

# **Studies in Computational Intelligence**

Volume 989

## **Series Editor**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

Rahul Kumar Sevakula · Nishchal K. Verma

# Improving Classifier Generalization

Real-Time Machine Learning based  
Applications

Rahul Kumar Sevakula  
Signal Processing  
Whoop Inc  
Boston, MA, USA

Nishchal K. Verma  
Department of Electrical Engineering  
Indian Institute of Technology Kanpur  
Kanpur, India

ISSN 1860-949X ISSN 1860-9503 (electronic)  
Studies in Computational Intelligence  
ISBN 978-981-19-5072-8 ISBN 978-981-19-5073-5 (eBook)  
<https://doi.org/10.1007/978-981-19-5073-5>

© Springer Nature Singapore Pte Ltd. 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

Classification algorithms form the basis of decision-making in most pattern recognition problems, e.g. image recognition, speech and speaker recognition, iris recognition, and spam mail detection. With the horizon of their applications expanding at a fast pace, the need for further research has only increased. This fact becomes particularly true because (a) each application poses its own set of challenges and (b) one would always find a classifier with a particular improvisation that best suits the situation. No matter which classification approach is used, generalization is an important aspect. Generalization essentially indicates how well the trained classifier works in real time, i.e. on unseen test data.

This monograph begins with the fundamentals of classifiers, bias-variance tradeoff, statistical learning theory (SLT), probably approximate correct (PAC) framework, maximum margin classifiers, and popular methods which improve generalization like regularization, boosting, transfer learning, dropout in deep learning, etc. Furthermore, the monograph solves four independent problems that have great relevance for certain real-time applications.

The first part of the monograph aims at finding classifiers which exhibit extremely low variance. Classification algorithms are traditionally designed to simultaneously reduce errors caused by bias as well as variance. In many situations, low variance becomes extremely crucial for getting tangible classification solutions and even slight overfitting can have serious consequences on the test results. Classifiers with low variance have two main advantages: (1) the classifier statistically manages to keep the test errors close to the training error, and (2) the classifier learns effectively even with a small number of samples. The monograph introduces a class of classifiers called Majority Vote Point Classifier (MVPC), which on account of the lower Vapnik Chervonensis (VC) dimension can exhibit lower variance than even linear classifiers. The monograph contributes by estimating a trend for the MVP classifier's VC dimension, and validates its low variance on two real-time problems.

The monograph then focuses on the real-time application of condition-based monitoring of machines using acoustic and vibration measurements. Signal data acquired from machines are often found to change with time, wear and tear, and subsequent repair of the machine. Classifiers are typically trained to perform the

decision-making procedure during fault diagnosis/detection. Since data may change with time, low generalization error is essential to avoid overfitting during classification. Therefore, MVPC is seen to be best suited for this situation. However, MVPC has a limitation that it may not be able to fit the data sufficiently, and may have high training errors. The monograph presents a novel framework for pattern recognition, where novel procedures for optimal data source (sensitive position) identification, data acquisition, and feature selection are tailored to give the best possible training performance with the MVP classifier. The understanding here is that if MVP gives low training error, real-time fault diagnosis of machines becomes feasible with consistent accuracy. The introduced framework was experimentally implemented and tested for an air compressor condition monitoring application; associated real-time experiments showed a significant improvement in the reliability of fault detection.

The third part of the monograph focuses on dealing with class noise in the Fuzzy Support Vector Machine (FSVM) classifier. FSVM is considered to be a significant addition over soft margin SVM like C-SVM, for the former can guard against outlier sensitivity and the latter cannot. The ability of FSVM to absorb outliers strongly depends on how well the training samples are assigned fuzzy membership values (MVs). Traditionally, the membership functions (MFs) used for FSVM were custom-made for applications, and MFs used for one could in general not be used for others. To overcome the limitation, General Purpose Membership Functions (GPMFs) are defined in this monograph as those MFs which can universally be used for multiple applications, and which allow FSVM to statistically perform better than C-SVM. The monograph contributes to the GPMF literature in two stages. Firstly with help of convex hulls, it presents a few limitations that FSVM faces while treating all samples of a class with a single MF. Further, it recommends differential treatment of data by categorizing them into two fuzzy sets: one containing possible non-outliers and the other containing possible outliers. While possible outliers are modeled with a normal MF, possible non-outliers are recommended to have a constant MV of '1'. The chapter then introduces novel GPMFs which use clustering-based techniques to detect possible outliers, and use Hausdorff Distance and pt-set to characterize those possible outliers. To establish conclusions, the introduced GPMFs are thoroughly evaluated and statistically compared with earlier GPMFs on numerous real-world benchmark datasets. The results show that proposed GPMFs not only perform significantly better in treating class noise, but also execute with efficient run time complexity.

Finally, a novel scheme to introduce deep learning in Fuzzy Rule-based classifiers (FRCs) is presented. FRCs have gained prominence for their unique ability of giving good classification performance, and allowing existing expert knowledge to be used jointly with training data. Recent innovations in Deep Neural Networks are allowing researchers to tackle some very complex problems with improved theoretical and empirical justifications, e.g. image classification and audio classification. The monograph presents a scheme to incorporate stacked denoising sparse autoencoders within the FRC framework. While stacking of denoising sparse autoencoders helps learn the complex non-linear relationships among data and represent the input data in a reduced compact feature space, the framework built toward FRC allows users

to input expert knowledge to the system. To make denoising sparse autoencoders learn more effectively, data pre-processing strategies have been proposed. Further, to improve the classification performance and rule reduction performance of the FRC, three fine-tuning strategies have also been proposed. The proposed framework is tested across real-world benchmark datasets, and an elaborate comparison across literature shows that proposed methods are capable of building FRCs that provide state-of-the-art accuracies and/or a few rules, as per the user's demand.

The monograph ends with an epilogue, on the use of autoencoders in transfer learning, tumor classification, and condition monitoring problems. Furthermore, pertaining to the research contributions made herein, the directions for possible future work have also been discussed.

Boston, USA  
Kanpur, India

Rahul Kumar Sevakula  
Nishchal K. Verma

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Basics	1
1.2	Statistical Learning Theory and VC Dimension	3
1.2.1	VC Dimension	3
1.2.2	Probably Approximate Correct (PAC) Learning Framework	4
1.2.3	Growth Function	6
1.3	Bias-Variance Trade-Off	6
1.3.1	Occam's Razor	7
1.4	Outline	8
	References	9
<b>2</b>	<b>Methods Used to Improve Generalization Performance</b>	11
2.1	Maximum Margin Classifier	12
2.1.1	Upper Bound on VC Dimension of Linear Classifiers	12
2.1.2	Support Vector Machine (SVM)	13
2.2	Bagging and Boosting	15
2.3	Semi-Supervised Learning, Active Learning, and Transfer Learning	16
2.4	Regularization	17
2.5	Stochastic Gradient Descent, Momentum, Adaptive Learning Rates, Dropout, Skip Connections, Etc.	18
2.6	Conclusions	20
	References	20
<b>3</b>	<b>MVPC—A Classifier with Very Low VC Dimension</b>	23
3.1	Majority Vote Classifier	23
3.1.1	The Majority Vote Point (MVP) Classifier	24
3.1.2	Implementation of MVP Classifier	25



3.2	Evaluating and Comparing VC Dimension .....	25
3.2.1	Upper Bound on VC Dimension of MVP Classifier .....	25
3.2.2	Empirical Estimation of VC Dimension .....	29
3.2.3	Comparing with VC Dimension of Linear Classifiers .....	32
3.3	Case Studies .....	32
3.3.1	Rotating Machine Fault Diagnosis .....	32
3.3.2	Biological Data .....	35
3.4	Summary .....	38
	References .....	39
<b>4</b>	<b>Framework for Reliable Fault Detection with Sensor Data .....</b>	<b>41</b>
4.1	Literature Survey on Fault Diagnosis in Rotating Machines .....	41
4.2	Problem Formulation and Summary .....	44
4.3	Data Acquisition Framework to Simulate Real-Time Environment .....	44
4.4	Data Pre-processing .....	46
4.4.1	Normalization Robust to Outliers .....	47
4.5	Feature Extraction .....	51
4.6	Feature Selection Algorithm Using Graphical Indices .....	54
4.6.1	Feature Ranking .....	55
4.6.2	Dataset Rejection .....	60
4.6.3	Dataset Retrieval .....	62
4.6.4	Feature Selection Architecture .....	62
4.7	Classification .....	63
4.8	Sensitive Position Analysis (SPA) .....	64
4.8.1	Parameter Range Identification (PRI) .....	65
4.9	Case Study on Air Compressor Fault Detection .....	66
4.9.1	Leakage Inlet Valve (LIV) Fault Detection .....	66
4.9.2	Leakage Outlet Valve (LOV) Fault Detection .....	69
4.9.3	Offline Testing .....	70
4.9.4	Real-Time Testing .....	72
4.10	Summary .....	73
	References .....	74
<b>5</b>	<b>Membership Functions for Fuzzy Support Vector Machine in a Noisy Environment .....</b>	<b>77</b>
5.1	Dealing with Class Noise .....	78
5.2	Fuzzy Support Vector Machine (FSVM) .....	79
5.2.1	Available Membership Functions for FSVM .....	80
5.2.2	Limitations of Earlier Membership Functions .....	81
5.2.3	Convex Hull Analysis on FSVM .....	82
5.3	Set Measures—Distance Measure Between Points and Non-empty Sets .....	85
5.3.1	Distance Between a Point and a Non-empty Set .....	86
5.3.2	Hausdorff Distance (HD) .....	87

5.4	Proposed General Purpose Membership Functions .....	88
5.4.1	GPMFs with Density-Based Clustering .....	89
5.4.2	Proposed GPMFs with Fuzzy C-Means Clustering .....	96
5.5	Case Study and Performance Evaluation .....	99
5.5.1	Class Imbalance Learning in FSVM .....	99
5.5.2	Performance Evaluation .....	100
5.5.3	Experimentation .....	101
5.6	Results and Analysis .....	102
5.6.1	Statistical Analysis .....	102
5.7	Summary .....	108
	References .....	109
<b>6</b>	<b>Stacked Denoising Sparse Autoencoder-Based Fuzzy Rule Classifiers .....</b>	<b>113</b>
6.1	Fuzzy Rule Classifiers (FRC) .....	114
6.1.1	FRC by Ishibuchi et al. [6] .....	114
6.1.2	State-of-the-Art FRC Variations .....	116
6.2	Stacked Autoencoder (SAE)-Based FRC .....	116
6.2.1	Stacked Denoising Sparse Autoencoders (SDSAE) .....	117
6.2.2	Naive Stacked Autoencoder-Based FRC .....	120
6.3	Data Pre-processing Strategies for SDSAE-Based FRC .....	121
6.3.1	Pre-processing Strategy I (PP-I) .....	121
6.3.2	Gaussian Mixture Model (GMM) for Pre-processing .....	122
6.3.3	Pre-processing Strategy II (PP-II) .....	124
6.3.4	Pre-processing Strategy III (PP-III) .....	125
6.3.5	Pre-processing of Nominal Features .....	127
6.4	Fine-Tuning of Weights for FRC Modeling .....	128
6.4.1	Process Definitions .....	129
6.4.2	Fine-Tuning Strategy I (FT-I) .....	130
6.4.3	Fine-Tuning Strategy II (FT-II) .....	132
6.4.4	Fine-Tuning Strategy III (FT-III) .....	134
6.5	Integration with Expert Knowledge .....	138
6.6	Experimentation and Results .....	139
6.6.1	Dataset-Wise Observations .....	143
6.7	Summary .....	145
	References .....	145
<b>7</b>	<b>Epilogue .....</b>	<b>149</b>
7.1	Transfer Learning for Molecular Cancer Classification .....	149
7.1.1	Proposed Transfer Learning Model for Cancer Classification .....	150
7.2	Autoencoders for Time-Series Classification .....	155

7.3	Directions for Future Work .....	156
7.4	Conclusion .....	157
	References .....	157
<b>Appendix: Statistical Tests for GPMFs Analysis in FSVM .....</b>		<b>159</b>

## About the Authors

**Dr. Rahul Kumar Sevakula** has over 10 years of research experience in machine learning (ML) and deep learning (DL). He received his Bachelor's degree from the National Institute of Technology (NIT) Warangal, India in 2009 and later his Ph.D. degree from the Indian Institute of Technology (IIT) Kanpur, India in 2017. He is currently a Sr. Research Scientist at Whoop, and his research interests lie at the intersection of ML, physiological signals, cardiovascular health monitoring (medicine) and wearables. Prior to joining Whoop, he was an Instructor (junior research faculty) at Harvard Medical School and Massachusetts General Hospital, USA, and a Data Scientist at IBM India. He has filed multiple patent disclosures and has published over 45 research papers in international peer-reviewed journals and conferences. He is also a reviewer for several journals of national and international repute.

**Dr. Nishchal K. Verma** is a Professor in the Department of Electrical Engineering at Indian Institute of Technology (IIT) Kanpur, India. Dr. Verma's research interest falls in Artificial Intelligence (AI) related theories and its practical applications to inter-disciplinary domains like machine learning, deep learning, computer vision, prognosis and health management, bioinformatics, cyber-physical systems, complex and highly non-linear systems modeling, clustering, and classifications, etc. He has published more than 250 research papers in peer-reviewed reputed conferences and journals along with 4 books (edited/co-authored) in the field of AI. He has 20+ years of experience in the field of AI. He is currently serving as Associate Editor/ Editorial Board Member of various reputed journals and conferences. He has also developed several AI-related key technologies for The BOEING Company, USA.

# Acronyms

AE	Autoencoders
ANN	Artificial Neural Network
CBM	Condition-based Maintenance
CIL	Class Imbalance Learning
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DBSCAN	Density-based Spatial Clustering of Applications with Noise
DEC	Differential Error Cost
DFT	Discrete Fourier Transform
DL	Deep Learning
DNN	Deep Neural Network
DR	Dataset Rejection
DRL	Dataset Retrieval
DSAE	Denoising Sparse Autoencoders
EM	Expectation Maximization
FCM	Fuzzy C-Means
FDA	Fisher's Discriminant Analysis
FFN	Feed Forward Network
FR	Feature Ranking
FRC	Fuzzy Rule Classifier
FS	Feature Selection
FSVM	Fuzzy Support Vector Machine
GA	Genetic Algorithm
GM	Gaussian Mixture
GMM	Gaussian Mixture Model
GPMF	General Purpose Membership Function
HD	Hausdorff distance
ITER	Individual Training Error Reduction
LIV	Leakage Inlet Valve
LOV	Leakage Outlet Valve
MDL	Minimum Descriptive Length

MF	Membership Function
MV	Membership Value
MVP	Majority Vote Point
MWT	Morlet Wavelet Transform
NDM	Normalized distance of means
OPTICS	Ordering Points To Identify the Clustering Structure
PRI	Parameter Range Identification
RBF	Radial basis function
RBM	Restricted Boltzman Machine
RM	Ratio of Means
RNN	Recurrent Neural Network
SAE	Stacked Autoencoders
SC	Separation Count
SDM	Standard deviation of means
SDSAE	Stacked Denoising Sparse Autoencoders
SGD	Stochastic Gradient Descent
SLT	Statistical Learning Theory
SPA	Sensitive Position Analysis
SSAE	Stacked Sparse Autoencoders
SVM	Support Vector Machine
VC	Vapnik Chervonenkis
WPT	Wavelet Packet Transform
ZCR	Zero Crossing Rate

# Symbols

$\mathbb{R}$	Space of real numbers
$\Omega$	Sample space or Metric space
$S$	A subset of $\Omega$
$H$	Hypothesis space
$\Pi_H$	Growth function of $H$
$v, \Upsilon$	VC dimension
$x$	Scalar variable used for defining functions
$\mathbf{x}$	Input vector
$\mathbf{X}$	Input data matrix containing $m$ samples and $n$ features
$\mathbf{x}_i$	Feature vector of $i$ th sample
$\mathbf{x}_j$	Vector containing $j$ th feature values of all samples
$x_{ij}$	$j$ th feature value of $i$ th sample in $\mathbf{X}$
$\mathbf{y}$	Class label vector corresponding to $\mathbf{X}$
$y_i$	Class label of $i$ th sample
$\mathbf{O}$	Desired output matrix in neural network
$\mathbf{o}_i$	Desired output vector of $i$ th sample
$o_{ij}$	$j$ th feature value of $i$ th sample in $\mathbf{O}$
$i$	Data sample number
$j$	Feature number
$m$	Number of data samples
$n$	Number of features
$p$	$p$ th class in classification problem
$P$	Number of classes in classification problem
$\mathbf{w}$	Weight vector
$\mathbf{W}$	Weight matrix
$b$	Bias variable
$\alpha, \beta, \lambda$	Classifier and statistical test related parameters
$\zeta_i$	$i$ th Lagrange multiplier
$C$	Cost value in $C$ -SVM
$\gamma$	RBF kernel parameter
$\xi_i$	Slack variable of $i$ th sample in $C$ -SVM

$s_i$	Membership value of $i$ th sample for FSVM
$\Delta$	Margin between two classes w.r.t. a hyper-plane
$\delta$	A small value above 0
$r$	Radius of sphere and ellipsoid
$d$	Distance metric
$\varepsilon$	Radius of $\varepsilon$ neighborhood
$\tau$	Threshold value
$g$	Fuzzy rule number i.e. $g$ th fuzzy rule
$G$	Number of fuzzy rules
$k$	$k$ th cluster or fuzzy region
$K$	Number of clusters and number of fuzzy regions
$l$	$l$ th neuron layer in ANN
$L$	Number of layers in ANN
$\rho$	Sparsity parameter for sparsity constraint
$N$	Number of datasets
$\sigma$	Standard deviation
$\nu$	Mean
$\theta$	Parameters of GMM
$\mu$	Membership function
$E$	Number of epochs
$f$	Generic function
$R$	Risk function
TSDSAE	Function representing the training of SDSAE



# List of Figures

Fig. 1.1	Figures showing <b>a</b> sample classification problem, <b>b</b> multiple solutions available for classification, <b>c</b> sample regression problem, and <b>d</b> multiple solutions available for regression	4
Fig. 1.2	Figures showing that VC dimension for linear classifiers in 2D space is 3 [8]	5
Fig. 2.1	Classifier found using Hard Margin SVM	13
Fig. 2.2	Linearly non-separable problem becomes linearly separable after non-linear mapping facilitated by kernel trick [20]	14
Fig. 3.1	Illustration of VC dimension of a classifier in $\mathbb{R}$ . Circles and triangles represent two different classes, and the black colored cross represents the classification threshold. <b>a–d</b> show 2 points in $\mathbb{R}$ , being shattered by Point classifiers	26
Fig. 3.2	Illustration of VC dimension of a classifier in $\mathbb{R}$ : <b>a–h</b> show how 3 points in $\mathbb{R}$ cannot be shattered by Point classifiers (encircled cases would not be classified correctly)	26
Fig. 3.3	Illustration of <b>a</b> the two real branches of Lambert W function $W(x)$ , <b>b</b> lower bound for the VC dimension $\Upsilon$ of MVP classifier, <b>c</b> upper bound for the VC dimension $\Upsilon$ of MVP classifier, and <b>d</b> comparing the VC dimension of linear classifiers with the upper bound on VC dimension of MVP classifier; it indicates that the MVP classifier could statistically exhibit lower variance than that of linear classifiers	28
Fig. 3.4	Class prediction by point classifier does not depend on exact magnitudes	30
Fig. 3.5	Two instances with different input combinations, but same MVP prediction	30
Fig. 3.6	Plot of the ratio $\Pi_{H_{MV}}(m)/2^m$ versus number of samples $m$ . Each graph deviates from 1 at $m = 3$	31

Fig. 4.1	Generic process flow for condition-based monitoring .....	43
Fig. 4.2	Sample data acquisition setup for condition-based monitoring .....	45
Fig. 4.3	Training data organization as datasets .....	45
Fig. 4.4	<b>a</b> Flow of pre-processing and <b>b</b> clipping process with 50% overlap .....	46
Fig. 4.5	Histogram plots of sampled signal values of recording <b>a</b> before pre-processing and <b>b</b> after pre-processing .....	48
Fig. 4.6	<b>a</b> Plots of signal at every step of pre-processing and <b>b</b> Spectrogram plots of signal before and after pre-processing .....	50
Fig. 4.7	Sample feature graph for feature 30 .....	55
Fig. 4.8	Feature graphs: <b>a</b> has high ZCR value and <b>b</b> has low ZCR value .....	56
Fig. 4.9	Feature graphs: <b>a</b> has high SC value and <b>b</b> has low SC value ...	56
Fig. 4.10	Feature graph: <b>a</b> has high RM value and <b>b</b> has low RM value ...	57
Fig. 4.11	Feature graph: <b>a</b> has high $NDM$ value and <b>b</b> has low $NDM$ value .....	57
Fig. 4.12	Feature graph: <b>a</b> has high $SDM$ value and <b>b</b> has low $SDM$ value .....	58
Fig. 4.13	Flowchart for feature ranking with $Z$ graphical indices .....	60
Fig. 4.14	Flowchart for dataset rejection with $Z$ graphical indices .....	62
Fig. 4.15	Process flow for feature selection .....	63
Fig. 4.16	Data organization needed for sensitive position analysis .....	64
Fig. 4.17	Process flow for feature selection .....	65
Fig. 4.18	Figures showing <b>a</b> ELGI LG-02-100 reciprocating-type air compressor, <b>b</b> sensor positions on top of piston, <b>c</b> sensor positions on one side of compressor, and <b>d</b> sensor positions from second side of piston .....	67
Fig. 4.19	Figures showing <b>a–d</b> plots of selected features and <b>e–f</b> plots of features that are not selected during LIV fault detection .....	68
Fig. 4.20	Figures showing selected feature plots where recordings taken during pressure range (6–8) $N/m^2$ show better separation for most datasets .....	70
Fig. 4.21	Figures showing <b>a–d</b> plots of selected features with yellow column showing the identified corrupt dataset and <b>e–f</b> plots of features that are not selected, for recordings acquired in (6–8) $N/m^2$ air pressure range .....	71
Fig. 5.1	Figures showing <b>a</b> separable classification problem, <b>b</b> convex hulls with hard margin SVM, <b>c</b> non-separable classification problem with outlier, <b>d</b> convex hulls for $C$ -SVM with $C = 0.3$ , <b>e</b> for FSVM with $\mu_{lin}^{cen}$ and $C = 1$ , and <b>f</b> for FSVM with $\mu_{hd\_lin}^{prop}$ and $C = 1$ .....	83
Fig. 5.2	Visual depiction of $pt$ -set distance and $HD_{pt}$ distance .....	86

Fig. 5.3	Visual depiction of the two definitions of Hausdorff distance: <b>a</b> refers to first definition, <b>b</b> refers to second definition	88
Fig. 5.4	Outlier detection with <b>a</b> DBSCAN-based approach on class 2 of a toy dataset, <b>b</b> DBSCAN-based approach on class 1 and class 2 of the toy dataset, <b>c</b> OPTICS-OF on class 2 of the toy dataset, <b>d</b> OPTICS-OF on class 1 and class 2 of the toy dataset	90
Fig. 5.5	Plot that gives an idea of desired MF	92
Fig. 5.6	Visually depicting the effectiveness of proposed methodology: <b>a</b> non-linear and non-separable classification problem, <b>b</b> possible outliers and non-outliers identified using proposed strategy with DBSCAN	94
Fig. 5.7	Plots of a few membership functions for negative class of Haberman dataset: <b>a</b> $\mu_{exp}^{hyp}$ , <b>b</b> $\mu_{ptset\_exp}^{prop}$ , and <b>c</b> $\mu_{hyp\_exp}^{prop}$	95
Fig. 6.1	Learning rules for FRC where each box represents a rule	114
Fig. 6.2	Training layer by layer to give Stacked Autoencoder	118
Fig. 6.3	Plots showing the pre-processing functions on first feature of Sonar dataset with <b>a</b> Strategy I, <b>b</b> Strategy II, and <b>c</b> Strategy III	122
Fig. 6.4	Learning rules for FRC	128
Fig. 6.5	Expected effect of fine-tuning strategy I <b>a</b> feature data of samples before and after fine-tuning, <b>b</b> a possible fuzzy box arrangement after fine-tuning	131
Fig. 6.6	Desired rule generation scheme for Fine-Tuning 1	133
Fig. 6.7	Expected effect of fine-tuning strategy II: <b>a</b> feature data of samples before and after fine-tuning, and <b>b</b> a possible fuzzy box arrangement after fine-tuning	133
Fig. 6.8	<b>a</b> Framework for pre-processing, feature generation with SAE and fine-tuning, <b>b</b> flowchart of training phase, and <b>c</b> flowchart of test phase	137
Fig. 7.1	Flowchart of proposed approach	151
Fig. 7.2	Illustrating the median-based fine-tuning approach	155
Fig. A.1	ROC curves on Transfusion data for <b>a</b> C-SVM, <b>b</b> classifier from FSVM, <b>c</b> classifier from FSVM-OD, and <b>d</b> classifier from FSVM-ODSM	163
Fig. A.2	ROC curves on Ecoli data for <b>a</b> C-SVM, <b>b</b> classifier from FSVM, <b>c</b> classifier from FSVM-OD, and <b>d</b> classifier from FSVM-ODSM	164

# List of Tables

Table 3.1	Training and test accuracies for LIV fault detection . . . . .	36
Table 3.2	Test accuracies for prostate tumor data on each cross-validation set (in %) . . . . .	37
Table 4.1	Features selected for LIV fault detection . . . . .	67
Table 4.2	Features selected for LOV fault detection . . . . .	70
Table 4.3	Offline testing results for acoustic-based LIV fault detection . . . . .	72
Table 4.4	Offline testing results for vibration-based LOV fault detection . . . . .	72
Table 4.5	Results on real-time testing . . . . .	73
Table 5.1	Details of imbalanced datasets . . . . .	96
Table 5.2	Details of balanced datasets . . . . .	99
Table 5.3	Performance results for imbalanced datasets (All values in %) . . . . .	103
Table 5.4	Performance results for balanced datasets (All values in %) . . .	105
Table 5.5	Time complexity in calculating MFs . . . . .	107
Table 6.1	Details of experimented datasets . . . . .	139
Table 6.2	Detailed results with proposed methods (in %) . . . . .	141
Table 6.3	Comparing average test errors with literature (in %) . . . . .	142
Table A.1	Best performance values—Analysis I . . . . .	160
Table A.2	Ranking of classifiers—Analysis I . . . . .	160
Table A.3	AUC performance values—Analysis I . . . . .	161
Table A.4	Ranking of classifiers w.r.t. AUC—Analysis I . . . . .	162
Table A.5	Statistical comparison of $\mu_{FCM}$ and $\mu_{FCM}^{prop}$ . . . . .	165
Table A.6	Statistical comparison of set measures . . . . .	166