Meta-cognitive sequential learning in RBF network for diagnosis of neurodegenerative diseases

Giduthuri Sateesh Babu

2014

Giduthuri Sateesh Babu. (2014). Meta-cognitive sequential learning in RBF network for diagnosis of neurodegenerative diseases. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/61782>

<https://doi.org/10.32657/10356/61782>

META-COGNITIVE SEQUENTIAL LEARNING IN RBF NETWORK FOR DIAGNOSIS OF NEURODEGENERATIVE DISEASES

A thesis submitted to
School of Computer Engineering
Nanyang Technological University

by

Giduthuri Sateesh Babu

in fulfilment of the requirements
for the degree of Doctor of Philosophy

2014

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Suresh Sundaram for his intelligent guidance and patient nurture for the past years. I have learned so much from his ways of critical thinking and his analytic insights into the problems helped greatly in the accomplishment of this research work. I am proud to have such a great mentor on my way towards research and he has made my stay at Nanyang Technological University a truly valuable experience.

I want to thank Dr. R. Savitha and Dr. B. S. Mahanand for their numerous helpful comments and enlightening discussions throughout my research course. Their time and assistance were a valuable contribution to this research work.

I would also like to dedicate this special thanks to my family and friends, especially my parents, who have always been there for me. This research work would not have been possible without their love and encouragements.

Thanks are also dedicated to Center for Computational Intelligence for the research facilities. I also acknowledge Nanyang Technological University for the financial support and this precious opportunity of study. Finally, I pay my tributes to the God Almighty for blessing me in all my endeavors.

Contents

Acknowledgements	i
Abstract	vi
List of Figures	viii
List of Tables	x
List of Abbreviations	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	5
1.3 Thesis Contributions	7
1.4 Thesis Organization	10
2 Literature Review on Sequential Learning Algorithms in Neural Networks	13
2.1 Sequential Learning Algorithms	14
2.1.1 Error Driven Algorithms	14
2.1.2 Neuron Significance Based Algorithms	15
2.1.3 Extreme Learning Machine Based Algorithms	16
2.1.4 Spiking Neural Networks Algorithms	17
2.1.5 Incremental-Decremental SVM Algorithms	18
2.1.6 Kernel Least Mean Square Based Algorithms	18
2.1.7 Sequential Classification Algorithms	19
2.2 Summary	21

3	An Overview on Meta-cognition	22
3.1	Definitions of Important Concepts in Meta-cognition	22
3.2	Models of Meta-cognition	23
3.3	Motivation for Meta-cognitive Learning	24
3.4	Summary	25
4	Meta-cognitive Radial Basis Function Network and Its EKF Based Sequential Learning Algorithm for Classification Problems	26
4.1	Introduction	26
4.2	Classification Problem Definition	26
4.3	EKF-McRBFN Classifier	27
4.3.1	Cognitive Component of EKF-McRBFN	28
4.3.2	Meta-cognitive Component of EKF-McRBFN	29
4.4	EKF-McRBFN Algorithm	39
4.4.1	Guidelines for EKF-McRBFN Thresholds Initialization	39
4.5	Summary	42
5	Projection Based Learning Algorithm for Meta-cognitive RBF Network Classifier	44
5.1	Introduction	44
5.2	PBL-McRBFN Classifier	45
5.2.1	Cognitive Component of PBL-McRBFN	45
5.2.2	Meta-cognitive Component of PBL-McRBFN	47
5.3	PBL-McRBFN Algorithm	54
5.4	Salient Features of PBL-McRBFN Algorithm	54
5.5	Summary	57
6	Performance Evaluation of EKF-McRBFN and PBL-McRBFN Classifiers	58
6.1	Data Sets Description	58
6.2	Simulation Environment	60
6.3	Performance Measures	61

6.3.1	Statistical Significance Test	61
6.4	Performance Evaluation	62
6.4.1	Binary-class Data Sets	62
6.4.2	Multi-category Data Sets	64
6.4.3	Statistical Performance Comparison	66
6.4.4	10 Random Trial Results	69
6.5	Work Flow of Meta-cognitive Strategies	74
6.6	Study on the Effect of Meta-cognition	78
6.7	Summary	81
7	Alzheimer's Disease Diagnosis using PBL-McRBFN Classifier	82
7.1	Literature Review on Alzheimer's Disease	83
7.1.1	Region-of-Interest Approach	84
7.1.2	Whole Brain Morphometric Approach	85
7.2	Early Diagnosis of Alzheimer's Disease Based on MRI features	86
7.2.1	Materials	88
7.2.2	Voxel Based Morphometry Based Feature Extraction	90
7.2.3	Experimental Results	95
7.2.4	PBL-McRBFN Classifier Performance on the OASIS Data Set . .	96
7.2.5	PBL-McRBFN Classification Performance on the ADNI Data Set	98
7.2.6	Generalization Capability of the PBL-McRBFN Classifier for the Detection of AD	101
7.3	Identification of Imaging Biomarkers for AD	103
7.3.1	Imaging Biomarkers for AD in Complete OASIS Data Set	105
7.3.2	Imaging Biomarkers for AD Based on Age in OASIS Data Set . .	107
7.3.3	Imaging Biomarkers for AD Based on gender in OASIS Data Set .	110
7.4	Discussion	114
7.5	Summary	115

8	Parkinson's Disease Diagnosis using PBL-McRBFN Classifier	116
8.1	Literature Review on Parkinson's Disease	117
8.2	Materials	119
8.2.1	Microarray Gene Expression Data Set	119
8.2.2	MRI Data Set	120
8.2.3	Vocal Data Set	120
8.2.4	Gait Data Set	121
8.3	Early Diagnosis of Parkinson's Disease Based on Gene Expression Features	121
8.3.1	p -value Based Gene Selection	121
8.3.2	ICA Based Feature Reduction	122
8.3.3	Performance of PBL-McRBFN on ICA Reduced Features from Complete Genes	123
8.3.4	Performance of PBL-McRBFN on ICA Reduced Features from Sta- tistically Selected Genes	124
8.4	Early Diagnosis of Parkinson's Disease Based on MRI features	126
8.4.1	VBM Based Feature Extraction	126
8.4.2	Performance of PBL-McRBFN on VBM Features	127
8.4.3	Performance of PBL-McRBFN on Reduced features	129
8.4.4	Identification of Imaging Biomarkers for PD	130
8.5	PD Diagnosis Based on Vocal Features	133
8.6	PD Diagnosis Based on Gait Features	133
8.7	Summary	136
9	Conclusions and Future Works	137
9.1	Conclusions	137
9.2	Future Works	141
9.2.1	Plan of Work for McRBFN	141
9.2.2	Applications	142
	Publications	144
	References	146

Abstract

This research work focuses on the development of meta-cognitive sequential learning algorithms in Radial Basis Function (RBF) network classifiers, and their application to the early diagnosis of neurodegenerative diseases. The important issues in existing sequential learning algorithms are proper selection of training samples, finding minimal network structure and selection of an appropriate learning strategy. In addition, the random sequence of sample arrival influences the performance significantly. It has been reported in human learning that best learning strategies employ meta-cognition (meta-cognition means *cognition about cognition*) to address fundamental problems of *what-to-learn*, *when-to-learn* and *how-to-learn*. This thesis develops such meta-cognitive sequential learning algorithms in RBF network for classification problems. We call a RBF network employing meta-cognitive algorithm as ‘meta-cognitive RBF network’ (McRBFN).

McRBFN is developed based on Nelson and Narens model of meta-cognition for human learning. Accordingly, McRBFN has two components, namely cognitive and meta-cognitive components. A RBF network with evolving structure is the cognitive component and a self-regulatory learning mechanism is its meta-cognitive component. The meta-cognitive component controls the learning of cognitive component by choosing suitable learning strategies for each sample. When a new sample is presented, the meta-cognitive component either deletes the sample or learns the sample or reserves the sample for future use. Learning includes adding a new neuron or updating the parameters of the existing neurons using an extended Kalman filter (EKF). The McRBFN using EKF for parameter updates are referred as ‘EKF-McRBFN’.

EKF-McRBFN uses computationally intensive EKF based parameter update and does not utilize the past knowledge stored in the network. Therefore, an efficient Projection Based Learning (PBL) algorithm for McRBFN referred as PBL-McRBFN has been developed. When a neuron is added to the cognitive component, the Gaussian parameters

are determined based on the current sample and the output weights are estimated using the PBL algorithm. When a new neuron is added, existing neurons in the cognitive component will be used as pseudo-samples in PBL. There-by, the proposed algorithm exploits the knowledge stored in the network for proper initialization.

The performance of EKF-McRBFN and PBL-McRBFN has been evaluated using a number of benchmark classification problems. The statistical performance comparisons on multiple data sets clearly indicate the superior performance of the proposed PBL-McRBFN and EKF-McRBFN over existing popular classifiers. Experimental results also show that PBL-McRBFN performance is better than EKF-McRBFN classifier.

Another significant contribution of this thesis is in early diagnosis of neurodegenerative diseases. In this thesis, we employed PBL-McRBFN to early diagnosis of Alzheimer's disease (AD) and Parkinson's disease (PD).

The early diagnosis of AD problem from Magnetic Resonance Imaging (MRI) scans is formed as a binary classification problem. The performance of the PBL-McRBFN classifier has been evaluated on two well-known open access Open Access Series of Imaging Studies (OASIS) and Alzheimer's disease Neuroimaging Initiative (ADNI) data sets. Morphometric features are extracted from MRI scans using Voxel-Based Morphometry (VBM). The study results clearly show that the PBL-RBFN classifier produces a better generalization performance compared to the state-of-the-art AD detection results. Also, generalization conducted on ADNI data set with PBL-McRBFN classifier trained on OASIS data set shows that the proposed PBL-McRBFN can also achieve significant results on the unseen data set. Finally, PBL-McRBFN-RFE feature selection approach has been proposed to detect imaging biomarkers responsible for AD for different age groups and for both genders using OASIS data set.

The early diagnosis of PD problem is also formed as a binary classification problem. PBL-McRBFN classifier is used to predict PD using microarray gene expression data. Next, PBL-McRBFN classifier is used to predict PD from MRI scans. Further, imaging biomarkers responsible for PD are detected with the proposed PBL-McRBFN-RFE approach based on MRI scans. For completeness, PBL-McRBFN classifier is also used to detect PD from vocal and gait features. From the performance evaluation study, it is evident that the generalization performance of proposed PBL-McRBFN classifier is better than the state-of-the-art PD detection results.

List of Figures

1.1	Nelson and Narens model of meta-cognition	2
3.1	Nelson and Narens model of meta-cognition	24
4.1	(a) Nelson and Narens model of meta-cognition (b) EKF-McRBFN model	27
4.2	Schematic diagram of EKF-McRBFN	28
4.3	Cognitive component: RBF network	29
4.4	Schematic representation of training samples corresponding to overlapping/no-overlapping conditions	36
4.5	Error regions of various thresholds in EKF-McRBFN	41
6.1	Exemplification of sample deletion strategy in PBL-McRBFN for Image segmentation data set	75
6.2	Class-wise significance (a), and instantaneous hinge error with self-regulatory thresholds (b) in PBL-McRBFN for Image segmentation data set	76
6.3	History of number of hidden neurons (a), self-regulated addition (b), and update thresholds (c) in PBL-McRBFN for Image segmentation data set	77
7.1	Schematic diagram of the AD detection using PBL-McRBFN classifier . .	87
7.2	Schematic diagram of the stages in feature extraction based on the VBM analysis	90
7.3	Results of the unified segmentation and smoothing steps performed on MRI of an AD patient (from right: sagittal view, coronal view and axial view)	92
7.4	Maximum intensity projections from OASIS data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view	93

7.5	Maximum intensity projections from ADNI data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view	93
7.6	Gray matter volume change from OASIS data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view	94
7.7	Gray matter volume change from ADNI data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view	95
7.8	Schematic representation of the generalization capability study of OASIS trained PBL-McRBFN classifier on ADNI data set.	102
7.9	Comparison of gray matter volume change - Normal persons vs. AD patients from complete OASIS data set	106
7.10	Comparison of gray matter volume change - Normal persons vs. AD patients from 60-69 (a&b), 70-79 (c&d), 80-Above (e&f) age groups in OASIS data set	109
7.11	Comparison of gray matter volume change - Normal persons vs. AD patients from Male-OASIS data set	112
7.12	Comparison of gray matter volume change - Normal persons vs. AD patients from Female-OASIS data set	113
8.1	PBL-McRBFN classifier on ICA reduced features from: (a) Complete genes, (b) Selected genes	122
8.2	Schematic diagram of the stages in feature extraction based on the VBM analysis.	127
8.3	Maximum intensity projections from PPMI MRI data set - Normal persons vs. PD patients (a) sagittal view (b) coronal view (c) axial view	128
8.4	Gray matter volume change from PPMI MRI data set - Normal persons vs. PD patients (a) sagittal view (b) coronal view (c) axial view	129
8.5	Comparison of gray matter volume change - Normal persons vs. PD patients in Superior temporal gyrus region	132

List of Tables

2.1	Comparison of supervised sequential learning algorithms	20
6.1	Specification of benchmark binary and multi class data sets	59
6.2	Performance comparison of PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM on binary class data sets	63
6.3	Performance comparison on multi-category data sets	65
6.4	Ranks based on the overall testing efficiency (η_o)	67
6.5	Two-tailed critical values (F -distribution) for the Friedman test at 95% confidence level	67
6.6	Critical values for the Benferroni-Dunn test at 95% confidence level . . .	67
6.7	Ranks based on the average testing efficiency (η_a)	69
6.8	PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM classifiers 10 ran- dom trial results comparison on binary class data sets	70
6.9	10 random trial results comparison on multi-category data sets	72
6.10	One-tailed critical values (F -distribution) for the ANOVA test at 95% confidence level	73
6.11	Two-tailed critical values (t -distribution) for the Dunnett test at 95% con- fidence level	73
6.12	Effect of Meta-cognitive learning principles in the QKLMS algorithm . .	80
7.1	Demographic information of OASIS data used in our study	88
7.2	Demographic information of ADNI data used in our study	89
7.3	Classification performance of PBL-McRBFN on the OASIS data set . . .	96
7.4	Performance comparison with existing results on the OASIS data set . .	97
7.5	Classification performance of PBL-McRBFN on the ADNI Dataset . . .	99

7.6	Performance comparison with existing results on the ADNI data set . . .	100
7.7	Generalization performance of PBL-McRBFN classifier on unseen ADNI samples	101
7.8	VBM detected and PBL-McRBFN-RFE selected regions from complete OASIS data set	105
7.9	PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on complete OASIS data set	105
7.10	Generalization performance of PBL-McRBFN classifier on unseen ADNI samples with selected 906 features	106
7.11	VBM detected and PBL-McRBFN-RFE selected regions from age-wise OASIS data sets	108
7.12	PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on age-wise OASIS data sets	108
7.13	VBM detected and PBL-McRBFN-RFE selected regions from male-OASIS data set	111
7.14	PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on male-OASIS data set	111
7.15	VBM detected and PBL-McRBFN-RFE selected regions from female-OASIS data set	112
7.16	PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on female-OASIS data set	113
8.1	Demographic information of PPMI MRI data used in our study	120
8.2	Performance comparison on complete gene expression data set from an average of 10 trials	123
8.3	Performance comparison on selected gene expression data set with p -value < 0.05 from an average of 10 trials	124
8.4	Performance comparison on selected gene expression data set with p -value < 0.01 from an average of 10 trials	125
8.5	Performance comparison on 2981 VBM features data set from an average of 10 trials	128

8.6	Performance comparison on ICA reduced features data sets from an average of 10 trials	130
8.7	VBM detected and PBL-McRBFN-RFE selected regions responsible for PD131	
8.8	PBL-McRBFN classifier performance on VBM detected and PBL-McRBFN-RFE selected features from an average of 10 trials	132
8.9	Performance comparison on vocal data set from an average of 10 trials . .	133
8.10	PBL-McRBFN classifier performance comparison with studies in the literature on vocal data set	134
8.11	Performance comparison on gait data set from an average of 10 trials . .	135
8.12	PBL-McRBFN classifier performance comparison with studies in the literature using gait patterns	135

List of Abbreviations

RBF	Radial Basis Function
EKF	Extended Kalman Filter
RAN	Resource Allocation Network
RANEKF	Resource Allocation Network-Extended Kalman Filter
MRAN	Minimal Resource Allocation Network
EMRAN	Extended Minimum Resource Allocation Network
GAP-RBFN	Growing And Pruning Radial Basis Function Network
GGAP-RBFN	Generalized Growing And Pruning Radial Basis Function Network
FGAP-RBFN	Fast Growing And Pruning Radial Basis Function Network
SRAN	Self-adaptive Resource Allocation Network
SMC-RBFN	Sequential Multi-Category Radial Basis Function Network
ELM	Extreme Learning Machine
OS-ELM	On-line Sequential Extreme Learning Machine
I-ELM	Incremental Extreme Learning Machine
CI-ELM	Convex Incremental Extreme Learning Machine
EI-ELM	Enhanced Incremental Extreme Learning Machine

SNN	Spiking Neural Networks
SVM	Support Vector Machine
SIDSVM	Single Incremental Decremental Support Vector Machine
MIDSVM	Multiple Incremental Decremental Support Vector Machine
LMS	Least Mean Squares
RKHS	Reproducing Kernel Hilbert Space
KLMS	Kernel Least Mean Square
KLMS-CG	Kernel Least Mean Square with Constrained Growth
QKLMS	Quantized Kernel Least Mean Square
QKLMS-FB	Quantized Kernel Least Mean Square with Fixed-Budget
McRBFN	Meta-cognitive Radial Basis Function Network
EKF-McRBFN	Extended Kalman Filter based Meta-cognitive Radial Basis Function Network
PBL	Projection Based Learning
PBL-McRBFN	Projection Based Learning Meta-cognitive Radial Basis Network
RFE	Recursive Feature Selection
PBL-McRBFN- RFE	Projection Based Learning Meta-cognitive Radial Basis Network with Recursive Feature Selection
AD	Alzheimer's disease
PD	Parkinson's disease

MCI	Mild Cognitive Impairment
MRI	Magnetic Resonance Imaging
VBM	Voxel-Based Morphometry
SPM	Statistical Parametric Mapping
LD	Liver Disorders
PIMA	Pima Indian diabetes
BC	Breast Cancer
HEART	Heart disease
ION	Ionosphere
IRIS	Iris classification
IS	Image segmentation
WINE	Wine determination
AE	Acoustic Emission classification
VC	Vehicle Classification
GI	Glass Identification
GCM	Global Cancer Mapping using micro-array gene expression
LETTER	Letter recognition
SI	Satellite Image classification
LAND	Landsat Satellite
ANOVA	Analysis of Variance

CT	Computed Tomography
SPECT	Single-Photon Emission Computed Tomography
PET	Positron Emission Tomography
TCS	Transcranial Brain Sonography
CDR	Clinical Dementia Rating
MMSE	Mini-Mental State Examination
MPRAGE	Magnetization-Prepared Rapid-Acquisition Gradient Echo
ADNI	Alzheimer's Disease Neuroimaging Initiative
OASIS	Open Access Series of Imaging Studies
PPMI	Parkinson's Progression Markers Initiative

Chapter 1

Introduction

1.1 Motivation

Over the past decade, a number of supervised learning algorithms have been developed for pattern classification applications. Artificial neural networks have been widely used in the fields of pattern classification and they show their advantages over other methods in their learning, generalization and adaptation capability as well as their unique power for nonlinear mapping. In most of the practical applications especially in medical diagnosis, the complete training data describing the input-output relationship is not available a priori. For these problems, classical batch-learning algorithms are rather infeasible and instead sequential learning is employed [1].

In a sequential learning framework, the training samples arrive one-by-one and the samples are discarded after the learning process. Hence, it requires less memory and computational time during the learning process. In addition, sequential learning algorithms automatically determine the architecture that can accurately approximate the true decision function described by a stream of training samples. Samples from data stream do not follow the same static underlying distribution. Traditionally all sequential learning algorithms uses all training samples for learning and does not regulate the learning process. This inspires us to study human learning and develop learning algorithm which mimic the best learning strategy in neural networks.

In the families of artificial neural networks, Radial Basis Function (RBF) neural networks have been extensively used in a sequential learning framework due to its universal approximation ability and simplicity of architecture. Hence, in this thesis, we consider

RBF network for classification problems. Many sequential learning algorithms in RBF framework are available in the literature to solve classification problems and detailed review on these algorithms is provided in chapter 2.

On the other hand, educational psychologists have studied human learning for years and suggested that the learning process is effective when the learners adopt self-regulation in learning process using meta-cognition [2, 3, 4]. Cognition is defined as a group of symbolic mental activities and mental representations that includes attention, memory, producing and understanding language, learning, reasoning, problem solving, and decision making. Meta-cognition means *cognition about cognition*. The term meta-cognition was first coined by Flavell [5]. He defined meta-cognition as the *thoughts about one's own thoughts process and cognitions*. Precisely the learner should control the learning process, by planning and selecting learning strategies and monitor their progress by analyzing the effectiveness of the proposed learning strategies [6]. When necessary, these strategies should be appropriately adapted. Meta-cognition present in human-being provides a means to address *what-to-learn*, *when-to-learn* and *how-to-learn*, i.e., the ability to identify the specific piece of required knowledge, judge when to start and stop learning by emphasizing best learning strategy.

There are several meta-cognition models available in human psychology and a brief survey of various meta-cognition models is reported in [7]. Among the various models, the model proposed by Nelson and Narens in [8] is simple and clearly highlights the various actions in human meta-cognition as shown in Fig. 1.1.

Figure 1.1: Nelson and Narens model of meta-cognition

Nelson and Narens [8] model has two components, the cognitive component and the meta-cognitive component. The information flow from the cognitive component to meta-cognitive component is considered “monitoring”, while the information flow in the reverse direction is considered “control”. The basic notion underlying control is that the meta-cognitive component modifies the cognitive component based on these signals. The information flowing from the meta-cognitive component to the cognitive component either changes the state of the cognitive component or changes the cognitive component itself. “Monitoring” informs the meta-cognitive component about the state of cognitive component, thus continuously updating the meta-cognitive component’s model of cognitive component, including, ‘no change in state’.

In neural networks, the current state-of-the algorithms address pure cognitive aspect of human learning inspired from human brain, the concept of self-regulated learning using meta-cognition is not exploited. Self-regulated learning refers to the ability of a learning system to decide *what-to-learn when-to-learn* and *how-to-learn*. The current state-of-the-art neural network algorithms address only the *how-to-learn* component of human learning. In the neural networks field, a few algorithms have been developed which address some of the meta-cognitive learning aspects [9, 10]. One of the first works in neural networks to deal with meta-cognition is Self-adaptive Resource Allocation Network (SRAN) [9]. SRAN is a sequential learning algorithm and also addresses the *what-to-learn* component of meta-cognition by selecting significant samples using misclassification error and hinge loss error. The complex version of above algorithm is Complex-valued Self-regulating Resource Allocation Network (CSRAN) [10]. It has been shown in [9, 10], that selecting significant samples and removing repetitive samples in learning helps to improve the generalization performance. Therefore, it is apparent that emulating the three components of meta-cognition with suitable learning strategies would improve the generalization ability of a neural network. The drawbacks in the above algorithms are:

- The selection of significant samples from stream of training data are based on simple error criterion which is not sufficient.
- The allocation of new hidden neuron center without considering the amount of overlap with already existing neuron centers leads to misclassification.

- Knowledge gained from past trained samples is not utilized in further learning.
- These algorithms use computationally intensive Extended Kalman Filter (EKF) for parameter update.

Hence, there is a need to develop a learning algorithm which automatically selects appropriate samples for learning and adopt best learning strategy to learn them accurately. This thesis shall deal with the development of meta-cognitive sequential learning algorithm for RBF network classifiers which overcomes the above drawbacks. Also, the interaction between meta-cognitive and cognitive components and their influence on RBF network learning will be dealt with, accordingly. We evaluate the performance of the developed classifier and compare with existing classifiers. Also, we use proposed classifier to detect neurodegenerative diseases.

Another important motivation of the thesis is to identify imaging biomarkers for early diagnosis of neurodegenerative diseases. Neurodegenerative diseases are generally considered as a group of diseases that seriously and progressively impair the functions of the nervous system through selective neuronal vulnerability of specific brain regions. Depending on their type, neurodegenerative diseases can be serious or life-threatening and most of them have no cure. The goal of treatment for such diseases is usually to improve symptoms, relieve pain and increase mobility. In modern world, management and the development of computerized systems of patient care for chronic neurodegenerative diseases is critical due to growing data sets. Also, the sample imbalance in the classes of neurodegenerative problems makes machine learning techniques difficult to solve using them. Alzheimer's disease (AD) is the most common neurodegenerative disease and is one of the leading cause of death worldwide with the associated estimated cost of care exceeding \$200 billion annually [11]. It is estimated that there are more than 5.4 million Americans and about 30 million people worldwide suffering from AD, with the number expected to increase dramatically as the global population ages. Today, AD remains the largest unmet medical need in neurology, with the disease expected to afflict 100 million by 2050.

Parkinson's disease (PD) is the second most common neurodegenerative disease, after AD. It is estimated that there are more than 1 million Americans and about 10 million

people worldwide living with PD. Incidence of PD increases with age, but an estimated four percent of people with PD are diagnosed before the age of 50 [12].

The early diagnosis of most common neurodegenerative diseases particularly AD and PD using non-invasive brain imaging techniques such as Magnetic Resonance Imaging (MRI) will help to slow down the progress of diseases. MRI is the most important brain imaging procedure that provides accurate information about the shape and volume of the brain. MRI accurately monitors and identifies the tissue volume changes at all anatomical regions in the brain. MRI helps to detect AD at an early stage - before irreversible damage has been done [13]. As with the increased number of elderly people, there will be many cases of AD and PD, the databases of clinical study of these diseases are also increasing. Also, identifying the most relevant and meaningful imaging biomarkers with a predictive power for neurodegenerative diseases detection is important [14]. Hence, there is a need to develop algorithms which can handle these growing medical data sets and identification of imaging biomarkers to predict these neurodegenerative diseases in the early stages. This thesis shall also deal with handling these growing medical data sets for detection of AD and PD, and identification of imaging biomarkers responsible for AD and PD, which is a challenging task for the machine learning community.

1.2 Objectives

The main aim of this research work is to develop a generic framework of human meta-cognitive learning mechanism in RBF network architecture. The research discusses and evaluates the inter-relationship between the meta-cognitive knowledge, control and monitoring signals such that the learning in RBF is efficient, this in turn improve the performance of RBF classifier significantly. The main research objectives could be summarized as follows:

- **Sequential Meta-cognitive Learning Algorithm for Radial Basis Function Network:** Develop a meta-cognitive learning algorithm for a RBF network based on generic framework of meta-cognition proposed by Nelson and Narens model to handle data one by one and only one. Aspects of meta-cognitive monitoring such as ease-of-learning, judgement-of-learning and feeling-of-knowing need to be explored

in a machine learning framework. The major objectives of meta-cognitive sequential learning algorithm are fast and efficient learning, and must use past knowledge properly for self-regulation of learning. RBF networks is chosen due to its localization property of Gaussian function, and widely used in classification problems. The algorithm must also handle growing data sets of neurodegenerative disease particularly AD and PD. To handle the growing data sets it must handle the samples in sequential mode with less computation effort and also network architecture must be adaptive to the change in data distribution.

- **Alzheimer’s Disease Diagnosis Problem:** AD is the most common neurodegenerative disease. Early diagnosis of AD using non-invasive brain imaging methods plays a major role in providing treatment that may slow down its progress. MRI is the most important brain imaging procedure that provides accurate information of the brain with high spatial resolution and can detect minute abnormalities. Hence, early diagnosis of AD using MRI scans is another objective of this research work. This objective is two-folds, first is AD classification using MRI scans with accurate prediction rate and second is selection of relevant imaging biomarkers for AD using MRI scans. The wrapper based feature selection method which depends on the classification algorithm to detect imaging biomarkers for AD is needed. In medical literature, it is reported that gender and age may be an important modifying factor in AD’s development and expression. Hence, finding imaging biomarkers for AD for different age groups and genders is another objective.
- **Parkinson’s Disease Diagnosis Problem:** PD is the second most common neurodegenerative disease, after AD. In literature, PD diagnosis using machine learning methods had been reported using vocal and gait features. Research studies on PD discovered that, early diagnosis of PD using vocal and gait features is impossible because tremor and slow movements develop in PD patients only after approximately 70% of vulnerable dopaminergic neurons in the substantia nigra have already died. Recent studies on gene expression analysis found that there is a profound change in gene expression for individuals affected by PD. Also, early diagnosis of PD using non-invasive brain imaging techniques such as MRI is needed. In literature, machine learning techniques were not employed to early diagnosis of PD using gene

expression and MRI features. Hence, some other objectives of this research work are

- Early diagnosis of PD using gene expression features.
- Early diagnosis of PD and identification of imaging biomarkers for PD using MRI scans.

1.3 Thesis Contributions

This thesis brings several contributions to the field of neural networks and early diagnosis of neurodegenerative diseases. Major contributions of this thesis are categorized into two parts, (I) Algorithm part and (II) Application part. First, we highlight contributions in algorithm part and then present contributions in applications part.

I. Meta-cognitive Radial Basis Function Network

To incorporate human meta-cognitive principles in neural networks, we implemented a meta-cognitive radial basis function network based on Nelson and Narens model. If a RBF network analyzes its cognitive process and chooses suitable learning strategies adaptively to improve its cognitive process then it is referred to as ‘Meta-cognitive Radial Basis Function Network’ (McRBFN). McRBFN has two components, namely the cognitive component and the meta-cognitive component. A RBF network with evolving structure is the fundamental building block of the cognitive component. The meta-cognitive component devices sample deletion, neuron growth, parameter update and sample reserve strategies which directly addresses the basic principles of self-regulated human learning (i.e., *what-to-learn*, *when-to-learn* and *how-to-learn*). The meta-cognitive part controls the sequential learning process by selecting one of the above learning strategies for the new training sample. The strategies are also adapted to accommodate coarse knowledge first followed by fine tuning. Sample delete strategy removes redundant sample to avoid over-training.

- (1) **EKF Based Meta-cognitive Radial Basis Function Network:** Extended Kalman Filter (EKF) based sequential learning algorithm has been proposed for McRBFN

referred as EKF-McRBFN for classification problems. The novel contributions from this algorithm are listed below

- Hinge loss error function is used for better estimation of the posterior probability.
- Spherical potential is used to calculate the class-wise significance of sample.
- Overlapping criteria are introduced to initialize the new hidden neuron parameters.
- EKF is used to estimate the network parameters.

(2) **Projection Based Learning Meta-cognitive Radial Basis Function Network:** Above EKF based sequential algorithm uses computationally EKF parameter update and does not utilize the past knowledge stored in the network. Hence, less computationally intensive Projection Based Learning (PBL) sequential algorithm has been proposed for McRBFN referred as PBL-McRBFN for classification problems.

- The efficient PBL algorithm is implemented based on the principle of minimization of hinge error function and finds the optimal network output parameters for which the error function is minimum.
- Existing hidden neurons in the network are used as pseudosamples to exploit the past knowledge.

II. Early Diagnosis of Neurodegenerative Diseases

Neurodegenerative diseases, of which AD and PD are the most common, take an enormous toll on affected patients and their families. In the application part, the proposed PBL-McRBFN classifier is used in the early diagnosis of AD based on MRI scans. Also, PBL-McRBFN classifier is used in the early diagnosis of PD based on micro-array gene expression features and MRI scans.

3. **Alzheimer's Disease Diagnosis Problem:** PBL-McRBFN classifier is used to solve one of the most common neurodegenerative AD diagnosis problem. The early diagnosis of AD problem from MRI scans formed as a binary classification

problem. For this, morphometric features are extracted from MRI scans using Voxel-Based Morphometry (VBM). VBM is one of the widely used, fully automated, whole brain morphometric analysis. VBM is based on the Statistical Parametric Mapping (SPM) method, often employed for the investigation of tissue volume changes between the brains MRI scans of the diseased group versus the normal persons. We have used VBM analysis to identify probability of the gray matter in a given voxel, where a voxel is defined as a volume element representing the intensity of a point in a three-dimensional space. In this study, the contributions are two-fold

- PBL-McRBFN classifier is used to predict AD from morphometric features set obtained from the VBM analysis.
- Recursive Feature Selection (RFE) is incorporated in PBL-McRBFN and proposed feature selection scheme to identify critical imaging biomarkers relevant to AD using MRI scans called “PBL-McRBFN-RFE”.

The imaging biomarkers identified using PBL-McRBF-RFE are in the parahippocampal gyrus, the hippocampus, the superior temporal gyrus, the insula, the precentral gyrus and the extra-nuclear regions. Next, the PBL-McRBFN-RFE has also been used to identify imaging biomarkers for AD from the OASIS gender-wise and age-wise analysis. In medical literature [15, 16], it is reported that age and gender may be important modifying factors in AD’s development and expression. To verify this, we conducted imaging biomarkers detection analysis based on age and gender.

The results from the imaging biomarkers detection analysis based on age are:

- In the 60-69 age group AD patients, gray matter atrophy is observed in the superior temporal gyrus region which is responsible for processing sounds.
- In the 70-79 age group AD patients, gray matter atrophy is observed in the parahippocampal gyrus and the extra-nuclear regions which are responsible for memory encoding and retrieval.

- In the 80-89 age group AD patients, gray matter atrophy is observed in the hippocampus, the parahippocampal gyrus and the lateral ventricular regions which are responsible for short-term memory to long-term memory, spatial navigation, memory encoding and retrieval.

The results from the imaging biomarkers detection analysis based on gender are:

- In male AD patients, gray matter atrophy is observed in the insula region which is responsible for emotion and consciousness.
- In female AD patients, gray matter atrophy is observed in the parahippocampal gyrus and the extra-nuclear regions which are responsible for memory encoding and retrieval.

4. **Parkinson's Disease Diagnosis Problem:** Another most common neurodegenerative PD diagnosis problem has been handled by employing PBL-McRBFN classifier. The early diagnosis of PD problem is also formed as a binary classification problem. In this study, the contributions are

- PBL-McRBFN classifier is used to predict PD using microarray gene expression features obtained from genes with different significance levels.
- PBL-McRBFN classifier is used to predict PD from morphometric features set obtained from MRI scans using VBM analysis. We also used PBL-McRBFN-RFE approach to identify critical imaging biomarkers relevant to PD using MRI scans. The superior temporal gyrus brain region detected by PBL-McRBFN-RFE imaging biomarkers detection analysis may play more significant role than others in PD.
- PBL-McRBFN classifier is also used to detect PD using the standard vocal and gait PD data sets.

1.4 Thesis Organization

The thesis is organized as follows:

- In **chapter 2**, a literature review on the existing supervised sequential learning algorithms for neural networks and motivation for meta-cognitive learning is presented. In here, sequential learning algorithms are classified based on the framework: error driven, neuron significance, extreme learning machine, spiking neural networks, incremental-decremental SVM algorithms, kernel least mean square and sequential classification algorithms. A brief review of each class of sequential learning algorithms is presented.
- In **chapter 3**, an overview of the human meta-cognition including the main concepts of meta-cognition and models of meta-cognition available in literature are discussed. Next, motivation of meta-cognitive learning is explained.
- In **Chapter 4**, the EKF-McRBFN is proposed and its sequential learning algorithm for classification problems is presented in detail. A self-regulatory sequential learning mechanism which decides *what-to-learn*, *when-to-learn* and *how-to-learn* efficiently is presented. In EKF-McRBFN algorithm, sample delete strategy address the *what-to-learn* by deleting insignificant samples from data stream, neuron growth strategy and parameters update strategy address the *how-to-learn* by which the cognitive component learns from the samples, and self-adaptive nature of meta-cognitive thresholds in addition to the sample reserve strategy address the *when-to-learn* by presenting the samples in the learning process according to the knowledge present in the sample.
- In **Chapter 5**, an efficient sequential PBL-McRBFN classifier is presented. PBL-McRBFN classifier uses computationally less intensive PBL algorithm. PBL algorithm accurately estimates the output weight by direct minimization of error. PBL-McRBFN allows the network to ‘reuse’ the knowledge gained from the past samples in new hidden neuron parameters initialization and output weights estimations.
- In **Chapter 6**, the performance of two developed algorithms EKF-McRBFN and PBL-McRBFN is evaluated on real-world benchmark binary and multi-category classification problems with a wide range of imbalance factor. The performance

of proposed EKF-McRBFN and PBL-McRBFN algorithms is compared with the best performing sequential learning algorithm reported in the literature (SRAN) [17], batch ELM [18] and standard Support Vector Machine (SVM). A quantitative performance analysis, based on number of samples used in training, number of hidden neurons and average/overall testing efficiency is performed. A qualitative performance study based on Friedman test is also conducted.

- In **Chapter 7**, PBL-McRBFN classifier developed in the earlier chapter is employed to solve one of the most common neurodegenerative AD diagnosis problem from MRI scans. In here, morphometric features are extracted from MRI scans using VBM. The performance of the PBL-McRBFN classifier has been evaluated on two well-known open access OASIS and ADNI data sets. The performance of PBL-McRBFN classifier has been compared with other state-of-the-art methods in the literature in AD diagnosis problem using these data sets. Next, we have demonstrated the generalization capability of the PBL-McRBFN classifier by training with the OASIS data set being tested on the unseen ADNI data set. Finally, we have proposed PBL-McRBFN-RFE approach to identify the imaging biomarkers for AD on complete, different age group and gender OASIS data sets.
- In **Chapter 8**, PBL-McRBFN classifier is employed to solve another common neurodegenerative PD diagnosis problem. In here, diagnosis of PD using PBL-McRBFN classifier with microarray gene expression, MRI, vocal and gait features are presented. The obtained complete microarray gene expression data set consists of 22283 genes expression information from subjects. Since the complete gene expression data set consists of large number of redundant genes, we also conducted PD prediction using PBL-McRBFN classifier on the most informative genes with p -value selection at values less than 0.05/0.01. The quantitative performance comparison of PBL-McRBFN with SVM classifier has been presented in PD diagnosis problem. Further, proposed PBL-McRBFN-RFE approach is used to identify the imaging biomarkers for PD using MRI scans.
- **Chapter 9** summarizes the conclusions and provides directions for future research work.

Chapter 2

Literature Review on Sequential Learning Algorithms in Neural Networks

In this chapter, key concepts in sequential learning algorithm for neural networks are reviewed. The literature review on neurodegenerative diseases is provided in chapters 7 and 8. The review discusses briefly on different types of learning models in neural networks to handle sequential data.

In a sequential learning framework, the training samples arrive one-by-one and the samples are discarded after the learning process. Hence, it requires less memory and computational time during the learning process. In addition, sequential learning algorithms automatically determine the minimal architecture that can accurately approximate the decision function described by a stream of training samples. In the families of neural networks, Radial Basis Function (RBF) neural networks have been extensively used in a sequential learning framework due to its universal approximation ability and simplicity of architecture. Hence, in this thesis, we consider radial basis function neural network for classification problems. Many sequential learning algorithms in radial basis function framework are available in the literature to solve classification problems and depending on training method of the network and the structure, these learning algorithms could be broadly classified as belonging to one the these: error driven algorithms, neuron significance based algorithms, extreme learning machine based algorithms, spiking neural networks algorithms, kernel least mean square based algorithms, and sequential classifica-

tion algorithms. These are discussed in detail next, along with incremental-decremental SVM Algorithms.

2.1 Sequential Learning Algorithms

2.1.1 Error Driven Algorithms

Resource Allocation Network (RAN) [19] is the first sequential learning algorithm introduced in the literature. RAN evolves the network architecture required to approximate the true function using novelty based neuron growth criterion. In RAN, the novelty of a sample is determined based on the error and distance to the nearest neuron. If the novelty criterion is satisfied then a new hidden neuron is added to the network otherwise network parameters are updated using Least Mean Squares (LMS) algorithm.

Enhancement of RAN, known as RAN Extended Kalman Filter (RANEKF) algorithm has been proposed in [20]. In RANEKF, extended Kalman filter (EKF) is used rather than the LMS algorithm for updating the network parameters.

One drawback of both RAN and RANEKF is that once a hidden neuron is added, it can never be removed. Thus, both RAN and RANEKF could produce networks in which some hidden neurons, although active initially, may subsequently end up contributing little to the network output [21]. To overcome the above drawback, Minimal Resource Allocation Network (MRAN) algorithm has been proposed in [21]. In MRAN, the RANEKF is augmented with pruning strategy. The pruning strategy in MRAN removes those hidden neurons that consistently make little contribution to the network output. MRAN uses a sliding window of training samples in the growing and pruning criteria to identify the hidden neurons that contribute relatively little to the network output. Selection of the appropriate sizes for these windows critically depends on the distribution of the training samples.

MRAN updates all network parameters, the EKF require storage of huge covariance matrix and inverse of the same. Hence, computational effort and memory requirements in training phase are quite high. To overcome the above weakness for real-time implementation, an algorithm has been proposed called Extended Minimum Resource Allocation Network (EMRAN) in [22]. EMRAN is an improved version of the MRAN. In EMRAN,

a ‘winner neuron’ strategy is incorporated. In EMRAN, winner neuron is defined as that hidden neuron in the network which is closest to the training sample. The main contribution of the EMRAN algorithm is that, in every step, only those parameters that are related to the winner neuron are updated by the EKF algorithm.

2.1.2 Neuron Significance Based Algorithms

In [23], neuron significance with respect to input distribution is used as a criterion for growing and pruning the RBF network architecture, and is called Growing And Pruning RBFN (GAP-RBFN). A new hidden neuron is added if its significance exceeds a threshold value, whereas existing hidden neurons are pruned if their significance is less than certain threshold. In GAP-RBFN algorithm, significance of a neuron is defined based on the contribution made by that neuron to the network output averaged over all the training samples received so far and this requires the knowledge of the input data distribution. The significance of neuron is calculated with simplified piecewise linear approximation to the Gaussian function to reduce the computational effort. Similar to EMRAN, in GAP-RBFN also only nearest hidden neuron parameters are updated by EKF algorithm.

In GAP-RBFN algorithm, neuron significance is calculated based on the assumption that the training samples are uniformly distributed. If the training samples distribution is non-uniform then the performance of GAP-RBFN will be affected. To overcome the above drawback, Generalized version of GAP-RBFN (GGAP-RBFN) algorithm has been presented in [24]. The GGAP-RBFN algorithm can be used for any arbitrary input distribution.

The performance of GAP-RBFN algorithm for classification problems has been evaluated in [25], improvements to GAP-RBFN for enhancing its performance in both accuracy and training speed for classification problems has been presented in Fast GAP-RBFN (FGAP-RBFN) algorithm [25]. In FGAP-RBFN algorithm Decoupled EKF (DEKF) is used, whereas EKF is used in MRAN and GAP-RBFN algorithms for network parameters update. EKF requires more computational efforts and large memory for large input dimensional problems. On the other hand, DEKF only considers the pair wise interdependence of the parameters from the same decouple group, rather than the interdependence of all the parameters in the network. When the number of hidden neurons becomes large,

DEKF results in a significant reduction in computational cost per training sample and in storage requirements for error covariance matrix.

2.1.3 Extreme Learning Machine Based Algorithms

Extreme Learning Machine (ELM) [26] is a well known fast learning neural network paradigm. It is a batch learning algorithm for a single-hidden layer feed forward neural network. ELM randomly chooses input weights and analytically determines the output weights using minimum norm least-squares. A complete survey of research works in ELM framework are presented in [27]. Sequential version of ELM using recursive least squares has been presented in [28], referred as an On-line Sequential Extreme Learning Machine (OS-ELM). OS-ELM can handle samples one-by-one or chunk-by-chunk with varying chunk size. In the OS-ELM algorithm, the input weights are selected randomly and output weights are calculated analytically using the least square error. For sequential learning, the output weights are updated using recursive least. In the output weight calculations, OS-ELM uses a small chunk of initial training data. The small chunk of initial training data affects the training performance of OS-ELM. In case of sparse and imbalance data sets, the random selection of input weights with fixed number of hidden neurons in the OS-ELM affects the performance significantly as shown in [29].

Incremental versions of ELM have been proposed in [30, 31, 32]. In [30], Incremental ELM (I-ELM) algorithm is presented. In I-ELM, every time only one hidden neuron is randomly generated and added to the existing network. However, I-ELM does not recalculate the output weights of all the existing hidden neurons when a new hidden neuron is added. To improve the convergence of I-ELM, Convex Incremental ELM (CI-ELM) algorithm is presented in [31]. In CI-ELM, the convergence rate of I-ELM is improved by recalculating the output weights of the existing hidden neurons based on a convex optimization method when a new hidden neuron is randomly added. CI-ELM obtains a faster convergence rate and more compact network architecture while retaining the I-ELM's simplicity and efficiency. The performance of I-ELM is further improved in Enhanced I-ELM algorithm (EI-ELM) [32]. In EI-ELM, every time k hidden neurons are randomly generated and among the k randomly generated hidden neurons only the

most appropriate hidden neuron is added to the existing network. However I-ELM, CI-ELM and EI-ELM requires complete training data so these algorithms cannot handle sequential data.

2.1.4 Spiking Neural Networks Algorithms

Spiking Neural Networks (SNN) known as the third generation of neural network models, are more closely related to their biological neurons compared to classical artificial neural networks from the previous generations. A sequential learning algorithm in SNN framework has been presented in [33] for a four layer hierarchical neural network of two-dimensional integrate-and-fire neuronal maps. In this algorithm, the training is performed through synaptic plasticity and adaptive network structure. An event driven approach is used to optimize computation speed in order to simulate networks with large number of spiking neurons. Another sequential learning algorithm for SNN and its application to taste recognition has been presented in [34]. This algorithm is developed based on integrate-and-fire neurons with rank order coded inputs. The influence of information encoding in a population of spiking neurons on the performance of SNN was also explored.

There are number of unaddressed issues in above sequential algorithms for SNN [33, 34], such as fine tuning of learning parameters, automatic update of learning parameters in continuously changing environments (as these are set manually), improving learning speed for large size data sets, and the effect of handling imbalanced data sets on the training performance.

In [35], a novel self-adaptation system has been presented to train a real mobile robot for optimal navigation in dynamic environments by training a number of SNNs having the Spike Timing Dependent Plasticity (STDP) property. The spike response model is used and the trained SNNs are stored in a tree-type memory structure that are used as experiences for the robot to enhance its navigation ability in new and previously trained environments. The memory was designed to have a simple searching mechanism. Forgetting and online dynamic clustering techniques are used in order to control the memory size. The system used the minimum network structure required for performing obstacle avoidance task and its synaptic weights are changed online. However, more

experimental data needs to be collected in order to demonstrate the robot navigation ability in a dynamic office environment. The complete review on sequential learning algorithms for SNN is presented in [36].

2.1.5 Incremental-Decremental SVM Algorithms

Support Vector Machines (SVM) [37] is widely used algorithm to solve classification problems. A sequential learning version in support vector machine framework is called an incremental SVM has been presented in [38]. Incremental and decremental SVM algorithm referred as single incremental decremental SVM (SIDSVM) has been presented in [39]. It uses an on-line recursive algorithm for training SVMs, and it handles one sample at a time by retaining Karush-Kuhn-Tucker conditions on all previously seen training samples, while ‘adiabatically’ adding a new training samples to the solution. This approach is adapted to other variants of kernel machines in [38, 40].

The drawback of SIDSVM algorithm is when multiple training samples are added/removed it will repeat the updating operation for each single training sample. It often requires higher computational cost for real-time implementation. To overcome the above drawback, a multiple incremental decremental SVM (MIDSVM) algorithm has been proposed in [41]. MIDSVM algorithm is developed based on multi-parametric programming in the optimization literature [42]. Here, multiple samples are added or removed simultaneously and is faster than the conventional incremental decremental support vector machines presented in [39].

2.1.6 Kernel Least Mean Square Based Algorithms

Kernel least mean square based learning algorithms are also candidates for on-line kernel based learning other than SVM based learning algorithms. One of the first Kernel Least Mean Square (KLMS) algorithm is presented in [43]. In KLMS algorithm, LMS has been extended to a Reproducing Kernel Hilbert Space (RKHS), resulting in an adaptive filter built from a weighted sum of kernel functions evaluated at each incoming data sample. The mean square convergence study of the KLMS algorithm is presented in [44]. The major drawback of KLMS algorithm is with time, the size of the filter as well as the computational effort and requirement of memory increases. To overcome the above

drawback an efficient method to constrain the increase in the length of RBF is proposed in [45] referred as KLMS with Constrained Growth (KLMS-CG). KLMS-CG algorithm uses sequential Gaussian elimination method to test the linear dependency of the each new sample feature vector with all the previous samples feature vectors. The extended and quantized versions of KLMS algorithm are presented in [46, 47, 48, 49, 50, 51, 52]. A nonparametric information theoretic approach based KLMS using surprise criterion is proposed in [51], here surprise quantifies the amount of information a sample contains given a system state. Quantized KLMS (QKLMS) algorithm is proposed in [49], which is based on simple online vector quantization method. In QKLMS, quantization is applied to compress the input (or feature) space of the kernel adaptive filters so as to control the growth of the RBF structure. In Fixed-Budget QKLMS (QKLMS-FB) [52], a pruning and growing strategy is proposed based on significance measure to constrain its network size.

2.1.7 Sequential Classification Algorithms

In literature, some algorithms developed to solve only classification problems. Sequential Multi-Category Radial Basis Function network (SMC-RBFN) [53] algorithm has been developed exclusively to solve classification objectives. As the training samples are presented, SMC-RBFN adds the new hidden neurons or updates the network parameters. In the growth criterion, SMC-RBFN considers the similarity measure within class, misclassification rate and prediction error. SMC-RBFN uses the hinge loss function instead of the mean square loss function for a more accurate estimate of the posterior probability. New hidden neuron parameters are allocated similar to RAN algorithm. Otherwise, network parameters of nearest hidden neuron of the same class are updated using DEKF.

Another sequential classification algorithm has been presented in Self-adaptive Resource Allocation Network (SRAN) [9]. As each training sample is presented to the SRAN, based on the sample hinge error, the sample is either used for network training (growing/update) immediately or pushed to the rear end of the stack for learning in future or deleted from the data set. SRAN uses self-adaptive error based control parameters to identify reduced training sample sequence with significant information and removes the less significant samples (which are similar to the stored knowledge in the

Table 2.1: Comparison of supervised sequential learning algorithms

Algorithms	Architecture	Activation Function	Features	Complexity	Sample Selection
RAN	Self-adaptive	Gaussian	Novelty based neuron growth	Least mean square method	NO
RANEKF	Self-adaptive	Gaussian	EKF parameter update	EKF parameter update	NO
MRAN	Self-adaptive	Gaussian	Pruning strategy	EKF parameter update	NO
EMRAN	Self-adaptive	Gaussian	Winner neuron strategy	EKF parameter update	NO
GAP-RBFN	Self-adaptive	Gaussian	Neuron significance concept	EKF parameter update	NO
GGAP-RBFN	Self-adaptive	Gaussian	Any arbitrary input sampling distribution	EKF parameter update	NO
FGAP-RBFN	Self-adaptive	Gaussian	Decoupled EKF parameter update	Decoupled EKF parameter update	NO
OS-ELM	Fixed	Additive, Radial	Learn data one-by-one or chunk-by-chunk	Random selection of neuron parameters	NO
SMC-RBFN	Self-adaptive	Gaussian	Hinge loss function	EKF parameter update	NO
SRAN	Self-adaptive	Gaussian	Sequence alteration	EKF parameter update	YES
SIDSVM	Self-adaptive	Kernels	On-line SVM learning, Adiabatic increments	Parametric programming	NO
MIDSVM	Self-adaptive	RBF kernel	Handle multiple samples at a time	Mutli-parametric programming	NO
KLMS	Self-adaptive	RBF kernel	LMS extended to a RKHS	Stochastic gradient in RKHS	NO
QKLMS	Self-adaptive	RBF kernel	Quantized feature space	Stochastic gradient in RKHS	NO
QKLMS-FB	Self-adaptive	RBF kernel	Significance measure based growing and pruning	Stochastic gradient in RKHS	NO

network) to avoid overtraining. In growth/update criterion, SRAN considers explicit misclassification error and hinge loss error. New hidden neuron parameters are allocated similar to other sequential algorithms and all the network parameters are updated using EKF. Otherwise, the sample is pushed to the rear end of the stack, to be presented to the network in future. These reserved samples can be used to fine-tune the network parameters.

The Table 2.1 summarizes the key difference among the all the sequential learning algorithms discussed above.

2.2 Summary

In this chapter, we presented the different sequential learning algorithms for neural networks. We then categorized sequential learning algorithms as: error driven algorithms, neuron significance based algorithms, ELM based algorithms, spiking neural networks algorithms, KLMS based algorithms, and sequential classification algorithms are discussed along with incremental decremental SVM Algorithms. All the sequential learning algorithms for neural networks presented in literature addresses the technique used to learn the information contained in the training sample, efficiently. But they do not self-regulate their learning. In literature, it has been shown that a self-regulated learner using meta-cognition is the best learner. In next chapter, we give an overview of human meta-cognitive learning and introduce a literature review of models of meta-cognition which encompasses meta-cognitive knowledge and self-regulation.

Chapter 3

An Overview on Meta-cognition

In the previous chapter, a complete literature survey on sequential learning algorithms for neural networks was presented. Existing sequential learning algorithms for radial basis function neural networks use all the samples in the training data set to gain knowledge about the information contained in the samples. In other words, they possess information-processing abilities of humans, including perception, learning, remembering, judging, and problem-solving, and these abilities are cognitive in nature. However, recent studies on human learning have revealed that the learning process is effective when the learners adopt self-regulation in learning process using meta-cognition [3, 4]. Meta-cognition means *cognition about cognition*. In a meta-cognitive framework, human-beings think about their cognitive processes, develop new strategies to improve their cognitive skills and evaluate the information contained in their memory. This chapter gives an overview of human meta-cognition and motivation for meta-cognitive learning. First we define important concepts relevant to meta-cognition. Next, we present a brief review on models of meta-cognition in literature. Finally, we give motivation for meta-cognitive learning for a radial basis function neural network.

3.1 Definitions of Important Concepts in Meta-cognition

- **Cognition** : The mental process of knowing, including aspects such as awareness, perception, reasoning, and judgment.
- **Meta-cognition** : The term meta-cognition is defined in [5] as ‘one’s knowledge concerning one’s own cognitive processes or anything related to them’. The recent

definition of meta-cognition is given in [54] as the awareness and knowledge of one's mental processes such that one can monitor, regulate, and direct them to a desired goal.

- **Major concepts of meta-cognition :** The three major concepts of meta-cognition that have been investigated extensively are: meta-cognitive knowledge, meta-cognitive monitoring, and meta-cognitive control. The definitions of these terms are as follow:
 - *Meta-cognitive Knowledge:* Defined as declarative knowledge about cognition. Declarative knowledge is composed of facts, beliefs, and episodes that can be stated and used to access conscious awareness [55].
 - *Meta-cognitive Monitoring:* Defined as assessing the current state of a cognitive activity. Such as, judging whether you are approaching the correct solution to a problem and assessing how well you understand? What you are reading? [5].
 - *Meta-cognitive Control:* Defined as regulating ongoing cognitive activity. Such as, stop the process, continue it or change the process [5].

3.2 Models of Meta-cognition

There are several meta-cognition models available in human physiology and a brief survey of various meta-cognition models are reported in [7]. Among the various models, the model proposed by Nelson and Narens in [8] is simple and clearly highlights the various actions in human meta-cognition as shown in Fig. 3.1. The model is analogous to the meta-cognition in human-beings and has two components, the cognitive component and the meta-cognitive component. The information flow from the cognitive component to meta-cognitive component is considered monitoring, while the information flow in the reverse direction is considered control. The information flowing from the meta-cognitive component to the cognitive component either changes the state of the cognitive component or changes the cognitive component itself. Monitoring informs the meta-cognitive component about the state of cognitive component, thus continuously updating the meta-cognitive component's model of cognitive component, including, 'no change in state'.

Figure 3.1: Nelson and Narens model of meta-cognition

3.3 Motivation for Meta-cognitive Learning

Recent studies in human learning suggested that the learning process is effective when the learners adopt self-regulation in learning process using meta-cognition [3, 4]. In the meta-cognitive learning, learner controls the learning process by planning and selecting learning strategies and monitors their progress by analyzing the effectiveness of the proposed learning strategies. When necessary, these strategies should be adapted appropriately. Meta-cognition present in human-being provides a means to address *what-to-learn*, *when-to-learn* and *how-to-learn*, i.e., the ability to identify the specific piece of required knowledge, judge when to start and stop learning by emphasizing best learning strategy. Hence, there is a need to develop a meta-cognitive neural network classifier that is capable of deciding *what-to-learn*, *when-to-learn* and *how-to-learn* the decision function from the training data by emulating the human self-regulated learning.

In the existing sequential learning algorithms, Self-adaptive Resource Allocation Network (SRAN) [9] address the *what-to-learn* component of meta-cognition by selecting significant samples using misclassification error and hinge loss error. It has been shown that the selecting appropriate samples for learning and removing repetitive samples help in improving the generalization performance. Therefore, it is evident that emulating the three components of human learning with suitable learning strategies would improve the generalization ability of a neural network. The drawbacks in the existing sequential learning algorithms are: a) The samples for training are selected based on simple error criterion which is not sufficient to address the significance of samples; b) The new hidden

neuron center is allocated independently which may overlap with already existed neuron centers leading to misclassification; c) Knowledge gained from past samples is not used; and d) Uses computationally intensive parameter update.

In this thesis, to overcome the above drawbacks we develop a meta-cognitive learning algorithm for a radial basis function neural network based on generic framework of meta-cognition proposed by Nelson and Narens.

3.4 Summary

In this chapter, we presented an overview of meta-cognition including the definitions of major concepts. Next, the models of meta-cognitive learning are reviewed. Finally, motivation for meta-cognitive learning in neural networks framework is explained. In this thesis, a radial basis function neural network which can self-regulate its learning based on its meta-knowledge is termed as a meta-cognitive radial basis function neural network. In the next chapter, we introduce such a meta-cognitive radial basis function network and present its sequential learning algorithm to perform classification tasks.

Chapter 4

Meta-cognitive Radial Basis Function Network and Its EKF Based Sequential Learning Algorithm for Classification Problems

4.1 Introduction

In the previous chapter, an overview of human meta-cognition, models of meta-cognition and motivation for meta-cognitive learning in neural network framework was presented. In a meta-cognitive framework, human-beings think about their cognitive processes, develop new strategies to improve their cognitive skills and evaluate the information contained in their memory. If a radial basis function network analyzes its cognitive process and chooses suitable learning strategies adaptively to improve its cognitive process then it is referred to as ‘Meta-cognitive Radial Basis Function Network’ (McRBFN). This chapter focuses on the development of McRBFN and its Extended Kalman Filter (EKF) based sequential learning algorithm. First, we define the classification problem in sequential framework. Next, we present the learning algorithm.

4.2 Classification Problem Definition

The classification problem in a sequential learning framework can be defined as:

Given stream of training data samples, $\{(\mathbf{x}^1, c^1), \dots, (\mathbf{x}^t, c^t), \dots\}$, where $\mathbf{x}^t = [x_1^t, \dots, x_m^t]^T \in \mathbb{R}^m$ is the m -dimensional input of the t^{th} sample, and $c^t \in (1, n)$ is its class label. Where n

Figure 4.1: (a) Nelson and Narens model of meta-cognition (b) EKF-McRBFN model

is the total number of classes. The coded class labels $\left(\mathbf{y}^t = [y_1^t, \dots, y_j^t, \dots, y_n^t]^T\right) \in \mathfrak{R}^n$ are given by:

$$y_j^t = \begin{cases} 1 & \text{if } c^t = j \\ -1 & \text{otherwise} \end{cases} \quad j = 1, \dots, n \quad (4.1)$$

The objective of a classifier is to approximate the underlying decision function that maps $\mathbf{x}^t \in \mathfrak{R}^m \rightarrow \mathbf{y}^t \in \mathfrak{R}^n$.

4.3 EKF-McRBFN Classifier

In this section, we present the architecture of the EKF based Meta-cognitive Radial Basis Function Network (EKF-McRBFN) classifier and its working principles. EKF-McRBFN architecture is developed based on the Nelson and Narens meta-cognition model [8]. Fig. 4.1(a) shows the Nelson and Narens meta-cognition model which is analogous to the meta-cognition in human-beings and has two components, a cognitive component and a meta-cognitive component. The information flow from the cognitive component to meta-cognitive component is considered monitoring, while the information flow in the reverse direction is considered control. EKF-McRBFN is developed based on the Nelson and Narens meta-cognition model as shown in Fig. 4.1(b). Similar to Nelson and Narens model, EKF-McRBFN has two components as shown in schematic diagram in Fig. 4.2, namely the cognitive component and the meta-cognitive component. The cognitive component of EKF-McRBFN is a three layered feed forward radial basis function network with Gaussian activation function in the hidden layer as shown in Fig. 4.3. The meta-cognitive component contains copy of the cognitive component. When a new training sample arrives, the meta-cognitive component of EKF-McRBFN predicts the class label

Figure 4.2: Schematic diagram of EKF-McRBFN

and estimates the knowledge present in the new training sample with respect to the cognitive component. Based on this information, the meta-cognitive component selects a suitable learning strategy, for the current sample. Thereby, addressing the three fundamental issues in learning process: a) *what-to-learn* b) *when-to-learn* and c) *how-to-learn*. Meta-cognitive component is a regulatory system, which helps the adaptive cognitive component to learn the input-output relationship efficiently. It is similar to feedback system, where the meta-cognition provides appropriate learning strategies based on the monitory signals from the cognitive component.

EKF-McRBFN begins with zero hidden neuron and selects suitable strategy for each sample to achieve the objective. First, we present the cognitive component and next we highlight various learning strategies of the meta-cognitive component.

4.3.1 Cognitive Component of EKF-McRBFN

The cognitive component of EKF-McRBFN is a three layered feed forward radial basis function network. The input layer maps all features to the hidden layer without doing any

Figure 4.3: Cognitive component: RBF network

transformation, the hidden layer employs Gaussian activation function and the output layer uses a linear activation function as shown in Fig. 4.3.

Without loss of generality, we assume that the meta-cognitive learning algorithm builds K Gaussian neurons from $t - 1$ training samples. For given training sample $\mathbf{x}^t$, the predicted output ($\hat{\mathbf{y}}^t = [\hat{y}_1^t, \dots, \hat{y}_j^t, \dots, \hat{y}_n^t]^T$) of EKF-McRBFN classifier with K hidden neurons is

$$\hat{y}_j^t = \sum_{k=1}^K w_{kj} h_k^t, \quad j = 1, \dots, n \quad (4.2)$$

where w_{kj} is the weight connecting the k^{th} hidden neuron to the j^{th} output neuron and h_k^t is the response of the k^{th} hidden neuron to the input $\mathbf{x}^t$ is given by

$$h_k^t = \exp\left(-\frac{\|\mathbf{x}^t - \boldsymbol{\mu}_k^l\|^2}{(\sigma_k^l)^2}\right), \quad k = 1, \dots, K \quad (4.3)$$

where $\boldsymbol{\mu}_k^l \in \Re^m$ is the center and $\sigma_k^l \in \Re^+$ is the width of the k^{th} hidden neuron. Here, the superscript l represents the corresponding class of the hidden neuron.

The objective is to estimate the number of hidden neurons (K), neuron centers ($\boldsymbol{\mu}_k^l$), widths (σ_k^l) and output weights (w) of the network, such that network approximates the decision function accurately.

4.3.2 Meta-cognitive Component of EKF-McRBFN

The meta-cognitive component contains a dynamic model of the cognitive component, knowledge measures and self-regulated thresholds. During the learning process, meta-

cognitive component monitors the cognitive component and updates its dynamic model of the cognitive component. When a new training sample (t^{th}) sample is presented to the EKF-McRBFN, the meta-cognitive component of EKF-McRBFN estimates the knowledge present in the new training sample with respect to the cognitive component using its knowledge measures. The meta-cognitive component uses predicted class label ($\hat{c}^t$), maximum hinge error (E^t), confidence of classifier ($\hat{p}(c^t|\mathbf{x}^t)$) and class-wise significance (ψ_c) as the measures of knowledge in the new training sample. Self-regulated thresholds are adapted to capture the knowledge presented in the new training sample. Using the knowledge measures and self-regulated thresholds, the meta-cognitive component constructs two sample based learning strategies and two neuron based learning strategies. One of these strategies is selected for the new training sample such that the cognitive component learn them accurately and achieves better generalization performance.

The meta-cognitive component knowledge measures are defined as shown below:

Predicted class label ($\hat{c}^t$): Using the predicted output ($\hat{\mathbf{y}}^t$), the predicted class label ($\hat{c}^t$) can be obtained as

$$\hat{c}^t = \arg \max_{j \in 1, \dots, n} \hat{y}_j^t \quad (4.4)$$

where n is the total number of classes.

Maximum hinge error (E^t): The objective of the classifier is to minimize the error between the predicted output ($\hat{\mathbf{y}}^t$) and actual output ($\mathbf{y}^t$). In classification problems, it has been shown in [56, 57] that the classifier developed using hinge loss error estimates the posterior probability more accurately than the classifier developed using mean square error. Hence, in EKF-McRBFN, we use the hinge loss error ($\mathbf{e}^t = [e_1^t, \dots, e_j^t, \dots, e_n^t]^T$) $\in \mathbb{R}^n$ defined as

$$e_j^t = \begin{cases} 0 & \text{if } y_j^t \hat{y}_j^t > 1 \\ y_j^t - \hat{y}_j^t & \text{otherwise} \end{cases} \quad j = 1, \dots, n \quad (4.5)$$

where y_j^t is the actual output and $\hat{y}_j^t$ is the predicted output at the j^{th} -neuron for t^{th} -sample.

The maximum absolute hinge error (E^t) is given by

$$E^t = \max_{j \in 1, 2, \dots, n} |e_j^t| \quad (4.6)$$

where e_j^t is the hinge error at the j^{th} -neuron for t^{th} -sample.

Confidence of classifier ($\hat{p}(c^t|\mathbf{x}^t)$): The confidence level of classification or predicted posterior probability is given as

$$\hat{p}(c^t|\mathbf{x}^t) = \frac{\min(1, \max(-1, \hat{y}_j^t)) + 1}{2}, \quad j = c^t \quad (4.7)$$

where $\hat{y}_j^t$ is the predicted output at the j^{th} -neuron and actual class label for t^{th} -sample.

Class-wise significance (ψ_c): In general, the input feature ($\mathbf{x}^t$) is mapped on to a hyper-dimensional spherical feature space $\mathbb{S}$ using K Gaussian neurons, i.e., $\mathbf{x}^t \rightarrow \phi(\mathbf{x}^t)$. Where $\phi(\mathbf{x}^t)$ is input feature ($\mathbf{x}^t$) in feature space. Therefore, all $\phi(\mathbf{x}^t)$ lie on a hyper-dimensional sphere as shown in [58]. The knowledge or spherical potential of any sample in original space is expressed as a squared distance from the hyper-dimensional mapping $\mathbb{S}$ [59].

In EKF-McRBFN, the center (μ) and width (σ) of the Gaussian neurons describe the feature space $\mathbb{S}$. Let the center of the K -dimensional feature space be $\phi_0 = \frac{1}{K} \sum_{k=1}^K \phi(\mu_k)$. The knowledge present in the new data $\mathbf{x}^t$ can be expressed as the potential of the data in the original space, which is squared distance from the K -dimensional feature space to the center ϕ_0 . The potential (ψ) is given as

$$\psi = \|\phi(\mathbf{x}^t) - \phi_0\|^2 \quad (4.8)$$

As shown in [59], the above equation can be expressed as

$$\psi = h(\mathbf{x}^t, \mathbf{x}^t) - \frac{2}{K} \sum_{k=1}^K h(\mathbf{x}^t, \mu_k^l) + \frac{1}{K^2} \sum_{k,r=1}^K h(\mu_k^l, \mu_r^l) \quad (4.9)$$

Where, Gaussian kernel $h(\mathbf{x}^t, \mu_k^l)$ is expressed as $\exp(-\|\mathbf{x}^t - \mu_k^l\|^2/(\sigma_k^l)^2)$. From the above equation, we can see that for Gaussian function the first term ($h(\mathbf{x}^t, \mathbf{x}^t)$) and last term ($\frac{1}{K^2} \sum_{k,r=1}^K h(\mu_k^l, \mu_r^l)$) are constants. Since potential is a measure of novelty, these constants may be discarded and the potential can be reduced to

$$\psi \approx -\frac{2}{K} \sum_{k=1}^K h(\mathbf{x}^t, \mu_k^l) \quad (4.10)$$

Since we are addressing classification problems, the class-wise distribution plays a vital role and it will influence the performance of the classifier significantly [53]. Hence, we use the measure of the spherical potential of the new training sample $\mathbf{x}^t$ belonging to class $\mathbf{c}$ with respect to the neurons associated to same class (i.e., $\mathbf{l} = \mathbf{c}$). Let K^c be the number of neurons associated with the class $\mathbf{c}$, then class-wise spherical potential or class-wise significance (ψ_c) is defined as

$$\psi_c = \frac{1}{K^c} \sum_{k=1}^{K^c} h(\mathbf{x}^t, \mu_k^c) \quad (4.11)$$

We can observe that negative sign and constant $1/2$ in Eq. 4.10 is removed in Eq. 4.11, because the measure of spherical potential is not effected. Also note that the measure of the spherical potential is different from the potential function method referred in [60]. The spherical potential explicitly indicates the knowledge contained in the sample, a higher value of spherical potential (close to one) indicates that the sample is similar to the existing knowledge in the cognitive component and a smaller value of spherical potential (close to zero) indicates that the sample is novel.

4.3.2.1 Learning Strategies

Meta-cognitive component devices various learning strategies using the knowledge measures and self-regulated thresholds, which directly addresses the basic principles of self-regulated human learning (i.e., *what-to-learn*, *when-to-learn* and *how-to-learn*). The meta-cognitive part controls the learning process in cognitive component by selecting one of the following four learning strategies for the new training sample.

- **Sample delete strategy:** If the new training sample contains information similar to the knowledge present in the cognitive component, then delete the new training sample from the training data set without using it in the learning process.
- **Neuron growth strategy:** Use the new training sample to add a new hidden neuron in the cognitive component. During neuron addition, sample overlapping conditions are identified to allocate a new hidden neuron appropriately.
- **Parameter update strategy:** The new training sample is used to update the parameters of the cognitive component. EKF is used to update the parameters.

- **Sample reserve strategy:** The new training sample contains some information but not significant, they can be used at later stage of the learning process for fine tuning the parameters of the cognitive component. These samples may be discarded without learning or used for fine tuning the cognitive component parameters in a later stage.

Most of the existing sequential learning algorithms address only neuron addition/pruning and parameter update. In case of proposed EKF-McRBFN classifier, these learning strategies help to achieve the best human learning ability. Also, these strategies are adapted such that it suits to the current training sample. Since, the meta-cognitive component addresses *what-to-learn*, *when-to-learn* and *how-to-learn*, it improves the generalization ability of the cognitive component.

The principle behind these four learning strategies is described in detail below:

- **Sample delete strategy:** Prevents similar samples from being learnt, which avoids over-training and reduces the computational effort. When the predicted class label of the new training sample is same as the actual class label and the confidence level (estimated posterior probability) is greater than expected value then the new training sample does not provide additional information to the classifier and can be deleted from training sequence without being used in learning process. The *sample delete* criterion is given by

$$\hat{c}^t == c^t \text{ AND } \hat{p}(c^t|x^t) \geq \beta_d \quad (4.12)$$

The deletion threshold (β_d) controls the number of samples participating in the learning process. If one selects β_d close to 1 then no sample will be deleted and all the training samples participates in the learning process which results in over-training with similar samples. Reducing β_d beyond the desired accuracy results in deletion of too many samples from the training sequence. But, the resultant network may not satisfy the desired accuracy. Hence, it is fixed at the expected accuracy level. In our simulation studies, it is selected in the range of [0.9 - 0.95]. The sample deletion strategy prevents learning of samples with similar information, and thereby, avoids over-training and reduces the computational effort.

- **Neuron growth strategy:** When the new training sample contains significant information and the estimated class label is different from the actual class label then one need to add new hidden neuron to capture the knowledge. The neuron growth criterion is given by

$$\hat{c}^t \neq c^t \text{ AND } \psi_c(\mathbf{x}^t) \leq \beta_c \text{ AND } E^t \geq \beta_a \quad (4.13)$$

where β_c is the knowledge threshold and β_a is the addition threshold. The thresholds β_c and β_a allows samples with significant knowledge for learning first then uses the other samples for fine tuning. If β_c is chosen closer to zero and the initial value of β_a is chosen closer to the maximum value of hinge error (i.e., 2 because class labels are coded to -1 or 1), then very few neurons will be added to the network. Such a network will not approximate the function properly. If β_c is chosen closer to one and the initial value of β_a is chosen closer to the minimum value of hinge error, then the resultant network may contain many neurons with poor generalization ability. Hence, the range for the knowledge threshold can be selected in the interval [0.3 - 0.7] and the range for the initial value of addition threshold can be selected in the interval [1.3 - 1.7].

The β_a is adapted as follows

$$\beta_a^t = \delta \beta_a^{t-1} + (1 - \delta) E^t \quad (4.14)$$

where δ is the slope that controls rate of self-adaptation and is set close to 1. β_a allows samples with significant knowledge for learning first then uses the other samples for fine tuning. The hypothesis behind Eq. (4.14) is as the learning process progresses the network uses samples with higher hinge error than initial for neuron growth.

If growth criterion given in Eq. (4.13) is satisfied, then a new hidden neuron $K + 1$ is added and its parameters are initialized as explained below. Existing learning algorithms in the literature do not consider overlapping and distinct cluster criterion in assigning the new neuron parameters. However, the overlapping condition will significantly influence the performance. The new training sample may have overlap

with other classes or will be from a distinct cluster far away from the nearest neuron in the same class. Hence, EKF-McRBFN measures inter/intra class nearest neuron distances from the current sample in assigning the new neuron parameters. The

Let $\mathbf{nrS}$ be the nearest hidden neuron in the intra-class (i.e, $l == c$) with center ($\mu_{\mathbf{nrS}}^c$) and width ($\sigma_{\mathbf{nrS}}^c$) parameters, and $\mathbf{nrI}$ be the nearest hidden neuron in the inter-class (i.e, $l \neq c$) with center ($\mu_{\mathbf{nrI}}^l$) and width ($\sigma_{\mathbf{nrI}}^l$) parameters. They are defined as

$$\mathbf{nrS} = \arg \min_{l==c; \forall k} \|\mathbf{x}^t - \mu_k^l\| \quad (4.15)$$

$$\mathbf{nrI} = \arg \min_{l \neq c; \forall k} \|\mathbf{x}^t - \mu_k^l\| \quad (4.16)$$

Let the Euclidian distances between the new training sample to $\mathbf{nrS}$ and $\mathbf{nrI}$ be given as follows

$$d_S^t = \|\mathbf{x}^t - \mu_{\mathbf{nrS}}^c\| \quad (4.17)$$

$$d_I^t = \|\mathbf{x}^t - \mu_{\mathbf{nrI}}^l\| \quad (4.18)$$

Using the nearest neuron distances, we can determine the overlapping/no-overlapping conditions as in four categories, Fig. 4.4 shows pictorially the distribution of intra-class (same class), inter-class (different class) and four different samples for each overlapping condition:

- *Distinct sample*: when a new training sample is far away from both intra/inter class nearest neurons ($d_S^t \gg \sigma_{\mathbf{nrS}}^c$ AND $d_I^t \gg \sigma_{\mathbf{nrI}}^l$) then the new training sample does not overlap with any class cluster, and is forms a new distinct cluster. In Fig. 4.4, square symbol represents this case of training sample. In this case, the new hidden neuron center (μ_{K+1}^c), width (σ_{K+1}^c) and weight ($\mathbf{w}_{K+1}$) parameters are determined as

$$\mu_{K+1}^c = \mathbf{x}^t; \quad \sigma_{K+1}^c = \max(0.00001, \kappa \sqrt{\mathbf{x}^{tT} \mathbf{x}^t}); \quad \mathbf{w}_{K+1} = \mathbf{e}^t \quad (4.19)$$

where κ is a positive constant which controls the overlap of the responses of the hidden units in the input space, which lies in the range $0.5 \leq \kappa \leq 1$.

Figure 4.4: Schematic representation of training samples corresponding to overlapping/no-overlapping conditions

- *No-overlapping*: When a new training sample is close to the intra-class nearest neuron then the sample does not overlap with the other classes, i.e., the intra/inter class distance ratio is less than 1, then the sample does not overlap with the other classes. In Fig. 4.4, plus symbol represents this case of training sample. In this case, the new hidden neuron center (μ_{K+1}^c), width (σ_{K+1}^c) and weight ($\mathbf{w}_{K+1}$) parameters are determined as

$$\mu_{K+1}^c = \mathbf{x}^t; \quad \sigma_{K+1}^c = \kappa \|\mathbf{x}^t - \mu_{nrS}^c\|; \quad \mathbf{w}_{K+1} = \mathbf{e}^t \quad (4.20)$$

- *Minimum overlapping with the inter-class*: when a new training sample is close to the inter-class nearest neuron compared to the intra-class nearest neuron, i.e., the intra/inter class distance ratio is in range 1 to 1.5, then the sample has minimum overlapping with the other class. In Fig. 4.4, cross symbol represents this case of training sample. In this case, the center of the new hidden neuron is shifted away from the inter-class nearest neuron and shifted towards the intra-class nearest neuron, and is initialized as

$$\mu_{K+1}^c = \mathbf{x}^t + \zeta(\mu_{nrS}^c - \mu_{nrI}^l); \quad \sigma_{K+1}^c = \kappa \|\mu_{K+1}^c - \mu_{nrS}^c\| \quad (4.21)$$

where ζ is center shift factor which determines how much center has to be shifted from the new training sample location. It lies in range [0.01-0.1].

In this case, the center new hidden neuron shifted from the position of new training sample, the weight parameter of the new hidden neuron is calculated as

$$\mathbf{w}_{K+1} = \mathbf{e}^t / \mathbf{h}_{K+1}^t \quad (4.22)$$

where

$$\mathbf{h}_{K+1}^t = \exp \left(- \frac{\|\mathbf{x}^t - \boldsymbol{\mu}_{K+1}^c\|^2}{(\boldsymbol{\sigma}_{K+1}^c)^2} \right) \quad (4.23)$$

- *Significant overlapping with the inter-class*: When a new training sample is very close to the inter-class nearest neuron compared to the intra-class nearest neuron, i.e., the intra/inter class distance ratio is more than 1.5, then the sample has significant overlapping with the other class. In Fig. 4.4, triangle symbol represents this case of training sample. In this case, the center of the new hidden neuron is shifted away from the inter-class nearest neuron and is initialized as

$$\boldsymbol{\mu}_{K+1}^c = \mathbf{x}^t - \zeta(\boldsymbol{\mu}_{nrI}^l - \mathbf{x}^t); \quad \boldsymbol{\sigma}_{K+1}^c = \kappa \|\boldsymbol{\mu}_{K+1}^c - \boldsymbol{\mu}_{nrI}^l\|; \quad (4.24)$$

The weight parameter of new hidden neuron is calculated as given in Eq. (4.22).

The above mentioned center and width determination conditions help in minimizing the misclassification in EKF-McRBFN classifier.

- **Network parameters update strategy**: Cognitive component parameters ($\boldsymbol{\alpha} = [\mathbf{w}_1, \boldsymbol{\mu}_1^l, \boldsymbol{\sigma}_1^l, \dots, \mathbf{w}_K, \boldsymbol{\mu}_K^l, \boldsymbol{\sigma}_K^l]^T$) are updated if the following criterion is satisfied.

$$\widehat{\mathbf{c}}^t == \mathbf{c}^t \text{AND } \mathbf{E}^t \geq \beta_u \quad (4.25)$$

where β_u is the update threshold. If β_u is chosen closer to **50%** of maximum hinge error (i.e., 1), then very few samples will be used for adapting the network

parameters and most of the samples will be pushed to the end of the training sequence. The resultant network will not accurately approximate the function. If a lower value is chosen, then all samples will be used in updating the network parameters without altering the training sequence. Hence, the range for the initial value of update threshold can be selected in the interval $[0.4 - 0.7]$. The β_u is adapted based on the prediction error as:

$$\beta_u^t = \delta \beta_u^{t-1} + (1 - \delta) E^t \quad (4.26)$$

where δ is the slope that controls the rate of self-adaption of update threshold and is set close to 1. The advantage of self-adaptive thresholds is that, they help in selecting the samples for adding as a hidden neuron or to update parameters.

EKF-McRBFN uses extended Kalman filter to update the cognitive component parameters

$$\alpha^t = \alpha^{t-1} + G^t e^t \quad (4.27)$$

where e^t is the error obtained from the hinge loss function for t^{th} sample and $G^t \in \mathbb{R}^{z \times n}$ is the Kalman gain matrix given by:

$$G^t = P^t B^t [R + (B^t)^T P^t B^t]^{-1} \quad (4.28)$$

where $z = K(m + n + 1)$, $R = r_0 I_{n \times n}$ is the variance of the measurement noise, $P^t \in \mathbb{R}^{z \times z}$ is the error covariance matrix, B^t is the partial derivatives for the output with respect to the parameters (α) given by

$$B^t = \left[h_1^t I_{n \times n}, h_1^t \frac{2w_1}{(\sigma_1^t)^2} (x^t - \mu_1^l)^T, h_1^t \frac{2w_1}{(\sigma_1^t)^3} \|x^t - \mu_1^l\|^2 \dots, \right. \\ \left. h_K^t I_{n \times n}, h_K^t \frac{2w_K}{(\sigma_K^t)^2} (x^t - \mu_K^l)^T, h_K^t \frac{2w_K}{(\sigma_K^t)^3} \|x^t - \mu_K^l\|^2 \right]^T \quad (4.29)$$

The error covariance matrix is updated by

$$P^{t+1} = [I_{z \times z} - G^t (B^t)^T] P^t + q_0 I_{z \times z} \quad (4.30)$$

The addition of artificial process noise (q_0) helps in avoiding convergence to local minima.

When a new hidden neuron is added, the dimensionality of error covariance matrix $\mathbf{P}^t$ is increased to

$$\begin{bmatrix} \mathbf{P}_{z \times z}^{t-1} & \mathbf{0}_{z \times (m+n+1)} \\ \mathbf{0}_{(m+n+1) \times z} & \mathbf{p}_0 \mathbf{I}_{(m+n+1) \times (m+n+1)} \end{bmatrix} \quad (4.31)$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ is the zero matrix and $\mathbf{p}_0$ is the initial estimated uncertainty.

- **Sample reserve strategy:** If the new training sample does not satisfy either the deletion or the neuron growth or the cognitive component parameters update criterion, then the sample is pushed to the rear of the training sequence. Since, EKF-McRBFN modifies the strategies based on current sample knowledge, these samples may be used in later stage.

Ideally, training process stops when no further sample is available in the data stream. However, in real-time, stopping criterion is when samples in the reserve remain same.

4.4 EKF-McRBFN Algorithm

To summarize, the EKF-McRBFN algorithm in a pseudo code form is given in Pseudo code 1:

In EKF-McRBFN, sample delete strategy addresses the *what-to-learn* by deleting insignificant samples from data stream, neuron growth strategy and parameters update strategy addresses the *how-to-learn* by which the cognitive component learns from the samples, and self-adaptive nature of meta-cognitive thresholds in addition to the sample reserve strategy addresses the *when-to-learn* by presenting the samples in the learning process according to the knowledge present in the sample.

4.4.1 Guidelines for EKF-McRBFN Thresholds Initialization

In this section, we explain the influence of β_c , β_d , β_a and β_u thresholds on the performance of the EKF-McRBFN and provide some guidelines for their initialization.

Pseudocode 1 Pseudo code for the EKF-McRBFN classification algorithm.

Input : Present the training data one-by-one to the network from data stream.

Output : Decision function that estimates the relationship between feature space and class label.

START

Initialization : Assign the first sample as the first neuron ($K=1$).
The parameters of the neuron are chosen as shown in Eq. (4.19).

For each training sample ($\mathbf{x}^t, y^t$)

DO

Meta-cognitive component computes the significance of the sample with respect to the cognitive component :

 Compute the cognitive component output $\hat{\mathbf{y}}^t$ using Eq. (4.2).

 Find the predicted class label $\hat{\mathbf{c}}^t$, maximum hinge error $\mathbf{E}^t$, confidence of classifier $\hat{p}(\mathbf{c}^t | \mathbf{x}^t)$ and class-wise significance ψ_c using Eqs.(4.4),(4.6), (4.7) and (4.11).

Based on above calculated measures the meta-cognitive component selects one of the following strategies :

Sample delete strategy :

IF $\hat{\mathbf{c}}^t == \mathbf{c}^t$ AND $\hat{p}(\mathbf{c}^t | \mathbf{x}^t) \geq \beta_d$ THEN

 Delete the sample from the sequence without learning.

Neuron growth strategy :

ELSEIF $\hat{\mathbf{c}}^t \neq \mathbf{c}^t$ AND $\psi_c(\mathbf{x}^t) \leq \beta_c$ AND $\mathbf{E}^t \geq \beta_a$ THEN

 Add a neuron to the network ($K = K+1$).

 Choose the parameters of the new hidden neuron using Eqs. (4.19) to (4.24). Update the addition threshold according to Eq. (4.14) and increase the dimensionality of P.

Parameters update strategy :

ELSEIF $\hat{\mathbf{c}}^t == \mathbf{c}^t$ AND $\mathbf{E}^t \geq \beta_u$ THEN

 Update the parameters of the cognitive component using EKF according to Eq. (4.27). Update the update threshold according to Eq. (4.26).

Sample reserve strategy :

ELSE

 The current sample $\mathbf{x}^t, y^t$ is pushed to the rear end of the sample stack to be used in future. They can be later used to fine-tune the cognitive component parameters.

ENDIF

The cognitive component executes the above selected strategy.

ENDDO

END

Figure 4.5: Error regions of various thresholds in EKF-McRBFN

The knowledge threshold (β_c) helps in identifying novelty of current sample and it depends on spherical potential, the range of spherical potential is between 0-1. If spherical potential is close to 1 then sample is similar to existing knowledge, lesser value of spherical potential indicates that the sample is novel. If one selects the threshold β_c close to zero then the network does not allow addition of neurons. Similarly, if one selects the threshold β_c close to one then all samples will be identified as novel sample. Hence, β_c can be selected in the range $[0.3-0.7]$.

The deletion threshold (β_d) prevents over training by removing samples predicted accurately with high confidences. The self-regulated addition threshold (β_a) and update threshold (β_u) are used to select appropriate samples for efficient learning. β_d , β_a and β_u thresholds depends on hinge error (E^t). Note that the hinge error E^t is between $[0, 2]$. The characteristics of thresholds and their influence on EKF-McRBFN performance can be explained by dividing the error region into three sub regions, namely, the sample deleting region, the parameter updating region, and the neuron growing region, as shown in Fig. 4.5.

If confidence of classifier (estimated posterior probability) is greater than β_d and

predicted class label ($\hat{\mathbf{c}}^t$) is same as actual class label then the current sample is similar to existing knowledge in the cognitive component. In that case the current sample is deleted from the training sequence without being used in the learning process. Confidence of classifier decreases from 1 to 0 as hinge error ($\mathbf{E}^t$) increases from 0 to 2 as shown in Fig. 4.5. Suppose one selects the deletion threshold to be 0.5 then many samples satisfying the condition will be deleted without being used in learning. Hence, the resultant classifier may not provide better generalization performance. If one selects close to 1, say 0.99 then most of the similar samples will be used in learning, which will result in over training. Hence, the β_d can be selected in the range [0.9-0.95].

The addition threshold β_a is combined with other conditions, which measure misclassification and knowledge. The minimum possible prediction error one can get when there is a misclassification is 1. Hence, β_a should be greater than 1. If one selects the threshold β_a close to 1 then all samples misclassified will be used for neuron addition. If one selects the threshold β_a close to 2 then very few neurons will be added and the resultant network may not approximate the decision surface. Note that the meta-cognitive component adapts addition threshold such that new hidden neuron added for a sample with higher error. Hence, the initial value of addition threshold (β_a) can be selected in the range [1.3-1.7].

EKF-McRBFN updates the parameters when the predicted class is accurate and the hinge error ($\mathbf{E}^t$) is greater than β_u . When predicted class label is accurate then value of $\mathbf{E}^t$ is in between 0 to 1. If one selects the threshold β_u close to 1 then no sample will be used in updating and hence the resultant network does not approximate the decision function. If one selects the threshold β_u close to 0 then all samples will be used for updating. EKF-McRBFN updates the parameters using samples, which produces higher error. Hence, the initial value of update threshold (β_u) can be selected in the range [0.4-0.7].

4.5 Summary

In this chapter, we have presented the sequential learning algorithm for meta-cognitive radial basis function network using EKF based on human meta-cognitive learning principles for classification problems. The meta-cognitive component in McRBFN is helpful in

choosing suitable strategy for training the cognitive component in EKF-McRBFN. Also, the learning strategies consider sample overlapping condition for proper initialization of new hidden neurons, which helps in minimization of misclassification. The meta-cognitive component appropriately adapts the learning strategies and hence it efficiently decides *what-to-learn*, *when-to-learn* and *how-to-learn*. In addition, the overlapping conditions present in neuron growth strategy helps in proper initialization of new hidden neurons parameters and also minimize the misclassification error. The main drawbacks in EKF-McRBFN classifier is knowledge gained from past samples is not used properly and also uses computationally intensive extended Kalman filter for parameter update.

In the next chapter, a fast and efficient projection based sequential learning algorithm for meta-cognitive radial basis function network classifier is presented.

Chapter 5

Projection Based Learning Algorithm for Meta-cognitive RBF Network Classifier

5.1 Introduction

In the previous chapter, an EKF based sequential learning algorithm for meta-cognitive radial basis function network (EKF-McRBFN) classifier based on the principles of human meta-cognition was presented. Therein, the meta-cognitive learning helps the radial basis function network achieve better performance by controlling *what-to-learn*, *when-to-learn* and *how-to-learn*. The sample overlapping conditions for allocation of new hidden neuron parameters minimizes the misclassification. Also, the *knowledge measures* and *self-regulated thresholds* help the network to approximate the underlying function efficiently, by employing a compact network structure.

However, EKF-McRBFN does not use the knowledge gained from past samples and also uses a computationally intensive EKF for parameter update. To overcome these drawbacks, in this chapter we introduce a fast and efficient Projection Based Learning (PBL) algorithm for McRBFN. The McRBFN using the PBL to obtain the network parameters is referred to as, ‘Projection Based Learning algorithm for a Meta-cognitive Radial Basis Network’ (PBL-McRBFN).

5.2 PBL-McRBFN Classifier

In PBL-McRBFN, when a neuron is added to the cognitive component the Gaussian parameters (center and width) are determined based on the current sample and the output weights are estimated using the projection based algorithm. When a new neuron is added, existing neurons in the cognitive component will be used as pseudo-samples in projection based learning. There-by, the proposed algorithm exploits the knowledge stored in the network for proper initialization. The problem of finding optimal weights is first formulated as a linear programming problem using the principles of minimization and real calculus. The Projection Based Learning (PBL) algorithm then converts the linear programming problem into a system of linear equations and provides a solution for the optimal weights, corresponding to the minima of the error function.

We present a detailed description of the cognitive and the meta-cognitive components of PBL-McRBFN in the following sections:

5.2.1 Cognitive Component of PBL-McRBFN

In PBL-McRBFN, the cognitive component uses Projection Based Learning (PBL) algorithm for learning process instead of computationally intensive extended Kalman filter. The PBL algorithm is described as follows.

Projection based learning algorithm: The projection based learning algorithm works on the principle of minimization of error function and finds the optimal network output parameters for which the error function is minimum, i.e, the network achieves the minimum error point of the error function.

The considered error function is the sum of squared hinge loss error at output neurons. The error function for t^{th} sample is defined as

$$J_t = \sum_{j=1}^n \left(e_j^t\right)^2, \quad t = 1, 2, \dots \quad (5.1)$$

Where $\left(\mathbf{e}^t = \left[e_1^t, \dots, e_j^t, \dots, e_n^t\right]^T\right) \in \Re^n$ is the hinge loss error defined as

$$e_j^t = \begin{cases} 0 & \text{if } y_j^t \hat{y}_j^t > 1 \\ y_j^t - \hat{y}_j^t & \text{otherwise} \end{cases} \quad j = 1, \dots, n \quad (5.2)$$

From the definition of hinge loss error in Eq. (5.2), $\mathbf{J}_t$ becomes zero when $\mathbf{y}_j^t \hat{\mathbf{y}}_j^t > 1$ and when $\mathbf{y}_j^t \hat{\mathbf{y}}_j^t < 1$, $\mathbf{J}_t$ becomes

$$\mathbf{J}_t = \sum_{j=1}^n \left(\mathbf{y}_j^t - \sum_{k=1}^K \mathbf{w}_{kj} \mathbf{h}_k^t \right)^2 \quad (5.3)$$

For all t training samples, the overall error function becomes

$$\mathbf{J}(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^t \mathbf{J}_i = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^n \left(\mathbf{y}_j^i - \sum_{k=1}^K \mathbf{w}_{kj} \mathbf{h}_k^i \right)^2 \quad (5.4)$$

where $\mathbf{h}_k^i$ is the response of the k^{th} hidden neuron for i^{th} training sample.

The optimal output weights ($\mathbf{W}^* \in \Re^{K \times n}$) are estimated such that the total error reaches its minimum.

$$\mathbf{W}^* := \arg \min_{\mathbf{W} \in \Re^{K \times n}} \mathbf{J}(\mathbf{W}) \quad (5.5)$$

The optimal $\mathbf{W}^*$ corresponding to the minima of the error function ($\mathbf{J}(\mathbf{W}^*)$) is obtained by equating the first order partial derivative of $\mathbf{J}(\mathbf{W})$ with respect to the output weight to zero, i.e.,

$$\frac{\partial \mathbf{J}(\mathbf{W})}{\partial \mathbf{w}_{pj}} = 0, \quad p = 1, \dots, K; \quad j = 1, \dots, n \quad (5.6)$$

Equating the first partial derivative to zero and re-arranging we get

$$\sum_{k=1}^K \sum_{i=1}^t \mathbf{h}_k^i \mathbf{h}_p^i \mathbf{w}_{kj} = \sum_{i=1}^t \mathbf{h}_p^i \mathbf{y}_j^i \quad (5.7)$$

Eq. (5.7) can be written as

$$\sum_{k=1}^K \mathbf{a}_{kp} \mathbf{w}_{kj} = \mathbf{b}_{pj}, \quad p = 1, \dots, K; \quad j = 1, \dots, n \quad (5.8)$$

which can be represented in matrix form as

$$\mathbf{A} \mathbf{W} = \mathbf{B} \quad (5.9)$$

where the projection matrix $\mathbf{A} \in \Re^{K \times K}$ is given by

$$\mathbf{a}_{kp} = \sum_{i=1}^t \mathbf{h}_k^i \mathbf{h}_p^i, \quad k = 1, \dots, K; \quad p = 1, \dots, K \quad (5.10)$$

and the output matrix $\mathbf{B} \in \Re^{K \times n}$ is

$$b_{pj} = \sum_{i=1}^t h_p^i y_j^i, \quad p = 1, \dots, K; \quad j = 1, \dots, n \quad (5.11)$$

Eq. (5.8) gives the set of $K \times n$ linear equations with $K \times n$ unknown output weights $\mathbf{W}$. The proof for $\mathbf{A}$ matrix invertible is given at the end of this chapter.

The solution for $\mathbf{W}$ obtained as a solution to the set of equations as given in Eq. (5.9) is minimum, if $\frac{\partial^2 J}{\partial w_{lp}^2} > 0$. The second derivative of the error function (J) with respect to the output weights is given by,

$$\frac{\partial^2 J(\mathbf{W})}{\partial w_{lp}^2} = \sum_{i=1}^t h_p^i h_p^i = \sum_{i=1}^t |h_p^i|^2 > 0 \quad (5.12)$$

As the second derivative of the error function $J(\mathbf{W})$ is positive, the following observations can be made from Eq. (5.12):

1. The function J is a convex function.
2. The output weight $\mathbf{W}^*$ obtained as a solution to the set of linear equations (Eq. (5.9)) is the weight corresponding to the minima of error function (J).

If the projection matrix $\mathbf{A}$ is positive definite symmetric matrix then it is invertible. The solution for the system of equations in Eq. (5.9) can be determined as follows:

$$\mathbf{W}^* = \mathbf{A}^{-1} \mathbf{B} \quad (5.13)$$

5.2.2 Meta-cognitive Component of PBL-McRBFN

In PBL-McRBFN, the functioning of meta-cognitive component is same as McRBFN, however the principles underlying the learning strategies are modified for PBL algorithm.

5.2.2.1 Learning Strategies

The principle behind the learning strategies in PBL-McRBFN is described in detail below:

- **Sample delete strategy:** Prevents similar samples from being learnt, which avoids over-training and reduces the computational effort. When the predicted

class label of the new training sample is same as the actual class label and the confidence level (estimated posterior probability) is greater than expected value then the new training sample does not provide additional information to the classifier and can be deleted from training sequence without being used in learning process. The *sample delete* criterion is given by

$$\hat{c}^t == c^t \text{ AND } \hat{p}(c^t | \mathbf{x}^t) \geq \beta_d \quad (5.14)$$

The deletion threshold (β_d) controls the number of samples participating in the learning process. The sample deletion strategy prevents learning of samples with similar information, and thereby, avoids over-training and reduces the computational effort.

- **Neuron growth strategy:** When a new training sample contains significant information and the predicted class label is different from the actual class label then one need to add a new hidden neuron to represent the knowledge contained in the sample. The neuron growth criterion is given by

$$(\hat{c}^t \neq c^t \text{ OR } E^t \geq \beta_a) \text{ AND } \psi_c(\mathbf{x}^t) \leq \beta_c \quad (5.15)$$

where β_c is the knowledge threshold and β_a is the addition threshold. The thresholds β_c and β_a allows samples with significant knowledge for learning first then uses the other samples for fine tuning.

The PBL-McRBFN growth strategy in Eq. (5.15) is slightly different from EKF-McRBFN, here a new neuron is added when the class labels are different even if the hinge error is low. This change is will reduce the misclassification error.

The β_a is adapted as follows

$$\beta_a^t = \delta \beta_a^{t-1} + (1 - \delta) E^t \quad (5.16)$$

where δ is the slope that controls rate of self-adaptation and is set close to one. The β_a adaptation allows PBL-McRBFN to add neurons only when presented samples to the cognitive network contains significant information.

The center and width parameters of the new neuron ($K + 1$) are initialized based on overlapping conditions as discussed in chapter 4: Growth strategy. For the better continuity, the overlapping conditions are also described in this section. Existing learning algorithms in the literature do not consider overlapping and distinct cluster criterion in assigning the new neuron parameters. However, the overlapping condition will significantly influence the performance. The new training sample may have overlap with other classes or will be from a distinct cluster far away from the nearest neuron in the same class. Hence, PBL-McRBFN measures inter/intra class nearest neuron distances from the current sample in assigning the new neuron parameters.

Let nrS be the nearest hidden neuron in the intra-class (i.e, $l == c$) with center (μ_{nrS}^c) and width (σ_{nrS}^c) parameters, and nrI be the nearest hidden neuron in the inter-class (i.e, $l \neq c$) with center (μ_{nrI}^l) and width (σ_{nrI}^l) parameters. They are defined as

$$nrS = \arg \min_{l==c; \forall k} \|x^t - \mu_k^l\| \quad (5.17)$$

$$nrI = \arg \min_{l \neq c; \forall k} \|x^t - \mu_k^l\| \quad (5.18)$$

Let the Euclidian distances between the new training sample to nrS and nrI are given as follows

$$d_S^t = \|x^t - \mu_{nrS}^c\| \quad (5.19)$$

$$d_I^t = \|x^t - \mu_{nrI}^l\| \quad (5.20)$$

Using the nearest neuron distances, we can determine the overlapping/no-overlapping conditions as follows:

- *Distinct sample*: when a new training sample is far away from both intra/inter class nearest neurons ($d_S^t \gg \sigma_{nrS}^c$ AND $d_I^t \gg \sigma_{nrI}^l$) then the new training sample does not overlap with any class cluster, and is forms a new distinct cluster. In this case, the new hidden neuron center (μ_{K+1}^c) and width (σ_{K+1}^c) parameters are determined as

$$\mu_{K+1}^c = x^t; \quad \sigma_{K+1}^c = \max(0.00001, \kappa \sqrt{x^{tT} x^t}) \quad (5.21)$$

where κ is a positive constant which controls the overlap of the responses of the hidden units in the input space, which lies in the range $0.5 \leq \kappa \leq 1$.

- *No-overlapping*: When a new training sample is close to the intra-class nearest neuron then the sample does not overlap with the other classes, i.e., the intra/inter class distance ratio is less than 1, then the sample does not overlap with the other classes. In this case, the new hidden neuron center (μ_{K+1}^c) and width (σ_{K+1}^c) parameters are determined as

$$\mu_{K+1}^c = \mathbf{x}^t; \quad \sigma_{K+1}^c = \kappa \|\mathbf{x}^t - \mu_{nrS}^c\| \quad (5.22)$$

- *Minimum overlapping with the inter-class*: when a new training sample is close to the inter-class nearest neuron compared to the intra-class nearest neuron, i.e., the intra/inter class distance ratio is in range 1 to 1.5, then the sample has minimum overlapping with the other class. In this case, the center of the new hidden neuron is shifted away from the inter-class nearest neuron and shifted towards the intra-class nearest neuron, and is initialized as

$$\mu_{K+1}^c = \mathbf{x}^t + \zeta(\mu_{nrS}^c - \mu_{nrI}^l); \quad \sigma_{K+1}^c = \kappa \|\mu_{K+1}^c - \mu_{nrS}^c\| \quad (5.23)$$

where ζ is center shift factor which determines how much center has to be shifted from the new training sample location. It lies in range [0.01-0.1].

- *Significant overlapping with the inter-class*: When a new training sample is very close to the inter-class nearest neuron compared to the intra-class nearest neuron, i.e., the intra/inter class distance ratio is more than 1.5, then the sample has significant overlapping with the other class. In this case, the center of the new hidden neuron is shifted away from the inter-class nearest neuron and is initialized as

$$\mu_{K+1}^c = \mathbf{x}^t - \zeta(\mu_{nrI}^l - \mathbf{x}^t); \quad \sigma_{K+1}^c = \kappa \|\mu_{K+1}^c - \mu_{nrI}^l\|; \quad (5.24)$$

The above mentioned center and width determination conditions help in minimizing the misclassification in PBL-McRBFN classifier.

The existing sequential learning algorithms initialize output weight as error based on the current sample. The influence of past samples is not considered in weight initialization. Hence, it will affect the performance of the classifier significantly. The above mentioned issue is dealt in this chapter as existing knowledge of past samples stored in the network as neuron center is used to initialize the weight of new neuron. When a neuron is added to PBL-McRBFN, based on the existing knowledge of past samples stored in the network the output weights are estimated using the PBL as follows:

The size of matrix $\mathbf{A}$ is increased from $\mathbf{K} \times \mathbf{K}$ to $(\mathbf{K} + 1) \times (\mathbf{K} + 1)$

$$\mathbf{A}_{(\mathbf{K}+1) \times (\mathbf{K}+1)}^t = \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{K} \times \mathbf{K}}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t & \mathbf{a}_{\mathbf{K}+1}^T \\ \hline \mathbf{a}_{\mathbf{K}+1} & \mathbf{a}_{\mathbf{K}+1, \mathbf{K}+1} \end{array} \right] \quad (5.25)$$

where $\mathbf{h}^t = [h_1^t, h_2^t, \dots, h_K^t]$ is a vector of the existing $\mathbf{K}$ hidden neurons response for new (t^{th}) training sample. In sequential learning samples are discarded after learning, but the information present in the past samples is stored in the network. The centers of neuron provide the distribution of past samples in feature space. These centers can be used as pseudo-samples to capture the effect of past samples. Hence, existing hidden neurons are used as pseudo-samples to calculate $\mathbf{a}_{\mathbf{K}+1}$ and $\mathbf{a}_{\mathbf{K}+1, \mathbf{K}+1}$ terms.

$\mathbf{a}_{\mathbf{K}+1} \in \mathbb{R}^{1 \times \mathbf{K}}$ is assigned as

$$\mathbf{a}_{\mathbf{K}+1, p} = \sum_{i=1}^{\mathbf{K}+1} h_{\mathbf{K}+1}^i h_p^i, \quad p = 1, \dots, \mathbf{K} \text{ where } h_p^i = \exp\left(-\frac{\|\mu_i^l - \mu_p^l\|^2}{(\sigma_p^l)^2}\right) \quad (5.26)$$

and $\mathbf{a}_{\mathbf{K}+1, \mathbf{K}+1} \in \mathbb{R}^+$ value is assigned as

$$\mathbf{a}_{\mathbf{K}+1, \mathbf{K}+1} = \sum_{i=1}^{\mathbf{K}+1} h_{\mathbf{K}+1}^i h_{\mathbf{K}+1}^i \quad (5.27)$$

The size of matrix $\mathbf{B}$ is increased from $\mathbf{K} \times n$ to $(\mathbf{K} + 1) \times n$

$$\mathbf{B}_{(\mathbf{K}+1) \times n}^t = \left[\begin{array}{c} \mathbf{B}_{\mathbf{K} \times n}^{t-1} + (\mathbf{h}^t)^T (\mathbf{y}^t)^T \\ \mathbf{b}_{\mathbf{K}+1} \end{array} \right] \quad (5.28)$$

and $\mathbf{b}_{\mathbf{K}+1} \in \mathbb{R}^{1 \times n}$ is a row vector assigned as

$$\mathbf{b}_{\mathbf{K}+1, j} = \sum_{i=1}^{\mathbf{K}+1} h_{\mathbf{K}+1}^i \tilde{y}_j^i, \quad j = 1, \dots, n \quad (5.29)$$

where $\tilde{\mathbf{y}}^i$ is the pseudo-output for the i^{th} pseudo sample or hidden neuron (μ_i^l) given as

$$\tilde{\mathbf{y}}_j^i = \begin{cases} 1 & \text{if } l = j \\ -1 & \text{otherwise} \end{cases} \quad j = 1, \dots, n \quad (5.30)$$

Finally, the output weights are estimated as

$$\begin{bmatrix} \mathbf{W}_K^t \\ \mathbf{w}_{K+1}^t \end{bmatrix} = \left(\mathbf{A}_{(K+1) \times (K+1)}^t \right)^{-1} \mathbf{B}_{(K+1) \times n}^t \quad (5.31)$$

The above can be expanded as

$$\begin{bmatrix} \mathbf{W}_K^t \\ \mathbf{w}_{K+1}^t \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{A}_{K \times K}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t & \mathbf{a}_{K+1}^T \\ \hline \mathbf{a}_{K+1} & \mathbf{a}_{K+1, K+1} \end{array} \right]^{-1} \begin{bmatrix} \mathbf{B}_{K \times n}^{t-1} + (\mathbf{h}^t)^T (\mathbf{y}^t)^T \\ \mathbf{b}_{K+1} \end{bmatrix} \quad (5.32)$$

where $\mathbf{W}_K^t$ is the output weight matrix for K hidden neurons, and $\mathbf{w}_{K+1}^t$ is the vector of output weights for new hidden neuron after learning from t^{th} sample.

The inverse of a matrix $\mathbf{A}_{(K+1) \times (K+1)}^t$ is calculated recursively using matrix identities as

$$\left(\mathbf{A}_{(K+1) \times (K+1)}^t \right)^{-1} = \left[\begin{array}{c|c} (\mathbf{A}_{K \times K}^t)^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] + \frac{1}{\Delta} \begin{bmatrix} (\mathbf{A}_{K \times K}^t)^{-1} \mathbf{a}^T \\ \mathbf{1} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{K \times K}^t)^{-1} \mathbf{a}^T \\ \mathbf{1} \end{bmatrix}^T \quad (5.33)$$

where $\Delta = \mathbf{a}_{K+1, K+1} - \mathbf{a}_{K+1} \left(\mathbf{A}_{K \times K}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t \right)^{-1} \mathbf{a}_{K+1}^T$, $\mathbf{A}_{K \times K}^t = \mathbf{A}_{K \times K}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t$, and $(\mathbf{A}_{K \times K}^t)^{-1}$ is calculated as

$$(\mathbf{A}_{K \times K}^t)^{-1} = (\mathbf{A}_{K \times K}^{t-1})^{-1} - \frac{(\mathbf{A}_{K \times K}^{t-1})^{-1} (\mathbf{h}^t)^T \mathbf{h}^t (\mathbf{A}_{K \times K}^{t-1})^{-1}}{1 + \mathbf{h}^t (\mathbf{A}_{K \times K}^{t-1})^{-1} (\mathbf{h}^t)^T} \quad (5.34)$$

After calculating inverse of matrix in Eq. (5.32) using Eqs. (5.33)&(5.34), the resultant equations are

$$\begin{aligned} \mathbf{W}_K^t &= \left[\mathbf{I}_{K \times K} + \frac{(\mathbf{A}_{K \times K}^{t-1})^{-1} \mathbf{a}_{K+1}^T \mathbf{a}_{K+1}}{\Delta} \right] \left[\mathbf{W}_K^{t-1} + (\mathbf{A}_{K \times K}^{t-1})^{-1} (\mathbf{h}^t)^T (\mathbf{y}^t)^T \right] \\ &\quad - \frac{(\mathbf{A}_{K \times K}^{t-1})^{-1} \mathbf{a}_{K+1}^T \mathbf{b}_{K+1}}{\Delta} \end{aligned} \quad (5.35)$$

$$\mathbf{w}_{K+1}^t = -\frac{1}{\Delta} \left[\mathbf{a}_{K+1} \left(\mathbf{W}_K^{t-1} + (\mathbf{A}_{K \times K}^{t-1})^{-1} (\mathbf{h}^t)^T (\mathbf{y}^t)^T \right) - \mathbf{b}_{K+1} \right] \quad (5.36)$$

- **Parameters update strategy:** The current (t^{th}) training sample is used to update the output weights of the cognitive component ($\mathbf{W}_K = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]^T$) if the following criterion is satisfied.

$$\mathbf{c}^t == \hat{\mathbf{c}}^t \text{ AND } E^t \geq \beta_u \quad (5.37)$$

The β_u is adapted based on the hinge error as:

$$\beta_u^t = \delta \beta_u^{t-1} + (1 - \delta) E^t \quad (5.38)$$

where δ is the slope that controls the rate of self-adaption of parameter update and is set close to one.

When a sample is used to update the output weight parameters, the PBL algorithm updates the output weight parameters as follows:

$$\frac{\partial J_{1,t}(\mathbf{W}_K^t)}{\partial w_{pj}} = \frac{\partial J_{1,(t-1)}(\mathbf{W}_K^t)}{\partial w_{pj}} + \frac{\partial J_t(\mathbf{W}_K^t)}{\partial w_{pj}} = 0, \quad p = 1, \dots, K; \quad j = 1, \dots, n \quad (5.39)$$

Equating the first partial derivative to zero and re-arranging the Eq. (5.39), we get

$$\left(\mathbf{A}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t \right) \mathbf{W}_K^t - \left(\mathbf{B}^{t-1} + (\mathbf{h}^t)^T (\mathbf{y}^t)^T \right) = \mathbf{0} \quad (5.40)$$

By substituting $\mathbf{B}^{t-1} = \mathbf{A}^{t-1} \mathbf{W}_K^{t-1}$ & $\mathbf{A}^{t-1} + (\mathbf{h}^t)^T \mathbf{h}^t = \mathbf{A}^t$ and adding/subtracting the term $(\mathbf{h}^t)^T \mathbf{h}^t \mathbf{W}_K^{t-1}$ on both sides the Eq. (5.40) reduced to

$$\mathbf{W}_K^t = (\mathbf{A}^t)^{-1} \left(\mathbf{A}^t \mathbf{W}_K^{t-1} + (\mathbf{h}^t)^T \left((\mathbf{y}^t)^T - \mathbf{h}^t \mathbf{W}_K^{t-1} \right) \right) \quad (5.41)$$

Finally the output weights are updated as

$$\mathbf{W}_K^t = \mathbf{W}_K^{t-1} + (\mathbf{A}^t)^{-1} (\mathbf{h}^t)^T (\mathbf{e}^t)^T \quad (5.42)$$

where $\mathbf{e}^t$ is the hinge loss error for t^{th} sample obtained from Eq. (5.2).

- **Sample reserve strategy:** PBL-McRBFN reserve strategy is same as discussed in chapter 4. If the new training sample does not satisfy either the deletion or the neuron growth or the cognitive component parameters update criterion, then the current sample is pushed to the rear of the training sequence. Since PBL-McRBFN

modifies the strategies based on the current sample knowledge, these samples may be used in later stage.

Ideally, training process stops when no further sample is available in the data stream. However, in real-time, stopping criterion is when samples in the reserve remain same. The guidelines to select parameters β_c , β_d , β_a and β_u in PBL-McRBFN algorithm is same as in EKF-McRBFN algorithm as described in subsection 4.4.1.

5.3 PBL-McRBFN Algorithm

To summarize, the PBL-McRBFN algorithm in a pseudo code form is given in Pseudo code 2:

In PBL-McRBFN, sample delete strategy address the *what-to-learn* by deleting insignificant samples from training data set, neuron growth strategy and parameters update strategy address the *how-to-learn* efficiently by which the cognitive component learns from the samples, and self-adaptive nature of meta-cognitive thresholds in addition to the sample reserve strategy address the *when-to-learn* by presenting the samples in the learning process according to the knowledge present in the sample.

5.4 Salient Features of PBL-McRBFN Algorithm

In this section we list the similarities and dissimilarities of EKF-McRBFN and PBL-McRBFN learning algorithms.

Similarities of EKF-McRBFN and PBL-McRBFN:

- The sample deletion strategy in both of the algorithms is same. The sample deletion strategy helps the network to avoid over-training and save computational effort. The sample deletion strategy address the *what-to-learn*, thus *what-to-learn* in both of the learning algorithms (EKF-McRBFN and PBL-McRBFN) is same.
- The sample reserve strategy in both of the algorithms is same. The sample reserve strategy address the *when-to-learn* in addition to the self-adaptive nature of meta-cognitive thresholds. The meta-cognitive addition and update thresholds are also

Pseudocode 2 Pseudo code for the PBL-McRBFN classification algorithm.

Input : Present the training data one-by-one to the network
from data stream.

Output : Decision function that estimates the relationship
between feature space and class label.

START

Initialization : Assign the first sample as the first neuron ($K=1$).
The parameters of the neuron are chosen as shown in Eq. (5.21).

For each training sample ($\mathbf{x}^t, y^t$)

DO

*Meta-cognitive component computes the significance of the sample with
respect to the cognitive component :*

 Compute the cognitive component output $\hat{y}^t$ using Eq. (4.2).
 Find the predicted class label $\hat{c}^t$, maximum hinge error E^t ,
 confidence of classifier $\hat{p}(c^t | \mathbf{x}^t)$ and class-wise significance ψ_c
 using Eqs.(4.4),(4.6),(4.7) and (4.11).

*Based on above calculated measures the meta-cognitive
component selects one of the following strategies :*

Sample delete strategy :

IF $\hat{c}^t == c^t$ AND $\hat{p}(c^t | \mathbf{x}^t) \geq \beta_d$ THEN

 Delete the sample from the sequence without learning.

Neuron growth strategy :

ELSEIF ($\hat{c}^t \neq c^t$ OR $E^t \geq \beta_a$) AND $\psi_c(\mathbf{x}^t) \leq \beta_c$ THEN

 Add a neuron to the network ($K = K+1$).
 Choose the center and width parameters of the new hidden neuron
 using Eqs. (5.21) to (5.24) and estimate the new hidden neuron output
 weights using Eq. (5.36). Update the existing hidden neuron output
 weights using Eq. (5.35). Update the self-adaptive meta-cognitive
 addition threshold according to Eq. (5.16)

Parameters update strategy :

ELSEIF $c^t == \hat{c}^t$ AND $E^t \geq \beta_u$ THEN

 Update the parameters of the cognitive component using Eq. (5.42)
 Update the self-adaptive meta-cognitive update threshold according
 to Eq. (5.38)

Sample reserve strategy :

ELSE

 The current sample $\mathbf{x}^t, y^t$ is pushed to the rear
 end of the sample stack to be used in future. They can be
 later used to fine-tune the cognitive component parameters.

ENDIF

The cognitive component executes the above selected strategy.

ENDDO

END

adapted in same form in both of the algorithms. Thus *when-to-learn* is same in both of the learning algorithms.

Dissimilarities of EKF-McRBFN and PBL-McRBFN:

- The neuron growth strategy criterion in PBL-McRBFN algorithm is different from EKF-McRBFN. In EKF-McRBFN algorithm a new hidden neuron is added when the predicted class label of sample different from actual class label and maximum hinge error is greater than the knowledge threshold in addition to the novelty condition. While in PBL-McRBFN algorithm, a new hidden neuron is added either the predicted class label of sample is different from actual class label or maximum hinge error is greater than the knowledge threshold in addition to the novelty condition. Hence, the PBL-McRBFN algorithm may add slightly more hidden neurons than EKF-McRBFN algorithm.
- In EKF-McRBFN algorithm, the new hidden neuron output weights are estimated based on instantaneous prediction error ($\mathbf{e}^t$). While in PBL-McRBFN algorithm, the new hidden neuron output weights are estimated using existing knowledge of past trained samples stored in the network. Thus knowledge gained from past trained samples is used in further learning in PBL-McRBFN algorithm. Hence, execution of neuron growth strategy in PBL-McRBFN is different from EKF-McRBFN algorithm.
- In EKF-McRBFN algorithm, network parameters are updated using extended Kalman filter algorithm. In PBL-McRBFN, the existing neuron output weights are updated using projection based learning algorithm. Hence, execution of parameter update strategy in PBL-McRBFN is different from EKF-McRBFN algorithm.
- Neuron growth and parameter strategies in both algorithms are different, thus the *how-to-learn* is different in both of the learning algorithms. PBL-McRBFN algorithm uses past knowledge of trained samples in *how-to-learn*, thus its classification performance will be better than the EKF-McRBFN algorithm.

5.5 Summary

In this chapter, we have presented a Projection Based Learning (PBL) algorithm for Meta-cognitive Radial Basis Function Network (McRBFN) classifier. Projection based learning accurately estimates the output weight by direct minimization of hinge loss error. Knowledge gained from the past samples is used in new hidden neuron parameters initialization and output weights estimations.

In the next chapter, the performance of proposed EKF-McRBFN and PBL-McRBFN classifiers have been evaluated using a number of benchmark multi-category, binary classification problems and compared with other standard classifiers.

Chapter 6

Performance Evaluation of EKF-McRBFN and PBL-McRBFN Classifiers

In the previous chapters 4 and 5, we proposed two sequential learning algorithms for meta-cognitive radial basis function neural network which are EKF-McRBFN and PBL-McRBFN. Compared to the EKF-McRBFN classifier, the PBL-McRBFN classifier is fast and efficient.

In this chapter, we present the performance comparison of the proposed EKF-McRBFN and PBL-McRBFN with the best performing sequential learning algorithm reported in the literature (SRAN) [9], batch ELM [18] and standard Support Vector Machine (SVM) [37] on real-world benchmark binary and multi-category classification data sets from the UCI machine learning repository [61].

6.1 Data Sets Description

In order to extensively verify the performance of the proposed algorithms, we have chosen data sets with small and large number of samples, low and high dimensional features, and balanced and unbalanced data sets in both binary and multi classification problems. The detailed specifications of 5 binary and 10 multi classification data sets are given in Table 6.1. Note that the data sets are taken from UCI machine learning repository, except for satellite imaging [56], global cancer map using micro-array gene expression

Table 6.1: Specification of benchmark binary and multi class data sets

Data sets	# Features	# Classes	# Samples		I.F		Random Trial
			Training	Testing	Training	Testing	
IS	19	7	210	2100	0	0	Yes
IRIS	4	3	45	105	0	0	Yes
WINE	13	3	60	118	0	0.29	Yes
SI	6	9	9108	262144	0	0.87	No
LETTER	16	26	13333	6667	0.06	0.1	No
VC	18	4	424	422	0.1	0.12	Yes
AE	5	4	62	137	0.1	0.33	Yes
GCM	98	14	144	46	0.22	0.39	Yes
LAND	36	6	4435	2000	0.43	0.26	No
GI	9	6	336	105	0.68	0.77	Yes
HEART	13	2	70	200	0.14	0.1	Yes
LD	6	2	200	145	0.17	0.14	Yes
PIMA	8	2	400	368	0.22	0.39	Yes
BC	9	2	300	383	0.26	0.33	Yes
ION	34	2	100	251	0.28	0.28	Yes

[62] and acoustic emission [63] data sets. The sample imbalance in training and testing is measured using imbalance factor (I.F) and is defined as

$$\text{I.F} = 1 - \frac{n}{N} * \min_{j=1 \dots n} N_j \quad (6.1)$$

where N_j is the total number of training samples in class j and $N = \sum_{j=1}^n N_j$.

For efficient comparison, we present them under the following categories as described below:

- *Binary class data sets:* All considered binary class data sets have high sample imbalance and are grouped into two categories
 - *Low dimensional:* Liver Disorders (LD), Pima Indian diabetes (PIMA) and Breast Cancer (BC) are having low dimensional features with relatively smaller number of training samples.
 - *High dimensional:* Heart disease (HEART) and Ionosphere (ION) data sets are having smaller number of training samples with high dimensional features.

- *Multi class data sets*: Considered 9 multi class data sets are grouped into three categories:
 - *Well balanced*: Iris classification (IRIS), Image segmentation (IS) and Wine determination (WINE) data sets have equal number of training samples per class. These data sets are having varying number of features and training/testing samples.
 - *Imbalanced*: Acoustic Emission classification (AE), Vehicle Classification (VC) and Glass Identification (GI) data sets have lower dimensional features and highly imbalanced training samples. The Global Cancer Mapping using micro-array gene expression (GCM) is having high dimensional features with high sample imbalance.
 - *Large number of samples*: Letter recognition (LETTER), Satellite Image classification (SI) and Landsat Satellite (LAND) data sets have relatively large number of samples and classes.

6.2 Simulation Environment

For this performance comparison study, experiments are conducted for the PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM classifiers on all the data sets in MATLAB 2011 on a desktop PC with Intel Core 2 Duo, 2.66GHz CPU and 3GB RAM. The tuneable parameters of PBL-McRBFN, EKF-McRBFN, SRAN are chosen using cross-validation on the training data sets. For ELM classifier [18], the number of hidden neurons are obtained using the constructive-destructive procedure presented in [64]. The simulations for batch SVM with Gaussian kernels are carried out using the LIBSVM package in C [65]. For SVM classifier, the parameters $(\mathbf{c}, \gamma)$ are optimized using grid search technique. For PBL-McRBFN and EKF-McRBFN classifiers, the parameters $(\beta_d, \beta_a, \beta_c, \beta_u$ and $\kappa)$ are also optimized using grid search technique by cross-validating results on the training samples. Simulations on large data sets LETTER, SI and LAND are conducted in a high-performance computer with Intel Xeon, 3.16 GHz CPU and 16 GB RAM.

6.3 Performance Measures

The class-wise performance measures like overall/average efficiencies and a statistical significance test on performance of multiple classifiers on multiple data sets are used for performance comparison. The confusion matrix $\mathbf{Q}$ is used to obtain the class-level performance and global performance of the various classifiers. Class-level performance is measured by the percentage classification (η_j) which is defined as:

$$\eta_j = \frac{q_{jj}}{N_j} \times 100\% \quad (6.2)$$

where q_{jj} is the total number of correctly classified samples in the class j . The global measures used in the evaluation are the average per-class classification accuracy (η_a), the over-all classification accuracy (η_o) and the geometric classification accuracy (η_g) defined as:

$$\eta_a = \frac{1}{n} \sum_{j=1}^n \eta_j \quad (6.3)$$

$$\eta_o = \frac{\sum_{j=1}^n q_{jj}}{N} \times 100\% \quad (6.4)$$

$$\eta_g = \sqrt[n]{\eta_1 \eta_2 \cdots \eta_n} \quad (6.5)$$

6.3.1 Statistical Significance Test

The classification efficiency itself is not a conclusive measure of a classifier performance [66]. Since the developed classifier is compared with multiple classifiers over multiple data sets, the Friedman test followed by the Benferroni-Dunn test are used to establish the statistical significance of PBL-McRBFN classifier. A brief description of the conducted test is given below.

The Friedman test is used to compare multiple classifiers (U) over multiple data sets (V). Let r_i^j be the rank of the j^{th} classifier on the i^{th} data set. The Friedman test compares the average ranks of classifiers, $R_j = \frac{1}{V} \sum_i r_i^j$. Under the null-hypothesis, which states that all the classifiers are equivalent and so their ranks R_j should be equal, the Friedman statistic is given by

$$\chi_F^2 = \frac{12V}{U(U+1)} \left[\sum_j R_j^2 - \frac{U(U+1)^2}{4} \right] \quad (6.6)$$

which follows the χ^2 (Chi-square distribution) distribution with $U - 1$ degrees of freedom. A χ^2 distribution is the distribution of a sum of squares of U independent standard normal variables.

Iman and Davenport showed that Friedman's statistic (χ_F^2) is more conservative and derived a better statistic in [67]. It is given by

$$F_F = \frac{(V - 1)\chi_F^2}{V(U - 1) - \chi_F^2} \quad (6.7)$$

which follows the F -distribution with $U - 1$ (df1) and $(U - 1)(V - 1)$ (df2) degrees of freedom is used in this paper. F-distribution is defined as the probability distribution of the ratio of two independent χ^2 distributions over their respective degrees of freedom. The aim of the statistical test is to prove that the performance of PBL-McRBFN classifier is substantially different from the other classifiers with a confidence level of value $1 - \alpha$. If calculated $F_F > F_{\alpha/2, (U-1), (U-1)(V-1)}$ or $F_F < F_{1-\alpha/2, (U-1), (U-1)(V-1)}$, then the null-hypothesis is rejected. The statistical tables for critical values can be found in [68].

The Benferroni-Dunn test [69] is a post-hoc test that can be performed after rejection of the null-hypothesis. It is used to compare PBL-McRBFN classifier against all the other classifiers. This test assumes that the performances of two classifiers are significantly different if the corresponding average ranks differ by at least the critical difference (CD)

$$CD = q_\alpha \sqrt{\frac{U(U + 1)}{6V}} \quad (6.8)$$

where critical values q_α are based on the studentized range statistic divided by $\sqrt{2}$ as given in [66].

6.4 Performance Evaluation

The performance evaluation on binary and multi-category data sets presented separately as follows:

6.4.1 Binary-class Data Sets

The performance measures such as overall (η_o), average (η_a) testing efficiencies, number of neurons and samples used for PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM

Table 6.2: Performance comparison of PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM on binary class data sets

Data set	Classifier	# Neurons	# Samples used (%)	Training Time (sec)	Testing	
					η_o	η_a
HEART	SVM	42 ^a	100	0.03	75.5	75.10
	ELM	36	100	0.15	76.50	75.91
	SRAN	28	80	0.53	78.50	77.52
	EKF-McRBFN	26	65.7	0.55	80.50	79.65
	PBL-McRBFN	20	98.5	0.53	81.50	81.47
LD	SVM	141 ^a	100	0.30	71.03	70.21
	ELM	100	100	0.15	72.41	71.41
	SRAN	91	75.5	3.37	66.84	65.77
	EKF-McRBFN	68	55	2.05	73.79	71.60
	PBL-McRBFN	87	58	2.95	73.10	72.63
PIMA	SVM	221 ^a	100	0.20	77.45	76.33
	ELM	100	100	0.20	76.63	75.25
	SRAN	97	57.5	12.24	78.53	74.90
	EKF-McRBFN	76	48.2	6.45	80.16	77.31
	PBL-McRBFN	100	41.2	5.95	79.62	79.13
BC	SVM	24 ^a	100	0.04	96.61	97.06
	ELM	66	100	0.12	96.35	96.48
	SRAN	7	30.3	0.17	96.87	97.26
	EKF-McRBFN	9	9	0.62	97.39	97.85
	PBL-McRBFN	13	15	0.60	97.39	97.85
ION	SVM	43 ^a	100	0.20	91.24	88.51
	ELM	32	100	0.18	89.64	87.52
	SRAN	21	86	3.71	90.84	91.88
	EKF-McRBFN	20	39	1.02	95.62	95.60
	PBL-McRBFN	18	58	0.52	96.41	96.47

^aNumber of support vectors

classifiers on all the 5 binary class data sets are reported in Table 6.2. From the performance comparison results in Table 6.2, one can see that in case of low dimensional LD and PIMA data sets, the proposed PBL-McRBFN uses fewer samples for training and achieves better generalization performance approximately 1-2% improvement over EKF-McRBFN, 4-7% improvement over SRAN, 1-4% improvement over ELM and SVM with less number of neurons. In case of simple BC data set, PBL-McRBFN uses fewer samples for training and achieves slightly better generalization performance approximately 1% improvement over SRAN, ELM and SVM and same performance as EKF-McRBFN. In case of large dimensional HEART and ION data sets, PBL-McRBFN uses fewer samples for training and achieves better generalization performance 1-2% improvement over EKF-McRBFN, 4-5% improvement over SRAN, 6-9% improvement over SVM and ELM. The overlapping conditions and class specific criterion in learning strategies of PBL-McRBFN, EKF-McRBFN helps in capturing the knowledge accurately in case of high sample imbalance problems. We can notice that the PBL-McRBFN takes less computational time than EKF-McRBFN on all data sets.

6.4.2 Multi-category Data Sets

The testing efficiencies η_o and η_a , number of neurons and samples used for PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM classifiers on all the 10 multi-category data sets are reported in Table 6.3. From Table 6.3, one can see that in case of well balanced IS, IRIS, and WINE data sets, PBL-McRBFN uses only 42-50% training samples to achieve better generalization performance approximately 2-3% improvement over EKF-McRBFN, SRAN, SVM and ELM with the less number of neurons. Sample deletion criterion in meta-cognitive component helps in removing redundant samples from the training set and thereby improves the generalization performance. For highly unbalanced data sets, the proposed PBL-McRBFN is able to achieve significantly better performance than the other classifiers. In case of VC and GI data sets, PBL-McRBFN uses fewer samples and achieves better generalization performance approximately 1-5% improvement over EKF-McRBFN and ELM, 3-12% improvement over SRAN and SVM. In case of low dimensional AE data set, PBL-McRBFN achieves slightly better generalization performance 1% improvement over SVM, similar to EKF-McRBFN, SRAN and

Table 6.3: Performance comparison on multi-category data sets

Data set	Classifier	# Neurons	# Samples used (%)	Training Time (sec)	Testing	
					η_o	η_a
IS	SVM	127 ^a	100	0.72	91.38	91.38
	ELM	49	100	0.22	90.23	90.23
	SRAN	47	53.8	22	92.29	92.29
	EKF-McRBFN	49	44.2	10.67	93.38	93.38
	PBL-McRBFN	50	42.3	1.69	94.19	94.19
IRIS	SVM	13 ^a	100	0.02	96.19	96.19
	ELM	10	100	0.01	96.19	96.19
	SRAN	8	64.4	0.08	96.19	96.19
	EKF-McRBFN	5	48.8	0.47	97.14	97.14
	PBL-McRBFN	6	44.4	0.38	98.10	98.10
WINE	SVM	36 ^a	100	0.06	97.46	98.04
	ELM	10	100	0.03	97.46	98.04
	SRAN	12	76.6	0.16	96.61	97.19
	EKF-McRBFN	9	45	0.55	98.30	98.49
	PBL-McRBFN	11	48.3	0.45	98.31	98.69
VC	SVM	340 ^a	100	1.3	70.62	68.51
	ELM	150	100	0.36	77.01	77.59
	SRAN	113	70.4	55	75.12	76.86
	EKF-McRBFN	146	57.9	562.12	77.72	78.72
	PBL-McRBFN	175	51.2	21.83	78.91	79.09
AE	SVM	22 ^a	100	0.11	98.54	97.95
	ELM	10	100	0.16	99.27	98.91
	SRAN	10	62.9	0.09	99.27	98.91
	EKF-McRBFN	5	32.2	0.69	99.27	98.91
	PBL-McRBFN	5	14.5	0.45	99.27	98.91
GCM	SVM	137 ^a	100	0.03	76.08	74.76
	ELM	55	100	0.03	76.08	80.23
	SRAN	92	77	345.68	78.26	71.42
	EKF-McRBFN	71	76.3	325.11	76.08	73.57
	PBL-McRBFN	72	75.6	0.81	93.47	91.67
GI	SVM	183 ^a	100	0.92	70.47	75.61
	ELM	80	100	0.38	81.31	87.43
	SRAN	59	47.3	28	86.21	80.95
	EKF-McRBFN	73	34.8	13.61	85.71	87.03
	PBL-McRBFN	71	34.2	2.90	84.76	92.72
SI	SVM	1298 ^a	100	-	92.21	90.14
	ELM	1500	100	-	88.39	-
	PBL-McRBFN	1243	32.1	2118.98	90.76	92.17
LETTER	SVM	4429 ^a	100	-	92.94	-
	ELM	-	100	-	93.51	-
	PBL-McRBFN	1654	25	9875.15	95.42	95.44
LAND	SVM	981 ^a	100	4.64	87.90	86.08
	ELM	380	100	2.27	87.45	84.20
	PBL-McRBFN	245	20.6	26.75	89.35	88.56

^aNumber of support vectors

ELM. In case of large dimensional GCM data set, PBL-McRBFN achieves significantly better generalization performance approximately 11-20% improvement over SVM, ELM, SRAN and EKF-McRBFN. In case of large samples LETTER, SI and LAND data sets, the generalization performance of PBL-McRBFN is better than ELM and SVM by approximately 2%, 2-3% and 2-4% respectively using less number of training samples and neurons. We can notice that the PBL-McRBFN takes less computational time than EKF-McRBFN on all data sets. Due to computational complex EKF parameter update, EKF-McRBFN and SRAN experiences memory problem for large problems like Letter, SI and LAND. Hence, the results for EKF-McRBFN and SRAN on these problems are not presented here.

From the Tables 6.2 and 6.3, we can say that the proposed PBL-McRBFN improves generalization performance under wide range of sample imbalance data sets.

6.4.3 Statistical Performance Comparison

In order to compare the performance of the proposed PBL-McRBFN over EKF-McRBFN, SRAN, ELM and SVM classifiers on various benchmark data sets, we employ a non-parametric Friedman test followed by the Benforroni-Dunn test as described in [66]. The Friedman test compares whether the mean of individual experimental condition differs significantly from the aggregate mean across all conditions. If the measured F -statistic is greater than the critical F -Statistic at 95% confidence level then one rejects the equality of mean hypothesis (the classifiers used in our study perform similarly on different data sets). If Friedman test rejects the equality hypothesis, then pair-wise post-hoc should be conducted to test which mean is different from others. We have used 5 different classifiers ranks based on η_o and η_a testing efficiencies on 12 different data sets from the Table 6.1.

- *Comparison based on the overall testing efficiency (η_o)* : Ranks of all 5 classifiers based on the overall testing efficiency for each data set are provided in Table 6.4. The Friedman statistic (χ_F^2 as in Eq. (6.6)) is **25.49** and modified (Iman and Davenport) statistic (F_F as in Eq. (6.7)) is **13.78**. For 5 classifiers and 12 data sets, the modified statistic is distributed according to the F -distribution with **4** and **44** degrees-of-freedom. The critical value for rejecting the null hypothesis at 95%

Table 6.4: Ranks based on the overall testing efficiency (η_o)

Data sets	PBL-McRBFN	EKF-McRBFN	SRAN	ELM	SVM
HEART	1	2	3	4	5
LD	2	1	5	3	4
PIMA	2	1	3	5	4
BC	1.5	1.5	3	5	4
ION	1	2	4	5	3
IS	1	2	3	5	4
IRIS	1	2	4	4	4
WINE	1	2	5	3.5	3.5
VC	1	2	4	3	5
AE	2.5	2.5	2.5	2.5	5
GI	3	2	1	4	5
GCM	1	4	2	4	4
Average rank ($\mathbf{R}_j$)	1.5	2.00	3.29	4.00	4.20

Table 6.5: Two-tailed critical values (F -distribution) for the Friedman test at 95% confidence level

	df1=1	2	3	4	5	6	7	8	9	10
df2=44	5.385	4.016	3.429	3.093	2.871	2.711	2.591	2.496	2.419	2.355

confidence level ($F_{4,44,0.025}$) is 3.09 and the corresponding reference F -distribution table is given in Table 6.5. Since, modified statistic is greater than the critical value ($\mathbf{13.78} \gg \mathbf{3.09}$), we can reject the null hypothesis at a confidence of level 95%. Hence, we can say that the performance of all 5 classifiers is different on these 12 data sets based on overall testing efficiency.

Next, we conduct a pair-wise comparison using a Benforroni-Dunn test to highlight the performance significance of PBL-McRBFN classifier with respect to other classifiers. Here, the proposed PBL-McRBFN classifier is used as a control. From the

Table 6.6: Critical values for the Benferroni-Dunn test at 95% confidence level

	# Classifiers=2	3	4	5	6	7	8	9	10
$\mathbf{q_{0.05}}$	1.960	2.241	2.394	2.498	2.576	2.638	2.690	2.724	2.773

Eq. (6.8), the critical difference (CD) is calculated is 1.61 at at 95% confidence level and the reference q_α values for different number of classifiers are provided in Table 6.6. From the Table 6.4, we can obtain the average ranks for all five classifiers are PBL-McRBFN: 1.50, EKF-McRBFN: 2.00, SRAN: 3.29, ELM: 4.00, and SVM: 4.20. The difference in average rank between the proposed PBL-McRBFN classifier and the other four classifiers are PBL-McRBFN&EKF-McRBFN: 0.50, PBL-McRBFN&SRAN: 1.79, PBL-McRBFN&ELM: 2.50 and PBL-McRBFN&SVM: 2.70. Note that difference in average rank for PBL-McRBFN&SRAN, PBL-McRBFN&ELM and PBL-McRBFN&SVM pairs are greater than critical difference (CD) at 95% confidence level, i.e, $1.79 > 1.61$, $2.50 > 1.61$ and $2.70 > 1.61$. The difference in average rank for PBL-McRBFN&EKF-McRBFN pair is less than CD at 95% confidence level, i.e, $0.50 < 1.61$. Hence, we can say that PBL-McRBFN performs slightly better than the EKF-McRBFN classifier and significantly better than SRAN, ELM and SVM classifiers with a confidence of 95% based on the overall testing efficiency.

- *Comparison based on the average testing efficiency (η_α)* : Ranks of all 5 classifiers based on the average testing efficiency for each data set are provided in Table 6.7. The Friedman statistic (χ_F^2 as in Eq. (6.6)) is **27.69** and modified statistic (F_F as in Eq. (6.7)) is **16.99**. Since, modified statistic is greater than the critical value (**16.99** >> **3.09**), we can reject the null hypothesis at a confidence of level 95%. Hence, we can say that the performance of all 5 classifiers is different on these 12 data sets based on average testing efficiency.

From the Table 6.7, we can obtain the average ranks for all five classifiers and are PBL-McRBFN: 1.16, EKF-McRBFN: 2.25, SRAN: 3.87, ELM: 3.58, and SVM: 4.12. The difference in average rank between the proposed PBL-McRBFN classifier and the other three classifiers are PBL-McRBFN&EKF-McRBFN: 1.09, PBL-McRBFN&SRAN: 2.71, PBL-McRBFN&ELM: 2.42 and PBL-McRBFN&SVM: 2.96. Note that difference in average rank for PBL-McRBFN&SRAN, PBL-McRBFN&ELM and PBL-McRBFN&SVM pairs are greater than critical difference (CD) at 95% confidence level, i.e, $2.71 > 1.61$, $2.42 > 1.61$ and $2.96 > 1.61$. The difference in average rank for PBL-McRBFN&EKF-McRBFN pair is less than CD

Table 6.7: Ranks based on the average testing efficiency (η_a)

Data sets	PBL-McRBFN	EKF-McRBFN	SRAN	ELM	SVM
HEART	1	2	3	4	5
LD	1	2	5	3	4
PIMA	1	2	5	4	3
BC	1.5	1.5	3	5	4
ION	1	2	3	5	4
IS	1	2	3	5	4
IRIS	1	2	4	4	4
WINE	1	2	5	3.5	3.5
VC	1	2	4	3	5
AE	2.5	2.5	2.5	2.5	5
GI	1	3	4	2	5
GCM	1	4	5	2	3
Average rank (R_j)	1.16	2.25	3.87	3.58	4.12

at 95% confidence level, i.e, $1.09 < 1.61$. Hence, we can say that PBL-McRBFN performs slightly better than the EKF-McRBFN classifier and significantly better than SRAN, ELM and SVM classifiers with a confidence of 95% based on the average testing efficiency.

6.4.4 10 Random Trial Results

For this performance comparison study, 10 random trial experiments are conducted for the PBL-McRBFN, EKF-McRBFN, ELM and SVM classifiers on 12 different data sets from the Table 6.1 except LETTER, SI and LANDSAT data sets. For this study, 10 random trial data sets are generated by maintaining the imbalance factor for 12 data sets.

Binary class data sets : The performance measures such as overall (η_o), geometric (η_g) testing efficiencies and F -score, number of neurons and samples used for PBL-McRBFN, EKF-McRBFN, ELM and SVM classifiers on all the 5 binary class data sets are reported in Table 6.8. From the performance comparison results in Table 6.8, one can see that

Table 6.8: PBL-McRBFN, EKF-McRBFN, SRAN, ELM and SVM classifiers 10 random trial results comparison on binary class data sets

Data set	Classifier	# Neurons		# Samples used (%)		Testing					
		Mean	Dev	Mean	Dev	η_o		η_g		$F - score$	
						Mean	Dev	Mean	Dev	Mean	Dev
HEART	SVM	44 ^a	7.39	100	0	77.3	2.69	77.35	2.56	0.76	0.02
	ELM	46.5	2.41	100	0	73.2	2.79	73.18	2.93	0.71	0.03
	EKF-McRBFN	20.2	3.96	67.5	10.87	78.25	3.29	78.01	3.18	0.76	0.02
	PBL-McRBFN	28.2	2.39	81.2	12.2	81.7	1.13	81.37	1.02	0.79	0.01
LD	SVM	157.5 ^a	4.72	100	0	69.21	2.1	62.91	4.17	0.75	0.02
	ELM	127	15.67	100	0	64.55	3.8	63.52	4.04	0.68	0.04
	EKF-McRBFN	64.9	8.65	53.2	8.83	67.79	1.86	64.46	4.15	0.73	0.01
	PBL-McRBFN	78.1	8.55	65.1	8.22	70.82	1.08	69.87	1.57	0.74	0.01
PIMA	SVM	252.7 ^a	42.28	100	0	76.76	1.45	66.78	6.58	0.57	0.07
	ELM	172	25.73	100	0	70.86	1.44	65.50	2.67	0.53	0.03
	EKF-McRBFN	79.8	6.66	46.1	7.90	75.67	1.67	73.56	1.35	0.63	0.01
	PBL-McRBFN	100.8	8.67	46.4	7.39	74.45	2.12	74.41	1.00	0.63	0.01
BC	SVM	27.7 ^a	3.6	100	0	96.55	0.57	96.41	0.67	0.94	0.01
	ELM	37.9	1.34	100	0	96.97	0.47	96.89	0.87	0.95	0.01
	EKF-McRBFN	9.4	3.47	13.4	4.99	97.72	0.52	98.12	0.35	0.96	0.0
	PBL-McRBFN	12.3	4.02	35.8	11.61	97.77	0.60	97.95	0.37	0.96	0.01
ION	SVM	70.9 ^a	10.27	100	0	91.24	1.11	90.10	2.69	0.87	0.02
	ELM	46	1.76	100	0	80.26	2.3	74.63	2.59	0.69	0.03
	EKF-McRBFN	18.4	6.16	51.5	11.08	90.63	3.31	90.47	2.92	0.87	0.04
	PBL-McRBFN	20.6	3.68	47.5	11.14	93.90	1.31	93.74	1.78	0.91	0.01

^aNumber of support vectors

in case of low dimensional LD and PIMA data sets, the proposed PBL-McRBFN uses fewer samples for training and achieves significantly better generalization performance approximately 4-9% improvement over ELM and SVM with less number of neurons, and achieves better generalization performance approximately 1-5% improvement over EKF-McRBFN with slightly more number of neurons. In case of simple BC data set, PBL-McRBFN and EKF-McRBFN performs similarly, uses fewer samples for training and achieves slightly better generalization performance approximately 1% improvement over ELM and SVM with less number of neurons. In case of large dimensional HEART and ION data sets, PBL-McRBFN uses fewer samples for training and achieves better generalization performance 3-4% improvement over SVM and 18-19% improvement over ELM, and achieves better generalization performance approximately 3% improvement over EKF-McRBFN.

Multi-category data sets : Overall (η_o) and geometric (η_g) testing efficiencies, num-

ber of neurons and samples used for PBL-McRBFN, EKF-McRBFN, ELM and SVM classifiers on all the 10 multi-category data sets are reported in Table 6.9. From Table 6.9, we can see that PBL-McRBFN performs similar to EKF-McRBFN and significantly better than the ELM and SVM on all the **10** multi-category data sets. In case of well balanced IS, IRIS, and WINE data sets, PBL-McRBFN and EKF-McRBFN uses only 45-55% training samples to achieve 2-3% more generalization performance over SVM and ELM with the less number of neurons. Meta-cognitive sample deletion criteria helps in removing redundant samples from the training set and thereby improving the generalization performance. For highly unbalanced data sets, one can see that the proposed PBL-McRBFN is able to achieve significantly better performance than the other classifiers. In case of VC and GI data sets, PBL-McRBFN uses fewer samples and achieves better generalization performance approximately 2-8% improvement over EKF-McRBFN, 9% improvement over SVM and 3-6% improvement over ELM. In case of low dimensional AE data set, PBL-McRBFN and EKF-McRBFN achieves slightly better generalization performance 1% improvement over SVM and ELM. In case of large dimensional GCM data set, with less number of neurons and using few samples PBL-McRBFN achieves better generalization performance approximately 3% improvement over EKF-McRBFN and significantly better generalization performance approximately 13% improvement over SVM and ELM.

From the Tables 6.8 and 6.9, we can see the better generalization performance of PBL-McRBFN than other classifiers based on 10 random trial results.

Statistical comparison based geometric testing efficiency (η_g):

In order to compare the performance of the proposed PBL-McRBFN classifier over EKF-McRBFN, ELM and SVM classifiers on various benchmark data sets statistically based geometric testing efficiency (η_g), we employ an one-way repeated analysis of variance (ANOVA) measure followed by a pair wise comparison using post-hoc Dunnett test [70]. The ANOVA measure compares whether the mean of individual experimental condition differs significantly from the aggregate mean across all conditions. If the measure F score is greater than the F -Statics at 95% confidence level then one rejects the equality of mean hypothesis (the classifiers used in our study perform similarly on different data sets).

Table 6.9: 10 random trial results comparison on multi-category data sets

Data set	Classifier	# Neurons		# Samples used(%)		Testing			
		Mean	Dev	Mean	Dev	η_o		η_g	
						Mean	Dev	Mean	Dev
IS	SVM	107.3 ^a	10.91	100	0	90.92	0.43	90.13	0.46
	ELM	55.59	10.91	100	0	90.13	0.72	89.40	0.83
	EKF-McRBFN	36.8	5.63	54.1	8.60	91.64	0.95	91.16	1.08
	PBL-McRBFN	48.9	4.67	44.3	5.60	91.50	1.31	91.07	1.39
IRIS	SVM	16.6 ^a	0.93	100	0	95.99	1.17	95.90	1.22
	ELM	16	4.59	100	0	96.38	1.08	96.29	1.16
	EKF-McRBFN	5.2	1.93	44.6	16.51	96.95	1.17	96.86	1.23
	PBL-McRBFN	7.8	1.61	47.7	20.55	97.14	1.18	97.08	1.21
WINE	SVM	37 ^a	3.02	100	0	96.02	1.2	96.77	0.93
	ELM	15.2	2.09	100	0	94.83	1.35	95.58	1.08
	EKF-McRBFN	5.5	2.36	50	14.68	98.64	0.71	98.86	0.66
	PBL-McRBFN	10.6	1.77	53.6	14.43	98.31	1.16	98.62	1.17
VC	SVM	246.8 ^a	58.2	100	0	73.2	2.01	69.4	3.61
	ELM	150	25.3	100	0	77.09	3.44	75.16	3.83
	EKF-McRBFN	140	10.98	78.7	7.03	78.96	1.32	76.34	2.85
	PBL-McRBFN	210.8	7.28	68.39	11.65	79.62	1.28	78.33	1.55
AE	SVM	21.7 ^a	4.48	100	0	97.66	1.02	97.62	1.43
	ELM	14.5	3.68	100	0	97.88	0.87	97.34	1.11
	EKF-McRBFN	5.7	1.70	34.5	11.35	98.90	0.51	98.52	0.72
	PBL-McRBFN	7.2	2.69	30.4	9.56	98.31	1.31	98.13	1.25
GCM	SVM	120.6 ^a	5.78	100	0	76.53	2.80	63.91	5.59
	ELM	114.4	10.2	100	0	91.80	2.83	63.39	4.71
	EKF-McRBFN	71	1.58	73.6	4.13	71.73	0.03	73.17	7.79
	PBL-McRBFN	68.2	2.82	70.0	5.02	83.04	4.55	76.78	5.42
GI	SVM	176.6 ^a	18.9	100	0	77.23	5.01	85.83	3.62
	ELM	86	9.17	100	0	80.09	3.05	88.81	2.28
	EKF-McRBFN	73.2	9.17	41.8	3.45	80.00	4.62	86.67	4.18
	PBL-McRBFN	82.4	9.17	40.0	6.55	89.71	1.94	94.40	1.08

^aNumber of support vectors

Table 6.10: One-tailed critical values (F -distribution) for the ANOVA test at 95% confidence level

	df1=1	2	3	4	5	6	7	8	9	10
df2=33	4.139	3.284	2.891	2.658	2.502	2.389	2.302	2.234	2.178	2.132

Table 6.11: Two-tailed critical values (t -distribution) for the Dunnett test at 95% confidence level

df	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
33	1.692	2.034	2.348	2.378	3.008

If one-way repeated ANOVA test rejects the equality hypothesis, then pair-wise post-hoc should be conducted to test which mean is different from others. In our study, we have used 4 different classifiers and 12 different data sets. For a given data set and a classifier, 10 random trials are conducted to measure the mean and variance of the classifiers performance. The geometric testing efficiencies (η_g) of 4 classifiers on 12 data sets are organized as four groups and ANOVA monitors three kinds of variations in the data, viz., within-group variations, between-group variations and the total variation. The F -score obtained using repeated measures of one-way ANOVA test is 7.42, which is greater than the F -statistic at 95% confidence level ($F_{3,33,0.05}$ is 2.89), i.e., $7.42 > 2.89$ and the corresponding reference F -distribution table is given in Table 6.10. Hence, one can reject mean equality hypothesis at a confidence of level 95%, i.e., performances of 4 classifiers are different on different data sets.

Next, we conduct a pair-wise comparison using a parametric Dunnett test to highlight the performance significance of PBL-McRBFN classifier with respect to other classifiers. Here, the proposed PBL-McRBFN classifier is used as a control. The t -observed obtained for individual pairs are PBL-McRBFN & EKF-McRBFN: 1.93, PBL-McRBFN & ELM: 4.31 PBL-McRBFN & SVM: 3.63. Note that the t -observed PBL-McRBFN&ELM and PBL-McRBFN&SVM pairs are greater than t -critical ($t_{33,0.025}$ is 2.34 at 95% confidence level; The corresponding reference t -distribution table is given in Table 6.11) and the t -observed PBL-McRBFN&EKF-McRBFN pair is less than t -critical. Hence, we can say that PBL-McRBFN performs slightly better than the EKF-McRBFN classifier and significantly better than ELM and SVM classifiers with a confidence of 95% based on 10 random trial performance evaluation study.

6.5 Work Flow of Meta-cognitive Strategies

In this section we exemplify the dynamic working nature of meta-cognitive strategies (delete, growth, update and reserve) using PBL-McRBFN for Image segmentation (IS) data set. IS is a seven classes multi-category data set contains 210 training samples and 2100 testing samples. The tuneable parameters are chosen as follows: deletion threshold (β_d) = 0.9, knowledge threshold (β_c) = 0.5117, addition threshold (β_a) = 1.3, update threshold (β_u) = 0.6769, self-adaption slope control parameter (δ) = 0.9927, hidden neuron overlap constant (κ) = 0.655 and hidden neuron center shifting factor (ζ) = 0.0111. We shall now consider how the different strategies of meta-cognitive learning in PBL-McRBFN aid to its accurate classification.

- *Sample delete strategy:* When the predicted class label of the new training sample is same as the actual class label and the confidence level (estimated posterior probability) is greater than expected value then the new training sample does not provide additional information to the classifier and can be deleted from training sequence without being used in learning process. We shall exemplify the working of this strategy with a Fig. 6.1. Fig. 6.1 gives a snap-shot of prediction confidence of PBL-McRBFN classifier for samples in range 100-150 along with the deletion threshold (β_d). Those samples with confidence level greater than β_d are deleted without participating in the PBL-McRBFN learning process. In Fig. 6.1, the confidence of PBL-McRBFN classifier for sample at instant 130 is higher than deletion threshold ($\beta_d = 0.9$) and is thus deleted without participating in the learning process. By deleting correctly classified samples with negligible novel information, the sample deletion strategy helps the network to avoid over-training and save computational effort.
- *Neuron growth strategy:* When a new training sample contains novel information and the predicted class label is different from the actual class label then a new hidden neuron is added. The novelty of sample is determined by class-wise significance (ψ_c) and maximum hinge error (E^t). Let us study the effect of these measures in the IS problem by considering a snap-shot of 50-100 training samples. The class-wise significance for these samples and knowledge threshold (β_c) are given in Fig.

Figure 6.1: Exemplification of sample deletion strategy in PBL-McRBFN for Image segmentation data set

6.2(a), whereas Fig. 6.2(b) gives the hinge loss error, self-regulated addition and deletion threshold for these 50-100 training samples. A sample contains the novel information if the class-wise significance is less than the knowledge threshold ($\beta_c = 0.5117$) and hinge error is greater than the self-regulatory addition threshold ($\beta_a = 1.3$). It could be noticed from the Figs. 6.2(a)&(b) that even though the sample is novel, a new hidden neuron is added to the network if the hinge error criterion is not satisfied. Let us consider sample at instant 61, since both the class-wise significance measured is lower than the knowledge threshold and the instantaneous hinge error is higher than the addition threshold, a new hidden neuron is added to the network. The history of neuron and addition threshold in PBL-McRBFN learning process for IS data set are given in Fig. 6.3(a)&(b). The neuron history is plotted against the only samples used in training samples; PBL-McRBFN uses only 89 samples out of 210 training samples. One could notice from Fig. 6.3(b) that the self-regulatory addition threshold adapts its value based on the predicted maximum hinge error.

(a) Class-wise significance

(b) Instantaneous hinge error

Figure 6.2: Class-wise significance (a), and instantaneous hinge error with self-regulatory thresholds (b) in PBL-McRBFN for Image segmentation data set

(a) Neuron history

(b) Self-regulated addition threshold history

(c) self-regulated update threshold history

Figure 6.3: History of number of hidden neurons (a), self-regulated addition (b), and update thresholds (c) in PBL-McRBFN for Image segmentation data set

- *Parameters update strategy*: A correctly classified sample (i.e., Predicted class label is same as the actual class label) is used to update the network parameters in PBL-McRBFN for IS data set when the hinge error is greater than the self-regulatory update threshold ($\beta_a = 0.6769$). In Fig. 6.2(b), the sample at instant 79 does not contain significant novel information and the hinge error is greater than the update threshold, thus it is used to update the network parameters. The self-regulatory parameter update threshold adapts its value based on the predicted maximum hinge error as shown in Fig. 6.3(c).
- *Sample reserve strategy*: The samples which does not satisfy the either the deletion or the neuron growth or the parameters update criterion are reserved by network to be considered for learning later. These samples, by the virtue of the self-regulatory nature of addition and parameter update thresholds, may be used in the learning process at a later stage. In Fig. 6.2(b), the sample at instant 76 is reserved to be later in the learning process. There are a few such reserve samples in PBL-McRBFN training process for IS data set. It could be noticed that these samples will be pushed to the rear of the data-stream to be learnt later. It should be noted that these samples may be used in the learning process at a later stage to fine tune the network.

6.6 Study on the Effect of Meta-cognition

In this section, we analyze the effect of meta-cognitive strategies in PBL-McRBFN algorithm using Quantized Kernel Least Mean Square (QKLMS) [49] algorithm as a baseline. QKLMS algorithm is a recently developed online kernel adaptive filtering algorithm which is based on simple online vector quantization method. In QKLMS, quantization is applied to compress the input (or feature) space of the kernel adaptive filters so as to control the growth of the RBF structure. For this purpose, we have conducted three different experiments on the QKLMS algorithm using nine different data sets.

- (1) Effect of *how-to-learn*: The QKLMS algorithm is trained using original samples and its performance is compared with PBL-McRBFN algorithm. This experiment emphasizes the advantage of neuron growth and update strategies in PBL-McRBFN.

- (2) Effect of *when-to-learn*: The QKLMS algorithm is trained using samples selected by PBL-McRBFN in training, i.e., the selected samples in PBL-McRBFN training are sequentially presented in the same order in which they are used. We call this algorithm QKLMS*. This experiment emphasizes advantage of the self-adaptive nature of meta-cognitive thresholds and the sample reserve strategy in PBL-McRBFN.
- (3) Effect of *what-to-learn*: The QKLMS algorithm is trained using selected samples followed by deleted samples by PBL-McRBFN in training. We call this algorithm as QKLMS**. This experiment emphasizes the advantage of sample delete strategy in PBL-McRBFN.

The overall and average efficiencies of training and testing and the number of hidden neurons are given in Table 6.12 for all the nine benchmark data sets. From the results one can make the following observations.

- (1) PBL-McRBFN classifier training and testing performance is better than QKLMS on all the nine data sets. In case of large dimensional binary class ION data set, PBL-McRBFN uses fewer hidden neurons and achieves better testing performance 23% improvement over QKLMS. Also in case of unbalanced multi-category GI data set, PBL-McRBFN uses fewer hidden neurons and achieves better testing performance 15% improvement over QKLMS. This clearly shows that the meta-cognitive learning principles in PBL-McRBFN classifier help to achieve better performance results.
- (2) QKLMS* uses less training samples selected by PBL-McRBFN and achieves 1-3% improvement in testing over QKLMS with less number of hidden neurons. On large dimensional binary class ION data set, QKLMS* testing performance is 15% improved over QKLMS. This show *when-to-learn* the training sample is important in training a classifier.
- (3) QKLMS** uses all samples (selected samples followed by deleted samples by PBL-McRBFN), its training performance is similar or better than the QKLMS* algorithm. However, the testing performance is slightly lower than the QKLMS*. In case of large dimensional binary class ION data set, QKLMS** uses more hidden neurons and

Table 6.12: Effect of Meta-cognitive learning principles in the QKLMS algorithm

Data set	Classifier	# Neurons	Training		Testing	
			η_o	η_a	η_o	η_a
HEART	QKLMS	50	91.43	90.42	78.00	77.07
	QKLMS*	50	78.57	76.67	80.00	78.68
	QKLMS**	25	77.00	77.78	77.00	77.77
	PBL-McRBFN	20	100	100	81.50	81.47
PIMA	QKLMS	150	79.25	79.15	68.75	68.82
	QKLMS*	140	80.25	78.07	73.09	70.23
	QKLMS**	150	80.25	79.61	70.65	71.67
	PBL-McRBFN	100	92.25	91.19	79.62	79.13
BC	QKLMS	20	98.67	98.75	96.34	96.86
	QKLMS*	4	97.33	96.76	97.38	97.45
	QKLMS**	7	97.33	96.77	97.12	97.25
	PBL-McRBFN	13	99.00	98.83	97.39	97.85
ION	QKLMS	20	91.00	90.54	78.08	73.85
	QKLMS*	14	87.00	84.98	82.47	78.98
	QKLMS**	20	91.00	87.50	80.08	73.69
	PBL-McRBFN	18	99.00	99.22	96.41	96.47
IS	QKLMS	207	100	100	93.09	93.09
	QKLMS*	89	100	100	93.67	93.67
	QKLMS**	207	98.09	98.09	92.09	92.09
	PBL-McRBFN	50	98.57	98.57	94.19	94.19
IRIS	QKLMS	13	100	100	96.19	96.19
	QKLMS*	9	100	100	98.10	98.10
	QKLMS**	13	100	100	96.19	96.19
	PBL-McRBFN	6	100	100	98.10	98.10
WINE	QKLMS	20	96.67	96.67	92.37	93.91
	QKLMS*	14	96.67	96.67	95.76	96.19
	QKLMS**	19	95.00	95.00	93.22	94.03
	PBL-McRBFN	11	100	100	98.31	98.69
VC	QKLMS	300	88.68	88.64	70.85	70.95
	QKLMS*	298	93.39	93.29	73.46	73.58
	QKLMS**	620	92.92	92.71	69.19	69.19
	PBL-McRBFN	175	96.46	96.47	78.91	79.09
GI	QKLMS	80	89.91	95.02	73.33	77.21
	QKLMS*	68	87.15	93.48	78.09	78.88
	QKLMS**	80	94.49	95.80	72.38	74.45
	PBL-McRBFN	71	94.49	97.29	84.76	92.72

achieves 4% improvement in training and 5% decrement in testing over QKLMS*. Also in case of unbalanced multi-category GI data set, QKLMS** uses more hidden neurons and achieves 2% improvement in training and 4% decrement in testing over QKLMS*. This shows *what-to-learn* in training is important in a learning algorithm.

The PBL-McRBFN algorithm uses the meta-cognitive principles by implementing different (delete, growth, update and reserve) strategies and addresses the *what-to-learn*, *when-to-learn* and *how-to-learn* efficiently. The aforementioned results also clearly highlight that the use of meta-cognitive principles present in PBL-McRBFN improves the performance of the QKLMS algorithm significantly.

6.7 Summary

In this chapter, we have presented the performance evaluation study of proposed EKF-McRBFN and PBL-McRBFN using a number of benchmark multi-category, binary classification problems with a wide range of imbalance factor. The qualitative and quantitative performance analysis using multiple data sets clearly indicate the superior performance of the proposed PBL-McRBFN and EKF-McRBFN classifiers over other classifiers considered in this study. Results also show that PBL-McRBFN classifier performance is better than EKF-McRBFN classifier. Hence, in the next chapters, PBL-McRBFN is used in the early diagnosis of neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease.

Chapter 7

Alzheimer’s Disease Diagnosis using PBL-McRBFN Classifier

In this chapter we present an application of the proposed PBL-McRBFN classifier in the area of medical informatics particularly for early diagnosis of neurodegenerative diseases. Neurodegenerative diseases are generally considered as a group of diseases that seriously and progressively impair the functions of the nervous system through selective neuronal vulnerability of specific brain regions. Depending on their type, neurodegenerative diseases can be serious or life-threatening and most of them have no cure. The goal of treatment for such diseases is usually to improve symptoms, relieve pain and increase mobility. Alzheimer’s disease (AD) is the most common neurodegenerative disease [71]. Parkinson’s disease (PD) is the second most common neurodegenerative disease, after AD. The prevalence of AD and PD is increasing in the elderly [72].

In this chapter, we use PBL-McRBFN classifier for the early diagnosis of AD using MRI scans. Since, the classifier developed using PBL-McRBFN accurately approximates the decision boundary, we also propose a Recursive Feature Elimination approach (called PBL-McRBFN-RFE) to identify the most relevant and meaningful imaging biomarkers with a predictive power for diagnosis of AD.

AD is a progressive neurodegenerative disease that causes memory loss, problems in learning, confusion and poor judgment. AD is considered to be one of the most common causes of dementia among elderly persons. Dementia is a clinical syndrome characterized by significant loss or decline in memory and other cognitive abilities. Around 60–80% of age related dementia is caused due to AD [73]. The only way to make a definitive

diagnosis of AD is from a brain autopsy revealing the characteristics of the neurofibrillary tangles and amyloid plaques that define AD. Early detection of AD using non-invasive neuroimaging techniques will help in providing assistance to them and thereby one can slowdown the progression.

The literature review on AD detection is presented in the next section.

7.1 Literature Review on Alzheimer's Disease

Early detection of AD using non-invasive methods play a major role in providing treatment that may slow down its progress. One such non-invasive method of early detection of AD is by brain imaging. Commonly used brain imaging techniques for this purpose are: Computed Tomography (CT) [74, 75], Single-Photon Emission Computed Tomography (SPECT) [76, 77], Positron Emission Tomography (PET) [77] and Magnetic Resonance Imaging (MRI) [78, 79].

Studies using CT scans for AD diagnosis have been described in [74, 75]. However, due to a lower spatial resolution and the possibility of unreliable structural change detection in the early stages of the disease, CT has been employed only in very few cases. SPECT and PET are functional brain imaging techniques which use a radioactive substance to detect the changes in the blood flow and metabolism in the brain. For AD diagnosis, several studies using SPECT and PET images have been reported in [76, 77]. Both PET and SPECT involve the use of ionizing radiation and they are harmful if these are repeatedly used for the study. Hence, use of PET and SPECT in normal persons is typically limited to a single scan which may not provide adequate information for a proper diagnosis. Also, lack of spatial resolution in the SPECT images influence the accuracy in the detection of AD.

MRI is one of the most important brain imaging procedures that provides accurate information about the shape and volume of the brain. Compared to CT, SPECT and PET scans, MRI provides a high spatial resolution and can detect minute abnormalities in the brain. Usage of MRI for the accurate detection of AD has recently become very active [80, 79]. MRI helps to detect AD at an early stage - before irreversible damage has been done [13]. Early detection of AD from MRI requires appropriate methods to detect,

locate and quantify tissue atrophy in the brain. Primarily, the visual assessment of the degree of atrophy in the neuroanatomical structures is performed by an expert using MRI. However, this may be adequate in a normal clinical setting, but is not enough to get quantitative measures such as the fine incremental grades of atrophy and overall brain volume [81].

The early detection of AD from MRI can be cast as a binary classification problem and one can employ machine learning techniques to automatically detect the AD [78, 79, 82]. The main idea behind using machine learning techniques is to relate the brain volume changes to the onset of AD. Two major ways of estimating the brain volume changes from the MRI are: i) Region-of-Interest (ROI) approach; ii) Whole brain morphometric approach.

7.1.1 Region-of-Interest Approach

ROI approach has been traditionally used to obtain a regional measurement of the brain volume and to investigate the abnormal tissue structures with AD [83]. In the ROI approach, a volumetric analysis is performed by manually delineating specific brain regions. In practice, *a priori* knowledge about abnormal regions is not always available. However, in AD diagnosis, a lot of studies rely on manual tracing of hippocampus and entorhinal cortex which is laborious and time consuming [84, 85]. In [86], the volumes of manually segmented hippocampus and entorhinal cortex are measured to discriminate between the AD patients and normal persons. Major shortcomings in the use of the manual ROI approach are: it is dependent on the tracer’s expertise, is time-consuming and error-prone. Recently, an automatic method for the segmentation of the hippocampus using probabilistic and anatomical priors have been proposed for the detection of AD patients [82]. In [82], automatically segmented hippocampus volumes have been used to classify AD patients and normal persons. Although, ROI techniques for AD analysis have been widely used, they are difficult to accurately identify the brain volume changes in the AD patients when the tissue loss is generally smaller. To overcome these shortcomings, several approaches that enable the assessment of the whole brain have been reported in the literature [87, 88, 89, 90].

7.1.2 Whole Brain Morphometric Approach

Voxel Based Morphometry (VBM) is one of the widely used, fully automated, whole brain morphometric analysis [90]. VBM is based on the Statistical Parametric Mapping (SPM) method, often employed for the investigation of tissue volume changes between the brain MRI scans of the diseased group versus the normal persons. In the VBM analysis, the brain MRI undergo various preprocessing steps before undergoing the voxel-wise parametric tests [91]. The preprocessing steps involved in the VBM analysis are: normalization, segmentation, modulation and smoothing. VBM analysis identifies probability of the gray matter, white matter and cerebrospinal fluid tissue classes in a given voxel, where a voxel is defined as a volume element representing the intensity of a point in a three-dimensional space [90].

In the literature, AD classification studies have been conducted using morphometric features and Support Vector Machine (SVM) classifier [78, 79, 92, 78, 93, 94]. These methods use different morphological features and different data sets for AD detection. In [78], 90 samples (**33** probable mild AD patients and **57** normal) from Rochester, Minnesota is used. A statistical parametric map on gray matter tissues is obtained using these 90 samples and this map is used to extract the features for a SVM classifier. In [79], a mass-preserving Regional Analysis of Volumes Examined in Normalized Space (RAVENS) is used to extract the features from a smaller set of Baltimore longitudinal study data. Here, 15 probable mild AD patients and 15 normal person’s samples are used for the AD detection. RAVENS based feature extraction and SVM classifier provides good performance, but the feature extraction process is computationally intensive. For completely mild AD patients data, the computational effort in RAVENS increases further and it influences the accuracy of the features extracted which affects the SVM classifier performance.

SVM is based on the evaluation of discrimination power for classification, hence it has limitation in dealing with noisy data which is the case for neuroimaging data. Also, high dimensional VBM features make AD classification difficult and hence feature reduction techniques have been increasingly used for dimensionality reduction in neuroimage classification studies [79, 95, 96, 97]. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are the widely used feature construction techniques. PCA is

a subspace learning method and transfers the original features into a new linear subspace [79, 97]. ICA, as one of the important techniques of blind signal separation, has been shown to provide a powerful method for neuroimaging data [98, 99]. However, these techniques involve a careful selection of parameters such as the number of components to preserve the important subsets of the feature space. Also, the reduced features do not provide any information on the original voxels and hence the regions in the brain responsible for AD. Feature selection problem is also addressed using genetic algorithms, Integer Coded Genetic Algorithm (ICGA) has been used along with a neural network in [100]. However, the selection of size of population and other parameters affect the convergence of the genetic algorithm. Recursive Feature Elimination (RFE) is a computationally less intensive wrapper based feature selection method - selection of the features depends upon the classifier model.

Machine learning algorithms for AD detection requires samples with significant information as training samples. Ideal classifier for AD detection must incorporate sample selection in training for effective learning. Proposed PBL-McRBFN in chapter 5 selects the samples for learning and found to be effective for learning. Hence, in this chapter we propose a RFE approach with efficient classification method PBL-McRBFN (referred as PBL-McRBFN-RFE) to AD classification and identify critical imaging biomarkers relevant to AD at the same time.

In the next section, we present AD detection using PBL-McRBFN classifier with morphometric features obtained from MRI scans.

7.2 Early Diagnosis of Alzheimer’s Disease Based on MRI features

The framework of our method is shown in Fig. 7.1. First, from all MRI scans the morphometric features are extracted using the VBM. Next, high dimensional VBM features are used for classification using PBL-McRBFN. The following sections present a description of the MRI data, VBM analysis for feature extraction and performance results of PBL-McRBFN classifier in AD detection.

Figure 7.1: Schematic diagram of the AD detection using PBL-McRBFN classifier

7.2.1 Materials

7.2.1.1 OASIS data set

In our study, the publicly available Open Access Series of Imaging Studies (OASIS) data set has been used [101]. OASIS data set has a cross-sectional collection of 416 persons covering the adult life span, aged between 18 to 96 including individuals with AD in an early-stage. The data includes 218 persons aged between 18 to 59 years and 198 persons aged between 60 to 96 years. Of the 198 older persons, 98 had no AD i.e., with a Clinical Dementia Rating (CDR) of 0, 70 persons have been diagnosed with a very mild AD (CDR=0.5), 28 persons are diagnosed with mild AD (CDR=1) and 2 persons with moderate AD (CDR=2). The AD patients has scores between 14–30 on the Mini-Mental State Examination (MMSE) and normal persons have MMSE scores between 25–30. In our study, we have considered 198 elderly persons comprising of 98 normal persons and 100 AD patients. For each person, the whole-brain T1-weighted 3-dimensional MPRAGE (Magnetization-Prepared Rapid-Acquisition Gradient Echo) images has been acquired on a Siemens 1.5T scanner. The acquired volumes had 128 sagittal 1.25 mm slices without gaps and pixel resolution of 256×256 (1×1 mm). The OASIS data set demographics and the dementia details are summarized in Table 7.1.

Table 7.1: Demographic information of OASIS data used in our study

Group	Normal Persons	AD Patients
No. of persons	98	100
Percentage of Male	26.5%	41.0%
Age(mean $\pm$ std)	75.92 $\pm$ 8.99	74.76 $\pm$ 7.12
MMSE(mean $\pm$ std)	28.96 $\pm$ 1.21	24.32 $\pm$ 4.17
CDR 0/0.5/1/2	98/0/0/0	0/70/28/2

7.2.1.2 ADNI data set

In our study, we also used the data obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data set [102]. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies

and non-profit organizations, as a USD 60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid the researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and persons have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI is to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see (www.adni-info.org).

In our study, we have considered all the 432 elderly persons (232 normal persons and 200 AD patients) available in the ADNI data set as of February 2012. Standard 1.5T screening/baseline T1-weighted images obtained using volumetric 3D MPRAGE protocol with resolutions ranging from $0.9 \text{ mm} \times 0.9 \text{ mm} \times 1.20 \text{ mm}$ to $1.3 \text{ mm} \times 1.3 \text{ mm} \times 1.20 \text{ mm}$ are included from the ADNI data set. The detailed information of the MRI protocols and preprocessing steps are presented in [103]. The demographics for the 432 elderly persons used in our study is shown in Table 7.2.

Table 7.2: Demographic information of ADNI data used in our study

Group	Normal Persons	AD Patients
No. of persons	232	200
Percentage of Male	51.7%	51.5%
Age(mean $\pm$ std)	76.01 $\pm$ 5.00	75.65 $\pm$ 7.70
MMSE(mean $\pm$ std)	29.11 $\pm$ 1.00	23.29 $\pm$ 2.05
CDR 0/0.5/1/2	232/0/0/0	0/98/102/0

Figure 7.2: Schematic diagram of the stages in feature extraction based on the VBM analysis

7.2.2 Voxel Based Morphometry Based Feature Extraction

A feature extraction method based on the VBM is employed in this work [100]. The flow diagram of the feature extraction process is shown in Fig. 7.2. VBM is a voxel-wise comparison of local tissue volumes of gray matter within a group or across groups of persons using MRI scans. In our study, VBM is used to detect significant gray matter differences between the AD patients and normal persons. The VBM detected voxels locations of significant regions are further used as masks in order to extract the features from all gray matter segmented MRI scans. VBM is performed on the OASIS and ADNI data sets using the Statistical Parametric Map (SPM) software package [91].

In a VBM analysis, the brain MR images undergo various preprocessing steps before the voxel-wise parametric tests are carried out on them. In our study, VBM analysis based on a recently proposed unified segmentation model is performed [104]. The steps involved in VBM analysis are: unified segmentation, smoothing and statistical testing, in that order. The unified segmentation step is a generative modeling approach, in which tissue segmentation, bias correction and image registration are combined in a single model [104]. The unified segmentation framework combines deformable tissue probability maps with a Gaussian mixture model. The MR brain images of both the AD patients

and normal persons are segmented into gray matter tissue class. The segmented and normalized gray matter images are then smoothed by convolving them with an isotropic Gaussian kernel. In our approach, a 10 mm full-width at half-maximum Gaussian kernel is employed.

For better understanding of the VBM analysis, we have shown three planar views (sagittal, coronal and axial) of the original images and images after every stage in the VBM analysis in Fig. 7.3(a-c). Fig. 7.3(a) shows the different planar views of the MRI scan. From the MRI scan, one has to perform a bias correction, tissue segmentation and register the segmented image in a standard template for removing non-uniform artifacts. The images after undergoing these steps are shown in Fig. 7.3(b). From Fig. 7.3(b), we can see that the unified segmentation in the VBM analysis efficiently identifies the gray matter from the MRI scans. These segmented images are now smoothed by convolving them with an isotropic Gaussian kernel and the resultant images are shown in Fig. 7.3(c). The smoothing process averages the concentration of the gray matter around the voxel and this helps considerably in the subsequent voxel-by-voxel statistical analysis [104].

The smoothed brain volumes of AD patients and normal persons are used in the statistical analysis to identify regions of gray matter concentration that are significantly related to the AD. These regions will be used to extract the features for accurate identification of AD. For the statistical analysis, a general linear model is used to detect the volumetric changes in gray matter across the AD patients and normal persons. In our statistical analysis, estimated total intracranial volume is used as a covariate in the design matrix of the general linear model. Also a two-sample t -test is performed on the smoothed images of normal persons and AD patients and a multiple comparison correction method, namely, a family wise error with $\mathbf{P} < \mathbf{0.05}$ has been applied. Following the application of the general linear model and statistical tests, the significance of any differences in gray matter volume is ascertained using the theory of Gaussian random fields [105]. These tests result in a maximum intensity projection map, which will be then used to extract the features from individual segmented gray images for further analysis. For better understanding, we show the maximum intensity projections of the significant voxels in sagittal, coronal and axial views in Fig. 7.4 and Fig. 7.5.

From Fig. 7.4 and Fig. 7.5, it can be noted that there are significant areas with decreased gray matter density in the AD patients relative to the normal persons indicating

Figure 7.3: Results of the unified segmentation and smoothing steps performed on MRI of an AD patient (from right: sagittal view, coronal view and axial view)

Figure 7.4: Maximum intensity projections from OASIS data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view

Figure 7.5: Maximum intensity projections from ADNI data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view

Figure 7.6: Gray matter volume change from OASIS data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view

that for the AD patients gray matter in these locations are lower. A total of 19879/23797 features are extracted from OASIS/ADNI data sets using the above VBM analysis and these features are then used for classification of AD patients. It is found in the literature that the VBM produces different significant areas with increased gray matter density in the brain, when the voxel-wise statistics with different groups of persons (e.g., male Vs female, only female etc.) and different covariates (e.g., gender, age etc.) are used [78, 106]. This also implies that by employing the above VBM analysis one can obtain different sets (with varying numbers) of feature vectors.

To locate the above regions with respect to the spatial locations in the brain, these regions are overlaid on the sliced sections of the commonly used Montreal Neurological Institute (MNI) brain template and the results of the same are shown in Fig. 7.6 and Fig. 7.7. From Fig. 7.6 and Fig. 7.7, one can infer the regions of the brain which get affected significantly for the AD patients. In other words, during the MRI scans if we can notice that the gray matter in these specific locations are lower, one can infer a good likelihood of developing AD later in these patients.

Figure 7.7: Gray matter volume change from ADNI data set - Normal persons vs. AD patients (a) sagittal view (b) coronal view (c) axial view

7.2.3 Experimental Results

In this section, we present the systematic studies that have been carried out on AD detection using the PBL-McRBFN classifier with MRI. Before, we present the details of the study results, we highlight the sequence in which the studies have been carried-out. First, we present the performance results of PBL-McRBFN classifier for AD detection using OASIS/ADNI data sets and also compare the performance with existing results in the literature. Next, the generalization capability of the PBL-McRBFN classifier is shown by testing the ADNI samples on PBL-McRBFN classifier developed using the OASIS data set. Further, we present a method to identify the imaging biomarkers for AD using proposed a RFE method for feature reduction and the PBL-McRBFN classifier. Finally, we present the detailed studies based on age/gender groups in OASIS data to identify the imaging biomarkers for AD.

Table 7.3: Classification performance of PBL-McRBFN on the OASIS data set

Feature Type	# Features	# Neurons	Training			Testing		
			Accuracy(%) mean (std)	Sensitivity(%) mean (std)	Specificity(%) mean (std)	Accuracy(%) mean (std)	Sensitivity(%) mean (std)	Specificity(%) mean (std)
VBM+ICA	50	49	93.47(1.71)	95.92(2.59)	91.02(3.09)	72.33 (2.48)	72.00(2.00)	72.65(6.85)
VBM	19879	43	92.86(1.90)	91.83(6.92)	93.88(4.99)	75.80 (1.01)	71.20(7.82)	80.41(7.98)

7.2.4 PBL-McRBFN Classifier Performance on the OASIS Data Set

The complete OASIS data set consists of 198 samples. For each sample 19879 morphometric features were extracted using the VBM. However, as the feature space is large, all features may not be responsible for AD. Hence, the obtained morphometric features from VBM analysis are further reduced statistically by using the ICA. We employed the FastICA (fixed-point) algorithm [107] - the FastICA package for MATLAB [108] is used on the VBM detected morphometric features and reduced to 50 features.

We conducted 10 random trials of experiments using PBL-McRBFN classifier on the complete 19879 feature set and reduced 50 features set with each trial containing 50% total samples for training and the remaining for testing. The training/testing accuracy, sensitivity and specificity obtained from the PBL-McRBFN classifier is presented in Table 7.3. PBL-McRBFN produces the testing accuracy using 50 reduced features is 72.33%. PBL-McRBFN testing accuracy on complete feature set is 75.8% which is **3%** more than the 50 reduced feature set, this is because the considered binary classification problem consists of MRI scans of 100 ‘very mild to moderate AD’ patients and 98 healthy elder persons. Hence, PBL-McRBFN classifier requires all morphometric features to classify AD patients group consists of wide range of CDR from **0.5 – 2** and healthy elder persons.

Next, we compare the PBL-McRBFN classification results based on the OASIS data set with other reported results in the literature.

Comparison with Related Works on the OASIS Data Set:

In [109], Yong Fan et. al proposed a method called “Integrated feature extraction and selection for neuroimage classification”, which presents an integrated feature extraction

Table 7.4: Performance comparison with existing results on the OASIS data set

Feature Type	Algorithm	# Features	Testing Accuracy(%)
VBM+ICA	PBL-McRBFN 50-50% train-test data Single/10 random trial	50	74.79/72.33
VBM	PBL-McRBFN 50-50% train-test data Single/10 random trial	19879	76.90/75.80
VBM+PCA [109]	SVM 50-50% train-test data 4 random trials	-	66.9
VBM+IPCA [109]	SVM 50-50% train-test data 4 random trials	-	69.7
VBM+ICA [110]	SVM 50-50% train-test data Single trial	200	62.8

and selection algorithm that contains two iterative steps, viz. : a constrained subspace learning based feature extraction method and SVM based feature selection. Here, VBM is used to extract features from MRI scans and an integrated feature extraction and selection algorithm (IPCA) which is used to select features in conjunction with the SVM based classification.

In [110], Wenlu Yang et. al proposed a method based on ICA called “ICA based feature extraction and automatic classification of AD related MRI data”. Here, features are extracted from VBM followed by ICA with the SVM based classification.

The above two methods used the same OASIS data set as described in Table 7.1. First method reported 4 random trial results results and second method reported single trial result. Hence, we compared the single and 10 random trial results obtained from PBL-McRBFN with the results from methods in [109, 110] and are presented in Table 7.4.

From the Table 7.4, it is observed the PBL-McRBFN based classification of AD patients and normal persons (both single trial and 10 random trial) is 6-14% than the

results from methods proposed in [109, 110]. PBL-McRBFN gives better results on complete features set extracted VBM than reduced features after ICA. The PBL-McRBFN classification efficiency on complete features set is 10% more than the PCA based SVM classification efficiency, 7% more than classification efficiency of the IPCA algorithm based SVM classification efficiency proposed in [109] and 14% more than the ICA based SVM classification efficiency proposed in [110]. PBL-McRBFN performs better on 50 features set is 8% more than the PCA based SVM classification efficiency, 5% more than classification efficiency of the IPCA algorithm based SVM classification efficiency proposed in [109] and 12% more than the ICA based SVM classification efficiency proposed in [110]. Since PBL-McRBFN uses sample selection for proper learning of decision function. The performance of PBL-McRBFN classifier is better than the results reported using well-known SVM classifier.

7.2.5 PBL-McRBFN Classification Performance on the ADNI Data Set

The performance of the PBL-McRBFN classifier has also been evaluated using the ADNI data set [103]. The complete ADNI data set consists of 232 normal persons and 200 AD patients. After verification of the unified segmentation results, 6 normal persons and 4 AD patients were excluded (due to bad segmentation) from the VBM analysis. In our study we considered 422 samples, for each sample 23797 morphometric features were obtained from the VBM analysis. Here also, the obtained morphometric features are further statistically reduced to 200 features by ICA [107]. In our classification study, for each of the 10 random trial experiments, 50% samples are randomly chosen for training and the remaining used for testing. PBL-McRBFN produces the testing accuracy using 200 reduced features is 82.38%. PBL-McRBFN testing accuracy on complete feature set is 85.28% which is **3%** more than the 200 reduced feature set. The classification performance of the PBL-McRBFN classifier using both the complete and 200 reduced feature set are given in Table 7.5.

Next, we compare the PBL-McRBFN classification results using the ADNI data set with other reported results in the literature.

Comparison with Related Works on the ADNI Data Set:

Table 7.5: Classification performance of PBL-McRBFN on the ADNI Dataset

Feature Type	# Features	# Neurons	Training			Testing		
			Accuracy(%) mean (std)	Sensitivity(%) mean (std)	Specificity(%) mean (std)	Accuracy(%) mean (std)	Sensitivity(%) mean (std)	Specificity(%) mean (std)
VBM+ICA	200	67	94.85(2.64)	93.06(4.10)	96.63(2.20)	82.38 (0.53)	77.14(3.11)	87.61(2.42)
VBM	23797	64	96.35(0.94)	95.71(1.32)	96.99(1.00)	85.27 (1.02)	82.03(2.34)	88.49(1.77)

Here, we compare the results of the PBL-McRBFN classifier results with some recent results reported in the literature that are also based on MRI ADNI data set for AD classification. In particular, four recent methods have been compared in Table 7.6. In [111], the automatic diagnostic capabilities of four structural MRI feature extraction methods (manifold based learning, hippocampal volume, cortical thickness and tensor-based morphometry) are compared using a SVM classifier. The best obtained result using the tensor-based morphometry is provided in Table 7.6. In [112], the Linear Program (LP) boosting method with a novel additional regularization have been proposed to incorporate the spatial smoothness of MRI feature space into the learning process. In [14], ten methods, which include five voxel-based methods, three cortical thickness based methods, and two hippocampus based methods are compared using a SVM classifier. The best result obtained using the voxel-wise Gray Matter (GM) features is provided in Table 7.6. In [94], 93 volumetric features extracted from the 93 ROI in GM density maps of MRI data have been used for classification.

From Table 7.6, it can be seen that among the VBM based features method, PBL-McRBFN’s performance is 3% more than that of the LP boosting method [112] and 2% lower than that of the SVM method [14]. This may be due to the fact that the SVM method in [14] uses a lower number of subjects in the study. Comparing the performance of the PBL-McRBFN classifier using the VBM features with the method of SVM using the 93 ROI features [94] and also the method using the tensor based morphometry features [111], one can see that PBL-McRBFN’s performance is similar.

Table 7.6: Performance comparison with existing results on the ADNI data set

Feature Type	Algorithm	Subjects	Testing Accuracy(%)
VBM	PBL-McRBFN 50-50% train-test data Single/ 10 random trial	226 Normal 196 AD	86.02/85.27
VBM	PBL-McRBFN 75-25% train-test data Single trial	226 Normal 196 AD	87.22
VBM	PBL-McRBFN 95-5% train-test data Single trial	226 Normal 196 AD	91.67
VBM [112]	LP boosting leave-N-out cross- validation	94 Normal 89 AD	82.00
VBM [14]	SVM 50-50% train-test data Single trial	162 Normal 137 AD	88.58
93 ROI [94]	SVM 10-fold cross- validation	52 Normal 51 AD	86.20
Tensor-based morphometry [111]	SVM 95-5% train-test data leave-N-out cross- validation, 100 times	231 Normal 198 AD	87.00

Table 7.7: Generalization performance of PBL-McRBFN classifier on unseen ADNI samples

% of ADNI Samples in Adaptation	# Neur ons	Training			Testing		
		Accu- racy(%)	Sensit- ivity(%)	Specif- icity(%)	Accu- racy(%)	Sensit- ivity(%)	Specif- icity(%)
Without	48	94.90	93.88	95.92	62.39	95.92	31.42
With 25%	81	91.66	85.45	97.87	77.27	97.87	83.43

7.2.6 Generalization Capability of the PBL-McRBFN Classifier for the Detection of AD

The aim of this study is to evaluate the generalization capability of the PBL-McRBFN classifier trained with the OASIS data set being tested on the unseen samples from the ADNI data set. The logical schematic diagram of the generalization capability study of PBL-McRBFN classifier is shown in Fig. 7.8. For this study, the Maximum Intensity Projection (MIP) of the gray matter voxel locations obtained from the OASIS data set (possible brain regions for AD) are used to extract the morphometric features from the ADNI data set in VBM feature extraction procedure apart from the normal processes of the unified segmentation and smoothing. VBM selected 19879 voxel locations from OASIS data set and the same 19879 voxel locations are used to extract the morphometric features from ADNI data set. These unseen ADNI samples with 19879 morphometric features are tested with the best classifier developed using the OASIS training samples. From the description of OASIS and ADNI data sets, we can see that these data sets are collected from different demographic people with different geographic locations. Hence, the data sets represent a wide variation in the data distribution. Hence, 25% of ADNI samples are used further for adaptation of OASIS trained PBL-McRBFN classifier and the same is tested using the remaining ADNI samples. Such a generalization capability of the classifier will prevent computationally intensive VBM feature extraction process to be done again and unify and simplify the diagnosis mechanism. The generalization performance of OASIS trained PBL-McRBFN classifier on the unseen ADNI data set is given in Table 7.7.

Figure 7.8: Schematic representation of the generalization capability study of OASIS trained PBL-McRBFN classifier on ADNI data set.

To evaluate the PBL-McRBFN generalization capability, first PBL-McRBFN classifier is trained on OASIS training data set (50% OASIS samples) and is tested on the 422 samples from the ADNI data set. This experiment is called ‘Without’ because all ADNI samples are tested using PBL-McRBFN classifier trained on OASIS samples and without ADNI samples for adaptation. The classification accuracy on unseen ADNI samples from the above experiment is 62.39%. Hence, we can say that PBL-McRBFN classifier for AD detection trained with VBM features using one data set (OASIS) can classify unseen samples from the other data set (ADNI). Further, 25% of samples from the ADNI data set were used to adapt the PBL-McRBFN classifier using meta-cognitive principles and the same is tested on the remaining 75% of samples and the testing accuracy is 77.27%. This experiment is called ‘With 25%’ because 25% of ADNI samples are used for adaptation in PBL-McRBFN classifier trained on OASIS samples. Hence, we can say that, with minor adaptation, PBL-McRBFN can classify unseen samples from other data sets accurately.

Results shows that the MIP of the gray matter voxels locations generated using VBM from one data set samples (OASIS data set) are able to discriminate samples from other

data set (ADNI data set). For growing data sets, the sequential learning PBL-McRBFN classifier is able to capture the functional relationship between the VBM features and the class labels (Disease status). Results shows that PBL-McRBFN classifier trained on one data set (OASIS data set) with minor adaptation using fewer samples from other data set (ADNI data set) achieves significant testing accuracy on a larger unseen samples.

7.3 Identification of Imaging Biomarkers for AD

In this section, we present the identification of imaging biomarkers for AD using the OASIS data set. VBM extracted the gray matter voxels which are statistically variant between the normal persons and AD patients. All the voxels generated from VBM may not be responsible to detect AD. Therefore, we propose PBL-McRBFN-RFE to find the minimal set of features among the VBM generated voxels that maximizes the detection of AD. The selected minimal set of features can be termed as the imaging biomarkers for AD. In literature, many feature selection techniques have been proposed, in general the goal of feature selection are to reduce the dimensionality. Filter and wrapper methods are two kinds of well-known feature selection techniques for high dimensional data set [113]. In the filter method, features are selected on the basis of feature separability of training samples, which is independent of the learning algorithm. The separability only takes into account the correlations between the features, so the selected features may not be optimal. Wrapper methods search for critical features based on the learning algorithm, and often result in better results than filter methods.

Recursive Feature Elimination (RFE) is a computationally less intensive wrapper based feature selection method. The basic principle of RFE is to include initially all features of a large region, and to gradually exclude features, that do not contribute in discriminating patterns from different classes. Whether a feature in the current feature set contributes enough to be kept is determined by the discriminative power of a feature resulting from training a classifier with the current set of features. In order to increase the likelihood that the “best” feature are selected, feature elimination progresses gradually and includes cross-validation steps. In each feature elimination step, a single feature is discarded until a core set of features remains with the highest discriminative power.

In this study, informative features are selected by the method of RFE utilizing a PBL-McRBFN classifier. RFE utilizing PBL-McRBFN is referred as PBL-McRBFN-RFE. PBL-McRBFN-RFE conducts feature selection in a sequential elimination manner, which starts with all the features and discards one feature at a time from the top. The discard feature added to the feature set again if the training/testing efficiency decreases. The PBL-McRBFN-RFE feature selection algorithm runs till the number of selected features in current iteration (s) is more than predefined minimum limit for number of features to be selected (r) and is not equal to the number of selected features in previous iteration (p). To summarize, the PBL-McRBFN-RFE feature selection algorithm in a pseudo code form is given in Pseudo code 3.

Pseudocode 3 Pseudo code for the PBL-McRBFN-RFE feature selection algorithm.

Input : N data samples $\{(x^t, y^t)\}$,
 r predefined minimum limit for number of features to be selected.
Output : S selected features set.

START

Initialization : Initialize the set of selected features (S) to the full feature set. Assign the number of selected features in current iteration (s) to the number of features in set S and the number of selected features in previous iteration p to zero.

WHILE $s > r$ AND $s \neq p$ DO

 Assign p to the number of features from set S .

 FOR each feature value in set s DO

 Remove first feature value in set S .

 Train PBL-McRBFN classifier with remaining features in set S and calculate training and testing efficiencies.

 IF *Training OR Testing efficiency decreases* THEN

 Insert the removed first feature value at the rear end of the set S .

 ENDIF

 ENFFOR

 Assign s to the number of features in set S .

ENDWHILE

RETURN selected features set S .

END

First we present the identification of imaging biomarkers for AD from the complete

Table 7.8: VBM detected and PBL-McRBFN-RFE selected regions from complete OASIS data set

Feature Type	# Features	Identified Regions
VBM	19879	Parahippocampal gyrus, Amygdala, Hippocampus, Superior temporal gyrus, Insula, Sub-gyrus, Precentral gyrus, Extra-nuclear
VBM + PBL-McRBFN-RFE	906	Parahippocampal gyrus, Hippocampus, Superior temporal gyrus, Insula, Precentral gyrus, Extra-nuclear

Table 7.9: PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on complete OASIS data set

# Features	Training			Testing		
	Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)	Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)
19879 (VBM)	94.90	93.87	95.91	76.90	64	89.79
906 (VBM+ PBL-McRBFN-RFE)	91.84	91.83	91.83	84.96	74	95.91

OASIS data set using PBL-McRBFN-RFE. In medical literature it is reported that there is difference between the regions effected by AD among male and female persons [15, 16]. Hence, we also conducted an age-wise and gender-wise analysis to identify imaging biomarkers for AD using PBL-McRBFN-RFE.

7.3.1 Imaging Biomarkers for AD in Complete OASIS Data Set

In imaging biomarkers identification analysis from the complete OASIS data set, the minimal set of voxels which are more relevant to AD are found using PBL-McRBFN-RFE. Brain regions corresponding to the VBM detected voxels (19879 voxels with 198 OASIS samples) are reported in Table 7.8. PBL-McRBFN-RFE selected 906 voxels among 19879 voxels, corresponding brain regions to 906 voxels are reported in Table 7.8.

MNI templates of the complete 19879 and 906 voxel regions are shown in Fig. 7.9. The testing performance of PBL-McRBFN on this selected 906 features set is 84.95%

Figure 7.9: Comparison of gray matter volume change - Normal persons vs. AD patients from complete OASIS data set

Table 7.10: Generalization performance of PBL-McRBFN classifier on unseen ADNI samples with selected 906 features

% of ADNI Samples in Adaptation	# Neurons	Training			Testing		
		Accuracy(%)	Sensitivity(%)	Specificity(%)	Accuracy(%)	Sensitivity(%)	Specificity(%)
Without	38	91.83	91.83	91.83	52.59	92.34	12.83
With 25%	50	80.75	63.63	97.87	75.86	62.22	89.50

as shown in Table 7.9. To check the discriminating capability of the selected 906 voxels, we have conducted a generalization capability study as in section 7.2.6. The generalization performance of PBL-McRBFN classifier on unseen ADNI samples with selected 906 features is given in Table 7.10.

We found these selected voxels by PBL-McRBFN-RFE are located at the brain regions such as: the superior temporal gyrus, the insula, the precentral gyrus and the extranuclear. These regions are consistent with those reported in the medical literature as the biomarkers for AD [114, 115, 116]. Hence, the gray matter atrophy in the brain regions detected by PBL-McRBFN-RFE among VBM features are may be more relevant to the

detection of AD.

7.3.2 Imaging Biomarkers for AD Based on Age in OASIS Data Set

The brain regions affected in AD patients may differ based on their ages. To verify this, an analysis is conducted in the OASIS data set based on the ages of the patients. Among the 198 persons in the OASIS data set, 40 persons are in the age group of 60-69, 83 persons are in the age group of 70-79 and 75 persons are in the ages of 80 and above. We have separately conducted this analysis based on different age groups.

Study of the 60-69 age group : VBM extracted 292 features with 40 persons on this age group, with a 50-50% training and testing split PBL-McRBFN obtained the testing performance on 292 features set of 100%. After performing PBL-McRBFN-RFE on 292 features set, 25 features are selected and the testing performance of PBL-McRBFN on this selected features set is 100%.

Study of the 70-79 age group : VBM extracted 3298 features with 83 persons on this age group, with a 50-50% training and testing split PBL-McRBFN obtained the best testing performance on 3298 features set of 91.67%. After performing PBL-McRBFN-RFE on 3298 features set, 90 features are selected and the testing performance of PBL-McRBFN on this selected features set is 95.83%.

Study of the Above 80 age group : VBM extracted 1047 features with 75 persons on this age group, with a 50-50% training and testing split PBL-McRBFN obtained the best testing performance on 1047 features set of 89.47%. After performing PBL-McRBFN-RFE on 1047 features set, 154 features are selected and the testing performance of PBL-McRBFN on this selected features set is 94.59%.

The VBM detected and PBL-McRBFN-RFE selected brain regions of voxels for different age groups are listed in Table 7.11 and the corresponding PBL-McRBFN performance

Table 7.11: VBM detected and PBL-McRBFN-RFE selected regions from age-wise OASIS data sets

Age Group	Feature Type	# Features	Identified Regions
60-69	VBM	292	Superior temporal gyrus, Postcentral gyrus
	VBM + PBL-McRBFN-RFE	25	Superior temporal gyrus
70-79	VBM	3298	Parahippocampal gyrus, Amygdala, Extra-nuclear, Uncus, Third Ventricle
	VBM+ PBL-McRBFN-RFE	90	Parahippocampal gyrus , Extra-nuclear
80-Above	VBM	1047	Hippocampus, Parahippocampal gyrus, Lateral Ventrical, Extra-nuclear
	VBM + PBL-McRBFN-RFE	154	Hippocampus, Parahippocampal gyrus, Lateral Ventrical

Table 7.12: PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on age-wise OASIS data sets

Age Group	# Features	Training			Testing		
		Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)	Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)
60-69	292 (VBM)	93.75	87.5	100	100	100	100
	25 (VBM+ PBL-McRBFN-RFE)	93.75	87.5	100	100	100	100
70-79	3298(VBM)	100	100	100	91.67	83.33	100
	90 (VBM+ PBL-McRBFN-RFE)	95.14	95.83	94.44	95.83	91.67	100
80-Above	1047 (VBM)	92.10	100	84.21	89.47	100	78.94
	154 (VBM+ PBL-McRBFN-RFE)	91.81	88.88	94.73	94.59	94.44	94.73

Figure 7.10: Comparison of gray matter volume change - Normal persons vs. AD patients from 60-69 (a&b), 70-79 (c&d), 80-Above (e&f) age groups in OASIS data set

results are shown in Table 7.12. MNI templates of complete VBM and PBL-McRBFN-RFE selected voxel regions are shown in Fig. 7.10. In each age group analysis, the selected voxels from PBL-McRBFN-RFE gives better classification accuracy than the VBM detected regions.

From the age-wise analysis, we can see that PBL-McRBFN classifier detects AD accurately in the 60-69 age group and 25 voxels selected from the PBL-McRBFN-RFE are still able to classify AD accurately. The PBL-McRBFN-RFE detected brain region responsible for AD in the 60-69 age group is the superior temporal gyrus which contains the primary auditory cortex and is responsible for processing sounds. Hence, we can conclude that, AD patients in the 60-69 age group may have auditory related problems as indicators of AD. In the 70-79 age group, the detected brain regions responsible for AD are the parahippocampal gyrus and the extra-nuclear which are responsible for memory encoding and retrieval. Hence, we can conclude that, AD patients in the 70-79 age group may have memory related problems. In the 80-89 age group, the detected brain regions responsible for AD are the hippocampus, the parahippocampal gyrus and the lateral ventricle which are responsible for short-term memory to long-term memory, spatial navigation, memory encoding and retrieval. Hence, we can conclude that, AD patients in the 80-89 age group may have major difficulties in memory.

7.3.3 Imaging Biomarkers for AD Based on gender in OASIS Data Set

In the medical literature [15, 16], it is reported that gender may be an important modifying factor in AD's development and expression. To verify this, an analysis is conducted gender-wise in the OASIS data sets. Among the 198 persons in the OASIS data set, 67 persons are male and 131 persons are female. We have conducted the analysis using the PBL-McRBFN-RFE on both the male and female persons separately.

Male persons study : Here, AD imaging biomarkers identification analysis is conducted considering 67 male persons alone. VBM extracted 1239 voxels with 67 male persons. The corresponding brain regions are shown in Table 7.13. PBL-McRBFN obtained best testing performance on the complete 1239 features is 79.8% with 50-50% training and testing data set split as shown in Table 7.14. After performing PBL-McRBFN-RFE on

Table 7.13: VBM detected and PBL-McRBFN-RFE selected regions from male-OASIS data set

Feature Type	# Features	Identified Regions
VBM	1239	Parahippocampal gyrus, Traverse temporal gyrus, Insula, Superior temporal gyrus, Sub-gyrus, Extra-nuclear, Inferior parietal lobule
VBM + PBL-McRBFN-RFE	31	Insula

Table 7.14: PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on male-OASIS data set

# Features	Training			Testing		
	Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)	Accu-racy(%)	Sensit-ivity(%)	Specif-icity(%)
1239 (VBM)	96.15	100	92.30	79.81	75	84.61
31 (VBM+ PBL-McRBFN-RFE)	100	100	100	89.81	95	84.61

OASIS male complete data set, 31 voxels were selected and the testing performance of PBL-McRBFN on this reduced features set is 89.8% as shown in Table 7.14.

The corresponding brain regions of the 31 voxels are listed in Table 7.13. MNI templates of the complete 1239 and 31 voxel regions are shown in Fig. 7.11. All the 31 voxels are from the insular cortex region which is responsible for emotion and consciousness. The insula region is also reported in AD research studies [117, 118] and it is associated with hypometabolism. Hence, we can conclude that the male AD patients may have emotion related problems.

Female persons study : Here, AD imaging biomarkers identification analysis is conducted considering 131 female persons alone. VBM extracted 15203 voxels with 131 female persons. The corresponding brain regions of 15203 voxels are shown in Table 7.15. PBL-McRBFN obtained best testing performance on complete 15203 features is 79.93 % with 50-50% training and testing data set split as shown in Table 7.16.

Figure 7.11: Comparison of gray matter volume change - Normal persons vs. AD patients from Male-OASIS data set

Table 7.15: VBM detected and PBL-McRBFN-RFE selected regions from female-OASIS data set

Feature Type	# Features	Identified Regions
VBM	15203	Parahippocampal gyrus, Amygdala, Superior temporal gyrus, Inferior temporal gyrus, Middle temporal gyrus, Insula, Sub-gyrus, Extra-nuclear,
VBM + PBL-McRBFN-RFE	294	Parahippocampal gyrus , Extra-nuclear

Table 7.16: PBL-McRBFN classifier performance comparison with VBM detected and PBL-McRBFN-RFE selected features on female-OASIS data set

# Features	Training			Testing		
	Accu- racy(%)	Sensit- ivity(%)	Specif- icity(%)	Accu- racy(%)	Sensit- ivity(%)	Specif- icity(%)
15203 (VBM)	90.99	93.10	88.88	79.93	79.31	80.55
294 (VBM+ PBL-McRBFN-RFE)	90.33	86.20	94.44	85.44	93.10	77.77

Figure 7.12: Comparison of gray matter volume change - Normal persons vs. AD patients from Female-OASIS data set

After performing PBL-McRBFN-RFE on the OASIS female data set, 294 voxels were selected and the testing performance of PBL-McRBFN on this selected features set is 85.44 % as shown in Table 7.16. The corresponding brain regions of 294 voxels are listed in Table 7.15. MNI templates of the complete 15203 and 294 voxel regions are shown in Fig. 7.12. We found that these selected 294 voxels are located at the brain regions such as the parahippocampal gyrus and the extra-nuclear regions, which are responsible for memory encoding and retrieval. Hence, we can conclude that female AD patients may have memory related problems.

The above detailed study results indicate the superior performance of the proposed PBL-McRBFN classifier. Also, the PBL-McRBFN-RFE approach identifies the imaging biomarkers for the onset of AD.

7.4 Discussion

The AD detection performance of PBL-McRBFN is better than the existing results in the literature for OASIS and ADNI data sets. The generalization capability of the PBL-McRBFN classifier has been demonstrated by testing the unseen samples from ADNI data set using the trained PBL-McRBFN classifier with samples from OASIS data set. Using the proposed PBL-McRBFN-RFE, we have identified the imaging biomarkers (critical regions in the brain) responsible for AD using the OASIS data set. Based on our study, the gray matter atrophy identified in AD patients in the superior temporal gyrus, the insula, the precentral gyrus and the extra nuclear regions, which have also been highlighted in the medical literature [114, 115, 116]. Further, we have carried out a detailed analysis based on the age and gender. Based on our analysis, the indicators that emerge for the onset of AD are:

- *In the age group 60-69*: Degradation in sound processing capability (primary auditory cortex)
- *In the age group 70-79*: Memory related problems (parahippocampal gyrus and extra nuclear)
- *In the age group 80-89*: Problem in short-term/long-term memory, encoding/retrieval and spatial navigation (parahippocampal gyrus and lateral ventricular)

- *In male persons:* Emotion related problems usually associated as a hypometabolism (insula)
- *In female persons:* Mainly memory related problems (parahippocampal gyrus and extra nuclear)

Thus, the proposed approach provides imaging biomarkers identification mechanism for AD and this approach can be used for other similar problems.

7.5 Summary

In this chapter, AD diagnosis problem is solved by employing PBL-McRBFN classifier. Morphometric features were extracted from MRI scans using VBM. For simulation studies, we have used well-known OASIS and ADNI data sets. The performance of the PBL-McRBFN classifier has been evaluated on complete morphometric features set obtained from the VBM analysis and also on reduced features sets from ICA. Since, the data sets contain very mild AD patients and fewer samples, AD detection using complete VBM features provide better performance than ICA reduced features. The performance of the proposed method is compared against state-of-the-art methods reported in literature. Next, the performance evaluation on ADNI data set with PBL-McRBFN classifier trained on OASIS data set shows that the proposed PBL-McRBFN can also achieve significant results on the unseen data set. Finally, imaging biomarkers responsible for AD are detected with PBL-McRBFN-RFE approach using OASIS data set. Imaging biomarkers responsible for AD were found for different age groups, and for both genders.

Parkinson's disease is the second widely reported neurodegenerative disease, next only to AD. Hence, in the next chapter, diagnosis of Parkinson disease using PBL-McRBFN classifier is presented.

Chapter 8

Parkinson's Disease Diagnosis using PBL-McRBFN Classifier

In this chapter we use PBL-McRBFN classifier for the diagnosis of Parkinson's disease based on microarray gene expression, MRI scans, vocal and gait features.

Parkinson's Disease (PD) is characterized by progressive degeneration of dopaminergic neurons in the pars compacta of the substantia nigra [119]. Most important symptoms of PD include muscle rigidity, tremors, and changes in speech and gait [119, 120]. PD is more common in elderly people over the age of 50, which has influenced millions of people worldwide. According to the global declaration for PD, 6.3 million people are affected by this disease worldwide, and affect all races and cultures in 2013. Albeit significant research advances have been made, including the recent identification of possible genetic and environmental risk factors for PD, further research is required to illustrate the underlying causes of PD and to discover ameliorated treatments. At present there is no cure for PD and the diagnosis of PD is based on medical history and neurological examination conducted by interviewing and observing the patient in person using the disease rating scales. Unified Parkinson's Disease Rating Scale (UPDRS), Hoehn and Yahr scale, Schwab and England Activities of Daily Living (ADL) scale, PDQ-39, PD Non-motor symptoms (NMS) questionnaire, NMS survey are most commonly used PD rating scales. The reliable diagnosis of PD using these scales is difficult, especially in its early stages [121]. As the symptoms of PD are comorbid with other neurological diseases, only 75% of clinical diagnoses of PD are confirmed to be idiopathic PD at autopsy. Thus, automatic approaches based on machine learning techniques are needed to increase the diagnosis accuracy and to assist physicians to make better decisions.

The literature review on machine learning approaches for diagnosis of PD is presented in the next section.

8.1 Literature Review on Parkinson's Disease

In literature, machine learning approaches for PD classification were undertaken by detecting dysphonia and tremor symptoms. Nearly one third of PD patients exhibit a group of vocal impairment symptoms, which is known as dysphonia [122]. Machine learning approaches for PD classification by detecting dysphonia using acoustic measurements are experimented with a data set created in [122], consisting of sustained vowel phonation's from 31 people of whom 23 with PD. In [123], kernel-support vector machine is used for PD classification. Here, exhaustive search process is implemented to select the best kernel width and penalty value. In [124], Support Vector Machine (SVM) classifier is used with features selection using maximum-relevance-minimum-redundancy (mRMR) criterion. For mRMR, all the available samples are used in mutual information computations. Also, the above two approaches uses all the available data samples to optimize the SVM parameters which is inevitable when working with such a small data set. In [125], four independent classification approaches (neural networks, data mining neural, logistic regression and decision trees) are compared for diagnosis of PD. Among the four approaches, neural networks classifier (Multi-layer feed-forward neural network with Levenberg-Marquardt algorithm) yields the best performance. The drawback in the above neural network approach is random initialization of weights and heuristic determination of number of hidden neurons, which affects the classification performance significantly. Also, it requires retraining when the training samples changes with time. In [126], parallel neural networks approach is used for prediction of PD. The training time and complexity of the parallel network approach do increase as the number of parallel networks increases. In [127], adaptive neuro-fuzzy classifier with linguistic hedges is used for feature selection and classification. Linguistic hedges feature selection will require the optimization search from huge set of theoretically possible encoding combinations for each feature and hence computationally intensive. In [128], a fuzzy c-means clustering-based feature weighting with k -NN classification approach is presented. The choice of

number of clusters or selection of k-neighbors significantly affects the performance. In addition, increasing the number of training samples may further complicate the choice of number of clusters.

PD patients exhibit large gait variability compared to the normal persons [129]. Gait analysis is routinely used in clinical settings to assess these gait disorders. Gait analysis is a locomotion study and typically consists measurements of spatial-temporal parameters of the gait cycle, motion of joints and segments, force/moments, and electromyography patterns of muscle activation. Machine learning approaches using the different gait features have been reported in the literature for PD classification [130, 131, 121]. In [130], image data is obtained from plantar pressure measurements of right foot during heel to motion from 17 control and 21 PD patients, and SVM is applied to distinguish gait pattern. Here, other important basic, kinetic and kinematic features are not used in gait analysis. In [131], a data set from 20 control and 12 PD patients consists of basic spatiotemporal, kinematic and kinetic gait features is used, and the ability of ANN and SVM classifiers is discussed. The above two studies on gait features use their own proprietary data, with less number of subjects. In [121], data collected from sensors located under the feet of 73 controls and 93 PD patients is used. In their approach, wavelet transform has been employed to extract the relevant features and neural networks with weighted fuzzy membership have been used to approximate the functional relationship between the extracted features and class label.

Recent studies on gene expression analysis found that there is a profound change in gene expression for individuals affected by PD [132]. These studies discovered that, diagnosis of early stage PD using vocal and gait features is impossible because tremor and slow movements develop in PD patients only after approximately 70% of vulnerable dopaminergic neurons in the substantia nigra have already died [132]. However, machine learning approaches for PD classification based on gene analysis have been limited. Therefore, there is a need for devising a new machine learning approach for PD classification based on gene analysis.

Over the past two decades, neuroimaging techniques such as Positron Emission Tomography (PET) [133], Single-Photon Emission Computed Tomography (SPECT) [134], Magnetic Resonance Imaging (MRI) [135] and Transcranial Brain Sonography (TCS)

[136] have increasingly been employed to predict PD, to strengthen the neuropathological mechanisms and compensatory responses underlying symptoms and treatment associated complications, and to monitor disease progression [137]. MRI is far more widely available than PET and SPECT and is most commonly used in clinical practice to differentiate PD from normal persons [135]. However, machine learning approaches for PD classification based on MRI scans have been limited. Therefore, there is a need for devising a new machine learning approach for PD classification based on MRI scans.

8.2 Materials

We have considered four possible ways to diagnosis PD using PBL-McRBFN classifier

- (a) Prediction of PD from microarray gene expression features.
- (b) Prediction of PD from MRI scans.
- (c) Detection of PD from vocal features.
- (d) Detection of PD from gait features.

8.2.1 Microarray Gene Expression Data Set

In this study, the normalized microarray gene expression data is obtained from ParkDB database [138] under the accession number E-GEOD-6613. ParkDB database is the first queryable database dedicated to gene expression in PD. ParkDB database contains a complete set of re-analyzed, curated and annotated microarray data sets. The considered data set is obtained by transcriptional profiling from the RNA extracted from whole blood of 50 PD patients at early stage and 22 controls [132]. The extracted 22283 oligonucleotide probe sets (short section of gene) on microarrays is used to analyze the difference in expression of the genes between PD patients and control. Robust Multi-array Analysis (RMA) method in Limma package [139] is used to normalize and summarize the probe intensity measurements. Thus, the complete gene expression data set contains 72 subjects with 22283 genes expression information.

8.2.2 MRI Data Set

MRI data is obtained from the Parkinson’s Progression Markers Initiative (PPMI) data set (www.ppmi-info.org/data). Standard 1.5T baseline 3D volumetric T1-weighted brain MR images were selected from the PPMI data set. We have considered 239 persons (112 normal persons and 127 PD patients) available in the data set as of April 2012. Among 239 persons, 31 normal persons and 34 PD patient’s MR images were excluded due to failure of the segmentation method.

Whole brain T1-weighted, 3D MPRAGE MR images were acquired using at least 1.5 Tesla scanner with repetition time between 5-11 ms and echo delay time between 2-6 ms. The acquired volumes have slice thickness ranging from 1-1.5 mm and voxel dimensions of 1.0 mm $\times$ 1.0 mm $\times$ 1.20 mm. The detailed information of the MRI protocols and preprocessing steps are presented in [140]. The demographics for the data used in our study is shown in Table 8.1.

Table 8.1: Demographic information of PPMI MRI data used in our study

Group	Normal Persons	PD Patients
No. of persons	112	127
Sex(M/F)	64/48	86/41
Age(mean $\pm$ std)	58.35 $\pm$ 11.31	61.83 $\pm$ 9.67

8.2.3 Vocal Data Set

Vocal data set is obtained from voice recording originally done at University of Oxford by Max Little [122] has been used to PD classification by detecting dysphonia. The recording consists of 195 entries collected from 31 people of whom 23 are suffering from PD. From the 195 samples, 147 are of PD patients and 48 controls. Averages of six phonations were recorded from each subject, ranging from 1 to 36 sec in length. The 22 attributes used in this prediction task can be broadly classified into jitter (variation in fundamental frequency), shimmer (variation in amplitude), harmonic/noise ratio (amplitude of noise relative to tonal components in the speech), fundamental frequency, descriptive statistics and correlation factors (non-linear measures) [122].

8.2.4 Gait Data Set

Gait data set provided by PhysioBank [141] has been used to discriminate Parkinson gait and normal gait. This data set consists 166 samples, contains measures of gait from 93 PD patients and 73 controls. The data set collected by 8 sensors underneath each foot, includes the vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. The 10 attributes used in this prediction task are left swing interval (sec), right swing interval (sec), left swing interval (% of stride), right swing interval (% of stride), double support interval (sec), double support interval (% of stride), left stride variability, right stride variability, cadence and speed. The 4 swing interval measures and 2 double support interval measures are ranked the top attributes with maximum relevance and less redundancy for gait analysis [142].

8.3 Early Diagnosis of Parkinson’s Disease Based on Gene Expression Features

In this section, we present the performance evaluation of PBL-McRBFN on PD classification using microarray gene expression data set. The PBL-McRBFN classifier performance is evaluated using gene expression features in two scenarios as shown in Fig. 8.1. In first scenario, its performance is evaluated on ICA reduced features from complete 22283 genes as shown in Fig. 8.1(a). Next, its performance is evaluated on ICA reduced features from selected 1594/412 genes with significance levels $p < 0.05/0.01$ as shown in Fig. 8.1(b). We have conducted 10 random trials of experiments for every ICA reduced feature set. In each trial, randomly 75% of total samples are selected for training and 25% for testing. The classification performance of PBL-McRBFN is compared with the standard SVM classifier.

8.3.1 p -value Based Gene Selection

The complete gene expression data set consists of large number of redundant genes, which affects the classifier performance on PD prediction. Hence, we select the most informative genes based on p -value selection from ParkDB database. When less constraints are

Figure 8.1: PBL-McRBFN classifier on ICA reduced features from: (a) Complete genes, (b) Selected genes

incorporated, with gene fold change greater than 1.5 (on a binary logarithmic scale) and p -value less than 0.05, 1594 genes are extracted. When more stringent constraints are incorporated, with the same fold change (1.5) and increased level significance level (p -value less than 0.01), 412 genes are extracted. The above two sets of selected genes expression features of the same 72 subjects are considered as selected gene expression data sets.

However, as the feature space in complete gene expression data set and selected gene expression data sets is high compared to the number of samples, it will be difficult to predict PD accurately. Hence, the obtained complete and selected genes expression features are further reduced statistically by ICA [107].

8.3.2 ICA Based Feature Reduction

The basic goal of independent component analysis is to find a transformation in which the components of transformed data are statistically as independent from each other as possible. ICA can be applied for blind source separation, exploratory data analysis and feature extraction. Feature extraction using the ICA is a promising application. The extracted feature vectors from the ICA analysis are as independent from each other as possible, i.e. the extracted features do not contain mutual information about other features.

In our study, we employ FastICA (fixed-point) algorithm [107], the FastICA package for MATLAB [108] is used to reduce complete and selected genes expression features to combinations of 10, 25, 50 features.

The mean classification performance measures (average/overall) testing efficiencies and F -score obtained from 10 random trials of experiments with PBL-McRBFN and SVM using ICA reduced feature data sets from complete and selected gene expression data sets are presented in Tables 8.2-8.4. From the Tables 8.2-8.4, it is evident that PBL-McRBFN generalization performance is better than SVM classifier on both complete and selected gene expression data sets.

8.3.3 Performance of PBL-McRBFN on ICA Reduced Features from Complete Genes

Table 8.2: Performance comparison on complete gene expression data set from an average of 10 trials

Features Type	# Features	Algorithm	Testing Accuracy		
			Overall ($\eta_o\%$) mean (std)	Average ($\eta_a\%$) mean (std)	F -score mean (std)
ICA Reduced	10	SVM	72.78 (7.15)	66.87 (7.84)	0.8053 (0.0656)
		PBL-McRBFN	85.55 (4.68)	84.94 (3.66)	0.8896 (0.0424)
ICA Reduced	25	SVM	72.78 (7.61)	67.99 (9.39)	0.8037 (0.0620)
		PBL-McRBFN	83.89 (4.86)	83.75 (5.55)	0.8763 (0.0438)
ICA Reduced	50	SVM	71.67 (8.05)	60.94 (6.50)	0.8166 (0.0498)
		PBL-McRBFN	84.44 (5.73)	83.58 (3.41)	0.8802 (0.0536)
Complete	22283	SVM	86.67 (7.03)	72.03 (9.99)	0.9201 (0.0439)
		PBL-McRBFN	88.89 (5.24)	88.53 (6.21)	0.9226 (0.0393)

On complete gene expression data set, PBL-McRBFN classifier achieves better generalization performance with 10 ICA reduced features than 25, 50 features as shown in the Table 8.2. From the Table 8.2, we can see that on 10 ICA reduced features η_a of PBL-McRBFN is 8% more than SVM with better F -score value. Similarly on 25, 50 features, η_a of PBL-McRBFN is more than SVM classifier. The η_a of PBL-McRBFN is reduced by 1% on 25 and 50 features compared to 10 features, this is due to the redundancy of ICA features. On 25, 50 features, SVM classifier performance is reduced significantly than PBL-McRBFN classifier. We can also see that the PBL-McRBFN performance on original 22283 features is more than ICA reduced features and 2% more when compared to SVM performance.

8.3.4 Performance of PBL-McRBFN on ICA Reduced Features from Statistically Selected Genes

On selected gene expression data set with p -value < 0.05 , PBL-McRBFN classifier achieves better generalization performance with 10 ICA reduced features than 25, 50 features as shown in Table 8.3. From the Table 8.3, we can see that on 10 ICA reduced features, η_a of PBL-McRBFN is 7% more than SVM with better F -score value. The better performance of both classifiers on the selected gene expression data set with p -value < 0.05 can be observed when compared to the performance on complete gene expression data set. The η_a of PBL-McRBFN on the selected gene expression data set with p -value < 0.05 is 13% more than complete gene expression data set, this is due to presence of more redundant genes information relative to PD in the complete gene expression data set.

Table 8.3: Performance comparison on selected gene expression data set with p -value < 0.05 from an average of 10 trials

Features Type	# Features	Algorithm	Testing Accuracy		
			Overall (η_o) mean (std)	Average (η_a) mean (std)	F -score mean (std)
ICA Reduced	10	SVM	86.67 (5.37)	84.17 (7.53)	0.9047 (0.0388)
		PBL-McRBFN	96.67 (4.68)	97.17 (4.32)	0.9769 (0.0318)
ICA Reduced	25	SVM	83.39 (5.11)	80.42 (7.09)	0.8862 (0.0494)
		PBL-McRBFN	88.33 (7.14)	89.25 (7.26)	0.9097 (0.0596)
ICA Reduced	50	SVM	78.33 (8.47)	71.44 (10.26)	0.8499 (0.0619)
		PBL-McRBFN	84.99 (6.95)	85.51 (5.00)	0.8820 (0.0665)
Complete	1594	SVM	95.55 (3.51)	96.27 (3.17)	0.9687 (0.0258)
		PBL-McRBFN	100 (0)	100 (0)	1 (0)

On selected gene expression data set with p -value < 0.01 , PBL-McRBFN classifier achieves better generalization performance with 10 ICA reduced features than 25, 50 features as shown in Table 8.4. From Table 8.4, we can see that on 10 ICA reduced features data set η_a of PBL-McRBFN is 30% more than SVM with better F -score value. The minor reduction in performance of both classifiers on the selected gene expression data set with p -value < 0.01 can be observed when compared to the performance on selected gene expression data set with p -value < 0.05 and better when compare to the performance on the complete gene expression data set. The η_a of PBL-McRBFN on the

selected gene expression data set with p -value < 0.01 is 1% less than the selected gene expression data set with p -value < 0.05 , this is due to absence of few informative genes relative to PD in the selected gene expression data set with p -value < 0.01 .

Table 8.4: Performance comparison on selected gene expression data set with p -value < 0.01 from an average of 10 trials

Features Type	# Features	Algorithm	Testing Accuracy		
			Overall (η_o) mean (std)	Average (η_a) mean (std)	F -score mean (std)
ICA Reduced	10	SVM	72.78 (7.15)	66.87 (7.84)	0.8974 (0.0660)
		PBL-McRBFN	95.55 (3.51)	96.02 (3.32)	0.9676 (0.0260)
ICA Reduced	25	SVM	90.00 (6.83)	86.43 (10.27)	0.9312 (0.0454)
		PBL-McRBFN	94.44 (5.23)	94.64 (6.24)	0.9582 (0.0391)
ICA Reduced	50	SVM	73.33 (5.74)	68.03 (10.17)	0.8120 (0.0381)
		PBL-McRBFN	83.89 (4.09)	85.28 (6.33)	0.8762 (0.0375)
Complete	412	SVM	96.67 (3.88)	97.19 (3.25)	0.9753 (0.0292)
		PBL-McRBFN	100 (0)	100 (0)	1 (0)

From the Tables (8.2 to 8.4), we can see that the PBL-McRBFN classifier achieves best performance on the 10 ICA features data set obtained from selected gene expression data set with p -value < 0.05 . The η_a of PBL-McRBFN classifier on the 10 ICA features data set obtained from selected gene expression data set with p -value < 0.05 is 97.17% which is 1% more than the η_a of PBL-McRBFN classifier on the selected gene expression data set with p -value < 0.05 without ICA feature reduction (96.87%). The PBL-McRBFN classifier performance on all the three gene expression data sets without ICA feature reduction is same. Thus, we can observe that the changes in the performance of PBL-McRBFN classifier on the three gene expression data sets with ICA reduced features is due to bad performance of ICA. We can also see that the PBL-McRBFN performance on original 1594 and 412 features is more than ICA reduced features and better when compared to SVM performance. PBL-McRBFN accurately classifies PD with original 1594 and 412 statistically selected gene expression features.

8.4 Early Diagnosis of Parkinson’s Disease Based on MRI features

In this section, we present the performance evaluation of PBL-McRBFN on PD classification using MRI data set. First time in the literature, we conducted PD prediction study using MRI scans. Hence, no study available in the literature on PD prediction based on MRI scans. We also propose PBL-McRBFN-RFE approach to identify imaging biomarkers (critical brain regions) responsible for PD. In PBL-McRBFN-RFE approach, PBL-McRBFN is used for PD classification and RFE method is used for feature selection. RFE uses the training algorithm (PBL-McRBFN) recursively to eliminate irrelevant features one at a time. RFE seeks to improve generalization performance by eliminating the least important feature whose elimination will have the no effect on classification performance.

8.4.1 VBM Based Feature Extraction

The VBM analysis is used in this study to identify the regional differences in gray matter between PD patients and normal persons, and to extract morphometric features from MRI scans. VBM analysis used in this study is as described in section 7.2.2. The flow diagram of the feature extraction process is shown in Fig. 8.2. For better understanding, we show the maximum intensity projections of the significant voxels in sagittal, coronal and axial views in Fig. 8.3.

To locate the above regions with respect to the spatial locations in the brain, these regions were overlaid on the sliced sections of the commonly used MNI brain template and the results of the same are shown in Fig. 8.4. From Fig. 8.3 and Fig. 8.4, it is inferred that there is a significant gray matter volume differences in the superior temporal gyrus, middle temporal gyrus, parahippocampal gyrus, sub-gyral and insula regions of the brain, which have also been highlighted in the medical literature [144].

The voxels locations of the VBM detected significant regions are used as mask in order to extract the features from all the segmented gray matter images. The feature extraction process computes a vector with all the gray matter segmentation values for the voxels locations included in each VBM identified region.

Figure 8.2: Schematic diagram of the stages in feature extraction based on the VBM analysis.

A total of 2981 features (gray matter tissue probability values) are extracted from the VBM identified region and are then used as an input to the PBL-McRBFN classifier. However, as the feature space in VBM detected features set is high compared to the number of samples, it will be difficult to predict PD accurately. Hence, the extracted VBM features are further reduced statistically by ICA [107]. In our study, we employ FastICA (fixed-point) algorithm [107]. The FastICA package for MATLAB [108] is used to reduce complete 2981 VBM detected features to combinations of 10 and 50 features.

The PBL-McRBFN classifier performance is evaluated using VBM detected features and ICA reduced features. We have conducted 10 random trials of experiments for VBM feature set and for every ICA reduced feature set. In each trial, 75% and 25% samples are randomly chosen for training and testing respectively. The classification performance of PBL-McRBFN is compared with the SVM classifier [37].

8.4.2 Performance of PBL-McRBFN on VBM Features

The complete MRI data set consists of 239 samples with 2981 morphometric features. The mean, standard deviation of training/testing accuracy, sensitivity and specificity obtained during the 10 random trials for the PBL-McRBFN and the SVM classifiers on

Figure 8.3: Maximum intensity projections from PPMI MRI data set - Normal persons vs. PD patients (a) sagittal view (b) coronal view (c) axial view

2981 VBM features set are presented in Table 8.5. In each trial, randomly 75% of total samples are selected for training and 25% for testing. From Table 8.5, we can see that testing accuracy of PBL-McRBFN is 3% more than SVM with better sensitivity and specificity values. Thus, PBL-McRBFN classifier performs an efficient classification of the VBM morphometric features from MRI scans for prediction of PD.

Table 8.5: Performance comparison on 2981 VBM features data set from an average of 10 trials

Algorithm	Training			Testing		
	Accu- racy(%) mean (std)	Sensit- ivity(%) mean (std)	Specif- icity(%) mean (std)	Accu- racy(%) mean (std)	Sensit- ivity(%) mean (std)	Specif- icity(%) mean (std)
SVM	96.40 (2.18)	98.52 (1.55)	94.00 (3.61)	79.06 (3.63)	83.04 (6.30)	74.5 (9.55)
PBL-McRBFN	93.04 (2.45)	95.88 (2.92)	89.83 (6.20)	82.32 (2.50)	83.47 (6.41)	81.00 (4.59)

Figure 8.4: Gray matter volume change from PPMI MRI data set - Normal persons vs. PD patients (a) sagittal view (b) coronal view (c) axial view

8.4.3 Performance of PBL-McRBFN on Reduced features

The feature reduction using ICA is performed on the morphometric features obtained from VBM analysis. A total 2981 morphometric features extracted from the VBM were reduced to different combinations of 10 and 50 features using FastICA [108]. The mean, standard deviation of training/testing accuracy, sensitivity and specificity obtained during the 10 random trials for the PBL-McRBFN and the SVM classifiers on different ICA reduced features sets are presented in Table 8.6. In each trial, randomly 75% of total samples are selected for training and 25% for testing. From Table 8.6, we can see that the PBL-McRBFN classifier achieves better generalization performance with 10 ICA reduced features than 50 ICA reduced features as shown in Table 8.6. On 10 ICA reduced features, testing accuracy of PBL-McRBFN is 4% more than SVM with better sensitivity and specificity values. Similarly on 50 ICA reduced features, testing accuracy of PBL-McRBFN is 5% more than SVM classifier with better sensitivity and specificity values.

The testing accuracy of PBL-McRBFN is reduced by 3% on 50 features compared to 10 features, this is due to the redundancy of ICA features.

Table 8.6: Performance comparison on ICA reduced features data sets from an average of 10 trials

# ICA Features	Algorithm	Training Performance			Testing Performance		
		Accu-racy(%) mean (std)	Sensit-ivity(%) mean (std)	Specif-icity(%) mean (std)	Accu-racy(%) mean (std)	Sensit-ivity(%) mean (std)	Specif-icity(%) mean (std)
10	SVM	94.76 (2.52)	97.94 (1.02)	91.16 (4.97)	67.44 (3.46)	75.65 (8.50)	58.00 (10.85)
	PBL-McRBFN	93.28 (2.18)	94.55 (3.18)	91.83 (4.04)	71.39 (4.25)	63.91 (12.64)	80.00 (9.71)
50	SVM	98.35 (2.88)	98.52 (2.77)	98.16 (3.08)	63.25 (3.06)	67.82 (8.24)	58.00 (8.88)
	PBL-McRBFN	92.50 (9.69)	92.79 (9.59)	92.16 (1.27)	68.83 (2.94)	68.26 (7.68)	69.50 (10.12)

From Tables 8.5 and 8.6, it is observed that PBL-McRBFN classifier gives better results with 2981 morphometric features extracted from VBM analysis than reduced features from VBM and ICA analysis. PBL-McRBFN classification accuracy on VBM features set is 82.32% which is 11% more than the classification accuracy on ICA reduced 10 features set, this is because the considered binary classification problem consists of small number MRI volumes with high dimensional VBM features of PD patients and normal persons from both genders with different ages.

8.4.4 Identification of Imaging Biomarkers for PD

In this section, we present the identification of imaging biomarkers responsible for PD using PPMI MRI data set. In the previous section, PBL-McRBFN classifier performance has been evaluated on features obtained from VBM analysis and further reduced ICA features. VBM involves voxel-wise statistical analysis of MRI volumes and infers regions in which brain volume differs between PD and normal persons. All the inferred regions from VBM may not be useful to predict PD and further reduced ICA features do not provide any information relevant to the critical brain regions relevant to PD. In this study, we have conducted an analysis to identify most significant brain regions (imaging biomarkers) for PD. Identification of imaging biomarkers for PD can be considered a general feature selection problem. Feature selection techniques attempt to remove as

many irrelevant and redundant features as possible and to find a feature subset such that, with reduced low dimensional data, a machine learning classifier can achieve better performance. Filter and wrapper methods are two kinds of well-known feature selection techniques for high dimensional data [113]. In the filter method, features are selected on the basis of feature separability of training samples, which is independent of the learning algorithm. The separability only takes into account the correlations between the features, so the selected features may not be optimal. Wrapper methods search for critical features based on the learning algorithm, and often result in better results than filter methods. RFE is a computationally less intensive wrapper based feature selection method. In this study, we used RFE feature selection utilizing a PBL-McRBFN classifier. RFE utilizing PBL-McRBFN is referred as PBL-McRBFN-RFE. PBL-McRBFN-RFE conducts feature selection in a sequential elimination manner, which starts with all the features and discards one feature at a time.

Table 8.7: VBM detected and PBL-McRBFN-RFE selected regions responsible for PD

Feature Type	# Features	Identified Regions
VBM	2981	Superior temporal gyrus, Middle temporal gyrus, Parahippocampal gyrus, Sub-gyral, Insula
VBM + PBL-McRBFN-RFE	19	Superior temporal gyrus

VBM analysis detected a total of 2981 features. The corresponding brain regions to the 2981 VBM detected features are reported in Table 8.7. The mean testing performance of PBL-McRBFN on 2981 VBM features from 10 random trials is 82.32% as shown in Table 8.8. To identify most significant brain regions responsible for PD, the minimal set of features are found using PBL-McRBFN-RFE. After performing feature selection using PBL-McRBFN-RFE, 19 features among 2981 were selected. The corresponding brain regions to the 19 PBL-McRBFN-RFE selected features are reported in Table 8.7. MNI templates of the complete 2981 and 19 features regions are shown in Fig. 8.5. The mean testing performance of PBL-McRBFN on this selected 19 features from 10 random trials

is 87.21% as shown in Table 8.8. From Table 8.8, we can see that the selected 19 features from PBL-McRBFN-RFE approach give better prediction rate than 2981 features.

Table 8.8: PBL-McRBFN classifier performance on VBM detected and PBL-McRBFN-RFE selected features from an average of 10 trials

# Features	Training			Testing		
	Accu- racy(%) mean (std)	Sensit- ivity(%) mean (std)	Specif- icity(%) mean (std)	Accu- racy(%) mean (std)	Sensit- ivity(%) mean (std)	Specif- icity(%) mean (std)
2981 (VBM)	93.04 (2.45)	95.88 (2.92)	89.83 (6.20)	82.32 (2.50)	83.47 (6.41)	81.00 (4.59)
19 (VBM+ PBL-McRBFN-RFE)	92.50 (9.69)	92.79 (9.59)	92.17 (12.77)	87.21 (3.67)	87.39 (4.32)	87.00 (10.32)

Figure 8.5: Comparison of gray matter volume change - Normal persons vs. PD patients in Superior temporal gyrus region

We found PBL-McRBFN-RFE selected 19 voxels are located at the superior temporal gyrus brain region. The superior temporal gyrus is one of three (sometimes two) gyri in the temporal lobe of the human brain. The superior temporal gyrus is involved in auditory processing, including language and social cognition. Superior temporal gyrus is consistently reported in medical research studies as the biomarker of PD [137, 143, 144]. Hence, the brain region detected by PBL-McRBFN-RFE among VBM features from MRI volumes may play more significant role than others in PD.

Table 8.9: Performance comparison on vocal data set from an average of 10 trials

Algorithm	Testing Accuracy		
	Overall (η_o)	Average (η_a)	F -score
SVM	96.94 (2.40)	96.67 (2.16)	0.9221 (0.0286)
PBL-McRBFN	98.97 (1.07)	99.35 (0.68)	0.9934 (0.0069)

8.5 PD Diagnosis Based on Vocal Features

Research on PD has shown that approximately 90% of patients exhibit some form of vocal impairment [123]. PD patients display a constellation of vocal symptom that includes impairment in the normal production of vocal sounds, which is dysphonia. The voice of people with dysphonia will sound hoarse, strained or effortful. Telemonitoring of the PD disease using measurements of dysphonia has a vital role in its early diagnosis as the symptoms of PD occur gradually and mostly targeting the elderly people for whom physical visits to the clinic are costly and inconvenient.

The mean, standard deviation of testing efficiencies and F -score obtained during the 10 random trials with PBL-McRBFN and SVM classifiers are presented in Table 8.9. In each trial, randomly 75% of total samples are selected for training and 25% for testing. From the Table 8.9, we can see that on vocal data set, η_a of PBL-McRBFN is 3% more than SVM classifier with better F -score value. The reported results in the literature on the same vocal data is given in Table 8.10. From the Table 8.10, it is evident that PBL-McRBFN classifier generalization performance is also better than reported results in the literature on the same vocal data set. On 50-50% train-test combination, the best PBL-McRBFN generalization performance is 98.63% which is approximately 1% more than k -NN approach with fuzzy c-means clustering (97.93%) [128]. Thus, PBL-McRBFN classifier performs an efficient classification of the vocal features for prediction of PD.

8.6 PD Diagnosis Based on Gait Features

PD greatly influences the patients gait by reducing speed, stride length and total range of movement during walking. Gait analysis is a systematic study of human motion from the measurements of spatial-temporal parameters of the gait cycle, motion of joints and seg-

Table 8.10: PBL-McRBFN classifier performance comparison with studies in the literature on vocal data set

Study	Algorithm	Testing Accuracy
M. A. Little et. al[123]	SVM, 50 trials with bootstrap	91.4 (4.4)
Shahbaba et. al[145]	Dirichlet process multinomial logit, 5-fold cross validation	87.7 (3.3)
Psorakis et. al[146]	non-sparse Expectation- Maximization, 10-fold cross validation	89.5 (6.6)
Sakar et. al[124]	SVM, 50 trials with bootstrap	92.8 (1.2)
Pei-Fang Guo et. al[147]	Genetic Programming and Expectation Maximization, 10-fold cross validation	93.1 (2.9)
Resul Das et. al[125]	Neural network, 65-35% train-test data,	92.9
Mehmet et. al[127]	Adaptive Neuro-Fuzzy classifier with linguistic hedges, 10 random trials, 50-50% train-test data	94.72
F. Strom et. al[126]	9 Parallel neural networks, 30 random trials, 60-40% train-test data	91.2 (1.6)
PBL-McRBFN Approach	PBL-McRBFN, 10 random trials, 50-50% train-test data	96.83 (0.97)
PBL-McRBFN Approach	PBL-McRBFN, 10 random trials, 60-40% train-test data	97.67 (1.31)
PBL-McRBFN Approach	PBL-McRBFN, 10 random trials, 75-25% train-test data	99.35 (0.68)

Table 8.11: Performance comparison on gait data set from an average of 10 trials

Algorithm	Testing Accuracy		
	Overall (η_o)	Average (η_a)	F -score
SVM	77.56 (3.78)	77.37 (3.68)	0.7803 (0.0508)
PBL-McRBFN	83.90 (2.86)	84.36 (2.42)	0.8519 (0.0340)

Table 8.12: PBL-McRBFN classifier performance comparison with studies in the literature using gait patterns

Study	Algorithm	Testing Accuracy
Sang-Hong Lee et. al[121] 93 PD patients, 73 control	Neural network with weighted fuzzy membership functions, Single trial, 50-50% train-test data	77.33
PBL-McRBFN approach 93 PD patients, 73 control	PBL-McRBFN, Single trial, 50-50% train-test data	82.52
PBL-McRBFN approach 93 PD patients, 73 control	PBL-McRBFN, 10 random trials 75-25% train-test data	84.36 (2.42)

ments, force/moments, and electromyography patterns of muscle activation. Application of gait analysis in the assessment of PD is an important application.

The mean, standard deviation of testing efficiencies and F -score obtained during the 10 random trials for PBL-McRBFN and SVM classifiers are presented in Table 8.11. In each trial, randomly 75% of total samples are selected for training and 25% for testing. From the Table 8.11, we can see that η_a of PBL-McRBFN is 7% more than SVM with better F -score value. The reported result in the literature on the same gait features data is given in Table 8.12. From Table 8.12, we can see that on the same gait data set with 50-50% train-test combination, PBL-McRBFN generalization performance is approximately 5% more than neural network approach with weighted fuzzy membership functions on wavelet based feature extraction [121]. Thus, PBL-McRBFN classifier performs an efficient classification of the gait features for prediction of PD.

8.7 Summary

In this chapter, PD diagnosis problem was solved by employing PBL-McRBFN classifier. The early diagnosis of PD based on microarray gene expression and MRI features, and detection of PD based on vocal and gait features are presented. The quantitative comparison with SVM classifier and existing literature results clearly indicates the superior performance of the proposed PBL-McRBFN classifier for prediction of individuals with or without PD. Imaging biomarkers responsible for PD are detected with PBL-McRBFN-RFE approach using PPMI MRI data set. Identification of genes responsible for PD using feature selection techniques will help doctor's to track the development of the disease. This will be undertaken as a future study.

In the next chapter, we shall summarize the work done in this thesis and conclude it by giving plans for future directions.

Chapter 9

Conclusions and Future Works

9.1 Conclusions

This thesis focuses on the development and application of meta-cognitive sequential learning algorithms in radial basis function network for classification problems with fewer samples, high dimensional feature set, and high sample imbalance. First time in the literature, human meta-cognition principles are integrated in radial basis function network. Human like self-regulated learning helps to radial basis function network in achieving better generalization by proper selection of samples and strategies in learning. In thesis, two such meta-cognitive algorithms EKF-McRBFN and PBL-McRBFN developed to handle classification problems. One of the developed meta-cognitive learning algorithms PBL-McRBFN is applied to the early diagnosis of neurological diseases - Alzheimer's disease and Parkinson's disease. To summarize, the major contributions of this thesis are:

- (a) Development of an Extended Kalman Filter based Meta-cognitive Radial Basis Function Network (EKF-McRBFN) classifier.
- (b) Development of a Projection Based Learning algorithm for a Meta-cognitive Radial Basis Function Network (PBL-McRBFN) classifier.
- (c) Application of PBL-McRBFN classifier to the early diagnosis of Alzheimer's disease based on MRI scans and development of PBL-McRBFN-RFE feature selection approach for imaging biomarkers detection of Alzheimer's disease based on MRI scans.

- (d) Application of PBL-McRBFN to the early diagnosis of Parkinson's disease based on micro-array gene expression, MRI scans, gait and vocal features and application of PBL-McRBFN-RFE approach for imaging biomarkers detection of Parkinson's disease based on MRI scans.

The major conclusions from the above studies are:

(1) **EKF-McRBFN Classifier :**

McRBFN classifier has been developed based on principles of human meta-cognition and its sequential learning algorithm has been derived using Extended Kalman Filter (EKF). The McRBFN using EKF is referred to as, 'EKF-McRBFN'. The McRBFN has a cognitive and a meta-cognitive component that monitors and controls the learning ability of the cognitive component. A radial basis function with the Gaussian activation function is the cognitive component, and a self-regulatory learning mechanism is its meta-cognitive component. The cognitive component begins with zero hidden neurons and adds a neuron or updates an existing neuron based on the learning strategy chosen by the meta-cognitive component for every sample in the training data set. Thus meta-cognitive component controls the cognitive component by deciding *what-to-learn*, *when-to-learn* and *how-to-learn*. It realizes *what-to-learn* by deleting samples that contain knowledge similar to that already learnt by the network. It decides *how-to-learn* by choosing the sample to either add a neuron or by updating an existing neuron. While a sample is used to add a new neuron, the parameters of the new neuron are initialized based on the sample overlapping conditions. When a sample is used to update an existing neuron, its parameters are updated using the EKF. The meta-cognitive component decides *when-to-learn* by reserving the sample for future use. Performance study of EKF-McRBFN on benchmark classification problems shows superior performance of EKF-McRBFN, in comparison to existing learning algorithms.

(2) **PBL-McRBFN Classifier :**

Here, a PBL algorithm has been developed for a McRBFN. The McRBFN using

the PBL algorithm is referred to as, ‘PBL-McRBFN’. Similar to EKF-McRBFN, the PBL-McRBFN also has a cognitive and meta-cognitive component. An RBF network with a Gaussian activation function is the cognitive component of McRBFN and a self-regulatory learning mechanism is its meta-cognitive component. However, when a new neuron is added to the network, the projection based learning algorithm of PBL-McRBFN initializes the input parameters based on the distance criterion, and computes the optimum output weights by minimizing a hinge-loss error function. The problem of minimizing the error function is solved as a linear programming problem and the output weights are obtained by solving a set of simultaneous linear equations. Thus, while adding a new neuron, the existing neurons are used as pseudo-samples to represent knowledge of the past samples. Thus, the PBL-McRBFN explicitly uses knowledge in the past samples and a computationally efficient algorithm to map the input-output relationship defined by the training data set. Performance study of PBL-McRBFN on benchmark classification problems shows superior performance of PBL-McRBFN, in comparison to EKF-McRBFN and existing learning algorithms.

(3) Early Diagnosis of Alzheimer’s Disease using PBL-McRBFN based on MRI scans :

The developed PBL-McRBFN classifier is used for automatic, early diagnosis of AD from MRI scans using the OASIS [101] and ADNI [102] data sets. Study results show that PBL-McRBFN classifier accurately distinguishes AD patients from normal subjects in both the OASIS and ADNI data sets. It is observed from the results that the AD diagnosis using complete voxel-based morphometric features provide better performance than ICA reduced features, as the data sets contain very mild AD patients and fewer samples. Thus, from the studies conducted on OASIS and ADNI data sets, we can infer that human meta-cognitive principles in machine learning algorithm improve the classification performance significantly. Next, the generalization ability of PBL-McRBFN is studied by training the PBL-McRBFN on OASIS data set, and testing its performance on the ADNI data set. Performance results on this generalization ability study shows that the PBL-McRBFN is capable of generalization on unseen data set. Finally, PBL-McRBFN-RFE feature selection approach is

proposed to identify imaging biomarkers for AD. Imaging biomarkers responsible for AD are detected with the proposed PBL-McRBFN-RFE approach using OASIS data set. The imaging biomarkers identified using PBL-McRBF-RFE approach are in the parahippocampal gyrus, the hippocampus, the superior temporal gyrus, the insula, the precentral gyrus and the extra-nuclear regions. These regions are also indicated as biomarkers for AD in the medical literature [114, 115, 116].

Next, the PBL-McRBFN-RFE approach has also been used to identify imaging biomarkers for AD from the OASIS gender-wise and age-wise analysis. The results indicate the following:

In the 60-69 age group AD patients, gray matter atrophy is observed in the superior temporal gyrus region which is responsible for processing sounds. In the 70-79 age group AD patients, gray matter atrophy is observed in the parahippocampal gyrus and the extra-nuclear regions which are responsible for memory encoding and retrieval. In the 80-89 age group AD patients, gray matter atrophy is observed in the hippocampus, the parahippocampal gyrus and the lateral ventricular regions which are responsible for short-term memory to long-term memory, spatial navigation, memory encoding and retrieval. In male AD patients, gray matter atrophy is observed in the insula region which is responsible for emotion and consciousness. The insula region is also reported in AD research studies [117, 118] and it is associated with hypometabolism. In female AD patients, gray matter atrophy is observed in the parahippocampal gyrus and the extra-nuclear regions which are responsible for memory encoding and retrieval.

(4) **Early Diagnosis of Parkinson's Disease using PBL-McRBFN :**

PBL-McRBFN classifier is used in early diagnosis of PD using micro-array gene expression, MRI scans, gait and vocal features. Here, PBL-McRBFN classifier is used in diagnosis of PD in the following ways:

- Early detection of PD using PBL-McRBFN based on micro-array gene expression features data obtained from ParkDB database [138].

- Early detection of PD using PBL-McRBFN based on MRI scans obtained from the PPMI data set [140].
- Detection of PD using PBL-McRBFN based on vocal features data [122].
- Detection of PD using PBL-McRBFN based on gait features data [141].

In comparison with existing results in the literature the performance results from the above studies on these data sets show that the PBL-McRBFN performs better than existing results. Finally, imaging biomarkers responsible for PD are detected with the proposed PBL-McRBFN-RFE feature selection approach using PPMI MRI data set. PBL-McRBFN-RFE approach results shows that the superior temporal gyrus brain region plays a more significant role than others in detecting PD. The superior temporal gyrus is involved in auditory processing, including language and social cognition. Superior temporal gyrus is also consistently reported in medical research studies as the biomarker of PD [137, 143, 144].

9.2 Future Works

9.2.1 Plan of Work for McRBFN

Monitoring Signals for McRBFN :

In this thesis, the meta-cognitive learning considers only the error, nearest neuron distance and the predicted class labels as the monitoring signals. However, in the literature of human meta-cognition, the *feel-of-knowing* has been used as the monitoring signal. The term *feel-of-knowing* (FOK) stated as, an individual may fail to recall an item from memory but still feel that it would be recognized on a later test, a retrieval state, the original FOK definition proposed by Hart [148] as a composite of two criteria:

- A feeling that the sought-after information is known.
- A feeling that the sought-after information can be correctly identified on a later criterion test.

Hence, the addition of *feel-of-knowing* (FOK) measure using a combination of class-wise significance measure and reliability of the meta-cognitive component, as a monitoring

signal for PBL-McRBFN will be undertaken as a future study.

Selection of Radial Basis Function in McRBFN :

In this thesis, Gaussian radial basis function is considered in McRBFN. However, the Cauchy radial basis function is preferred in applications like image retrieval [149] and computerized tomography [150], while inverse multi-quadratic radial basis function is preferred in real-time signal processing application [151]. A q -Gaussian function parameterizes standard Gaussian distribution by replacing exponential expressions with q -exponential expressions [152]. Thus, the modification of the q -parameter allows the representation of different basis functions (Gaussian, Cauchy and inverse multi-quadratic etc) and helps the q -Gaussian function to match the shape of the kernel and the distribution of the distances better [152]. The q -parameter helps the q -Gaussian RBF to reproduce different RBF's [152]. For example, when $q \rightarrow 1$, the q -Gaussian converges to a Gaussian RBF, while $q \rightarrow 2$ and $q \rightarrow 3$ converges to a Cauchy RBF, and inverted multi-quadratic RBF, respectively. Thus, the q -Gaussian helps to realize different radial basis functions for different values of the parameter q . Therefore, it is desirable to employ an activation function like the q -Gaussian function that helps to employ radial basis functions with kernels, whose shape can be relaxed or contracted.

Feature Selection :

One important issue in the diagnosis of neurodegenerative diseases is the curse of dimensionality, where the data sets have fewer samples with huge dimensional feature set. Hence, there is a need to select appropriate feature set for better performance, and to select the best feature subset that are non-redundant and more relevant to the class distributions. We plan to work along this direction for better performance results.

9.2.2 Applications

Alzheimer's Disease Diagnosis :

In the present work, PBL-McRBFN classifier has been used to predict AD patients from normal persons based on MRI scans with VBM features extracted from MRI scans and PBL-McRBFN-RFE approach is used to AD imaging biomarkers detection based on

MRI scans. Similar approach can be used to predict Mild Cognitive Impairment (MCI) patients from normal persons. MCI is early stage of AD and it increases the risk of developing AD. If one were able to successfully treat MCI such that the progression of these individuals to AD could be delayed by one year, there would be a significant saving.

Parkinson's Disease Diagnosis :

In the present work, PBL-McRBFN classifier has been used to predict PD patients from normal persons based on micro-array gene expression and MRI scans, and PD detection based on gait and vocal features. Also, PBL-McRBFN-RFE approach is used to detection imaging biomarkers for PD based on MRI scans. Similar PBL-McRBFN-RFE feature selection approach can be used to detect biomarkers based on gene expression features.

Publications List

Journals

1. G. Sateesh Babu, S. Suresh and B. S. Mahanand, "A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for Parkinson's disease", *Expert Systems with Applications*, vol. 41(2), pp:478-488, 2014.
2. G. Sateesh Babu and S. Suresh, "Sequential Projection-Based Metacognitive Learning in a Radial Basis Function Network for Classification Problems", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24(2), pp: 194-206, 2013.
3. G. Sateesh Babu and S. Suresh, "Parkinson's Disease Prediction Using Gene Expression - A Projection Based Learning Meta-cognitive Neural Classifier Approach", *Expert Systems with Applications*, vol. 40(5), pp: 1519-1529, 2013.
4. G. Sateesh Babu and S. Suresh, "Meta-cognitive RBF Network and Its Projection Based Learning algorithm for classification problems", *Applied Soft Computing*, vol. 13(1), pp: 654-666, 2013.
5. G. Sateesh Babu and S. Suresh, "Meta-cognitive Neural Network for classification problems in a sequential learning framework", *Neurocomputing*, vol. 81, pp: 86-96, 2012.

Conference Proceedings

1. G. Sateesh Babu, S. Suresh and B. S. Mahanand, "Meta-cognitive q-Gaussian RBF Network for Binary Classification: Application to Mild Cognitive Impairment (MCI)", *Intl. Joint Conf. Neural Networks (IJCNN)*, Dallas (Texas, USA), pp: 1-8, 2013.

2. G. Sateesh Babu, S. Suresh and B. S. Mahanand, “A Sequential Projection Based Learning Meta-cognitive RBF Network Classifier: Application to Alzheimer’s disease detection”, *International Conference on Machine Learning (ICML): Workshop on Machine Learning for Clinical Data Analysis*, Edinburgh (Scotland) 2012.
3. G. Sateesh Babu, R. Savitha and S. Suresh, “A Projection Based Learning in Meta-cognitive Radial Basis Function Network for classification problems”, *Intl. Joint Conf. Neural Networks (IJCNN)*, Brisbane (Australia), pp: 2907-2914, 2012.
4. G. Sateesh Babu, S. Suresh and B. S. Mahanand, “Alzheimer’s disease detection using a Projection Based Learning Meta-cognitive RBF Network”, *Intl. Joint Conf. Neural Networks (IJCNN)*, Brisbane (Australia), pp: 408-415, 2012.
5. G. Sateesh Babu, S. Suresh, K. Uma Sangumathi and H. J. Kim, “A Projection Based Learning Meta-cognitive RBF Network Classifier for effective diagnosis of Parkinson’s disease”, *International Symposium on Neural Networks (ISNN)*, Shenyang (China) Part II, LNCS 7368, pp: 611-620, Springer-Verlag Berlin Heidelberg, 2012.

Bibliography

- [1] T. G. Dietterich, “Machine learning for sequential data: A review,” in *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 15–30, New York: Springer Verlag, 2002.
- [2] A. Wenden, *Learner strategies for learner autonomy*. Great Britain: Prentice Hall, 1998.
- [3] W. Rivers, “Autonomy at all costs: An ethnography of meta-cognitive self assessment and self management among experience language learners,” *Modern Language Journal*, vol. 85, no. 2, pp. 279 – 290, 2001.
- [4] R. Isaacson and F. Fujita, “Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning,” *Journal of the Scholarship of Teaching and Learning*, vol. 6, no. 1, pp. 39–55, 2006.
- [5] J. H. Flavell, “Meta-cognition and cognitive monitoring: A new area of cognitive-developmental inquiry,” *American Psychologist*, vol. 34, no. 10, pp. 906 – 911, 1979.
- [6] D. P. Josyula, F. C. Hughes, H. Vadali, B. J. Donahue, F. Molla, M. Snowden, J. Miles, A. Kamara, and C. Maduka, “Meta-cognition for self-regulated learning in a dynamic environment,” in *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop (SASOW)*, pp. 261 – 268, 2010.
- [7] M. T. Cox, “Meta-cognition in computation: A selected research review,” *Artificial Intelligence*, vol. 169, no. 2, pp. 104 – 141, 2005.
- [8] T. O. Nelson and L. Narens, “Metamemory: A theoritical framework and new findings,” *Psychology of Learning and Motivation*, vol. 26, pp. 125 – 173, 1990.

- [9] S. Suresh, K. Dong, and H. Kim, “A sequential learning algorithm for self adaptive resource allocation network classifier,” *Neurocomputing*, vol. 73, no. 16 – 18, pp. 3012 – 3019, 2010.
- [10] S. Suresh, R. Savitha, and N. Sundararajan, “A sequential learning algorithm for complex valued self-regulatory resource allocation network,” *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1061 – 1072, 2011.
- [11] “Alzheimer’s disease facts and figures,” 2013. http://www.alz.org/alzheimers_disease_facts_and_figures.asp.
- [12] “Statistics on Parkinson’s disease,” 2013. http://www.pdf.org/en/parkinson_statistics.
- [13] G. C. Chiang, P. S. Insel, D. Tosun, N. Schuff, D. Truran-Sacrey, S. Raptentsetsang, C. R. Jack, M. W. Weiner, and F. the Alzheimer’s Disease Neuroimaging Initiative, “Identifying Cognitively Healthy Elderly Individuals with Subsequent Memory Decline by Using Automated MR Temporoparietal Volumes,” *Radiology*, vol. 259, no. 3, pp. 844–851, 2011.
- [14] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, “Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database,” *NeuroImage*, vol. 56, no. 2, pp. 766 – 781, 2011.
- [15] D. N. Ripich, S. A. Petrill, P. J. Whitehouse, and E. W. Ziolo, “Gender differences in language of AD patients: A longitudinal study,” *Neurology*, vol. 45, no. 2, pp. 299–302, 1995.
- [16] B. R. Ott, K. L. Lapane, and G. Gambassi, “Gender differences in the treatment of behavior problems in Alzheimer’s disease,” *Neurology*, vol. 54, no. 2, pp. 427–432, 2000.
- [17] G. Sateesh Babu and S. Suresh, “Meta-cognitive Neural Network for classification problems in a sequential learning framework,” *Neurocomputing*, vol. 81, pp. 86 – 96, 2012.

- [18] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme Learning Machine for Regression and Multiclass Classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [19] J. C. Platt, “A resource allocation network for function interpolation,” *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.
- [20] V. Kadirkamanathan and M. Niranjan, “A Function Estimation Approach to Sequential Learning with Neural Networks,” *Neural Computation*, vol. 5, no. 6, pp. 954–975, 1993.
- [21] L. Yingwei, N. Sundararajan, and P. Saratchandran, “A sequential learning scheme for function approximation using minimal radial basis function neural networks,” *Neural Computation*, vol. 9, no. 2, pp. 461–478, 1997.
- [22] Y. Li, N. Sundararajan, and P. Saratchandran, “Analysis of minimal radial basis function network algorithm for real-time identification of nonlinear dynamic systems,” *IEE Proceedings: Control Theory and Applications*, vol. 147, no. 4, pp. 476–484, 2000.
- [23] G.-B. Huang, P. Saratchandran, and N. Sundararajan, “An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks,” *IEEE transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, vol. 34, no. 6, pp. 2284–2292, 2004.
- [24] G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation,” *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 57–67, 2005.
- [25] R. Zhang, G.-B. Huang, N. Sundararajan, and P. Saratchandran, “Improved GAP-RBF network for classification problems,” *Neurocomputing*, vol. 70, no. 16-18, pp. 3011 – 3018, 2007.
- [26] G.-B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: A new learning scheme of feedforward neural networks,” in *IEEE International Joint Conference on Neural Networks. Proceedings*, vol. 2, pp. 985–990, 2004.

- [27] G.-B. Huang, D. Wang, and Y. Lana, “Extreme Learning Machines: A Survey,” *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [28] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks.,” *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411–1423, 2006.
- [29] S. Suresh, R. V. Babu, and H. J. Kim, “No-reference Image Quality Assessment using Modified Extreme Learning Machine Classifier,” *Applied Soft Computing*, vol. 9, no. 2, pp. 541–552, 2009.
- [30] G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [31] G.-B. Huang and L. Chen, “Convex incremental extreme learning machine,” *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, 2007.
- [32] G.-B. Huang and L. Chen, “Enhanced random search based incremental extreme learning machine,” *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, 2008.
- [33] S. Wysoski, L. Benuskova, and N. Kasabov, “On-Line Learning with Structural Adaptation in a Network of Spiking Neurons for Visual Pattern Recognition,” in *Artificial Neural Networks - ICANN 2006* (S. Kollias, A. Stafylopatis, W. Duch, and E. Oja, eds.), vol. 4131 of *Lecture Notes in Computer Science*, pp. 61–70, Springer Berlin Heidelberg, 2006.
- [34] S. Soltic, S. Wysoski, and N. Kasabov, “Evolving spiking neural networks for taste recognition,” in *IEEE International Joint Conference on Neural Networks, 2008*, pp. 2091–2097, 2008.
- [35] F. Alnajjar, I. Bin Mohd Zin, and K. Murase, “A Spiking Neural Network with dynamic memory for a real autonomous mobile robot in dynamic environment,” in *IEEE International Joint Conference on Neural Networks, 2008*, pp. 2207–2213, 2008.

- [36] J. Wang, A. Belatreche, L. Maguire, and M. McGinnity, “Online versus offline learning for spiking neural networks: A review and new strategies,” in *2010 IEEE 9th International Conference on Cybernetic Intelligent Systems (CIS)*, pp. 1–6, 2010.
- [37] C. Cortes and V. Vapnik, “Support vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273 – 297, 1995.
- [38] P. Laskov, C. Gehl, S. Krüger, and K.-R. Müller, “Incremental Support Vector Learning: Analysis, Implementation and Applications,” *Journal of Machine Learning Research*, vol. 7, pp. 1909–1936, 2006.
- [39] G. Cauwenberghs and T. Poggio, “Incremental and decremental support vector machine learning,” in *Advances in Neural Information Processing Systems (NIPS 2000)*, vol. 13, pp. 409–415, MIT Press, Cambridge, MA, 2001.
- [40] J. Ma, J. Theiler, and S. Perkins, “Accurate Online Support Vector Regression,” *Neural Computation*, vol. 15, pp. 2683–2703, 2003.
- [41] M. Karasuyama and I. Takeuchi, “Multiple incremental decremental learning of support vector machines,” *IEEE Transactions on Neural Networks*, vol. 21, no. 7, pp. 1048–1059, 2010.
- [42] E. Pistikopoulos, M. Georgiadis, and V. Dua, *Process Systems Engineering: Volume 1: Multi-Parametric Programming*. No. 2 in Process Systems Engineering, John Wiley & Sons, 2007.
- [43] W. Liu, P. P. Pokharel, and J. C. Príncipe, “The Kernel Least-Mean-Square Algorithm,” *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 543–554, 2008.
- [44] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, “Mean square convergence analysis for kernel least mean square algorithm,” *Signal Processing*, vol. 92, no. 11, pp. 2624 – 2632, 2012.

- [45] P. P. Pokharel, W. Liu, and J. C. Príncipe, “Kernel least mean square algorithm with constrained growth,” *Signal Processing*, vol. 89, no. 3, pp. 257 – 265, 2009.
- [46] W. Liu, I. M. Park, Y. Wang, and J. C. Príncipe, “Extended Kernel Recursive Least Squares Algorithm,” *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3801–3814, 2009.
- [47] A. R. C. Paiva, I. Park, and J. C. Príncipe, “A reproducing kernel Hilbert space framework for spike train signal processing,” *Neural Computation*, vol. 21, no. 2, pp. 424–449, 2009.
- [48] S. Van Vaerenbergh, I. Santamaria, W. Liu, and J. C. Príncipe, “Fixed-budget kernel recursive least-squares,” in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1882–1885, 2010.
- [49] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, “Quantized Kernel Least Mean Square Algorithm,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, 2012.
- [50] P. Zhu, B. Chen, and J. C. Príncipe, “A novel extended kernel recursive least squares algorithm,” *Neural Networks*, vol. 32, pp. 349 – 357, 2012.
- [51] S. Zhao, and J. C. Príncipe, “A nonparametric information theoretic approach for change detection in time series,” in *The International Joint Conference on Neural Networks (IJCNN)*, pp. 1281–1284, 2011.
- [52] S. Zhao, B. Chen, P. Zhu, and J. C. Príncipe, “Fixed budget quantized kernel least-mean-square algorithm,” *Signal Processing*, vol. 93, no. 9, pp. 2759 – 2770, 2013.
- [53] S. Suresh, N. Sundararajan, and P. Saratchandran, “Sequential Multi-Category Classifier using Radial Basis Function Networks,” *Neurocomputing*, vol. 71, no. 7-9, pp. 1345–1358, 2008.
- [54] T. Harris and R. Hodges, eds., *The literacy dictionary: The vocabulary of reading and writing*. Newark, DE: International Reading Association, 1995.

- [55] L. R. Squire, “Mechanisms of memory,” *Science*, vol. 232, no. 4758, pp. 1612–1619, 1986.
- [56] S. Suresh, N. Sundararajan, and P. Saratchandran, “Risk Sensitive Loss Functions for Sparse Multi-category Classification Problems,” *Information Sciences*, vol. 179, no. 21, pp. 2621–2638, 2008.
- [57] T. Zhang, “Statistical behavior and consistency of classification methods based on convex risk minimization,” *Annals of Statistics*, vol. 32, no. 1, pp. 56–85, 2003.
- [58] B. Scholkopf and A.-J. Smola, *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.
- [59] H. Hoffmann, “Kernel PCA for novelty detection,” *Pattern Recognition*, vol. 40, no. 3, pp. 863–874, 2007.
- [60] M. A. Aizerman, E. A. Braverman, L. Rozonoer, “Theoretical foundations of the potential function method in pattern recognition learning,” *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.
- [61] C. Blake and C. Merz, “UCI repository of machine learning databases,” University of California, Irvine, Department of Information and Computer Sciences, 1998.
- [62] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub, “Multiclass cancer diagnosis using tumor gene expression signatures,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 26, pp. 15149 – 15154, 2001.
- [63] S. N. Omkar, S. Suresh, T. R. Raghavendra, and V. Mani, “Acoustic emission signal classification using fuzzy c-means clustering,” in *Intl. Conf. Neural Information Processing*, vol. 4, pp. 1872–1831, 2002.
- [64] S. Suresh, S. N. Omkar, V. Mani, and T. N. Guru Prakash, “Lift coefficient prediction at high angle of attack using recurrent neural network,” *Aerospace Science and Technology*, vol. 7, no. 8, pp. 595 – 602, 2003.

- [65] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [66] J. Demsar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.
- [67] R. L. Iman and J. M. Davenport, “Approximations of the critical region of the Friedman static,” *Communication in Statistics*, vol. 9, no. 6, pp. 571 – 595, 1980.
- [68] J. H. Zar, *Biostatistical analysis (4th Ed.)*. Englewood Clifs, New Jersey: Prentice-Hall, 1999.
- [69] O. J. Dunn, “Multiple comparisons among means,” *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52 – 64, 1961.
- [70] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 2011.
- [71] M. K. Aronson, W. L. Ooi, D. L. Geva, D. Masur, A. Blau, and W. Frishman, “Dementia: Age-dependent incidence, prevalence, and mortality in the old old,” *Archives of Internal Medicine*, vol. 151, no. 5, pp. 989–992, 1991.
- [72] M. M. Hoehn and M. Yahr, “Parkinsonism: onset, progression and mortality.,” *Neurology*, vol. 17, no. 5, pp. 427–442, 1967.
- [73] W. W. Barker, C. A. Luis, A. Kashuba, M. Luis, D. G. Harwood, D. Loewenstein, C. Waters, P. Jimison, E. Shepherd, S. Sevush, N. Graff-Radford, D. Newland, M. Todd, B. Miller, M. Gold, K. Heilman, L. Doty, I. Goodman, B. Robinson, G. Pearl, D. Dickson, and R. Duara, “Relative frequencies of Alzheimer disease, lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the state of Florida brain bank,” *Alzheimer Disease and Associated Disorder*, vol. 16, no. 4, pp. 203–212, 2002.

- [74] O. L. Lopez, J. T. Becker, C. A. Jungreis, D. Rezek, C. Estol, F. Boller, and S. T. DeKosky, "Computed tomography—but not magnetic resonance imaging—identified periventricular white-matter lesions predict symptomatic cerebrovascular disease in probable Alzheimer's disease," *Archives of Neurology*, vol. 52, no. 7, pp. 659–664, 1995.
- [75] K. A. Jobst, L. P. Barnetson, and B. J. Shepstone, "Accurate prediction of histologically confirmed Alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, X-ray CT, and Apo E4 in medial temporal lobe dementias. Oxford Project to Investigate Memory and Aging," *International Psychogeriatrics*, vol. 10, no. 3, pp. 271–302, 1998.
- [76] J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, and C. G. Puntonet, "Early detection of the Alzheimer's disease combining feature selection and kernel machines," in *Advances in Neuro-Information Processing* (M. Köppen, N. Kasabov, and G. Coghill, eds.), vol. 5507, pp. 410–417, Springer Berlin / Heidelberg, 2009.
- [77] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río, and the Alzheimer's Disease Neuroimaging Initiative, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 12160–1271, 2011.
- [78] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, 2008.
- [79] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of MRI," *Neurobiology of Aging*, vol. 29, pp. 514–523, 2008.

- [80] P. M. Thompson, K. M. Hayashi, R. A. Dutton, M. C. Chiang, A. D. Leow, E. R. Sowell, G. D. Zubicaray, J. T. Becker, O. L. Lopez, H. J. Aizenstein, and A. W. Toga, "Tracking Alzheimer's disease," *Annals Of The New York Academy Of Sciences*, vol. 1097, pp. 183–214, 2007.
- [81] P. Vemuri and C. Jack Jr., "Role of structural MRI in Alzheimer's disease," *Alzheimer's Research and Therapy*, vol. 2, no. 4, pp. 1–10, 2010.
- [82] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, O. Colliot, and the Alzheimer's Disease Neuroimaging Initiative, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, no. 6, pp. 579–587, 2009.
- [83] N. R. Giuliani, V. D. Calhoun, G. D. Pearlson, A. Francis, and R. W. Buchanan, "Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter disturbances in schizophrenia," *Schizophrenia Research*, vol. 74, no. 2–3, pp. 135–147, 2005.
- [84] C. R. Jack Jr, R. C. Petersen, Y. C. Xu, P. C. ÓBrien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1403, 1999.
- [85] R. J. Killiany, B. T. Hyman, T. Gomez-Isla, M. B. Moss, R. Kikinis, F. Jolesz, R. Tanzi, K. Jones, and M. S. Albert, "MRI measures of entorhinal cortex vs hippocampus in preclinical AD," *Neurology*, vol. 58, no. 8, pp. 1188–1196, 2002.
- [86] G. B. Frisoni, M. P. Laakso, A. Beltramello, C. Geroldi, A. Bianchetti, H. Soininen, and M. Trabucchi, "Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease," *Neurology*, vol. 52, no. 1, pp. 91–100, 1999.
- [87] A. Fornito, M. Yücel, S. J. Wood, C. Adamson, D. Velakoulis, M. M. Saling, P. D. McGorry, and C. Pantelis, "Surface-based morphometry of the anterior cingulate cortex in first episode schizophrenia," *Human Brain Mapping*, vol. 29, no. 4, pp. 478–489, 2008.

- [88] J. Ashburner, C. Hutton, R. S. J. Frackowiak, I. Johnsrude, C. Price, and K. J. Friston, "Identifying global anatomical differences: deformation-based morphometry," *NeuroImage*, vol. 6, no. 5–6, pp. 348–357, 1998.
- [89] N. Lepore, C. A. Brun, M. C. Chiang, Y. Y. Chou, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson, "Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 129–141, 2008.
- [90] J. Ashburner and K. J. Friston, "Voxel-Based Morphometry-The Methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
- [91] SPM8, "Wellcome trust center for neuroimaging, Institute of neurology, UCL, London, UK," 2011. <http://www.fil.ion.ucl.ac.uk/spm/>.
- [92] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 93 –105, 2007.
- [93] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Péligrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.
- [94] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856 – 867, 2011.
- [95] Y. Fan, D. Shen, and C. Davatzikos, "Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM," in *Proceedings of the 8th international conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I, MICCAI'05*, (Berlin, Heidelberg), pp. 1–8, Springer-Verlag, 2005.

- [96] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. J. Jr., “Alzheimer’s disease diagnosis in individual subjects using structural MR images: Validation studies,” *NeuroImage*, vol. 39, no. 3, pp. 1186 – 1197, 2008.
- [97] U. Yoon, J.-M. Lee, K. Im, Y.-W. Shin, B. H. Cho, I. Y. Kim, J. S. Kwon, and S. I. Kim, “Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia,” *NeuroImage*, vol. 34, no. 4, pp. 1405 – 1415, 2007.
- [98] M. J. McKeown and T. J. Sejnowski, “Independent component analysis of fMRI data: Examining the assumptions,” *Human Brain Mapping*, vol. 6, no. 5-6, pp. 368–372, 1998.
- [99] L. Xu, G. Pearlson, and V. D. Calhoun, “Joint source based morphometry identifies linked gray and white matter group differences,” *NeuroImage*, vol. 44, no. 3, pp. 777 – 789, 2009.
- [100] B. S. Mahanand, S. Suresh, N. Sundararajan, and M. A. Kumar, “Alzheimer’s disease detection using a Self-adaptive Resource Allocation Network classifier,” in *The International Joint Conference on Neural Networks (IJCNN)*, pp. 1930–1934, 2011.
- [101] D. S. Marcus, T. H. Wang, H. Parker, J. . G. Csernansky, J. C. Morris, and R. L. Buckner, “Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults,” *Journal of Cognitive Neuroscience*, vol. 19, pp. 1498–1507, 2007.
- [102] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, “The Alzheimer’s Disease Neuroimaging Initiative,” *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869 – 877, 2005.

- [103] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, L. W. Jennifer, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, K. Gunnar, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, “The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods,” *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.
- [104] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, pp. 839–851, 2005.
- [105] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak, “Statistical parametric maps in functional imaging: A general linear approach,” *Human Brain Mapping*, vol. 2, pp. 189–210, 1994.
- [106] M. García-Sebastián, A. Savio, M. Graña, and J. Villanúa, “On the use of morphometry based features for Alzheimer’s disease detection on MRI,” in *Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence (IWANN2009)*, (Salamanca, Spain), pp. 957–964, 2009.
- [107] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [108] H. Gavert, J. Hurri, J. Sarela, and A. Hyvarinen, “The FastICA package for MATLAB,” 2005. <http://www.cis.hut.fi/projects/ica/fastica/>.
- [109] Y. Fan and D. Shen, “Integrated feature extraction and selection for neuroimage classification,” in *Medical Imaging 2009: Image Processing*, (Lake Buena Vista, FL, USA), p. 72591U, 2009.
- [110] W. Yang, H. Xia, B. Xia, L. M. Lui, and X. Huang, “ICA-based feature extraction and automatic classification of AD-related MRI data,” in *Sixth International Conference on Natural Computation (ICNC)*, vol. 3, pp. 1261–1265, Aug. 2010.

- [111] R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D. P. Zhang, D. Rueckert, H. Soininen, J. Lötjönen, and the Alzheimer's Disease Neuroimaging Initiative, "Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease," *PLoS ONE*, vol. 6, no. 10, p. e25446, 2011.
- [112] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138 – 149, 2009.
- [113] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [114] H. Hampel, K. Burger, S. J. Teipel, A. L. Bokde, H. Zetterberg, and K. Blennow, "Core candidate neurochemical and imaging biomarkers of Alzheimer's disease," *Alzheimer's and Dementia*, vol. 4, no. 1, pp. 38 – 48, 2008.
- [115] J. A. Harasty, G. M. Halliday, J. J. Kril, and C. Code, "Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease," *Brain*, vol. 122, no. 4, pp. 675–686, 1999.
- [116] G. W. Van Hoesen, J. C. Augustinack, J. Dierking, S. J. Redman, and R. Thangavel, "The Parahippocampal Gyrus in Alzheimer's Disease: Clinical and Preclinical Neuroanatomical Correlates," *Annals of the New York Academy of Sciences*, vol. 911, no. 1, pp. 254–274, 2000.
- [117] A. L. Foundas, C. M. Leonard, S. M. Mahoney, O. F. Agee, and K. M. Heilman, "Atrophy of the hippocampus, parietal cortex, and insula in Alzheimer's disease: A volumetric magnetic resonance imaging study," *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, vol. 10, no. 2, pp. 81–89, 1997.
- [118] D. Bonthius, A. Solodkin, and G. Van Hoesen, "Pathology of the insular cortex in Alzheimer disease depends on cortical architecture," *Journal of Neuropathology and Experimental Neurology*, vol. 64, no. 10, pp. 910–922, 2005.

- [119] M. Arianna, C. Flavia, V. Augusto, C. Gemma, L. Giovanni, and A. Enrico, "The application of Russell and Burch 3R principle in rodent models of neurodegenerative disease: The case of Parkinson's disease," *Neuroscience and Biobehavioral Reviews*, vol. 33, no. 1, pp. 18–32, 2009.
- [120] C.-W. Cho, W.-H. Chao, S.-H. Lin, and Y.-Y. Chen, "A vision-based analysis system for gait recognition in patients with Parkinson's disease," *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 7033–7039, 2009.
- [121] S.-H. Lee and J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7338–7344, 2012.
- [122] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection," *BioMedical Engineering OnLine*, vol. 6, no. 23, pp. 1–19, 2007.
- [123] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.
- [124] C. Sakar and O. Kursun, "Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia," *Journal of Medical Systems*, vol. 34, no. 4, pp. 591–599, 2010.
- [125] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568–1572, 2010.
- [126] F. Strom and R. Koker, "A parallel neural network approach to prediction of Parkinson's Disease," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12470–12474, 2011.
- [127] M. F. Caglar, B. Cetisli, and I. B. Toprak, "Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier," *Journal of Engineering Science and Design*, vol. 1, no. 2, pp. 59–64, 2010.

BIBLIOGRAPHY

- [128] K. Polat, "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering," *International Journal of Systems Science*, vol. 43, no. 4, pp. 597–609, 2012.
- [129] G. T. Stebbins and C. G. Goetz, "Factor structure of the Unified Parkinson's Disease Rating Scale: Motor Examination section," *Movement Disorders*, vol. 13, no. 4, pp. 633–636, 1998.
- [130] H.-S. Jeon, J. Han, W.-J. Yi, B. Jeon, and K. S. Park, "Classification of Parkinson gait and normal gait using Spatial-Temporal Image of Plantar pressure," in *Engineering in Medicine and Biology Society. 30th Annual International Conference of the IEEE*, pp. 4672–4675, 2008.
- [131] N. M. Tahir and H. H. Manap, "Parkinson Disease Gait Classification based on Machine Learning Approach," *Journal of Applied Sciences*, vol. 12, no. 2, pp. 180–185, 2012.
- [132] C. R. Scherzer, A. C. Eklund, L. J. Morse, Z. Liao, J. J. Locascio, D. Fefer, M. A. Schwarzschild, M. G. Schlossmacher, M. A. Hauser, J. M. Vance, L. R. Sudarsky, D. G. Standaert, J. H. Growdon, R. V. Jensen, and S. R. Gullans, "Molecular markers of early Parkinson's disease based on gene expression in blood," *Proceedings of the National Academy of Sciences*, vol. 104, no. 3, pp. 955–960, 2007.
- [133] H. Shinotoh and D. B. Calne, "The Use of PET in Parkinsons-Disease," *Brain and Cognition*, vol. 28, no. 3, pp. 297–310, 1995.
- [134] A. Winogrodzka, P. Bergmans, J. Booij, E. A. Van Royen, J. C. Stoof, and E. C. Wolters, "[^{123}I] β -CIT SPECT is a useful method for monitoring dopaminergic degeneration in early stage Parkinson's disease," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 74, no. 3, pp. 294–298, 2003.
- [135] A. Schrag, C. Good, K. Miszkiel, H. Morris, C. Mathias, A. Lees, and N. Quinn, "Differentiation of atypical parkinsonian syndromes with routine MRI," *Neurology*, vol. 54, no. 3, pp. 697–702, 2000.

- [136] G. Becker, J. Seufert, U. Bogdahn, H. Reichmann, and K. Reiners, "Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography," *Neurology*, vol. 45, no. 1, pp. 182–184, 1995.
- [137] P. Nicola and J. B. David, "Imaging neurodegeneration in Parkinson's disease," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1792, no. 7, pp. 722–729, 2009.
- [138] C. Taccioli, V. Maselli, J. Tegner, D. Gomez-Cabrero, G. Altobelli, W. Emmett, F. Lescai, S. Gustincich, and E. Stupka, "ParkDB: a Parkinson's disease gene expression database.," 2011. <http://database.oxfordjournals.org/content/2011/bar007.abstract>.
- [139] G. K. Smyth, "Linear models and empirical bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 1, 2004. Article 3.
- [140] M. Kenneth, J. Danna, L. Shirley, S. Andrew, T. Caroline, S. Tanya, C. Chris, K. Karl, F. Emily, C. Sohini, P. Werner, M. Brit, K. Paracelsus-Elena, S. Todd, F. Mark, M. Claire, R. Alice, C. Cindy, S. John, M. Susan, S. Norbert, Z. Ying, T. Arthur, C. Karen, A. Alison, P. D. Blasio, P. Michele, T. John, S. Les, S. Andrew, H. Keith, E. Jamie, B. Deborah, R. David, L. Laura, F. Stewart, S. Barbara, H. Penelope, P. Emily, W. Karen, S. David, G. Stephanie, H. Robert, D. Holly, J. Joseph, H. Christine, S. Matthew, T. Baochan, L. Jim, B. Marne, F. Sam, T. Cathi-Ann, R. Irene, D. Cheryl, R. Linda, S. Fabienne, L. Elisabeth, S. Holly, O. Sanja, F. Hubert, W. Adrienna, B. Daniela, G. Katharina, G. Douglas, F. Deborah, M. Zoltan, G. Melissa, B. David, M. Sophie, B. Paolo, L. Katia, C. Tom, R. Bernard, G. Igor, G. Kim, C. Michelle, L. W. Katherine, O. Suzanne, F. Paulo, H. Tony, L. Johan, B. Marcel van der, D. R. Alastair, and T. Peggy, "The Parkinson Progression Marker Initiative (PPMI)," *Progress in Neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.
- [141] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *European Journal of Neuroscience*, vol. 26, no. 8, pp. 2369–2375, 2007.

- [142] H. Zheng, M. Yang, H. Wang, and S. McClean, “Machine Learning and Statistical Approaches to Support the Discrimination of Neuro-degenerative Diseases Based on Gait Analysis,” in *Intelligent Patient Management* (S. McClean, P. Millard, E. El-Darzi, and C. Nugent, eds.), vol. 189 of *Studies in Computational Intelligence*, pp. 57–70, Springer Berlin / Heidelberg, 2009.
- [143] M. K. Beyer, C. C. Janvin, J. P. Larsen, and D. Aarsland, “A magnetic resonance imaging study of patients with Parkinson’s disease with mild cognitive impairment and dementia using voxel-based morphometry,” *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 78, no. 3, pp. 254–259, 2007.
- [144] C. Summerfield, C. Junqué, E. Tolosa, and et al, “Structural brain changes in parkinson disease with dementia: A voxel-based morphometry study,” *Archives of Neurology*, vol. 62, no. 2, pp. 281–285, 2005.
- [145] B. Shahbaba and R. Neal, “Nonlinear Models Using Dirichlet Process Mixtures,” *Journal of Machine Learning Research*, vol. 10, pp. 1829–1850, 2009.
- [146] I. Psorakis, T. Damoulas, and M. A. Girolami, “Multiclass Relevance Vector Machines: Sparsity and Accuracy,” *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1588–1598, 2010.
- [147] P.-F. Guo, P. Bhattacharya, and N. Kharma, “Advances in Detecting Parkinson’s Disease,” in *Medical Biometrics* (D. Zhang and M. Sonka, eds.), vol. 6165 of *Lecture Notes in Computer Science*, pp. 306–314, Springer Berlin / Heidelberg, 2010.
- [148] J. T. Hart, “Memory and the feeling-of-knowing experience.,” *Journal of Educational Psychology*, vol. 56, no. 4, pp. 208 – 216, 1965.
- [149] K. Shkurko and X. Qi, “A Radial Basis Function and Semantic Learning Space Based Composite Learning Approach to Image Retrieval,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 945 –948, 2007.
- [150] J. Zhang and H. Li, “A Reconstruction Approach to CT with Cauchy RBFs Network,” in *Advances in Neural Networks - ISNN* (F.-L. Yin, J. Wang, and C. Guo,

- eds.), vol. 3174 of *Lecture Notes in Computer Science*, pp. 531–536, Springer Berlin Heidelberg, 2004.
- [151] A. Saranli and B. Baykal, “Complexity reduction in radial basis function (RBF) networks by using radial B-spline functions,” *Neurocomputing*, vol. 18, no. 1-3, pp. 183 – 194, 1998.
- [152] F. Fernández-Navarro, C. Hervás-Martínez, P. A. Gutiérrez, J. M. Peña-Barragán, and F. López-Granados, “Parameter estimation of q-Gaussian Radial Basis Functions Neural Networks with a Hybrid Algorithm for binary classification,” *Neurocomputing*, vol. 75, no. 1, pp. 123 – 134, 2012.