Helsinki University of Technology Control Engineering Laboratory

Espoo 2004 Report 141

SUPPORT VECTOR MACHINE BASED CLASSIFICATION IN CONDITION MONITORING OF INDUCTION MOTORS

Sanna Pöyhönen

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D´HELSINKI

Helsinki University of Technology Control Engineering Laboratory

Espoo June 2004 Report 141

SUPPORT VECTOR MACHINE BASED CLASSIFICATION IN CONDITION MONITORING OF INDUCTION MOTORS

Sanna Pöyhönen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Automation and Systems Technology, for public examination and debate in Auditorium TU1 at Helsinki University of Technology (Espoo, Finland) on the 18th of June, 2004, at 12 noon.

Helsinki University of Technology

Department of Automation and Systems Technology

Control Engineering Laboratory

Distribution:
Helsinki University of Technology
Control Engineering Laboratory
P.O. Box 5500
FIN-02015 HUT, Finland
Tel. +358-9-451 5201
Fax. +358-9-451 5208
E-mail: control.engineering@hut.fi
http://www.control.hut.fi/

ISBN 951-22-7154-0 (printed)
ISBN 951-22-7155-9 (pdf)
ISSN 0356-0872

Picaset Oy
Helsinki 2004

Available on net at http://lib.hut.fi/Diss/2004/isbn9512271559

HELSINKI UNIVERSITY OF TECHNOLOGY P.O. BOX 1000, FIN-02015 HUT http://www.hut.fi	ABSTRACT OF DOCTORAL DISSERTATION

Author Sanna Pöyhönen

Name of the dissertation

Support Vector Machine Based Classification in Condition Monitoring of Induction Motors

Date of manuscript June 7, 2004	Date of the dissertation June 18, 2004
☐ Monograph	☑ Article dissertation (summary + original articles)

Department	Automation and Systems Technology
Laboratory	Control Engineering Laboratory
Field of research	Control Engineering
Opponent(s)	Professor Hannu Koivisto, Professor Jarmo Partanen
Supervisor	Professor Heikki Koivo, AS-74
(Instructor)	Professor Heikki Hyötyniemi, AS-74

Abstract

Continuous and trouble-free operation of induction motors is an essential part of modern power and production plants. Faults and failures of electrical machinery may cause remarkable economical losses but also highly dangerous situations. In addition to analytical and knowledge-based models, application of data-based models has established a firm position in the induction motor fault diagnostics during the last decade. For example, pattern recognition with Neural Networks (NN) is widely studied.

Support Vector Machine (SVM) is a novel machine learning method introduced in early 90's. It is based on the statistical learning theory presented by V.N. Vapnik, and it has been successfully applied to numerous classification and pattern recognition problems such as text categorization, image recognition and bioinformatics. SVM based classifier is built to minimize the structural misclassification risk, whereas conventional classification techniques often apply minimization of the empirical risk. Therefore, SVM is claimed to lead enhanced generalisation properties. Further, application of SVM results in the global solution for a classification problem. Thirdly, SVM based classification is attractive, because its efficiency does not directly depend on the dimension of classified entities. This property is very useful in fault diagnostics, because the number of fault classification features does not have to be drastically limited. However, SVM has not yet been widely studied in the area of fault diagnostics. Specifically, in the condition monitoring of induction motor, it does not seem to have been considered before this research.

In this thesis, a SVM based classification scheme is designed for different tasks in induction motor fault diagnostics and for partial discharge analysis of insulation condition monitoring. Several variables are compared as fault indicators, and forces on rotor are found to be important in fault detection instead of motor current that is currently widely studied. The measurement of forces is difficult, but easily measurable vibrations are directly related to the forces. Hence, vibration monitoring is considered in more detail as the medium for the motor fault diagnostics.

SVM classifiers are essentially 2-class classifiers. In addition to the induction motor fault diagnostics, the results of this thesis cover various methods for coupling SVMs for carrying out a multi-class classification problem.

Keywords	fault diagnostics, induction motor, support vector machine, classification		
UDC		Number of pages	63 + 45
ISBN (printed)	951-22-7154-0	ISBN (pdf)	951-22-7155-9
ISBN (others)		ISSN	0356-0872
Publisher	Helsinki University of Technology, Control Engineering Laboratory		
Print distribution	Helsinki University of Technology, Control Engineering Laboratory		
☑	The dissertation can be read at http://lib.hut.fi/Diss/		

Preface

This thesis was written in Control Engineering Laboratory, Helsinki University of Technology. I want to thank the head of laboratory, Professor Heikki Koivo, for supervising the thesis, and providing a relaxed and pleasant working environment in the laboratory. I am also deeply grateful to Professor Heikki Hyötyniemi for inspirational guidance throughout the work.

The work has been carried out in VAIVI research project funded by Tekes (the National Technology Agency) and industrial partners KCI, ABB, Kone and Kupari Mittaus. I would like to thank all the financiers, and I am also thankful to my co-workers in the project: Professor Antero Arkkio, Marian Negrea, Pedro Jover, and Jarmo Lehtonen. I highly appreciate their work. I also owe thanks to Marco Conti from the University of Bologna for his excellent guidance on the partial discharge monitoring.

In addition, I want to thank the reviewers of this thesis, Dr. Jari Hämäläinen and Professor Hannu Koivisto, whose valuable comments essentially improved the thesis.

The additional financial support for the work has been given by Finnish Academy Graduate School on Electronics, Telecommunications and Automation (GETA), the Finnish Society of Automation, the Finnish Cultural Foundation, the Technology Promotion Foundation, Jenny and Antti Wihuri Foundation, the Neles Foundation, Centre for International Mobility (CIMO) and Ulla Tuominen Foundation, which are gratefully acknowledged.

Further, I want to thank the whole personnel of Control Engineering Laboratory for being humorous and delightfully eccentric fellow workers. Especially, I want to thank Vesa for the bad jokes, Terhi for the attitude, Mikael for the mastery of ceremonies, Pekka for non-laundry-days, Petteri for personal training in addition to the world's best DVD rental, and Konsta for the most constructive comments on books, sports and particularly on this thesis.

Finally, I want to express my gratitude to the dearest persons in my life. My mother I want to thank for teaching me to appreciate knowledge and education. To Mari I am grateful for guidance in my studies and being the omniscient big sister. Kaisa I thank for being a true friend and highly enjoyable company here, there and everywhere. Most of all, I thank Ville for bedtime-stories, intense discussions on everything except induction motors, patient attempts to make me not to worry too much, and for being there with me also on rainy days.

Espoo, June 2004

Sanna Pöyhönen

List of Publications

P1 Pöyhönen, S., Negrea, M., Arkkio, A., Hyötyniemi, H., Koivo, H.: ”Support Vector Classification for Fault Diagnostics of an Electrical Machine”, *Proc. of the 6th Int. Conf. on Signal Processing*, ICSP’02, Vol.2, pp. 1719-1722, Beijing-China, August, 2002.

P2 Pöyhönen, S., Negrea, M., Arkkio, A., Hyötyniemi, H., Koivo, H.: ”Fault Diagnostics of an Electrical Machine with Multiple Support Vector Classifiers”, *Proc. of the 17th IEEE Int. Symp. on Intelligent Control*, ISIC'02, Vol. 1, pp. 373-378, Vancouver, British Columbia, Canada, October, 2002.

P3 Pöyhönen, S., Arkkio, A., Hyötyniemi, H.: ”Coupling Pairwise Support Vector Machines for Fault Classification”, *Proc. of the 5th IFAC Symp. on Fault Detection, Supervision and Safety of Technical Processes*, Safeprocess2003, pp. 705-710, Washington, D.C., USA, June, 2003.

P4 Pöyhönen, S., Negrea, M., Jover, P., Arkkio, A., Hyötyniemi, H.: ”Numerical Magnetic Field Analysis and Signal Processing for Fault Diagnostics of Electrical Machines”, *The Int. Journal for Computation and Mathematics in Electrical and Electronic Engineering*, COMPEL, Vol. 2, No.4, pp. 969-981, 2003.

P5 Pöyhönen, S., Jover, P., Hyötyniemi, H.: “Signal Processing of Vibrations for Condition Monitoring of an Induction Motor*”, Proc. of the 1st IEEE-EURASIP Int. Symp. on Control, Communications, and Signal Processing*, ISCCSP 2004, pp. 499-502, Hammamet, Tunisia, March 2004.

P6 Pöyhönen, S., Jover, P., Hyötyniemi, H.: ”Independent Component Analysis of Vibrations for Fault Diagnosis of an Induction Motor”, *Proc. of the IASTED Int. Conf. on Circuits, Signals, and Systems*, CSS 2003, Vol. 1, pp. 203-208, May 19-21, Cancun, Mexico, 2003.

P7 Pöyhönen S., Conti. M., Cavallini, A., Montanari, G.C., Filippetti, F.: “Insulation Defect Localization through Partial Discharge Measurements and Numerical Classification”, *Proc. of the IEEE Int. Symp. on Industrial Electronics*, ISIE 2004, pp. 417 – 422, Ajaccio, France, May 2004.

The author is the main writer and generated the most important results in all papers. The chapters considering finite element method (FEM) were written by M. Negrea and A. Arkkio, who also provided the FEM models applied in the papers [P1]-[P4]. P. Jover carried out the current and vibration measurements considered in the publications [P4]-[P6]. In [P7], M. Conti wrote the chapters considering partial discharge (PD) diagnostics, reference database and PD parameters description in general. The PD measurements were carried out in University of Bologna. Other authors gave valuable insights to the subject.

Contents

Nomenclature

Mathematical notations

ω_s	the synchronous speed of the revolving field
ω_r	the slip speed
ω_m	the rotor speed
P	the number of poles
f	the frequency of the current in the stator winding
s	the slip of the motor
V_1	the applied voltage on a per-phase basis
R_1	a per-phase stator winding resistance
X_1	a per-phase stator winding reactance
R_r	a per-phase rotor winding resistance
X_b	a per-phase rotor winding leakage reactance at $s = 1$
X_r	a per-phase rotor winding leakage reactance at any slip s
X_m	a per-phase magnetization reactance
R_c	a per-phase equivalent core-loss resistance
E_1	the per-phase induced EMF in the stator winding
E_b	the per-phase induced EMF in the rotor winding at $s = 1$
E_r	the per-phase induced EMF in the rotor winding at slip s
I_r	the per-phase rotor winding current
I_1	the per-phase current supplied by the source
I_Φ	the per-phase excitation current
I_c	the per-phase core loss current
I_m	the per-phase magnetization current
a	a transformation ratio
k_{w1}	the winding factor of the stator winding
k_{w2}	the winding factor of the rotor winding
N_1	turns per phase of the stator winding
N_2	turns per phase of the rotor winding.
v	the reluctivity of the material
$\mathbf{J}$	the current density
Φ	the electric scalar potential:
$\mathbf{A}$	the magnetic vector potential
σ	the conductivity of the material
u	the potential difference induced between the ends of the conductor
l	the length of the conductor
i	the total induced current
R	the d.c. resistance of the conductor
Δt	a short time interval
n	the dimension of the classified vector, i.e. the number of features
$\mathbf{x} = (x_1, ..., x_n)^T$	a sample vector to be classified
K	the number of classes
M	the number of the training samples
$p_i = P(C_i)$	a probability that $\mathbf{x}$ belongs to class i
p_{ij}	a conditional probability that $\mathbf{x}$ belongs to class i when $\mathbf{x}$ belongs either to i or j

$P(\mathbf{x}\|C_i)$	a conditional probability for getting an observation $\mathbf{x}$ from class i $P(C_i\|\mathbf{x})$ an a posteriori probability for sample to belong to class i.
$P(\mathbf{x})$	the unconditional probability for occurrence of $\mathbf{x}$
y_i	a label for a sample vector $\mathbf{x}_i$, $y = +1$ or $y = -1$
$\mathbf{w} = (w_1,\ldots,w_n)^{\mathrm{T}}$	a weighting vector
$f(\mathbf{x})$	the function defining a separating hyperplane
$h(\mathbf{x})$	$h(\mathbf{x})$=sign($f(\mathbf{x})$), the decision rule
b	a bias of the separating hyperplane
$\lambda \in \mathbb{R}^+$	a learning rate of a linear learning machine
d	the VC dimension
R	the risk for misclassification in the test set
R_{emp}	the empirical risk for misclassification in the test set
P_{error}	the probability of test error
η	the bound for risk of misclassification holds with probability $1 - \eta$, $0 \le \eta \le 1$
Δ	the margin of Δ-margin separating hyperplanes
$\mathbf{x}^+$	a sample belonging to positive class
$\mathbf{x}^-$	a sample belonging to negative class
$\mathbf{x}^{\mathrm{o}}$	a sample on a separating hyperplane
γ	the geometrical margin between positive and negative class
L	Lagrange function
W	a cost function in the dual representation of the Lagrange optimisation problem
$\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_M)^{\mathrm{T}}$	Lagrange multipliers for all samples
$\#SV$	the number of support vectors
$\mathbf{x}^s = (\mathbf{x}^{s1}, \mathbf{x}^{s2},\ldots\mathbf{x}^{s\#SV})^{\mathrm{T}}$	a group of support vectors
ξ_i	a slack variable introduced to the ith inequality condition, $\xi_i > 0$, $i = 1,\ldots,M$
C	the upper bound for elements of $\boldsymbol{\alpha}$, error penalty for SVM training
$\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}),\ldots,\varphi_l(\mathbf{x}))^{\mathrm{T}}$	a nonlinear vector function that maps $\mathbf{x}$ into the l-dimensional feature space
$K(\mathbf{x},\mathbf{z}) = \boldsymbol{\varphi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{z})$	a Kernel function that calculates inner products of feature space in original data space
$\mathbf{G}$	Gram matrix
$\mathbf{g}_n$	a vector containing outputs of n classifiers
$\mathbf{M}$	$K \times n$ mixture matrix
$\mathbf{g}_{\mathrm{mixed}} = \mathbf{M}\mathbf{g}_n$	$K \times 1$ vector containing fusion of the outputs of the classifiers
Q	an accuracy measure
σ	the width of the radial basis function
M	a polyspectral measure with magnitude $\|M\|$ and phase $\angle M$, the phase being bound by $-\pi$ and π
M_{enh}	an enhanced polyspectral measure
$a_1,\ldots,a_n$	coefficients of the nth order autoregressive model

Abbreviations

ANFIS	Adaptive Neuro-Fuzzy Inference System (or Adaptive Network based Fuzzy Inference System)
AR	AutoRegressive
BB	Broken rotor Bar
BR	Broken end Ring
BSS	Blind Signal Separation
DE	Dynamic rotor Eccentricity
EMF	ElectroMotive Force
FALCON	Fuzzy Adaptive Learning Control Network
FEM	Finite Element Method
FFT	Fast Fourier Transform
HOS	Higher Order Spectrum
HV	High Voltage
IC	Independent Component
ICA	Independent Component Analysis
IM	Induction Motor
k-NN	*k*-Nearest Neighbor
LV	Low Voltage
MANOVA	Multivariate ANalysis Of VAriance
MCSA	Motor Current Signature Analysis
MDA	Multiple Discriminant Analysis
MLP	MultiLayer Perceptron
MV	Medium Voltage
MVDA	MultiVariate Data Analysis
NF	No Fault
NN	Neural Network
PD	Partial Discharge
PC	Principal Component
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
PRPD	Phase Resolved Partial Discharge (diagram)
PSD	Power Spectrum Density
RC	shorted Coil in Rotor winding
RMS	Root Mean Square
RT	shorted Turn in Rotor winding
SC	shorted Coil in Stator winding
SE	Static rotor Eccentricity
ST	shorted Turn in Stator winding
STFT	Short Time Fourier Transform
SV	Support Vector
SVC	Support Vector Classification
SVM	Support Vector Machine
VA	Voltage Asymmetry
VC	Vapnik-Chervonenkis

1. Introduction

1.1 Background and motivation

Induction motors are widely used in the industry, and there is a strong demand for their reliable and safe operation. Faults and failures of critical electro-mechanical parts can lead to excessive downtimes and generate costs of millions of euros in reduced output, emergency maintenance and lost revenues. Thus, finding efficient and reliable fault diagnostics methods, especially for induction motors, is extremely important. In the industry, methods based on analytical models of the motor systems are still the most common choices for condition monitoring of electrical machinery. However, during the last decade also applications of different kinds of data-based models such as Neural Networks (NN) have established a firm position.

Support Vector Machine (SVM) is a modern computational learning method based on statistical learning theory presented by V. N. Vapnik [Vapnik00]. In SVM, the original input space is mapped into a high dimensional dot product space called feature space, and in the feature space the optimal hyperplane is determined to maximize the generalisation ability of the classifier. The optimal hyperplane is found by exploiting optimisation theory, and respecting insights provided by the statistical learning theory.

SVMs have several benefits compared to traditional classifiers. Their most important characteristic is that they can handle very large feature spaces. Their generalisation ability and computational efficiency are both independent of the dimension of the input space. That is why, it has been found to be especially efficient in high dimensional classification problems. Concerning fault classification this is a great benefit, because the number of fault features does not have to be limited. Aggressive feature selection could result in a loss of information. The other benefit compared to conventional algorithms is that the SVM results in a globally optimal solution for the problem under study. As a comparison one could consider the NNs, which may have many local minima leading to a not trustworthy solution. Further benefit is sparseness of the solution that enables efficient and fast computation.

In this thesis, the induction motor fault diagnostics is investigated using SVMs in co-operation with various signal processing tools. Different aspects of motor fault diagnostics are considered, concentrating especially on motor current signature analysis (MCSA), vibration monitoring and insulation system condition monitoring with partial discharge (PD) analysis. Vibration monitoring is a classic tool for induction motor fault diagnostics, but recently a great amount of research is oriented towards application of MCSA, because monitoring the motor current is a non-invasive method and it does not require any additional instrumentation such as vibration sensors. In some studies, it is claimed that motor current contains equal information on motor faults as the motor vibrations, which seems to justify the efforts put on the research of the subject [Benbouzid00]. In this thesis, several variables will be compared as fault indicators including the motor current and forces related to motor vibrations.

SVMs have been successfully applied in various classification and pattern recognition tasks, but in the area of fault diagnostics they have not been widely studied. Especially, in fault diagnostics of induction motors they do not seem to have been documented before this research. As stated above they are highly competitive with, e.g., NNs which are widely studied also in induction motor condition monitoring.

1.2 Summary of articles

The thesis consists of seven articles that consider different aspects of the fault diagnostics of electrical machines. MCSA, vibration monitoring and PD analysis are studied, and data-based analysis receives special attention. SVM based classification is applied in fault classification tasks and various feature extraction methods are considered in each motor condition monitoring approach. The methods are tested with both virtual measurement data generated with numerical magnetic field analysis and real measurements. Numerical magnetic field analysis has been carried out in the Laboratory of Electromechanis, HUT, for this study.

In [P1] and [P2], the MCSA approach is followed. In both articles, the estimates of power spectrum density (PSD) of the stator line current of a 15 kW induction motor are used as a medium of fault detection. In [P1], novel SVM based classifiers are trained to distinguish healthy spectrum from faulty spectra and faulty spectra from each other. In [P2], the outputs of pairwise SVMs are fused with a simple majority voting approach. The application of SVM made possible the utilisation of the whole PSD estimate as the feature vector in classification despite of the small number of measurements, and feature selection is not needed before training the classifier. Six different faults are studied in addition to the healthy operation of the motor. Numerical magnetic field analysis is used to provide virtual measurement data from the motor operation. PSD estimates of the stator current of the motor are calculated with Welch's method [Welch67]. The results are promising. Most of the faults can be separated correctly from each other. One can also see from the support vector percentages that the classification is successful in most of the fault cases. In [P2], also the influence of noise was studied. SVM had not been applied in induction motor fault diagnostics before these studies.

Without noise the classification structure performs well, but noise degrades the total classification rate. With noise filtering, the fault detection rate increases only slightly. This raises the question, whether the stator line current is the best choice for a fault detection medium. It is widely used, because the current measurements do not require access to the motor, but nevertheless the current may not contain enough information on motor faults to be applied as an efficient medium of fault diagnostics. However, some faults can easily be detected from stator line current regardless of the noise. With 3-class classification structure, detection rate of shorted coil and shorted turn operation are adequate even in noisy situations.

It is noticed that also in noiseless case, the errors in outputs of 2-class classifiers are cumulated in reconstruction of the final multi-class solution. The malfunction of some 2-class classifiers should be taken into account while reconstructing the final classification solution. In [P3], four different coupling techniques of the 2-class classifiers are studied to get the global decision of the motor condition. A properly tuned NN coupling results in the most accurate multi-class classification result, but with a much simpler reconstruction approach relying on a mixture matrix almost equal classification performance is obtained. In this application, linear coupling is a practical choice for the reconstruction scheme, because training and tuning a NN is an exhausting task, and the benefits of applying a nonlinear approach are marginal. Comparison of different coupling techniques of 2-class classifiers is important concerning the research of classification methods in general, and the mixture matrix approach had not been earlier applied exactly in this form.

In [P4], SVM based classification is applied to fault diagnostics of a 35 kW cage induction motor and a slip-ring generator. A mixture matrix coupling scheme is used to combine 2-class SVMs. Stator line current, circulating currents between parallel branches and forces acting on the machine's rotors are compared as fault indicators. Forces on the rotor and circulating currents between parallel branches are found to be superior indicators of faults compared to the stator

current. Measurement of the forces is difficult, but they are directly related to measurable motor vibrations that are widely used in the condition monitoring of induction motor.

In [P5], vibration monitoring is studied instead of MCSA. Several features of vibration signals are compared as indicators of broken rotor bar of a 35 kW induction motor. Regular fast Fourier transform (FFT) based PSD estimation is compared to signal processing with higher order spectra, cepstrum analysis and signal description with autoregressive (AR) modelling. The fault detection routine and feature comparison is carried out with SVM based classification. The best method for feature extraction seems to be the application of coefficients of AR model. The results are discovered from real measurement data from several motor conditions and load situations. It is noteworthy that the independent test set contains measurements from rotor faulted motor operation that are not exactly the same as the data used in the training of the classifier. This further ensures the generalisation of the fault classification when AR-coefficients are used as classification features.

Also [P6] deals with the vibration monitoring. Vibrations of the motor are measured with multiple sensors, and it is studied, whether the fault classification rate improves, if the information of measurements is fused and possible interference removed before the classification. This is carried out with independent component analysis (ICA). Only the PSD estimation is used as a feature extraction tool, but with the data fusion also this results in excellent fault classification: all test samples are correctly classified. For comparison, regular principal component analysis (PCA) is applied for data fusion and it also improves the results. It is obvious that the vibration data fusion benefits the fault classification, but it cannot be concluded that a broken rotor bar generates an independent component to the vibrations that could be detected and utilised in fault diagnostics. ICA could be more successfully applied in situations, where independent interference should be separated from vibrations before fault diagnostics.

In [P7], another aspect of induction motor fault diagnostics is studied: the condition monitoring of insulation systems. Partial discharge (PD) measurements are widely used for diagnosis of insulation condition in different high or medium voltage apparatus. This study does not cover the whole PD analysis process, but only a PD generating defect localization task is considered. Firstly, various parameters are extracted from PD distributions and statistical analysis is carried out to find good parameters for insulation defect localization. Then numerical classification is applied to build an automated localization tool. Because the most important features concerning the localization are found beforehand, also other classification tools can be applied in addition to SVM. When comparing *k*-nearest neighbor classifier (*k*-NN), probabilistic neural network and SVM, the best results are achieved with SVM. However, the application of a simple tool such as *k*-NN classification may be justified when considering tuning, updating and computational aspects. These are considerable simpler tasks with *k*-NN classifier than with SVM. SVM had not been studied regarding PD analysis before this research.

1.3 Contributions of the author

The author's contributions can be summarized as follows:

- SVM based classification routine for fault diagnostics of induction motors is implemented in a simulation environment using stator line current, motor vibrations or forces on the rotor as fault indicators. As far as the author knows SVM has not been used in induction motor condition monitoring earlier.
- Various measurements are compared for fault diagnostics of an induction motor, and the stator line current is found to be a poor indicator of faults unlike, for example, forces on the rotor. The

result is interesting, because recently a great part of induction motor fault diagnostics research has concentrated on using the current as the main information source.

- Various coupling schemes of pairwise classifiers are implemented and compared to solve a multi-class classification problem. The best scheme, mixture matrix coupling, was not applied in the literature in this form before this thesis.
- Of various signal processing tools, AR modelling was showed to reveal the influence of a broken rotor bar in a 35 kW induction motor in the best way, when vibrations are used as fault indicators. AR modelling as well as other signal processing tools considered has been documented in the literature before, but their comparison is new.
- Information fusion of multi-channel vibration measurements with multivariate data analysis was shown to enhance the fault detection rate of a 35 kW induction motor. Similar results can be found in the literature for other kinds of rotating machines.
- Partial discharge distribution parameters were analysed to find the quantities, which can be used in automated localization of defects of insulation systems. Various numerical classification methods for automated localization of the defects were implemented in software and compared. SVM based classification routine was shown to give the best results. SVM has not been documented in the partial discharge analysis earlier.

1.4 Structure and organisation of this thesis

The thesis consists of a summary and seven articles mentioned before. In Chapter 2, the performance of the induction motor is clarified. In Chapter 3, fault diagnostics of the induction motor is reviewed. Existing fault diagnostics methods are first considered in general and then three common aspects of induction motor condition monitoring are presented: vibration monitoring, MCSA and PD analysis. In Chapter 4, the performance of SVM in classification is presented starting from the influence of Vapnik-Chervonenkis dimension on the classifier's generalisation performance. Then the Δ-margin separating hyperplane is explained, and the maximal margin classification is considered in addition to application of kernel functions in nonlinear classification tasks. In Chapter 4, also the design and tuning of SVM are considered, and comparison of different methods of using SVM in multi-class classification tasks is carried out. The comparison of the multi-class classification methods is based on the publication [P3]. Further, enhanced version of SVM, least squares SVM (LS-SVM) algorithm is described. Chapter 5 summarizes the results of the publications [P1] and [P2], where the motor current signal analysis is carried out for a 15 kW induction motor with SVM. Chapter 6 bases on the publication [P4], where using the motor current as a fault indicator is compared to using two other variables: forces on the rotor and the circulating currents in the parallel branches. Chapter 7 summarizes the results from [P5] and [P6], where vibration monitoring is considered in the broken rotor bar detection of a 35 kW induction motor. Several signal processing methods are compared in revealing the influence of broken rotor bar on the motor vibrations. Then information fusion of the multi-channel vibration measurements is considered for improving the fault detection rate. Chapter 8 summarizes results of the publication [P7], where the partial discharge measurements and numerical classification are applied for the localization of defects in insulation systems that are also important part of the structure of induction motors. In Chapter 9, conclusions are made.

2. Induction motor

In this chapter, the performance of induction motor is presented. In addition to the general structure and the principle of operation (Chapter 2.1), a per-phase equivalent circuit of a balanced three-phase induction motor is clarified in Chapter 2.2. The balanced equivalent circuit is widely used to model the motor performance, but it is valid only for symmetric induction motors fed from a sinusoidal power source. Faults generate asymmetries to the motor, and therefore, the equivalent circuit cannot be used to imitate the behaviour of faulted machine. For faulted motors and non-sinusoidal power sources, more enhanced models are required. An example of numerical models for the induction motor is the analysis of magnetic field with finite element method (FEM), which is explained in Chapter 2.3. In Chapter 2.4, faults of the induction motor are reviewed in addition to the most common measurements used for the fault diagnostics.

2.1 Structure and the principle of operation

Of all the electrical machines, induction motors are the most common in industry due to their simplicity, rugged structure, cheapness and easy maintainability. Induction motors receive their power by induction. There are two basic types of induction motors: single-phase induction motors and polyphase induction motors. Polyphase induction motors cover a variety of horsepower ratings and their use is preferred, if a polyphase power source is available [Guru01].

A three-phase induction motor is the most popular polyphase induction motor. It has two main parts: stationary stator and a revolving rotor, which is separated from stator by a small air-gap. The induction motors studied in this thesis have a stator, which consists of an aluminium frame that supports a hollow, cylindrical core made up of stacked laminations. A number of evenly spaced slots, punched out of the laminations, provide the space for the stator winding. The rotor is also composed of punched laminations. These are stacked to create a series of rotor slots to provide space for rotor winding. When considering the squirrel cage induction motors the rotor is composed of bars, which are pushed into the slots. The opposite ends are welded to two end rings, so that all bars are short circuit together. The entire construction resembles a squirrel cage, which the motor has got its name from. An exploded view of a cage induction motor is presented in Fig. 2.1 [Wildi97].

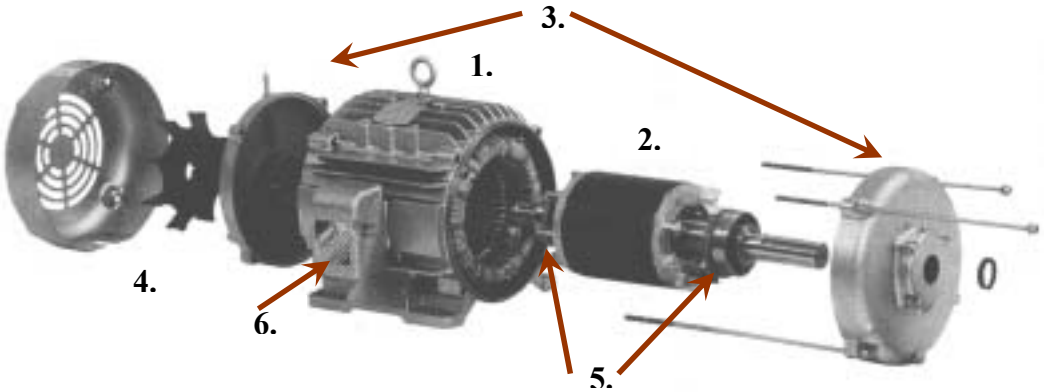


Figure 2.1. Construction of a squirrel cage motor showing the stator (1), rotor (2), end-caps (3), cooling fan (4), ball bearings (5) and terminal box (6) [Wildi97].

Another type of induction motors is the wound-rotor induction motor. It is, however, more expensive and less efficient than a squirrel-cage induction motor, and it is used only when a squirrel-cage induction motor cannot deliver the high enough starting torque [Guru01].

When the stator winding of a three-phase induction motor is connected to a three-phase power source, it produces a magnetic field that is a constant in magnitude and revolves around the rotor at the synchronous speed. If f is the frequency of the current in the stator winding and P is the number of poles, the synchronous speed of the revolving field is:

$$\omega_s = \frac{4\pi f}{P} \tag{2.1}$$

The revolving field induces electromotive force (EMF) in the rotor winding. Since the rotor winding forms a closed loop, the induced EMF in each coil gives rise to an induced current in that coil. When a current-carrying coil is in a magnetic field, it experiences a force that tends to rotate it. The rotor receives its power by induction only when there is a relative motion between the rotor speed and the revolving field. Since the rotor rotates at a speed lower than the synchronous speed of the revolving field, an induction motor is also called an asynchronous motor [Guru01].

The relative speed between stator and rotor is also called slip speed. If the rotor speed is ω_m, the slip speed is $\omega_r = \omega_s - \omega_m$. A common practice is to express the slip speed in terms of the slip, s, which is a ratio of the slip speed to the synchronous speed:

$$s = \frac{\omega_r}{\omega_s} \ . \tag{2.2}$$

2.2 An equivalent circuit

The performance of an induction motor is often described with a simple equivalent electrical circuit model. For example, a phase equivalent circuit of a balanced three-phase induction motor from [Guru01] is represented in Fig. 2.2. The induction motor is assumed to be symmetric, i.e. all phases are similar.

In the figure, V_1 is applied voltage on a per-phase basis, R_1 is a per-phase stator winding resistance, X_1 is a per-phase stator winding reactance, R_r is a per-phase rotor winding resistance, X_b is a per-phase rotor winding leakage reactance at $s = 1$ (i.e. the rotor is at rest), $X_r = s\,X_b$ is a per-phase rotor winding leakage reactance at any slip s, X_m is a per-phase magnetization reactance, R_c is a per-phase equivalent core-loss resistance, E_1 is a per-phase induced EMF in the stator winding, E_b is a per-phase induced EMF in the rotor winding at $s = 1$, $E_r = sE_b$ is a per-phase induced EMF in the rotor winding at slip s, I_r is a per-phase rotor winding current, I_1 is a per-phase current supplied by the source, $I_\Phi = I_c + I_m$ is a per-phase excitation current, I_c is a per-phase core loss current, and I_m is a per-phase magnetization current.

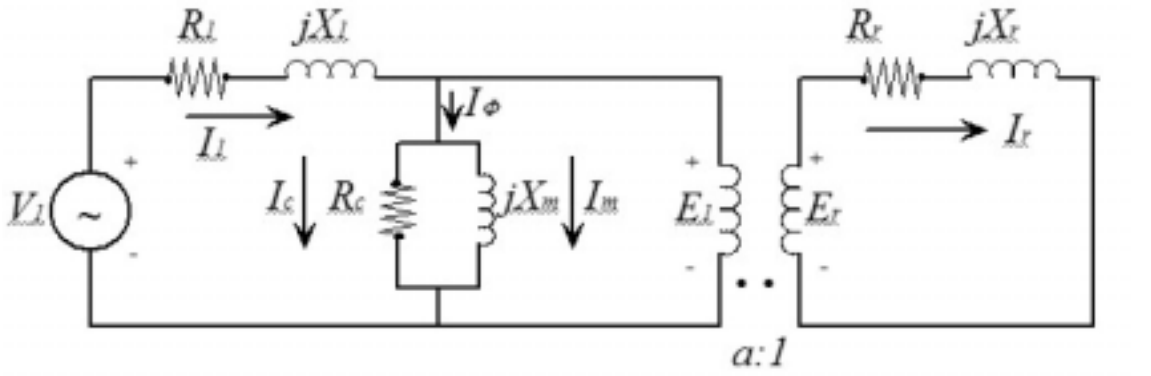


Figure 2.2. Per-phase equivalent circuit of a balanced three-phase induction motor [Guru01].

Let us define the ratio of transformation:

$$a = \frac{N_1 k_{w1}}{N_2 k_{w2}} \tag{2.3}$$

where k_{w1} is the winding factor of the stator winding, k_{w2} the winding factor of the rotor winding, N_1 turns per phase of the stator winding, and N_2 turns per phase of the rotor winding. With the ratio of transformation the equivalent circuit in the Fig. 2.2 can be transformed to the circuit presented in Fig. 2.3. This circuit is called the exact equivalent circuit of a balanced three-phase induction motor on a per-phase basis. In Fig. 2.3:

$$\begin{aligned} R_2 &= a^2 R_r \\ X_2 &= a^2 X_b \quad . \\ I_2 &= \frac{I_r}{a} \end{aligned} \tag{2.4}$$

In the circuit, R_2 represents the actual resistance of the rotor and $R_2/s(1\text{-}s)$ is called the load resistance or dynamic resistance. It depends on the speed of the motor and is said to be the electrical equivalent of a mechanical load on the motor.

A more detailed derivation of the exact equivalent circuit can be found in [Guru01].

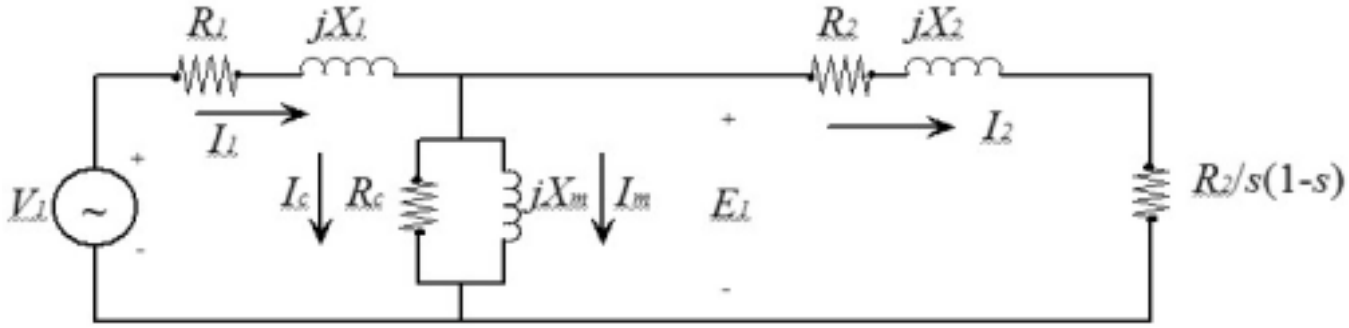


Figure 2.3. The exact equivalent circuit of a balanced three-phase induction motor on a per-phase basis [Guru01]

2.3 Finite element analysis of the magnetic field

Induction motor modelling based on the equivalent circuit is useful only, when the motor is healthy and fed with the sinusoidal power. If faulty models are needed or frequency converters are used between the power source and the motor, the equivalent circuit is not valid, and more enhanced models are required. An alternative for the equivalent circuit is calculation of the motor based on the finite element analysis (FEM) of the magnetic field. The approach is computationally challenging, but it allows highly accurate imitation of the motor performance.

The following is strongly based on [Arkkio90], where FEM of the magnetic field of the cage induction motor fed by static frequency converters is presented. The field is assumed to be two-dimensional, and the time-dependence of the field and the motion of the rotor are modelled by the Crank-Nicholson time-stepping method.

Magnetic vector potential **A** satisfies the equation:

$$\nabla \times (\nu \nabla \times \mathbf{A}) = \mathbf{J} \tag{2.5}$$

where ν is the reluctivity of the material and **J** is the current density. The current density can be expressed as a function of the vector potential and the electric scalar potential Φ:

$$\mathbf{J} = -\sigma \frac{\partial \mathbf{A}}{\partial t} - \sigma \nabla \Phi \tag{2.6}$$

where σ is the conductivity of the material. In the two-dimensional model, the vector potential and the current density have only the z-components:

$$\begin{aligned} \mathbf{A} &= A(x, y, t)\mathbf{e}_z \\ \mathbf{J} &= J(x, y, t)\mathbf{e}_z \end{aligned} \quad . \tag{2.7}$$

The scalar potential Φ has a constant value on the cross-section of a two-dimensional conductor and it is a linear function of the z-coordinate. The gradient of the scalar potential can be expressed by the aid of the potential difference u induced between the ends of the conductor. By substituting (2.6) in (2.5), the field equation becomes:

$$\nabla \times (\nu \nabla \times \mathbf{A}) + \sigma \frac{\partial \mathbf{A}}{\partial t} = \frac{\sigma}{l} u \mathbf{e}_z \ , \tag{2.8}$$

where l is the length of the conductor. A relation between the total current i and the potential difference u is obtained by integrating the current density (2.6) over the cross-section of the conductor:

$$u = Ri + R \int_{\mathrm{S}} \sigma \frac{\partial \mathbf{A}}{\partial t} \cdot d\mathbf{S} \ , \tag{2.9}$$

where R is the DC resistance of the conductor. The circuit equations for the rotor cage are constructed by applying Kirchhoff's laws and (2.8) for the potential difference. The end-region fields are modelled by constant end-winding impedances in the circuit equations. The equations of the stator winding are simplified by assuming constant current densities on the cross-sections of the coil sides.

A time-dependent field is solved by discretizing the time at short time intervals Δt and evaluating the field at times t_1, t_2, t_3,… ($t_{k+1} = t_k + \Delta t$). In the Crank-Nicholson method, the vector potential at time t_{k+1} is approximated:

$$\mathbf{A}_{k+1} = \frac{1}{2}\left\{ \frac{\partial \mathbf{A}}{\partial t}\bigg|_{k+1} + \frac{\partial \mathbf{A}}{\partial t}\big|_{k} \right\} \Delta t + \mathbf{A}_k \,. \tag{2.10}$$

By adding the field equations written at times t_k and t_{k+1} together and substituting the sum of derivatives from (2.10), the equation:

$$\nabla \times (\nu_{k+1} \nabla \times \mathbf{A}_{k+1}) + \frac{2\sigma}{\Delta t}\mathbf{A}_{k+1} = \frac{\sigma}{l} u_{k+1} \mathbf{e}_z - \left\{ \nabla \times (\nu_k \nabla \times \mathbf{A}_k) - \frac{2\sigma}{\Delta t}\mathbf{A}_k - \frac{\sigma}{l} u_k \mathbf{e}_z \right\} \tag{2.11}$$

is obtained. The potential difference vector in (2.9) is discretized in the same way as the field equation, the result being:

$$\frac{1}{2}(u_{k+1} + u_k) = \frac{1}{2} R (i_{k+1} + i_k) + R \int_{\mathrm{s}} \sigma \frac{\mathbf{A}_{k+1} - \mathbf{A}_k}{\Delta t} \cdot d\mathbf{S} \,. \tag{2.12}$$

Equations (2.11) and (2.12) form the basic system of equations in the step-by-step formulation. Starting from the initial values and successively evaluating the potentials and currents of the next time steps, the time-variation of the quantities is worked out. The magnetic field, the currents and the potential differences of the windings are obtained from the equations, and most of the other machine characteristics can be derived from these quantities. The electromagnetic force acting between the stator and rotor is computed from the air-gap field using the method developed by Coulomb [Coulomb83].

In a general time-stepping analysis of a running motor the equations for rotor and stator fields are written in their own coordinate systems. The solutions of the two field equations are matched each other in the air gap. The rotor is rotated at each time-step by an angle corresponding to the mechanical angular frequency. The rotation is accomplished by changing the finite element mesh in the air gap.

The finite element discretization leads to a large nonlinear system of equations in which the unknown variables are the nodal values of the vector potential and the currents or potential differences of the windings. The equations can be solved by the Newton-Raphson method.

In this thesis, FEM models are used to generate virtual measurement data for construction of the fault diagnostics system of induction motors. In addition to simulated currents and forces, also vibrations are studied as fault indicators. However, due to complexity of the interaction between forces on the rotor and vibration signals, vibrations are not included to the numeric model. Their behaviour in fault diagnostics is studied with real measurement signals from artificially damaged motors.

2.4 Faults and measurements for diagnostics

The faults in induction motors can occur in any of the three components of the motor: stator, rotor or bearings. In [Epri82], a large survey on faults in the motors is carried out. The survey contains 5000 motors, 97% of those three-phase squirrel cage induction motors. In Fig. 2.4, the occurrence of the individual faults is presented based on the survey. The most common fault is related to worn motor bearings, and it will generate extra vibrations, noise and possible misalignment of the rotor shaft. Most of the stator related faults are due to degraded insulation in stator windings causing an inter-turn, phase-to-phase or phase-to-ground short circuits. These are serious faults that result in a complete machine failure. Rotor faults can be divided into faults related to rotor eccentricity and

physical damage of the rotor, and they usually develop slowly although in the end the broken bars may damage the stator windings.

In this study, only rotor and stator related faults are considered. The bearing faults were left out, because there were neither simulation nor measurement data available for studying these fault types. This is a drawback for a developed fault diagnostics system, because bearing faults are common induction motor faults. However, trustworthy methods for detection of bearing faults based on motor vibrations are already reported. For example, in [Lindh03] such a method is presented.

The physical damages of the rotor are studied in the form of broken rotor bar (BB) and broken rotor end ring (BR), whereas dynamic and static eccentricity (DE, SE) of the rotor represent the eccentricity faults. Shorted turn (ST) and shorted coil (SC) in stator windings represent the stator related faults. In addition to these, the insulation system condition monitoring is considered in general. Small defects in insulation systems may give rise to small electrical discharges (partial discharges, PD), which cause electrical treeing and finally destroy the insulation. Successful detection and analysis of the small insulation flaws aids to prevent the more severe fault situations.

It is found out that a variety of measurements can be applied to collect information that is useful in the detection of induction motor faults. In this thesis, three of them are elaborated on: stator current of the motor, vibrations of the motor and PD in the insulation system of the motor. Vibration analysis has been used in motor fault detection for decades. Each fault in a rotating machine produces vibrations with distinctive characteristics that can be measured and compared with reference ones in order to perform the fault detection and diagnosis. Motor current monitoring is also called motor current signature analysis (MCSA) and it is widely studied, because no extra instrumentation is needed, if the faults can be detected based on the current. It is also claimed that MCSA give the same information on motor condition as vibration measurements [Benbouzid00]. PD analysis is applied in the detection of faults in motor insulation. Defects in the insulation system cause characteristic PD distributions that can be measured and analysed. The measurement system is presented in Chapter 3.4. Certain PD types are harmful and cause electrical treeing that finally breaks the insulation. In addition to these, also other sensors appear in the fault diagnostics of motors e.g. air-gap and external magnetic flux densities, rotor position and speed, internal and external temperatures or output torque [Benbouzid00].

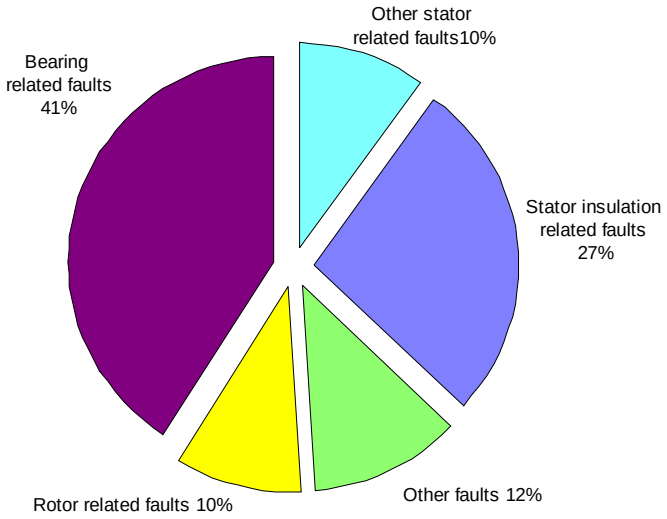


Figure 2.4. Occurrence of motor faults [Epri82]. Bearing related faults occur most often followed by stator insulation related faults.

3. Fault diagnostics and condition monitoring of the induction motor

Fault diagnostics of the induction motor is a wide subject of research. However, despite of the efforts invested to the subject a condition monitoring system is still missing, which could show various different faults in a trustworthy way and preferable in advance before complete failure. In this chapter, fault diagnostics and condition monitoring of the induction motor is considered. Firstly, the strategies for fault diagnostics are reviewed in general. It is noticed that data-based diagnostic strategies are the most interesting approaches for the induction motor fault diagnostics, and, especially, of all data-based modelling methods, support vector machine offers great benefits. Finally, the three induction motor monitoring approaches considered in the thesis: vibration monitoring, MCSA and PD diagnostics, and existing fault diagnostics strategies based on these measurements are reviewed.

3.1 Fault diagnostics methods in general

Automated fault diagnostics and condition monitoring are important parts of most of the world's industrial processes. Uncontrollable faults in the processes may cause considerable economical losses, degrade the quality of the process performance or even cause serious damage for human life or health and the environment. The fault diagnostics is usually divided in two parts: fault detection and fault identification. Early detection of faults prevents performance and quality degradation and damage of machinery or danger to human life. Identification of faults enables the right decisions for further repair or alarm actions. There exists a wide variety of techniques for carrying out the fault diagnostics tasks, and there are also several ways to categorize these techniques. In this section, the fault diagnostics techniques are divided in signal-based and model-based methods. Signal-based methods build the fundamentals for condition monitoring systems, and usually their performance can be enhanced with various model-based techniques.

When applying signal-based methods, the condition monitoring of a process is carried out just by monitoring process signals. One of the simplest examples is the measurement of liquid level in a tank and alarming, when it rises above or decreases below a certain limit. These kinds of fault diagnostics systems are implemented even in many technical devices that are in people's everyday use. In large industrial processes, the measured signals can be much more sophisticated, and they can be manipulated to distinctly reveal the characteristics of the process operation.

Signal monitoring can be carried out either in time domain, frequency domain or time-frequency domain. In time-domain, the fault related features of the signals may be extracted by calculation of various statistical figures such as means, variances or kurtosis of the signals. In frequency domain spectral analysis of signals are made and fault related frequencies are monitored. This is a common approach for example in condition monitoring of rotating electrical machines. In addition to traditional spectral analysis, higher order spectral (HOS) analysis can be carried out. Analysing the signal in frequency domain often gives valuable information of the process operation, but when transforming the signal from time domain into frequency domain information may be lost. For example, in frequency domain one cannot detect transients of process operation. To overcome this problem, signal monitoring methods in time-frequency domain are developed. Applying the short-time Fourier transform (STFT) or wavelet analysis suitable measures can be calculated to combine the information of the signal behaviour in both time and frequency domain.

Modern fault detection and identification usually start from signal monitoring, and various model-based methods are applied to create fully independent and intelligent condition monitoring systems. Model-based fault diagnosis methods take advantage of the plant models, and the idea is to calculate such quantities from the models that reflect inconsistencies between nominal and faulty system operation. The models can be divided into three classes: analytical [Isermann93], data-based [Patton99] and knowledge-based [Isermann98] models. Analytical models base on the known physical interactions in the diagnosed plant, whereas data-based models are built based on the data retrieved from the process under study. Knowledge-based models rely on human-like knowledge of the process and its faults.

Usually, the application of model-based methods can be divided in two parts: residual generation and decision making. In Fig. 3.1, a fault diagnostics scheme is presented. In the first part, process models in healthy and faulty operation are applied with real process measurements to generate residuals describing the current condition of the process. In the second part, the condition is decided based on the residuals. Both parts apply individual models that can be either data-based, knowledge-based or based on analytical models. The residual generation in the fault diagnostics scheme presented in Fig. 3.1 is based on model and process outputs, but residuals can be generated in several ways like with model parameters estimated from process measurements.

A typical example of complete fault diagnostics system is application of a data-based method to create the process models and residuals in different conditions, and application of a knowledge-based model in decision making. It is also possible to utilise simple signal manipulation methods to generate residuals, and apply more enhanced models only in the decision making part of the system. Basic signal monitoring is actually this kind of approach. A simple knowledge-based model is required for making conclusions on the process condition even if the diagnosis is based on process signals. In this thesis, the fault diagnostics system is built with various residual generation approaches, but the decision making part relies on a data-based classification model.

3.1.1 Analytical models

Analytical models are built based on known physical interactions in the process. Analytical models can be applied using observers (e.g., [Liu97]), parameter estimation (e.g., [Isermann93]) or parity equations (e.g., [Gertler92]). Using observers, the underlying idea is to estimate the system outputs with the system model from the available inputs and outputs of the system. Then the difference between the estimated and the actual outputs is calculated and fault diagnosis is based on this measure. If the difference is small, the model can be considered to describe the process operation. Parameter estimation approach makes use of the assumption that faults of a dynamic system reflect on the physical parameters of the process (e.g. friction, mass velocity resistance) and thus also on the model parameters. Faults are detected through the estimation or identification of model parameters. As described by [Gertler92], parity equations are mathematical relationships linking a number of variables, arranged in such a way that all terms appear on the same side of the equation. Parity equations can be statistical or dynamical, and the fault diagnostics bases on the output pattern of the equations.

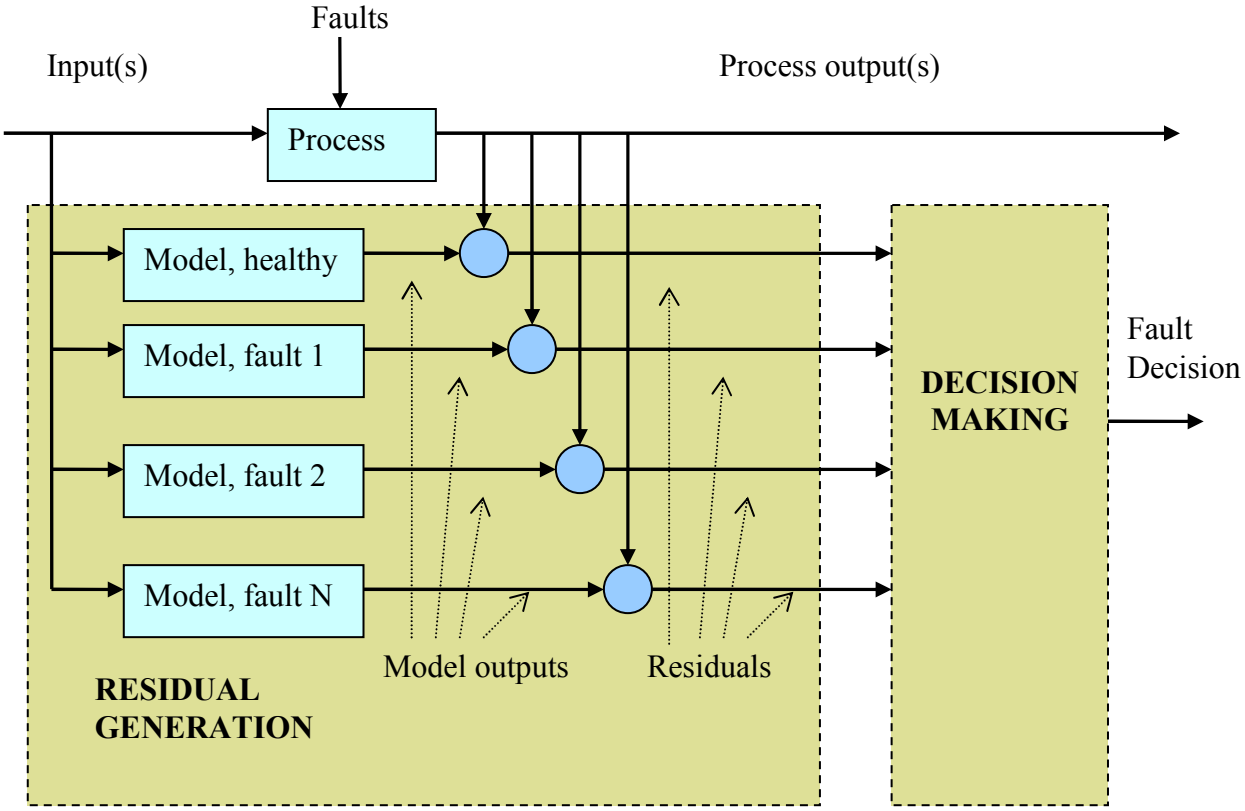


Figure 3.1. Model-based fault diagnostics scheme can be divided into residual generation and decision making tasks.

3.1.2 Knowledge-based models

Knowledge-based models like expert systems or decision trees apply human-like knowledge of the process for fault diagnosis. In fault diagnostics, the human expert could be a person who operates the diagnosed machine or process and who is very well aware of different kinds of faults occurring in it. Building the knowledge base can be done interviewing the human operator on faults occurring in the diagnosed machine and on their symptoms.

Traditional expert systems can be enhanced with fuzzy logic [Wang97]. Expert systems are usually suitable for problems, where a human expert can linguistically describe the solution. Typical human knowledge is vague and inexact, and handling this kind of information has often been a problem with traditional expert systems. For example, the limit, when the temperature in a sauna is too high, is vague in human mind. Fuzzy logic provides a systematic framework to process vague, qualitative knowledge. It is speculated that in the future, most of the expert systems use fuzzy sets and fuzzy logic instead of traditional crisp sets.

Considering fault diagnosis, fuzzy systems are useful, because fault diagnosis often needs a knowledge-based treatment. In practice, it is very difficult to obtain adequate representations of the complex and highly non-linear behaviour of faulty plants using quantitative models. The use of fuzzy qualitative models can also take account of the uncertainties associated with describing the system [Isermann98] [Dexter95].

3.1.3 Data-based models

When considering fault diagnostics of the induction motor it is difficult to develop an analytical model that adequately describes the motor performance in its all operation points with any power source. Knowledge-based models may be utilised together with a simple signal-based diagnostics, if the expert knowledge of the process is available. However, it is often impossible even for a human expert to distinguish faults from the healthy operation, and, also, multiple information sources may need to be used for trustworthy decision making. Thus, the data-based models are the most interesting approach for the induction motor diagnostics.

Data-based models are applied when the process model is not known in the analytical form and expert knowledge of the process performance under faults is not available. The data-based models can be created in numerous ways. The most traditional approach is time series analysis. The resulting models can be utilised in the same way as analytical models, although model parameters do not necessarily equal to any physical parameters of the process. Time series analysis results in linear models. Time series models can also be extended to nonlinear case, e.g. using Hammerstein models, neural networks (NN) or fuzzy systems.

During the last years NN based models like multilayer perceptrons (MLP), radial basis function (RBF) networks or self organising maps have been a popular research subject [Haykin99], and also their application in the data-based fault diagnosis is widely studied [Sorsa95] [Patton99]. With NN models it is possible to estimate a nonlinear function without requiring a mathematical description of how the output functionally depends on the input – NNs learn from examples. The most commonly mentioned advantages of NNs are their ability to model any non-linear system, the ability to learn, the highly parallel structure and the ability to deal with inconsistent or noisy data.

In the fault diagnostics, some of the difficulties of using analytical models can be overcome, and fault diagnosis algorithms can be made more applicable to real systems using NNs. NNs can be used both to generate residuals and to make a fault decision. One of the main features of NNs is their ability to learn from examples. Thus, NNs are often used in situations, where it is possible to get plenty of measurement data of the system. The large amount of numerical data from the system is also an essential requirement for training a NN. Difficulties occur in creating a reliable network, if there are not enough measurements available from all operation states of the process. Another disadvantage of NNs is that the net architecture with weighting factors is difficult to figure out by human. This may be a problem in tuning the system, or explaining the diagnosis results to a system operator.

Application of a NN in the decision making part of the fault diagnostics system is also called NN based fault classification or pattern recognition**.** Classification and pattern recognition are general names for data-based algorithms that classify or categorize things based on multiple numerical measurements, i.e. features. The classification methods can be applied in the fault diagnostics also without a distinct residual generation and decision making parts. The classifiers can be trained to represent direct relationships between measurement data of the system and certain fault conditions. In addition to NN based classification models, there exist numerous other classification algorithms such as the traditional Bayesian classifier, linear and quadratic discriminant analysis, simple distance based classifiers or the SVM [Duda01].

SVM gives refreshing views on conventional pattern recognition and classification systems. It has several benefits compared to e.g. statistical classifiers or MLPs. The most important benefit is its efficiency in high dimensional classification problems, where statistical classifiers often fail. Linear and quadratic discriminant analyses apply the inverse of covariance matrix of the vectors to be classified requiring estimation of the covariance matrix with high accuracy. To estimate high

dimensional covariance matrices well one needs an unpredictably large number of observations. When applying SVM, the generalisation ability of the classifier can be measured only with the number of samples locating on the border of two classes regardless of the dimension of the input space. Also the computations are independent of the dimension of the input space, because they are handled through Gram matrices of the input data. The other benefit of SVM compared to statistical classifiers is its general applicability to nonlinear problems. MLPs or RBF networks can also be applied in nonlinear problems, but SVM outperforms them when considering the globality of solution. Training of the SVM results in a global solution for the problem under study, whereas MLPs and RBF networks may have many local minima leading to not a trustworthy solution. Further benefit of SVM is the sparseness of solution. With a low number of samples near the class border, the actual classification task can be carried out very efficiently and fast. Due to these reasons, in this thesis, the SVM is chosen to build the data-based induction motor fault diagnostics.

SVMs have been successfully applied to various classification problems. For example to:

- text categorization e.g. in [Joachims97]
- image recognition e.g. in [Pontil98]
- phoneme classification e.g. in [Salomon01]
- hand written digit recognition e.g. in [Boser92]
- medicine, breast cancer prognosis e.g. in [Freiss98]
- bioinformatics, protein fold recognition e.g. in [Ding01]
- gene expression e.g. in [Brown97]

However, SVM based classification seems not to have been applied to the fault diagnostics of induction motors before this research. Although, SVM has shown good performance in different kinds of classification applications, its appropriateness even to fault diagnosis in general has not been widely studied. In addition to this thesis, the first studies on SVM for fault diagnosis have been published just recently. For example, Saunders & al. examine SVM to determine correct repairs for faults from past production history [Saunders00]. One of the first diagnostics applications is [Ypma992], where support vector data description is used in condition monitoring of a submersible pump. In [Rychetsky99], engine knock detection is carried out with SVM. Feng & al. [Feng02] apply SVMs to quality monitoring in robotized arc welding. In [Zöllner02], SVM is used in the diagnosis of large inspection datasets. In [Gao02] SVM is used in the valve fault diagnosis. Further, Yu & al. study SVM in the fault diagnosis of chemical process in [Yu02], and in [Ribeiro02], the injection molding machine diagnosis is carried out with SVM. In [Batur02] SVMs are applied in the fault detection of heat exchangers. In many of the studies, SVM is compared to MLPs or RBF networks and it seems to give very promising results.

3.2 Vibration monitoring

3.2.1 Spectral analysis

Spectral analysis of vibrations has been used in rotating machines fault diagnosis for decades [Betta01][Marcal00][Laggan99]. It is claimed that vibration monitoring is the most reliable method of assessing the overall health of a rotor system. Machines have complex mechanical structures that oscillate and coupled parts of machines transmit these oscillations. This results in a machine related frequency spectrum that characterizes healthy machine behaviour. When a mechanical part of the machine either wears or breaks up, a frequency component in the spectrum will change. In fact,

each fault in a rotating machine produces vibrations with distinctive characteristics that can be measured and compared with reference ones in order to perform the fault detection and diagnosis. Influence of different faults on the vibration spectrum are presented e.g. in [Betta01].

Examples of using neural networks in classification of faults based on vibration signals are presented in [Yang00], [Alguindigue93], [Penman94], and [Li00]. In these articles, NNs have been used as pattern recognition tools, and the training is carried out in frequency domain.

In addition to spectrum analysis, other signal processing tools are applied on vibrations to reveal influence of faults in the machine. For example applications of higher order spectrum (HOS) analysis, cepstrum analysis or regular time series analysis have given good results.

3.2.2 Higher order spectrum and cepstrum analysis

Higher order spectra (also known as polyspectra) are defined in terms of higher order statistics and they have attractive properties considering signal processing for purpose of condition monitoring. Firstly, in the HOS analysis, additive Gaussian noise is automatically suppressed. Secondly, information due to deviations from Gaussianity can be extracted, and finally, nonlinear systems as well as nonminumum phase systems can be identified [Nikias93]. In Fig. 3.2, the computation of HOS is presented. Notice that second order HOS equals to power spectrum of a signal. The third order HOS is also called bispectrum and the fourth order HOS is called trispectrum.

In [Arthur00], HOS has been studied in vibration monitoring of rotating machines. The authors also suggest an enhancement to HOS that summaries both magnitude and phase information of the original time series. Given a polyspectral measure M with magnitude $|M|$ and phase $\angle M$, the phase being bound by $-\pi$ and π, the enhanced measure M_{enh} is defined to be:

$$M_{enh} = |M| \times \frac{\pi - |\angle M|}{\pi} \quad . \tag{3.1}$$

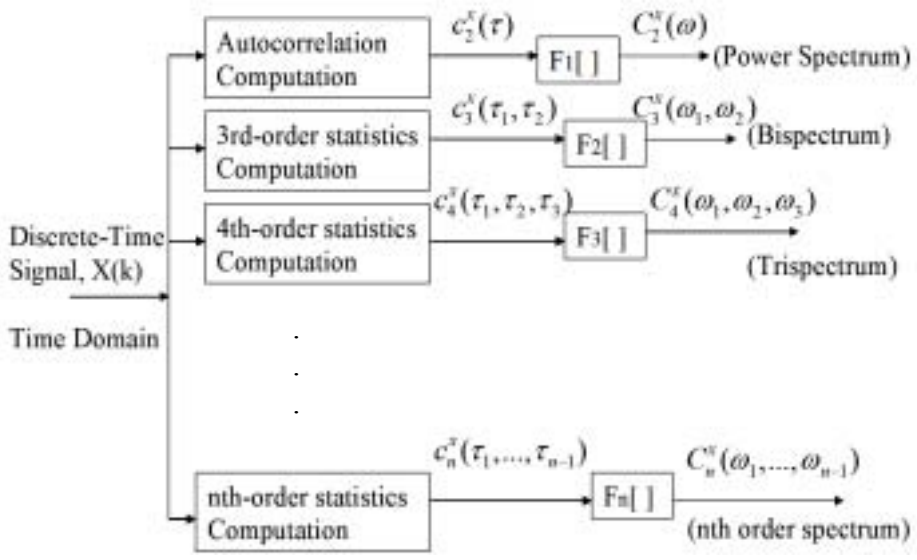


Figure 3.2. Calculus of higher order spectra. $F_n[\ ]$ denotes n-dimensional Fourier transform [Nikias93].

HOS and fuzzy logic based models have been applied in [Lasurt00], where HOS analysis is used as a pre-processing of a machine vibration signal. Due to complexity of HOS signatures data reduction and parameterization are carried out first, and fuzzy logic procedure is then applied to enable diagnosis of the machine fault.

One of the rare applications of genetic algorithms in condition monitoring of electrical machines can be found in [Jack00], where a genetic algorithm is used to isolate the features of input space that are able to indicate faults. The preliminary feature set contained HOS estimates with different time constants.

Cepstrum analysis has been utilised in vibration monitoring e.g. in [Merwe02]. One of its most important characteristics is the fact that any periodicities in a frequency spectrum will be shown as one or two specific components in the cepstrum. The cepstrum is the forward Fourier transform of logarithm of the spectrum of the signal. The real cepstrum of a signal z is defined as:

$$RCEPS(z) = real\left(F^{-1}\log(|F(z)|)\right), \tag{3.2}$$

where F denotes Fourier transform and F^{-1} inversed Fourier transform.

3.2.3 Analysis in time domain

When carrying out vibration analysis in time domain, some simple quantities can be utilised such as root mean square (RMS), crest factor, kurtosis and other statistical moments, but often they do not offer enough information on the vibrations for thorough diagnosis. A traditional tool for more enhanced signal description in time domain is time series analysis. Whereas HOS and cepstrum analyses are static signal description tools, the time series analysis is suitable for description of dynamic phenomenon – such as the behaviour of induction motor. For example in [Ypma992], AR modelling for vibration signal description is studied, although there the authors study condition monitoring of a submersible pump. The model of vibration signal is estimated in time domain and the model coefficients are used as indicators for the motor condition.

The general form of an nth order autoregressive (AR) model of a time series $x(t)$ is following:

$$\begin{aligned} & x(t) = a_1 x(t-1) + a_2 x(t-2) + \ldots + a_n x(t-n) + e(t) \\ & \Rightarrow x(t) - a_1 x(t-1) - a_2 x(t-2) - \ldots - a_n x(t-n) = e(t) \end{aligned}, \tag{3.3}$$

where $e(t)$ is noise and the parameters $a_1, \ldots, a_n$ are called AR coefficients. The coefficients can be estimated from a time series, for example, with the least squares method.

Another example of analysis in time domain can be found in [Loparo00], where analytical models are applied in vibration monitoring. A multiple model framework is used to develop monitoring, fault detection and diagnosis system in rotating machines. Each fault to be identified is associated with a certain vibrations and rolling element bearing model structure and parameters in the rotating machinery model. Fault diagnosis is based on statistical testing of residuals of the bank of stochastic non-linear observers. The residuals of the filters are monitored, and the conditional probability that each filter model is the process model is computed, and the filter with the highest probability is declared to match the current operating condition.

3.2.4 Information fusion of multi-channel vibration measurements

Vibration monitoring system requires storing of a large amount of data. Vibration is often measured with multiple sensors mounted on different parts of the machine. For each machine there are typically several vibration signals being analysed in addition to static parameters such as load. The examination of data can be tedious and sensitive to errors. Also, fault related machine vibration is usually corrupted with structural machine vibration and noise from interfering machinery. Further, depending on the sensor position, large deviations on noise may occur in measurements.

Due to these problems intelligent compression of the multi-channel measurement data may aid in the data management for fault diagnostics purpose. Multivariate data analysis (MVDA) is a general name for various methods that aim to find structure in large amount of multivariate data. Independent component analysis (ICA) is one of those methods. ICA may be used to compress measurements from several channels into a smaller amount of channel combinations that are statistically independent sources of the original vibration measurements, and that could clearly indicate faults in the machine. Calculations for finding the independent components are relatively straight-forward. The only assumptions of applying the method are that the original sources are statistically independent of each other and they are linearly mixed. Also, ICA or other blind source separation (BSS) methods may aid in separating machine related vibration signal from interfering vibration sources.

Methods closely related to ICA are e.g. projection pursuit and beamforming. Projection pursuit aims at finding "interesting" projections in data by assuming nongaussian distributions of a projection as more interesting than gaussian, whereas in ICA, the interestingness criteria may vary. The main difference between beamforming and blind source separation methods is the criterion they use for separating a source signal: beamforming needs the direction of the main source while blind source separation algorithms do not.

ICA and other BSS methods are studied to provide a robust and reliable fault diagnostics routine for rotating machines e.g. in [Ypma991], [Ypma992], [Ypma02], [Knaak01], [Knaak02] and [Gelle01]. The existing research can be divided into two categories. The first one is applying BSS to compress vibration measurements into independent components, which could indicate a fault in a more reliable way than pure vibrations such as in [Ypma991] and [Ypma992], where ICA of multiple vibration signals is successfully utilised in fault diagnostics of a submersible pump. The second category is using BSS on separating machine signature for fault diagnosis purpose from interfering sources or from other machine signals such as in [Knaak01], [Knaak02] and [Gelle01]. In [Knaak01] and [Knaak02], BSS is applied to acoustic signals instead of vibrations, but the behaviour of vibrations and acoustic signals are quite similar.

The second category is much more widely studied than the first one. This is probably due to assumption of independence: it is more likely that separate machines or interference sources are independent from each other than fault related signal component is independent from normal machine vibrations. However, even application of ICA or other BSS method to preprocess measurement data before fault diagnosis routine has shown to improve the performance of the fault diagnosis system.

Some articles (e.g. [Ypma02]) discuss assumptions valid for the mixing process in rotating machine applications. Instantaneous mixing means that sources are assumed to linearly mix into measurements without time delays. When using convolutive models, time delays in mixing are allowed. This leads to linear mixing in frequency domain and to a solvable separation problem despite of more complex calculations. Convolutive models should be used, if sources are physically far away from each other or severe reflections or distortions of signal may happen. This is

especially the case, when signatures of two separated machines or signature of a machine and a distant interfering source are aimed to separate. If BSS is used to compress vibration measurements of one machine, instantaneous mixing model should be true enough.

3.3 Motor current signature analysis (MCSA)

During last years, MCSA has been widely studied in addition to vibration analysis for induction motor fault detection. It has been claimed that electrical measures contain the same information on the faults of motor as the vibration measurement [Kliman92]. The main benefit of using motor current as the basis of fault detection system instead of vibrations is that no extra instrumentation is needed for measurements. In particular, a large amount of research has been directed towards detecting broken rotor bars and mechanical unbalance from the spectrum of stator current. This is an ironic fact, because the rotor related faults are actually quite rare compared to e.g. bearing faults. However, in many cases these specific faults can be quite easily detected from characteristic frequencies of the stator current.

In monitoring the stator current, quite similar signal processing tools are applied with the vibration monitoring tools. The traditional way to produce the current signature is the calculation of frequency spectrum with an FFT based method, but in addition to this there exist detection schemes in literature that apply cepstrum analysis, HOS or wavelets [Benbouzid00].

Various model-based MCSA fault diagnostics systems can be found in literature. An example is [Wieser98], where the sensitivity and robustness of the on-line model based Vienna monitoring method is addressed. The proposed condition monitoring method compares the outputs of a reference model, which represents an ideal machine, to a measurement model. Observing the deviations of these two models makes it possible to detect and even locate rotor faults. The method utilises a voltage and a current model structure, which respond differently to the faulty rotor bar. Differences of the model outputs are evaluated and clustered. The same researchers have studied the method also in [Kral00] and in [Wieser97].

A NN based MCSA method can be found e.g. in [Filippetti95], where an example of using NNs for modelling an induction motor is presented. There the faulted machine models used to formalize the knowledge base of the diagnostic system are formed with NNs.

In [Schoen95], an interesting NN based clustering approach for fault diagnostics of an electrical machine is presented. There NNs are used to learn on-line the spectral characteristics of a healthy motor current. A special frequency filter is used to pass only those harmonics, which are known to be of importance in fault detection, to a NN clustering algorithm. After a sufficient training period, the NN signals a potential failure condition, when a new cluster is formed and persisted for some time.

Also, in MCSA, fuzzy logic has become common, especially, in the decision making part of the diagnostics scheme. For example, in [Nejjari99], fuzzy logic is applied to induction motor's condition monitoring and its stator and phase conditions through the amplitude features of the stator currents. Further, in [Altug99], ANFIS (Adaptive Neuro Fuzzy Inference System) -based fault diagnostics system of an induction motor is compared with another adaptive neuro-fuzzy system FALCON (Fuzzy Adaptive Learning Control Network). Altug & al. have found out that both structures provide good fault diagnostics framework under varying operation conditions.

3.4 Partial discharge (PD) analysis

3.4.1 Description of the phenomenon

Partial discharge (PD) is defined by the standards [IEC01] as a localised electrical discharge that partially bridges the insulation between conductors and which may or may not occur adjacent to a conductor. The phenomenon manifests itself in a wide range of ways; for example, internal discharges within the bulk of the insulation, surface discharges at the surface or at the interfaces of two dielectric media or as corona in air. The role of PDs in degradation and eventual failure of electrical insulation systems at high or medium voltage (HV, MV) is a well-established fact, but the severity of the effect that PD has on the insulation depends on the nature and the location of the PD generating defect. Therefore, a key step in PD analysis is the identification of the defect, through the characterization of its PD activity.

PDs may be detected with different techniques, that is, with electrical, acoustic or optical sensors. In this thesis, electrical detection is considered. A PD event consists of a charge transfer in a limited portion of the insulation, causing a current pulse to flow through the electrodes. To detect such pulses, a circuit such as the one of Fig. 3.3 is used.

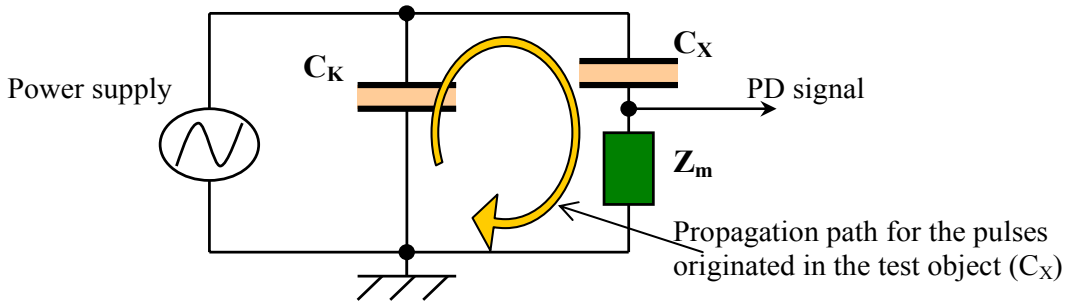


Figure 3.3. A direct PD measurement circuit [IEC01]. The capacitor C_x represents the test object, C_K is the coupling capacitor, and Z_m is the measurement impedance. The high frequency PD pulse flows only through the two capacitors.

In Fig. 3.3, the test object is represented by capacitor C_X, the coupling capacitor by C_K and the measurement impedance by Z_m. PD pulse, being a high frequency signal, flows through the two capacitors and does not travel toward the power supply, which is characterized by a strong inductive component. PD signal is actually the voltage drop of the PD current pulse at the measurement impedance. The coupling capacitor C_K is used to separate the PD signal from the test voltage. In the measurements used in this study Z_m is a 50 Ohm resistor and C_K = 1 nF, 80 pF… depending on the type of measurement object.

The circuit shown in Fig. 3.3 is called “direct”, because Z_m is in series to C_X. In alternative, an “indirect” circuit can be used [IEC01], with Z_m in series to C_k. In the latter case, pulses are detected with a sign that is opposite with respect to that of the supply voltage driving the discharges.

The PD acquisition provides four kinds of information:

- Pulse amplitude
- Pulse phase
- Pulse time of occurrence
- Pulse waveform

The shape of the PD pulse depends on the transfer function that the pulse faces when it travels from the PD source to the sensor. Hence, the pulse shape depends on the circuit. For this reason, PD pulse shape itself does not have any diagnostic meaning. However, within the same acquisition, the fact that groups of pulses show different shapes, thus facing different transfer functions, would mean that multiple sources are present, characterized by different nature and/or different location.

Partial discharge is essentially a stochastic phenomenon. Thus, statistical analysis needs to be carried out on phase, amplitude and time of occurrence of the discharges in order to characterize the discharge process. Each one of these three data sets is presented as a histogram. Usually, phase and amplitude histograms are summarized in a three dimensional representation, called phase resolved PD (PRPD) pattern (Fig.3.4), where the number of discharges is represented by means of colour. Note that, as regards the time of occurrence, a quantity called "intertime" is used for the histogram. Being t_k the time of occurrence of the kth PD, the kth intertime is $(t_k - t_{k-1})$.

Different sources may generate PD pulse signals with variety shapes and rise times as short as nanoseconds. In measurements, the sampling frequency should be larger than half of the equivalent frequency to avoid frequency aliasing. Also, the phenomenon is a stochastic process and a large amount of pulses needs to be stored to reveal the statistical features of the process. Further, the possible presence of multiple PD phenomena makes the required amount of pulses even larger. Due to these issues a considerable amount of data needs to be acquired, to accomplish a trustworthy condition monitoring process based on PDs. A great part of the PD research concentrates on developing measurement techniques that decrease the required memory buffer. For example, in [Contin02], instrumentation with a triggerable partitioned on-line memory buffer is used to limit the amount of data compared to using constant sampling frequency. Also packing data with different wavelet analysis based techniques has been widely studied in the literature. For example in [Ma02], the data from PD activity is compressed to approximately 5% of the original volume using wavelet and wavelet packet based analysis.

3.4.2 PD diagnostic strategy

Insulation systems diagnostic based on PDs can be divided in four subtasks: classification process and identification process with three levels. Classification is required for separating PD pulses to different homogeneous groups based on their sources. The steps of identification process are: identification of PD type, identification of defect characteristics, and, finally, the outputs of these are combined with the information available about the application under study (such as generator windings, high voltage cable etc.) to return the insulation condition.

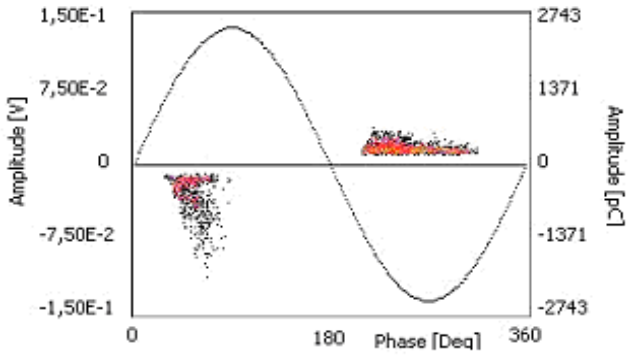


Figure 3.4. Phase and amplitude histograms are summarized in a three dimensional representation, called phase resolved PD (PRPD) pattern. The number of discharges is represented with colour.

The classification or separation of PD signals is needed to divide pulses into different groups according to their shapes. Noise signal should be separated from pulses originating from a defect and pulses from different defects should be separated from each other to achieve a homogeneous data set for further analysis. This can be done for example assuming that signals from different sources produce pulses that have different shapes such as in [Cavallini031] and [Cavallini032]. The classification is often carried out with a special classification map that is formed by mapping the PD pulses into the time-frequency plane by calculating so-called equivalent time-lengths and equivalent band-widths of the PD pulses; see e.g. [Contin02]. Each cluster in the classification map represents a certain PD generating phenomenon, and they are further analysed separately. The classification process is depicted in Fig. 3.5.

The first level of identification provides an indication about the nature of the PD with respect to five categories [Cavallini032]:

- Internal discharges are pulses which occur in air gaps delimited by dielectric surfaces, or solid dielectric and metallic electrodes, involving significant components of electric field orthogonal to the defect surface.
- Surface PDs are defined as discharges that develop on surfaces of solid insulating materials, involving significant field component tangential to the defect surface.
- Corona discharges are the PD generated in open air (gas) from a sharp edge.
- Noise can be either background noise or external disturbances, and it can be recognized by statistical tests.
- PD data that have no physical meaning or have not been successfully identified are categorized as invalid.

The three first categories correspond to the main defect types. Hence, a combination of these outputs identifies the basic nature of the defect. It is noteworthy that a defect may constitute an intermediate situation with respect to those basic categories.

The second level of identification provides a more detailed description of the defect. The defect is localized with respect to the electrodes, its shape is defined with respect to the electric field direction, the field distribution inside the defect is identified, and possible presence of electrical treeing is detected. As for the first level, output categories do not depend on the equipment under test.

The third level provides a detailed description of the defect for a specific class of equipment. A certain insulation defect may be severe for some electrical equipment and only a small drawback for some other. Therefore, the third level output is not generally valid for insulation system of any electrical apparatus, but a great deal of expertise is required at this level of the identification process to achieve trustworthy information of the insulation condition.

In addition to typical statistical measures such as means, minimums and maximums, skewness and kurtosis derived from PD diagrams, the Weibull function has been found to be important in analysis of the PD amplitude distribution. The scale and shape factors of Weibull function that fit the amplitude distribution data are found to be consistent when the PDs have the same nature [Contin94]. Also, modern analysis methods can be found from the literature. For example, in [Abdul00], different texture analysis methods are applied to three-dimensional PRPD diagram, and in [Satish95] fractal features of the distribution are applied.

During last years also different soft computing methods have been applied to PD monitoring. Different kinds of fuzzy logic based systems have shown to be suitable for PD diagnosis, because the PD phenomenon is very complex, and it often requires knowledge of a human expert to be analysed [Cavallini02], [Cavallini032], [Wenzel94], [Salama00]. It should also be noted that usually the categories defined as output of the various identification tasks are essentially fuzzy, in the sense that any situation intermediate among the PD categories is possible.

Also, a lot of research is done applying artificial neural networks on PD diagnosis. In [Danikas03], a review of using different kinds of NN models on PD monitoring can be found. The authors conclude that high recognition rates with NN methods are reported in literature, although these are often obtained either in small sample sets with well-defined artificial defects or in the absence of interference. Authors make four important observations on applying NNs for PD pattern recognition. Firstly, they note that the PD parameters with which a NN is trained or adopted are of paramount importance concerning the success of classification in addition to the NN type and structure. Secondly, they argue that stochastic behaviour of PD may lead to misclassifications, if one does not look closely at the physical processes of PD and related parameters during classification process. Thirdly, they emphasize that the problem of multiple defects should be taken into account when training a NN for PD monitoring, because in reality there hardly exists isolated defects. Finally, they suggest that separate NNs should be considered and trained for different applications instead of trying to build a universal model for monitoring PD in any possible application.

In this thesis, a task of the second level PD identification is considered: localization of the PD generating defect with respect to the electrodes. Various numerical classification methods are studied for the automated localization concentrating on SVM based classification.

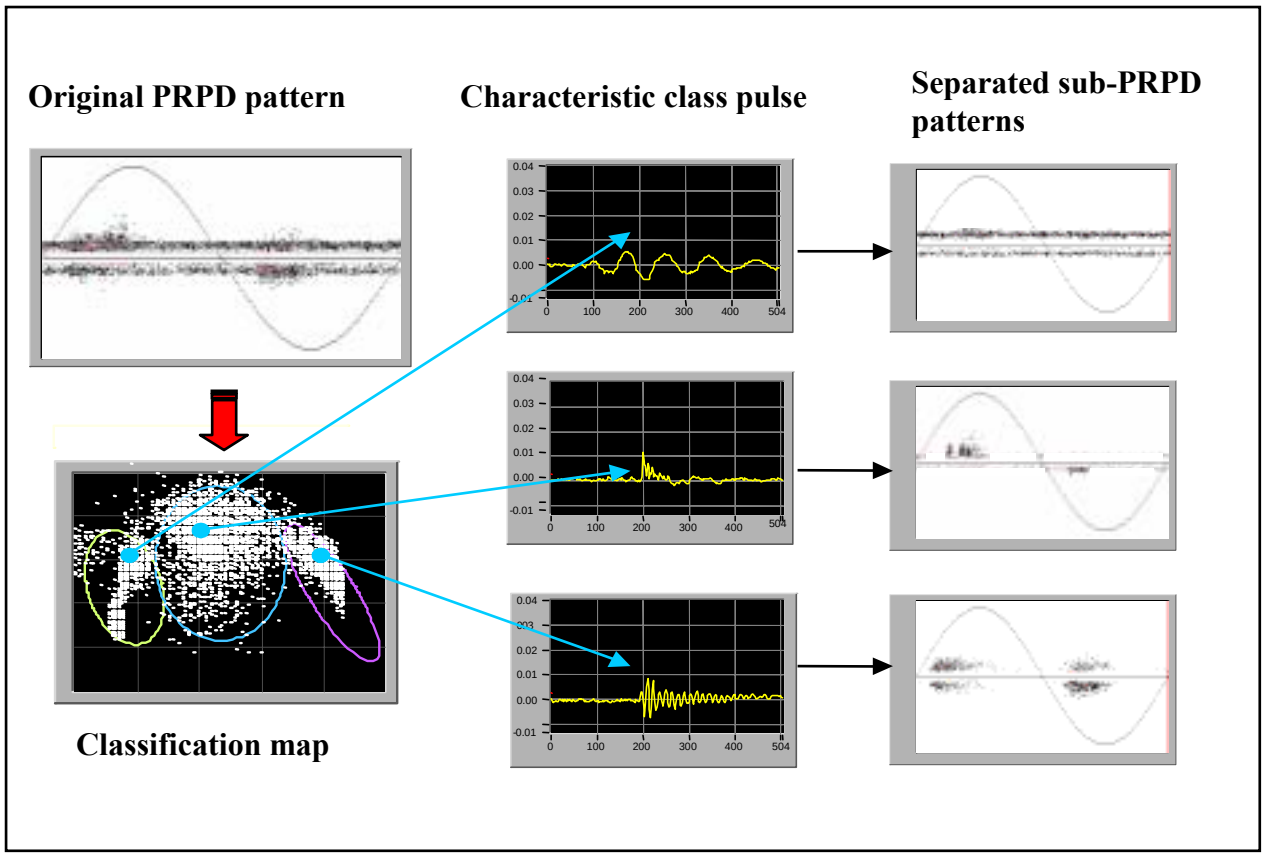


Figure 3.5. Example of the PD classification process [Contin02].

4. Support vector machine (SVM) for classification

In this thesis, support vector machine based classification is applied in three induction motor condition monitoring approaches: MCSA, vibration monitoring, and PD analysis. SVM is used to classify electromechanical faults of the motor based on motor vibrations or motor current. In addition to this, the automated insulation system defect localization is designed with application of PD measurements and SVM. In this chapter, the theory of SVM is explained, and its application in multi-class classification problems is clarified. In Chapter 4.3, also results from the publication [P3] are summarized.

4.1 Introduction to SVM

4.1.1 Vapnik-Chervonenkis dimension

It was already in the 14th century, when a logician and Franciscan friar; William of Occam stated that: "Plurality should not be posited without necessity." This issue is often referred as Occam's razor. In machine learning, Occam's razor means that the simplest hypothesis that fits the data is the best. Generally, as the model complexity grows the fit improves in the training set, but over-learning with a loss of generalization may occur in the test set. Alternatives for model selection instead of simple cross validation methods are e.g. application of Akaike information criterion, Bayesian information criterion or Vapnik-Chervonenkis dimension. The last one reveals important aspects on the classification with SVMs, and it will be considered here in more detail.

Consider a set of continuous, classifying functions $h(\mathbf{x},\mathbf{w})$, which learn to map $\mathbf{x} \rightarrow y$ by adjusting $\mathbf{w}$. Vapnik-Chervonenkis (VC) dimension, d, is a property of $h(\mathbf{x},\mathbf{w})$. VC dimension measures the capacity of the classifying functions i.e. how much complexity in the data the function is able to model. The VC dimension for a set of functions is defined as the maximum number of training samples that can be shattered by h. Function $h(\mathbf{x},\mathbf{w})$ can shatter a set of points $\mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_M$ if and only if for every possible training set of the form $(\mathbf{x}_1,y_1)$, $(\mathbf{x}_2,y_2)\ldots(\mathbf{x}_M,y_M)$ there exists a value $\mathbf{w}$ that results in zero for training error [Vapnik00]. For linear learning machines the VC dimension is equal to $n + 1$, where n is the dimension of the space, but it should be noted that this is not the case with all sets of classifying functions. Consider, for example, the family of one-parameter functions defined by $h(x,w) = \text{sign}[(\sin(wx))]$, $w, x \in \mathbb{R}^1$. The set of functions has an infinite VC dimension although it has only one free parameter [Vapnik00].

Let us study the influence of the capacity of the classifier through the risk of misclassification in the test set. If the testing and training data are independently drawn and identically distributed with cumulative probability distribution $P(\mathbf{x},y)$, the expected risk for misclassification in the test set is [Vapnik00]:

$$R(\mathbf{w}) = \int \frac{1}{2}\left|y - h(\mathbf{x},\mathbf{w})\right| dP(\mathbf{x},y)\,. \tag{4.1}$$

In typical cases, $P(\mathbf{x},y)$ is not available and most conventional training algorithms for learning machines aim to minimise the empirical risk, R_{emp}, instead of the expected risk when adjusting $\mathbf{w}$:

$$R_{emp}(\mathbf{w}) = \frac{1}{M}\sum_{i=1}^{M}\left|y - h(\mathbf{x}_i, \mathbf{w})\right|, \tag{4.2}$$

where M is the number of training samples. These kinds of algorithms do not consider the capacity of the learning machine, and this can result in overfitting, i.e. using a learning machine with too much capacity for a particular problem. The structural risk minimization principle is developed to overcome this problem. Instead of minimizing the empirical risk, the goal of structural risk minimization is to find a classifier that compromises between low empirical risk and small capacity. With $0 \leq \eta \leq 1$ the following bound holds with probability 1 - η [Vapnik00]:

$$R(\mathbf{w}) \leq R_{emp}(\mathbf{w}) + \sqrt{\left(\frac{d(\log(2M/d)+1) - \log(\eta/4)}{M}\right)}, \tag{4.3}$$

where d is the VC dimension of the classifier. The right hand side of the equation is also called "risk bound" and the second term is called "VC confidence". Notice that the equation does not contain $P(\mathbf{x},y)$ and it is easy to calculate if d is known. Hence, if a sufficiently small η is chosen and a classifier that minimizes the right hand side is found, the classifier is the one that gives the lowest upper bound on the actual risk. This is the fundamental of structural risk minimization.

Notice that the VC confidence increases when the VC dimension increases and it decreases when the number of samples increase; see Fig. 4.1. Hence, when considering linear leaning machines, the higher is the dimension of the feature space the more training data should exist to achieve good generalisation of the classifier. The computational difficulties resulting from high dimensional feature vectors are obvious, but this result shows that they also influence on the generalisation ability of the classifier. The phenomenon is called the curse of dimensionality.

Support vector machines are learning machines that are able to avoid the curse of dimensionality in both computationally and in generalisation. The SVM theory is presented thoroughly e.g. in the books [Vapnik00], [Cristianini00], [Kecman01], and in a tutorial [Burges98]. Their application in classification is briefly summarized in the next chapter.

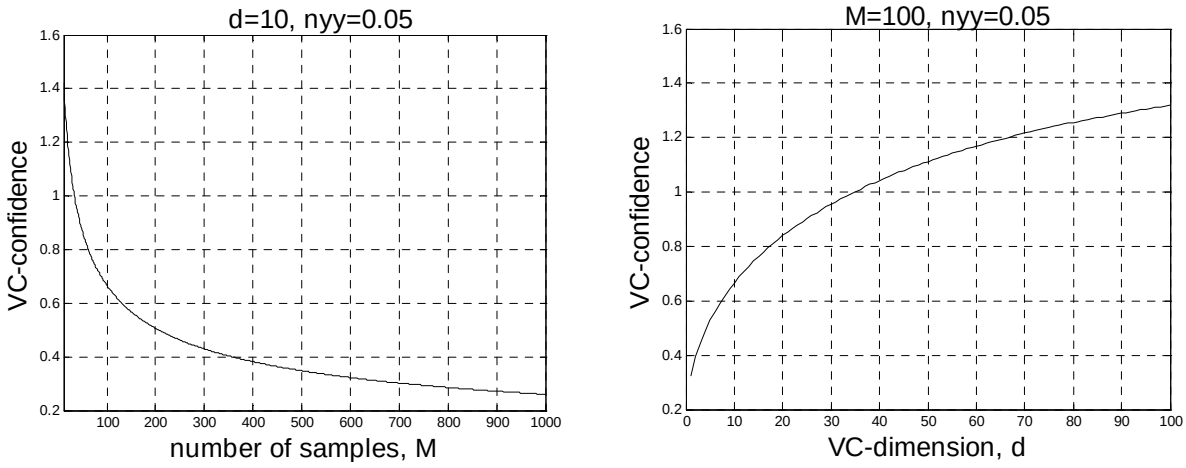


Figure 4.1. VC confidence decreases with the increasing number of samples and increases with the increasing VC dimension: a) varying VC dimension (M=100) and b) varying number of samples (d=10).

4.1.2 Δ-margin separating hyperplane

Let $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ni})^{\mathrm{T}}$, $i = 1,\ldots,M$, be a sample of $\mathbf{x} \in \mathbb{R}^n$ and belong to Class I or Class II. For linearly separable data, it is possible to determine a hyperplane that separates the data leaving one class on one side of the hyperplane, the other on the other side. This plane can be described by the equation:

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b = \sum_{j=1}^{n} w_j x_j + b = 0 \quad , \tag{4.4}$$

where $\mathbf{w} \in \mathbb{R}^n$ is a weight vector and b is a scalar. The vector $\mathbf{w}$ and the scalar b determine the position of the separating hyperplane.

Let us define the label y_i associated to $\mathbf{x}_i$ as $y_i = 1$ if $\mathbf{x}_i$ belongs to Class I, $y_i = -1$ for Class II. A separating hyperplane satisfies the constraints $f(\mathbf{x}_i) \geq 0$, if $y_i = +1$, and, $f(\mathbf{x}_i) < 0$, if $y_i = -1$. A separating hyperplane is called a Δ-margin separating hyperplane, if it satisfies $\|\mathbf{w}\| = 1$ and:

$$y = \begin{cases} 1, & if\ \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \geq \Delta \\ -1, & if\ \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \leq -\Delta \end{cases} \quad . \tag{4.5}$$

In the previous chapter, it was noted that the VC dimension of the set of hyperplanes is equal to $n + 1$, where n is the dimension of the space. However, Vapnik has shown [Vapnik00] that, if $\mathbf{x} \in \mathbb{R}^n$ belong to a sphere of radius R, the VC dimension of Δ-margin separating hyperplanes is bounded by following:

$$d \leq \min\left(\left[\frac{R^2}{\Delta^2}\right], n\right) + 1 \quad . \tag{4.6}$$

This is an important result, because this states that the VC dimension of Δ-margin separating hyperplanes can be considerably less than $n + 1$, if the margin Δ is large. Further, this means that the VC confidence in the equation of structural risk (4.3) decreases with increasing Δ, when considering separating hyperplanes.

4.1.3 Maximal margin classification

When formulating SVMs, Δ-margin separating hyperplanes are considered with aim to maximize the margin and setting $\Delta = 1/\|\mathbf{w}\|$, so that $f(\mathbf{x}_i) \geq +1$, if $y_i = +1$, and $f(\mathbf{x}_i) \leq -1$, if $y_i = -1$. The inequalities can be combined with (4.4) so that:

$$y_i f(\mathbf{x}_i) = y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1, \quad \text{for} \quad i = 1,\ldots,M \,. \tag{4.7}$$

The margin can also be calculated through the geometrical margin γ, which is defined as half of the sum of the distances between arbitrary separating hyperplane and the nearest negative and positive datum ($\mathbf{x}^-$ and $\mathbf{x}^+$):

$$\gamma = \frac{1}{2}\left(\left(\frac{-\mathbf{w}^{\mathrm{T}}(\mathbf{x}^{\mathrm{o}} - \mathbf{x}^+)}{\|\mathbf{w}\|_2}\right) - \left(\frac{-\mathbf{w}^{\mathrm{T}}(\mathbf{x}^{\mathrm{o}} - \mathbf{x}^-)}{\|\mathbf{w}\|_2}\right)\right) = \frac{1}{2\|\mathbf{w}\|_2}\left((\mathbf{w}^{\mathrm{T}}\mathbf{x}^+) - (\mathbf{w}^{\mathrm{T}}\mathbf{x}^-)\right) \quad , \tag{4.8}$$

where $\mathbf{x}^{\mathrm{o}}$ is a point on a hyperplane. The separating hyperplane that maximizes the margin is called the optimal separating hyperplane. An example of optimal separating hyperplane of two datasets is

presented in Fig. 4.2. The optimal separating hyperplane can be searched among the so-called canonical hyperplanes, which fulfil $\mathbf{w}^{\mathrm{T}}\mathbf{x}^{+}+b=1$ and $\mathbf{w}^{\mathrm{T}}\mathbf{x}^{-}+b=-1$ [Cristianini00], leading to the following simplified expression for the geometrical margin:

$$\gamma=\frac{1}{\|\mathbf{w}\|_2} \quad . \tag{4.9}$$

Since the optimal hyperplane maximizes the margin, it can be found by solving the following convex quadratic optimisation problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to } \mathrm{y_i}(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+b)\geq 1 \quad . \end{aligned} \tag{4.10}$$

Notice that the solution for this optimization problem is global. This is a great benefit compared to e.g. MLPs or RBF networks that may have many local minima meaning that the globally optimal solution is not usually guaranteed.

If the number of attributes of data examples is large, the calculations can be considerably simplified by converting the problem to the equivalent Lagrange dual problem. The Lagrange function for (4.10) is:

$$L(\mathbf{w},b,\boldsymbol{\alpha})=\frac{1}{2}(\mathbf{w}^{\mathrm{T}}\mathbf{w})-\sum_{i=1}^{M}\alpha_i\left[y_i\left(\left(\mathbf{w}_i{}^{\mathrm{T}}\mathbf{x}_i\right)+b\right)-1\right], \tag{4.11}$$

where $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_M)^{\mathrm{T}}$ is the Lagrange multiplier. The dual problem is:

$$\begin{aligned} &\text{maximize } L(\mathbf{w},b,\boldsymbol{\alpha}) \\ &\text{subject to } \alpha_i\geq 0, \quad i=1,\ldots,M \quad . \end{aligned} \tag{4.12}$$

By differentiating (4.11) with respect to $\mathbf{w}$ and b and imposing stationarity, we get:

$$\begin{aligned} &\frac{\partial L}{\partial \mathbf{w}}(\mathbf{w},b,\boldsymbol{\alpha})=\mathbf{w}-\sum_{i=1}^{M}y_i\alpha_i\mathbf{x}_i=\mathbf{0} \\ &\frac{\partial L}{\partial b}(\mathbf{w},b,\boldsymbol{\alpha})=\sum_{i=1}^{M}y_i\alpha_i=\mathbf{0} \quad . \end{aligned} \tag{4.13}$$

From (4.11)-(4.13) the dual representation of the optimisation problem is achieved:

$$\begin{aligned} &\text{maximize } W(\boldsymbol{\alpha})=\sum_{\mathrm{i=1}}^{\mathrm{M}}\alpha_i-\frac{1}{2}\sum_{i,k=0}^{M}\alpha_i\alpha_k y_i y_k \mathbf{x}_i{}^{\mathrm{T}}\mathbf{x}_k \\ &\text{subject to } \sum_{\mathrm{i=1}}^{\mathrm{M}}y_i\alpha_i=0 \\ &\qquad\qquad \alpha_i\geq 0, \quad i=1,\ldots,M \end{aligned} \quad , \tag{4.14}$$

where $\boldsymbol{\alpha}$ is the Lagrange multiplier vector. Let us assume that optimal solution for the dual problem is $\boldsymbol{\alpha}^*$ and b^*. According to the Karush-Kuhn-Tucker theorem, the equality condition in (4.10) holds for the training input-output pair $(\mathbf{x}_i,y_i)$ only if the associated α_i^* is not 0. In this case, the training example $\mathbf{x}_i$ is called a support vector (SV), i.e. the SVs are such training samples that are on the margin of two datasets. In Fig.4.2, SVs are bolded. These samples give the name to this learning machine, because they show to be very important in the classification both computationally and concerning the generalisation. Firstly, they provide a sparse solution to the classification problem, and, secondly, Vapnik has presented a simple and powerful result that connects the number of SVs

and the generalisation ability of the classifier [Vapnik00]. If the training data set contains M samples that are separated by the maximal margin hyperplane, the bound for expectation of the probability of test error can be calculated as:

$$E[P_{error}] \leq \frac{\#SV}{M} \quad , \tag{4.15}$$

where $\#SV$ is the number of SVs. This gives an easy way to estimate the generalisation ability of the classifier with a bound that is independent on the dimensionality of the input space.

The optimal Lagrange multipliers $\boldsymbol{\alpha}^*$ can also be used for outlier detection in the training data set. The larger is a specific element α_i^* in the multiplier vector, the more difficult a sample $\mathbf{x}_i$ has been concerning the classification. If α_i^* is considerable large compared to other multipliers, $\mathbf{x}_i$ should be removed from the data set.

Solving (4.14) for $\boldsymbol{\alpha}$, the SVs are obtained for classes I and II. The optimal separating hyperplane situates at equal distance from the SVs for classes I and II, and b^* is given by:

$$b^* = -\frac{1}{2} \sum_{\text{Support vectors}} y_i \alpha_i^* (\mathbf{x}^{s1\mathrm{T}} \mathbf{x}_i + \mathbf{x}^{s2\mathrm{T}} \mathbf{x}_i) \quad , \tag{4.16}$$

where $\mathbf{x}^{s1} = (x_1^{s1}, x_2^{s1}, ..., x_n^{s1})^{\mathrm{T}}$ and $\mathbf{x}^{s2} = (x_1^{s2}, x_2^{s2}, ..., x_n^{s2})^{\mathrm{T}}$ are arbitrary SVs for Class I and Class II, respectively. Notice that only the terms associated with the SVs are summed, because the elements of optimal Lagrange multiplier $\boldsymbol{\alpha}^*$ corresponding to other samples are equal to zero.

So far it has been assumed that the training data is linearly separable. In the case, where the training data cannot be linearly separated, non-negative slack variables ξ_i are introduced to inequality conditions in (4.10), and the sum of the slack variables multiplied by the parameter C is added to the objective function given in (4.10). This corresponds to adding the upper bound C to the elements of $\boldsymbol{\alpha}$. The optimization problem will be:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M} \xi_i \\ &\text{subject to } \mathrm{y_i}(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad . \end{aligned} \tag{4.17}$$

In both separable and nonseparable cases, the form of the decision function is the same and given by:

$$f(\mathbf{x}) = \sum_{\text{Support vectors}} \alpha_i^* y_i \mathbf{x}_i^{\mathrm{T}} \mathbf{x} + b^* \quad . \tag{4.18}$$

Then unknown data example $\mathbf{x}$ is classified as follows:

$$\mathbf{x} \in \begin{cases} \text{Class I,} & \text{if } f(\mathbf{x}) > 0 \\ \text{Class II,} & \text{otherwise} \quad . \end{cases} \tag{4.19}$$

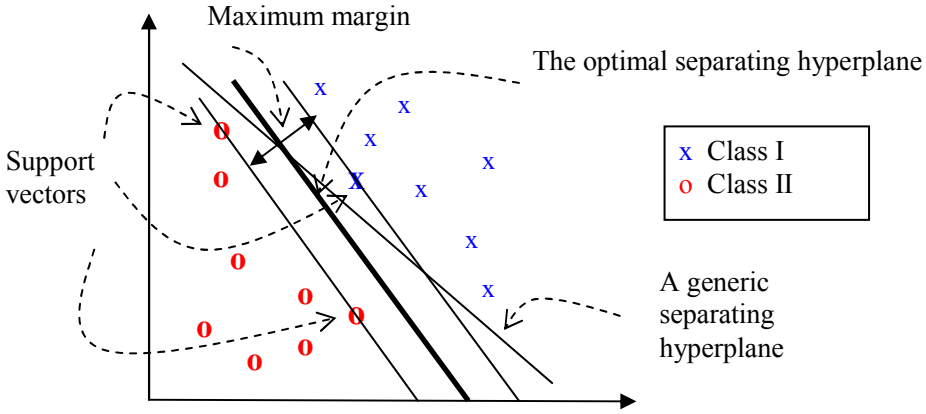


Figure 4.2. The optimal separating hyperplane maximizes generalisation ability of the classifier.

4.1.4 Nonlinear classification

In addition to linear classification, SVM can be applied to non-linear classification problems. When applying SVM in nonlinear problems, nonlinear mapping is used to generate the classification features from the original data. The nonlinearly separable data to be classified is mapped onto a high-dimensional feature space, where the data can be linearly classified; see Fig. 4.3. [Cristianini00].

Using a non-linear vector function $\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}),\ldots, \varphi_l(\mathbf{x}))^{\mathrm{T}}$ ($l >> n$) to map the n-dimensional input vector $\mathbf{x}$ into the l-dimensional feature space, the linear decision function in dual form is given by (compare Eq.(4.18)):

$$f(\mathbf{x}) = \sum_{\text{Support vectors}} \alpha_i^* y_i \boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}) + b^* . \qquad (4.20)$$

Working in the high-dimensional feature space enables the expression of complex functions, but it also generates problems. Computational problems occur due to large vectors and the danger of overfitting also exists due to high dimensionality. The latter problem is solved above with application of the maximal margin classifier, and so-called kernels give solution to the first problem. Notice that in (4.20) as well in the optimisation problem (4.14), the data occur only in inner products. A function that returns a dot product of the feature space mappings of original data points is called a kernel, $K(\mathbf{x},\mathbf{z}) = \boldsymbol{\varphi}(\mathbf{x})^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{z})$. When applying a kernel function, the learning in the feature space does not require explicit evaluation of $\boldsymbol{\varphi}$. Using a kernel function, the decision function will be:

$$f(\mathbf{x}) = \sum_{\text{Support vectors}} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) \quad , \qquad (4.21)$$

and the unknown data example is classified as before. The values of $K(\mathrm{x}_i,\mathrm{x}_j)$ over all training samples, $i, j = 1, \ldots, M$, form the kernel matrix, which is a central structure in the kernel theory. Mercer's theorem states that any symmetric positive definite matrix can be regarded as a kernel matrix that is an inner product matrix in some space [Cristianini00]. For example, polynomial learning machines of degree q have inner product kernel $K(\mathbf{x},\mathbf{z}) = \left(\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1\right)^q$, and RBF machines

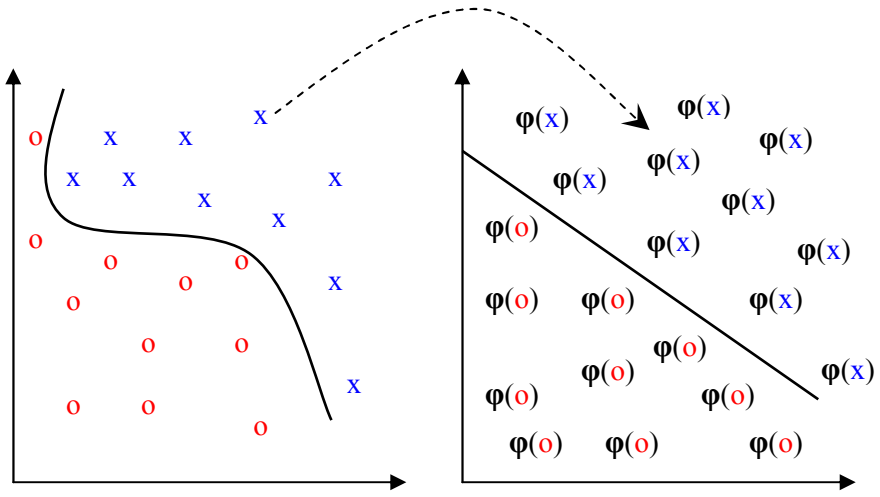


Figure 4.3. Mapping the input space to the feature space, where linear classification is possible.

have the inner product kernel $K(\mathbf{x},\mathbf{z})=\exp\left\{-\frac{|\mathbf{x}-\mathbf{z}|^2}{2\sigma^2}\right\}$, where σ defines the width. When applying the first order polynomial kernel an ordinary linear classifier is achieved. Application of RBF kernel results in a similar classifier structure with RBF networks, but most of the tuning is carried out during the training in contrast to classical RBF methods which require heuristic tuning e.g. in the determination of the number of RBF centers in the hidden layer.

4.2 Design and tuning

Design of SVM for a classification task consists of two tasks: choosing the kernel function and setting a value for the parameter C [Chin98]. The parameter C is also called an error penalty, because it deals with the trade-off between maximum margin and the classification error during training. A high error penalty will force the SVM training to avoid classification errors. It is clear that with high error penalty, the optimiser gives a boundary that classifies all the training points correctly. This, however, can give very irregular boundaries that may not lead good performance of the classifier in the test set. The selection of kernel function has also influence on the decision boundary. Usually RBFs are favored instead of polynomial kernel functions, because they are not sensitive to outliers and do not require inputs to have equal variances. However, in some cases polynomial kernels result in an excellent classification performance. In addition to the choice of the kernel function, various tuning parameters of the kernel should be chosen. When using polynomial kernel function, the order of the polynomial needs to be chosen, and when using RBF the spread, σ, needs to be decided. A large σ-value will give a smooth decision surface and regular decision boundary.

In Fig. 4.4 a) – e), the performance of various kernel functions is shown in a same classification problem. A two-dimensional problem is studied to be able to graphically depict the different cases. The considered noisy data set is generated artificially. In the figures, points represent decision regions and training data is plotted with circles and crosses. One can see that the more complex kernel function is the more irregular decision boundary results from training. The number of support vectors is an important measure of generalisation when choosing the kernel. The less SVs the better generalisation. In this case, a second order polynomial results in the smallest SV

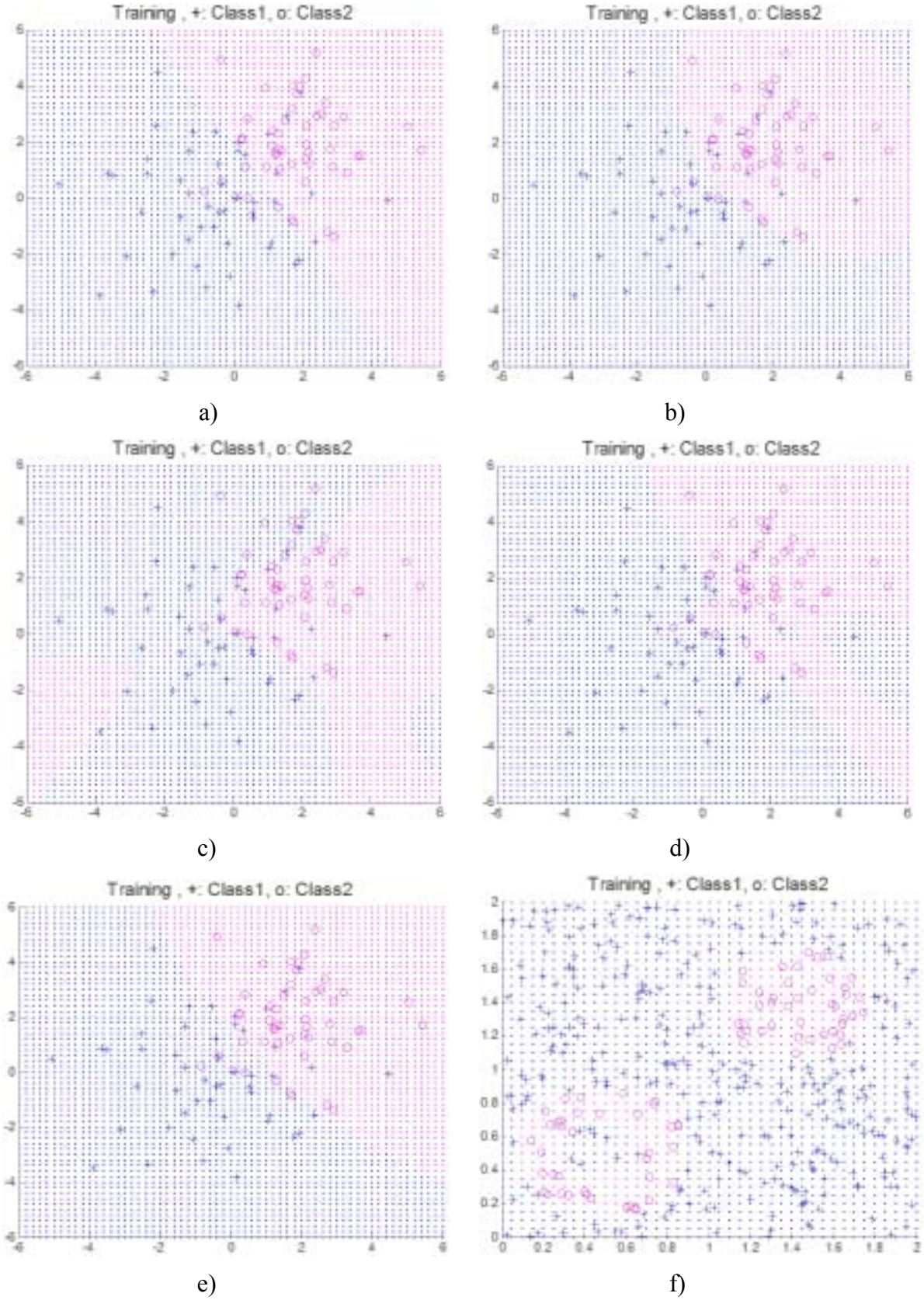


Figure 4.4. The choice of the kernel function has a strong impact on the complexity of the decision boundary. Solutions for a two-dimensional classification problem with various kernels: a) first order polynomial, b) second order polynomial, c) fourth order polynomial, d) RBF with σ = 0.5, e) RBF with σ = 5 and f) RBF with σ = 0.1 in another classification problem. In all cases, C=0.5.

percentage (44%). A sample was considered to be a SV if the corresponding element of $\boldsymbol{\alpha}^*$ was smaller than 10^{-6}. Even the smallest number of SVs is large, because the classes are not easily separated due to noise. However, also kernel function corresponding the first order polynomial results in only 46% SVs of all data. The more complex kernels result in higher percentage of SVs (78% in c), 92% in d), and 62% in e)), and when looking at the decision boundaries, especially in c) and d), one can also clearly see that they would not perform well in the presence of noise. However, RBF kernel with large value for σ, seems to result in a quite regular boundary, although it is very close to the one achieved with the simple first order kernel function. Based on this analysis, the kernel corresponding to the first order polynomial should be chosen for the kernel function.

In Fig. 4.4 f), a solution for another classification problem is shown. The problem is more complex than the first one, because the classes to be separated are merged into each other. However, with RBF kernel function with $\sigma = 0.1$ and $C = 1$, an adequately performing decision boundary can be found (SV percentage 53%). This example shows that with proper tuning SVM can solve very difficult classification problems, but one should always pay attention to real complexity of the problem, before applying very complicated kernel functions.

Choosing a kernel function is an application dependent task, and thorough guidelines do not exist. Usually the kernel function is chosen based on trial and error or a cross validation method. If there is any a priori knowledge on the problem under study that should be taken account when tuning the SVM based classifier. Selection of the kernel function is similar task with the selection of the network architecture when applying NNs.

4.3 Least squares support vector machine (LS-SVM)

One of the drawbacks of SVMs is tedious computation in the training phase due to the quadratic optimization problem. Suykens & al. have reformulated the standard SVMs to avoid this problem and developed Least Squares Support Vector Machines (LS-SVM) [Suykens02]. The cost function is a regularized least squares function with equality constraints leading to linear Karush-Kuhn-Tucker systems. The solution can be found efficiently by iterative methods such as conjugate gradient algorithm. LS-SVMs do not lead sparse solutions such as SVMs but a solution for the optimization problem is found very fast, and pruning techniques can be easily applied to enhance the sparsity.

LS-SVM algorithm is derived in a following way [Suykens02]. The classification problem can be solved from optimization problem with equality constraints:

$$\begin{aligned} &\min_{w,b,e} J(\mathbf{w},b,e) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma\frac{1}{2}\sum_{i=1}^{M} e_i^2 \\ &\text{subject to } y_i[\mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i)+b] = 1-e_i,\ i=1,...,M \end{aligned} \quad . \tag{4.22}$$

Denotation corresponds to Chapter 4.1, but e_i denotes the error in the classification of the sample $\mathbf{x}_i$. One defines the Lagrangian:

$$L(\mathbf{w},b,\mathbf{e};\boldsymbol{\alpha}) = J(\mathbf{w},b,\mathbf{e}) - \sum_{i=1}^{M}\boldsymbol{\alpha}_i\left\{y_i\left[\mathbf{w}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_i)+b\right]-1+e_i\right\} \quad , \tag{4.23}$$

where α_i are Lagrange multipliers, which can be either positive or negative due to equality constraints. The conditions for optimality:

$$\begin{aligned}
&\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^{M} \alpha_i y_i \boldsymbol{\varphi}(\mathbf{x}_i) \\
&\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{M} \alpha_i y_i = 0 \\
&\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad i = 1,...,M \\
&\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i[w^T \boldsymbol{\varphi}(\mathbf{x}_i) + b] - 1 + e_i = 0, \quad i = 1,...,M
\end{aligned} \tag{4.24}$$

can be written as the solution to the following set of linear equations:

$$\begin{bmatrix} \mathbf{I} & 0 & 0 & -\mathbf{Z}^{\mathrm{T}} \\ 0 & 0 & 0 & -\mathbf{Y}^{\mathrm{T}} \\ 0 & 0 & \gamma \mathrm{I} & -\mathbf{I} \\ \mathbf{Z} & \mathbf{Y} & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ b \\ \mathbf{e} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{\mathbf{1}} \end{bmatrix} \quad , \tag{4.25}$$

where $\mathbf{Z} = (\boldsymbol{\varphi}(\mathbf{x}_1)^{\mathrm{T}} y_1, \ldots, \boldsymbol{\varphi}(\mathbf{x}_M)^{\mathrm{T}} y_M)^{\mathrm{T}}$, $\mathbf{Y} = (y_1, \ldots, y_M)^{\mathrm{T}}$, $\vec{\mathbf{1}} = (1, ..., 1)^{\mathrm{T}}$, $\mathbf{e} = (e_1, \ldots, e_M)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^{\mathrm{T}}$. The solution is also given by:

$$\begin{bmatrix} 0 & -\mathbf{Y}^{\mathrm{T}} \\ \mathbf{Y} & \mathbf{Z}\mathbf{Z}^{\mathrm{T}} + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{\mathbf{1}} \end{bmatrix} \quad . \tag{4.26}$$

Mercer's theoremcan be applied again to matrix $\boldsymbol{\Omega} = \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ where:

$$\Omega_{i,l} = y_i y_l \boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_l) = y_i y_l K(\mathbf{x}_i, \mathbf{x}_l) \quad . \tag{4.27}$$

Hence the classifier (4.22) is found by solving the linear set of equations (4.26)-(4.27) instead of quadratic programming. The support values α_i are proportional to the errors at the data points, while in standard SVM most values are equal to zero. This is a drawback of LS-SVM, but application of pruning techniques is relatively easy because of the globally optimal solution and very short computation time.

4.4 Multi-class classification

4.4.1 Coupling schemes

SVMs are essentially binary classifiers. They are designed to separate only two classes from each other. However, in most of the real applications, multi-class classification is required. For example, in the fault classification of an induction motor, there exist several fault classes in addition to healthy operation.

A solution is to decompose a multi-class problem to several 2-class problems, train classifiers to solve these problems, and then couple the classifiers to reconstruct the solution of the multi-class problem from outputs of the individual classifiers. One of the simplest multi-class classification structures is the so-called one-against-others approach. In this method, K classifiers are built in the way that each classifier separates one class from all the others. However, in many applications, this

approach has been found to be inferior to a pairwise coupling approach, where $\frac{1}{2}K(K-1)$ 2-class classifiers are built, each separating one class from another ignoring all the other classes. Outputs of the pairwise classifiers are then fused to find the global solution to the *K*-class problem. In this approach, more 2-class classifiers are needed than in the former case, but using it, the total classification performance can usually be highly improved. For example, in [Suykens02], various multi-class extensions of SVM are compared in several classification problems.

There exist numerous schemes to reconstruct the final classification solution from the outputs of pairwise classifiers' solutions. The simplest methods are based on majority voting [Friedman96]. Pairwise classifiers give votes for classes and the class that gets most of the votes is selected to be a final class decision for a sample considered. If rough reconstruction is used, the classifiers can only give binary votes (-1 or 1), but in soft reconstruction the exact outputs of the classifiers are considered as votes. The higher is the exact output of a SVM based classifier, the more likely the specific sample to be classified belongs to the positive class, and correspondingly, the smaller is the output, the more likely the sample belongs to the negative class. If the output is close to zero, the classification decision is unreliable.

An important problem occurs when applying majority voting. For a given sample **x**, the voting scheme weights equally the outputs of all pairwise classifiers, without considering their significance. Of course, the relevant classifiers concerning the success of the classification are not known in advance. However, redundancy of some pairwise classifiers may be considered with a so called mixture matrix. With this approach, the outputs of classifiers are linearly combined with the mixture matrix created, for example, with least squares estimation, to minimize the error between the correct class decision and the linear combination of the pairwise classifiers' outputs. The mixture matrix approach is proposed in [Mayoraz99], but it has been considered there in scaling the outputs of one-against-others type of classifiers. An interesting question also is, whether a nonlinear coupling – e.g. with a NN – can improve the performance of the classification structure.

Other coupling schemes suggested in the literature are, for example, binary trees [Schwenker00] and a fuzzy logic based method [Inoue01]. When applying binary trees, a proper hierarchy of classifiers should be known before training the classifiers. This requires a priori knowledge of the solution of the classification problem or implementation of sophisticated clustering or vector quantisation algorithms. When using the fuzzy logic approach, choosing and tuning of the membership functions is an application dependent task, and may sometimes be quite time-consuming.

4.4.2 Comparison of the coupling schemes

In the publication [P3], the performance of four coupling schemes is tested in the fault classification of a 35 kW induction motor: majority voting with soft and rough reconstruction, mixture matrix, and a NN. The results are summarized here. The classification of faults is based on spectral information of circulating currents in parallel branches of the motor. Virtual measurement data is used, and the simulations are based on time-stepping, finite-element analysis. The studied machine is an inverter-fed 35 kW squirrel cage induction motor with a star connection of the stator winding. The main frequency of the input voltage is 100 Hz. Three load conditions are studied: no load, half load and full load. The circuit equations of the stator and rotor windings are modified to implement the faults (Table 4.1).

Table 4.1 Division of faults to classes

Class 1	Healthy machine (NF)
Class 2	Broken rotor bar (BB)
Class 3	Broken end-ring in rotor cage (BR)
Class 4	Shorted coil in stator winding (SC)
Class 5	Shorted turn in stator winding (ST)
Class 6	Static eccentricity in rotor (SE)
Class 7	Dynamic eccentricity in rotor (DE)

Welch's method [Welch67] is used to calculate the PSD estimates of circulating currents of the induction motor. Hanning window is used with 500 samples window, and number of overlapping samples is 250. Before applying PSD estimation, measurement noise is imitated by adding normally distributed noise to the current. Mean value of noise is equal to zero and its standard deviation is 5% of the amplitude of current. PSD estimates were calculated 32 times from different parts of a circulating current signal in each fault case. Thus, a spectrum sample set with 224 samples is achieved for three different load cases. Half of the samples were chosen for training the classifier, and half of the samples were left for testing the generalisation ability of the classifier. An average healthy spectrum from the training set was chosen to be a reference, and all the other spectra were scaled with it. The difference values from the reference created the sample set.

All pairwise SVM classifiers were designed with the same simple inner product kernel function corresponding to a first order polynomial in the feature space. Practically, similar results were achieved with a kernel function corresponding a properly parametrized radial basis function and higher order polynomials, whose degree was an odd number. The error penalty C was equal to 1.

As a nonlinear approach a NN is studied. A regular feed-forward MLP network is trained to minimize the error between the correct class decision and the non-linear combination of outputs of pairwise classifiers. The network is chosen to have one hidden layer and log-sigmoid transfer functions. The log-sigmoid transfer function is chosen because its output range (from zero to one) suits to output Boolean values. The number of neurons in a hidden layer has a strong impact on the performance of the network. With a large number of hidden neurons, it is possible to achieve excellent performance in the training set, but this does not necessarily lead to good generalisation ability and high accuracy in the evaluation set. After numerous tests, the number of hidden neurons that gave the best overall accuracy was chosen. For no load classification structure the number was 12, for half load classification structure 18 and for full load classification structure 8.

The classification results with these structures are presented in Fig. 4.5.a). These are the best results achieved, but with arbitrarily chosen number of hidden neurons, the results could be much worse. In Fig. 4.5.b), correct classification percentages are presented, when majority voting with rough reconstruction is used. In Fig. 4.5.c), classification results are shown, when soft reconstruction is used. In Fig. 4.5.d), the results with linear coupling are displayed. The best total classification results over all load situations and over all classes are gained with NN coupling (95.8%) and the worst results with majority voting with rough reconstruction (87.6%). Majority voting with soft reconstruction (93.9%), mixture matrix coupling (95.5%) and NN coupling were highly competitive compared to it. The difference between the coupling methods can especially be seen in detection of SE faults. With majority voting and rough reconstruction, SE fault samples tend to get equal amount of votes with ST fault class. Thus, the total accuracy of the classifier degrades. In no load situation, problems occur also in detection of other faults in addition to SE, and these problems cannot be totally solved by changing the coupling scheme. It is obvious that faults are more easily detected from a motor that is working under load. If differences in a faulty spectrum sample compared to the average healthy sample are hidden due to noise, classification does not succeed.

The NN approach and the mixture matrix approach resulted in a classification structure with almost the same accuracy. When comparing the results, it should be taken into account that with arbitrarily chosen number of hidden neurons, the NN could result in much worse classification results. Also, training and tuning the NN is an exhausting task. A linear combination of SVMs with a mixture matrix already gives excellent results, so using nonlinear reconstruction is not necessary in this application.

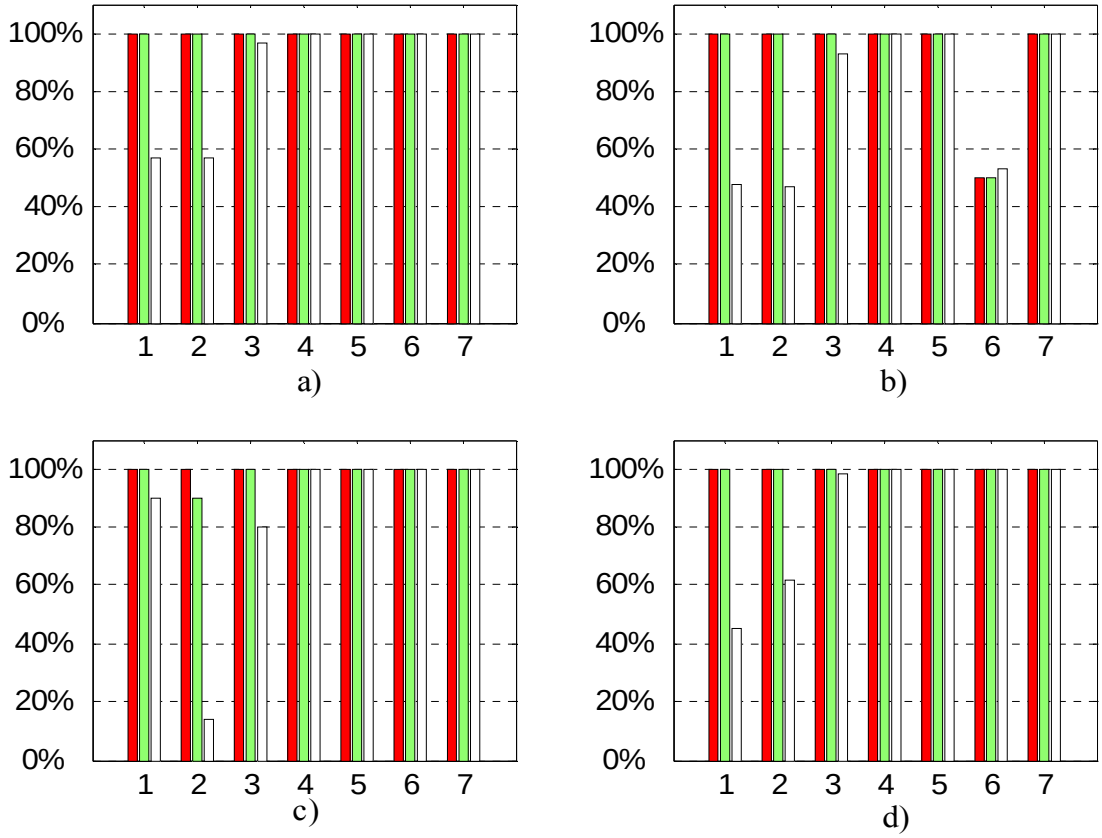


Figure 4.5. Correct classification percentages in a testing set with a) neural network coupling, b) majority voting and rough reconstruction c) majority voting and soft reconstruction and with d) mixture matrix coupling. Each group of bars represents accuracies Q_i for each class, and in each group, the first bar represents full load situation, the second bar half load situation and the third bar no load situation.

5. Motor current signature analysis of a 15 kW induction motor

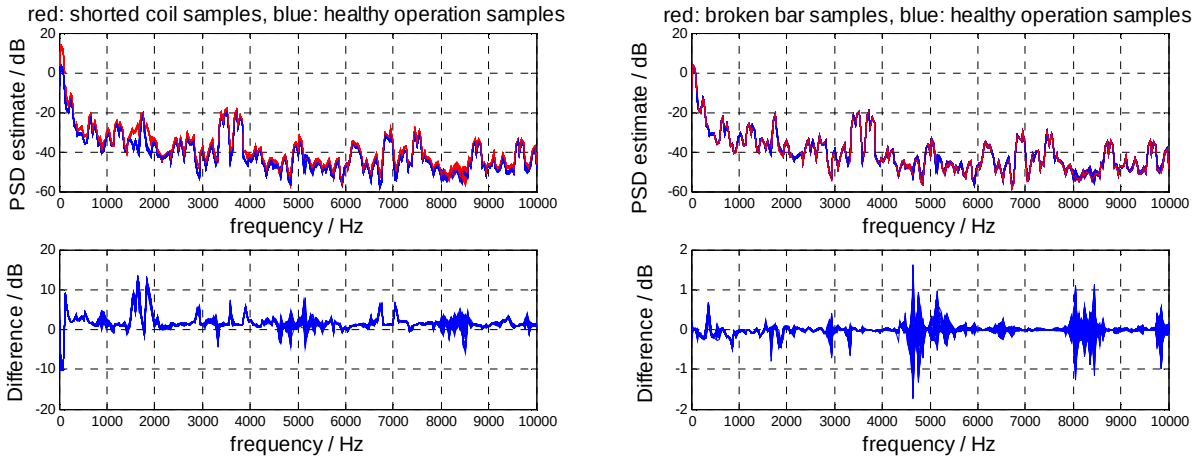
In this chapter, the results from the publications [P1] and [P2] are summarized. MCSA of a 15 kW induction motor is carried out with PSD estimation and SVM based classification. The machine studied is an inverter-fed four-pole cage induction motor. The stator winding is delta connected, and the coils of a stator phase are all connected in series (no parallel branches). Three different load conditions are considered: no load, half load, and full load. The method is tested with virtual measurement data retrieved from finite-element analysis of the motor [Arkkio90]. In the finite-element analysis, the circuit equations of the stator and rotor windings are modified to implement the faults (Table 5.1). A shorted coil in stator winding is obtained by extracting a coil from the series connected coils forming a healthy phase and forcing the voltage of the extracted coil to be zero. A shorted turn is constructed in a similar manner. A broken bar or end-ring in the rotor cage is obtained by adding a large resistance properly in the circuit equations of the rotor.

Table 5.1. The motor states studied

NF	No fault
BB	Broken rotor bar
BR	Broken end-ring in rotor cage
SC	Shorted coil in stator winding
ST	Shorted turn in stator winding

PSD estimates of the stator currents have often been used as a medium of fault detection of induction motors [Benbouzid00]. Main disadvantage of classical spectral estimation techniques, such as FFT, is the impact of side lobe leakage due to windowing of finite data sets. Window weighting decreases the effect of side lobes. Further, in order to improve statistical stability of the spectral estimate, averaging by segmenting the data can be applied. The more segments are used the more stable the estimate is. However, the signal length limits the number of segments used, but with overlapping segments the number of segments can be increased. In this thesis, Welch's method is used to calculate the PSD estimates of a stator current of the induction motor. The method applies both the window weighting and the averaging over overlapping segments to estimate the PSD. In this study, Hanning window sized 500 samples is used, and number of overlapping samples is 250. The length of resulting PSD estimate vector is 257 and the whole vector is considered as an entity to be classified.

The PSD estimates are calculated 80 times from different parts of a stator current signal in each fault case. Thus, we got a spectra sample set with 400 samples for three different load cases. SVMs were trained and tested separately in different load situations. The first 200 samples form the training data set and the rest are left for testing the generalisation ability of the classifier. An average healthy spectrum from the training set is chosen to be a reference, and all the other spectra are scaled with it. The difference values from the reference create the sample set. As an example, the sample set of shorted coil fault is plotted in no load situation in Fig. 5.1.a). In the first picture, PSD estimates of a stator current in healthy operation and in shorted coil operation are plotted. In the second picture, the difference between the average healthy spectrum and shorted coil spectra are plotted.



a) b)

Figure 5.1. Sample set of Welch's PSD estimates of a stator current in a) healthy and in shorted coil operation and b) in healthy and in broken rotor bar operation.

Of all the faults, the shorted coil has the most distinct impact on the stator current. For comparison, a sample set of PSD estimates of a stator current in broken rotor ring operation is shown in Fig. 5.1.b). The differences between PSD estimates in broken rotor bar operation and in average healthy operation are less than 10% from the differences between estimates in shorted coil operation and in average healthy operation. Also, the peaks of differences do not always occur in same frequencies, which may make the classification difficult.

When applying pairwise coupling scheme, ten 2-class classifiers need to be designed. All the classifiers are designed with a kernel function corresponding to a first order polynomial. Almost similar results are achieved with a radial basis function and with higher order polynomials, whose degree is an odd number. The upper bound C for the Lagrange multipliers is chosen to be 10.

In Table 5.2, the correct classification percentages between different classes in the test set are presented in full load situation. Shorted turn and shorted coil in stator windings are always distinguished from other faults and healthy situation, but there are difficulties in distinguishing broken rotor bar from broken end ring and the healthy operation.

In Table 5.3, the percentages of SVs in the training set are presented in different classification cases in full load operation. A sample is chosen to be a SV, if the corresponding Lagrange multiplier is larger than 10^{-3}. The shorted coil operation has been easy to distinguish from other fault classes: the numbers of SVs in these classification cases are relatively small.

Table 5.2. Correct classification percentages in a testing set and in full load operation

%	BB	BR	ST	SC
NF	63	100	100	100
BB	-	94	100	100
BR	-	-	100	100
ST	-	-	-	100

Table 5.3. Percentages of support vectors of the training set in full load operation

%	BB	BR	ST	SC
NF	85	68	50	25
BB	-	63	41	16
BR	-	-	28	16
ST	-	-	-	26

Training of the pairwise classifiers also shows the appropriateness of SVM in high dimensional classification problems. The number of training samples is only 80 in each classification case, whereas the whole PSD estimate creates the feature set. However, one can see from the SV percentages that especially the cases, where shorted coil fault is involved, the training of the classifier is highly successful. The correct classification rates in the test set are congruent with the SV percentages.

After training the 10 pair-wise classifiers for each load situation (30 classifiers in total), their co-operation is tested. The majority voting approach is used. 200 samples (40 representatives from each fault case and from the healthy case) are tested. In Table 5.4, the correct classification percentages of each class are presented in each load situation.

Table 5.4. Correct classification percentages in a testing set with combined classifiers

%	NF	BB	BR	ST	SC	Total
Full Load	50	65	98	100	100	83
Half Load	83	100	50	100	100	87
No Load	100	100	100	100	100	100

In reconstruction the final class decision, faults in pairwise classifiers performance are cumulated. In no load operation, results are excellent, but on other load levels, some fault classes cannot be perfectly separated from the healthy class and from each other. However, in total, over 80 % of all samples are correctly classified in all load situations.

Next random measurement noise is added to the stator current before training the classifiers. The standard deviation of normally distributed measurement error is assumed to be 1% of the amplitude of line current and its mean value is set to zero. In Table 5.5, the classification results in the presence of noise in all load situations are showed. Classification results degrade considerably. Only shorted coil fault is always correctly classified regardless of noise.

Table 5.5. Correct classification percentages in a testing set with combined classifiers, while noise is present.

%	NF	BB	BR	ST	SC	Total
Full Load	33	20	35	28	100	43
Half Load	38	40	38	30	100	49
No Load	13	45	53	7.5	100	44

Dropping such pairwise classifiers whose performance is not adequate out of the classification structure, the detection rate of certain faults can be increased also in noisy situation (see Table 5.6).

Table 5.6. Correct classification percentages in a testing set with 3-class classification structure

%	NF	ST	SC	Total
Full Load	73	88	100	87
Half Load	83	98	100	93
No Load	95	38	100	78

This indicates that the malfunction of some pairwise classifiers should be able to be taken into account in the reconstruction of the final classification solution. Mixture matrix coupling is a better choice for the reconstruction instead of majority voting.

6. Comparison of motor variables as fault indicators

The results of the previous chapter indicate that the stator current is not a good indicator of all faults. In the publication [P4], three motor variables are compared as the medium of induction motor fault detection: the stator line current, the circulating currents in parallel branches and the force on the rotor. The results are summarized in Tables 6.1-6.3. In addition to previous fault cases, also static eccentricity (SE) and dynamic eccentricity (DE) of the rotor are studied. The results are based on the performance of a 35 kW induction motor, which is simulated with finite element analysis. The standard deviation of normally distributed measurement error is assumed to be 3% of the amplitude of the signal and its mean value is set to zero. A simple predictive noise filtering is also applied. The training and testing data sets are generated in a similar way as before, and various SVM tunings are considered to find the best classifier. All classifiers are designed with a radial basis kernel function width equal to 11 except when studying forces as indicators of faults. In those cases, the first order polynomial kernel function is used.

Table 6.1. Correct classification percentages in a testing set, stator line current as a medium of fault detection

%	NF	BB	BR	ST	SC	SE	DE	Total
Full Load	20	45	23	100	100	5.0	7.5	43
Half Load	28	18	20	100	100	38	38	49
No Load	70	85	70	100	100	50	68	78
Total	39	49	38	100	100	31	38	**57**

Table 6.2. Correct classification percentages in a testing set, circulating currents between parallel branches as media of fault detection

%	NF	BB	BR	ST	SC	SE	DE	Total
Full Load	100	100	100	100	100	100	100	100
Half Load	100	100	100	100	100	100	100	100
No Load	83	75	100	100	100	100	100	94
Total	94	92	100	100	100	100	100	**98**

Table 6.3. Correct classification percentages in a testing set, force on the rotor as a medium of fault detection

%	NF	BB	BR	ST	SC	SE	DE	Total
Full Load	100	100	100	100	100	100	100	100
Half Load	100	100	100	100	100	100	100	100
No Load	100	100	100	100	100	100	100	100
Total	100	100	100	100	100	100	100	**100**

The results indicate that the circulating currents and the force on the rotor are superior media for fault detection compared to the stator line current. The application of stator current is widely studied only because its monitoring does not require extra instrumentation and it is non-invasive method.

The force on the rotor is the best fault indicator. The measurement of forces is difficult, but they are directly related to vibrations that are measurable and used in condition monitoring of rotating machines for decades. Because of these results, the vibrations are studied more thoroughly in fault diagnostics of a 35 kW cage induction motor.

7. Broken rotor bar detection of a 35 kW induction motor with vibration monitoring

In this chapter, results of the publications [P5] and [P6] are summarized. Firstly, signal processing of motor vibrations is studied for detection of broken rotor bar in a 35 kW induction motor. Then, information fusion of multi-channel vibration measurements is considered to improve the broken rotor bar detection.

7.1 Signal processing of vibrations

In this chapter, signal processing of vibrations is studied for revealing broken rotor bar faults in an induction motor. Real measurement data from an artificially damaged 35 kW cage induction motor is used. The motor is fed from a Vacon inverter, and a DC generator is the motor load. The switching frequency of the inverter is fixed at 3 kHz. The signal given by the vibration sensor mounted on the back part of the motor is amplified through charge amplifiers Bruel & Kjaer 2635. The amplified signals are the transient recorder inputs. The recorded transient is calibrated using a true root mean square voltmeter connected in the amplifier output. This calibration is made in such a way that the recorded measurements in the transient recorder are in acceleration units.

Measurements are carried out with motor in healthy condition and with motor under three rotor fault situations: one broken rotor bar, two broken rotor bars and three broken rotor bars plus end ring broken. In addition to these, one broken rotor bar operation is also measured with external interference present. The sampling frequency is 40 kHz and the number of samples in each data set is 20 thousands. Three load situations are considered: no load, half load and full load. The feature samples are calculated 30 times from each motor condition and each load situation – in total 450 samples. Half of the samples from healthy situation and broken rotor bar situation are used in training the classifier (90 samples) and half are left for testing the classifier's generalization ability in addition to the all samples of more serious fault situations (360 samples). The latter samples are important to be included to the test set, because success of the classification with these samples implies, whether the selected features are able to indicate a broken rotor bar fault despite of the interference due to variations in measurement installation and variations in the nature of the rotor bar fault.

Four feature extraction techniques are applied to various sections of the vibration signals. In Fig. 7.1.a), examples of real cepstra in healthy motor operation and in broken rotor bar operation are presented. In the figure, also the difference of the cepstra is presented, and there is a distinct difference between the signals. In Fig. 7.1.b), the 13^{th} order AR model coefficients of the vibration signals are presented in the same motor conditions. The biggest difference can be seen in the 3^{rd} and 4^{th} coefficients. Such a high model order was chosen, because it resulted in the best classification performance. When applying HOS analysis, the enhanced PSD, bispectrum and trispectrum measures are calculated at the critical frequencies, and their concatenation is the feature vector to be classified.

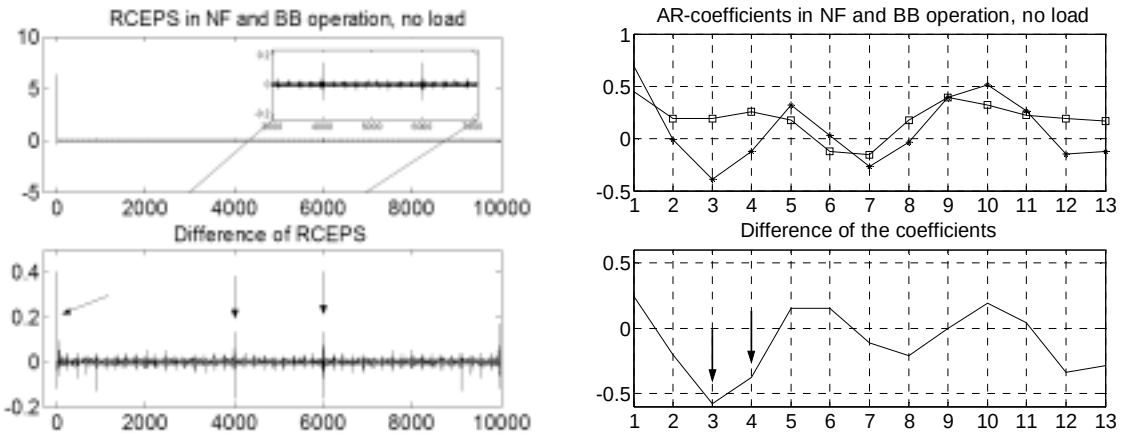
In SVM classification, radial basis kernel functions with various widths from 0.05 to 10 were considered in addition to a kernel function corresponding to the first order polynomial. Error penalty C was varied from 0.1 to 10. A proper SVM tuning varied depending on the features under study. The best classification results (100%) were achieved with application of AR coefficients, a

kernel function corresponding to the first order polynomial, and $C = 2$. The second best results were achieved with application of Welch's PSD estimates or cepstrum analysis (81.8%). With the latter approaches, the results degrade, because the three broken rotor bar and broken end ring fault was mostly not detected. In this study, the enhanced HOS analysis was not a competitive feature generation approach (14.2%). This may be due to inaccurate estimation of the critical frequencies.

The number of SVs also influence on the generalization ability of the classifier. The generalization is the better the less SVs there are. When AR signal description was applied the SV percentage of all training samples was only 11.1%, when Welch's PSD estimates were applied the percentage was 82.7%, and when cepstrum analysis was applied 74.4%. It is possible that the results gained with the cepstrum analysis and Welch's PSD estimates might be improved, if feature selection algorithms were applied before classification.

In [P5], the order of AR-model was chosen to be 13, because that seemed to result in the best classification performance, when detecting the broken rotor bar. Also some lower degrees resulted in the 100% correct classification rate, but with lower degrees the number of SVs increased. However, after publication of [P5] a robustness analysis was carried out in the presence of noise, and it showed that coefficients of the third order model are more robust features for the broken rotor bar detection than coefficients of the 13^{th} order model. When applying the 13^{th} order model coefficients, the correct classification rate started to decrease, if noise was included to vibrations measurements with the variance 2% of the maximum value of vibration signal and the mean of zero. When applying the coefficients of the third order model, the correct classification rate did not decrease until the variance of noise was over 12% of the maximum value of original vibration. Although the vibration measurements did not originally contain much noise, they will very likely vary between measurement sessions, and, thus, the more robust model is a practical choice for feature extraction.

It should be noticed that the fault detection rate of 100% with the features generated by AR modelling is an excellent result, because most of the test samples were totally independent from the training data set. The reinstallation of the measurement system always influences on the vibration signals. Also, some rotor bar fault types that were included to the test set were not shown to the classifier in its training phase. This indicates that the selected features are very likely able to show the rotor bar fault of induction motor even despite of more severe interference than noise. However, the study concentrated on the detection of rotor bar faults of an induction motor. For detection of other kinds of faults additional features of vibration signal may show to be relevant.



a) b)

*Figure 7.1. a) Real cepstra of vibration signals in healthy and broken rotor bar operation. b) AR-coefficients of vibration signals in healthy and broken rotor bar operation. (▫: healthy, *: broken rotor bar).*

7.2 Data fusion and interference removal with ICA

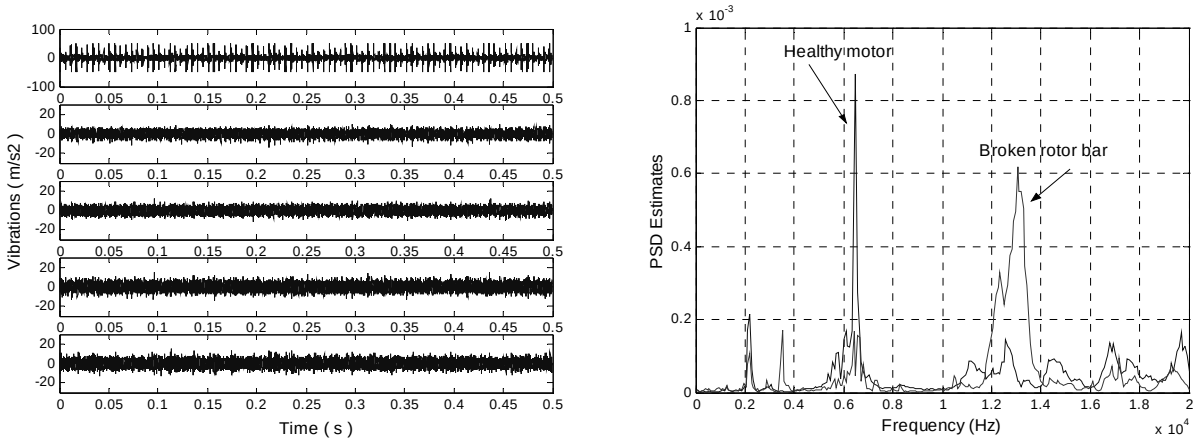
In the previous chapter, only one sensor was used to measure the motor vibrations. An interesting question is how the multiple sensors influence on the diagnostics process. In this chapter, vibrations are measured with five sensors and their fusion is used as the fault indicator. The chapter bases on the publication [P6].

The measurement set-up is similar with the one in the previous chapter, but now five acceleration vibration sensors are placed in different parts of the motor. Three of them are placed in the cooling surface of the motor frame. One is placed in the covering surface of the frontal bearing and the other one near the cooling fan in the back part of the motor. The last sensor placement equals the one used in the last chapter. As an example, the vibration measurements in a healthy situation with no load are presented in Fig. 7.2.a). One of the vibration measurements is considerably higher than measurements from the other sensors. This signal is measured with the sensor placed in the back part of the motor.

FastICA algorithm [Hyvärinen99] is used to calculate independent components (IC) of vibration measurements. A FastICA MATLAB-package developed at the Laboratory of Computer and Information Science in the Helsinki University of Technology was applied. Before applying FastICA, the vibration measurements are whitened with PCA. In Fig. 7.b), Welch's PSD estimates of the first IC in healthy and broken rotor bar situation are presented in the same picture (Hanning window sized 500 samples with 250 overlapping samples). There is a distinct difference between them.

LS-SVM based classifier is built to discriminate between the healthy and broken rotor bar condition. It is also tested with measurements from two broken bar situation and three broken bar and broken end ring situation. LS-SVM has a kernel function that corresponds to the first order polynomial and the upper bound for Lagrange multipliers was chosen to be equal to 10. Due to another classification algorithm the results presented in this chapter are not fully comparable with the results in Chapter 7.1.

Training data is formed by calculating Welch's PSD estimates from different parts of the first IC of vibrations in the healthy and broken rotor bar situations. In the earlier figures, only no load situations are plotted, but also two other load situations are taken into account in training: half load and full load. The training data consists of 2x3x15 = 90 samples so that from both motor conditions and from all load situations there exist 15 samples. The other faults than one broken rotor bar fault are used only in testing the classifier, so that in total there are 240 samples for testing: 45 from healthy situation, 45 from one broken rotor bar situation and 90 from two broken rotor bars and 60 from three broken rotor bars and broken end ring. In the last case, full load measurements were not available. All of the test samples are correctly classified.



a) b)

Figure 7.2. a) Vibration measurements (m/s^2) in a healthy situation with no load. The sensor placed in the back part of the motor measures considerable bigger vibration than the others. The main vibration signal is depicted in the first plot. b) Welch's PSD estimates of the first IC of vibrations are different in healthy situation and in broken rotor bar situation (no load).

For comparison, a classifier that uses the main vibration component as a fault indicator instead of the first IC of all vibrations was also trained. Training and testing data sets were formed in a same way as earlier. At this time 5% of test samples were wrongly classified. All of these samples were two broken rotor bar samples that were wrongly classified to be healthy.

Further, a classifier that uses the first principal component (PC) of vibrations as fault indicator was trained. Again, all the test samples were correctly classified. Even if the difference between using pure vibrations or ICs or PCs of vibrations in fault detection is quite small, this could indicate that faults are more easily detected from fused vibration measurements than pure vibrations. ICs and PCs contain more information on all vibration measurements than any of the individual vibration measurements. Using all vibration components for classification might improve the results without any data fusion, but at the same time computation would become heavier.

In this case, both data fusion methods, ICA and PCA, resulted in equally excellent performance of broken rotor bar detection. However, further studies are required with measurements from other faults to conclude overall usefulness of data fusion of vibration measurements, because the broken rotor bar was quite easily detected also based on the main vibration measurement. Also, building a multi-class classifier for detection of several faults may degrade the classification results.

8. Insulation defect localization with PD analysis and numerical classification

In this chapter, the results from the publication [P7] are summarized. The publication considers analysis of PD signals to localize defects in insulation systems. The task of automatic defect localization with respect to electrodes has a wide range of industrial applications. In fact, depending on the apparatus type, risk assessment is remarkably affected by defect location with respect to the electrodes. In this study, various parameters are first extracted from PD distributions, and statistical analysis is performed to select the most significant parameters concerning localization. Then, the localization process is carried out through numerical classification. Three different classification methods are compared to find the best approach for this application: a *k*-nearest neighbour classifier (*k*-NN), a probabilistic neural network (PNN) and a SVM based classifier.

Two types of samples were under study: artificial specimens and HV apparatus with artificial defects. The first set of samples is important because the defect characteristics (geometry, materials, etc.) are well known, therefore the target likelihood is large. The drawback of these samples is that they are poor representatives of real situations. The second set of specimens is important because it represents quite well real situations, although the likelihood of its targets is sometimes small. The latter set contained samples from HV and MV cables and rotating machine insulations systems. In total, 17 samples with defect close to the high voltage electrode (denoted “HV samples”) and 26 with defect close to the low voltage electrode (denoted “LV samples”) were under study. Measurements were performed with an ultra wide band digital instrument that allows automatic noise rejection and separation of simultaneous PD activities.

The majority of the parameters were extracted from amplitude, phase and time of occurrence histograms. In particular, parameters were first defined for a single polarity distribution, i.e. the set of discharges occurred when the electric field internal to the defect had a positive or negative sign. As an example, consider Fig. 8.1, which shows a schematic PRPD pattern (i.e. a graphical representation of PD density in the phase-amplitude plane): positive distribution is constituted by discharges occurred in the first half period of the applied voltage, and negative distribution is constituted by discharges occurred in the second half period. Afterwards, the parameters relevant to each polarity were combined and normalized, in order to account for the dissymmetry of the PD activity. In fact, the location of a defect in proximity of a metal surface produces, in general, a dissymmetry in the PD activity. In addition, a new parameter from the intertime distribution was considered. The parameter is related to the dissymmetry of PD activity in terms of intertimes. It takes into account both PD polarity, which is determined by the phase of the pulses, and the pulses chronological order, which determines the intertime distribution.

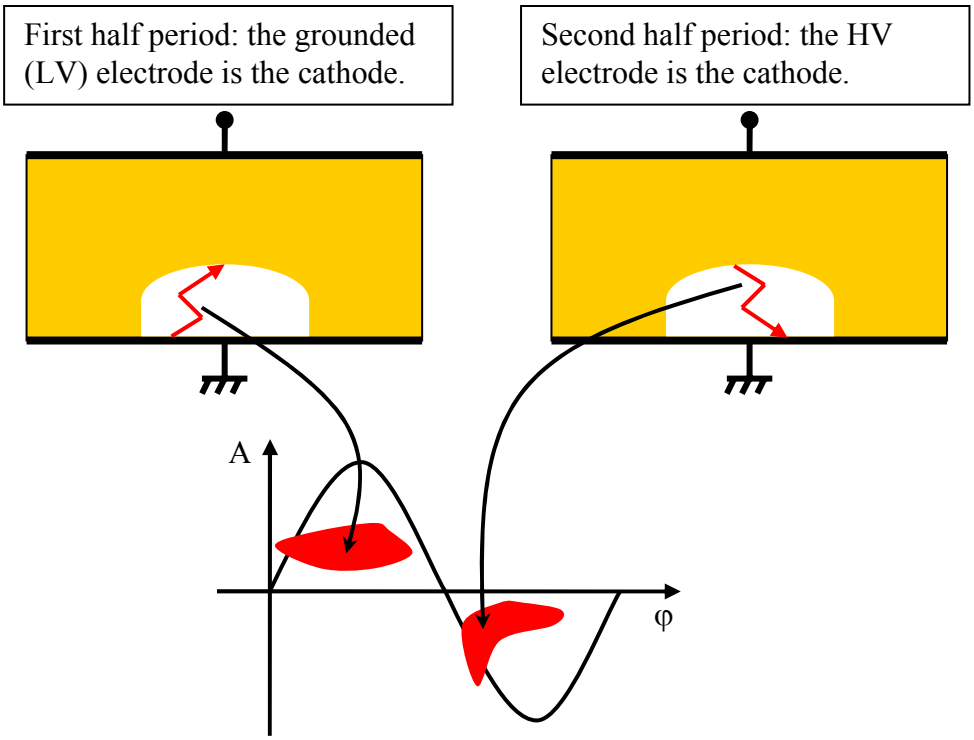


Figure 8.1. Schematization of the discharge process with respect to PD polarity, for a defect localised close to LV electrode. Most of the parameters are first calculated independently for each polarity, and then combined to indicate the dissymmetry between the polarities.

For the localization process, 17 parameters were taken in consideration:

- Combined minimum phase of a PD distribution.
- Combined mean phase of a PD distribution.
- Combined phase interval of a PD distribution.
- Combined skewness of the amplitude distribution.
- Shape factor of the Weibull function that fits the positive amplitude distribution data.
- Shape factor of the Weibull function that fits the negative amplitude distribution data.
- Combined shape factor of the Weibull function that fits the amplitude distribution data.
- Combined scale factor of the Weibull function that fits the amplitude distribution data.
- Combined skewness of the phase distribution.
- Combined minimum absolute value for an amplitude distribution.
- Fuzzy membership to the “Internal” category (first identification level output).
- Fuzzy membership to the “Surface” category (first identification level output).
- Fuzzy membership to the “Corona” category (first identification level output).
- Combined number of discharges of a distribution.
- Combined maximum absolute value for an amplitude distribution.
- Combined standard deviation of an amplitude distribution.
- Combined mean value of the intertime distribution, with reference to polarity.

Multivariate analysis of variance (MANOVA) [Krzanowski88] is a statistical multivariate method that can be applied to a set of grouped data to determine, whether there are significant differences between the independent variables depending on the groups. MANOVA can be divided in two tasks: significance tests and canonical correlation. The significance test is usually carried out

with Wilks' lambda, which is a direct measure of group difference of the centroids of means of independent variables. The smaller is the lambda, the greater is the difference. Bartlett's transformation can be used to compute the significance of lambda.

The significance test often gives the only results one is interested in, but an optional step in MANOVA is the canonical correlation that can be used to better understand the nature of group differences. In order to test the hypothesis that groups differ significantly on weighted combinations of the observed independent variables, MANOVA carries out a multiple discriminant analysis (MDA). With MDA the variance of dependent variables is partitioned into components also called canonical roots. The canonical roots correspond to principal components in PCA, except they seek to maximize the variance between the groups.

Some assumptions should be noticed when applying MANOVA:

- Observations should be independent of one another.
- Group sizes should not be very unequal.
- There should be adequate number of samples. At a minimum, every group should have more cases than there are variables.
- The variance-covariance matrix is the same for each group. Also within group variances should be similar.
- Variables should follow multivariate normal distributions
- No outliers should be in the data.

There are several ways to utilize MANOVA. Firstly, it can be used to compare the characteristics of the groups formed by multiple variables. Secondly, it can be used for reducing a set of variables to a smaller, more easily modelled number of variables. Thirdly, it can be used to identify the variables, which differentiate the groups most. In this study, MANOVA was applied to find parameters that differentiate two sample sets: PD samples from a defect attached to the LV electrode and samples from a defect attached to the HV electrode. MATLAB function *manova1* is applied for the analysis. The sample set is carefully formed by normalising the variables and removing the outliers, so the assumptions of MANOVA are mostly fulfilled, although the number of observations could be larger. The normality of the distributions cannot be guaranteed due to the small number of observations, but the test should give adequate results even if the distributions are not purely normal.

The 17 parameters described before were extracted from the measurements, and MANOVA was applied to find a combination of parameters that gives the best separation between the HV and LV samples. The database was too small to perform multivariate analysis for all parameters at the same time. Thus, all possible combinations of parameter pairs were tested. The couple of parameters that gave the best combination were chosen to be the first in the parameter set. Then, other parameters were added to the set depending on their influence on the separation process. According to the test, most of the parameters did not seem to give any information concerning the separation of HV and LV samples. However, there were three parameters that always made the mean values of the groups significantly different. The first parameter was the new quantity that takes into account the dissymmetry of the PD activity in terms of time between discharges. The second parameter was a combination of minimum phase of occurrence for positive and negative polarity distributions. The third parameter was the shape factor of the Weibull function that fits the distribution of positive discharge amplitude.

Three classification algorithms were trained and tested to classify samples to be either in the vicinity of HV electrode or LV electrode. Inputs of the classifiers were the three parameters that seemed to be the best separators of the classes. In addition to the classifier input selection, the

classification performance depends on tuning the classifier. Each of the classifiers has several tuning parameters that have influence on the results.

Some of the simplest and accurate classification algorithms are based on *k*-NN rule [Cover67]. In this approach, each training data vector is labelled with the class it belongs, and treated as a reference vector, when a new sample is to be classified. During classification process, *k* nearest reference vectors of the new sample are found based on a distance measure (e.g. Euclidean distance) and the class of the sample under study is defined by voting among nearest reference vectors. There are several benefits on using *k*-NN classification approach. It is simple to implement and easy to update, because it actually does not require training. It also outputs a measure that is related to the probability that the classification decision is correct, i.e. the number of votes divided by *k*. If the sample to be classified is on the border of two classes, it gets somewhat equal amount of votes from each class. However, problems occur when applying *k*-NN classification, if the entity to be classified has redundant features concerning the classification. That is why a careful feature selection is important when applying this approach unlike when applying SVM. In this thesis, various features were studied for defect classification based on location, and the best features were chosen with careful analysis. Regarding *k*-NN classifier, the distance function and the size of the neighbourhood, *k*, are chosen while tuning the classifier. Here only Euclidean distance function is used and *k* is varied.

PNNs [Specht88] are radial basis networks suitable for classification problems. The network consists of two layers. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector, whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to output a vector of probabilities. Finally, in the output of the second layer, the maximum of these probabilities is picked up, and the class decision is made based on such a maximum. The activation function in the second layer neurons is a radial basis function (RBF):

$$RBF(x) = \exp\left(\frac{-\|x\|^2}{2\sigma^2}\right) \tag{8.1}$$

Its maximum value is one, and when x grows, it approaches to zero. If the spread σ is near zero, the network should act as a nearest neighbour classifier. As the spread becomes larger, the designed network will take into account several nearby training samples in classification. The optimum value for σ should be between $[0,\infty]$.

When considering SVM, the kernel function and the error penalty *C* need to be chosen. Usually radial basis kernel functions are favoured instead of polynomial kernel functions, because they are not so sensitive to outliers and do not require inputs to have equal variances. However, in this study also polynomial kernels are studied, because the data is normalised and do not contain outliers. When using polynomial kernel function, the order of the polynomial has to be chosen, and when using RBF the spread, σ, has to be decided. A large σ-value will give a smooth decision surface and regular decision boundary.

5-fold cross validation results, obtained by the above-discussed classifiers and with different tunings, are presented in Table 8.1. Almost equally good performance is achieved by several different structures, but the best results (correct classification rate 78.4%) are gained with SVM that has the fourth order polynomial kernel function and error penalty equal to 0.5. The second best results (correct classification rate 73.4%) are achieved with several different SVM classifiers and a *k*-NN classifier with *k* equal to 13. PNN was not competitive with two others.

In Table 8.1, also the comparison of computation times in MATLAB is presented. The computation times should be considered indicative only, because they are highly dependent on the

software and implementation details. The time for set-up corresponds to the training time of SVMs and the time for setting up a new neural network structure when considering PNNs. The classification time of a sample is an average over various test sets in the 5-fold cross validation. The classification of samples is relatively fast with all classification methods, although PNN gives the shortest times. The time required for the set-up is the largest when applying SVM and the shortest when applying the *k*-NN classifier.

Although the SVM based classifiers gave the best classification results, it should be emphasized that the *k*-NN classifier is much simpler to implement and update than SVM, and it also outputs an estimate of the probability that a sample belongs to a certain class. Thus, it is easy to neglect such classification decisions that are not trustworthy enough. For example, if we use a *k*-NN classifier with $k = 14$ and ignore such classification results that have probability equal to 0.5, there will be 26.1% unclassified samples in addition to 59.1% correct classification decisions. By setting a threshold for probability, it is possible to influence on sensitivity and reliability of the monitoring system. With high treshold the classification is more trustworthy but less sensitive and with low treshold opposite. Although SVM also outputs a measure that has increasing value as the likelihood of the sample to be correctly classified increases, it is not as easy to set the threshold that indicates trustworthy classification decision for this measure as it is for the output of the *k*-NN classifier.

Table 8.1. Classification results with different techniques: SVM (support vector machine), k-NN (k-nearest neighbour), PNN (probabilistic neural network)

Classifier	Correct classification rate	Time for set-up	Time for classification
SVM, poly 1, $C = 4$	70.9%	0.30 s	0.0020 s
SVM, poly 4, $C = 0.5$	78.4%	0.26 s	0.0020 s
SVM, poly 3, $C = 0.5$	73.4%	0.20 s	0.0020 s
SVM, rbf $\sigma = 0.5$, $C = 1$	73.4%	0.29 s	0.0020 s
SVM, rbf $\sigma = 0.75$, $C = 2$	73.4%	0.27 s	0.0020 s
k-NN, $k = 11$	70.2%	-	0.0026 s
k-NN, $k = 12$	61.6%	-	0.0026 s
k-NN, $k = 13$	73.4%	-	0.0020 s
k-NN, $k = 14$	59.1%	-	0.0033 s
PNN, $\sigma = 0.01$	55.9%	0.10 s	0.0013 s
PNN, $\sigma = 0.1$	68.4%	0.10 s	0.0013 s
PNN, $\sigma = 0.13$	70.9%	0.11 s	0.0013 s
PNN, $\sigma = 1$	60.2%	0.10 s	0.0013 s

It can be stated that the results with either SVM or *k*-NN are fairly good. The benefits of SVM are the slightly better classification results, and non-requirement of the feature selection before building up the classifier. The benefits of *k*-NN are the non-existence of training phase, and the better usability in practise. It could be concluded that both numerical classification methods can be used in insulation systems PD analysis as a defect localization tool, if classification is based on properly selected features of PD distributions.

9. Conclusions

Induction motors play an important part in the world's industry. Their fault diagnostics and condition monitoring is a widely researched subject. In addition to traditional fault diagnostics methods that are based on analytical models of the diagnosed plant, different kinds of data-based methods have become popular in the area of fault diagnostics of electrical machinery. For example, a wide variety of artificial neural network based applications can be found from the literature. Support vector machine is a modern and highly promising machine learning method, and although it has been successfully applied to numerous classification and pattern recognition problems, its utilization in fault diagnostics is low. In fault diagnostics of induction motors, SVM does not seem to have been applied before this research, even if it showed to be an efficient and reliable way to do the classification of motor faults.

This thesis considered different aspects of induction motor condition monitoring. The developed methods could be applied in on-line condition monitoring of induction motors, but also in testing of recently manufactured motors. In the testing phase, more extensive measurement instrumentation can be applied compared to on-line monitoring. Data based classification tools are especially appropriate for testing motors that are manufactured in large series, because in those cases, a large amount of measurement data can be collected from numerous motor individuals. When considering motors that are manufactured in small series, accurate simulation models need to be used instead to generate virtual measurement data. However, most of the results presented in this thesis are qualitative. Good parameters are found to indicate the motor condition, and various appropriate signal processing tools are considered to further reveal the faults from specific motor variables.

Firstly, a popular motor current signature analysis was studied. Current signatures were formed with FFT based PSD estimation and SVM was applied as a pattern recognition tool to categorize the signatures according to the fault situation. The success of the fault detection was good in noiseless situation, but after a thorough comparison of several motor variables as media of fault diagnostics, it was found out that for example forces on the rotor were more trustworthy indicators of faults than the current. This is an interesting result, because a great deal of all induction motor fault diagnostics research is oriented towards application of the motor current as the indicator of faults.

The forces are difficult to measure, but they are directly related to vibrations that are traditionally used in rotating machines condition monitoring in addition to acoustic signals. Vibration monitoring based induction motor fault diagnostics was further enhanced with various signal processing tools such as calculation of higher order spectra, signal description with autoregressive modeling and cepstrum analysis. Also, information fusion of multi-channel vibration measurements was considered with multivariate data analysis. The vibration based fault diagnostics was tested with real measurement data from healthy and rotor faulted motors, and the results further ensured the suitability of vibrations for induction motor condition monitoring. The best signal processing tool before application of SVM seemed to be the signal description with autoregressive modelling, but it is possible that other faults require other kinds of tools to be revealed.

In addition to detection of faults in electromechanical parts of the motor, condition monitoring of the motor insulation system was considered, and the results presented in the thesis can be extended to other HV and MV apparatus such as cables or generators. The role of partial discharges in degradation and eventual failure of electrical insulation systems at high or medium voltage is a well-established fact, and thus the condition monitoring of insulation is often based on PD analysis. The severity of the effect that PD has on the insulation depends on the nature and the location of the PD generating defect. The PD studies in this thesis did not cover the whole PD diagnostic process,

but it was concentrated on localization of the defects to the vicinity of either of the electrodes. Firstly, feature selection was carried out among various parameters describing the PD distributions, and the most important parameters were chosen to be the basis of the automated localization. The final localization was carried out with several numerical classification methods, and SVM was found to give the best classification results. However, it was noticed that, if the number of classification features can be decreased to only few important parameters, also simpler classification methods such as a *k*-nearest neighbor classifier can be used. The SVMs were not applied neither in PD analysis research before this thesis.

Although SVMs have given excellent results in various classification tasks, one of their drawbacks is that they are essentially 2-class classifiers. In many real applications, such as in fault classification of an induction motor, a multi-class classification problem needs to be solved. In the thesis, also this problem was considered. Four different coupling techniques of the 2-class classifiers were studied to get the global decision of the motor condition. A mixture matrix coupling was found to be the best technique in this application. In this thesis, the coupling schemes are considered with SVMs, but they can also be used with any other pairwise classifiers. Comparison of different coupling techniques is important knowledge concerning the research of classification methods in general, and the mixture matrix approach had not been earlier applied in this form.

In this thesis, data-based models were chosen to be the basis of the fault diagnostics, because simple analytical models of induction motors are not usually accurate enough for description of various motor faults. Enhanced numerical models of the motor, such as the finite element analysis of the magnetic field, can accurately imitate the motor performance, but despite of the increasing efficiency of the computational hardware, the required computation times for detailed numerical models are still too long for on-line implementation. Knowledge-based models were also left out from this thesis, because their construction requires reliable expert knowledge of the motor performance and the influence of faults. Induction motor faults may also manifest themselves in such a complex way that it is difficult even for a highly skilled expert to diagnose all the faults.

SVMs were chosen for building the data-based fault classification models, because they have several benefits compared to conventional data-based modeling methods. Firstly, SVMs are claimed to have better generalisation properties than traditional classifiers. Further, application of SVM results in the global solution for a classification problem. Thirdly, SVM based classification is attractive, because it is very efficient in high dimensional problems: neither its generalisation ability nor computational efficiency is dependent on the dimension of the problem. This property is very useful in fault diagnostics, because the number of utilised features does not have to be drastically limited. It is also easy to remove outliers from the training data sets based on the optimal Lagrange multipliers resulting from SVM training.

At the first glimpse, SVMs seem to give a perfect approach for carrying out classification and pattern recognition tasks, but there also exist drawbacks. Perhaps the most important drawback is that a formal proof that SVMs really minimize the structural risk in high dimensional spaces is still missing, although Vapnik has presented plausible reasons that this happens for example by bounding the generalisation ability with the number of support vectors (Eq. (4.15)). See more on this e.g. in [Burges98]. Second drawback is that SVMs are essentially binary classifiers despite of the wide research done to cope with this problem. Finally, it should be pointed out that choosing the kernel function and setting parameters for SVM is not a straightforward task. Choosing the kernel function is related to choosing the network structure when applying NNs, and one should have prior knowledge on the problem under study to have the best kernel. Usually, the kernel function is chosen by trial and error or cross validation. However, SVM is still an evolving machine learning technique, and the basic SVM method presented in this thesis is continuously studied and improved.

Although, SVM has many benefits compared to for example statistical classifiers, MLPs or RBF networks, it is still a purely data-based modelling method. The success of all data-based methods is highly dependent of the data they are based on. To build a trustworthy data-based model, there should be representative data available from the whole operation area. Especially, in construction of the fault diagnostics system, this may be a problem, because often reliable data lack from the faulty process operation. In many cases, fault models have to be built to generate virtual measurement data from abnormal process operation.

Also, the results of this thesis are discovered without the data from naturally developing fault situations. Simulated data from enhanced numerical induction motor models and real measurement data from artificially damaged motors are used instead. The lack of entirely reliable data may degrade the diagnostics performance in real operation. In implementation of the method for on-line condition monitoring of induction motors of real industrial processes, also model calibration and updating issues may cause problems. Some differences may occur in behaviour of individual motors, and utilised models should be calibrated for each motor. It would also be very useful to update the models after occurrence of any fault, but this is not possible in most of the applications.

This thesis did not cover all possible faults of an induction motor. Although bearing faults are the most common induction motor faults, they were not considered because of the lack of data from bearing faulted machines. Also, in vibration monitoring only rotor related faults were studied. The considerations of wider variety of faults are left for future study. In PD analysis, only a part of the third level identification was considered. Construction of a thorough PD identification system is also left for future study. Further, the complete testing of the developed diagnostics method in real on-line operation of an induction motor will be carried out in the future.

References

[Abdul00]: Abdul Rahman M.K., Arora, R., Srivastava, S.C, "Partial Discharge Classification Using Principal Component Analysis", IEE Proc.-Sci.Meas. Technol., pp. 7-13, Vol. 147, No.1, January 2000.

[Alguindigue93]: Alguindigue, I.E., Loskiewicz-Buczak, Uhrig, R., "Monitoring and Diagnosis of Rolling Element Bearings Using Artificial Neural Networks", IEEE Transactions on Industrial Electronics, Vol. 40, No.2, pp. 209-217, April 1993.

[Altug99]: Altug, S., Mo-Yuen Chow, Trussell, H.J. "Fuzzy Inference Systems Implemented on Neural Architectures for Motor Fault Detection and Diagnosis", IEEE Transactions on Electronics, Vol. 46 Issue: 6, pp.1069 –1079, Dec. 1999.

[Arkkio90]: Arkkio, A., "Finite Element Analysis of Cage Induction Motors Fed by Static Frequency Converters. IEEE Transactions on Magnetics, 26, 2, pp. 551-554, 1990.

[Arthur00] Arthur, N., Penman, J., "Induction Machine Condition Monitoring with Higher Order Spectra", IEEE Trans. on Industrial Electronics, Vol. 47, No. 5, pp. 1031-1041, 2000.

[Batur02]: Batur, C.; Zhou, L.; Chan, C.C; "Support Vector Machines for Fault Detection", Proc. of the 41st IEEE Conf. on Decision and Control, Vol.2 , pp. 1355 -1356, Dec.2002.

[Benbouzid00]: Benbouzid, M., "A Review of Induction Motor Signature Analysis as a Medium for Faults Detection", IEEE Transactions on Industrial Electronics, Vol. 47, No. 5, pp. 984-993, October 2000.

[Betta01]: G. Betta, C. Liguori, A. Paolillo, A. Pietrosanto, A., "A DSP-based FFT Analyzer for the Fault Diagnosis of Rotating Machine Based on Vibration Analysis", Proc. IEEE Conf. on Instrumentation and Measurement Technology, pp. 572-577, Hungary, 2001.

[Boser92]: Boser, B.E., Guyon, I.M., Vapnik, V.N., "A Training Algorithm for Optimal Margin Classifiers", Proc. Of the 5th Annual ACM Workshop on Computational Learning Theory, pp. 142-152, ACM Press, 1992.

[Brown97]: Brown, M.P.S., Grundy, W., Lin, D., Cristianini, N., Sugnet, C., Furey, T.S., Ares, M. Jr., Haussler, D.: "Knowledge-based Analysis of Microarray Gene Expression Data Using Support Vector Machines", Proceedings of the National Academy of Sciences, Vol. 1, pp. 262-267, 1997.

[Burges98]: Burges C., A Tutorial on Support Vector Machines for Pattern Recognition, Journal of data Mining and Knowledge Discovery, 2(2),121-167, 1998.

[Cavallini031]: Cavallini, A., Contin, A., Montanari, G. C., Puletti, F. "Advanced PD Inference in On-Field Measurements. Part 1: Noise Rejection", IEEE Trans. on DEI, Vol. 10, n. 2, pp. 216-224, April 2003.

[Cavallini032]: Cavallini, A., Conti, M., Contin, A., Montanari, G. C. "Advanced PD Inference in On-Field Measurements. Part 2: Identification of Defects in Solid Insulation Systems", IEEE Trans. on DEI, Vol. 10, n. 3, pp. 528-538, June 2003.

[Cavallini02]: Cavallini, A., Conti, M., Contin, A., Montanari, G. C. "Inferring Partial Discharge Identification through Fuzzy Tools", IEEE CEIDP, pp. 698-702, October 2002.

[Chin98]: Chin, K.K., "Support Vector Machines applied to Speech Pattern Classification**",** MPhil. Thesis, Department of Engineering, University of Cambridge, 1998.

[Contin94]: Contin, A.; Cacciari, M.; Montanari, G.C.; "Estimation of Weibull Distribution Parameters for Partial Discharge Inference" IEEE Conf. on Electrical Insulation and Dielectric Phenomena, pp. 71 -78, Oct. 1994.

[Contin02]: A. Contin, A. Cavallini, G. C. Montanari, G. Pasini, F. Puletti, "Digital Detection and Fuzzy Classification of Partial Discharge Signals", IEEE Trans. on DEI, Vol.9, N.3, pp.335-348, June 2002.

[Coulomb83] J.L. Coulomb, "A Methodology for the Determination of Global Electromechanical Quantities from the Finite Element Analysis and Its Application to the Evaluation of Magnetic Forces, Torques and Stiffness". IEEE Transactions on Magnetics, MAG-19, (6), pp. 2514-2519, 1983.

[Cover67]: Cover, T. M., Hart, P. E.: "Nearest Neighbor Pattern Classification," IEEE Trans. Inform. Theory, Vol. IT-13, pp. 21-27, Jan. 1967.

[Cristianini00]: Cristianini, N., Shawe-Taylor, J., "Support Vector Machines and Other Kernel-Based Learning Methods", Cambridge University Press, 2000

[Danikas03]: Danikas, M.G., Gao, N., Aro. M. "Partial Discharge Recognition Using Neural Networks: a Review", Electrical Engineering, Vol. 85, No. 2, pp. 87-93, May 2003.

[Dexter95]: Dexter, A.L., "Fuzzy Model Based Fault Diagnosis", IEE Proc. Control Theory Appl., Vol. 142, No. 6, pp. 545-550, November 1995.

[Ding01]: Ding, C. H.Q, Dubchak, I., "Multi-Class Protein Fold Recognition Using Support Vector Machines and Neural Networks", Bioinformatics, Vol.17, No.4, pp.349-358, April 2001.

[Duda01]: Duda, R.O., Hart, P.E., Stork, D.G. "Pattern Classification", John Wiley & Sons, Inc., 2001.

[Epri82]: EPRI publication EL-2678, "Improved Motors for Utility Applications", Vol.1, October 1982.

[Feng02]: Feng, Y., Lun, S.Y., Di, L., Zong, L.Y., "Application of Support Vector Machines to Quality Monitoring in Robotized Arc Welding", Int. Joint Conf. On Neural Networks, Vol. 3, pp. 2321-2326, 2002.

[Filippetti95]: Filippetti, F., Franceschini, G., Tassoni, C.: "Neural Network Aided On-Line Diagnostics of Induction Motor Rotor Faults," IEEE Trans. Ind. Applicat., vol. 31, pp. 892–899, July/Aug. 1995.

[Freiss98]: Freiss, T., Cristianini, N., Campbell, C., "The Kernel-Adatron Algorithm: A Fast and and Simple Learning Procedure for Support Vector Machines", Proc. Of the 15th Int. Conf. On Machine Learning, ICML'98, pp. 180-196, 1998.

[Friedman96]: Friedman, J., "Another Approach to Polychotomous Classification", Technical Report, Stanford University, Department of Statistics, 1996.

[Gao02]: Gao, J.; Shi, W.; Tan, J.; Zhong, F.; "Support Vector Machines Based Approach for Fault Diagnosis of Valves in Reciprocating Pumps", IEEE Canadian Conf. on Electrical and Computer Engineering, Vol. 3, pp. 1622 -1627 May 2002.

[Gelle01]: Gelle, G., Colas, M., Serviere, C.: "Blind Source Separation: A Tool for Rotating Machine Monitoring by Vibration Analysis", Journal of Sound and Vibration, 248(5), pp. 865-885, 2001.

[Gertler92]: Gertler, J.J., Anderson, K.C., "An Evidential Reasoning Extension to Quantitative Model-Based Failure Diagnosis", IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, Issue 2, pp. 275 –289, March-April 1992.

[Guru01]: Guru, B.S., Hiziroglu, H.R., "Electric Machinery and Transformers", Oxford University Press, 2001.

[Haykin99]: Haykin, S.,"Neural Networks – A Comprehensive Foundation", Prentice Hall International, Inc., 1999.

[Hyvärinen99]: Hyvärinen, A. "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis, IEEE Trans. on Neural Networks, 10(3), 1999, 626-634.

[IEC01]: IEC 60270, Partial Discharge Measurements, 3rd edition, March 2001.

[Inoue01]: Inoue, T., Abe, S., "Fuzzy Support Vector Machines for Pattern Classification", Proc. of Int. Joint Conference on Neural Networks, IJCNN '01, Vol. 2, pp. 1449 -1454, 2001.

[Isermann93]: Isermann, R., "Fault Diagnosis of Machines via Parameter Estimation and Knowledge Processing", Automatica, Vol.29, No. 4, pp. 815-835, 1993.

[Isermann98]: Isermann, R., "On Fuzzy Logic Applications for Automatic Control, Supervision, and Fault Diagnosis", IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, Vol. 28, No. 2, pp. 221-235, March 1998.

[Jack00]: Jack, L.B., Nandi, A.K., "Genetic Algorithms for Feature Selection in Machine Condition Monitoring with Vibrating Signals", IEEE Proc. Vis. Image Signal Processing, Vol.147, No.3, pp.205-212, June 2000.

[Joachims97]: Joachims, T., "Text Categorization with Support Vector Machines: Learning with Many Relevant Features", University of Dortmund, Technical Report, 1997.

[Kecman01]: Kecman, V. "Learning and Soft Computing; Support Vector Machines, Neural Networks and Fuzzy Logic Models", The MIT Press, 2001.

[Kliman92]: Kliman, G.B., et al. "Methods of Motor Current Signature Analysis", Elect. Mach. Power Syst., vol. 20, no.5, pp. 463-474, Sept 1992..

[Knaak01]: Knaak, M, Fausten, M., Filbert, D.: "Acoustical Machine Monitoring Using Blind Source Separation", 4th International Conference on Acoustical and Vibratory Surveillance Methods and Diagnostics Technique, Compiègne, France, Sept. 2001.

[Knaak02]: Knaak, M., Kunter, M., Filbert, D.: "Blind Source Separation for Acoustic Machine Diagnosis" Procceedings of 14th IEEE Conference on Digital Signal Processing, DSP2002, Vol. I, Santorini, Greece, July, 2002.

[Kral00]: Kral, C., Pirker, F., Pascoli, G. "Detection of Rotor Faults in Squirrel Cage Induction Machines at Standstill for Batch Tests by Means of the Vienna Monitoring Method", Conf. Record of IEEE Conf. On Industry Applications, Vol. 1, pp. 499 –504, 2000.

[Krzanowski88]: Krzanowski, W. J. Principles of Multivariate Analysis. Oxford University Press, 1988.

[Laggan99]: Laggan, P.A. "Vibration Monitoring", Proc. IEE Colloquium on Understanding your Condition Monitoring, pp. 1-11, 1999.

[Lasurt00]: Lasurt, I., Stronach, A.F., Penman, J., "A Fuzzy Logic Approach to the Interpretation of Higher Order Spectra Applied to Fault Diagnosis in Electrical machines", 19th Int. Conf. Of the North American Fuzzy Information Processing Society, NAFIPS., pp. 158 –162, 2000.

[Li00]: Li, B., Chow, M-Y, Tipsuwan, Y., Hung, J.C., "Neural-Network-Based Motor Rolling Bearing Fault Diagnosis", IEEE Transactions on Industrial Electronics, Vol. 47, No. 5, pp. 1060-1069, October 2000.

[Lindh03]: Lindh, T. "On the Condition Monitoring of Induction Machines", Doctoral Thesis, Lappeenranta University of Technology, 2003.

[Liu97]: Liu, B., Si, J. "Fault Isolation Filter Design for Linear Time-Invariant Systems", IEEE Transactions on Automatic Control, Vol.42, Issue 5, pp. 704 - 707, May 1997.

[Loparo00]: Loparo, K.A., Adams, M.L., Lin, W., Abdel-Magied, M.F., Afshari, N.: "Fault Detection and Diagnosis of Rotating Machinery", IEEE Trans. on Industrial Electronics, Vol. 47, Issue: 5, pp. 1005 –1014, Oct. 2000.

[Ma02]: Ma, X.; Zhou, C.; Kemp, I.J.; "Interpretation of Wavelet Analysis and Its Application in Partial Discharge Detection", IEEE Trans. on DEI, Vol.9 Issue: 3, pp.446 –457, June 2002.

[Marcal00]: Marcal, R.F.M., Negreiros, M., Susin, A.A., Kovaleski, J.L. "Detecting Faults in Rotating Machines", IEEE Instrumentation & Measurement Magazine, *3*(4), 2000, 24-26.

[Mayoraz99]: Mayoraz, E., Alpaydin E., "Support Vector Machines for Multi-Class Classification", Int. Workshop on Artificial Neural Networks (IWANN), Vol. 2, pp. 833-842, 1999.

[Merwe02]: van der Merwe, N.T., Hoffman, A.J.: "A Modified Cepstrum Analysis Applied to Vibrational Signals", 14th Int. Conf. on Digital Signal Processing, Vol. 2, pp. 873 -876, 2002.

[Nejjari99]: Nejjari, H., Benbouzid, M.E.H., "Application of Fuzzy Logic to Induction Motors Condition Monitoring" IEEE Power Engineering Review, Vol. 19, Issue 6, pp. 52 –54, June 1999.

[Nikias93] Nikias, C.L., Mendel, J.M.: "Signal Processing with Higher Order Spectra", IEEE Signal Processing magazine, pp.10-37, Vol. 10, Issue 3, July 1993.

[Patton99]: Patton, R.J., Lopez-Toribio, C.J., Uppal, F.J., "Artificial Intelligence Approaches to Fault Diagnosis", Condition Monitoring: IEE Colloquium on Machinery, External Structures and Health (Ref. No. 1999/034), pp.5/1-5/181999.

[Penman94]: Penman, J., Yin, ,C. M. "Feasibility of Using Unsupervised Learning NN for the Condition Monitoring of Electrical Machines," Proc. IEE—Elect. Power Applicat., vol. 141, no. 6, pp. 317–322, 1994.

[Pontil98]: Pontil, M., Verri, A., "Object Recognition with Support Vector Machines", IEEE Trans. On PAMI, Vol.20, pp. 637-646, 1998

[Ribeiro02]: Ribeiro, B, "Support Vector Machines and RBF Neural Networks for Fault Detection and Diagnosis", Proc. of The 8th Int. Conf. on Neural Information Processing, paper 191, 2002.

[Rychetsky99]: Rychetsky, M.; Ortmann, S.; Glesner, M.; "Support Vector Approaches for Engine Knock Detection", Int. Joint Conf. on Neural Networks, Vol. 2, pp. 969 -974, July 1999.

[Salama00]: Salama, M. M. A., Bartnikas, R. "Fuzzy logic applied to PD pattern recognition", IEEE Trans. on DEI, Vol.7, No. 1, pp. 118-123, February, 2000.

[Salomon01]: Salomon, J., "Support Vector Machines for Phoneme Classification", M.Sc Thesis, University of Edinburgh, 2001.

[Satish95]: Satish, L.; Zaengl, W.S.; "Can Fractal Features be Used for Recognizing 3-D Partial Discharge Patterns?", IEEE Trans. on DEI, Vol.2, No. 3 , pp. 352 -359 June 1995.

[Saunders00]: Saunders, C., Gammerman, A., Brown, H., Donald, G., "Application of Support Vector Machines to Fault Diagnosis and Automated Repair", Proc. Of the 11th Int. Workshop on Principles of Diagnosis (DX'00), June 8-10, 2000.

[Schoen95]: Schoen, R.R., Lin, B.K., Habetler, T.G.; Schlag, J.H.; Farag, S. "An Unsupervised, On-Line System for Induction Motor Fault Detection Using Stator Current Monitoring", IEEE Transactions on Industry Applications, Vol. 31 Issue: 6, pp. 1280 –1286, Nov.-Dec. 1995.

[Sorsa95]: Sorsa, T., "Neural Network Approach to Fault Diagnosis", Doctoral Thesis, Tampere University of Technology Publications 153, 1995.

[Specht88]: Specht, D.F. "Probabilistic Neural Networks for Classification, Mapping, or Associative Memory", in Proc. of the IEEE Int. Conf. on Neural Networks, IEEE ICNN, 1988, Vol. 1, pp. 525 -532.

[Suykens02]: Suykens, J.A.K., Van Gestel, T., De Brabanter, J. ,. De Moor, B, Vandewalle, J. "Least Squares Support Vector Machines", World Scientific, Singapore, 2002.

[Schwenker00]: Schwenker, F., "Hierarchical Support Vector Machines for Multi-Class Pattern Recognition", Proc. of 4th Int. Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, Vol. 2, pp. 561 –565, 2000.

[Vapnik00]: Vapnik, V.N., "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 2000.

[Wang97]: Wang, L-X, "A Course in Fuzzy Systems and Control", Prentice Hall PTR, 1997.

[Welch67]: Welch, P.D, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Transactions on Audio Electroacoustics, Vol. AU-15, pp. 70-73, June 1967.

[Wenzel94]: D, Wenzel, H, Borsi, E. Gockenbach, "Partial Discharge Recognition and Localization on Transformers via Fuzzy Logic", IEEE ISEI, pp. 233-236, June 1994.

[Wieser97]: Wieser, R., Kral, C., Pirker, F., Schagginger, M. "Condition Monitoring of Inverter Fed Induction Machines by Means of State Variable Observation", Eigth Int. Conf. on Electrical Machines and Drives, No. 444, pp. 336 –340, 1997.

[Wieser98]: Wieser, R., Kral, C., Pirker, F., Schagginger, M. "Robust Induction Machine Cage Monitoring Technique for Highly Distorted Voltage and Current Waveforms, the Vienna method", Seventh Int. Conf. on Power Electronics and Variable Speed Drives, No. 456, pp. 194 –199, 1998.

[Wildi97]: Wildi, T., "Electrical Machines, Drives and Power Systems", Prentice Hall Int., 1997.

[Yang00]: Yang, D-M, Penman, J.: "Intelligent Detection of Induction Motor Bearing Faults Using Current and Vibration Monitoring", Proc. of the Int. Cong. On Condition Monitoring and Diagnostics Engineering Management, Vol. 1, pp. 461-470, December 2000

[Ypma991]: A. Ypma, P. Pajunen, P, "Rotating Machine Vibration Analysis with Second-Order Independent Component Analysis", Proc. of the Workshop on ICA and Signal Separation, Aussois, France, 1999, 37-42.

[Ypma992]: A. Ypma, D.M.J. Tax, R.P.W Duin, "Robust Machine Fault Detection with Independent Component Analysis and Support Vector Data Description", Proc. IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing, Madison, Wisconsin, USA, 1999, 67–76.

[Ypma02]: Ypma, A., Leshem, A., Duin, R.P.W.: “Blind Separation of Rotating Machine Sources: Bilinear Forms and Convolutive Mixtures”, Neurocomputing - Special Issue on ICA/BSS, volume 49/1-4, pp. 349-368, December 2002

[Yu02]: Yu, S.; Ma, F-Y; Chen, J-X; Yin, X-G; Shi H-B; “Unsteady Fault Diagnosis Method for Chemical Process Based on SVM”, Proc. of Int. Conf. on Machine Learning and Cybernetics, Vol. 2, pp. 772 -775, Nov. 2002.

[Zöllner02]: Zöllner, J.M.; Berns, K.; Dillmann, R.; “Diagnosis of Large Inspection Datasets Using an Adaptive, Learning System”, Intl. conf. on Multisensor Fusion and Integration for Intelligent Systems, pp. 31 -36, Aug. 2001.