied. A state set of measurements consists of 51 measurement signals of the size of 5000 points. The rotational velocity of the rotor remains constant on 9000 rpm.

To apply the DWT, an analyzing window of a suitable length is shifted across the data stream generating the wavelet parameters at a suitable level of the wavelet analysis. In this contribution the SVM method is used, therefore after a training phase a model for the classification of test data is developed. The corresponding transformation defines a dependency (mapping) between the indicating features and the system state using a separating hyperplane with a maximum separation. The main advantage of the SVM is its generalization ability. Here the maximum margin criterion in the process of selecting the separating hyperplane can be realized. Another advantage of SVM is its robustness against signal-related outliers using the so called penalty parameter, which allows controlling the misclassification error.

The applied data set is divided into two data subsets. The first one is a training data set of four states; no crack and three different small sizes of crack (hr = 0,0001, 0,0005, and 0,001). The second data subset includes one state of comparatively bigger size of crack leading to chaotic behavior of the rotor. An enlarged example of the four measured signals is illustrated in Fig. 5.14 for stationary crack-free behavior and hr=0.022 cracked rotor. The signals are understood as measurements taken from the system from those nodes which can be measured. In general, all signals shown, result from simulations of the cracked and uncracked rotor.

The two data subsets are tested in order to investigate the influence of different artificial disturbances on the classification ability. Two kinds of disturbances are used. The signal $S(t_i)$ is disturbed by a random noise as

$$S_{Rnd}(t_i) = S(t_i) + Random\{-10^{-5}...10^{-5}\} \tag{5.6}$$

and by an amplitude-dependent disturbance as

$$S_{Amp}(t_i) = S(t_i)(1 + 0,5Random\{0...1\}). \tag{5.7}$$

The influence of the disturbances on the signal Lx is enlarged as shown in Fig. 5.15.

The Cross Validation (CV) is used as a measure for the quality of classification, if no test dataset is available. The application of scanning window is adjusted to generate 51 measurements for each crack state in order to have sufficient CV estimation for the classification ability.

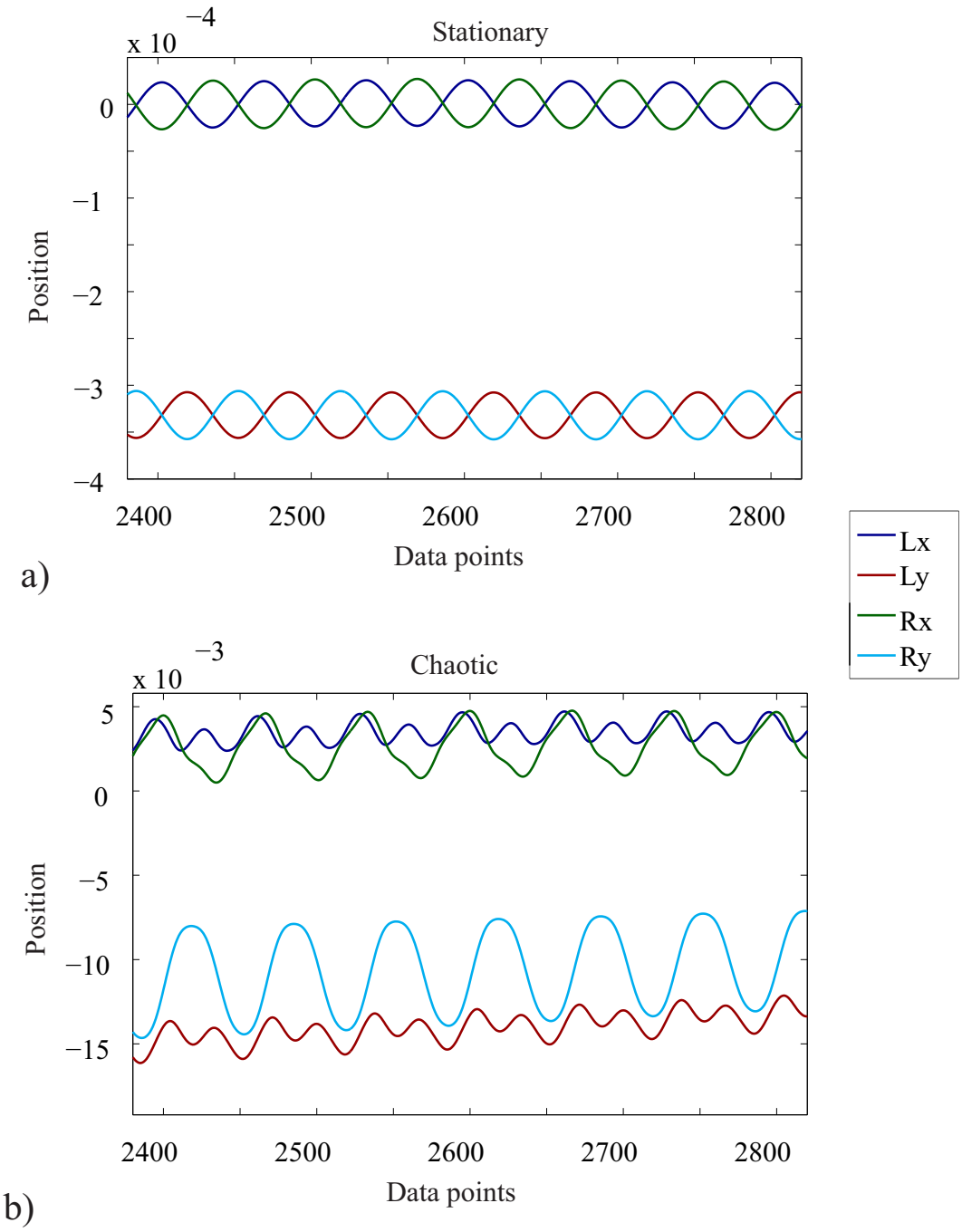


Figure 5.14: Signal sample with a) no-crack stationary behavior, and b) crack chaotic behavior (hr = 0,022)

The four crack states are considered as four classes for training. The classification results are summarized in Table 5.1. The best classification ability (100%) is achieved by the vertical measurements (y) in case of random noise (Rnd.). The CV

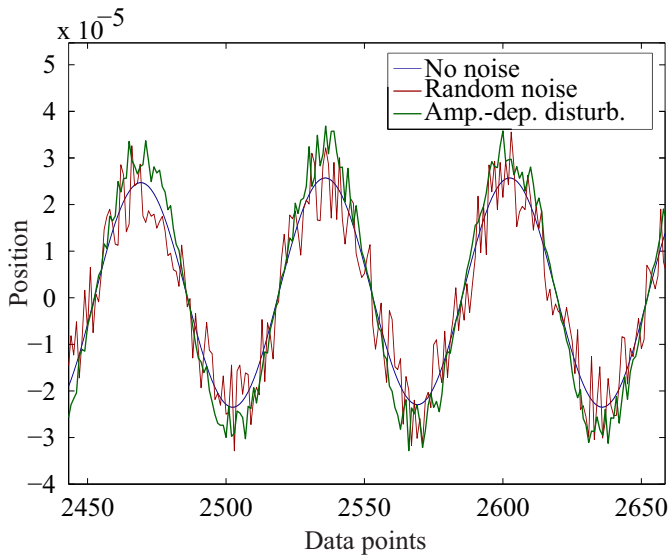


Figure 5.15: Signal sample Lx with different types of noise

of the measurements in the right side of the rotor (Rx and Ry) is generally better, in which are the levels 5 and 6 give the best scale level of the DWT. These two levels are generally better also in case of amplitude-dependent disturbance (Amp.), in which the horizontal measurements (x) perform better.

Furthermore, the three crack states are combined in one class as cracked rotor. By not considering the differences between the crack sizes, the focus is put on size-independent classification. The classification results are summarized in Table 5.2. The same previous number of measurements is considered, thus 51 measurements for the class "no crack" and 153 measurements for the class "crack", accordingly, the worst CV accuracy is 75%. The results are coinciding with those presented in Table 5.1. In addition to the approximation level of the vertical measurements (y), the detail levels 4-6 are best candidates for classification. As a conclusion, the classification between the states no crack and small sized crack is considered reliable.

Additionally, a larger size of crack (hr=0,022) leading to chaotic behavior of the rotor is generated by 51 measurements in order to investigate classification ability against a state of no crack. As a result, and considering all the features together, a CV accuracy of 100% is achieved independent of the state of the disturbance. In order to investigate the robustness of the classification, the classifier is trained by the four classes mentioned in Table 5.1 and tested by the bigger sized crack mentioned above. Individual features and minimum CV accuracy of 70% are considered. The allocation of the 51 test measurements into the four classes is introduced in Table 5.3. With the exception of Rx and one Ry measurements with random noise, all the other individual features classify the test measurements as cracked. This is an indication of reliable parameters for classification. Inconsistently, the test measurements in the table are most frequently classified as 0,0005 class crack.

Table 5.1: CV results for classification of 4 crack classes with different noises (%)

Feature	**Lx Rnd.**	**Lx Amp.**	**Ly Rnd.**	**Ly Amp.**	**Rx Rnd.**	**Rx Amp.**	**Ry Rnd.**	**Ry Amp.**
RMS	25	25,9	29,9	24	40,2	62,7	64,2	29,9
Detail 1	36,3	36,2	29,9	53,9	25	50	28,9	37,7
Detail 2	33,3	25	27,9	42,1	35,2	50,9	30,8	38,7
Detail 3	30,8	38,7	37,7	35,7	25,9	43,1	35,2	40,1
Detail 4	28,4	71,1	32,3	33,3	49	72,5	60,2	32,8
Detail 5	44,1	77,4	47,5	30,8	70	89,2	83,3	46
Detail 6	37,2	63,7	50	32,3	75,9	89,7	87,7	44,1
Detail 7	25,9	25,9	27,4	26,4	35,2	37,2	37,2	29,9
Detail 8	31,3	44,6	32,8	34,8	33,8	52,4	32,8	25
Detail 9	29,9	36,2	25	35,2	34,8	50	27,9	31,8
Detail 10	29,9	37,7	33,3	26,9	34,8	36,2	32,8	33,3
Approx.	32,8	27,9	100	93,6	28,4	35,7	100	88,7
All feat.	36,2	55,4	58,8	65,6	57,8	79,4	82,8	54,9

A signal sample for a period of 4 seconds of rotor run is considered, including the stage of starting up, in order to further investigate the separability of the size of crack states and the remaining service life. To apply the DWT, an analyzing window of a suitable length is required. Here the most appropriate wavelet mother function is used. Different wavelet mother functions are tested on the system in different levels and parameters in order to find the most crack state indicators separating the states. The root mean square (RMS), as a measure of the magnitude of varying quantities, is used to quantify the wavelet parameters generated within the analyzing window. The tested wavelet mother functions include haar, dmey, sym, and db. The best results are obtained in the decomposition level 6 using the discrete meyer (dmey) wavelet mother function (Fig. 5.16), characterized by the highest separability of the crack indicators and a homogenous applicability independent of the place of application within the data. It should be noted that the choice of the mother wavelet as well as the observed result that using the coefficients from decomposition level 6 will give the best results, can not be generalized. This result is (as usual using these approaches) obtained by practical comparisons.

The resulting RMS measures and a related moving average smoothing of the noisy signal Ry are shown in Fig. 5.17. The results are presented for 4 different crack size levels of the considered rotor. It can be seen that the separability of the crack levels is not affected by the non-stationary startup of the rotor. It can also be seen, that based on fewer measurements, good results can be achieved.

For comparison, alternative four different results are shown in Fig. 5.18. The pre-

Table 5.2: CV results for classification of 2 crack classes with different noises (%)

Feature	**Lx Rnd.**	**Lx Amp.**	**Ly Rnd.**	**Ly Amp.**	**Rx Rnd.**	**Rx Amp.**	**Ry Rnd.**	**Ry Amp.**
RMS	75	75	75	75	75	75	75	75
Detail 1	75	75	75	75	75	75,4	75	75
Detail 2	75	75	75	75	75	75	75	75
Detail 3	75	75	75	75	75	75	75	75
Detail 4	75	75	75	75	79,4	81,3	82,3	75,9
Detail 5	75	79,9	75	75	76,9	88,7	81,3	75
Detail 6	75	75	75	75	78,9	91,6	86,2	75
Detail 7	75	75	75	75	76,9	75	75,9	75
Detail 8	75	75	75	75	75	75	75	75
Detail 9	75	75	75	75	75	75	75	75
Detail 10	75	75	75	75	75	75	75	75
Approx.	75	75	100	95,1	75	75	100	88,7
All feat.	75	75,9	75,9	75,4	75	79,4	81,8	75

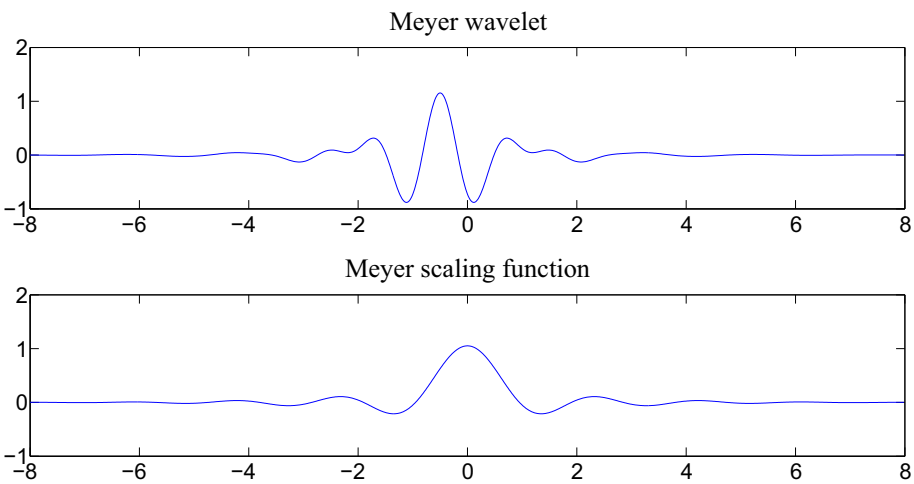


Figure 5.16: Meyer wavelet mother function

sented figures result from the same dataset used in Fig. 5.17 but using less appropriate DWT parameters and mother wavelets. In the results presented in Fig. 5.18(a) and Fig. 5.18(b), the high level of fluctuation and the non-homogenous solution prevent the applicability of the solution. The other two wavelet functions presented in Fig. 5.18(c) and Fig. 5.18(d) result in RMS range of cracks 1.82e-6 and 1.83e-6 respectively, whereas the selected wavelet (Fig. 5.17) results in higher RMS range (2.02e-6) which means better separability of the different crack sizes and accordingly more accurate results of the crack detection and evaluation system. In a window of

Table 5.3: Classification of the chaotic crack state in 4 crack classes

Measur. Disturb.	**Feature**	**CV** [%]	**Crack size**			
			hr=0,0	**hr=0,0001**	**hr=0,0005**	**hr=0,001**
Lx Amp.	Detail 4	71,1	0	0	0	51
	Detail 5	77,4	0	0	0	51
Ly Rnd.	Approx.	100	0	0	51	0
Ly Amp.	Approx.	93,6	0	0	51	0
Rx Rnd.	Detail 5	70	19	8	23	1
	Detail 6	75,9	4	17	30	0
Rx Amp.	Detail 4	72,5	0	0	38	13
	Detail 5	89,2	0	0	51	0
	Detail 6	89,7	0	7	44	0
Ry Rnd.	Detail 5	83,3	0	0	51	0
	Detail 6	87,7	0	0	51	0
	Approx.	100	1	0	50	0
Ry Amp.	Approx.	88,7	0	1	50	0

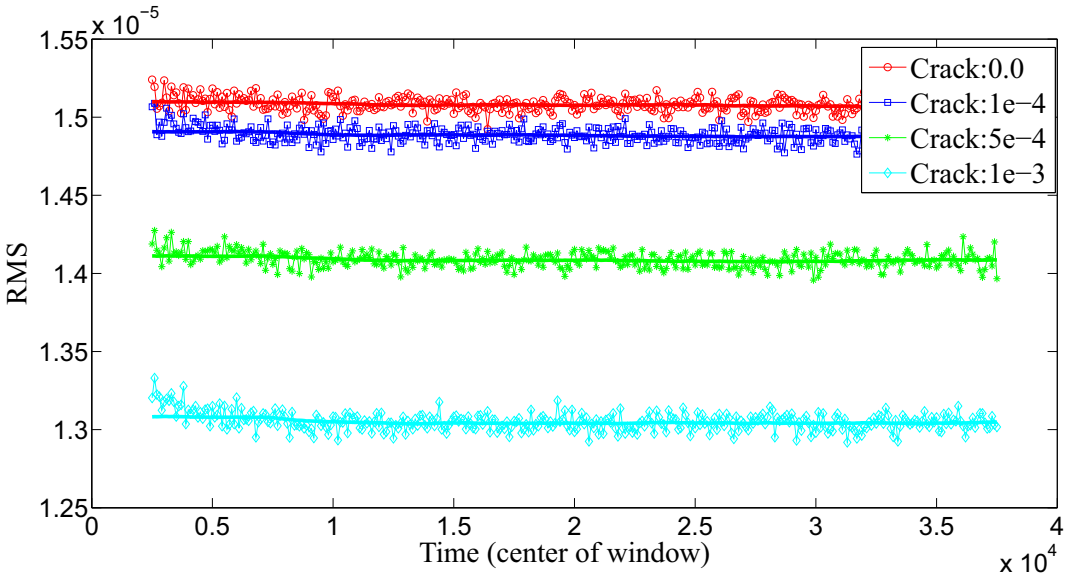
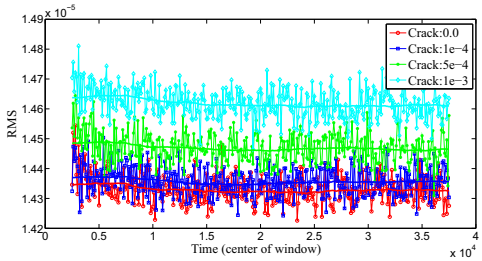


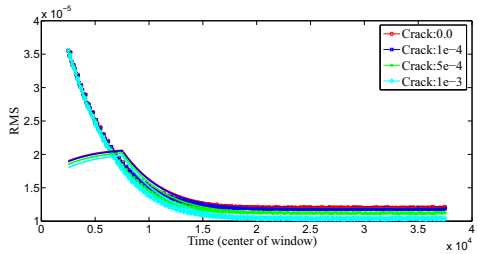
Figure 5.17: RMS and smoothing; Ry, dmey, level 6 [SWWS16]

0.5 second at the end of the data, the effect of changing the rotor system damping is presented as minimum in Fig. 5.19 for the selected wavelet (Fig. 5.17).

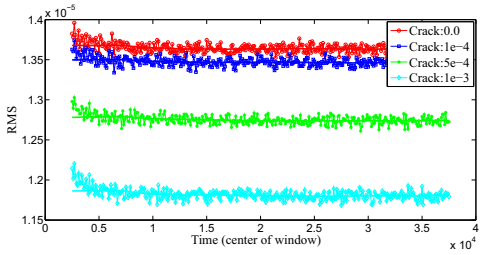
The task of the SVM classification module is to evaluate the extracted features which contain the indicators of the system state in order to achieve a statement about the existence and size of the cracks and faults to be detected. The fusion function of the



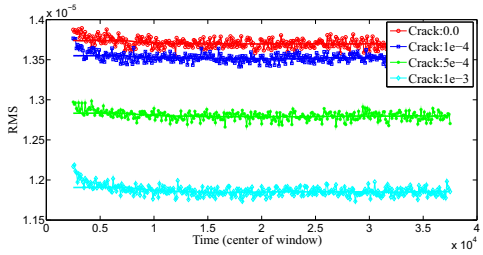
(a) RMS and smoothing; Ly, dmey, 6



(b) RMS and smoothing; Ry, haar, 6



(c) RMS and smoothing; Ry, db4, 6



(d) RMS and smoothing; Ry, sym, 6

Figure 5.18: RMS and smoothing of different combinations for comparison [SWWS16]

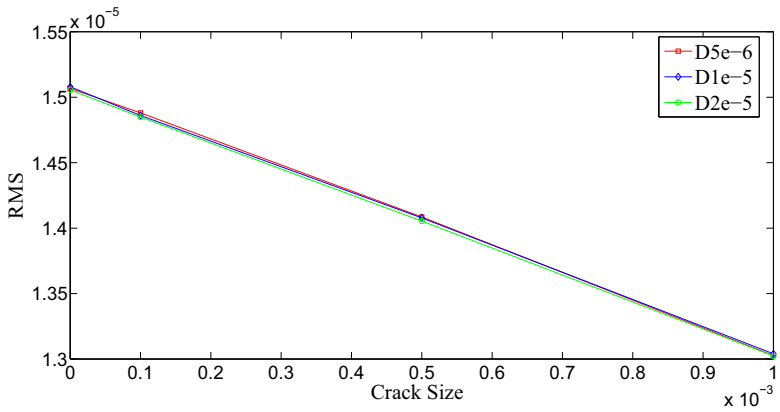


Figure 5.19: Effect of damping; Ry, dmey, 6 [SWWS16]

SVM classifier helps to obtain a more reliable complementary sensor array of the 4 filtered sensor data provided by the rotor system.

The required evaluation of the extracted features can be provided in two ways: binary classification and multiclass classification. In the binary classification the classifier is trained using binary training data to classify two classes; here: cracked

and non-cracked rotor. In this case the required quantification of the crack size is calculated using the position and distance of the specified state from the separating hyperplane in the feature space. In the case of using the multiclass classification, the training data used to train the classifier are grouped according to the size of the crack into segments coinciding with the required scale and accuracy of the size of the cracks. Furthermore, fine tuning of the results can be done using the distance of the state from the corresponding separating hyperplane in the feature space.

The training result of the multiclass classification of the previously mentioned rotor implementation is given in a two-dimensional view in Fig. 5.20. The class boarders of the crack size classes can be clearly detected. It can also be seen that the data resemble one-directional trajectory in the direction of the increasing crack size.

The decision function of the binary classification of the rotor implementation is presented in Fig. 5.21. The decision boarder in the figure represents the separating hyperplane between the two classes of the training data; cracked and non-cracked rotor. The 4 training datasets of the 4 different crack sizes which are mentioned in Fig. 5.20 are recognized in Fig. 5.21 as different distances from the decision boarder. In case of more data samples available representing all the crack sizes in between, the decision function would resemble a one-directional trajectory in the direction of the increasing crack size. In this case the size of the crack can be defined implicitly by using the distance from the decision boarder.

For remaining life prognostic purposes, the decision function can be undergone an extrapolation process to reach a pre-trained crack safety limit, as shown in [SS11].

The previous discussion indicates that the binary classification may be more suitable for the crack size quantization. This is concluded from the one-directional trajectory of the decision function in the direction of the increasing crack size. This conclusion is supported by the fact that the SVM classification is originally and more efficiently a binary classification [Abe10].

5.5 Discussion of the approaches to be compared

Many publications have been published presenting the successful applications of many methods representing the two mentioned approaches of monitoring system design. Indeed, it is quite difficult to estimate how successful an approach is, compared to the other for many reasons. In the literature, the successful implemented algorithms have generally been limited to faults which are basic in comparison. Very few publications have dealt with realistic behaviors of faults and cracks which might appear even simultaneously. Additionally, most of the published systems have been implemented and validated on laboratory structures which could essentially differ from the real world industrial systems. The lack of benchmark measurements and

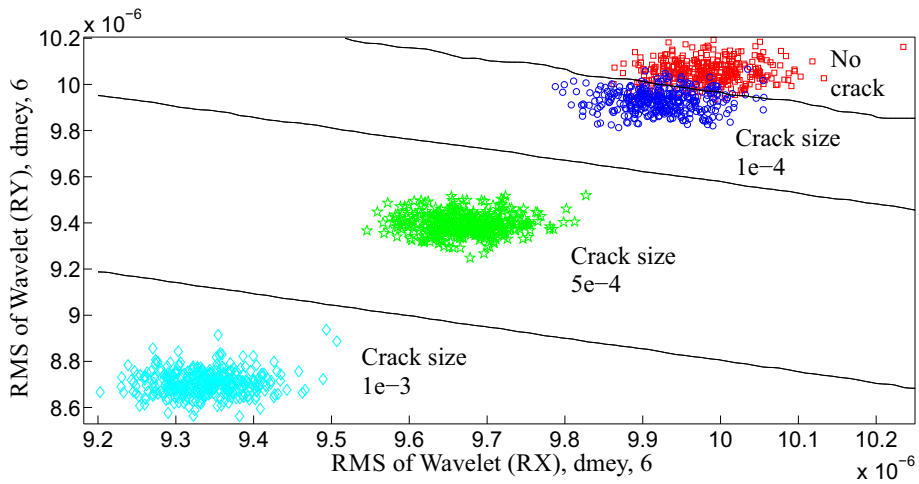


Figure 5.20: The SVM feature space [SWWS16]

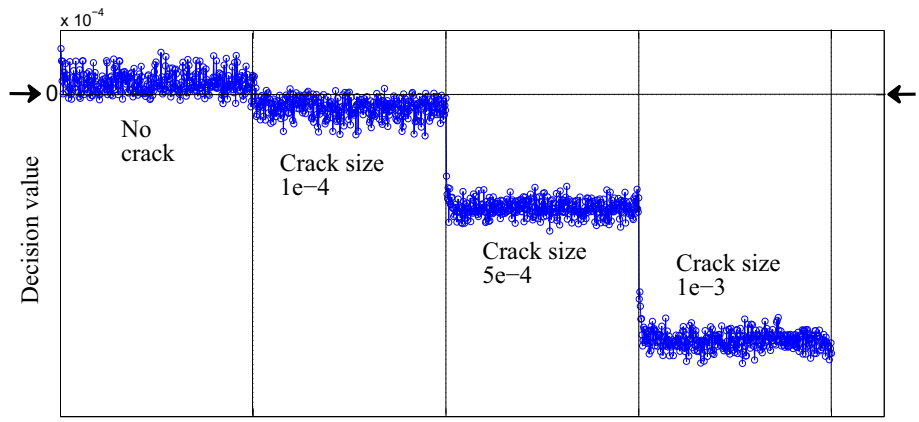


Figure 5.21: The SVM decision function [SWWS16]

assessment criteria for the systems make it even harder to compare. However, some general inferences can be concluded from the study of both approaches.

The main advantage of model-based approaches is that the usually available, very detailed, and physical-oriented understanding of the fault and crack effects acting to the system is preserved and can be used by the approach to compensate measurements. The typical and known interpretation problem of signal-based approaches is avoided using problem-related indicators like the introduced stiffness change in rotating coordinates, showing a measurement-based reconstruction of the physical breathing of the cracked rotor. The introduced approaches allow also the implicit definition of the crack depth, if the measurements are noise-free. One of the drawbacks of this kind of approaches is the necessity of fault models, as well as the assumed hypotheses about the location of the fault or the crack. In general the success of all methods of analytical redundancy is essentially defined by the quality of the

model. Modeling faults inevitably lead to errors which could lead with the possibly existing real effects from the rotor dynamics to the fact that corresponding faults cannot be distinguished in principal. Indeed, a successful rotor modeling should consider the dynamics of the related foundation which is much more complicated.

Modern signal-based approaches, on the other hand, are based on robust machine learning and feature extraction techniques, making the detection more reliable and robust even for small cracks. The used methods are more robust against disturbances and noises. This advantage makes them more suitable for realistic conditions in the industry. Additionally, the simplicity of the signal-based techniques makes them more appropriate for real time systems.

Considering the targets of the detection systems, more reliable fault localization could be achieved by model-based systems, although multiple fault localization requires the establishment of multiple model system with locations hypothesis. On the other hand, fault severity estimation could be more effective using modern feature extraction and classification techniques adopted by signal-based systems.

The field of sensor and decision fusion is one possible method to combine the two main approaches [ASSS14]. Another alternative is to use a model-based system to recognize the changing service conditions and operation setpoint, in order to adjust the adaptable feature extraction and classification in the concurrent signal-based system.

5.6 Short summary and conclusion

The two main approaches used for crack detection and prediction in rotating machinery; model-based and signal-based approaches, are investigated by one typical example. Several strength and weak points are discussed and compared for the two approaches using two representative applicable methods, in order to achieve a comparative overview of the available techniques.

Beside an observer-based solution predicting crack depth related information, a new signal-based/data-driven approach is introduced to improve the detection problem with respect to noise.

The PI-observer-based method is considered as a modern model-based technique, to give indication about possibilities and limitations of such kind of methods. A novel signal-based approach is introduced, based on SVM and wavelets as modern machine learning techniques.

Modern machine learning techniques are found more robust against disturbances and noises, whereas the model-based techniques are more adaptable with load changes of the system and more able to be connected with system physics and modeling parameters.

6 Summary and outlook

In this chapter, a summary of the thesis is introduced with brief conclusions of the work. The scientific contribution of the thesis is presented briefly, and some future prospects are mentioned at the end.

6.1 Summary and conclusion

In this work, the application of pattern recognition techniques is considered for different kinds of fault diagnosis and prognosis problems and applications. The investigated applications represent real industrial applications, in which different measurement characteristics (such as cyclic, impulsive, and periodic signals), different recognition objective characteristics (such as accumulative and one-time events), different operational conditions and parameters of the machine, and different faults and detection system requirements (such as wear, crack, and object detection; system state and remaining life time) are challenging the existence of modular pattern recognition procedures and techniques. Different approaches are investigated and applied such as SVM, DWT, WPT, and CWT, and many concepts and solutions are proposed and verified, in order to achieve a reliable condition monitoring system, which supports the maintenance planning of the machine and adds value to the production quality and cost.

In the first application, a production machine related supervision task is investigated over a long duration. An approach for developing a fault diagnosis and prognosis system to support condition-based maintenance of wear parts is presented. The system is used as a prewarning module to detect the necessity for replacing wear parts of production machines and to evaluate the remaining life time of the supervised part. Wear parts failure is detected before scuffing or seizing lead to serious failure of the machine. The sensor signals encountered for processing are nondeterministic with cyclic nature related to the operation cycle of the machine. Real industrial process and sensor signals are considered in this application. However, in principle the developed approaches can include arbitrary data acquisition techniques. By applying the proposed methods, a considerable improvement in the accuracy and robustness of the solution is achieved.

In the second application, the goal is to monitor a production process for online detection of a target object with the lowest possible false alarm rate. The application is a stationary process with a continuous stream of sensor signals. The individual events with no consequent and accumulating effects investigated in this work represent many possible process and fault related incidents. The signals encountered in the system of this work are characterized with nonstatinary impulsive one-time events representing the goal object. Another characteristic of the sensor cluster signals is the partly simultaneous stimulation of events which prevents the advantage

of using exact simultaneous events for fusing the data in the features level, which requires the use of suitable decision fusion techniques. Real industrial process and sensor signals are considered in this application. However, in principle the developed approaches can include arbitrary data acquisition techniques. The results of the proposed methods reveal improved detection and false alarm rates.

In the last application, two main approaches used for crack detection and prediction in rotating machinery; model based and signal based, are investigated, in order to achieve a prewarning technique for rotor cracks to be applied for online monitoring in turbo-machinery. Several strength and weak points are discussed and compared, in order to achieve a comparative overview of the available techniques. The concepts fault severity estimation and remaining life time prediction are considered in the application. The signals encountered are periodic vibration signals with accumulative impact of the fault incident. Realistic simulation example is considered in this chapter. However, in principle the developed approaches can include arbitrary data acquisition techniques.

As a general conclusion of the work, it can be stated that Wavelets and SVM are reliable tools for feature extraction and classification in the field of condition monitoring, and the feature space of SVM is useful for remaining life prediction. However; specific application oriented solutions and tricks are necessary, considering the diversity of fault diagnosis and prognosis problems and difficulties. More different realistic applications and benchmark data are still required, in order to achieve sufficient insight into this diverse field.

6.2 Scientific contributions

Several scientific papers have been published based on the studies and results of the work in this thesis. These papers are mentioned as references in the beginning of the main chapters; Chapter 3, 4, and 5.

The applications investigated in this work represent real industrial applications, in which different situations and circumstances are challenging the existence of modular pattern recognition procedures and techniques. The investigated situations involve measurement characteristics, recognition objective characteristics, operational conditions of the machine, and detection system requirements. The diversity of the investigated problems helps to aware of difficulties and obstacles encountered in real life industry.

More specifically, some of the major contributions of the work are summarized in the following.

A concept for state evaluation and remaining life prediction based on the decision value of the SVM is proposed (Chapter 3 and Chapter 5). The proposed concept is

supported by the solidity and simplicity of the feature space concept in the SVM. The decision boarder of the SVM represents the separating hyperplane in the feature space. Consequently, the decision value is a measure of how far a measurement is from the separating hyperplane. Accordingly, the decision value is an indication of how much the information of a system state contributes to a particular class. Moreover, the measurements (and hence the system state) move between the class zones across the feature space. This movement, which occurs over time coinciding with the deterioration of the monitored system, is an indication of the rate of deterioration. The proposed concept provides a useful tool for severity quantification of faults and can provide a reliable solution to predict the remaining service life of a mechanical system.

Moreover, the proposed Change Index (CI) (Chapter 3 and Chapter 5) helps for online estimation of state and provides better indication of the critical parts at any time point of operation. The proposed index is more convenient for practical use as an indication of how "worn" a part is, and (in case of enough operation parameters provided) how long it will last. It is a measure function of the distance between the current position of the decision value and the maximum allowed position of the decision value before changing of the monitored part is necessary.

For the case of cyclic machine operation, the method of feature-based resampling is proposed (Chapter 3), in order to deal with the difficulties of the changing lengths of operation cycles. This kind of operation and behavior is familiar in many production systems. The proposed method suggests more reliable features as reference for the process of resampling, which is necessary for the construction of the input matrix of the classifier. The reliable features are based on a selective reconstruction process of DWT components, in order to provide more comparable operation cycles with more comparable features, which allow better localizing of the segments of the cycle for resampling. The proposed method considers the disturbed time scale of the resampled cycle by considering the original positions of the key features as part of the input matrix of the classifier.

Finally, a decision fusion technique is proposed (Chapter 4). The technique is developed as scale-invariant, based on cascade SVM classifiers, in order to deal with the inevitable and varying time shift between the event stimulation of the individual sensors. The technique is developed based on cascade SVM classifiers, as scale-invariant, for the fact that the effect of noise and disturbance signals to the system and sensors is inevitable, and the intensity of the sensor stimulation is not related to the size of the object to be detected. The technique combines the individual preliminary decisions to obtain the highest possible detection rate for the lowest possible false alarm rate. This proposed decision fusion technique can be applied to monitor various specific complex systems with similar characteristics.

6.3 Future prospects

As a continuation of the current work, some guidelines, which might be helpful, are worth mentioning. For the concepts of remaining life prediction and the Change Index (CI), more specific framework is required, in order to construct an extensive technique for remaining life modeling, considering changing operation parameters and different types of faults and incidents. It is also essential for the improvement of the concepts, to develop more reliable measurement techniques for the SVM feature space. It could be helpful to adopt mathematical techniques developed for multidimensional spaces, and techniques developed for dimensionality reduction. For the method of feature based resampling, and the method of scale invariant decision fusion, the applicability for various complex problems could be tested for better generalization of the techniques.

Bibliography

[AAN08] M. Ayat, K. Assaleh, and H. Nashash, "Fetal ecg extraction from a single abdominal ecg signal using svd and polynomial classifiers," in *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, Oct 2008, pp. 250–254.

[Abe10] S. Abe, *Support vector machines for pattern classification*, S. Singh, Ed. London, England: Springer-Verlag London Limited, 2010.

[AC09] G. Acuna and M. Curilem, "Comparison of neural networks and support vector machine dynamic models for state estimation in semiautogenous mills," in *MICAI 2009: Advances in artificial intelligence*, ser. Lecture Notes in Computer Science, A. Aguirre, R. Borja, and C. Garcia, Eds. Springer Berlin Heidelberg, 2009, vol. 5845, pp. 478–487.

[ACSK11] D. An, J.-H. Choi, T. L. Schmitz, and N. H. Kim, "In situ monitoring and prediction of progressive joint wear using bayesian statistics," *Wear*, vol. 270, pp. 828–838, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043164811000627

[Add05] P. S. Addison, "Wavelet transforms and the ECG: a review," *Physiological Measurement*, vol. 26, no. 5, p. R155, 2005. [Online]. Available: http://stacks.iop.org/0967-3334/26/i=5/a=R01

[AKJ02] S. Antani, R. Kasturi, and R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video," *Pattern Recognition*, vol. 35, no. 4, pp. 945–965, 2002. [Online]. Available: http://dblp.uni-trier.de/db/journals/pr/pr35.html#AntaniKJ02

[ALPM+13] F. Amato, A. López, E. M. Peña-Méndez, P. Vanhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine*, vol. 11, no. 2, pp. 47 – 58, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1214021X14600570

[ARK10] I. Arel, D. C. Rose, and T. P. Karnowski, "Research frontier: Deep machine learning–a new frontier in artificial intelligence research," *Comp. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, Nov. 2010. [Online]. Available: http://dx.doi.org/10.1109/MCI.2010.938364

[AS08] F. Alonso and D. Salgado, "Analysis of the structure of vibration signals for tool wear detection," *Mechanical Systems and Signal*

Processing, vol. 22, no. 3, pp. 735–748, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007001902

[AS11] H. Aljouma and D. Söffker, "Multi-class approach based on fuzzy-filtering for condition monitoring," *IAENG International Journal of Computer Science*, vol. 38, no. 3, pp. 66–73, 2011.

[ASSS11] L. Al-Shrouf, M.-S. Saadawia, N. Szczepanski, and D. Söffker, "Adaptive classification based on multisensoric decision fusion," in *8th Int. Workshop on Structural Health Monitoring 2011*, F. Chang, Ed., Stanford, USA, September 2011, pp. 127–134.

[ASSS14] L. Al-Shrouf, M.-S. Saadawia, and D. Söffker, "Improved process monitoring and supervision based on a reliable multi-stage feature-based pattern recognition technique," *Information Sciences*, vol. 259, no. 0, pp. 282–294, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025513007202

[ASSSS10] L. Al-Shrouf, M.-S. Saadawia, N. Szczepanski, and D. Söffker, "Multisensorfusionsbasierte Prozessüberwachung," in *Berichtsband 8. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2010, pp. 413–422.

[ASSSS12] ——, "Event detection using multisensor fusion and filtering techniques based on CWT and SVM," in *6th European Workshop on Structural Health Monitoring 2012*, C. Boller, Ed., Dresden, Germany, July 2012, pp. 206–213.

[BADB14] D. Bhatt, P. Aggarwal, V. Devabhaktuni, and P. Bhattacharya, "A novel hybrid fusion algorithm to bridge the period of {GPS} outages using low-cost {INS}," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2166–2173, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741741300746X

[BBK10] J. Basu, D. Bhattacharyya, and T.-H. Kim, "Use of artificial neural network in pattern recognition," *International Journal of Software Engineering and Its Applications*, vol. 4, no. 2, pp. 23–34, 2010.

[BBST13] L. Batista, B. Badri, R. Sabourin, and M. Thomas, "A classifier fusion system for bearing fault diagnosis," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6788 – 6797, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413004284

[BCV13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[BD12] T. Banerjee and S. Das, "Multi-sensor data fusion using support vector machine for motor fault detection," *Information Sciences*, vol. 217, no. 0, pp. 96–107, 2012.

[Ben09] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[BGLD12] G. Bin, J. Gao, X. Li, and B. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets: Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, no. 0, pp. 696–711, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327011003207

[BGRS99] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *Proc. Int. Conf. on Database Theory*, 1999, pp. 217–235.

[Bis06] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[BKL+06] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder, *Diagnosis and Fault-Tolerant Control*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[BLS+13] S. Bagavathiappan, B. Lahiri, T. Saravanan, J. Philip, and T. Jayakumar, "Infrared thermography for condition monitoring: A review," *Infrared Physics & Technology*, vol. 60, no. 0, pp. 35–55, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449513000327

[BPT10] N. Bachschmid, P. Pennacchi, and E. Tanzi, *Cracked Rotors: A Survey on Static and Dynamic Behaviour Including Modelling and Diagnosis*. Springer-Verlag, 2010. [Online]. Available: http://books.google.de/books?id=FbPiJDxOG2AC

[Bru82] Bruel & Kjaer, "Measuring vibration," 1982. [Online]. Available: http://www.bksv.com/doc/br0094.pdf

[BTvdH08] G. Bailador, G. Trivino, and A. van der Heide, "Fuzzy Sets of Quasi-Periodic Signals," in *XIV Congreso español sobre tecnología y lógica fuzzy (ESTYLF 2008)*, Sep. 2008.

[Bur98] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[BWR13] G. Bhatnagar, Q. Wu, and B. Raman, "Discrete fractional wavelet transform and its application to multiple encryption," *Information Sciences*, vol. 223, pp. 297–316, 2013.

[CDZT14] F. Cao, T. Dai, Y. Zhang, and Y. Tan, "Compressed classification learning with markov chain samples," *Neural Networks*, vol. 50, no. 0, pp. 90–97, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608013002724

[CF04] E. P. Carden and P. Fanning, "Vibration based condition monitoring: A review," *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, 2004. [Online]. Available: http://shm.sagepub.com/content/3/4/355.abstract

[CL11] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[CLS05] P. Chen, C.-J. Lin, and B. Schölkopf, "A tutorial on ν-support vector machines," Department of Computer Science, National Taiwan University, Tech. Rep., 2005. [Online]. Available: http://www.csie.ntu.edu.tw/\~cjlin/papers/nusvmtutorial.pdf

[CR00] J. Chapa and R. Rao, "Algorithms for designing wavelets to match a specified signal," *Signal Processing, IEEE Transactions on*, vol. 48, no. 12, pp. 3395–3406, Dec 2000.

[CV95] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[CV08] F. Camastra and A. Vinciarelli, *Machine Learning for Audio, Image and Video Analysis: Theory and Applications*, ser. Advanced Information and Knowledge Processing. Springer, 2008. [Online]. Available: http://books.google.de/books?id=YXPqVIHaVYsC

[DDDLV07] A. De Decker, G. De Lannoy, and M. Verleysen, "Functional SOM for variable-length signal windows," in *Proceedings of the 6th International Workshop on Self-Organizing Maps WSOM 2007*. University of Bielefeld, 2007.

[DHS00] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[DZG+14] J.-X. Du, C.-M. Zhai, Y.-L. Guo, Y.-Y. Tang, and P. C. C. Lung, "Recognizing complex events in real movies by combining

audio and video features," *Neurocomputing*, no. 0, pp. –, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214002343

[ELF98] S. Edwards, A. W. Lees, and M. I. Friswell, "Fault diagnosis of rotating machinery," *The Shock and Vibration Digest*, vol. 30, no. 1, pp. 4–13, 1998.

[EPK06] S. Ebersbach, Z. Peng, and N. Kessissoglou, "The investigation of the condition and faults of a spur gearbox using vibration and wear debris analysis techniques," *Wear*, vol. 260, pp. 16–24, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043164804004521

[FGRX07] L. Fortuna, S. Graziani, A. Rizzo, and M. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*, ser. Advances in Industrial Control. Springer, 2007. [Online]. Available: http://books.google.de/books?id=q0PPVrZAsNIC

[FMD14] M. A. Ferrer, A. Morales, and A. Díaz, "An approach to {SWIR} hyperspectral hand biometrics," *Information Sciences*, vol. 268, no. 0, pp. 3–19, 2014, new Sensing and Processing Technologies for Hand-based Biometrics Authentication. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025513007366

[FPSV14] P. Foggia, G. Percannella, A. Saggese, and M. Vento, "Pattern recognition in stained hep-2 cells: Where are we now?" *Pattern Recognition*, vol. 47, no. 7, pp. 2305 – 2314, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320314000284

[GA08] S. Georgantzinos and N. Anifantis, "An insight into the breathing mechanism of a crack in a rotating shaft," *Journal of Sound and Vibration*, vol. 318, no. 1-2, pp. 279–295, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X08003398

[Gas93] R. Gasch, "A survey of the dynamic behaviour of a simple rotating shaft with a transverse crack," *Journal of Sound and Vibration*, vol. 160, no. 2, pp. 313–332, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X83710266

[Gas08] ——, "Dynamic behaviour of the laval rotor with a transverse crack," *Mechanical Systems and Signal Processing*, vol. 22, no. 4, pp. 790–804, 2008, special Issue: Crack Effects in Rotordynamics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007002270

[GC05] I. Green and C. Casey, "Crack detection in a rotor dynamic system by vibration monitoring - part i: Analysis," *Journal of Engineering for Gas Turbines and Power*, vol. 127, no. 2, pp. 425–436, 2005.

[GE03] I. Guyon and I. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, March 2003.

[GMR00] K. B. Goode, J. Moore, and B. J. Roylance, "Plant machinery working life prediction method utilizing reliability and condition-monitoring data," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 214, no. 2, pp. 109–122, 2000. [Online]. Available: http://pie.sagepub.com/content/214/2/109.abstract

[GP07] D. Guo and Z. Peng, "Vibration analysis of a cracked rotor using hilbert-huang transform," *Mechanical Systems and Signal Processing*, vol. 21, no. 8, pp. 3030–3041, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007000787

[GSK+06] R. C. Guido, J. F. W. Slaets, R. Köberle, L. O. B. Almeida, and J. C. Pereira, "A new technique to construct a wavelet transform matching a specified signal with applications to digital, real time, spike, and overlap pattern recognition," *Digital Signal Processing*, vol. 16, no. 1, pp. 24 – 44, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200405000618

[GY11] R. Gao and R. Yan, *Wavelets theory and applications for manufacturing*, Wang, Lihui, and et al., Eds. New York, USA: Springer Science + Business Media, 2011.

[Han05] J. Han, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[HCL10] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2010, tech. rep., Department of Computer Science, National Taiwan University. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers.html

[HH02] Y. H. Hu and J.-N. Hwang, Eds., *Handbook of neural network signal processing*. Boca Raton, Florida: CRC Press LLC, 2002.

[HHC+09] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renewable and Sustainable Energy Reviews*,

vol. 13, no. 1, pp. 1–39, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032107001098

[HKK06] T. M. Huang, V. Kecman, and I. Kopriva, *Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-supervised, and Unsupervised Learning.*, ser. Studies in Computational Intelligence. Springer, 2006, vol. 17. [Online]. Available: http://dx.doi.org/10.1007/3-540-31689-2

[HL13] K. Hung and K. Lin, "Long-term business cycle forecasting through a potential intuitionistic fuzzy least-squares support vector regression approach," *Information Sciences*, vol. 224, pp. 37–48, 2013.

[HOT06] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1162/neco.2006.18.7.1527

[HTF08] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2008. [Online]. Available: http://scholar.google.com/scholar.bib?q=info:roqIsr0iT4UJ:scholar.google.com/&output=citation&hl=en&ct=citation&cd=0

[HVG11] D. Hammonda, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[HYLW10] C. Huang, D. Yu, H. Guan, G. Li, and Q. Wu, "Feature extraction of frictional vibration based on CWT time-frequency image," in *Intelligent Control and Information Processing (ICICIP), 2010 International Conference on*, Dalian, China, August 2010, pp. 192–196.

[Ish08] Y. Ishida, "Cracked rotors: Industrial machine case histories and nonlinear effects shown by simple jeffcott rotor," *Mechanical Systems and Signal Processing, Special Issue: Crack Effects in Rotordynamics*, vol. 22, no. 4, pp. 805–817, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007002282

[ISS11] S. Ismail, A. Shabri, and R. Samsudin, "A hybrid model of self-organizing maps (som) and least square support vector machine (lssvm) for time-series forecasting," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10 574–10 578, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417411003137

[IT13] I. Ilhan and G. Tezel, "A genetic algorithm support vector machine method with parameter optimization for selecting the tag snps," *Journal of Biomedical Informatics*, vol. 46, no. 2, pp. 328–340, 2013.

[Iva07] O. Ivanciuc, "Applications of Support Vector Machines in Chemistry," in *Reviews in Computational Chemistry*. John Wiley & Sons, Inc., 2007. [Online]. Available: http://www.ivanciuc.org/Files/Reprint/Ivanciuc_Applications_of_Support_Vector_Machines_in_Chemistry.pdf

[JEEL92] O. Jun, H. Eun, Y. Earmme, and C.-W. Lee, "Modelling and vibration analysis of a simple rotor with a breathing crack," *Journal of Sound and Vibration*, vol. 155, no. 2, pp. 273–290, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022460X9290511U

[JG08] O. Jun and M. Gadala, "Dynamic behavior analysis of cracked rotor," *Journal of Sound and Vibration*, vol. 309, no. 1, pp. 210–245, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X07004993

[JLZD11] X. Jin, L. Lin, S. Zhong, and G. Ding, "Rotor fault analysis of classification accuracy optimition base on kernel principal component analysis and {SVM}," *Procedia Engineering*, vol. 15, no. 0, pp. 5279–5283, 2011, {CEIS} 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705811024799

[JMZ11] R. Juranek, S. Machalik, and P. Zemcik, "Analysis of wear debris through classification," in *Proceedings of Advanced Concepts of Inteligent Vision Systems ACIVS 2011*, ser. LNCS 6915. Springer Verlag, 2011, pp. 273–283.

[Kec05] V. Kecman, "Support vector machines - an introduction," in *Support vector machines: Theory and applications*, ser. Studies in Fuzziness and Soft Computing, L. Wang, Ed. Berlin, Germany: Springer Berlin Heidelberg, 2005, vol. 177, pp. 1–47.

[KH11] S. R. Kulkarni and G. Harman, "Statistical learning theory: a tutorial," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, no. 6, pp. 543–556, 2011. [Online]. Available: http://dx.doi.org/10.1002/wics.179

[KKL13] S. Kim, S. Kavuri, and M. Lee, "Deep network with support vector machines." in *ICONIP (1)*, ser. Lecture Notes in Computer Science, M. Lee, A. Hirose, Z.-G. Hou, and R. M. Kil, Eds., vol. 8226. Springer, 2013, pp. 458–465. [Online]. Available: http://dblp.uni-trier.de/db/conf/iconip/iconip2013-1.html#KimKL13

[KLB+13] I. Khelf, L. Laouar, A. M. Bouchelaghem, D. Rémond, and S. Saad, "Adaptive fault diagnosis in rotating machines using indicators selection," *Mechanical Systems and Signal Processing*, vol. 40, no. 2, pp. 452 – 468, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327013003129

[KR09] C. Kumar and V. Rastogi, "A brief review on dynamics of a cracked rotor," *International Journal of Rotating Machinery*, vol. 2009, pp. 1–6, 2009, Article ID 758108.

[KS09] D. Kim and J. Sung, *Automated Face Analysis: Emerging Technologies and Research*, ser. Premier Reference Source. Medical Information Science Reference, 2009. [Online]. Available: http://books.google.de/books?id=v3WzmAEACAAJ

[KTFME10] T. Kolodziejczyk, R. Toscano, S. Fouvry, and G. Morales-Espejel, "Artificial intelligence as efficient technique for ball bearing fretting wear damage prediction," *Wear*, vol. 268, no. 1, pp. 309–315, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043164809005031

[LDN+07] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, and J. Lee, "Similarity based method for manufacturing process performance prediction and diagnosis," *Computers in Industry*, vol. 58, no. 6, pp. 558–566, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166361506001850

[LF06] A. Lees and M. Friswell, "Where next for condition monitoring of rotating machinery?" *Advances in Vibration Engineering*, vol. 5 (4), pp. 263–277, 2006, other: http://www.tvi-in.com/.

[LS12] Y. Liu and D. Söffker, "Variable high-gain disturbance observer design with online adaption of observer gains embedded in numerical integration," *Mathematics and Computers in Simulation*, vol. 82, no. 5, pp. 847–857, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037847541200002X

[LT12] H. Lin and L. Tseng, "Prediction of disulfide bonding pattern based on a support vector machine and multiple trajectory search," *Information Sciences*, vol. 199, pp. 167–178, 2012.

[LTLQ10] J. Li, H. Tu, Q. Liu, and Y. Qian, "Research on state recognition and wear lifetime predication of tribo-system," in *Prognostics and Health Management Conference, 2010. PHM '10.*, Jan 2010, pp. 1–3.

[MB04] A. Meyer-Bäse, *Pattern Recognition for Medical Imaging*, ser. Electronics & Electrical. Elsevier Academic Press, 2004. [Online]. Available: http://books.google.de/books?id=DlIeOwNGxpUC

[MD84] I. Mayes and W. Davies, "Analysis of the response of a multi-rotor-bearing system containing a transverse crack in a rotor," *Journal of Vibration, Acoustics Stress and Reliability in Design*, vol. 106, no. 1, pp. 139–145, 1984. [Online]. Available: http://dx.doi.org/10.1115/1.3269142

[Met00] Metso Menerals, "Lindemann RAS scrap baling presses," 2000. [Online]. Available: www.metsominerals.com/metalwasterecycling

[MIO11] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[MLS07] S. Makino, T. Lee, and H. Sawada, *Blind Speech Separation*, ser. Signals and Communication Technology. Springer, 2007. [Online]. Available: http://books.google.de/books?id=LzReh0N4bTUC

[MMJP09] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd ed. Springer Publishing Company, Incorporated, 2009.

[MP43] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 5, no. 4, pp. 115–133, Dec. 1943. [Online]. Available: http://dx.doi.org/10.1007/bf02478259

[MR06] D. Mba and R. Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: Bearings, pumps, gearboxes, engines, and rotating structures," *The Shock and Vibration Digest*, vol. 38, no. 1, pp. 3–16, 2006.

[MTPP12] F. P. G. Marquez, A. M. Tobias, J. M. P. Perez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, no. 0, pp. 169–178, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148112001899

[NDSL13] R. Navarathna, D. Dean, S. Sridharan, and P. Lucey, "Multiple cameras for audio-visual speech recognition in an automotive environment," *Computer Speech and Language*, vol. 27, no. 4, pp. 911 – 927, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230812000538

[Nie09] T. Nieß, "Automatic metal and stone detection on bucket-wheel excavators used in RWE power AGs opencast lignite mines," *World of Mining*, vol. 61, pp. 226–233, 2009.

[NJR14] M. Namdari and H. Jazayeri-Rad, "Incipient fault diagnosis using support vector machines based on monitoring continuous decision functions," *Engineering Applications of Artificial Intelligence*, vol. 28, pp. 22–35, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197613002303

[NL11] K. Ngan and H. Li, *Video Segmentation and Its Applications*, ser. SpringerLink : Bücher. Springer, 2011. [Online]. Available: http://books.google.de/books?id=1PyYDv-oupsC

[NNRMR09] C. Nagaraju, K. Narayana Rao, and K. Mallikarijuna Rao, "Application of 3d wavelet transforms for crack detection in rotor systems," *Sadhana*, vol. 34, no. 3, pp. 407–419, 2009. [Online]. Available: http://dx.doi.org/10.1007/s12046-009-0024-y

[OPYM+12] G. Orru, W. Pettersson-Yeo, A. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1140–1152, 2012.

[OSAC14] C. S. Ooi, K. P. Seng, L.-M. Ang, and L. W. Chew, "A new approach of audio emotion recognition," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5858–5869, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414001638

[PD87] C. Papadopoulos and A. Dimarogonas, "Coupled longitudinal and bending vibrations of a rotating shaft with an open crack," *Journal of Sound and Vibration*, vol. 117, no. 1, pp. 81–93, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022460X87904378

[PDD02] W. Perrizo, Q. Ding, and A. Denton, "Lazy classifiers using p-trees," in *Proc. Int. Conf. on Computer Applications in Industry and Engineering*, San Diego, California, USA, 2002.

[Pel14] M. Pelillo, "Alhazen and the nearest neighbor rule," *Pattern Recognition Letters*, vol. 38, no. 0, pp. 34–37, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865513004145

[PF03] J. Penny and M. I. Friswell, "Simplified modelling of rotor cracks," *Key Engineering Materials*, vol. 245-246, no. 2003, pp. 223–232, 2003. [Online]. Available: http://www.scientific.net/KEM.245-246.223

[PK98] Z. Peng and T. Kirk, "Automatic wear-particle classification using neural networks," *Tribology Letters*, vol. 5, no. 4, pp. 249–257, 1998.

[PK05] F. Petrich and U. Köhler, "Stone detection and stone excavation - results and current directions of work with regard to technical stone handling in lusatian opencast mines," *World of Mining*, vol. 57, pp. 412–420, 2005.

[PSA13] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: A systematic review," *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129–135, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0208521613000168

[RA11] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics: A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327010002530

[RHH+13] J. Rodriguez, Y. Hamzaoui, J. Hernandez, J. Garcia, J. Flores, and A. Tejeda, "The use of artificial neural network (ann) for modeling the useful life of the failure assessment in blades of steam turbines," *Engineering Failure Analysis*, vol. 35, no. 0, pp. 562–575, 2013, special issue on {ICEFA} V- Part 1. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350630713001787

[RLBSA+13] I. Rodriguez-Lujan, G. Bailador, C. Sanchez-Avila, A. Herrero, and G. V. de Miguel, "Analysis of pattern recognition and dimensionality reduction techniques for odor biometrics," *Knowledge-Based Systems*, vol. 52, no. 0, pp. 279 – 289, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705113002323

[RNWV10] T. Rauber, E. Nascimento, E. Wandekoken, and F. Varejao, "Pattern recognition based fault diagnosis in industrial processes: Review and application," in *Pattern Recognition Recent Advances*, A. Herout, Ed. InTech, 2010, pp. 483–508.

[ROG99] M. Rychetsky, S. Ortmann, and M. Glesner, "Support vector approaches for engine knock detection," in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol. 2, Washington, USA, July 1999, pp. 969–974.

[Ros58] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[Ryt93] A. Rytter, *Vibrational Based Inspection of Civil Engineering Structures*, ser. Fracture and Dynamics. Dept. of Building Technology and Structural Engineering, Aalborg University, 1993, Ph.D.-Thesis defended publicly at the University of Aalborg, April 20, 1993 PDF for print: 206 pp.

[SA08] C. Stoisser and S. Audebert, "A comprehensive theoretical, numerical and experimental approach for crack detection in power plant rotating machinery," *Mechanical Systems and Signal Processing*, vol. 22, no. 4, pp. 818–844, 2008, special Issue: Crack Effects in Rotordynamics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007002294

[SAHO14] H. M. Sim, H. Asmuni, R. Hassan, and R. M. Othman, "Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5390–5404, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414001316

[Sal10] A. Salah, "Machine learning for biometrics," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, E. Olivas, J. Guerrero, M. Martinez-Sober, J. Magdalena-Benedito, and A. López, Eds. IGI Global, 2010, pp. 539–560.

[SAS12] D. Söffker and L. Al-Shrouf, "Merkmalsbasierte Multisensorik zur Objekt und Steinerkennung," Duisburg, 2012, Tech. rep., Chair of Dynamics and Control, University of Duisburg-Essen.

[SBM93] D. Söffker, J. Bajkowski, and P. C. Müller, "Detection of cracks in turborotors: A new observer based method," *Journal of Dynamic Systems, Measurement, and Control*, vol. 115, no. 3, pp. 518–524, 1993. [Online]. Available: http://dx.doi.org/10.1115/1.2899130

[Sek03] A. S. Sekhar, "Identification of a crack in a rotor system using a model-based wavelet approach," *Structural Health Monitoring*, vol. 2, no. 4, pp. 293–308, 2003. [Online]. Available: http://shm.sagepub.com/content/2/4/293.abstract

[SEMJ+11] M. Seyagh, M. El-Mostapha, A. Jarid, D. Cherqaoui, A. Schmitzer, and D. Villemin, "Pattern recognition: Application of support vector

machines, artificial neural networks and decision trees for anti-HIV activity prediction of organic compounds," in *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*, Marrakech, Morocco, April 2011, pp. 1–6.

[SGBD00] C. Saunders, A. Gammerman, H. Brown, and G. Donald, "Application of support vector machines to fault diagnosis and automated repair," in *11th Int. Workshop on Principles of Diagnostics (DX '00)*, Beijing, China, August 30 - September 1 2000, pp. 8–10.

[SHJ09] G. Schaefer, A. Hassanien, and J. Jiang, *Computational intelligence in medical imaging: techniques and applications.* Boca Raton, FL: CRC Press, 2009.

[Sin09] J. Sinou, "A review of damage detection and health monitoring of mechanical systems from changes in the measurement of linear and non-linear vibrations," in *Mechanical Vibrations: Measurement, Effects and Control*, R. C. Sapri, Ed. Nova Science Publishers, Inc., 2009, pp. 643–702.

[SKKQ04] G. Sabnavis, R. G. Kirk, M. Kasarda, and D. Quinn, "Cracked shaft detection and diagnostics: A literature review," *The Shock and Vibration Digest*, vol. 36, pp. 287–296, 2004.

[SL12] Y. Shao and R. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 70, no. 0, pp. 78–87, 2012.

[SLS11] M.-S. Saadawia, G. Li, and D. Söffker, "Wavelet-based svm system for evaluation of wear states and remaining life time," in *8th Int. Workshop on Structural Health Monitoring 2011*, F. Chang, Ed., Stanford, USA, September 2011.

[SMS07] J. Saß, F. Matare, and T. Schubert, *Metallsuchgerät und Steinerfassungssystem, Bagger 258 und 261 Tagebau Garzweiler, Bagger 259, 260 und 290 Tagebau Hambach, Bagger 255 Tagebau Inden*, Siemens AG, April 2007.

[SPH12] J. Sanz, R. Perera, and C. Huerta, "Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks," *Applied Soft Computing*, vol. 12, no. 9, pp. 2867–2878, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494612001688

[SS09] M.-S. Saadawia and D. Söffker, "SVM-based fault diagnosis system for materials change detection," in *Proc. IAR/ACD Workshop on Advanced Control and Diagnosis 2009*, Zielona Gora, Poland, November 2009.

[SS10a] ——, "Anwendung der Support Vector Machine für die Bewertung des Verschleisszustandes und der Restlebensdauer," in *Berichtsband 8th Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2010, pp. 271–281.

[SS10b] ——, "Application of Support Vector Machine for evaluation of wear state and remaining life time," in *Proc. EUROSIM Congress on Modelling and Simulation 2010*, Prague, 2010.

[SS11] ——, "Wavelet-based SVM system for evaluation of wear states and remaining life time," in *Chang, F.K. (Ed.): Structural Health Monitoring 2011*, Stanford, USA, September 2011, pp. 1349–1356.

[SS12] ——, "Feature-based resampling for classification using discrete wavelet transform for diagnostic purposes of industrial processes with periodic data," in *6th European Workshop on Structural Health Monitoring 2012*, C. Boller, Ed., Dresden, Germany, July 2012.

[SSL09] J. Sawicki, A. Sen, and G. Litak, "Multiresolution wavelet analysis of the dynamics of a cracked rotor," *International Journal of Rotating Machinery*, vol. 2009, pp. 1–8, 2009, article ID 265198.

[SSW12] D. Söffker, M.-S. Saadawia, and C. Wei, "Model- and feature-based diagnostics in rotating machinery," in *Berichtsband 9. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2012.

[SSW13] ——, "Model- and feature-based diagnostics in rotating machinery," in *DINAME 2013 - XV International Symposium on Dynamic Problems of Mechanics, Brazil*, 2013.

[Sta05] G. Stachowiak, *Wear: Materials, Mechanisms and Practice*, ser. Tribology in Practice Series. Wiley, 2005. [Online]. Available: http://books.google.de/books?id=r8tRAAAAMAAJ

[Sta09] R. Stapelberg, *Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design.* Springer, 2009. [Online]. Available: http://books.google.de/books?id=b4UudC1EIIoC

[Sto10] J. Stojek, "Application of time-frequency analysis for diagnostics of valve plate wear in axial-piston pump," *Archive of Mechanical*

Engineering, vol. 57, no. 3, pp. 309–322, 2010. [Online]. Available: http://versita.metapress.com/content/KJ488046382N7048

[Swe98] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, Mar. 1998. [Online]. Available: http://dx.doi.org/10.1137/S0036141095289051

[SWWS16] D. Söffker, C. Wei, S. Wolff, and M.-S. Saadawia, "Detection of rotor cracks: Comparison of an old model-based approach with a new signal-based approach," *Nonlinear Dynamics*, vol. 83, no. 3, pp. 1153–1170, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11071-015-2394-5

[SY06] L.-K. Shark and C. Yu, "Design of matched wavelets based on generalized mexican-hat function," *Signal Processing*, vol. 86, no. 7, pp. 1451 – 1469, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168405002604

[SYM95] D. Söffker, T.-J. Yu, and P. C. Müller, "State estimation of dynamical systems with nonlinearities by using proportional-integral observer," *International Journal of Systems Science*, vol. 26, no. 9, pp. 1571–1582, 1995. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00207729508929120

[SZRT13] M. Silani, S. Ziaei-Rad, and H. Talebi, "Vibration analysis of rotating systems with open and breathing cracks," *Applied Mathematical Modelling*, vol. 37, no. 24, pp. 9907–9921, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0307904X13003612

[Tan13] Y. Tang, "Deep learning using support vector machines," *Computing Research Repository (CoRR)*, vol. abs/1306.0239, 2013. [Online]. Available: http://arxiv.org/abs/1306.0239

[TBR09] M. Tahir, A. Bouridane, and M. Roula, "Prostate cancer classification using multispectral imagery and metaheuristics," in *Computational intelligence in medical imaging: techniques and applications*, G. Schäfer, A. Hassanien, and J. Jiang, Eds. Taylor & Francis, 2009, pp. 139–166.

[TK08] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Burlington, USA: Academic Press, 2008.

[TLLW13] X. Tao, C. Lu, C. Lu, and Z. Wang, "An approach to performance assessment and fault diagnosis for rotating machinery equipment," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1,

2013. [Online]. Available: http://asp.eurasipjournals.com/content/2013/1/5

[URB97] G. Uytterhoeven, D. Roose, and A. Bultheel, "Wavelet transforms using the lifting scheme," 1997, technical report, Department of Computer Science, Catholic University of Leuven.

[VG12] P. Varney and I. Green, "Crack detection in a rotor dynamic system by vibration monitoring - part ii: Extended analysis and experimental results," *Journal of Engineering for Gas Turbines and Power*, vol. 134, no. 11, 2012.

[VLR+06] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley, 2006. [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-047172999X.html

[WCDB09] H. WANG, T.-Y. CHAI, J.-L. DING, and M. BROWN, "Data driven fault diagnosis and fault tolerant control: Some advances and possible new directions," *Acta Automatica Sinica*, vol. 35, no. 6, pp. 739–747, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1874102908600932

[WG10] L. Wang and R. Gao, *Condition Monitoring and Control for Intelligent Manufacturing*, ser. Springer Series in Advanced Manufacturing. Springer, 2010. [Online]. Available: http://books.google.de/books?id=r-WpcQAACAAJ

[WP08] H. Wang and C. P., "Recent research results on intelligent methods for condition dignosis of rotating machinery," in *4th World Congress on Maintenance*, November 2008.

[WPJ13] R. Walker, S. Perinpanayagam, and I. Jennions, "Rotordynamic faults: Recent advances in diagnosis and prognosis," *International Journal of Rotating Machinery*, vol. 2013, pp. 1–12, 2013, Article ID 856865. [Online]. Available: http://www.hindawi.com/journals/ijrm/2013/856865/

[Wu11] Q. Wu, "A self-adaptive embedded chaotic particle swarm optimization for parameters selection of wv-svm," *Expert Systems with Applications*, vol. 38, no. 1, pp. 184–192, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410005427

[WY07] A. Widodo and B. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, August 2007.

[XWYG12] J. Xiao, H. Wang, X. Yang, and Z. Gao, "Multiple faults diagnosis in motion system based on SVM," *Machine Learning and Cybernetics*, vol. 3, no. 1, pp. 77–82, 2012.

[XZCH08] J. Xiang, Y. Zhong, X. Chen, and Z. He, "Crack detection in a shaft by combination of wavelet-based elements and genetic algorithm," *International Journal of Solids and Structures*, vol. 45, no. 17, pp. 4782–4795, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020768308001777

[YG09] R. Yan and R. Gao, "Energy-based feature extraction for defect diagnosis in rotary machines," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 9, pp. 3130–3139, 2009.

[YMM02] H. Yang, J. Mathew, and L. Ma, "Intelligent diagnosis of rotating machinery faults: A review," in *Proc. 3rd Asia-Pacific Conference on System Integrity and Maintenance ACSIM*, Cairns, Australia, September 2002, pp. 385–392. [Online]. Available: http://eprints.qut.edu.au/17942/

[YMM03] ——, "Vibration feature extraction techniques for fault diagnosis of rotating machinery : a literature survey," in *Proc. Asia-Pacific Vibration Conference*, Gold Coast, Australia, 2003, pp. 801–807. [Online]. Available: http://eprints.qut.edu.au/17941/

[Yu11] J. Yu, "Bearing performance degradation assessment using locality preserving projections and gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327011000707

[YWW11] Z. Yu, H. Wong, and G. Wen, "A modified support vector machine and its application to image segmentation," *Image and Vision Computing*, vol. 29, no. 1, pp. 29–40, 2011.

[YWY+13] S. Yen, Y. Wu, J. Yang, Y. Lee, C. Lee, and J. Liu, "A support vector machine-based context-ranking model for question answering," *Information Sciences*, vol. 224, pp. 77–87, 2013.

[Zan12] E. Zanaty, "Support vector machines (SVMs) versus multilayer perception (MLP) in data classification," *Egyptian Informatics Journal*, vol. 13, no. 3, pp. 177–183, 2012.

[ZJZL14] M. Zhao, X. Jin, Z. Zhang, and B. Li, "Fault diagnosis of rolling element bearings via discriminative subspace learning: Visualization and classification," *Expert Systems with Applications*,

vol. 41, no. 7, pp. 3391–3401, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413009469

[ZS11] Z.-B. Zhu and Z.-H. Song, "A novel fault diagnosis system using pattern classification on kernel {FDA} subspace," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6895–6905, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410013977

[ZZZ13] X. Zhu, Y. Zhang, and Y. Zhu, "Bearing performance degradation assessment based on the rough support vector data description," *Mechanical Systems and Signal Processing*, vol. 34, pp. 203–217, 2013.

The thesis is based on the results and development steps presented in the following previous publications:

[SS12] M.-S. Saadawia and D. Söffker, "Feature-based resampling for classification using discrete wavelet transform for diagnostic purposes of industrial processes with periodic data," in *6th European Workshop on Structural Health Monitoring 2012*, C. Boller, Ed., Dresden, Germany, July 2012.

[SLS11] M.-S. Saadawia, G. Li, and D. Söffker, "Wavelet-based svm system for evaluation of wear states and remaining life time," in *8th Int. Workshop on Structural Health Monitoring 2011*, F. Chang, Ed., Stanford, USA, September 2011.

[SS10a] M.-S. Saadawia and D. Söffker, "Anwendung der Support Vector Machine für die Bewertung des Verschleisszustandes und der Restlebensdauer," in *Berichtsband 8th Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2010, pp. 271-281.

[SS10b] M.-S. Saadawia and D. Söffker, "Application of Support Vector Machine for evaluation of wear state and remaining life time," in *Proc. EUROSIM Congress on Modelling and Simulation*, Prague, 2010.

[SS09] M.-S. Saadawia and D. Söffker, "SVM-based fault diagnosis system for materials change detection," in *Proc. IAR/ACD Workshop on Advanced Control and Diagnosis*, Zielona Gora, Poland, November 2009.

[ASSS14] L. Al-Shrouf, M.-S. Saadawia, and D. Söffker, "Improved process monitoring and supervision based on a reliable multi-stage feature-based pattern recognition technique," *Information Sciences*, vol.259, no.0, pp. 282-294, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025513007202

[ASSSS12] L. Al-Shrouf, M.-S. Saadawia, N. Szczepanski, and D. Söffker, "Event detection using multisensor fusion and filtering techniques based on CWT and SVM," in *6th European Workshop on Structural Health Monitoring 2012*, C. Boller, Ed., Dresden, Germany, July 2012, pp. 206-213.

[ASSS11] L. Al-Shrouf, M.-S. Saadawia, N. Szczepanski, and D. Söffker, "Adaptive classification based on multisensoric decision fusion," in *8th Int. Workshop on Structural Health Monitoring 2011*, F. Chang, Ed., Stanford, USA, September 2011, pp. 127-134.

[ASSSS10] L. Al-Shrouf, M.-S. Saadawia, N. Szczepanski, and D. Söffker, "Multisensorfusionsbasierte Prozessüberwachung," in *Berichtsband 8. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2010, pp. 413-422.

[SWWS16] D. Söffker, C. Wei, S. Wolff, and M.-S. Saadawia, "Detection of rotor cracks: Comparison of an old model-based approach with a new signal-based approach," *Nonlinear Dynamics*, vol. 83, no. 3, pp. 1153-1170, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11071-015-2394-5.

[SSW13] D. Söffker, M.-S. Saadawia, and C. Wei, "Model- and feature-based diagnostics in rotating machinery," in *DINAME 2013 - XV International Symposium on Dynamic Problems of Mechanics*, Brazil, 2013.

[SSW12] D. Söffker, M.-S. Saadawia, and C. Wei, "Model- and feature-based diagnostics in rotating machinery," in *Berichtsband 9. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2012.

[SAJS08] D. Söffker, H. Al-Joumaa, and M.-S. Saadawia, "Defining features for diagnosis and prognosis - Part II: Data driven adaption of diagnosis filter," in *Proc. 9th International Conference on Motion and Vibration Control MOVIC*, Munich, 2008.

[SAJSD08] D. Söffker, H. Al-Joumaa, M.-S. Saadawia, and K.-U. Dettmann, "Beschreibung des Maschinenzustandes durch komplexe Merkmale zur Schadenfrüherkennung und zum Condition Monitoring," in *Berichtsband 7. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung AKIDA*, Aachen, Germany, 2008.

In the context of the research projects at the Chair of Dynamics and Control the following student theses have been supervised by Mahmud-Sami Saadawia and Univ.-Prof. Dr.-Ing. Dirk Söffker. Development steps and results of the research projects and the student theses are integrated to each other and hence are also part of this thesis.

[Qiu09] Yingying Qiu, "Design of GUI for transformation and visualization of real industrial sensory data", Bachelor thesis, University of Duisburg-Essen, Duisburg, Germany, November 2009.

[Bat10] Kürsat Batmaz, "Signalanalyse und visuelle Fehlerklassifikation mit Matlab", Project thesis, University of Duisburg-Essen, Duisburg, Germany, September 2010.

[Li11] Gang Li, "Fault detection and classification using Wavelets and Support Vector Machine", Master thesis, University of Duisburg-Essen, Duisburg, Germany, February 2011.

[Laf12] Asmaa Lafta, "Object Detection Based on Wavelet Transform and Artificial Neural Networks", Master thesis, University of Duisburg-Essen, Duisburg, Germany, April 2012.

[Wol13] Stefan Wolff, "Diagnose komplexer Maschinenzustände auf Basis realer Daten mechatronischer Systeme", Master thesis, University of Duisburg-Essen, Duisburg, Germany, September 2013.