

# **Towards a Novel Medical Diagnosis System for Clinical Decision Support System Applications.**

Thesis submitted in accordance with the requirements of  
the University of Stirling for the degree of Doctor of  
Philosophy

by

Summrina Kanwal

Division of Computing Science and Mathematics  
School of Natural Sciences  
University of Stirling  
Scotland, UK

October 2016

# Acknowledgements

I consider this thesis to be the joint venture of a number of individuals, and without their support I would not have been able to accomplish this task. My heartfelt and deepest dedication goes to my father, Ghulam Habib, and mother, Gul Zaren. It has been solely due to their continuous hard work, commitment and dedication to our education that this achievement has been possible. I would like to thank my family members, especially my husband, Wajid, for his continual patience and encouragement, especially as he had to drive me a number of times to meetings with my supervisor. I would like to express my heartfelt thanks to my beloved children, Waheeba, Rohma, Awwab and Raed, for their patience throughout my tight schedule due to my PhD work and my job at the same time. You have been my inspiration, and have made it possible for me to accomplish this task. Last but not least, my sisters as well as my in-laws' cooperation has been unimaginable, and special thanks go to my father-in-law. It has been solely due to their continuous encouragement and support in every way possible that I have been able to achieve this goal. Thank you.

I am very much indebted to Prof Amir Hussain for his generous offer of the PhD position so that I could fulfil my dream of doctoral study in his prestigious Cognitive Big Data Informatics (CogBID, formerly COSIPRA) research lab. I have special thanks to offer as you have helped me achieve the best out of myself. I have only been able to explore my real strengths because of your continuous guidance and assignments. You have enabled me to realize how much I enjoy research endeavours. Thank you for the most moving and radical

ideas in our countless meetings, unparalleled guidance and support in writing and publishing the high quality research contributions — this PhD would definitely not have been possible without you!

Furthermore, I would like to thank Dr David Cairns, Dr Anwar Mirza (my master's degree supervisor) and Dr Anthony Maeder (Flinders university) for their support and encouragement. Special thanks go to Dr Kamran Farooq and Information Services, who printed and bound this document.

Lastly, I would like to thank my colleagues from the COSIPRA Lab, with whom I have had the opportunity to share and discuss research prospects. I would like to offer special thanks to Dr Kamran Farooq, whose has acted as my mentor throughout the PhD programme. I would also like to thank Grace, Lynn, Linda and Gemma for providing support throughout my PhD.

# Declaration

I understand the nature of plagiarism, and I am aware of the University's policy on this. I certify that this dissertation reports original work by me during my University project. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

---

Signature

---

Date

# Abstract

Clinical diagnosis of chronic disease is a vital and challenging research problem which requires intensive clinical practice guidelines in order to ensure consistent and efficient patient care. Conventional medical diagnosis systems inculcate certain limitations, like complex diagnosis processes, lack of expertise, lack of well described procedures for conducting diagnoses, low computing skills, and so on. Automated clinical decision support system (CDSS) can help physicians and radiologists to overcome these challenges by combining the competency of radiologists and physicians with the capabilities of computers. CDSS depend on many techniques from the fields of image acquisition, image processing, pattern recognition, machine learning as well as optimization for medical data analysis to produce efficient diagnoses.

In this dissertation, we discuss the current challenges in designing an efficient CDSS as well as a number of the latest techniques (while identifying best practices for each stage of the framework) to meet these challenges by finding informative patterns in the medical dataset, analysing them and building a descriptive model of the object of interest and thus aiding in medical diagnosis.

To meet these challenges, we propose an extension of conventional clinical decision support system framework, by incorporating artificial immune network (AIN) based hyper-parameter optimization as integral part of it.

We applied the conventional as well as optimized CDSS on four case studies (most of them comprise medical images) for efficient medical diagnosis and compared the results.

The first key contribution is the novel application of a local energy-based shape histogram (LESH) as the feature set for the recognition of abnormalities in mammograms. We investigated the implication of this technique for the mammogram datasets of the Mammographic Image Analysis Society and INbreast. In the evaluation, regions of interest were extracted from the mammograms, their LESH features were calculated, and they were fed to support vector machine (SVM) and echo state network (ESN) classifiers. In addition, the impact of selecting a subset of LESH features based on the classification performance was also observed and benchmarked against a state-of-the-art wavelet based feature extraction method.

The second key contribution is to apply the LESH technique to detect lung cancer. The JSRT Digital Image Database of chest radiographs was selected for research experimentation. Prior to LESH feature extraction, we enhanced the radiograph images using a contrast limited adaptive histogram equalization (CLAHE) approach. Selected state-of-the-art cognitive machine learning classifiers, namely the extreme learning machine (ELM), SVM and ESN, were then applied using the LESH extracted features to enable the efficient diagnosis of a correct medical state (the existence of benign or malignant cancer) in the x-ray images. Comparative simulation results, evaluated using the classification accuracy performance measure, were further benchmarked against state-of-

the-art wavelet based features, and authenticated the distinct capability of our proposed framework for enhancing the diagnosis outcome.

As the third contribution, this thesis presents a novel technique for detecting breast cancer in volumetric medical images based on a three-dimensional (3D) LESH model. It is a hybrid approach, and combines the 3D LESH feature extraction technique with machine learning classifiers to detect breast cancer from MRI images. The proposed system applies CLAHE to the MRI images before extracting the 3D LESH features. Furthermore, a selected subset of features is fed to a machine learning classifier, namely the SVM, ELM or ESN, to detect abnormalities and to distinguish between different stages of abnormality. The results indicate the high performance of the proposed system. When compared with the wavelet-based feature extraction technique, statistical analysis testifies to the significance of our proposed algorithm.

The fourth contribution is a novel application of the (AIN) for optimizing machine learning classification algorithms as part of CDSS. We employed our proposed technique in conjunction with selected machine learning classifiers, namely the ELM, SVM and ESN, and validated it using the benchmark medical datasets of PIMA India diabetes and BUPA liver disorders, two-dimensional (2D) medical images, namely MIAS and INbreast and JSRT chest radiographs, as well as on the three-dimensional TCGA-BRCA breast MRI dataset. The results were investigated using the classification accuracy measure and the learning time. We also compared our methodology with the benchmarked multi-objective genetic algorithm (ES)-based optimization technique. The results authenticate the potential of the AIN optimised CDSS.

# Glossary of Abbreviations

AIN	Artificial Immune Network
AIS	Artificial Immune System
ANOVA	Analysis of variance
AUROC	Area Under ROC
BIRADS	Breast Imaging-Reporting and Data System
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CDSS	Clinical Decision Support Systems
CLAHE	Contrast Limited Adaptive Histogram Equalisation
DCE	Dynamic-limited Contrast Enhanced
EHR	Electronic Health Records
ELM	Extreme Learning Machine
ESN	Echo State Network
HIS	Hospital Information Systems
IS	Immune System
JSRT	Japanese Society of Radiological Technology
LESH	Local Energy-based Shape Histogram (LESH)
MIAS	Mammographic Image Analysis Society
MLP	multi-layer perceptron
MRI	Magnetic Resonance Imaging

RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SLFN	Single Layer Feed Forward Network
SVM	Support Vector Machine
TCGA-BRCA	The Cancer Genome Atlas Breast Invasive Carcinoma
TCIA	The Cancer Imaging Archive

# Table of Contents

Chapter 1 - Introduction .....	1
1.1 Knowledge-Based Clinical Decision Support Systems.....	3
1.2 Non Knowledge-Based Clinical Decision Support Systems .....	3
1.3 Motivation.....	3
1.4 Original Contribution .....	4
1.5 Organisation of Thesis .....	6
1.6 Publications.....	7
1.6.1 Published.....	7
1.6.2 Submitted .....	8
Chapter 2 - Clinical Decision Support Systems (CDSS): State of the Art Systems and Methods.....	9
2.1 Clinical Decision Support Systems (CDSS) and their Variants .....	9
2.1.1 Computer Aided Detection (CADe).....	9
2.1.2 Computer Aided Diagnosis (CADx) .....	10
2.2 Historical Review of Clinical Decision Support Systems (CDSS).....	11
2.3 Medical Image Analysis using CDSSs .....	15
2.4 Survey on Techniques in CDSS Design.....	17
2.4.1 Medical Imaging Acquisition/Formation .....	18
2.4.2 Image Pre-processing Techniques .....	25
2.4.3 Segmentation Techniques .....	29
2.4.4 Shape Modelling and Pattern Recognition Techniques. ....	31
2.4.5 Classification Techniques .....	36
2.5 CDSS Evaluation Methods.....	42
2.6 The Role of CDSS in Patient Management.....	43
2.7 The Role of CDSS in Prognostic Tasks .....	43

2.8	Framework for Development of CDSS .....	44
Chapter 3 - The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications .....		
3.1	The Framework for CDSS Design.....	46
3.1.1	Dataset .....	47
3.1.2	Pre-processing .....	47
3.1.3	Image segmentation .....	53
3.1.4	Feature Extraction .....	55
3.1.5	Feature Selection .....	61
3.1.6	Classification Model Selection .....	64
3.2	Conclusion and Discussion .....	91
Chapter 4 - A Novel Application of a Local Energy-based Shape Histogram (LESH)–based CDSS for Breast Cancer Detection using Mammography and Magnetic Resonance Imaging.....		
4.1	Introduction .....	93
4.2	Local Energy-based Shape Histogram (LESH).....	94
4.2.1	Two-dimensional (2D) LESH Feature Extraction Technique .....	95
4.2.2	Three-dimensional (3D) Local Energy-based Shape Histogram (LESH) Feature Extraction .....	97
4.3	Simulation Results .....	101
4.3.1	Case Study I: Mammogram Data Sets .....	101
4.3.2	Case Study II: Lungs Cancer Detection using JSRT Data Set of Chest Radiographs .....	122
4.3.3	Case Study II: 3D Breast MRI images .....	132
4.4	Conclusion & Discussion.....	147
Chapter 5 - Nature-Inspired Algorithm for Machine Learning Classifier Optimisation 151		
5.1	The Biological Immune System (IS).....	152
5.1.1	Immune System Concepts .....	153

5.2	Artificial Immune Systems (AIS).....	155
5.2.1	Immune Engineering Framework to design AIS .....	156
5.2.2	Artificial Immune Networks (AIN).....	157
5.3	Artificial Immune Network System-Based Optimisation .....	159
5.3.1	Create the initial Search Space Population of the components of the application domain .....	163
5.3.2	Train the Classifier.....	164
5.3.3	Calculate the Fitness .....	164
5.3.4	Clone and Mutate the Cells .....	164
5.3.5	Rank Cells .....	165
5.3.6	One-dimensional Mutation.....	165
5.3.7	Gene Duplication .....	166
5.3.8	Suppress Cells .....	166
5.3.9	Create Population Diversity .....	167
5.4	Simulation and Results .....	168
5.4.1	Case Study I: Application to UCL Dataset .....	169
5.4.2	Case Study II: Application to 2D Medical Images.....	176
5.4.3	Case Study IV: Application to 3D breast MRI .....	180
5.5	Conclusions and Discussion .....	184
Chapter 6 -	Conclusion and Future Work.....	186
6.1	Benefits of the Developed CDSSs .....	187
6.2	Limitations of the Developed CDSSs .....	190
6.3	Future Expansions and Recommendations .....	191
6.3.1	Breast Cancer Detection with Gabor-filter based cascading network 191	
6.3.2	Towards a Novel Approach for Medical Image Segmentation ....	193
6.3.3	Sensitivity Analysis (SA) as a performance booster for LESH features extraction .....	193

6.3.4	Minimization of computation cost of AIN based optimization technique .....	193
6.3.5	Cognitively-inspired Iterative Learning System for an efficient clinical decision support system.....	194
6.3.6	State-of-the-art learning systems for radiologist training .....	196
6.3.7	Possible implications of Cognition-inspired Iterative Learning framework upon training radiologists. ....	196
6.3.8	Towards eMobile Clinical Decision Support Systems.....	197
	Bibliography .....	198

## List of Figures

Figure 2.1 Electronic Health Records (EHR) application in clinical trials [6].....	13
Figure 2.2 First ever introduced concept of the radio doctor appeared in news in 1924 as a cover photo of a popular magazine of that time [6].....	14
Figure 2.3 Medical system diagnosis model [17].....	18
Figure 2.4. Schematic representation of an x-ray [17].....	19
Figure 2.5. Samples of abnormalities in the MIAS database: (a) circumscribed mass (CIRC) mdb015, (b) ill-defined mass (MISC) mdb032, (c) speculated masses (SPIC) mdb145, (d) calcification (CALC) mdb241, and (e) architectural distortion (ARCH) mdb115. ....	21
Figure 2.6. Chest radiograph with common landmarks labelled [21].....	22
Figure 2.7. Schematic representation of MRI image acquisition. ....	23
Figure 2.8. Basic ESN architecture. Dashed lines are optional connections [8]. .....	40
Figure 2.9. Extreme Learning Machine [24]. ....	42
Figure 3.1. Overall framework for designing CDSS.....	46
Figure 3.2. MOGA for optimisation.....	72
Figure 3.3. A sample t-table .....	91
Figure 4.1. Overview of proposed methodology.....	103
Figure 4.2. LESH vector for different abnormalities obtained from software developed by (Zakir et al., 2011) (a) circumscribed masses (CIRC); (b) CIRC after enhancement; (c) LESH vector for CIRC; (d) ill-defined masses (MISC); (e) MISC after enhancement; (f) LESH vector for MISC; (g) speculated masses (SPIC); (h) SPIC after enhancement; (i) LESH vector for SPIC; (j) calcification (CALC); (k) CALC after enhancement; (l) LESH vector for CALC; (m) architectural distortion (ARCH); (n) ARCH after enhancement; (o) LESH vector for ARCH.....	105
Figure 4.3. ROC for classification of malignant/benign cases for MIAS data set. .....	111

Figure 4.4. ROC for classification of malignant/benign cases for INbreast data set. ....	112
Figure 4.5. ROC for classification of masses/calcifications cases for INbreast data set. ....	112
Figure 4.6 Samples of abnormalities in the MIAS database which has been misclassified (a) circumscribed mass (CIRC) mdb017, (b) ill-defined mass (MISC) mdb263, (c) speculated masses (SPIC) mdb179, (d) calcification (CALC) mdb231, and (e) architectural distortion (ARCH) mdb188.....	114
Figure 4.7. Classification accuracy vs number of features selection for INBreast dataset (malignant/Benign). ....	120
Figure 4.8. Classification accuracy vs number of features selection for MIAS dataset (malignant/Benign). ....	120
Figure 4.9. Sample chest radiograph from JSRT database (a) sample radiograph (b) segmented image (c) sample ROI containing malignant nodule (d) sample ROI containing benign nodule. ....	123
Figure 4.10. Comparison of different classifier performances with different-sized feature subset selections (nodule non-nodule lungs). ....	125
Figure 4.11. Comparison of different classifier performances with different-sized feature subset selections (malignant vs benign). ....	125
Figure 4.12. Workflow of the proposed methodology. ....	133
Figure 4.13. Sample T2-weighted MRI displaying Stage I from the TCGA-BRCA data set. ....	134
Figure 4.14. Sample T2-weighted MRI displaying Stage II from the TCGA-BRCA data set. ....	135
Figure 4.15. Sample T2-weighted MRI displaying Stage III from the TCGA-BRCA data set. ....	135
Figure 4.16. Graph depicting break down of the ROI into sub-regions, for 3D LESH feature vector computation. ....	136
Figure 4.17. ROC for classification between Abnormal/Normal cases for MRI data set (Linear SVM, and RBF SVM curves are overlapping, whereas Polynomial SVM curve is partially overlapping and thus some part of it is invisible.) ....	141
Figure 4.18. No of LESH Features selected vs. Accuracy for MRI dataset. ...	142

Figure 5.1. The overall system structure of the proposed algorithm.....	162
Figure 5.2. Learning curve epochs (time) for two datasets, using AIN-based ESN optimisation.....	172
Figure 5.3. Learning curve over epochs (time) for two datasets, using AIN-based ELM optimisation.....	173
Figure 5.4. Learning curve over epochs (time) for two datasets, using AIN-based SVM optimisation. ....	173
Figure 5.5 Grid search for SVM RBF for finding optimum values of (gamma and C) for PIMA dataset.....	174
Figure 5.6 Grid search for SVM RBF for finding optimum values of (gamma and C) for BUPA dataset.....	174
Figure 6.1. Three layer abnormality detection network.. ....	192
Figure 6.2. Detection system with Gabor Filter. ....	192

## List of Tables

Table 2.1. Medical images and their applications [17].....	19
Table 2.2. Cancer stages in MRI [22].....	24
Table 2.3. Common shape features for masses/nodules in medical images. ...	31
Table 3.1. Merits and demerits of difference enhancement techniques for mammograms. ....	51
Table 3.2. Merits and demerits of different segmentation techniques for medical images. ....	54
Table 3.3. Merits and demerits of different feature selection techniques. ....	63
Table 3.4. Merits and demerits of different optimization techniques.....	76
Table 3.5. Merits and demerits of different classification techniques. ....	82
Table 3.6. Confusion Matrix .....	87
Table 4.1. LESH feature extraction algorithm.....	97
Table 4.2. 3D LESH feature extraction algorithm.....	101
Table 4.3. LESH-based classification accuracy % (malignant/benign) for INbreast data set.....	108
Table 4.4. LESH-based classification $A_z$ value (malignant/benign) for INbreast data set. ....	109
Table 4.5. LESH-based classification accuracy % (malignant/benign) for MIAS data set. ....	109
Table 4.6. LESH-based classification $A_z$ value (malignant/benign) for MIAS data set. ....	109
Table 4.7. LESH-based classification accuracy % (mass vs micro calcification) for INbreast data set.....	110
Table 4.8. LESH-based classification $A_z$ value (mass vs micro calcification) for INbreast data set.....	110
Table 4.9. LESH feature with multiclass (one vs all) SVM classifier performance accuracy % for MIAS data set.....	110
Table 4.10. LESH-based features extraction binary classification (Using ELM ) results with 100 & 50 features selected in %.....	111

Table 4.11. Wavelet-based (Daubechies) feature extraction classification results for binary classification between malignant and benign cases for INbreast data set.....	113
Table 4.12. Wavelet-based (Daubechies) feature extraction classification (mass vs micro calcifications) results for INbreast data set. ....	113
Table 4.13. Wavelet-based feature extraction for multiclass (one vs all) SVM classifier performance accuracy % for MIAS data set. ....	113
Table 4.14. Wavelet-based (Daubechies) feature extraction binary classification results. ....	115
Table 4.15. Summarised results of other classifiers with lesh and wavelet features averaged over 10-fold cross-validation.....	115
Table 4.16. Gabor-based LESH feature extraction based classification results for binary classification (malignant/benign) for MIAS data set.....	116
Table 4.17. Gabor-based LESH feature extraction based classification results for binary classification (malignant/benign) for INbreast data set. ....	116
Table 4.18. Gabor-based LESH feature extraction based classification (mass vs micro calcifications) results for INbreast data set. ....	117
Table 4.19. Gabor-based LESH feature extraction based classification for multiclass (one vs all) SVM classifier performance accuracy % for MIAS data set. ....	117
Table 4.20. The results of t-test at level of significance = 0.05.....	117
Table 4.21. The results of t-test at level of significance = 0.05.....	119
Table 4.22. The results of t-test at level of significance = 0.05.....	121
Table 4.23. LESH features extraction based binary classification results for lungs with/without nodules (average of 10-fold cross-validation). ....	126
Table 4.24. LESH features extraction based binary classification results for malignant and benign nodules (average of 10-fold cross-validation). ....	126
Table 4.25. Wavelet (daubechies) features extraction based binary classification results for lungs radiograph data set with 100 selected feature coefficients (average of 10-fold cross-validation). ....	127
Table 4.26. Gabor-based LESH features extraction based binary classification results for lungs radiograph data set with 100 selected feature coefficients (average of 10-fold cross-validation). ....	128

Table 4.27. The results of t-test at level of significance = 0.05.....	129
Table 4.28. The results of t-test at level of significance = 0.05.....	130
Table 4.29. The results of t-test at level of significance = 0.05.....	132
Table 4.30. LESH-based classification accuracy % (abnormal/normal) for MRI data set. ....	139
Table 4.31. LESH-based classification Az value (abnormal/normal) for MRI data set. ....	139
Table 4.32. Selected LESH feature (50,100 & all) with multiclass (one vs all) SVM classifier performance accuracy % for MRI data set. ....	141
Table 4.33. The results of t-test at level of significance = 0.05.....	143
Table 4.34. Gabor-based LESH feature extraction-based classification results for binary classification between normal and abnormal cases for MRI data set. ....	144
Table 4.35. Gabor-based LESH feature extraction-based multiclass (one vs all) classification performance accuracy % for MRI data set. ....	144
Table 4.36. The results of t-test at level of significance = 0.05 for binary classification.....	145
Table 4.37. Wavelet (Daubechies) feature extraction-based classification results for binary classification between normal and abnormal cases for MRI data set. ....	145
Table 4.38. Wavelet-based feature extraction for multiclass (one vs all) classification performance accuracy % for MRI data set.....	146
Table 4.39. The results of t-test at level of significance = 0.05 for binary classification.....	147
Table 5.1. Example dataset .....	160
Table 5.2. Average of 10-fold cross validation results for classifier accuracy %/CPU Time for PIMA dataset.....	171
Table 5.3. Average of 10-fold cross validation results for classifier accuracy %/CPU Time for BUPA dataset.....	171
Table 5.4. The results of the t-test at the 0.05 level of significance.....	176
Table 5.5. Average of 10-fold classification results for classifier accuracy %/CPU Time for MIAS dataset. ....	177

Table 5.6. Average of 10-fold classification results for classifier accuracy %/CPU Time for INbreast dataset. ....	177
Table 5.7. Average of 10-fold classification results for classifier accuracy %/CPU Time for JSRT dataset.....	177
Table 5.8. The results of the t-test at the 0.05 level of significance.....	179
Table 5.9. Average of 10-fold classification results for classifier accuracy %/ learning time for Breast MRI dataset.....	181
Table 5.10. The results of the t-test at the 0.05 level of significance.....	183

## **Chapter 1 - Introduction**

This chapter presents an introduction to, and the motivation for, designing Clinical Decisions Support Systems (CDSSs), while emphasising the current state of healthcare and summarising the global trends.

Interest in CDSSs dates back to the 1960's and efforts to improve healthcare and enhance the process of medical decision making. Osheroff et al. expected CDSSs systems to do the following things: "provide the right information, to the right person, in the right format, through the right channel, at the right point in workflow to improve health and health care decisions and outcomes" [5].

Recognition of the variable quality of medical care provided to patients has drawn attention to enhanced medical care for patients. CDSSs help improve medical care in the diagnosis of the patient by analysing available data for a pathophysiological explanation of the disease symptoms. For this, CDSSs provide with a diagnosis process - a replacement of the traditional hectic process which comprises of completing a set of inquiry questions, prescribing tests, procedures, and assessments of the results [4]. Further, knowing the diagnosis, it is a challenge to suggest the treatment.

It is still a challenge to evolve CDSSs as replacement for physician's knowledge and experience: regarding the treatment plan while keeping the patient's response to therapy in view; choosing an alternative approach for medication; or the assessment of prognosis for cure or death risk [4].

Among the problems encountered in clinical practice which can benefit from the aid of CDSSs are computer programs that offer patients specific

## Introduction

recommendations, or management based on these specific diagnoses which can improve clinical decisions. Clinicians can stay abreast of the latest technology and apply it to caring for individual patients and plan a treatment accordingly [4].

The excellent CDSSs provided efficient decision-making based on the provision of [4]: access to accurate patient data in a suitable quantity for making decisions.

A lot of information hinders decisions making, as synthesising the knowledge intelligently and rapidly is difficult in that case. Additional data may confuse and is a major challenge to cope with. It is important to apply computation tools to this data to synthesise a useful outcome.

“CDSSs provide patient-specific, situation-specific alerts, reminders, physician order sets, or other recommendations for direct action; and also organize and present information in a way that facilitates problem solving and decision making, as in dashboards, graphical displays, documentation templates, structured reports, and order sets” [4].

These systems help make decisions in diverse areas, like patient management, diagnosis and treatment -which involves doctors and nurses-while the other area involves higher decision making for the hospital management like inventory management, finance management, and record keeping [1].

CDSSs are categorised as knowledge-based systems, or non-knowledge-based systems where the difference between them is summarised below.

## **1.1 Knowledge-Based Clinical Decision Support Systems**

These are expert systems aimed at building computer programs that can simulate human thinking. These systems have evolved over time to support real life patient care and help experts in decision making. These systems consist of three major parts: the knowledge base, the inference engine and the user interface. The knowledge base consists of compiled information in the form of if-then rules, while the inference engine comprises of formulas for combining knowledge rules with patient data [2].

## **1.2 Non Knowledge-Based Clinical Decision Support Systems**

These systems use pattern recognition, machine learning, and data mining methods to learn from the patterns in clinical data. They have a narrow focus on a list of systems and its diagnosis for a single disease, as opposed to the knowledge-based approach which is a rule-based method for the diagnosis of many diseases. Our research aims to focus on the study and development of non-knowledge-based CDSSs.

## **1.3 Motivation**

The development of twenty-first century CDSSs technology is vital to meet the changes in healthcare practices.

CDSSs have taken on increasing importance due to challenges in (1) handling bulky clinical knowledge and information, (2) meaningful use of medical datasets and (3) providing personalised healthcare services [4].

## Introduction

The aim of our research is to build non-knowledge based CDSSs that can process data for clinicians so as to provide insight into medical problems and solve them by creating formal computational models. The system can offer:

- Societal benefits, as it aids practitioners in better patient diagnosis, thus leading to better clinical outcomes.
- Developed Electronic Health Records (EHR) of patient information to carry out clinical research and eliminate the manual technique of medical data analysis. This makes it easy for clinicians to study patients' information based on currently available data.
- The creation of an infrastructure for clinicians to use the system in their practices to assist in patient care, hence providing patients with counsel on illness prevention.

The principal objective of our interdisciplinary research work is the development of robust non-knowledge-based clinical decision support systems for efficient medical diagnosis. Further, we have developed CDSSs based on that framework, for handling a variety of medical datasets while identifying the best practices of state of the art technologies for solving problems of medical diagnosis. We also investigated the latest trends in the field of CDSS, and how the state of the art technology might be directed to meet future needs.

### **1.4 Original Contribution**

Clinical Decision Support Systems (CDSS) play an important role in the early diagnosis of fatal diseases like cancer. These systems help in detecting or

## Introduction

diagnosing abnormal conditions in medical datasets that may consist of images or other disease relevant information for multiple patients. We aim to enhance the performance of current CDSSs by using state-of-the-art image processing and machine learning techniques, especially for breast and lung cancer diagnosis. For this, we recently proposed a novel application of a local energy-based shape histogram (LESH) as the feature set for recognition of abnormalities in X-rays, mammograms and MRI images, coupled with the optimised benchmark machine learning classifiers to enhance its classification performance.

The major contributions include:

1. An Application of the conventional CDSS framework for designing new generation applications which employ state-of-the-art machine learning and image processing as well as optimization techniques to enhance medical diagnosis.
2. A novel application of the two-dimensional LESH feature extraction technique to extract resilient features from mammograms for breast cancer diagnosis as well as from chest radiographs for lung cancer detection.
3. Novel formulation of a three-dimensional (3D) LESH model to detect breast cancer and its stages in volumetric medical images. It is a hybrid approach which combines the three-dimensional LESH feature extraction technique with machine learning classifiers to detect breast cancer from MRI images.

4. Application of the Artificial Immune Network (AIN) for optimising machine learning classifiers.
5. Validation of the proposed CDSS framework's performance by application to UCL data sets, two-dimensional mammograms, chest radiographs and three-dimensional MRI images.

### **1.5 Organisation of Thesis**

Chapter 2 provides a general introduction to clinical decision support systems, moving from definition, historical perspective, review of the existing literature concerning diverse methods in the field of image processing, pattern recognition, machine learning and optimisation; which are used as the baseline for the contribution to knowledge in this thesis. Finally, a brief analysis of the key features and benefits of such systems follows.

Chapter 3 discusses the conventional framework for developing CDSSs, while the following two chapters describe the CDSS applications developed on the basis of that framework and built around the author's publications. Chapter 4 presents a novel application of a local energy-based shape histogram (LESH) Feature Extraction Technique to Breast Cancer Detection using mammography, chest X-ray (radiograph) and magnetic resonance imaging (MRI). Chapter 5 describes the application of the proposed artificial immune network (AIN) based optimisation technique for enhancing machine learning classifiers' performance when applied to UCL machine learning repository, mammograms, radiographs and MRI datasets. Chapter 6 presents an analysis of the overall research work and discusses future directions.

## 1.6 Publications

The following papers are the outcome of the research presented in this thesis:

### 1.6.1 Published

1. Summrina Kanwal Wajid, Amir Hussain, and Bin Luo, "An efficient Computer Aided Decision Support System for breast cancer diagnosis using Echo State Network classifier," *Computational Intelligence in Healthcare and e-health (CICARE) 2014*, pp.17-24.
2. Summrina Kanwal Wajid and Amir Hussain, "Local energy-based shape histogram feature extraction technique for breast cancer diagnosis," *Expert Systems with Applications*, vol. 42, no. 20, pp. 6990-6999, 15 November 2015, ISI-SCI IF: 2.9.
3. Summrina Kanwal Wajid, Amir Hussain, Kaizhu Huang, and Wadii Boulila, "Lung Cancer detection using a novel application of a Local Energy-based Shape Histogram (LESH) Feature Extraction and Machine Learning Classification Techniques," *International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC) 2016*, pp. 22-23 Aug, 2016, Stanford University, USA.
4. Summrina Kanwal Wajid, Amir Hussain, and Bin Luo, "An Investigation of Machine Learning and Neural Computation Paradigms in the Design of Clinical Decision Support Systems (CDSSs)," *Brain-Inspired Cognitive Systems (BICS): 2016*.

**1.6.2 Submitted**

5. Summrina Kanwal Wajid and Amir Hussain, "Novel Artificial Immune Network-based Optimization of Machine Learning Classifiers," *Submitted to IEEE Transactions on Systems, Man, and Cybernetics: Systems*, ISI Impact Factor (IF): 1.598.
6. Summrina Kanwal Wajid and Amir Hussain, "Local Energy-based Shape Histogram (LESH)-Based Clinical Decision Support System for Breast Cancer Detection using Magnetic Resonance Imaging (MRI)," *Submitted to Expert Systems with Applications.*, ISI-SCI IF: 2.9.
7. Summrina Kanwal Wajid, Amir Hussain, Bin Luo, Kaizhu HUANG, "A Review of State-of-the-Art Machine-Learning and Neural Computation Methods for the Efficient Design of Clinical Decision Support Systems (CDSSs)," *Submitted to Cognitive Computation*, Impact Factor 1.933.

## **Chapter 2 - Clinical Decision Support Systems (CDSS): State of the Art Systems and Methods**

CDSS help physicians and radiologists to treat and analyse patients by combining their competency physicians with the capabilities of computers. This chapter introduces the general concept of the Clinical Decision Support System (CDSS) and its variants; its evolution over time augmented with the potential benefits in presenting medical diagnosis and reviewing state of the art techniques applied in designing next generation CDSS.

### **2.1 Clinical Decision Support Systems (CDSS) and their Variants**

Any computerised system that helps physicians and radiologists to treat and analyse patients falls under the category of CDSS. These systems improve the diagnosis of cancer or other similar diseases by combining the competency of radiologists and physicians with the capabilities of computers. They are known by different titles, depending on their special characteristics and contributions to the field of medical diagnosis [1]. Some of these titles are described below.

#### **2.1.1 Computer Aided Detection (CADe)**

These systems process medical images so that the radiologist can detect the locations of previously undetected abnormalities. Everyday radiologists have to practise observing images, as well as inferring their patterns. To reduce errors occurring in this practice, the concept of double reading was introduced for screening the images and for thus minimising the false negative and false positive rates. Double reading was not considered a practical alternative, due

to extensive labour requirement. Furthermore, CADe systems were introduced to reduce the manual overhead while reducing errors introduced due to observation oversight [226].

CADe systems recognise the patterns which a radiologist examines while analysing a medical image. In the case of mammograms, the CADe systems look for micro-calcifications and masses while in the case of chest x-rays they search for pulmonary nodules. Similarly, in the case of breast MRI images, they look for a possible region of abnormality, either malignant or benign [226].

### **2.1.2 Computer Aided Diagnosis (CADx)**

When compared to CADe systems, where the focus lies in marking the abnormal regions in the medical image, computer-aided diagnosis (CADx) systems aim to evaluate abnormal regions (e.g. whether they are cancerous/non-cancerous). These systems are also applied in the diagnostic process wherein the aim is to determine the intensity or stage of a detected abnormality, thus helping to reduce the biopsy rate.

It is a prime research area in medical imaging analysis where the computer output is considered as a “second opinion” when the physician is making a diagnosis decision. Hence CADx systems complement the physician in making a better diagnosis [227].

CADx systems make use of state-of-the-art computational algorithms to improve cancer or other disease diagnosis capabilities, specifically by analysing medical images like mammograms, x-rays, CT scans and so on [227] and have

thus become an integral part of routine patient clinical observation for cancer detection.

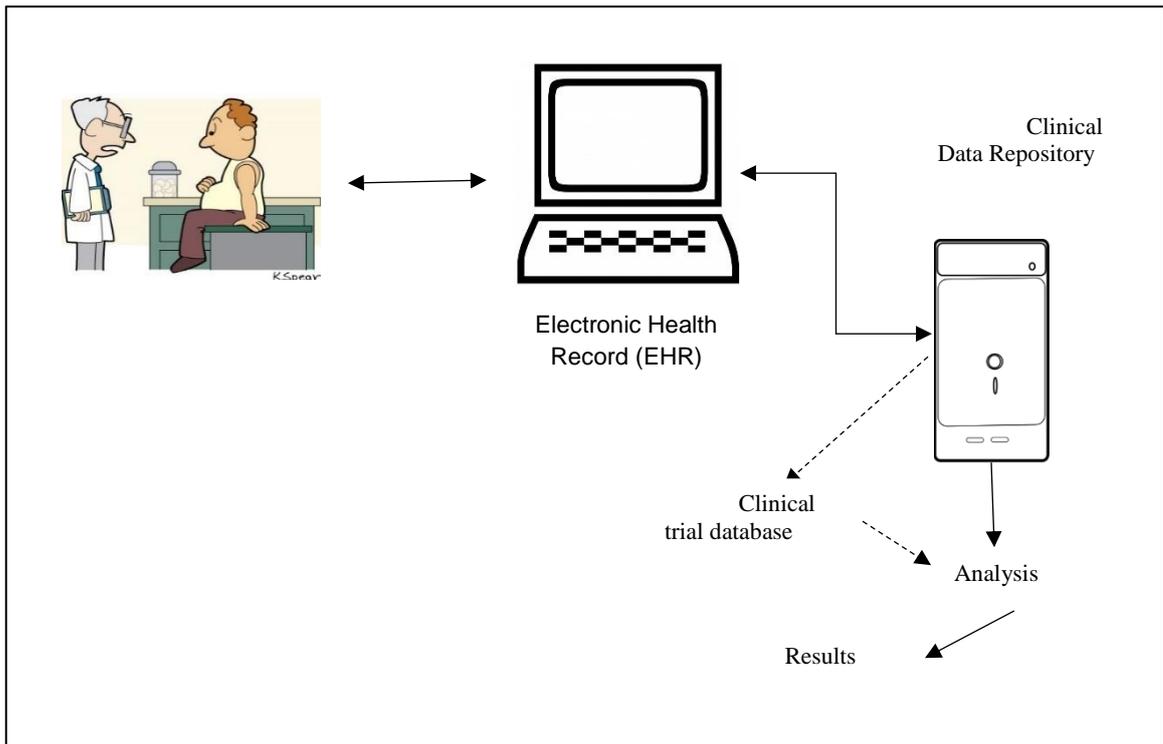
### **2.2 Historical Review of Clinical Decision Support Systems (CDSS)**

Clinical Decision Support Systems (CDSS) help in detecting or diagnosing abnormal conditions in medical datasets that may consist of images or other disease-relevant information for multiple patients. It is vital to integrate validated clinical information and convert it into knowledge that can be helpful in the diagnosis, prognosis and treatment of a patient. This entails the development of efficient CDSSs that can help health practitioners to provide medical health facilities [5].

Following the invention of the first digital computer in the 1940s, physicians realised the dramatic effect these devices may have on clinical practises. Further, the advent of the World Wide Web (WWW), smartphones, social networking and wireless communication, in the 1990s, revolutionised the concept of international health care systems. The bulk of medical data being produced was also making it difficult for public health planners and health institution managers to cope with planning with new health care environments, as this required a deep understanding of the role of information technology in the current environment. Most health care institutes were motivated to develop an integrated computer-based information management environment with tools to assist them with patient care. These medical tools were expected to assist in reporting test results, the direct entry of patient information into transcribed reports, telemedicine etc. as well as administration of financial topics like

patient tracking, inventory management, analysis of treatment and procedure outcomes, quality assurance, access to digital libraries and drug information databases as well as bibliographic searches. The key motive was to evolve an integrated environment for clinicians and patients to foster treatment and problem-solving [6] [208].

Previous research in the clinical environment was constrained by the clumsy error-prone paper-based methodology of acquiring data manually for clinical trials and statistical analysis. To improve upon this, and hence eliminate the need for traditional paper-based medical records, following the 19<sup>th</sup> century introduction of “lab notebooks” so clinicians could record their observations and plans for their respective patients, electronic health records (EHR) were suggested, as depicted in Figure 2.1. Henceforth many organisations have since attempted to integrated decision-making tools with their electronic health records (EHR) systems to provide computer-based diagnosis to practitioners [6].



**Figure 2.1 Electronic Health Records (EHR) application in clinical trials [6].**

Medical practitioners realised there was a need to develop systems to assist clinicians by learning from past experiences and envisioned an interconnected community of clinicians and institutions, thus they built digital data resource using EHR. The concept was introduced as a learning health care system by the Institute of Medicine (IOM) in USA [7]. Through the introduction of new communication technologies clinicians gained a creative way to provide high quality care to their patients interactively. The concept of telemedicine emerged in the early 20<sup>th</sup> century but was initially limited for the last 30 to 40 years due to lack of communication facilities (only telephone conversation was possible). The concept of telemedicine has grown rapidly after the introduction of telecommunications and the internet [6].



Figure 2.2 First ever introduced concept of the radio doctor appeared in news in 1924 as a cover photo of a popular magazine of that time [6].

Long before the introduction of television, many creative people used to think about how doctors and patients could communicate over long distances through advanced technologies. An example of their imagination was the concept of the radio doctor that appeared the in news in 1924 as the cover photo of a popular magazine at that time, as depicted in Figure 2.2. One of the

early attempts to enhance health facility was the development of hospital information systems (HIS), which although short-term was formidable. In the US, the beginning was the development of the MEDINET project at General Electric, followed by work in other cities like Boston, Massachusetts and so on. Furthermore, a number of applications were developed in Massachusetts General Hospital (MGH) by Barnett and his group in the 1960-80 era. This was further followed in Warner at Latter Day Saints (LDS) Hospital (Salt Lake City, Utah), Kaiser Permanente (Oakland, California), Stanford University (Stanford, California) and at Lockheed (Sunnyvale, California) [6]. Afterwards, a new generation of distributed HIS began to emerge in the 1980s [6].

In the late 1970s, with the advent of microprocessors and personal computers, a number of computer-based information management tools began to emerge as commercial products, thus the implications of personal computer (PC) in the application of patient care and clinical investigation was introduced [6].

A particular role of CDSS is “diagnosis”, which is carried out as part of a medical process by physicians before choosing a therapy for a particular patient. It is a cognitive task, which incorporates patient assessment, disease diagnosis and management, as well as analysis of treatment results and monitoring of disease progression [6].

### **2.3 Medical Image Analysis using CDSSs**

Early attempts to diagnose disease from medical image analysis using CDSS dates back to 1960 and were conducted by Lodwick et al. [8], Myers et al. [9] and Winsbarg et al. [10]. These studies were followed in the 1980s by the Kurt

Rossmann Laboratories for Radiologic Image Research in the Department of Radiology at the University of Chicago. The early research focused on the development of a picture archiving and communication system (PACS) [11]. The theory was extended for development of a system that would help radiologists in their daily work routines of reading images and diagnosing, which led to the concept of CDSS. Initially, the prime focus of the studies was diagnosis of breast cancer, cardiovascular disease and lung cancer. Fujita et al. [12] and Hoffmann et al. [13] focused on detection and quantitative analysis of lesions in vascular imaging. Giger et al. [14, 15] focused on detection of lung nodules in chest radiographs, while Chan et al. [16] explored mammograms for possible disease analysis [11].

By the mid-1980s, the basic structure of a CDSS for lung and breast cancer detection had been developed and applied at the Rossmann Laboratories. The system had a high false positive rate of, on average, 4 fp per mammogram for detecting clustered micro-calcifications, with a sensitivity of 85%. An observer performance study suggested that radiologists using CDSS performed better than those not using CDSS, and thus that CDSS systems could help improve diagnosis [11].

Initial steps to introduce computer diagnosis in radiology were unsuccessful, as computers were less powerful, advanced image processing techniques did not yet exist, and digital images were not easily available [11]. Since then, the performance of CDSS systems has improved by orders of magnitude. Recent advances provide a new look at CDSS with more effective and efficient algorithms integrated into the clinical workflow [11].

## 2.4 Survey on Techniques in CDSS Design

CDSS depend on many techniques from the fields of medical image acquisition, image processing and pattern recognition, machine learning, data mining, statistics, signal processing for data collection, organisation and analysis as well as optimisation to enhance system performance. These techniques try to find a pattern in the data, which can help to build a descriptive model of the object of interest and diagnose any abnormalities in the dataset [11]. Machine learning methods can distinguish between malignant and benign cases, data mining enables the extraction of information patterns from quantitative data and aids the diagnosis of new cases, and data reduction is useful for feature selection.

Medical image processing may be applied in three levels (Figure 2.3), which comprise the following steps:

- *Image acquisition/formation* is the process of forming an image from an image-capturing device, which is analogous to digital conversion and sampling.
- *Pre-processing* comprises of techniques, which help to better visualise objects in images with low resolution. Common methods include histogram transformations and filtering [17]. Filtering includes operations such as enhancement, de-blurring and image sharpening.
- *Segmentation* partitions the image into segments by assigning labels to pixels in the image such that pixels with the same label share common

characteristics. Among the most common methods for image segmentation are classical region growers and clustering algorithms [17].

- *Shape modelling* is applied when necessary to describe the shape of an object in more detail. It is a more compact representation of an object in the image, also called image features.
- *Classification* assigns abnormalities or diseases to a certain class based on the features selected from the image.

Medical imaging techniques and modalities have evolved over time, requiring improved computer vision and machine learning techniques for handling novel modalities.

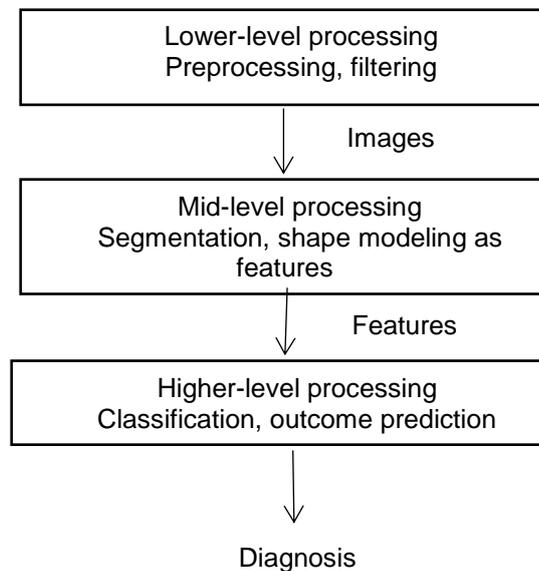


Figure 2.3 Medical system diagnosis model [17].

#### 2.4.1 Medical Imaging Acquisition/Formation

It is a technology that is applied to attain images of human body [216]. Medical images of human organs, such as the heart, breast, lungs and brain, aid medical diagnosis in radiology, psychiatry, cardiology and internal medicine

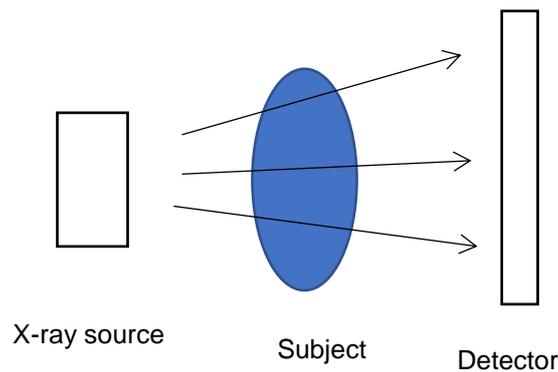
[17]. This is a rapidly advancing field and has resulted in the improvement of image quality and computerised image analysis. When integrated with patients' clinical information, it enriches the analysis of a case. These images come in different modalities, a few of which are listed and given in the table below.

**Table 2.1. Medical images and their applications [17].**

Medical Image Modality	Application
X-rays	Breast, lungs, bones
$\gamma$ -rays	Brain, heart
MRI (Medical Resonance Imaging)	Soft tissues
US (Ultra Sound)	Foetus

#### 2.4.1.1 *X-rays*

X-rays are two-dimensional images that result from the attenuation of x-rays in tissues, as depicted in Figure 2.4.



**Figure 2.4. Schematic representation of an x-ray [17].**

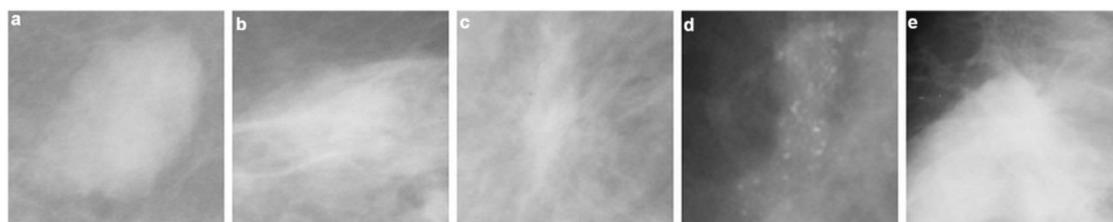
#### 2.4.1.2 *Mammograms*

X-rays of the breast are known as mammograms. Mammograms are offered as a regular breast-screening tool for women above the age of 40, as this has been shown to reduce the breast cancer mortality rate. In mammograms, signs of cancer are established by the presence of micro classifications (MCs) or masses. MCs are mostly associated with non-invasive in situ cancers, while masses are indicative of invasive malignancy [11].

The breast imaging-reporting and data system (BIRADS) describes masses, depending on their size, shape, margin and density, as [18]:

1. Circumscribed masses (CIRC): These are the most commonly found abnormalities in mammograms. They are dense and have circumscribed margins, whereas malignant masses are greatly dense with irregular borders.
2. Ill-defined masses (MISC): These masses have poor structure and are mostly malignant. Usually they are surrounded by fine tendrils.
3. Speculated masses (SPIC): These masses have an irregular appearance with speculation.
4. Calcification (CALC): These can be defined as benign or malignant, depending on their size, shape and distribution pattern. Their size may vary from minute to 3 mm in diameter.
5. Architectural distortion (ARCH): These are speculated masses without a central dense mass.

Examples of these normalities from the Mammographic Image Analysis Society (MIAS) dataset [19] are shown in Figure 2.5.



**Figure 2.5. Samples of abnormalities in the MIAS database: (a) circumscribed mass (CIRC) mdb015, (b) ill-defined mass (MISC) mdb032, (c) speculated masses (SPIC) mdb145, (d) calcification (CALC) mdb241, and (e) architectural distortion (ARCH) mdb115.**

Breast cancer originates mostly in the ductal system, and initially appears as micro-calcifications in the ducts or lobules, in mammograms. Afterwards they become invasive and appear as masses. Detectable masses are in the range of 0.5-2 cm in diameter, whereas diameters of less than 5 mm are hardly visible. Masses can be benign (round or oval with sharp boundaries) or malignant (irregular in shape with fuzzy boundaries). Detection of different types of abnormalities in mammograms is an active field of research. Various image processing techniques have been applied to distinguish among these masses [11].

Mammograms are produced in two projections (views): the craniocaudal (CC) and mediolateral oblique (MLO) projections. The breast is compressed diagonally, and x-rays run from the upper medial to lower lateral, with an average angle between these two views of 45 degrees. MLO views are considered better, as they show the upper outer quadrant and axilla more clearly, but CC views project the glandular disk more clearly. Masses appear

brighter than the tissue surrounding them, but are difficult to detect in cases where they are surrounded by dense tissues [11].

#### 2.4.1.3 *Chest X-rays (radiographs)*

Chest X-rays are used to diagnose the medical condition of the chest and nearby structures. Chest radiography is the result of an application of a radiation dose at the rate of 0.02 mSv (2 mrem) for the posterior-anterior and 0.08 mSv (8 mrem) from a latero-lateral view [20]. A sample chest x-ray is given in Figure 2.6.

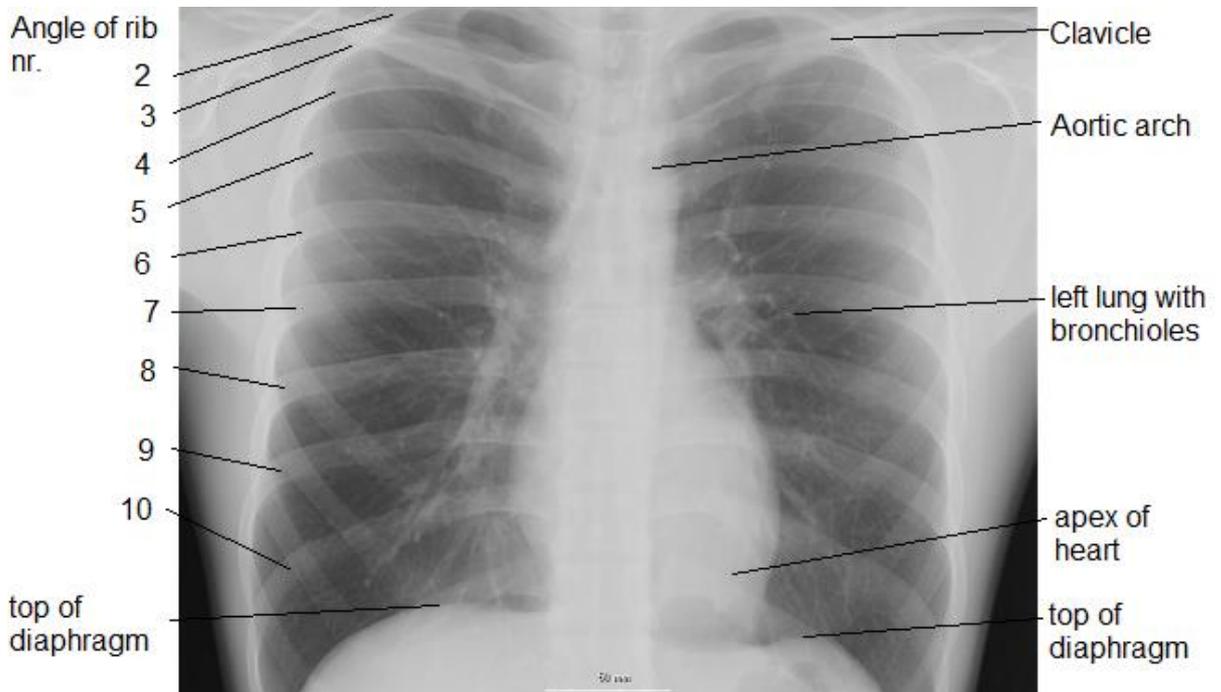


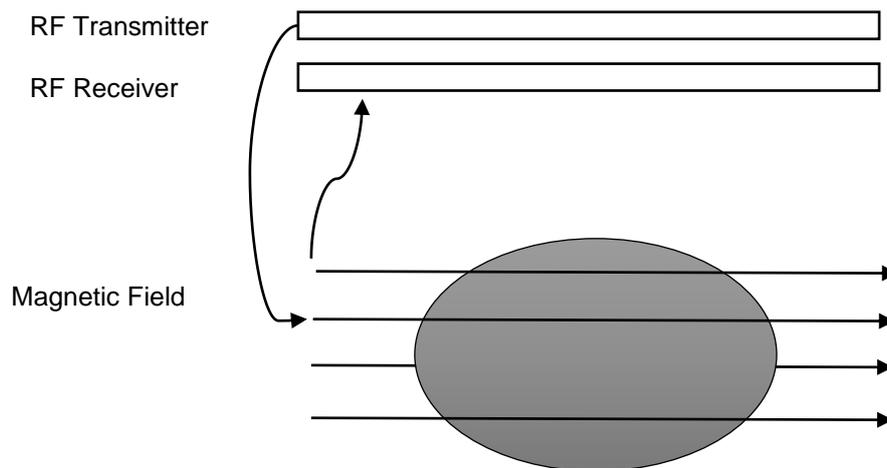
Figure 2.6. Chest radiograph with common landmarks labelled [21].

#### 2.4.1.4 *Magnetic Resonance Imaging (MRI)*

MRI depicts the water molecules present in the tissues. These molecules are randomly oriented, but when exposed to a magnetic field they turn themselves in the direction of the field and spin at a particular angular frequency called Larmour [17]. Here, supplementary energy is produced by a radio frequency

(RF) transmitter, tuned to the resonance frequency of water molecules producing short energy bursts. As a result, molecules align to the magnetic field. When returning to their original orientation, these molecules release energy as RF signals [17]. A schematic representation is given in Figure 2.7 below.

An MRI offers highly sensitive medical images specifically utilising dynamic limited contrast enhanced (DCE) acquisition, which enhances breast cancer detection and serves as imaging bio markers to guide clinical decision-making and avoid unnecessary biopsies.



**Figure 2.7. Schematic representation of MRI image acquisition.**

#### 2.4.1.4.1 TNM (Tumour, Node, Metastasis) system

Cancer stages in MRI are described using the TNM system, which is described below [22]:

## Clinical Decision Support Systems: State of the Art Systems and Methods

- Tumour (T): 'T' accompanied by a letter or number (from 0-4) representing the size of the tumour.
- Node (N): 'N' accompanied by a number, representing the expanding of lymph nodes. These lymph nodes may be located under the arm, around the collarbone and under the breastbone (regional lymph nodes). Lymph nodes that are expanded further away from these locations are called distant lymph nodes.
- Metastasis (M): Where and how much of the tumour has spread to other parts of the body.

Based on these systems, cancer stages are grouped as provided in the table below:

**Table 2.2. Cancer stages in MRI [22].**

Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
	T0	N1mi	M0
Stage IB	T1	N1mi	M0
	T0	N1	M0
Stage IIA	T1	N1	M0
	T2	N0	M0
	T2	N1	M0
Stage IIB	T3	N0	M0
	T0	N2	M0
	T1	N2	M0
Stage IIIA	T2	N2	M0

	T3	N1	M0
	T3	N2	M0
	T4	N0	M0
	T4	N1	
Stage IIIB	T4	N2	
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

## 2.4.2 Image Pre-processing Techniques

Loss of information can result from the introduction of artefacts by components of the image-capturing device. It is important to identify and erase the outer effects of these artefacts, as they can hamper the diagnosis procedure [23]. Pre-processing of medical images includes image orientation, label and artefact removal, image enhancement and de-noising. Extensive research has been conducted to solve these issues. Some of the research may only focus on noise removal algorithms, while others will apply a method that will suppress noise and enhance image features simultaneously. Image enhancement techniques must keep a trade-off between local contrast and overall image intensity, since increasing image intensity may also increase noise. [23]. Some of the frequently used techniques for medical enhancements are discussed below.

### 2.4.2.1 *Histogram Equalisation (HE)*

This is a conventional technique to adjust image intensities, which applies a non-linear, monotonic mapping of intensity values of the pixels so that the transformed image has a uniform distribution of intensity values. Let  $I$  be the image with pixel intensity values ranging from 0 to  $L-1$ , where  $L$  is 256; then the probability of occurrence of intensity value  $r_k$  can be written as [24]:

$$p_r(r_k) = \frac{\text{number of pixels with intensity } k}{\text{total number of pixels in the image}} \quad k: 0 \text{ to } L - 1 \quad (2.1)$$

The Cumulative Distribution Function corresponding to  $p_r$  is given as:

$$cdf(k) = \sum_{i=0}^k p_r(r_k) \quad (2.2)$$

Thus, the histogram equalisation acts as the image transformation so that the transformed image is obtained by mapping each pixel with intensity  $r_k$  to corresponding level  $s_k$  given as:

$$s_k = (L - 1) * cdf(k) \quad (2.3)$$

This transformation is called histogram equalisation.

#### 2.4.2.2 ***Contrast Limited Adaptive Histogram Equalisation (CLAHE)***

CLAHE applies histogram equalisation (HE) to the sub-regions of an image by first dividing an image into contextual blocks (tiles). It then creates a histogram for each block using a specific number of bins and clips the histogram at a certain threshold. It maps each region according to the new histogram results. Finally, it interpolates gray-level mapping to reconstruct the final CLAHE image [25].

This approach is the most widely used technique for enhancing medical images and is based on histogram equalisation.

#### 2.4.2.3 ***(HM-LCE) Histogram Modified Local Contrast Enhancement***

Sundaram et al. proposed a modified histogram equalisation technique, named Histogram Modified Local Contrast Enhancement (HM-LCE) [26]. This technique allows control of the level of contrast to an image by tuning the value

of contrast parameter where, in case of classical HE, there is no such control provided. It controls the level of contrast such that the enhanced image histogram is uniformly distributed and the difference between the input and the enhanced image histogram is minimal. The method works in two stages: first it modifies the histogram using enhancement parameters, and then applies local contrast enhancement to the histogram's modified image.

#### 2.4.2.4 *Local Contract Modification (LCM)\_CLAHE*

Shelda et al. suggested another variation of CLAHE in [209] for image enhancement. This technique first applies Local Contract Modification (LCM) to the image  $I$  given as:

$$T = E. \frac{\mu_M}{\sigma} \quad (2.4)$$

$$I_h = T * (I - \mu_m) + \mu_m \quad (2.5)$$

Where  $I$  and  $I_h$  are original and LCM enhance histogram images, respectively.  $\mu_M$  is the global mean,  $\mu_m$  and  $\sigma$  are the mean and standard deviation of local window of size  $m \times m$ . Next it applies CLAHE to the modified image.

#### 2.4.2.5 *Unsharped Masking (UM)*

It is a method that increases the contrast of an image,  $I$ , by using a blurring mask of the original image, given as:

$$I_{h_i}' = I_i + f_u(I_i - s_i(\sigma)) \quad (2.6)$$

Where  $I_{h_i}'$  is the processed pixel,  $I_i$  is the original pixel at  $i^{th}$  location,  $s_i(\sigma)$  is the smoothed image pixel value and  $f_u$  is the function which determines level of

enhancement. Blurring is performed using Gaussian smoothing where the scale of smoothing function defines the limit of contrast enhancement. UM is capable of enhancing the high frequency component of an image, which is why it is specifically used for micro calcification enhancement.

Siddharth et al. [27] proposed a modified UM scheme combined with nonlinear enhancement for mammogram images. It first segments the image in low-detail, medium-detail and high-detail regions corresponding to low, medium and high frequencies, respectively. Then it convolves the image as:

$$I_{h_i}'(x, y) = I(x, y) + \beta(x, y)g(x, y) \quad (2.7)$$

where  $I_{h_i}'$  is the final enhanced image,  $I(x, y)$  is the original image and  $g(x, y)$  is the output of the linear high pass filter at  $(x, y)$  location and  $\beta(x, y)$  is the constant parameter for this method, which can adopt a value within the range of (0-1).

The image is simultaneously processed through UM and the nonlinear enhancement technique. After that both resultant images are linearly combined as single enhanced image. The method is useful to detect masses from the images since it enhances mass borders.

#### 2.4.2.6 *Adaptive Contrast Enhancement*

Sundaram et al. [36] proposed an adaptive contrast enhancement technique for image enhancement. This technique first calculates the min/max/average intensities from the local window (of size  $m \times n$ ) of the image.

The window size remains fixed for the entire image. The new adapted intensity values are then calculated using the conditional propagation scheme, given as:

$$(2.10) \quad \begin{cases} lmin_{m,n} = (1 - C) * lmin_{m,n} + C \times lmin_{m-1,n} & \text{if } lmin_{m-1,n} < lmin_{m,n} \\ lmax_{m,n} = (1 - C) * lmax_{m,n} + C \times lmax_{m-1,n} & \text{if } lmax_{m-1,n} < lmax_{m,n} \\ lavg_{m,n} = (1 - C) * lavg_{m,n} + C \times lavg_{m-1,n} & \text{otherwise} \end{cases}$$

Where  $C$  is the conductivity factor ranging from 0 to 1 and  $lavg$  is the local average map initialised by the original image values.  $lmin$  and  $lmax$  are local min/max maps. Next a transfer function is applied to stretch local image intensities. The transfer function is given as [36]:

$$I_h = \begin{cases} \omega_0 - \sqrt{\omega_0^2 - I^2} & \text{if } x \leq \omega_0 \\ \omega_0 + \sqrt{(255 - \omega_0)^2 - (255 - I)^2} & \text{otherwise} \end{cases} \quad (2.11)$$

Where  $I$  is the original input image range,  $\omega_0$  is a fixed value and  $I_h$  is the range of output image.

### 2.4.3 Segmentation Techniques

Segmentation may be conducted manually, or it can be automatic via application of state of the art techniques. Here we present some of the popular medical image segmentation methods. Further detail and comparison is provided in chapter 3.

#### 2.4.3.1 Thresholding

It is the simplest approach to segmentation in which one chooses a range of grey level values that represent an object to segment. The pixels in that range

are retained while the rest of them are set to zero. For quantitative image analysis, it is required that the threshold is chosen objectively, e.g. maximum pixel value in the image or some values based on image histogram analysis [42].

### 2.4.3.2 *Region Growing*

The method starts with an interactive or automatic selection of the range of grey level values to represent a region or object of interest, then choosing the pixel to be part of a region of interest called the seed point. Consequently, an outline is automatically drawn around the set of pixels.

The methods which rely on gray levels are subject to gray level inhomogeneity and the partial volume effect, meaning that some tissue may differ in their gray scale around the image or medical image pixel, while the voxel may be large and be comprised of more than one type of object, or there is a possibility that the image is blurred and the boundaries are not well defined [42].

### 2.4.3.3 *Active Contours and Snakes*

This is computational segmentation techniques based on complex mathematics, which involves using an outline or contour for representing the region of interest. The outline, initially, starts at a certain position and then calculation is performed to drive the contour towards points in the image, which satisfies predefined conditions based on known properties. The method requires understanding of the data and the selection of suitable parameter values. Additional conditions can be applied to develop contours around the

realistic shape. Among the most common variants of this method is the active shape model [42].

#### 2.4.4 Shape Modelling and Pattern Recognition Techniques.

Some of the pattern recognition techniques, which aim to extract useful texture and intensity information (features) from the image, are given below.

##### 2.4.4.1 Geometric and Shape Features

Geometric and shape features are commonly used for detecting the malignancy of masses. They are comprised of the geometric and morphological aspects of the mass. There is a number of them but the most commonly used features are described in the Table 2.3. below:

**Table 2.3. Common shape features for masses/nodules in medical images.**

Measure	Expression
Mass area A	$A= R $ where R is the number of pixels inside the mass region.
Perimeter P	Number of pixels around the boundary of a mass.
Compactness C	It is a measure of mass contour complexity. The higher the compactness the rougher the mass contour is: $C = \frac{P^2}{4\pi A} \quad (17)$
Normalised radial length	Euclidean distance from mass centre to each boundary coordinates normalised by dividing it by maximum radial length.
Minimum and maximum axis	Shortest path connecting two points on a mass boundary but passing through centre is the minimum axis. The longest path connecting two points through a mass boundary is called the maximum axis.
Eccentricity	It is the length of region of interest (ROI)

Circularity	It is a measure of similarity of the mass with respect to a circle. [43]
Convexity	Ratio of the perimeter of convex hull by perimeter of the original contour. [44]
Dispersion	Measure of irregularity of the mass. [43]
Equivalent diameter	Diameter of the circle which has the same area as ROI. [44]
Solidity	A measure of how convex or concave the mass is. [44]

---

---

Other measures are roughness, orientation, statistical mean, variance and moments of the ROI. Delogu et al. [45] extracted 16 different shape features, such as the mass perimeter, mass circularity, mean of the normalised radial length etc., from segmented masses and studied the different combinations of these features. The features were further selected using Feature Discriminating Power and Linear Correlation Interplay techniques and fed to a Multi-Layer Perceptron Neural Network for classification. The method achieved a classification accuracy of 97.8%.

### 2.4.4.2 *Texture Features*

Image texture is a set of matrices, used to define the spatial arrangement of colour and intensities inside an image [46]. There are multiple approaches that can define texture features. Some common approaches are structural and statistical feature extraction techniques, as described below.

#### 2.4.4.2.1 Grey-level first-order statistics (GLFOS)

These features are calculated considering individual pixel values, giving no weight to neighbourhood pixels. These features are the analysis of grey level distribution in the image / calculated as [47]:

$$\text{Minimum Intensity } I_{min} = \text{Min}\{I(x, y)\} \quad (2.18)$$

$$\text{Maximum Intensity } I_{max} = \text{Max}\{I(x, y)\} \quad (2.19)$$

$$\text{Standard Deviation } \sigma = \frac{1}{mn - 1} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y) - \text{Mean})^2 \quad (2.20)$$

$$\text{Where Mean} = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y)) \quad (2.21)$$

$$\text{Standard Deviation} = \sqrt{\sigma} \quad (2.22)$$

$$\text{Skewness} = \frac{1}{mn(\sigma^3)} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y) - \text{Mean})^3 \quad (2.23)$$

$$\text{Kurtosis} = \frac{1}{mn - 1(\sigma^4)} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y) - \text{Mean})^4 \quad (2.24)$$

$$\text{Modified Standard Deviation} \quad (2.25)$$

$$= \sqrt{\frac{1}{mn - 1} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y) - \text{Mean})^2 P(I(i, j))}$$

$$\text{Modified Skew} = \frac{1}{mn(\sigma^3)} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (I(x, y) - \text{Mean})^3 \quad (2.26)$$

#### 2.4.4.2.2 Grey Level Co-Occurrence Matrix (GLCM)

One of the common statistical approaches is the Grey Level Co-Occurrence Matrix (GLCM). It extracts second order statistical features from the image. GLCM is a matrix with number of rows and columns equal to the grey levels  $G$  in the image. The matrix elements  $P(i, j | \Delta x, \Delta y)$  are the measure of the relative frequency with which two pixels at the distance  $(\Delta x, \Delta y)$  from each other appear within a neighbourhood. Here  $i, j$  denotes the intensities of the pixels. The elements of the GLCM matrix can also be represented as  $P(i, j | d, \theta)$ ; the

element suggests statistical probability that there is a change in grey level between  $i$ th and  $j$ th at a particular distance  $d$  and at an angle  $\theta$ [47].

### 2.4.4.2.3 Laws Texture Energy Measure

This approach to extract features applies local masks convolution kernels to the image. A set of nine 5X5 convolution masks are used to compute texture energy. The result is stored in a vector of 9 numbers for each pixel in the image [49].

Karahaliou et al. [50] found that the texture properties of the area surrounding the micro-calcification were significant when detecting malignancies in mammograms. This approach considered features such as the Grey-Level First order statistics, Grey Level Co-occurrence Matrices and Laws Texture Energy Measures, which were extracted from the Surrounding Tissue of Regions Of Interest (ST-ROI). A redundant Discrete Wavelet Transform (DWT) was then applied to the image, and the wavelet coefficients' first order statistics and co-occurrence matrices for ST-ROI were then used as features. This combination of features resulted in an  $A_z$  performance of 0.9989.

### 2.4.4.3 *Multi-resolution Image Analysis*

It is a powerful tool to analyse images which are comprised of a broad range of algorithms. Multi-resolution methods are of great interest to researchers in image processing, analysis, biological and other fields. Some of them are briefly described below.

### 2.4.4.3.1 Wavelet Features

The wavelet transform is used to detect features which may have been left undetected in other domains [51][217]. Cristiane et al. applied Daubechies' wavelets transform [53] to two-dimensional images by first decomposing the image into four levels, as demonstrated by Mallat [52]. This decomposition was the result of convolving the image with low and high pass wavelet band filters. These set of resultant coefficients at each level are given as:

- Low level frequency coefficients  $A_2^{dj} f$
- Vertical high level frequency coefficients  $D_2^{1j} f$
- Horizontal high level frequency coefficients  $D_2^{2j} f$
- High frequency coefficients in both directions  $D_2^{3j} f$

After the decomposition, only low level frequency coefficients  $A_2^{dj} f$  were fed to the classifier since they demonstrated the capability to enhance classification accuracy among multiple abnormality types as well as to detect abnormal cases in mammograms.

### 2.4.4.3.2 Curvelets

Curvelets were proposed by Candes and Donoho [54] as generalisations of wavelet transform. These are employed to represent images at different scales and angles. They represent a curve as a superposition of functions of various lengths and widths by obeying the scaling parabolic law as  $width \cong length^2$ . They represent edges and curves more efficiently compared to wavelets [54].

Eltoukhy et al. [55] applied curvelet transform to ROI to convert images to multi-scale decomposition levels. Further, a number of the largest coefficients were selected at each scale and combined as a feature vector. These features perform well to distinguish among different classes of abnormalities. The abnormality classes are micro-calcification, speculated mass, circumscribed mass, ill-defined mass (misclassified), architectural distortion, asymmetry and normal tissues. The accuracy achieved was 98.59%.

### 2.4.4.3.3 Contourlet

Do and Vetterli proposed contourlet transform in 2002, as an extension of the curvelet transform, to detect curves and to provide direction and anisotropy [56]. Contourlet transform can depict image contours efficiently using a minimum number of coefficients.

In the case of contourlet transform, multi-scale decomposition is performed using a Laplacian pyramid; further, a directional filter bank is applied to perform directional decomposition. It is suitable for applications which involve more contours [56].

## 2.4.5 Classification Techniques

Machine learning plays an important role in CDSS for clinical data analysis. Modern medicine relies on medical images, which contain complex patterns that are not easy to detect or discriminate by the naked eye. There is thus a need for a learning paradigm that can understand the useful information in the image, interpret it and help make decisions on the basis of this information. Medical images come in different modalities, and there is an increasing need to

understand, analyse and manipulate these different modalities; thus, the importance of machine learning is growing day by day.

Machine learning systems can change their behaviours based on their experiences. They extract a general model from the dataset and use training data to elucidate patterns within the dataset. The dataset can be in the form of a one-dimensional record or two- or three-dimensional image (either a region of interest or the entire image). Machine learning methods can be used in multiple stages of a CDSS, depending on the learning paradigm. Differences come from the information available in the training data and the nature of the model that needs to be extracted from the dataset [57]. In *supervised learning* techniques, the data takes the form of input instances and corresponding labels. If these labels are discrete, the learning is called a classification problem, while if they are continuous, it is considered a regression task [57]. If the training data does not have labels attached, the learning is *unsupervised* [57]. If some instances have labels while others do not, the unlabelled data cannot be discarded due to possible useful information attached to it, and the learning must be *semi-supervised* [57]. *Active learning* occurs when unlabelled data instances are actively queried by the user [57]. In general, a machine learning algorithm is applied to the training set multiple times to depict the model, which requires a great deal of computer memory. An *online learning* algorithm retains the training data in memory for a limited period of time, reducing memory usage [57]. Machine learning integrates techniques from a range of fields, including statistics, bio-medicine and pattern recognition [57].

#### 2.4.5.1 *Support Vector Machine (SVM)*

SVM proposed by Vapnik [58] is a supervised learning classification technique with high generalisation capability. Given a two-class problem (normal/abnormal class) cancer detection, it takes a set of features as input data and predicts to which possible class each feature vector belongs. For feature vectors of training set (of size  $K$ ) as  $U = \sum_{i=1}^K u_i$ , the SVM classifier draws a hyper plane, which separates them into two distinct classes: abnormal (class label being 1) and normal (class label being 0). The hyper plane is given as below:

$$g(U) = w^T U + w_0 = 0 \quad (2.28)$$

Here,  $w$  is a normal vector to the hyper plane. SVM is first trained to generate a model that can assign a specific class to a new unknown feature vector. In case features are not linearly separable, they are transformed to higher dimensional space using some kernel methods. We selected polynomial kernel for this purpose, which is given below.

##### 2.4.5.1.1 Linear Kernel

The linear kernel is given as follows:

$$k(U, U') = U^T U' + c \quad (2.29)$$

Where  $c$  is the optimal constant.

##### 2.4.5.1.2 Polynomial Kernel

The polynomial kernel with degree  $d$  can be written as:

$$k(U, U') = (\gamma U^T U' + c)^d \quad (2.30)$$

Where  $c$  is the adjustable constant and  $d$  is the degree of the polynomial.

#### 2.4.5.1.3 Gaussian, Radial Basis Function (RBF) Kernel

$$k(U, U') = \exp(-\gamma \|U - U'\|^2) \quad \gamma > 0 \quad (2.31)$$

Where  $\gamma$  is a positive parameter to control the radius.

#### 2.4.5.1.4 Hyperbolic Tangent (Sigmoid) Kernel or Multi-Layer Perceptron (MLP)

It was originally derived from neural networks as:

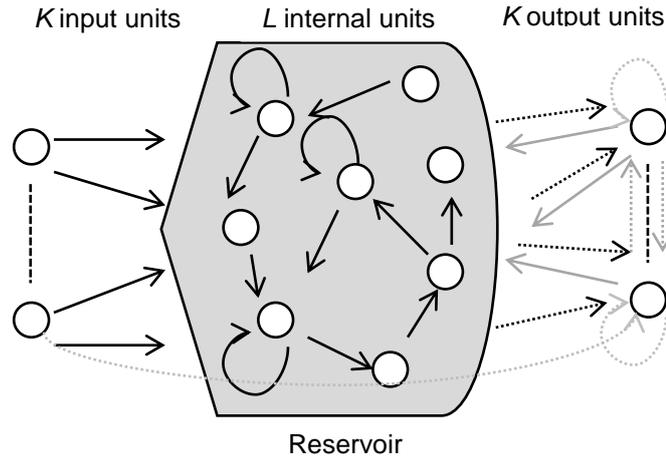
$$k(U, U') = \tanh(\gamma U^T U' + c) \quad \gamma > 0 \quad (2.32)$$

Where  $\gamma$  and  $c$  are adjustable constants.

#### 2.4.5.2 *Echo State Networks (ESN)*

Echo State Networks are supervised learning Recurrent Neural Networks where hidden layers are sparsely connected (almost 1% connectivity). In this network, connectivity and weights are randomly assigned and fixed [59].

ESN works by first generating a large, fixed size reservoir of RNN from input signals, which further generates a nonlinear response signal. The output signal is produced as a linear combination of these response signals [59]. An ESN with  $K$  input network units  $u(n) = (u_1(n), u_2(n), \dots, u_K(n))^T$ ,  $L$  internal units  $x(n) = (x_1(n), x_2(n), \dots, x_L(n))^T$  and  $K$  output units  $y(n) = (y_1(n), y_2(n), \dots, y_K(n))^T$  with activation of the units at a time step  $n$  can be modelled as seen in Figure 2.8.



**Figure 2.8. Basic ESN architecture. Dashed lines are optional connections [8].**

The units within ESN are connected through real valued connection weights given as [8]: For input units of size  $L \times K$

$$W^{in} = (w_{ij}^{in}) \quad (2.33)$$

For internal units:  $L \times L$  matrix

$$W = (w_{ij}) \quad (2.34)$$

For output units:  $K \times (K+L+K)$  matrix

$$W^{out} = (w_{ij}^{out}) \quad (2.35)$$

For connection projected back to internal units:  $L \times K$  matrix is given as:

$$W^{back} = (w_{ij}^{back}) \quad (2.36)$$

Through the learning process of ESN the activation of the internal units are updated as:

$$x(n+1) = f(W^{in}u(n+1) + Wx(n) + W^{back}y(n)) \quad (2.37)$$

Where  $f = (f_1 \dots \dots \dots f_L)$  is internal unit output function {mostly a sigmoid function tanh} at time step  $n$ . The output is calculated as:

$$y(n+1) = f^{out}(W^{out}(u(n+1), x(n+1), y(n))) \quad (2.38)$$

Where  $f^{out} = (f_1^{out} \dots \dots \dots f_K^{out})$  are output functions and are combinations of input, internal and previous output vectors at time step  $n$ . [8]

### 2.4.5.3 *Extreme Learning Machine (ELM)*

The ELM was initially suggested as a single-hidden-layer feed-forward neural network (SLFN) in which there was a hidden layer feature (mapping tuning is optional). In the ELM, the hidden layer neurons are randomly generated and are independent of one another as well as of the training dataset [25]. Given the  $K$  training samples  $(u_i, y_i)_{i=1}^k \in R^N \times R^1$ , the output of an ELM, with  $\tilde{N}$  hidden nodes, can be written as follows:

$$y_i = \sum_{j=1}^{\tilde{N}} \beta_j h_j(u_i) = \sum_{j=1}^{\tilde{N}} \beta_j h(u_i; a_j, b_j) \quad (2.39)$$

Where  $a_j \in [a_1, a_2 \dots \dots, a_n]^T$  and  $b_j$  are learning parameters for the  $j^{th}$  hidden neuron, and  $\beta_j \in [\beta_1, \dots \dots, \beta_{\tilde{N}}]^T$  is the weight vector between the hidden and output layer;  $h(u_i; a_j, b_j)$  is the activation function, which can be a non-linear piecewise continuous function, e.g. Sigmoid, Fourier, Hardlimit, Gaussian and Multiquadratics. ELM structure can be depicted as in Figure 2.9 as below.

$K$  input nodes/units       $\tilde{N}$  Random Hidden Neurons  
 $u = (u_1, u_2, \dots, u_K)^T$        $h_i(u) = f(a_j b_j, u_i)$

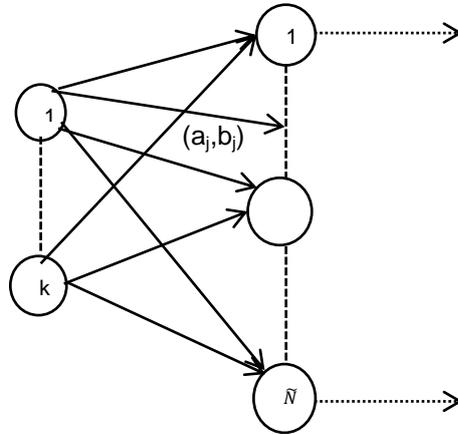


Figure 2.9. Extreme Learning Machine [24].

## 2.5 CDSS Evaluation Methods

The aforementioned algorithms must be evaluated for performance, as limited data availability makes correct segmentation or classification difficult [81]. Evaluation methods include hold-out sampling, bootstrap sampling, cross validation and leave-one-out [1].

In hold-out sampling, the data is divided into a training and test set (generally following 2/3 and 1/3 ratios, respectively). The algorithm is trained on the training set, then applied to the test set for prediction. These predicted values are then compared to the real output values to calculate the accuracy [81]. In cross-validation, the dataset is divided into several subsets of approximately equal size, the algorithm is applied to these subsets iteratively with the accuracy calculated for each, then the accuracy is averaged over all the subsets. An extension to cross-validation, the leave-one-out method, splits the

data into  $n$  subsets, with  $n$  being the size of the original dataset. The algorithm is trained and tested in  $n$  iterations when, in each iteration,  $n-1$  instances are used as part of training and the remaining instance is used for testing. Bootstrapping takes  $n$  random instances from the dataset of size  $n$  for training, and the remaining ones are used for testing [81]. These methods are discussed in detail in chapter 3.

### **2.6 The Role of CDSS in Patient Management**

It is expected that the role of CDSS will expand from providing a second opinion to the radiologist to carrying out the primary analysis of patient data. By joining medical images with other relevant information from a patient's history, a CDSS can help to assess a patient's condition. Tools from the Quantitative Imaging Biomarkers Alliance (QIBA) are effective novel tools that help with clinical decision-making and patient management by combining the efforts of researchers in the CDSS field and quantitative imaging systems. These systems will help to extract information about morphology, function, molecular structure and other characteristics in images of different modalities [1].

### **2.7 The Role of CDSS in Prognostic Tasks**

The main task of CDSS is being extended from discriminating among malignant or benign cases to prognostic estimation, which comprises a diagnosis of different cancer types as well as their stages. Joined with clinical information, this can help to determine the next step in patient management via computational analysis of histo-pathological data, DNA array autoradiographs and the morphological structure of tumours [1].

## **2.8 Framework for Development of CDSS**

Designing a CDSS requires several steps. At first, data is collected in the form of clinical information and medical images. Then, the data is pre-processed to erase artefacts, so that relevant features can be extracted from this pre-processed dataset. Clustering or classification of the features is then carried out based on similarity measures, and finally, the classification is interpreted with respect to the problem of interest.

In the following chapter, the overall framework for designing CDSS applications is presented, where the recent methods applied in each step are discussed with respect to their potential merits and demerits.

## **Chapter 3 - The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications**

In the previous chapter, a number of recent techniques for developing CDSSs have been highlighted. We extended this research further to develop a novel general framework which employs state of the art machine learning, image processing as well as optimization techniques to enhance the performance of CDSSs.

In this chapter, the conventional framework for developing CDSSs is discussed which has the ability to analyse the most widely used medical datasets. Although the proposed framework is being developed keeping in view the medical dataset analysis specifically, it is general enough to be adapted to the needs of data analysis from different domains of science.

Section 3.2 describes the general framework which consists of acquiring the dataset in the beginning (section 3.2.1) which is followed by pre-processing (section 3.2.2) as it is vital for medical images to be analysed. After pre-processing, features are extracted (from medical images) (discussed in section 3.2.3) and next a set of important features are selected (section 3.2.4) for further classification model development (section 3.2.5).

Overall system performance can be authenticated by the implication of different evaluation techniques discussed in section 3.2.6.4. Further statistical analysis may be applied to compare the performance of two or more CDSSs' performance (section 3.2.7). Finally, section 3.3 summarizes the overall system with the conclusion and discussion.

### 3.1 The Framework for CDSS Design

The overall proposed framework can be depicted as in Fig 3.1 below:

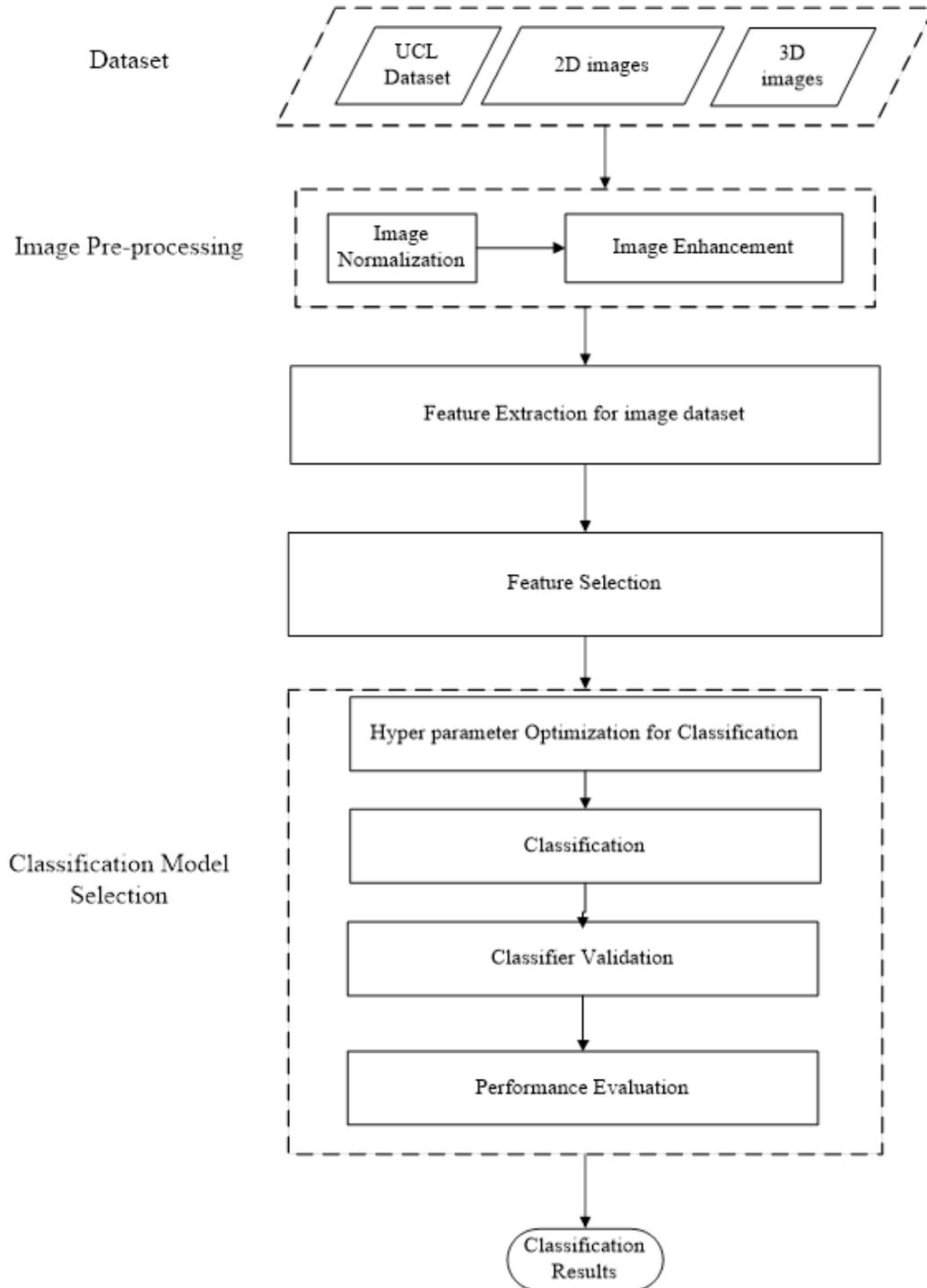


Figure 3.1. Overall framework for designing CDSS.

### **3.1.1 Dataset**

A typical dataset is a collection of a number of separate elements where each one comprises inputs as independent variables (representation of variation in dataset used to make predications) and dependent variables as output (whose variation has to be studied) [76].

Our research and experiments on medical diagnosis encompasses the benchmarked datasets retrieved from the internet. These datasets include UCL datasets such as PIMA and BUPA [77], as well as MIAS [78] and INbreast [79] mammograms and chest radiograph images [80]. We further extended our experiments to volumetric medical images and so applied the proposed framework to the three-dimensional (3D) MRI dataset from the cancer atlas for a breast cancer TCGA-BRCA dataset provided by the cancer imaging archive (TCIA) – the Frederick National Laboratory for Cancer Research [81]. The UCL datasets do not require pre-processing but medical images must be processed before they can be used for classification tasks. The detail of these datasets is available in the coming chapters where each experiment conducted has been elaborated.

### **3.1.2 Pre-processing**

Medical images acquired from various sources differ in their modalities. Some of the images are acquired in digital form, while others are acquired as analogue images which are later digitized. In the case of three-dimensional datasets, an image consists of a series of sub-images as slices and each pixel is called a voxel [42]. These images vary in their resolution. Taking into account

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

these differences, one needs to develop a system which can work on any digitized medical image. These medical images require fine tuning before they are used for analysis purposes. This pre-processing consists of image normalization and enhancement modules.

### 3.1.2.1 *Image Normalization*

Image normalization is a process to interpolate the medical images to a fixed intensity and resolution range. Since mostly 0.2 mm pixels or larger are used in studies of medical images, the intensity range of 12 bit/pixel is considered appropriate [11].

In the case of mammograms, the breast is compressed, thus causing breast thickness to decrease. All this develops a high gradient in image intensity and can lead to inaccurate system performance. It is necessary to correct this introduced gradient factor by enhancing the image intensities [11]. Let  $I_0$  be the original image then after normalization image,  $I$  is given as:

$$I = (I_0 - Min) \frac{newMax - newMin}{Max - Min} + newMin \quad (3.1)$$

Where the original image  $I_0$  has intensity range  $(Min, Max)$  and the new image  $I$  has intensity range  $(newMin, newMax)$  [24].

### 3.1.2.2 *Image Enhancement*

Image enhancement is applied to improve image quality by brightening and sharpening its details, removing noise or increasing the contrast. The improvement attained is subjective but quantitative measures like image to noise ratio or receiver operating characteristics (ROC) may be applied to calculate enhancement [42].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

Image enhancement is categorized as spatial domain filtering and frequency domain filtering. Spatial domain refers to all pixels an image is composed of. Spatial domain filtering methods are applied to the pixels directly. The process can be expressed as:

$$I_h(x, y) = T[I(x, y)] \quad (3.2)$$

Where  $I(x, y)$  is the input image,  $I_h(x, y)$  is the processed image, and  $T$  is an operator on  $I$  defined over some neighbourhood of  $(x, y)$  [82].

### 3.1.2.2.1 Spatial Domain Filter

This is a neighbourhood filter operation where the value of a certain pixel is changed by computation that involves the values of the pixel and its neighbourhood [42]. These filters are further categorized as rank filters and convolution filters.

### 3.1.2.2.2 Rank Filtering

In rank filtering algorithms, the neighbourhood pixels are arranged in ascending order in a list, and then the new value of the pixel of interest is the required value at the rank position in the list of neighbourhood pixels. The most popular ranks are median, maximum or minimum. The median filter replaces all the pixel values in the neighbourhood by the median value in the ranked list, which helps remove noise without blurring. In the case of minimum and maximum filters, each pixel value is replaced by the smallest or largest value in the neighbourhood [42]. A more recent and complex rank filtering method, adaptive contrast enhancement techniques, is discussed in section 2.4.2.7.

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

### 3.1.2.2.3 Convolution Filtering

In the case of convolution filtering, a square kernel is defined around the pixel of interest, where each location in the kernel is assigned a numerical value (weight), and the pixel at the centre is replaced by a weighted sum of the values around it. The kernel is moved around the entire image [42]. The most common technique is the Gaussian filter, which is a non-uniform kernel method in which the coefficients are samples from a two-dimensional Gaussian function.

### 3.1.2.2.4 Hybrid Filter

A hybrid filter applies both rank filtering and convolution filtering to the image. The most common example is unsharp mask (section 2.4.2.5) which enhances the image by subtracting the weighted copy of the smoothed version of the image from the original image. The other example is fuzzy rule based image filtering, discussed in section 2.4.2.6 [42].

### 3.1.2.2.5 Frequency Domain Filtering

These methods transform an image into the frequency domain by application of multi resolution algorithms (e.g. wavelet transform, discussed in section 2.4.2.8). Further in the frequency domain, the frequencies of interest are emphasized or de-emphasized. For sharpening the image, high frequencies are emphasized while for noise reduction low frequencies are de-emphasized. The enhancement operations are generally performed by multiplication of the frequency coefficients by certain weights, and then the enhanced image is obtained by application of the inverse transformation upon the transformed image [24].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

### 3.1.2.2.6 Histogram Processing

Histograms form the basis of numerous spatial domain image processing techniques. This is the plot of the rate of occurrence (frequency) of each grey-level from 0 (black) to 255 (white) [24]. Histogram-based image enhancement is an effective and popular technique in medical imaging.

The basic technique in histogram equalization (HE) dispenses pixel values uniformly so that the enhanced image has a linear cumulative histogram. There are multiple adaptive HE-based techniques suggested in literature and these are discussed in sections 2.4.2.1-4.

### 3.1.2.2.7 Comparison of Image Enhancement Techniques

The following table summarizes the merits and demerits of difference enhancement techniques for mammograms with an aim to find out the best technique to be incorporated as part of our developed CDSSs which are discussed in the next two chapters.

**Table 3.1. Merits and demerits of difference enhancement techniques for mammograms.**

<b>Technique</b>	<b>Advantages</b>	<b>Disadvantages</b>
Histogram Processing	HE	Increases background noise.
	Simple and straightforward and fast invertible operator.	Can produce undesirable effects for low colour depth images.
	Good for bright and dark areas in medical images. Intensity values are effectively spread out.	
	CLAHE	Fails to preserve some local information.
	Performs better than HE.	
	Reduces noise, edge-showing effect.	
	No noise enhancement since pixel	

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

---

	intensity limit is user selectable. Specially developed for medical imaging.	
	HM-CLAHE	
	Provides strong contrast enhancement while preserving local information.	
Rank Filtering	UM	Contrast in darker area enhanced more than contrast in lighter area.
	Very simple method, a powerful tool to sharpen the edges of the objects even in the presence of noise.	Also enhances the noise and digitized effects.
	Considers the human visual system response while enhancing the image.	
	NLUM	
	Suppresses the background noise as compared to UM, reduces the noise sensitivity of UM algorithm.	
Hybrid Filtering	Good for detecting micro calcifications.	Less effective for images with diverse features [83].
	Can manage ambiguity and vagueness efficiently. It smoothes noise while preserving important image details like edges, visually better result.	
Frequency Domain Filtering	Makes barely seen features obvious [83].	May increase false positive rate. [83]
	It represents both frequency and time domain information. Image is converted into different sub bands, where each sub band can be handled effectively. Noise in the high frequency sub band can easily be smoothed.	

---

Some of the other image enhancement techniques found in literature will be reviewed as part of our future research endeavours as we have time and space constraints. It is also common to apply a combination of techniques to improve medical images.

### **3.1.3 Image segmentation**

This is partitioning images into non-overlapping regions so that they are homogenous in certain respects, like intensity, texture, shape etc. If  $I$  is an image, then segmentation is a process of determining image subsets  $S_k \subset I$  so that their union is the whole image  $I$ , and their intersection is an empty set [84]. Image segmentation techniques fall into three major categories as described below.

#### **3.1.3.1 *Manual segmentation***

This is the most common technique in which some image processing tools are used to trace boundary of the region of interest (ROI) which is mostly an object like mass, lesion or an organ like a breast, kidney etc. It involves an expert to outline the object boundary. State of the art segmentation tools which help in drawing the outline around the region of interest (organ or abnormality) or completing the open contour around them are 3D Slicer [86], ImageJ [87], ITK-SNAP [85], etc.

Clear instructions are required to define object boundaries of anatomical features [42] lest this technique may lead to erroneous results and be time-consuming.

#### **3.1.3.2 *Semi-Automatic and Automatic Segmentation***

Semi-automatic and automatic techniques involve human intervention to perform segmentation. A number of methods in this category are described in section 2.4.3, including thresholding, region growing and active contours and snakes.

### 3.1.3.3 *Mathematical Morphology*

It is a set of techniques used to analyse the spatial structure of shapes, using set theory, integral geometry and lattice algebra [218]. Morphological operations are mostly performed using basic erosion and dilation. These techniques are also applied to refine the resultant binary image after segmentation. It helps to remove noise and create an outline of the objects in the image [42].

It was initially suggested by Matheron and Serra, in 1964, for enhancing petrography and mineralogy. The first experiments were for the analysis of binary images, by the application of simple set operations of translation, union intersection and so on [218].

### 3.1.3.4 *Comparison of Image Segmentation Techniques*

A comparative analysis of these techniques with their merits and demerits is given in the table which is followed by the discussion.

**Table 3.2. Merits and demerits of different segmentation techniques for medical images.**

Technique Name	Merits	Demerits
Manual segmentation	It is more accurate as it is easy to ensure the segment boundary manually.	Time-consuming method and results may not be scientifically reproducible [90].
Thresholding	It is a simple and fast algorithm with minimum computational cost [89].	Sensitive to noise, choosing an inappropriate value may lead to over or under segmentation. It can cause misleading edge detection [89].
Region Growing	Provides better results compared to other segmentation techniques as it is initialized with an inner point as a seed to grow to outer regions [89].	Incorrect choice of seed may lead to inappropriate segmentation and seed selection is manual. Similarly, stopping criteria selection is a tedious job [89].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

---

Active Contours and Snakes	Focuses on refining the boundary of the objects and thus generates closed contours [88].	Computationally expensive technique, which demands setting up multiples [88].
----------------------------	--	---

---

---

Segmentation, in the case of medical images, can be categorized as organ segmentation or abnormal region of interest (ROI) segmentation [91]. In the case of organ segmentation, it partitions the image into breast region and background region. It helps limit the search for abnormalities to a limited area in the image. Sometimes the technique is preceded by pectoral muscle removal (in the case of mammography) [91].

In the case of ROI segmentation, only the region of possible abnormality is segmented out of the whole mammogram. This segmentation can be manual, in which case information provided by radiologists is used to crop out the abnormal region. In the case of automatic segmentation, algorithms have to detect the abnormal region and crop accordingly.

Both techniques use almost similar algorithms for segmentation. We applied manual segmentation to crop the region of interest from mammogram images as well as from chest radiographs and breast MRIs. Detailed study of different segmentation techniques and their comparison is part of our future plan.

### 3.1.4 Feature Extraction

Feature extraction is an essential step for medical image analysis which is performed after the pre-processing stage. Human beings can derive a story from a picture based on the background knowledge that they have. Similarly, a computer can extract semantic knowledge from an image and build a model for

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

it. It is a process to retrieve highly significant information from a raw image so that extracted information defines the object of interest uniquely and precisely [92].

Feature extraction can be applied as a global, block-based, or region-based technique. These features represent the relevant shape in the image pattern so that it can be used to classify this pattern. It is also considered a form of dimensionality reduction, where the aim is to extract most relevant information from the original image and thus transform the information into feature vectors which are a reduced representation of the image pattern of interest. These features can be categorized as discussed below [92].

### 3.1.4.1 *Statistical Features*

These features are the result of applying statistical analysis of pixel distribution within the image. Grey-level first-order statistics (GLFOS) and grey level co-occurrence matrix (GLCM) are the example statistical features discussed in section 2.4.4.2. They are simple and high speed techniques applied to extract useful feature information from medical images.

### 3.1.4.2 *Global Transformation and Series Expansion Features*

These features are extracted by first transforming the image into a linear combination of a simple series of well-defined functions. The coefficients of such a linear combination are compact encoding known as series expansion. These features are invariant to translation and rotation operation upon the image. Most common are Fourier series, wavelets, Harr transform and soon, discussed in chapter 2.

### 3.1.4.3 *Geometrical and Topological Features*

These features encode knowledge about geometry of the object in the image and are developed by putting together geometric elements like points, lines, curves or surfaces. These features include corner, edge, ridges, blobs etc. The shape features discussed in section 2.4.4.1 come under this category.

### 3.1.4.4 *Discussion*

Transformation of the image into these feature sets helps in classification and recognition of ROI in medical images [92]. Intensive research is performed to develop newer techniques and methods that would help extraction of resilient features from medical images and so help improve the CDSS diagnosis system. These state of the art methods are discussed in chapter 2.

Some authors combined different types of features to see their impact on CDSS diagnosis performance. Rafayah et al. [94] and F. Moayedi et al. [95] & [96] experimented with three types of texture features. The first set of features corresponds to the maxima, mean, standard deviation, energy, entropy and skewness parameters for 4th level contourlet transform; second is a set of geometric features consisting of area, orientation, and centre of mass and the third comprises energy, correlation, inertia, entropy, inverse difference moment, sum average, sum of variance, sum entropy, difference average, variance, difference entropy, and information measure of correlation, extracted from co-occurrence matrix as a feature set. These features were fed to different types of classifiers and ultimately reached accuracy of 96% maximum. F. Moayedi et al. [97] used shape and texture features extracted from the region of interest to discriminate between malignant and benign cases. It selected eccentricity,

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

circularity, compactness, circular disproportion and circular density from shape features. For texture features Reply's K function [98] and Moran and Geary's Index have been used [99]. The results were computed using ROC whose value fell below 0.90.

Wang et al. [100] extracted features after applying daubechies-6 and daubechies-12 wavelet transform on the image. The features investigated include:

1. Intensity Features: contrast, invariant moment, mean grey and gradient of ROIs, standard derivation inside ROIs, higher order moments of ROIs, mean gradient of ROI's boundary.
2. Geometric Features: circularity, compactness, sphericity, Fourier descriptor
3. Texture Features: laws texture, co-occurrence matrix texture,
4. Wavelet transform texture

The system reached the accuracy of 90.6% and reduction in false positive rate of 3.6 marks/images.

Wang et al. [102] used different combinations of important features extracted from curvilinear and grey level co-occurrence matrix (16 texture features), Gabor features and statistical features extracted from Quincunx wavelet transform (from first 4 decomposition images at even levels) and fed them to structured SVM. Domínguez and Nandi [103] used multiple shape features to

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

discriminate among different types of masses and reached sensitivity of 80%. Eltoukhy et al. [104] classified malignant and benign masses using SVM for features selected from the ROI. The feature set included curvilinear, texture, Gabor and multi resolution features. Performance was measured using ROC and reached 0.97 with the maximum accuracy of 91.4%. Rashed et al. [105] experimented with different wavelet transforms to analyse their ability to discriminate among different classes of abnormalities such as clusters, speculated lesions, circumscribed masses, and ill-defined lesions. Moussa et al. [106] used wavelet (Daubechies) coefficients as feature vectors. This approach first extracted the horizontal, vertical and diagonal detailed coefficients from the image decomposition via Daubechies wavelets. It then normalized the coefficients and calculated the energy for each feature vector. These features were further reduced depending on the level of energy required and further used for classification purposes. Eltoukhy et al. [107] decomposed the ROI into 4 levels using wavelet functions Daubechies, Symlet and Bi-orthogonal. It also applied curvelet transform to decompose ROI into 4 levels. Next it selected the 100 largest coefficients at each decomposition level and combined them into a feature vector of 400 coefficient length. This feature vector is used to classify among benign, malignant and normal cases as well as for classification among different types of abnormalities available in the MIAS dataset. The result turned out to be promising. Sakka et al. [108] & Yang et al. [109] experimented with different families of wavelets at different levels of decomposition and compared their results. According to their results Daubechies at level 4 decomposition work best. Wajid et al. the LESH feature

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

extraction technique (chapter 4) in conjunction with SVM to detect malignancy as well as to distinguish between various types of malignancies in mammogram images [210]. The technique was further extended to an experiment with an ESN classifier for mammogram images [211]. An application of the LESH feature extraction technique combined with SVM, ESN and ELM classifiers was employed on chest radiographs to distinguish between lungs with and without nodules as well as to make a distinction between normal and abnormal lung nodules [214].

Different feature extraction techniques have their specific capabilities to highlight important aspects of information within the images. In general shape features are mostly used to analyse masses. Texture features are used for both masses and micro calcification detection and the same is the case with multi resolution features. It is important to note that for micro calcification, multi resolution features perform better than any other feature extraction technique. Wavelet features also turned out to be excellent in discriminating among different types of abnormalities in mammograms. Other feature extraction methods found in literature are intensity, morphological, grey level and temporal features, etc. We applied the LESH feature extraction technique, which is described in chapter 4. It turned out to be more resilient as compared to other state of the art feature extraction techniques, as is evident from the results in chapter 4.

### **3.1.5 Feature Selection**

The quality of features is important for better classification results. Studies have shown that high separate potential is possible with few feature sets. Feature selection is critical to the whole process of developing an efficient CDSS, which consists of detecting the relevant features and discarding the irrelevant features.

Lippman declared that: “Features should contain information required to distinguish between classes, be insensitive to irrelevant variability in the input, and also be limited in number, to permit efficient computation of discriminant functions and to limit the amount of training data required” [110].

This is the process of selecting the most informative features from the current feature set. Extracted features may contain redundant and irrelevant information which can hinder the predictor performance and increase computational cost [111].

In the presence of a large number of features, classification algorithm may over-fit, which will result in performance degeneration. To avoid the curse of dimensionality, dimensionality reduction has been applied which is to choose a small subset of relevant features from the original feature set based on certain evaluation criteria. The criteria may be accuracy of the classifier, low computational cost and so on [112].

Feature selection helps to:

- Improve prediction performance of the predictor/classifier.

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

- Provide cost effective classifier
- Provide better understanding of the process to generate data

Feature Selection techniques broadly fall into 3 categories as given below [112].

### 3.1.5.1 *Filter Model*

This selects features based on general characteristics of the data with no data mining technique involved. The data characteristics can be correlation coefficients, entropy measure etc. Some of the example algorithms are Iterative Search Margin Based Algorithm (Simba), and Greedy Feature Flip Algorithm (G-flip) [113]. These are faster methods with better generalization capability as they are independent of any induction algorithm [114].

### 3.1.5.2 *Wrapper Model*

This uses predefined data mining algorithms and performance for feature selection. Some of the common examples are forward and backward selection method.

### 3.1.5.3 *Embedded Model*

To overcome the shortcomings of the two above-mentioned methods (listed in Table 3.3 below), the embedded model was proposed, which incorporates some statistical criteria into select candidate features. Secondly, it chooses the subset of the features with highest classification accuracy. Hence, this method achieves accuracy and efficiency of both models thus achieving feature selection and model fitting simultaneously [115].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

Other techniques are statistical t-test based feature selection [117], Genetic Algorithm [72], Heuristic Reduct Algorithm [118] and so on.

We experimented with the feature selection technique suggested by [119]. In this technique they selected a set of the largest coefficients among the available wavelet coefficients and compared the performance of their selection of coefficients. We applied this technique for wavelet and LESH features and got promising results. Our future endeavours include detailed study of other feature selection techniques.

### 3.1.5.4 *Comparison of the Feature Selection Techniques*

**Table 3.3. Merits and demerits of different feature selection techniques.**

Technique Name	Merits	Demerits
Filter Model	Easily scalable to high-dimensional datasets.	They give the least consideration to interaction with a classifier.
	Simple and fast computational hazards.	They are univariate methods, meaning each feature is considered separately, and dependency between features is not considered [114].
	Independent of the classification algorithms [114].	
Wrapper Model	It takes into account the relationship between feature subset and model selection.	The technique is prone to over-fitting. It has high computational cost.
	Gives weight to features' interdependencies.	
Embedded Model	It considers interaction with the classification model.	It is specific to a given learning algorithm.
	It has lower computational cost compare to wrapper model	
	wrapper methods.	

### **3.1.6 Classification Model Selection**

Automated classification of medical images is an increasingly important tool for physicians in their daily activity. It includes two tasks: investigation of the relationship between image features and the object classes and selection of image feature set which helps to distinguish unambiguously between object classes with the least effort and minimum error. Even with the best available feature set, separation of objects into classes of interest is not possible. Sometimes transition from one class to the other is gradual, in that case we cannot find well-separated classes in feature space, even if we select the optimal feature set. There is a pressing need to select an appropriate model for classification which depends on the current dataset as well as on the classifiers which have been applied.

#### **3.1.6.1 *Hyper parameter Optimization for Classification***

Given a machine learning classifier, we aim to maximize its classification performance using optimization techniques. This goal is accomplished by searching for the optimum values of the hyper-parameters in a given search space (model selection). We studied the methods of enhancing the performance of CDSS by optimising classification performance using start of the art optimisation techniques. The objective function in this can be any of the classification performance measures, naming performance accuracy, receiver operating characteristic (ROC), Area Under ROC (AUROC), sensitivity, specificity etc.

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

Given a machine learning classifier  $A$ , we aimed to find a function  $f$  that can maximise the classification performance  $P(D;f)$  for a dataset  $D = \{(u_1, y_1), \dots, (u_n, y_n)\}$  using some optimisation techniques. This algorithm says  $\mathcal{A}$  maps the dataset  $D^{(train)}$  (a selected sample from  $D$ ) to a function  $f$ , through optimising training criterion, with respect to a parameter set  $\theta$ . Further, these algorithms comprise some values for their performance boost called hyper-parameters  $\lambda$ , as when chosen appropriately for learning upon a dataset becomes  $A_\lambda$ . So that  $f = A_\lambda (D^{(train)})$  provides us with best results. Choosing an appropriate value of  $\lambda$  is significant to improve classification performance. The problem of finding out the best values of this  $\lambda$  hyper-parameter is called hyper-parameter optimisation. Here we studied the optimisation of this performance via hyper-parameter optimisation, which is of practical importance to machine learning classification [60]. We can write this optimisation problem as:

$$\lambda^{(*)} = \underset{\lambda \in \Lambda}{\operatorname{argmax}} P[(D; A_\lambda (D^{(train)})] \quad (3.3)$$

The goal of obtaining generalisation performance  $P$  is that the algorithm  $A$  can be achieved via application of cross validation. This technique splits a given dataset  $D$  into disjoint sets  $D^{(train)}$  and  $D^{(Valid)}$  and classification performance is calculated as a mean of all the performances  $P$  applied on validation sets  $D^{(Valid)}$  so that the problem is transformed into:

$$\lambda^{(*)} = \underset{\lambda \in \Lambda}{\operatorname{argmax}} \operatorname{mean}_{d \in D^{(valid)}} P [(d; A_\lambda (D^{(train)}))] \quad (3.4)$$

Where  $A_\lambda$  is trained on  $D^{(train)}$  and evaluated on  $(D^{(valid)})$  when class validation is applied to split dataset into disjoint training and validation sets [60].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

Many algorithms have been suggested in the literature to find optimum values of the hyper-parameters in order to optimise the machine learning classifiers.

These algorithms address the problem of optimisation by solving:

$$\begin{aligned}\lambda^{(*)} &\equiv \operatorname{argmax}_{\lambda \in \Lambda} \Psi(\lambda) \\ &\approx \operatorname{argmax}_{\lambda \in \{\lambda^{(1)} \dots \lambda^{(S)}\}} \Psi(\lambda) \equiv \hat{\lambda} \quad (3.5)\end{aligned}$$

The above equation expresses the hyper-parameter optimisation with respect to response function,  $\Psi$  (also called response surface) [60].

In the case of different optimisation techniques, learning algorithms and datasets;  $\Lambda$  and  $\Psi$  will be different. So the aim of the learning algorithm is to find the optimum value of  $\lambda$  from among  $S$  trial points  $\{\lambda^{(1)} \dots \lambda^{(S)}\}$  drawn from the search space  $\Lambda$  by evaluating  $\Psi(\lambda)$  for each of them and then return the value of  $\lambda^{(i)}$  that works the best as  $\hat{\lambda}$ . [60]

We summarised some well-known hyper-parameter optimisation techniques below.

### 3.1.6.2 *Manual Search*

The straightforward technique for tuning hyper-parameters is the manual search. In case of manual searching, knowledge about a problem is used to identify the region in a search space  $\Lambda$  which is more promising to develop an intuition to choose certain values of  $S$  hyper-parameters as  $\lambda = \{\lambda^{(1)} \dots \lambda^{(S)}\}$ . This search method provides better insight into the surface function  $\Psi$  and there are no technical overheads involved [60].

### 3.1.6.3 *Grid Search*

A traditional method for optimisation is grid search, which is an exhaustive search through the manually selected subset of hyper-parameter search space  $\Lambda$ . This search is guided by the performance metric, which is calculated via cross-validation on training dataset. If a learning algorithm  $A$  has  $S$  hyper-parameters  $\lambda = \{\lambda^{(1)} \dots \lambda^{(S)}\}$  (to find the optimum value of) and each one of these is adopted among a set  $K$  of selected values  $\{L_i^1 \dots L_i^K\}$ ,  $i:1..S$ , then a set of trials will be formed by combining every possible combination of these values giving  $T = \prod_{k=1}^K |L^{(k)}|$  elements for trials. This makes grid search suffer from curse of dimensionality [61]. The grid search works better in the case of low dimensional spaces, like one- or two-dimensional spaces, and provides better results than manual searching [60].

### 3.1.6.4 *Random Search*

Random search is considered better than and grid searching [60]. It identifies the promising region in the search space by first gaining basic knowledge of the problem domain. Further, it independently draws values for trial set  $\{\lambda^{(1)} \dots \lambda^{(S)}\}$  by randomly choosing a combination of values from the range of each hyper-parameter value. This method is simple, and easy to implement trivial to parallelism [60], but also employs a big advantage of efficiency in high-dimensional search space because function  $\Psi$  has low effective dimensionality here (if a function can be approximated by the other function in low-dimension space, it has low effective-dimensionality).

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

The Grid search number of wasted trials is exponential (curse of dimensionality), which turned out to be irrelevant for a particular dataset. But random search thrives at low effective dimensionality, so it searches within the relevant dimensionalities [60].

### 3.1.6.5 *Bayesian Optimisation*

It is derivative-free technique used for the global optimisation of noisy black-box functions. When applied to hyper-parameter optimisation, it first constructs a probability model  $M$  of  $\lambda^{(*)}$  by evaluating it and by considering prior information about the dataset. This model is used to select subsequent configuration of  $\lambda$  using the acquisition function  $a_M: \Lambda \rightarrow R$ . This acquisition function quantifies knowledge about  $\lambda$  by calculating a predictive distribution of model  $M$  at arbitrary configuration of  $\lambda \in \Lambda$ . The function is maximised over  $\Lambda$  to find the next value of  $\lambda$  to be evaluated. There are different acquisition functions in research, like the Sequential Model-based Algorithm Configuration (SMAC) [62, 63], SPEARMINT [64], Tree Parzen Estimator (TPE) [65] and so on.

Other than these basic optimisation techniques, there have been heuristic optimisation algorithms suggested in the literature and we will discuss some of them here.

### 3.1.6.6 *Particle Swarm Optimisation (PSO)*

PSO is a population-based stochastic global optimisation technique based on biological systems (social systems), which is the study of the collective behaviour of individuals interacting with their environment and each other (swarm intelligence) [66]. It assumes the scenario that social interaction

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

produces information which evolves over time and distance and thus helps to provide solutions to certain problems [67].

PSO has been designed in keeping with the new concept of artificial life (A-life) imitations of bird flocking, fish schooling and swarm theory. It is also related to the concept of evolution theory [68].

Kennedy et al. simulated the search pattern of a bird flock searching for food in a given area. Assuming food is available at a random location in an area, not all birds know its whereabouts, so they get closer to the food during each iteration by following the bird which is nearest to the food. In PSO, each single solution is considered as a bird called a particle in search space. All particles have their fitness value (evaluated by the fitness function that needs to be optimised) and have velocities which directs their flight. These particles fly through the search space by following optimum particles [68].

### 3.1.6.6.1 PSO Based hyper-parameter Optimisation

Solving the problem of hyper-parameter optimisation using PSO aims to find the optimum value of  $\lambda$  from the search space  $\Lambda$ , by choosing  $S$  trial points  $\{\lambda^{(1)} \dots \lambda^{(S)}\}$ , which are drawn according to swarm optimisation technique and then return the value of  $\lambda^{(i)}$  that works the best as  $\hat{\lambda}$ . [60]

Taking each  $\lambda^{(i)}$  as a particle in PSO, its fitness value is the classification performance  $P$ . Each  $\lambda^{(i)}$  keeps a record of its previous position at best fitness value given as *pbest* in the global search space. Another best fitness value position is tracked by the swarm optimiser, which is obtained by any particle in

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

the search space  $nbest$ . When a particle takes its whole search space as its topological neighbour, its best position becomes global optimum  $gbest$ . Each particle moves in the direction of its best solution (position) and ultimately particles discover the best solution (position) in the swarm. Each particle updates its own velocity and position in each iteration as:

$$v = wv + c_1 \times r_1 \times (pbest - \lambda) + c_2 \times r_2 \times (gbest - \lambda) \quad (3.6)$$

$$\lambda = \lambda + v \quad (3.7)$$

Where  $w$  is the inertia coefficient introduced by Shi and Eberhart [69]. It remains in the interval  $[0, 1]$  and balances the global exploration and local exploitation;  $c_1$  and  $c_2$  are learning rates;  $r_1$  and  $r_2$  are random numbers in the interval  $[0, 1]$ ,  $v$  is within the interval  $[-vmax, vmax]$  and  $vmax$  is a designated maximum velocity. The search terminates according to the criteria of the maximum number of iterations or designated value of the fitness reached [70, 71].

### 3.1.6.7 **Multi-Objective Genetic Algorithm (MOGA)**

Multi-Objective Optimisation (MOO) is the application of optimisation to problems having more than one objective function to be optimised simultaneously, e.g. when we have to compromise between maximising the comfort of a car while minimising its cost. These problems do not have a single solution and there is a need to optimise each objective function simultaneously while keeping a trade-off between them. There are many possible pareto-optimal solutions (where one of the objective functions cannot be improved without degrading the performance of any other objective function) for such problems where all of them are considered equally good. Converting such problems to single objective optimisation is an important research field. Many

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

methods are presented to convert multi-objective problems to single objective optimisation, one of which is called scalarisation. Evolution-based MOGA is most popular among them and is presented here.

In the case of MOGA, it has both objective and fitness function. Objective function is the optimisation condition of GA (feature of the problem domain) whereas fitness function is the measure of quality of a particular solution of satisfying that condition and assigning a corresponding real-value to that solution [75].

### 3.1.6.7.1 MOGA as Optimiser

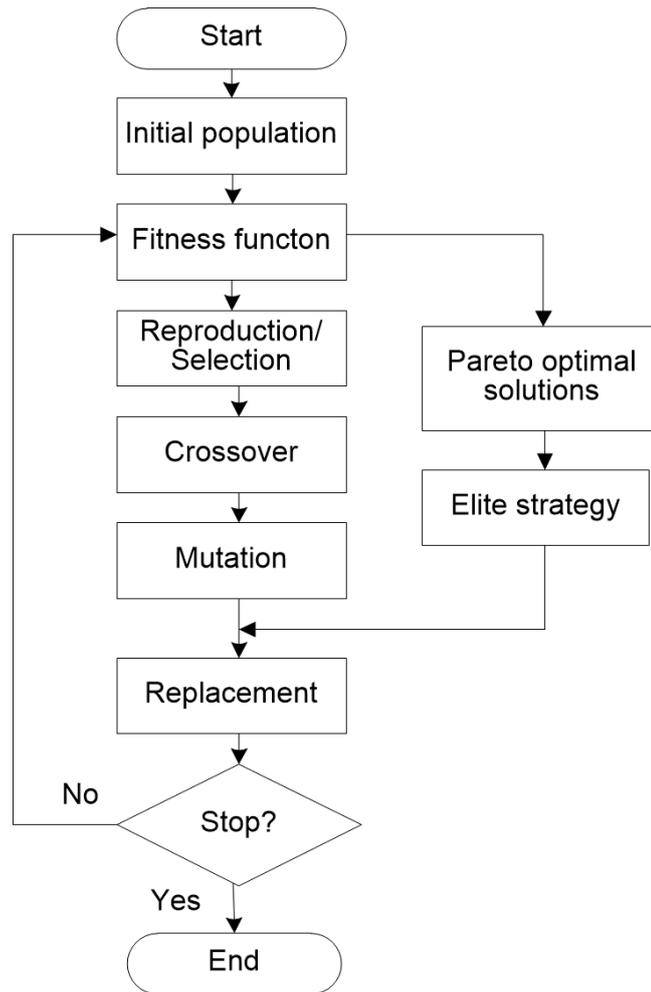
MOGA is based on the concept of pareto-dominance, which quantifies a relationship between two chromosomes. Given in a population of chromosomes of hyper-parameters, if  $\lambda^{(i)}$  is dominant  $\lambda^{(j)}$ , that means fitness of  $\lambda^{(i)}$  is better than fitness of  $\lambda^{(j)}$  for all objectives or, at least for one of objective,  $\lambda^{(i)}$  is strictly better than  $\lambda^{(j)}$ .

Implementing a pareto-based fitness assignment, MOGA was suggested by Fonseca and Fleming [76]. All individuals of the GA population are ranked according to the number of chromosomes in the current population, by which it is dominated, e.g. if a certain hyper parameter  $\lambda^{(i)}$  from the population at a certain generation  $t$  is dominated by  $p_i(t)$  individuals then rank is given as:

$$rank(\lambda^{(i)}, t) = 1 + p_i^{(t)} \quad (3.8)$$

The algorithm is given as:

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications



**Figure 3.2. MOGA for optimisation.**

Step 1: Apply binary or real-value encoding of hyper-parameters

Step 2: Generate initial population randomly as solutions for a certain problem.

Step 3: Calculate the objective values of chromosomes in the population and record the pareto-optimal solutions.

Step 4: Calculate objective values. In this case, the objective function  $f(\lambda^{(i)})$  is a combination of sub-objective functions. Given a chromosome

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

$\lambda^{(i)}$  with two sub objective functions then the overall objective function is represented as follows:

$$f(\lambda^{(i)}) = f_1(\lambda^{(i)}) + f_2(\lambda^{(i)}) \quad (3.9)$$

Where  $f_1(\lambda^{(i)})$  = performance and  $f_2(\lambda^{(i)})$  = time complexity

Step 5: Evaluate fitness as:

$$fitness(\lambda^{(i)}) = f(\lambda^{(i)}) \quad (3.10)$$

Step 6: Reproduce chromosomes using a standard selection method.

Step 7: Apply crossover.

Step 8: Apply mutation.

Step 9: Apply elite strategy: The elite strategy retains the top  $k$  solutions in order to keep quality solutions in each generation.

Step 10: Replace the new population generated with the previous one.

Step 11: Search for the pareto optimal solutions in the new population and update the old pareto optimal solutions with new ones.

Step 12: If the number of generations equals the pre-specified number then stop, otherwise go to step 4.

Other well-known methods in the literature for hyper-parameter optimisation are gradient-based method [77], ant colony optimisation [78], simulated annealing [79] and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) evolutionary algorithm [80] etc., which were not studied due to time constraints.

### 3.1.6.8 *Evolutionary Strategy (ES)*

Evolutionary strategy is a global optimization technique based on the theory of natural evolution. It belongs to the field of evolutionary computation, and is self-adaptation in nature, based on the mutation-selection theory given as  $(\mu, \lambda)$ -ES [219].

ES was first introduced by Rechenberg [220-221] and further modified and presented by Schwefel [222-223] where its first numerical experiment was performed by Hartmann [224]. ES turned out to be good at providing approximate solutions, even in cases of complicated optimization problems [219].

Rechenberg introduced the concept of convergence velocity in  $(1 + 1)$ -ES, as a scenario where mutation-selection was performed to produce one offspring in each generation by using Gaussian  $N(0,1)$  mutation. He further extended his theory to  $(\mu + 1)$ -ES which entails multiple members, where multiple offspring was produced and then better generation was selected over the worst parents. A further extension was  $(\mu+\lambda)$ -ES which performs as multiple parents versus multiple offspring [219].

#### 3.1.6.8.1 Evolution Strategy for Hyper-Parameter Optimization

Given a learning algorithm  $A$  for classification, ES is applied to find the best hyper-parameter value of  $\lambda$ , as  $\hat{\lambda}$ , so that equation (3.5) holds true for  $\lambda$  taken from search space  $\Lambda$  which consists of  $S$  trial points  $\{\lambda^{(1)} \dots \dots \{\lambda^{(S)}\}$ . ES performs the search using the three basic operations given below:

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

1. Mutation: It is the basic operation which introduces genetic variation among the population space. Although it is adapted and based on the problem at hand, it is conventionally performed by adding Gaussian-distribution-based random numbers to the search space given as:

$$\lambda^{(i)n+1} = \lambda^{(i)n} + N(0, \sigma^{i n}) \quad (3.11)$$

Where  $\sigma^i$  is the strategy operator which represents mutation step size and  $n$  is the generation number.

$$\sigma^{(i)n+1} = \sigma^{(i)n} \exp(\tau \cdot N(0,1)) \quad (3.12)$$

Here  $\tau$  represents the learning rate of the self-adaptation speed.

2. Recombination: Recombination is a process of combining the information of  $\mu$  number of parents to produce offspring. In the case of ES, recombination can either be discrete or intermediate. Discrete recombination combines randomly selected components of the  $\mu$  parent to produce offspring. Whereas intermediate recombination gives each of the  $\mu$  parents equal participation in the reproduction of the next generation offspring. Each offspring is taken as an average of the parent vectors [225].
3. Selection: It is performed to select the best candidate for the next generation. There are different selection operations in ES. The selection  $(\mu, \lambda)$  operator only takes  $\lambda$  offspring, while  $(\mu+\lambda)$  operator selects better candidates from both offspring as well as parent vectors [225].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

### 3.1.6.8.2 Comparative analysis of optimization techniques

The hyper parameter optimization techniques span from conventional methods like manual, grid and random search to heuristic algorithms like GA, ES, PSO and so on. All of these methods have their merits and demerits. Overall, we provide a summarized comparison below:

**Table 3.4. Merits and demerits of different optimization techniques.**

<b>Optimization Technique</b>	<b>Merits</b>	<b>Demerits</b>
Manual Search	No technical overhead involved and provides better insight into response function.	Cannot reproduce results, which is necessary for scientific research.
Grid Search	It provides better results than manual search and is reliable for low dimensional spaces.	Suffers from curse of dimensionality.
Random Search	Better than manual search as well as grid search with respect to time complexity.  Simple, easy to implement trivial to parallelism	Efficient in high-dimensional search space as response function has low effective dimensionality.
Bayesian Optimization	Faster than random and grid search (especially in higher dimensions).  Incorporates proper mathematical justification [121].	There is a high chance of fall in local minimum (exploitation).
Particle Swarm Optimization	Computationally efficient as compared to other population-based algorithms.  Easy to implement and few parameters to adjust [120].	The method easily suffers from partial optimism [120].
Multi-objective genetic algorithm	Solves complex multi-objective problems.	Low speed, perfect-optimization not guaranteed.

Evolutionary Strategy	It exhibits self-adaptation search capability; hence is faster to converge.	Evolutionary Strategy has individuals as a real number vector and is suitable for an application which deals with floating point values.
-----------------------	---	--

---

The machine learning classifiers are parameterized by a set of hyper parameters, which are used to configure various aspects of the classifier and have varying effect on the resulting model performance. The search for the best hyper-parameter values is a formidable optimization task (also called model selection).

Researchers have studied the impact of applying different optimization techniques to enhance machine learning classifiers performance. We consider starting with the basic manual search technique that manually examines a given set of hyper-parameter values for a possible solution. Next is the grid search, which explores each possible combination of hyper-parameter values for a potential solution. Although the grid search does not suffer from the demerits of a manual search (i.e., the manual search cannot reproduce its results scientifically), the grid search suffers from the curse of dimensionality. To overcome the weakness of the grid search, a random search was introduced by James et al. [60]. A random search arbitrarily chooses the elements from the given search space by selecting only those hyper-parameter dimensions that are effective.

Further research has extended the above results to discover the ability of heuristic search algorithms for a classifier's hyper-parameter optimization

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

(model selection). Among these possibilities is Particle Swarm Optimization (PSO), which is a stochastic global optimization technique that is based on biological systems (social systems) and involves the study of collective behaviour of individuals who are interacting with their environment and one another (swamp intelligence) [124]. This approach assumes the scenario that social interaction produces information that evolves over time and distances and, thus, helps to provide solutions to certain problems [125]. Kennedy et al. transformed this behaviour into a mathematical form and applied it successfully to optimization problems. In the list are evolutionary algorithm (EA) namely, genetic algorithms (GA) and multi-objective genetic algorithms (ES). These algorithms are based on the natural phenomena of adaptation and genetics in biological systems and how these phenomena can be imported to computer systems. Other well-known methods in the literature for hyper-parameter optimization are gradient-based methods [126], ant colony optimization [127], simulated annealing [128], and the covariance matrix adaptation evolution strategy (CMA-ES) evolutionary algorithm; these methods are not covered here due to space constraints.

Researchers have applied the following methods successfully to enhance the performance of machine learning classifiers. Hsu et al. applied a grid search to building a model for the SVM, with different kernels, to optimize the classification performance and reported a noticeable improvement in the classification performance [21]. James et al. applied random search to improve neural networks [129]. Imbault et al. applied genetic algorithms and simulated annealing techniques to optimize the hyper-parameters of an RBF kernel-based

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

SVM and proved that both of these techniques generate similar enhancement. Han et al. improved the ELM performance by optimizing the input weights and hidden biases using PSO selection [130], while Cho et al. applied the Bacterial Foraging Algorithm for the same purpose [131]. In the case of ESN, these methods work by tuning specific ESN parameters such as the reservoir size, input, reservoir and output weight scale, the sparsity of the weight matrices, and the leaking rate [132]. Sebastián Basterrech et al. applied the Particle Swarm Optimization (PSO) to find the optimum initial hidden weights for the ESN model [133]. They tested their algorithm on three-time series datasets (Sante Fe Laser, NARMA, Internet Traffic) for solving prediction problems. Krause et al. applied a ES to obtain an optimum ESN solution for an application to multiple motor-pattern learning [134]. Their application worked by tuning the ESN parameters: weight scale; sparsity ratio and reservoir size. Jiang et al. applied the CMA-ES evolutionary algorithm to optimize the ESN parameters [135]. The method was tested on the classic double-pole-balancing control problem. Their approach replaced gradient-based quadratic optimization (used in Jaeger's classical ESN algorithm to output the weights) with evolutionary algorithms. Han et al. proposed fast subspace decomposition (FSD) of an echo state network (FSDESN) as a new model of an ESN that extracts a subspace from the reservoir to remove the collinear components [136]. The method combines fast FSD with ESN to optimize the performance; it was tested on two multivariate chaotic benchmark tasks: prediction of the Henon Map and prediction of the Lorenz System. Devert et al. used an adaptive evolutionary methodology to optimize the weights of the network, using an artificial

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

embryogenic, and applied it to benchmark the disc and half-disc problems [137]. Yuan et al. applied GA and Particle Swarm Optimization (PSO) to select the best parameters and a suitable population size for SVM [138], whereas İlhan et al. applied GAs for enhancing the SVM performance by optimizing its parameters as well as selecting a subset of features [56]. Huang et al. proposed an optimized version of SVM, called Sparse Support Vector Classification (SSVC), which enhances the testing speed of the classifier by minimizing the number of required support vectors. It was accomplished by setting the value of irrelevant parameters to zero [139] in a Bayesian setting.

### 3.1.6.9 *Classification*

Classification is an integral part of many decision-making systems for prediction, forecasting, diagnosis and pattern recognition [141]. It is a procedure to assign a class/label (from among the predefined classes) to new cases in a dataset on the basis of the information that is incorporated into its feature set. A good classifier is judged by its accuracy, speed, comprehensibility, time to learn and avoidance of over-fitting [28]. In the literatures, there have been many famous proposals for constructing classifiers in machine learning; these proposals include generative models, neural networks, and discriminative models.

#### 3.1.6.9.1 Probabilistic and Statistical Models

Probabilistic and statistical models incorporate regression analysis, discriminant analysis, time series analysis, and principal component analysis, among many other methods. The most famous probabilistic models, Bayesian models, are based on the probabilistic Bayesian framework, which calculates the probability

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

of a sample belonging to a certain class given the data and prior assumptions. An example is the Naïve Bayes (NB) classifier, which assumes independence between features [11].

### 3.1.6.9.2 Bio-inspired Algorithms

Several machine learning algorithms are inspired by biology. Neural networks imitate the human nervous system by representing information as nodes that are connected by weighted links (synapses). The learning algorithm adjusts the weights in the neural network so that it can predict the class of unseen examples correctly [11]. Neural networks are distinguished by their architecture, such as back-propagation or feed-forward. The most famous neural networks are self-organising maps [142] and Hopfield networks [143]. Due to their predictive power, they are used for classification and clustering problems [11].

### 3.1.6.9.3 Evolution-based methods

Evolution-based methods are inspired by the Darwinian theory of evolution and include genetic algorithms and genetic programming. Other systems based on natural processes include artificial immune systems [145] and ant-colony optimization [144].

### 3.1.6.9.4 Symbolic learning and rule-induction methods

Symbolic learning and rule-induction methods represent a model as a set of rules and can be further categorized as learning by role, by being told, by analogy, by example, and by discovery. Algorithms include decision-tree induction, the ID 3 decision tree building algorithm, C 4.5, CART, and logic programming [146]. These algorithms search for the attributes that are best

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

suitable for dividing the dataset into classes by minimising the entropy, or information uncertainty. After classification, the results are represented as decision trees or production rules.

### 3.1.6.9.5 Analytic learning and fuzzy logic Models

Analytic learning and fuzzy logic represent knowledge as logical rules and perform reasoning on them. Advanced representation of these rules has been developed as fuzzy systems and fuzzy logics. These systems do not rely on binary assignments of 0 (false) or 1 (true), as in Boolean logic, but rather permit the full range of real values between 0 (completely false) and 1 (completely true), to make approximate reasoning [146].

### 3.1.6.9.6 Instance-based methods

Instance-based methods do not extract a model, but instead use the full training dataset to classify new samples. Methods such as KNN (k-nearest neighbours), weighted regression, and case-based reasoning come under this category [146]. Finally, kernel methods characterize each sample using a kernel function, which maps the samples from original space to high-dimensional feature space. Support vector machines (SVMs) are the most prominent.

### 3.1.6.9.7 Comparison of Classification Techniques

A brief comparison of machine learning techniques is provided in the table below.

**Table 3.5. Merits and demerits of different classification techniques.**

Technique Name	Merits	Demerits
Probabilistic and	Computational time for	Less accurate

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

statistical model[147]	training is minimal.	compared to other classifiers.
Neural Network Based Classification	Inherently parallel which can be used to speed up the computational process via parallel processing [147].	Requires long learning time.
Symbolic learning and rule-induction methods	Simple and fast. Easy to understand.	Long training time.
Instance-based methods	The most robust and accurate methods among all well-known classifiers[147].	Large storage requirements.

Classification plays an important role in improving the accuracy of the CADSSs. Researchers experimented with various classifiers. J.C. Fu [101] compared performance of two classifiers, General Regression Neural Network (GRNN) and Support Vector Machine (SVM) for detection of micro calcifications in mammogram images. It used texture features in the spectral and spatial domain. The results reported were 97% for SVM and 96% for GRNN. Wang et al. [100] used intensity, geometric and texture features to distinguish abnormalities from normal cases. The classification was performed using SVM and Relevance Feedback SVM (RF-SVM). The results reported were sensitivity within the range of 85.9%-90.6% and false-positive fall to 3.6 per image. F. Moayedi et al. [97] applied Support Vector Machine based Fuzzy Neural Network (SVFNN) as the classifier. F. Moayedi et al. [96] experimented with weighted SVM and Successive Enhancement Learning (SEL) weighted SVM for classification. The accuracy of the classifiers turned out to be within the

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

range of 0.73-0.86 for weighted SVM and 0.85-0.96 for SEL weighted SVM [101]. Wang et al. [102] introduced the Structured SVM (s-SVM) classifier for diagnosis of malignant cases. Eltoukhy et al. [107] applied a supervised classifier system based on Euclidean Distance to the multi resolution and curvelet features extracted from the segmented region in MIAS database mammograms. Classification Accuracy turned out to be 96.36% at the maximum for curvelet features and 95% for wavelet features [108] F. Moayedi et al. [95] applied SEL weighted SVM, SVFNN and kernel SVM classifiers to curvelet based texture features. The maximum accuracy was achieved with SEL weighted SVM which was 96.6%. Yu et al. [148] combined Markov Random Field based texture features with auxiliary features, such as mean pixel values, standard deviation and the edge density from suspicious micro-calcification regions. These features were fed to the Back Propagation Neural Network (BPNN) classifier and resulted in a maximum specificity of 98.8%.

Krishnan et al. applied two different kernel functions, the polynomial kernel function and Gaussian radial basis function, for classification purposes. The performance of the classification was analyzed using the overall classification accuracy, sensitivity and specificity measures [149]. Yua et al. proposed two-class SVM classification for the detection of masses in mammogram images [150]. This technique uses Moran's index and Geary's coefficient as feature vectors to be classified and uses radial basis as a kernel function. ED Übeyli compared the classification accuracy of different classifiers, including the multi-layer perception and support vector machine, on the Wisconsin dataset [151]. Huang et al. proposed a computer-aided diagnosis method for diagnosing the

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

gastroesophageal reflux disease (GERD) from endoscopic images of the esophageal-gastric junction using a hierarchical heterogeneous descriptor fusion support vector machine (HHDF-SVM) [155]. El-Naqa et al. used polynomial and Gaussian radial basis functions as kernels in SVMs [152]. Huang et al. applied SVMs with Radial Basis functions for classification of the malignant masses, but before applying the classifier, they found the best values of the RBF model ( $C$ ,  $\gamma$ ) by applying cross-validation [153]. Polat et al. suggested the Least Square Support Vector Machine (LS-SVM), which uses a linear equation for training the SVM [154]. Hussain et al. compared the performance of SVMs with different kernels to detect breast cancer on the Wisconsin dataset [156]. L. Chisci et al. predicted epileptic seizures from the online analysis of EEG data using SVM with 100% sensitivity [157].

### 3.1.6.10 *Classification Model Validation*

It is essential to validate the accuracy of a classification model to predict its future accuracy as well as to choose among a number of possible classifiers. This can be accomplished by applying some benchmark methods as described below.

#### 3.1.6.10.1 Holdout Method

This method partitions the dataset into two mutually exclusive subsets, training set and test set. A portion of the dataset is assigned as a training set, while the rest is considered test set [158]. The accuracy of a classification model tends to increase as more data points are incorporated into the training set, hence this is a pessimist estimator as only a selected subset of the whole dataset is provided for training. The more instances for test set, the larger the bias in estimate,

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

whereas leaving less data points for testing set will largely increase the variance of the accuracy estimation [158].

### 3.1.6.10.2 Cross-validation

The  $k$ -fold cross-validation technique divides the dataset  $D$  into  $k$  mutual exclusive subsets (folds) of almost the same size  $D_1, D_2, \dots, D_k$ . The classifier is trained and tested on these  $k$  subsets, and then an estimate of the accuracy is computed as the average of the accuracies computed for each fold. In the case of this cross validation scheme, increasing  $k \rightarrow n$  leads to leave-one-out cross validation, with decreased bias accuracy estimation. Decreasing  $k$  increases accuracy variance, due to increased instability of training sets [158].

Cross-validation is sometimes called stratified, in cases where each  $k$  partition reflects the distribution the same as among classes of the original dataset. When compared to regular model evaluation cross-validation, stratification turned out better than original in terms of bias and variance [158].

### 3.1.6.11 *Model Evaluation*

In order to evaluate the performance of CDSS, developed using supervised machine learning techniques, classification performance (binary or multiclass) is to be measured. In the case of the binary classification model, the aim is to classify each data point of the given dataset among two classes, positives (P) and negatives (N), according to a classification rule. In the clinical domain: a positive value normally indicates an abnormal condition which needs to be investigated, for example presence of a disease whereas in the case of multi-class analysis it is the class of the abnormality, or its stage [76].

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

Consequently, there are four possible outcomes when predicting the class label for each data element in the dataset, namely [76]:

TP (True Positive): A positive case is predicted correctly,

TN (True Negative): A negative case is predicted incorrectly,

FP (False Positive): A positive case is predicted incorrectly,

FN: False Negative: A negative case is predicted incorrectly.

Generally, the evaluation measures of classification algorithms are defined in the form of the matrix called the confusion matrix. The confusion matrix for a binary classification problem-with only two classes is shown in table below.

**Table 3.6. Confusion Matrix**

<b>Actual</b>	<b>Predicted</b>	
	<b><i>Negative</i></b>	<b><i>Positive</i></b>
<b>Negative</b>	TN	FP
<b>Positive</b>	FN	TP

Keeping the above in view *sensitivity*, the measure of accuracy of predicating true cases is given as:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3.13)$$

And *specificity*, the measure of accuracy of predicting false cases is given as:

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (3.14)$$

And

$$\text{Accuracy} = \frac{(TP+TN)}{(TN+FN+FP+TP)} \quad (3.15)$$

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

The above-mentioned performance measures lack the ability to characterize the predictive power of a classifier properly. For this reason, *receiver-operating characteristic (ROC)* curves are also used to analyse the classifier performance.

It is a simple graphical representation of classifier performance. Receiver operating characteristics (ROC) graphs are increasingly employed in medical decision making, machine learning and data-mining research for visualizing classification performance.

An ROC is a two-dimensional graph which depicts relative tradeoffs between benefits (true positives) and costs (false positives). In ROC, the FP rate is plotted on the X axis while the TP rate is plotted on the Y axis [159].

ROC performance can be expected with a single numeric value called the area under the ROC curve, abbreviated as AUC [160]. AUC has a value range between 0 and 1.0 since it is computed over the area of the unit square. In case the prediction is performed using random guessing, ROC may come up as a diagonal line between (0, 0) and (1, 1), which produces an AUC value of 0.5. In general, no realistic classifier has an AUC value below 0.5. The AUC value can be understood as the probability of a classifier ranking a randomly chosen positive instance higher than a randomly chosen negative instance [159].

### 3.1.6.12 *Statistical Tests for Comparison of Classifiers*

The statistical are generally applied by researchers to evaluate and then draw conclusions from their scientific research. There are a lot of them mentioned in

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

literature but we applied t-tests, and ANOVA to analyse and compare the performance of our classifiers.

There is no established procedure for comparing performance of more than one classifier over multiple datasets. There are a number of statistical techniques which are adopted by the researchers to decide whether the difference between classifiers is real or random.

### 3.1.6.12.1 t-test

The t-test, also called the Student's t-test, compares the performance of two classification algorithms by assessing whether the means of two classifiers' performance are statistically different from each other. It checks the null hypothesis ( $H_0$ ): means of the accuracy measure of two classifiers is the same with respect to alternate or research hypothesis ( $H_1$ ): means of the accuracy measure of two classifiers are different.

The t-test is based on two values, *t-score* or *t-ratio*, which are used to find out the p-value from a standard t-table. The t-score is computed as the difference of means of the datasets being tested divided by variability of data, represented as the standard error of the difference as mentioned in the equation below:

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2) - (\sum d)^2}{n-1}}} \quad (3.16)$$

Where  $\sum d =$  Sum of the differences between each pair of observations. We applied the paired t-test to dependant samples from the same dataset.

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

To perform the t-test, the following steps need to be followed: given two sample sets (assuming of same size  $n$ ):

- Calculate the t-ratio using equation (3.16),
- Find the corresponding p-value of the t-statistic in a t-table (Figure 3.3) at a certain level of significance alpha,
- In case the value of the t-statistic is less than alpha it is significant, thus concluding the sample sets are different [161].

# The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

**Table T** Critical Values of the *t* Distribution

<i>df</i>	One-Tail = .4 Two-Tail = .8	.25 .5	.1 .2	.05 .1	.025 .05	.01 .02	.005 .01	.0025 .005	.001 .002	.0005 .001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Source: From *Biometrika Tables for Statisticians*, Vol. 1, Third Edition, edited by E. S. Pearson and H. O. Hartley, 1966, p. 146. Reprinted by permission of the Biometrika Trustees.

Figure 3.3. A sample t-table <sup>1</sup>

## 3.2 Conclusion and Discussion

This chapter describes a general framework to develop efficient CDSSs with a focus on existing state of the art technique which has been adopted at each stage of the overall CDSS framework. Our review of the existing machine learning, pattern recognition and evolutionary algorithm techniques (in chapter 2) directed us towards building CDSSs with better medical diagnosis

<sup>1</sup> [www2.palomar.edu](http://www2.palomar.edu)

## The Conventional Framework for Development of Clinical Decision Support System (CDSS) Applications

capabilities (starting from chapter 3 as a general framework, with chapter 4 and chapter 5 as implications of this framework in conjunction with implication of novel and better techniques to develop CDSSs with better diagnosis performance. The next two chapters describe those systems with different applications in field of medical health care and diagnosis (chapter 4 and chapter 5).

## **Chapter 4 - A Novel Application of a Local Energy-based Shape Histogram (LESH)–based CDSS for Breast Cancer Detection using Mammography and Magnetic Resonance Imaging**

### **4.1 Introduction**

The previous chapter describes the general framework of CDSS for medical diagnosis. Building upon that framework, this chapter presents our proposed CDSSs based on the novel application of the LESH feature extraction technique for diagnosing cancer from mammograms, chest radiographs and MRI images. The LESH feature extraction technique has proved to be robust for solving problems involving pattern recognition. The LESH feature extraction technique was initially proposed by Sarfraz and Hellwich for face recognition systems of different face and head poses [163–165] and Zakir et al. applied LESH to automatically detect and recognise different road signs [166].

Looking at the capability of the LESH feature for discriminating between different medical abnormalities found in two-dimensional images, we further extended the technique to develop a three-dimensional (3D) LESH feature extraction technique for volumetric images and then applied it to detect cancer from breast MRI images.

Most of the algorithms mentioned in literature reviews focus on one type of abnormality (either mass or micro calcification) and try to diagnose its malignancy, whereas we aimed to propose a method which can efficiently deal with different medical image modalities (mammograms and MRI) as well as different types of

abnormalities with higher accuracy rates. The algorithm produces better results in discriminating between malignant and benign cases as well as different types of abnormalities. When combined with benchmarked machine-learning classifiers, our proposed system demonstrates prominent performance improvement compared to a state-of-the-art CDSS for breast cancer diagnosis.

The rest of the chapter is organised as follows: section 4.2 describes the details of the LESH feature extraction technique for two-dimensional and three-dimensional images. Next, the effectiveness of the proposed algorithm when incorporated as part of overall CDSSs (discussed in chapter 3) is demonstrated using two-dimensional and three-dimensional data sets and compared with a benchmark wavelet-based feature extraction technique in section 4.3. Lastly, the chapter concludes with the final remarks about the proposed technique in section 4.4.

## **4.2 Local Energy-based Shape Histogram (LESH)**

The LESH feature extraction technique works by computing a histogram of the local energy pattern within the image of interest. A histogram is a simple technique that forms the basis of many spatial domain image processing techniques [24]. It provides useful image statistics that can be used to further analyse and process the image. Details of LESH features are provided below for two-dimensional and three-dimensional images separately.

#### 4.2.1 Two-dimensional (2D) LESH Feature Extraction Technique

LESH works by converting an image into a combination of local energies along different orientations. Morrone and Owens suggested that features extracted at the points of maximum phase congruency can be helpful in image analysis [167]. The type of phase and amplitude of local maxima of the energy function determine the type, sign, and contrast of a feature. Local energies are calculated along different orientations using the phase congruency method suggested by Morrone and Owens [167]. Before calculating the phase congruency (PC), image  $I$  is first convolved with a bank of 2D log-Gabor filters with different orientations  $n$  and scales  $c$ . The transfer function for log-Gabor is given as:

$$G(\omega) = \exp \left\{ -\frac{\left( \log\left(\frac{\omega}{\omega_0}\right) \right)^2}{2 \left( \log\left(\frac{k}{\omega_0}\right) \right)^2} \right\} \quad (4.1)$$

Where  $\omega_0$  is filters central frequency and  $\frac{k}{\omega_0}$  is a constant that varies  $\omega_0$  central radial frequency.

Let  $G_{co}^{even}$  and  $G_{co}^{odd}$  be the even-symmetric and odd-symmetric filters at scale  $c$  and orientation  $n$ . Then, convolution with Fourier transform(fft) of image  $I$  results in (in frequency domain) the response vector given as:

$$[e_{cn}(x, y), o_{cn}(x, y)] = [fftI(x, y) * G_{cn}^{even}, fftI(x, y) * G_{cn}^{odd}] \quad (4.2)$$

where  $I(x, y)$  represents a location in the image. Hence, the amplitude of the response at a given scale and orientation can be computed as:

$$A_{cn} = \sqrt{(e_{cn}(x, y))^2 + (o_{cn}(x, y))^2} \quad (4.3)$$

And phase angle is given as:

$$\phi_{cn}(x, y) = a \tan(o_{cn}(x, y) / e_{cn}(x, y)) \quad (4.4)$$

Let  $\bar{\phi}_n(x, y)$  be the phase angle at orientation  $n$ , given as:

$$(\bar{\phi}_n^{even}(x, y), \bar{\phi}_n^{odd}(x, y)) = \frac{(\sum_c e_{cn}(x, y), \sum_c o_{cn}(x, y))}{\sqrt{(\sum_c e_{cn}(x, y))^2 + (\sum_c o_{cn}(x, y))^2}} \quad (4.5)$$

And a sensitive phase deviation measure is given as:

$$\Delta\Phi_n(x, y) = \cos(\phi_n(x, y) - \bar{\phi}_n(x, y)) - |\sin(\phi_n(x, y) - \bar{\phi}_n(x, y))| \quad (4.6)$$

Finally, local energy is computed as:

$$E(x, y) = \sqrt{(\sum_n e_{cn}(x, y))^2 + (\sum_n o_{cn}(x, y))^2} \quad (4.7)$$

Last, two-dimensional phase congruency for the image is computed as local energy normalised by the sum of Fourier amplitude components as:

$$PC(x, y) = \frac{\sum_n W(x, y) |A_n(x, y) \Delta\Phi_n(x, y) - T|}{\sum_n A_n + \epsilon} \quad (4.8)$$

where  $T$  is the noise cancellation factor and  $W(x, y)$  is the weighting of the frequency spread.  $A_n$  and  $\phi_n$  represent the amplitude and phase angle, respectively, of local complex value Fourier components at location  $(x, y)$  of the image of size, and  $\epsilon$  is a constant value incorporated to avoid division by zero [168].  $A_n$  and  $\phi_n$  are calculated using the logarithmic Gabor wavelets filter [167].

These filters are designed so that they can detect features at all orientations. The sum of energies for every orientation is normalised by overall sum and scales of amplitude [168]. The resultant LESH feature vector is calculated as follows:

$$h_{r,b} = \sum W_r \times PC(x, y) \times \delta_{r-b} \quad (4.9)$$

$$W_r = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{[(x-r_{x0})^2+(y-r_{y0})^2]}{\sigma^2}} \quad (4.10)$$

where  $W_r$  is the Gaussian weighting function of region  $r$  in the image,  $PC$  represents the local energy computed by the equation and  $\delta_{r-b}$ , represents Kronecker's delta of the orientation label map  $L$  and current bin  $b$  [163]. Table 4.1 depicts the overall algorithm for LESH feature extraction.

**Table 4.1. LESH feature extraction algorithm.**

---



---

**Algorithm: Let  $I$  be a two-dimensional medical image with a certain pixel location given as  $(x,y)$ .**

---

Begin:

1. Convolve the image  $I$  with a bank of 2D log-Gabor filters  $G(\omega)$  given in "(4.1)," with different orientations  $n$  and scales  $c$ . The resultant convolution response vector is given in "(4.2)".
2. Calculate the amplitude of the response using "(4.3)".
3. Calculate sensitive phase deviation measure using "(4.6)".
4. Calculate local energy using "(4.7)".
5. Calculate two-dimensional phase congruency for the image using "(4.8)".
6. Compute the LESH feature vector using "(4.9), & (4.10)".

End;

---



---

#### **4.2.2 Three-dimensional (3D) Local Energy-based Shape Histogram (LESH) Feature Extraction**

Our proposed 3D LESH is an extension of 2D LESH features by Sarfaraz and Hellwich [163], which we developed for extracting informative patterns from

volumetric medical images for medical diagnosis. LESH is calculated in the following steps:

#### 4.2.2.1 Convolution with 3D log Gabor

A Fourier transform (fft) of the 3D image  $I$  is convolved with a bank of 3D log-Gabor filters with different orientations  $n$  and scales  $c$  (in frequency domain). A 3D log-Gabor filter is devised as the product of one-dimensional log-Gabor function of the radial frequency and the Gaussian function of the angular distance with rotational symmetry of spherical coordinates given as:

$$G(\omega, \Theta) = \exp \left\{ -\frac{\left( \log\left(\frac{\omega}{\omega_i}\right) \right)^2}{2 \left( \log\left(\frac{\sigma\omega}{\omega_i}\right) \right)^2} \right\} \times \exp \left( -\frac{\alpha(\Theta)^2}{2 \sigma_\alpha^2} \right) \quad (4.11)$$

where  $\Theta = (\theta, \phi)$  is the filter orientation on a sphere of unit radius and  $\theta$  is the elevation angle and  $\phi$  is the azimuth angle.  $\omega_i$  is the central radial frequency of the  $i^{th}$  filter and  $\sigma_\omega$  and  $\sigma_\alpha$  are the standard deviations controlling the filter bandwidth [170, 178] and  $\alpha(\Theta) = \arccos\left(\frac{\omega \cdot v}{\|\omega\| \|\nu\|}\right)$ ,  $\nu = (\cos \phi \cdot \cos \theta, \cos \phi \cdot \sin \theta, \sin \theta)$  and  $\omega$  is the point in the frequency space in Cartesian space [170, 178]. Let  $G_{co}^{even}$  and  $G_{co}^{odd}$  be the even-symmetric and odd-symmetric filters at scale  $c$  and orientation  $n$ . The convolution (in the frequency domain) results in the response vector can be given as:

$$[e_n(s, \Theta), o_n(s, \Theta)] = [fftI(s) \cdot G(\omega, \Theta)_{cn}^{even}, fftI(s) \cdot G(\omega, \Theta)_{cn}^{odd}] \quad (4.12)$$

Where  $s$  indicates a location  $(x,y,z)$  in the image  $I$ . Hence, the amplitude of the response at a given scale and orientation can be computed as:

$$A_n(s, \theta) = \sqrt{(\mathbf{e}_n(s, \theta))^2 + (\mathbf{o}_n(s, \theta))^2} \quad (4.13)$$

And the sensitive phase deviation measure proposed by Kovési [168] is given as:

$$\begin{aligned} \Delta\Phi_n(s, \theta) &= \cos(\phi_n(s, \theta) - \bar{\phi}_n(s, \theta)) - |\sin(\phi_n(s, \theta) - \bar{\phi}_n(s, \theta))| \\ &= \mathbf{e}_n(s, \theta)\bar{\phi}_e(s, \theta) + \mathbf{o}_n(s, \theta)\bar{\phi}_o(s, \theta) - |\mathbf{e}_n(s, \theta)\bar{\phi}_o(s, \theta) - \mathbf{o}_n(s, \theta)\bar{\phi}_e(s, \theta)| \end{aligned} \quad (4.14)$$

Where

$$\bar{\phi}_e(s, \theta) = \sum_n \mathbf{e}_n(s, \theta) / E(s, \theta), \quad (4.15)$$

$$\bar{\phi}_o(s, \theta) = \sum_n \mathbf{o}_n(s, \theta) / E(s, \theta) \quad (4.16)$$

Finally, local energy is computed as:

$$E(s, \theta) = \sqrt{(\sum_n \mathbf{e}_n(s, \theta))^2 + (\sum_n \mathbf{o}_n(s, \theta))^2} \quad (4.17)$$

#### 4.2.2.2 Computation of 3D Phase Congruency

The 3D phase congruency is computed as local energy normalised by the sum of Fourier amplitude components as:

$$PC_{3D}(s, \theta) = \frac{\sum_n W(s, \theta) [A_n(s, \theta) \Delta\Phi_n(s, \theta) - T]}{\sum_n A_n(s, \theta) + \epsilon} \quad (4.18)$$

Where  $[.]$  means the enclosed measure is equal to itself if positive and is zero otherwise, for the spatial location  $s=(x, y, z)$  in the 3D MRI image, and  $\theta = (\theta, \Phi)$  is the filter orientation with unit radius for the sphere,  $\theta$  is elevation and  $\Phi$  is azimuth angle. Here,  $A_n(s, \theta)$  is the image energy at location  $s$ , computed using the 3D

Gabor filter at scale index  $n$ , which varies from 1 to  $N$  (total number of scales being used), and filter orientation  $\theta$ .  $T$  is the threshold which controls the noise level of the image energy map and  $\epsilon$  is a small constant to avoid division by zero [168].

$W(s, \theta) = \frac{1}{1+e^{\gamma(d-l(s,\theta))}}$  is the phase congruency weighting function devised by Kovesi [168–169], where  $d$  is the threshold value for the filter response spread. Below this, the phase congruency value is penalised and  $\gamma$  is the gain factor, which controls the sharpness of the filter response spread [170].

Given that  $l(s, \theta) = \frac{1}{N} \left( \frac{\sum_n A_n(s, \theta)}{A_{max}(s, \theta) + \epsilon} \right)$  is the filter response spread measure computed as the sum of amplitudes ( $A_n$ ) divided by the highest individual response ( $A_{max}$ ) the width of the distribution is obtained [170].

#### 4.2.2.3 *Local Energy-based Shape Histogram Feature Extraction*

The histogram of the local energy of each sub-region along each filter is generated and combined to preserve the relationship between different regions. The sum of energies for every orientation is normalised by the overall sum and scales of amplitude [169]. The resultant LESH feature vector is calculated as follows:

$$h_{r,b} = \sum W_{3d} \times PC_{3D}(s, \theta) \times \delta_{r-b} \quad (4.19)$$

$$W_r = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{[(x-r_{x0})^2+(y-r_{y0})^2+(z-r_{z0})^2]}{\sigma^2}} \quad (4.20)$$

where  $W_r$  is the Gaussian weighting function of region  $r$  in the 3D MRI image,  $PC_{3D}(s, \theta)$  represents the local energy computed by the equation and  $\delta_{r-b}$ ,

represents Kronecker's delta of the orientation label map  $L$  and current bin  $b$ . Table 4.2 illustrates the overall algorithm for 3D LESH feature extraction.

**Table 4.2. 3D LESH feature extraction algorithm.**

---

**Algorithm:** Let  $I$  be a 3D MRI image with a certain pixel location given as  $s=(x,y,z)$ .

---

Begin:

1. Convolve the image  $I$  with a bank of 3D log Gabor filter  $G(\omega, \theta)$  given in "(4.12)," with different orientations  $n$  and scale  $c$ . The resultant convolution response vector is given in "(4.12)".
  2. Calculate the amplitude of the response using "(4.13)".
  3. Calculate sensitive phase deviation measure using "(4.15)".
  4. Calculate local energy using "(4.18)".
  5. Calculate three-dimensional phase congruency for the image using "(4.18)".
  6. Compute the LESH feature vector using "(4.19), & (4.20)".
- End;
- 

## 4.3 Simulation Results

The effectiveness of our proposed LESH-based CDSS to identify cancer via medical image analysis is established using different 2D and 3D image modality data sets as described below.

### 4.3.1 Case Study I: Mammogram Data Sets

We selected two mammogram data sets to conduct this study, and they are discussed below.

#### 4.3.1.1 Mammogram Data Sets

We experimented with two publicly available mammogram data sets: Mammographic Image Analysis Society (MIAS) and INbreast data sets. The MIAS data set [78] contains images with 1024x1024 pixel resolution, which have been

digitised with 50-micron pixel edges; they are available in Portable Gray Map (PGM) format. The data set contains 322 mammogram cases from 161 patients. Of these 322 mammograms, 207 are normal, 51 are malignant, and 64 are benign. The images have been labelled by radiologists and provide a variety of abnormal cases: micro calcification (25 cases), circumscribed masses (23 cases), architectural distortion (19 cases), speculated masses (19 cases) and miscellaneous (15 cases). Almost 50% of the images are comprised of backgrounds with different types of noise.

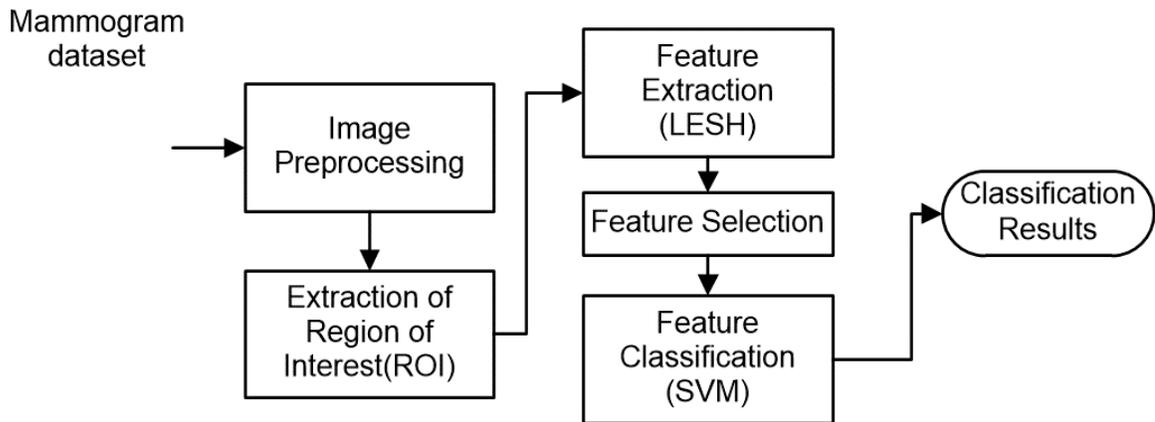
The INbreast data set contains 115 cases (410 images) with different types of abnormalities [79]. The images were acquired from the Centro Hospitalar de S. João [CHSJ], Breast Centre, Porto, with the permission of the Portuguese National Committee of Data Protection and Hospital's Ethics Committee. The images are in Digital Imaging and Communications in Medicine (DICOM) format, with annotations and metadata being available. The prominent abnormalities are masses (108 cases) and calcifications (308 cases) with only 3 cases of architectural distortion and 14 cases of asymmetry. Three hundred ninety-six ROI were gathered from the data set for experimentation; out of those, 117 were masses (some cases had multiple masses) and 279 were calcifications.

#### 4.3.1.2 *Experimental Set-up*

The region of interest (ROI) was manually segmented from the mammogram images using the information from the centre of abnormality and the radius of a circle enclosing it (provided with the dataset). All the ROIs were of different sizes

and thus were resized as 64x64. The dataset was further split into testing and training sets using ten-fold cross-validation, where 9 folds are used for training and the 10<sup>th</sup> fold is used for evaluation. The results are averaged and reported in Section 4.3.1.8.

The overall system framework is depicted in Figure 4.1 below.



**Figure 4.1. Overview of proposed methodology.**

#### 4.3.1.3 *Image Pre-processing*

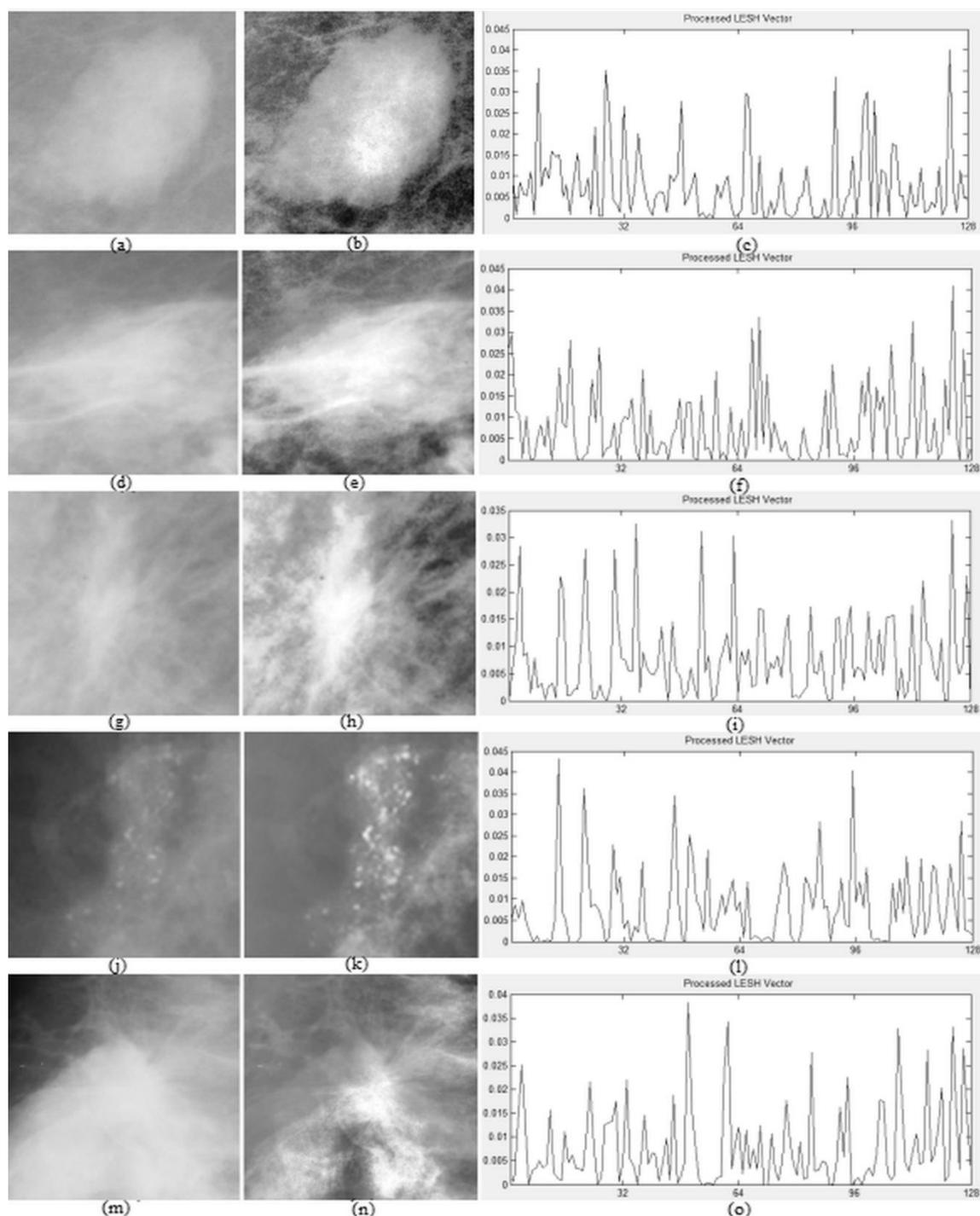
Mammograms are inherently noisy images. This noise hinders the true detection of small micro calcifications and masses. Mammograms were pre-processed to suppress the noise and enhance important image features. The pre-processing includes two steps: image normalisation and then application of Contrast Limited Adaptive Histogram Equalisation (CLAHE). Image normalisation and enhancement techniques are discussed in section 2.4.2.2.

#### 4.3.1.4 *Extraction of Region of Interest (ROI)*

After the pre-processing, ROIs were extracted from 113 abnormal cases of the MIAS data set and 396 abnormal cases from the INbreast data set [79]. The abnormalities included circumscribed masses (CIRC), ill-defined masses (MISC), speculated masses (SPIC), calcification (CALC) and architectural distortion (ARCH). Figure 2.5 shows the examples of these abnormalities from the MIAS data set. The ROI were extracted according to the information available in the database about the centre of the abnormality and the radius of the area within which the abnormality lies.

#### 4.3.1.5 *Feature Extraction*

A histogram of the local energy information, LESH, was obtained for 16 sub-regions of the ROI for 8 different orientations as the feature set. This resulted in a 128-dimensional feature vector ( $16 \times 8 = 128$ ). The histogram of the local energy of each sub-region along each filter was generated and combined to preserve the relationship between different regions. Next, an orientation map was generated by assigning an orientation that had the largest energy across all scales to each pixel. The local histogram was calculated as mentioned in [section 4.2](#) above.



**Figure 4.2. LESH vector for different abnormalities obtained from software developed by (Zakir et al., 2011) (a) circumscribed masses (CIRC); (b) CIRC after enhancement; (c) LESH vector for CIRC; (d) ill-defined masses (MISC); (e) MISC after enhancement; (f) LESH vector for MISC; (g) speculated masses (SPIC); (h) SPIC after enhancement; (i) LESH vector for SPIC; (j) calcification (CALC); (k) CALC after enhancement; (l) LESH vector for CALC; (m) architectural distortion (ARCH); (n) ARCH after enhancement; (o) LESH vector for ARCH.**

#### 4.3.1.6 *Feature Selection*

We selected a subset of  $N$ -largest coefficients, say  $h_N$ , from LESH feature vectors,  $h_{r,b}$ , and fed them to the benchmark machine-learning classifiers. Cristiane et al. [119] applied the same technique for dimensionality reduction and classification improvement. They experimented with different numbers of largest wavelet coefficients as features for classification and reported the results.

#### 4.3.1.7 *Classification*

Experiments have been conducted with MIAS and INbreast data sets to distinguish malignant and benign cases and to classify abnormalities according to their types.

#### 4.3.1.8 *Results and Discussion*

The selected LESH features,  $h_N$ , were fed to the SVM classifier with different kernel functions. SVMs are supervised learning models that are used for classification and regression analysis (section 2.4.5.3). The results were given with classification accuracy (for details see section 3.2.6.4) and the ROC curve as measures of performance. The ROC curve plots the fraction of true positives out of positives against the fraction of false positives out of negatives (for details see section 3.2.6.4.1). The ROC plots in Figures 4.3, 4.4, 4.5 depict classification performance for the MIAS and INbreast data sets for different SVM kernels. The area under the curve (AUC),  $A_z$ , signifies the classification performance and ranges between 0 and 1, with 1 being the highest performance. All reported results were averaged with 10-fold cross validation; they are presented in the tables below.

The results show that almost all SVM classifiers performed better with the LESH feature set. The prominent performance was achieved by the SVM linear classifier, which performed at a stable rate of 99.73% over all multiple feature selection sets (Table 4.3) and with an  $A_z$  value of  $0.9975 \pm 0.0010$  (Table 4.4) while using the INbreast data set. The results were verified for the MIAS data set when an accuracy of 100% was achieved by the SVM linear classifier (Table 4.5) and an  $A_z$  value of 1.0000 for 70 selected LESH features (Table 4.6). The second dominant classifiers were the polynomial and RBF; MLP showed an overall low performance accuracy.

The results in Table 4.7 verify that LESH is a good choice when discriminating between different types of abnormalities. The SVM linear classifier was capable of classifying mass or calcifications with no errors (Tables 4.7 and 4.8) for the INbreast data set. The polynomial classifier was second, followed by RBF, in terms of the performance score; MLP again showed a low performance. In the MIAS data set, a few cases existed for different abnormalities. Multiclass SVM classification was performed on the basis of the one-for-all scheme to distinguish between different types of abnormalities, as noted in Figure 4.2. Application of the proposed feature set in conjunction with SVM as well as ELM classifiers resulted in a higher performance (Table 4.9-4.10), which affirms the importance of LESH features. LESH features are invariant to the intensity of contrast exhibits in the mammogram image, and hence are suitable for medical image analysis. LESH is based on the phase congruency model, but is computed in a way that does not make them

rotation-invariant. Extending the research on how to compute them as rotation-invariant will be part of our future endeavours.

In summary, the results reported in Tables 4.3–4.9 suggest that LESH features have the power to emphasise important features in the images and that these features can be used to extract useful information. These features perform well with different kernel methods; selecting a subset from all available features does not deteriorate classification performance.

We conducted the experiments using the ESN classifier (described in section 2.4.5.2) code by Herbert Jaeger et al. [171]. The results have been reported using classification accuracy, sensitivity and specificity measures. We applied a 10-fold cross validation to authenticate the performance of our techniques. The results are displayed in the Table 4.14-4.15, which provides a summarised comparison of ESN with naive Bayes and k-nearest neighbours (kNN) classifiers. These results affirm the better performance of ESN as a classifier.

The selected LESH features,  $h_N$ , were fed to the benchmark machine learning classifiers and the results are reported in the tables (Table 4.3-4.10) below.

**Table 4.3. LESH-based classification accuracy % (malignant/benign) for INbreast data set.**

<b>Features Selected</b>	<b>SVM with RBF</b>	<b>SVM with Linear</b>	<b>SVM with MLP</b>	<b>SVM with Polynomial</b>
50	99.46	99.73	93.28	99.46
70	99.73	99.73	93.61	99.46
100	99.73	99.73	92.80	99.46

All	99.45	99.73	92.78	99.73
-----	-------	-------	-------	-------

**Table 4.4. LESH-based classification  $A_z$  value (malignant/benign) for INbreast data set.**

Features Selected	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
50	0.9955	0.9964	0.9472	0.9940
70	0.9975	0.9970	0.9487	0.9937
100	0.9980	0.9982	0.9440	0.9936
All	0.9955	0.9975	0.9448	0.9975

**Table 4.5. LESH-based classification accuracy % (malignant/benign) for MIAS data set.**

Features Selected	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
50	99.09	99.09	94.44	99.17
70	99.09	100.00	94.54	98.18
100	99.09	99.09	93.63	96.36
All	99.09	99.09	94.62	96.51

**Table 4.6. LESH-based classification  $A_z$  value (malignant/benign) for MIAS data set.**

Features Selected	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
50	0.9944	0.9944	0.9430	0.9900
70	0.9917	1.0000	0.9444	0.9733
100	0.9929	0.9929	0.9279	0.9616
All	0.9929	0.9936	0.9470	0.9583

**Table 4.7. LESH-based classification accuracy % (mass vs micro calcification) for INbreast data set.**

<b>Features Selected</b>	<b>SVM with RBF</b>	<b>SVM with Linear</b>	<b>SVM with MLP</b>	<b>SVM with Polynomial</b>
50	99.50	100.00	86.61	99.75
70	99.49	100.00	87.62	99.74
100	99.50	100.00	91.56	99.75
All	99.44	100.00	87.09	100.00

**Table 4.8. LESH-based classification  $A_z$  value (mass vs micro calcification) for INbreast data set.**

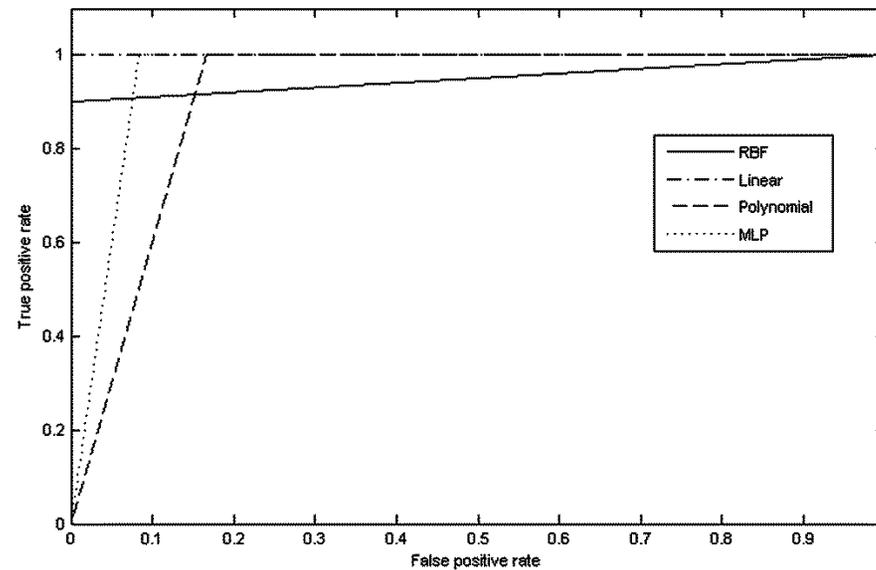
<b>Features Selected</b>	<b>SVM with RBF</b>	<b>SVM with Linear</b>	<b>SVM with MLP</b>	<b>SVM with Polynomial</b>
50	0.9967	1.0000	0.8990	0.9982
70	0.9964	0.9964	0.9172	0.9983
100	0.9965	1.0000	0.9057	0.9982
All	0.9960	1.0000	0.9087	1.0000

**Table 4.9. LESH feature with multiclass (one vs all) SVM classifier performance accuracy % for MIAS data set.**

<b>Abnormality Type</b>	<b>SVM with RBF</b>	<b>SVM with Linear</b>	<b>SVM with MLP</b>	<b>SVM with Polynomial</b>
ARCH	85.45	100.00	79.90	100.00
MISC	100.00	95.45	79.90	100.00
SPIC	95.67	82.17	90.00	95.45
CIRC	97.56	100.00	72.15	92.20
CALC	97.56	97.10	77.89	97.56

**Table 4.10. LESH-based features extraction binary classification (Using ELM ) results with 100 & 50 features selected in %.**

Performance Measure for ELM with 100 LESH features selected			Performance Measure for ESN with 100 LESH features selected			Performance Measure for SVM with 100 LESH features selected			Performance Measure for ELM with 50 LESH features selected			Performance Measure for ESN with 50 LESH features selected			Performance Measure for SVM with 50 LESH features selected		
Accur acy	Sensit ivity	Specif icity	Accur acy	Sensit ivity	Specif icity	Accur acy	Sensit ivity	Specif icity	Accur acy	Sensit ivity	Specif icity	Accur acy	Sensit ivity	Specif icity	Accur acy	Sensit ivity	Specif icity
97.34	98.71	97.56	99.05	100.0	97.22	97.00	92.58	100.0	98.18	97.87	97.32	84.29	88.11	81.28	99.17	98.83	99.70



**Figure 4.3. ROC for classification of malignant/benign cases for MIAS data set.**

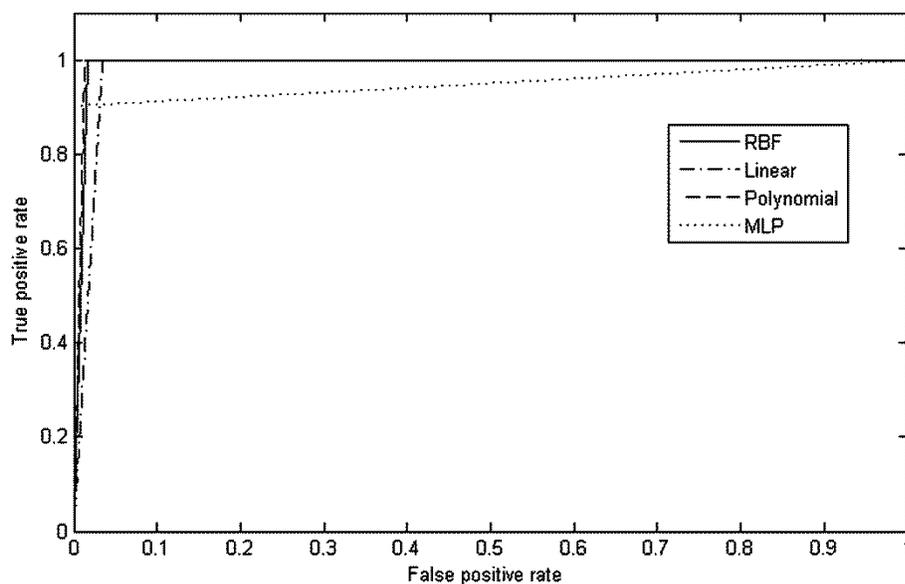


Figure 4.4. ROC for classification of malignant/benign cases for INbreast data set.

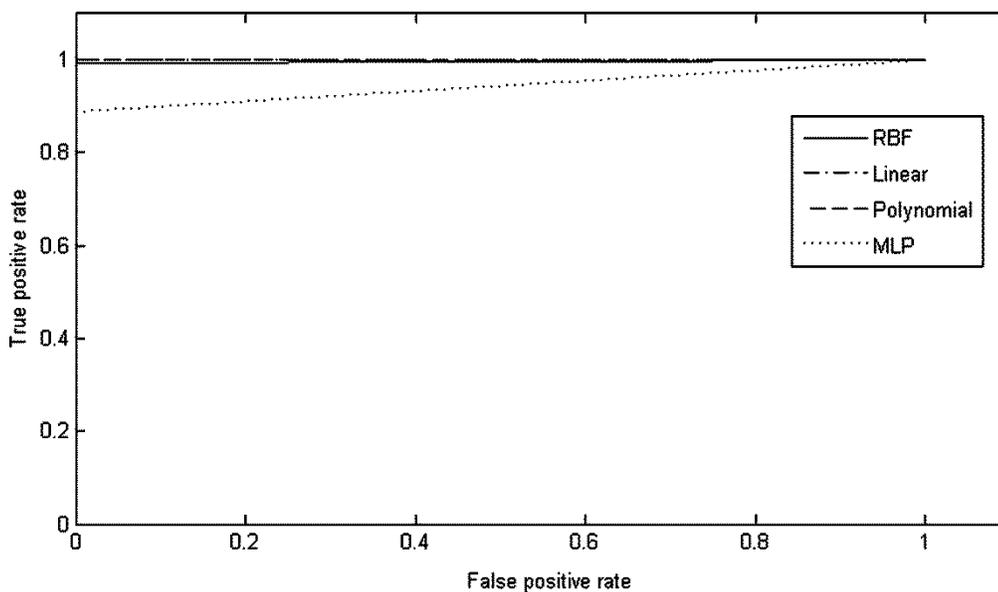


Figure 4.5. ROC for classification of masses/calcifications cases for INbreast data set.

#### 4.3.1.9 Comparative Analysis of LESH and Wavelet-based Feature Extraction for Classification Performance

In this study, LESH features were compared with wavelet transform-based feature extraction technique, which was used by Moussa et al. [106] and Cristiane et al.

[119]. In Cristiane et al. [119], the classification results of applying different types of wavelets to extract features from mammograms were compared. The *N*-largest wavelet coefficients were selected as feature vectors; later the SVM classifier was applied. These results were reproduced using Daubechies wavelets [37] for the 100 largest coefficients, as recommended by Cristiane et al. [119]. These results, shown in Tables 4.11–4.15, indicate superior performance using LESH features in both binary and multiclass classification cases.

**Table 4.11. Wavelet-based (Daubechies) feature extraction classification results for binary classification between malignant and benign cases for INbreast data set..**

Measure	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ACCURACY %	99.19	99.46	94.56	99.46
$A_z$	0.9928	0.9919	0.9433	0.9919

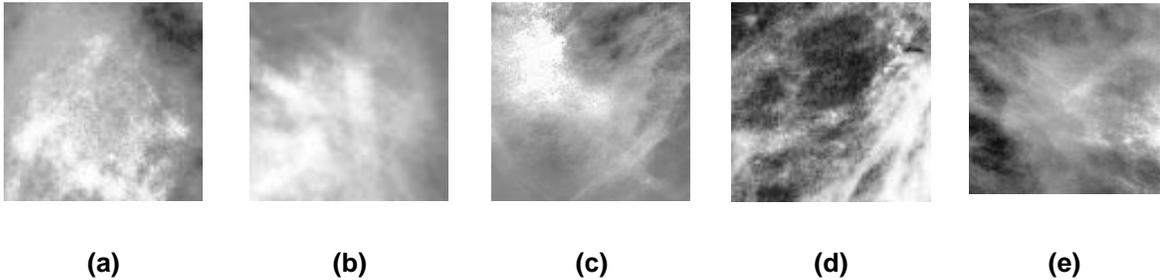
**Table 4.12. Wavelet-based (Daubechies) feature extraction classification (mass vs micro calcifications) results for INbreast data set.**

Measure	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ACCURACY %	98.73	100.00	95.67	99.75
$A_z$	0.9898	1.0000	0.9600	0.9980

**Table 4.13. Wavelet-based feature extraction for multiclass (one vs all) SVM classifier performance accuracy % for MIAS data set.**

Abnormality Type	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ARCH	80.10	94.10	80.10	92.44
MISC	78.34	76.67	86.56	71.89

SPIC	73.56	71.56	86.56	67.12
CIRC	67.12	79.90	46.15	98.00
CALC	76.67	100.00	96.10	100.00



**Figure 4.6** Samples of abnormalities in the MIAS database which have been misclassified (a) circumscribed mass (CIRC) mdb017, (b) ill-defined mass (MISC) mdb263, (c) speculated masses (SPIC) mdb179, (d) calcification (CALC) mdb231, and (e) architectural distortion (ARCH) mdb188.

Some cases of misclassification are shown in Figure.4.6. This misclassification is due to some of the texture pattern resembling within two or more different kinds of abnormalities. It was thus hard for the algorithm to discriminate between them e.g. in cases of circumscribed and speculated masses. On the other hand, different abnormalities may be found at the same region of interest e.g. a mass as well as a micro-classification are sometimes in the same place. Another possible reason for misclassification can be that the abnormality is covered in dense breast tissues so that its texture is not that recognizable.

**Table 4.14. Wavelet-based (Daubechies) feature extraction binary classification results.**

Performance Measure for ESN with all wavelet feature coefficients selected in %			Performance Measure for SVM with all wavelet feature coefficients selected in %			Performance Measure for ESN with 100 wavelet feature coefficients selected in %			Performance Measure for SVM with 100 wavelet feature coefficients selected in %		
Accuracy	Sensitivity	Accuracy	Sensitivity	Accuracy	Sensitivity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
98.04	98.30	98.04	98.30	89.60	98.30	58.27	74.93	66.06	94.90	92.95	97.61

**Table 4.15. Summarised results of other classifiers with lesh and wavelet features averaged over 10-fold cross-validation.**

Method	Performance Measure with 50 LESH features selected			Performance Measure with 100 LESH features selected			Performance Measure with 100 wavelet features selected			Performance Measure with all wavelet features selected		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
NaiveBayes	87.45	100.00	96.30	98.25	100.00	96.90	81.9	98.00	70.00	97.90	90.56	100.00
kNN	98.70	100.00	98.90	99.09	97.90	100.00	91.97	100.00	83.33	98.18	100.00	97.98

4.3.1.10 *Comparative Analysis of Classification Accuracy using LESH with log-Gabor (conventional LESH) vs. LESH with Gabor filter*

In conventional LESH feature extraction technique, the Fourier transform of the image is convolved with log-Gabor. We reformulated the framework LESH feature extraction by convolving the Fourier transform of the image with a simple Gabor filter in the frequency domain given as:

$$G(\omega) = \exp\left\{-\frac{\sigma^2}{2}(\omega - \omega_0)^2\right\} \quad (4.21)$$

Where  $\omega_0$  is the central radial frequency and the rest of the LESH framework is the same as described in Section 4.2.1. Some of the results of the Gabor filter-based LESH techniques are given in Tables 4.16-4.19 below.

**Table 4.16. Gabor-based LESH feature extraction-based classification results for binary classification (malignant/benign) for the MIAS data set.**

Measure	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ACCURACY %	80.08	94.45	90.00	95.45
$A_z$	0.8095	0.9475	0.8900	0.9500

**Table 4.17. Gabor-based LESH feature extraction-based classification results for binary classification (malignant/benign) for the INbreast data set.**

Measure	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ACCURACY %	99.73	99.46	91.44	99.73
$A_z$	0.9979	0.9958	0.9233	0.9979

**Table 4.18. Gabor-based LESH feature extraction-based classification (mass vs micro calcifications) results for the INbreast data set.**

Measure	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ACCURACY %	98.49	99.49	99.25	99.50
$A_z$	0.9964	0.9964	0.9923	0.9964

**Table 4.19. Gabor-based LESH feature extraction-based classification for multiclass (one vs all) SVM classifier performance accuracy % for the MIAS data set.**

Abnormality Type	SVM with RBF	SVM with Linear	SVM with MLP	SVM with Polynomial
ARCH	75.51	100.00	87.51	100.00
MISC	85.63	95.92	73.47	97.96
SPIC	71.43	91.56	87.76	93.88
CIRC	79.59	79.59	79.59	91.84
CALC	87.76	87.76	75.51	97.96

4.3.1.11 *Statistical Analysis of classification performance using log Gabor- and simple Gabor-based LESH feature extraction techniques.*

Given  $\mu_{lg}$  and  $\mu_g$  be the mean performance accuracies for SVM with Linear kernel

, when 100 LESH features were selected, we tested the hypothesis:

$$\begin{aligned}
 &H_0: \mu_{lg} - \mu_g = 0 \\
 &\text{Against} \\
 &H_a: \mu_{lg} - \mu_g > 0 \quad (4.22)
 \end{aligned}$$

**Table 4.20. The results of t-test at level of significance = 0.05.**

Method	Alternate Hypothesis $H_a$	P-value	T-value	Null Hypothesis $H_0$
--------	----------------------------	---------	---------	-----------------------

Method	Alternate Hypothesis $H_a$	P-value	T-value	Null Hypothesis $H_0$
MIAS (Malignant/Benign) case	$\mu_{lg} - \mu_g > 0$	0.001	2.31	Reject
INBreast (Malignant/Benign) case	$\mu_{lg} - \mu_g > 0$	0.031	1.85	Reject
INBreast Mass/Calc)	$\mu_{lg} - \mu_g > 0$	0.021	3.45	Reject

The results of the t-test (Table 4.20) suggest that the log Gabor-based LESH feature extraction technique performs better, when compared to Gabor filter-based LESH technique. This difference in performance is statistically significant; hence, the log Gabor-based LESH technique is a better choice for diagnosing malignancies.

#### 4.3.1.12 *Statistical Analysis of different classifiers performance using the t-test*

The t-test was applied to the performance accuracy resulting from 10-fold cross-validation applied to different feature sets for ESN and SVM classification results.

Let  $\mu_E$  and  $\mu_S$  be the mean performance accuracies for ESN and SVM classifiers, respectively. We tested the hypothesis:

$$\begin{aligned}
 &H_0: \mu_E - \mu_S = 0 \\
 &\text{Against} \\
 &H_a: \mu_E - \mu_S > 0 \\
 &\text{Or} \\
 &H_a: \mu_E - \mu_S < 0
 \end{aligned}
 \tag{4.23}$$

**Table 4.21. The results of t-test at level of significance = 0.05.**

Method	Alternate Hypothesis $H_a$	P-value	T-value	Null Hypothesis $H_0$
LESH with 100 features	$\mu_E - \mu_S > 0$	0.0901	1.4531	Accepted
LESH with 50 features	$\mu_E - \mu_S < 0$	0.0017	3.9322	Rejected
Wavelet with all features	$\mu_E - \mu_S > 0$	0.0004	4.9456	Rejected
Wavelet with 100 features	$\mu_E - \mu_S < 0$	0.0000	10.5156	Rejected

The result of the t-test is given in the tables above (Table 4.21). It also suggests that ESN performs as good as SVM, when 100 feature coefficients are selected from the LESH feature set, while SVM performs better when 50 feature coefficients are selected. Similarly, ESN performs better when all wavelet coefficients are fed to it as input, but its performance is weaker when a selected set of the largest wavelet coefficients is fed to it, in which case SVM performs better.

**4.3.1.13 Statistical Analysis of classifiers' performance for a different feature subset selection using the t-test**

To analyse the impact of a different feature subset selection on classification accuracy, we plotted the subset of the highest degree LESH coefficients selected vs. accuracy, using different classifiers in Figure 4.7-4.8. These plots indicate that there is a certain number of features which assist in diagnosing the abnormality with better performance.

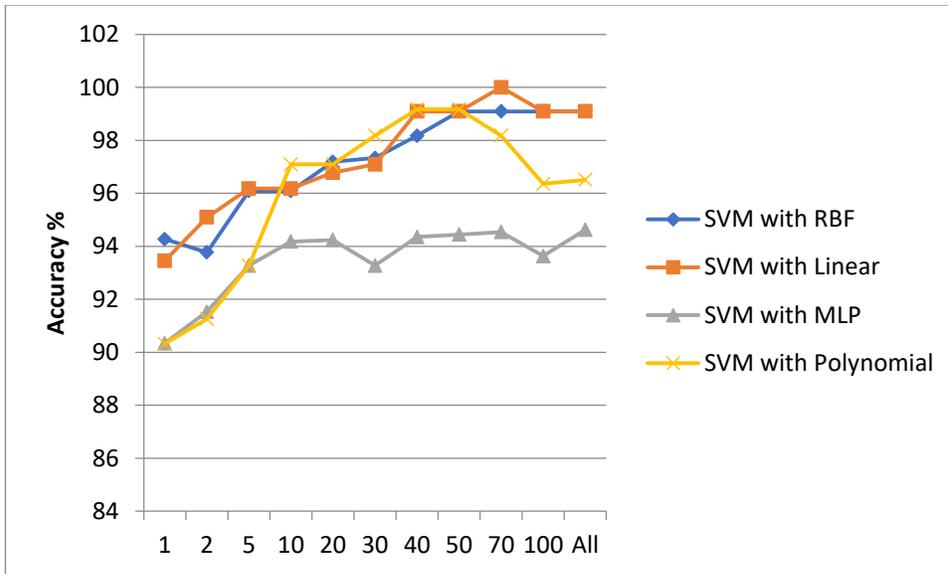


Figure 4.7. Classification accuracy vs number of features selection for INBreast dataset (malignant/Benign).

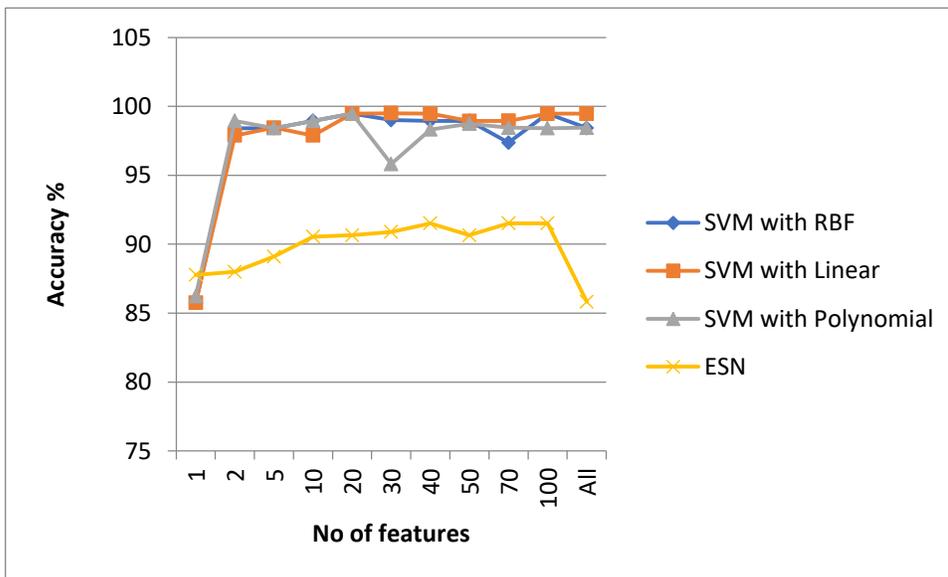


Figure 4.8. Classification accuracy vs number of features selection for MIAS dataset (malignant/Benign).

We applied a t-test to the performance accuracy resulting from selecting different feature subsets for the MIAS dataset. The results for a different subset of features

were computed using 10-fold cross-validation. For the experiment we selected only one feature set as (50, 70, 100, and all).

Let  $\mu_{50}$ ,  $\mu_{70}$ ,  $\mu_{100}$  and  $\mu_{all}$  be the mean performance accuracies for different feature subset selections. Then, we tested the null hypothesis against the alternate hypothesis, as given in Table 4.22 below:

**Table 4.22. The results of t-test at level of significance = 0.05.**

<b>Null Hypothesis <math>H_0</math></b>	<b>Alternate Hypothesis <math>H_a</math></b>	<b>P-value</b>	<b>T-value</b>	<b>Conclusion</b>
$\mu_{70} - \mu_{50} = 0$	$\mu_{70} - \mu_{50} > 0$	0.28	3.18	Accepted
$\mu_{100} - \mu_{70} = 0$	$\mu_{100} - \mu_{70} > 0$	0.03	2.54	Rejected
$\mu_{all} - \mu_{100} = 0$	$\mu_{all} - \mu_{100} > 0$	0.23	3.18	Rejected

The result of the t-test suggests that the 100 selected features are statistically significant in boosting the classification performance. The result can be further compared with each pair of feature subset selections, and we shall be keeping this as another of our future analysis endeavours, due to shortage of space and time.

### **4.3.2 Case Study II: Lungs Cancer Detection using JSRT Data Set of Chest Radiographs**

#### **4.3.2.1 Data Set**

Experimentation is conducted on a digital data set of chest radiographs provided by the Japanese Society of Radiological Technology (JSRT)<sup>2</sup>, which is available for free online. This data set contains images of 2048x2048 pixels, with pixel size being 0.175 mm. All images (with and without nodules) have been declared and classified as malignant or benign on the basis of histologic and cytologic examination. The data set consists of 154 lung nodules (100 malignant cases, 54 benign cases), and 93 non-nodules images [80].

#### **4.3.2.2 Image Pre-processing**

Chest radiographs have low contrast, which hinders detection of nodules in the images. We applied the CLAHE (Contrast Limited Adaptive Histogram Equalisation) method for image enhancement.

#### **4.3.2.3 Image Segmentation**

Next we applied the algorithm developed by Wong et al. [174] to extract lung areas from chest radiographs for application of algorithm which can discriminate between lungs with and without nodules. However, we applied manual segmentation to extract the ROI, i.e. region with nodules (from lungs with nodules), to further classify them as malignant or benign cases. Figure 4.9 shows a sample chest radiograph, its segmented form, and sample of extracted nodules. The ROI was

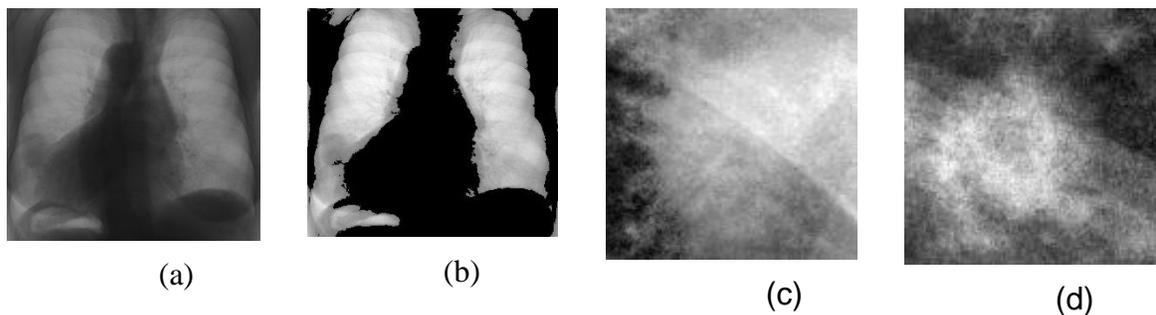
---

<sup>2</sup> <http://www.jsrt.or.jp/jsrt-db/eng.php>

extracted according to the information available in the database about the location of the nodule.

#### 4.3.2.4 *Feature Extraction and Selection*

From the segmented ROI, the histogram of the local energy information, LESH, was calculated for 16 sub-regions of segmented areas at 8 different orientations as the feature set. This resulted in a 128-dimensional feature vector ( $16 \times 8 = 128$ ). We selected a subset of  $N$ -largest coefficients,  $h_N$ , from LESH feature vectors,  $h_{r,b}$ , and fed them to the classifier.



**Figure 4.9.** Sample chest radiograph from JSRT database (a) sample radiograph (b) segmented image (c) sample ROI containing malignant nodule (d) sample ROI containing benign nodule.

#### 4.3.2.5 *Classification with selected Machine Learning Classifiers*

Classification is an integral part of CDSSs for prediction and diagnosis of a specific disease. It is a procedure to assign a class/label (from among the predefined classes) to new cases in a data set on the basis of the information that is incorporated into their feature set. We experimented with SVM, ESN and ELM machine learning classifiers and they are briefly described in section 2.4.5.

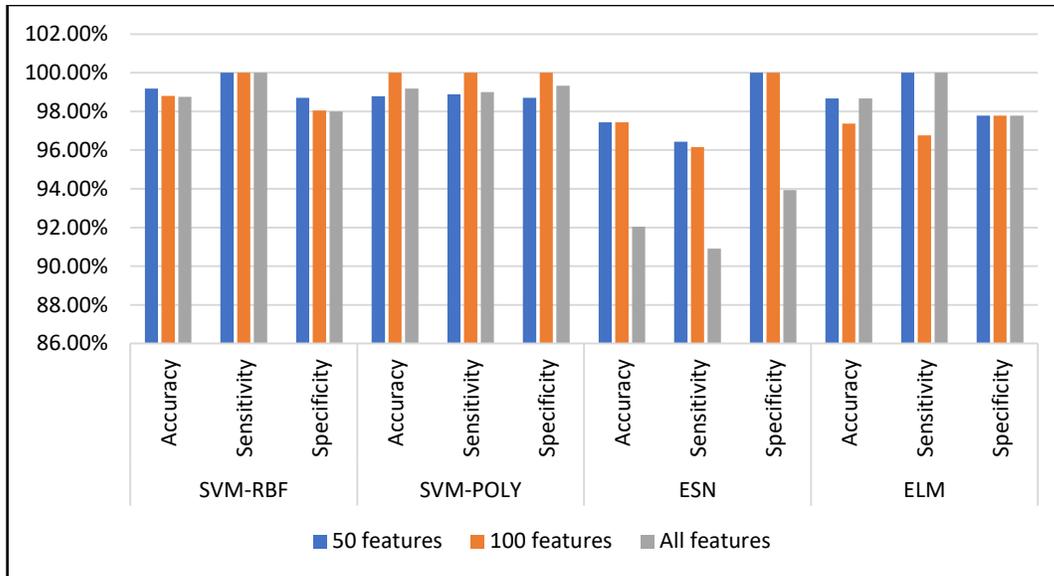
#### 4.3.2.6 *Results and Discussion*

We conducted experiments with JSRT chest radiographs to distinguish between lungs with and without nodules as well as to classify segmented nodules as malignant or benign. The results were evaluated using classification accuracy, sensitivity and specificity. Experiments were conducted using MATLAB. For application of ESN, we employed the version developed by Herbert Jaeger et al. [59], and, for ELM, we used the original model developed by G.B. Huang [19].

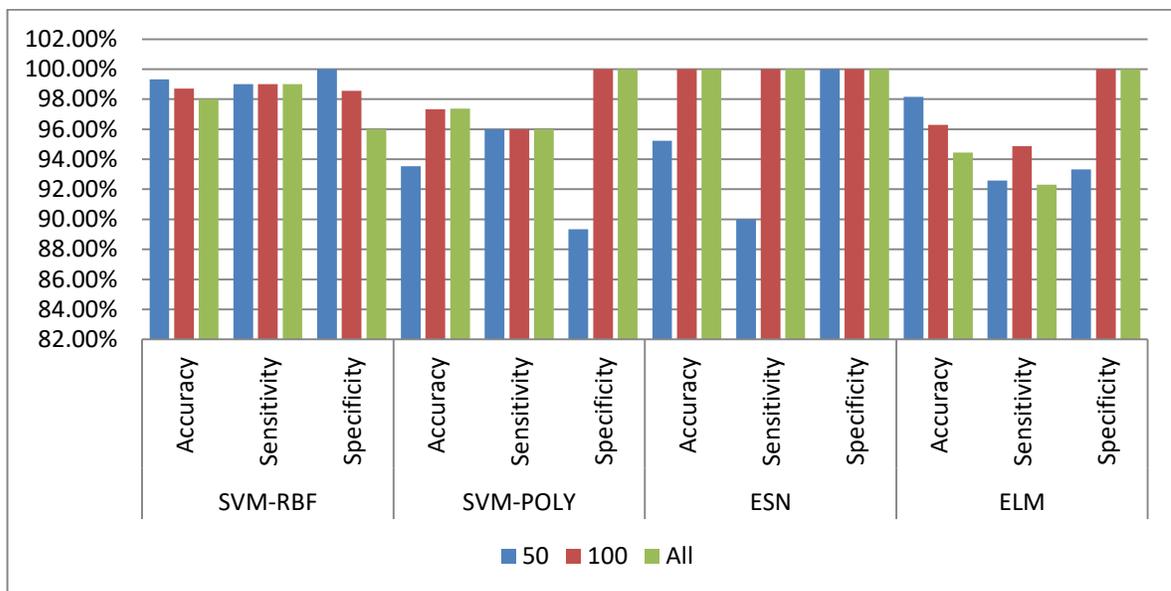
Once the classifier is trained with the training data set using 10-fold cross-validation, testing is performed and the results are evaluated using accuracy, sensitivity and the specificity performance measures (for different subsets of higher degree LESH coefficients) described and displayed in Tables 4.23 and 4.24. Although the wavelet-based feature extraction technique performs better in conjunction with SVM, when compared to any other classifier (Table 4.23), we note that LESH outperforms it- albeit with a little margin. As accurate detection of malignancies is critical in cancer diagnosis—to avoid unnecessary surgery—these results are considered as an improvement.

A comparison of the various classifiers' performances with respect to different subsets of selected features can be clearly seen in Figure 4.10 and Figure 4.11.

## A Novel Application of LESH-based CDSS...



**Figure 4.10. Comparison of different classifier performances with different-sized feature subset selections (nodule non-nodule lungs).**



**Figure 4.11. Comparison of different classifier performances with different-sized feature subset selections (malignant vs benign).**

**Table 4.23. LESH features extraction-based binary classification results for lungs with/without nodules (average of 10-fold cross-validation).**

Features Selected	Performance Measure for SVM with RBF			Performance Measure for SVM with Polynomial			Performance Measure for ESN Classifier			Performance Measure for ELM Classifier		
	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>
1	90.05%	93.00%	93.15%	91.00%	91.62%	92.00%	88.08%	89.20%	90.00%	90.01%	89.23%	90.53%
2	90.56%	93.00%	93.00%	94.05%	92.16%	93.15%	90.34%	93.56%	96.00%	93.34%	88.23%	90.23%
5	94.15%	95.66%	95.18%	94.34%	94.00%	94.34%	92.67%	93.56%	95.67%	95.45%	93.45%	94.67%
10	95.87%	97.00%	96.44%	94.33%	96.00%	95.13%	93.00%	94.33%	96.00%	96.00%	94.13%	95.00%
20	96.27%	96.00%	94.33%	94.13%	97.08%	95.13%	94.00%	95.15%	97.00%	95.37%	95.01%	97.08%
30	96.67%	97.00%	96.00%	95.42%	95.00%	96.00%	95.11%	94.33%	98%	96.71%	96.00%	97.08%
40	97.67%	97.00%	98.00%	97.42%	97.00%	98.00%	97.11%	96.33%	100%	97.71%	96.00%	99.08%
50	99.18%	100%	98.71%	98.78%	98.89%	98.71%	97.44%	96.43%	100%	98.68%	100%	97.78%
70	98.33%	99%	100%	96.75%	95%	100%	97.34%	98%	100%	97.37%	100%	98%
100	98.80%	100%	98.44%	100%	100%	100%	97.44%	96.15%	100%	98.67%	96.77%	97.78%
All	98.75%	100%	98%	99.18%	99%	99.33%	92.05%	90.91%	93.94%	98.68%	100%	97.78%

**Table 4.24. LESH features extraction-based binary classification results for malignant and benign nodules (average of 10-fold cross-validation).**

Features Selected	Performance Measure for SVM with RBF			Performance Measure for SVM with Polynomial			Performance Measure for ESN Classifier			Performance Measure for ELM Classifier		
	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>

A Novel Application of LESH-based CDSS...

	Accuracy	Sensitivity	Specificity									
1	90.29%	88.00%	100%	88.96%	83.00%	100%	89.23%	87.90%	94.46%	91.12%	94.00%	97.00%
2	93.08%	94.00%	97%	94.42%	93.00%	97%	90.44%	90.00%	94.45%	92.30%	91.18%	94.00%
5	93.75%	96.00%	100%	95.08%	94.00%	95.56%	91.45%	90.09%	92.55%	93.08%	94.00%	96.41%
10	95.00%	97.00%	100%	96.71%	95.00%	100%	95.56%	90.00%	95.60%	97.33%	96%	100%
20	96.38%	96.00%	100.00%	95.12%	97.00%	100%	95.33%	91.12%	97.45%	95.04%	97.00%	100%
30	97.33%	99.00%	100.00%	96.12%	96.00%	96.00%	96.33%	90.43%	98.00%	98.75%	98.00%	100.0%
40	98.67%	99.00%	98.00%	97.42%	97.00%	98.00%	97.11%	91.33%	100%	98.71%	98.00%	99.08%
50	99.33%	99%	100%	93.54%	96%	97.33%	95.24%	90%	100%	98.15%	92.59%	93.33%
70	97.33%	98.17%	99.09%	93.54%	97.21%	97.33%	96.78%	98.70%	99.09%	97.55%	97.00%	98.43%
100	99.71%	99%	100%	97.33%	96%	100%	100%	100%	100%	96.30%	94.87%	100%
All	98%	99%	96%	97.38%	96%	100%	100%	100%	100%	94.44%	92.31%	100%

**Table 4.25. Wavelet (daubechies) features extraction based binary classification results for lungs radiograph data set with 100 selected feature coefficients (average of 10-fold cross-validation).**

Classification	Performance Measure for SVM with RBF			Performance Measure for SVM with Polynomial			Performance Measure for ESN Classifier			Performance Measure for ELM Classifier		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Nodule/non-nodule	97.17%	96.71%	97.89%	98.73%	97.33%	100%	94.87%	87.50%	98.10%	94.74%	73.33%	98.34%

A Novel Application of LESH-based CDSS...

Classification	Performance Measure for SVM with RBF			Performance Measure for SVM with Polynomial			Performance Measure for ESN Classifier			Performance Measure for ELM Classifier		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Malignant/benign	98.67%	99%	98%	98.71%	98%	100%	95.24%	98.89%	97.75%	96.12%	92.45%	97.45%

**Table 4.26. Gabor-based LESH features extraction based binary classification results for lungs radiograph data set with 100 selected feature coefficients (average of 10-fold cross-validation).**

Classification	Performance Measure for SVM with RBF			Performance Measure for SVM with Polynomial			Performance Measure for ESN Classifier			Performance Measure for ELM Classifier		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Nodule/non-nodule	95.77%	97.00%	95.10%	96.05%	95.13%	98%	93.23%	88.60%	96.01%	93.75%	72.63%	97.35%
Malignant/benign	96.75%	95%	100%	97.46%	98%	96.67%	94.45%	96.19%	98.89%	95.52%	94.00%	97.03%

4.3.2.7 **Statistical Analysis of classifiers' performance for different feature subset selections using the t-test**

We compared the classification performance differences (taking SVM with RBF as the test case) due to different feature subset selection (50, 70, 100, and all) using the t-test and the results are given in Table 4.23-4.24 above. Let  $\mu_{50}$ ,  $\mu_{70}$ ,  $\mu_{100}$  and  $\mu_{all}$  be the mean classification performance accuracies with different feature subsets for differentiating between malignant and benign cases. Then, the null hypothesis tested against the alternate hypothesis is provided in Table 4.27 below:

**Table 4.27. The results of the t-test at level of significance = 0.05.**

<b>Null Hypothesis <math>H_0</math></b>	<b>Alternate Hypothesis <math>H_a</math></b>	<b>P-value</b>	<b>T-value</b>	<b>Conclusion</b>
$\mu_{70} - \mu_{50} = 0$	$\mu_{70} - \mu_{50} > 0$	0.20	3.18	Accepted
$\mu_{100} - \mu_{70} = 0$	$\mu_{100} - \mu_{70} > 0$	0.02	1.54	Rejected
$\mu_{all} - \mu_{100} = 0$	$\mu_{all} - \mu_{100} > 0$	0.14	3.18	Accepted

The above-mentioned t-test results suggest that N=100 selected features outperform the other selected feature sets in enhancing the classification accuracy for diagnosing malignant cases. Further analysis of the feature subset selection and its effect on classification accuracy for other classifiers is part of our future plan.

4.3.2.8 **Comparative Analysis of Classification Performance using log-Gabor based LESH Features (conventional LESH) vs. Gabor based LESH Features using t-test**

In the conventional LESH feature extraction technique, the Fourier transform of the image is convolved with the log-Gabor filter (section 4.2.1). We reformulated the framework for the computation of the LESH features by convolving the Fourier transform of the image with a simple Gabor filter given in equation (4.21) and the results are given in Table 4.26 above.

Given  $\mu_{lg}$  and  $\mu_g$  be the mean performance accuracies for SVM with RBF kernel, when 100 LESH features have been selected, we tested the hypothesis:

$$\begin{aligned} H_0: \mu_{lg} - \mu_g &= 0 \\ \text{Against} & \\ H_a: \mu_{lg} - \mu_g &> 0 \end{aligned} \quad (4.24)$$

**Table 4.28. The results of t-test at level of significance = 0.05.**

Case	Alternate Hypothesis $H_a$	P-value	T-value	Null Hypothesis $H_0$
(Malignant/Benign) case	$\mu_{lg} - \mu_g > 0$	0.004	2.31	Reject
Nodule/Non-Nodule	$\mu_{lg} - \mu_g > 0$	0.003	1.85	Reject

The results of the t-test (Table 4.28) suggest that in every case log-Gabor filter-based LESH techniques outperform simple Gabor-based LESH techniques. The test can further be expanded to analyse the performance of other classifiers.

#### 4.3.2.9 *Comparative Performance Analysis of LESH and Wavelet and SVM Classifiers using t-test.*

We compared the performance of the LESH feature extraction technique with that of the benchmarked wavelet features extraction technique suggested by Cristiane et al. [53]. They applied Daubechies wavelets transformed by first decomposing the image into four levels as prescribed by S. Mallat [119]. The decomposition was performed by convolving the image with low and high pass band filters. Afterwards, only low-level frequency coefficients were considered as features and fed to the classifier, since they demonstrate a higher capability to distinguish different patterns of abnormalities. We experimented with a selection of the largest 100 wavelet features so as to reproduce some results reported in Cristiane et al. [53], who experimented with different numbers of selected high-degree wavelet coefficients and applied SVM to classify them. Although wavelet based feature extraction technique performs better in conjunction with SVM as compare to any other classifier (Table 4.25), we note that LESH outperforms it- although with a little margin. As accurate detection of malignancies is critical in cancer diagnosis—to avoid unnecessary surgery—these results are considered as an improvement.

##### 4.3.2.9.1 t-test

We compared the performance of the LESH-based feature extraction technique with that of the wavelet-based technique using the t-test. It was accomplished by comparing the difference between two means of the resultant classification accuracy calculated using the LESH feature extraction technique (100 features

selected) and the wavelet feature extraction technique (100 features selected) at a significance level of 0.05.

Let  $\mu_L$  and  $\mu_W$  be the mean performance accuracies (mean is calculated from 10-fold cross-validation results) for LESH and wavelet classifiers, respectively. We tested the hypothesis:

$$\begin{aligned}
 &H_0: \mu_L - \mu_W = 0 \text{ (LESH performs same as wavelets)} \\
 &\quad \text{against} \\
 &H_a: \mu_L - \mu_W > 0 \text{ (LESH performs better than wavelets)} \\
 &H_a: \mu_E - \mu_S < 0 \tag{4.25}
 \end{aligned}$$

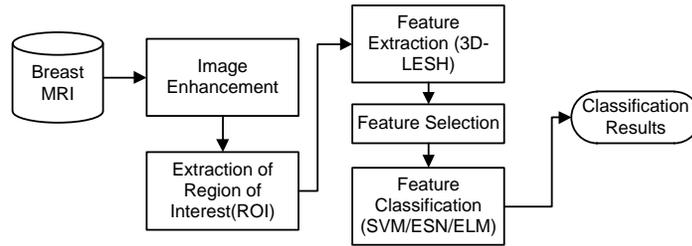
And the results are given in Table 4.29 below.

**Table 4.29. The results of t-test at level of significance = 0.05.**

Method	Alternate Hypothesis $H_a$	P-value	t-value	Null Hypothesis $H_0$
SVM (polynomial)	$\mu_L - \mu_W > 0$	0.47	1.83	Accept
SVM (RBF)	$\mu_L - \mu_W > 0$	0.19	1.83	Accept
ESN	$\mu_L - \mu_W > 0$	0.112	1.83	Accept
ELM	$\mu_L - \mu_W > 0$	0.45	1.83	Accept

### 4.3.3 Case Study II: 3D Breast MRI images

The workflow of the proposed breast cancer diagnosis scheme based on 3D LESH feature extraction technique is depicted in Figure. 4.12.



**Figure 4.12. Workflow of the proposed methodology.**

#### 4.3.3.1 *Breast MRI Data Set*

We collected 137 breast cancer MRI cases from the cancer atlas for breast cancer TCGA-BRCA data collection provided by The Cancer Imaging Archive (TCIA) from the Frederick National Laboratory for Cancer Research [81].

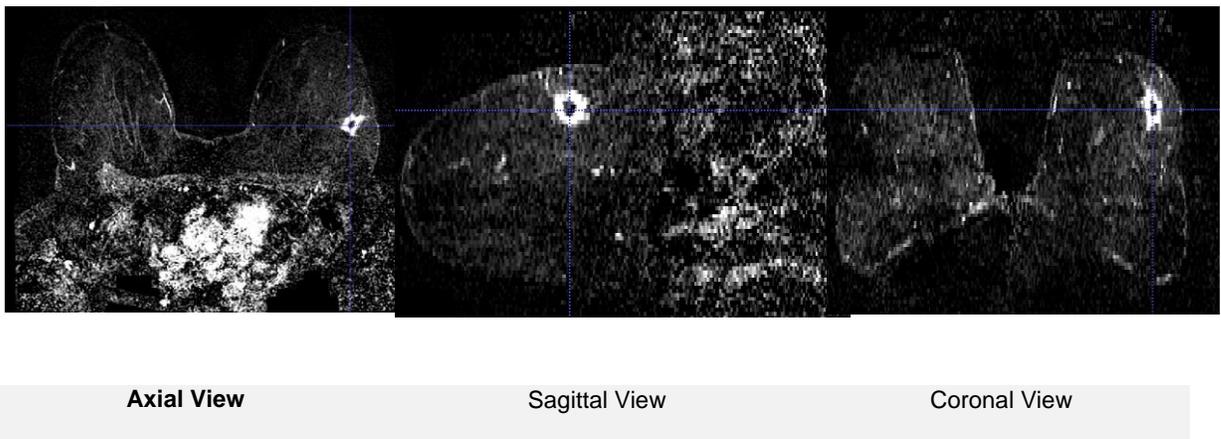
These MRI images are in Digital Imaging and Communications in Medicine (DICOM) format. DICOM combines images and metadata to create a rich description of a medical imaging procedure. The data set consists of T1-weighted, T2-weighted, pre-contrast, post-contrast images. The T2-weighted contrast-enhanced MR breast images were selected for the analysis.

#### 4.3.3.2 *Image Enhancement*

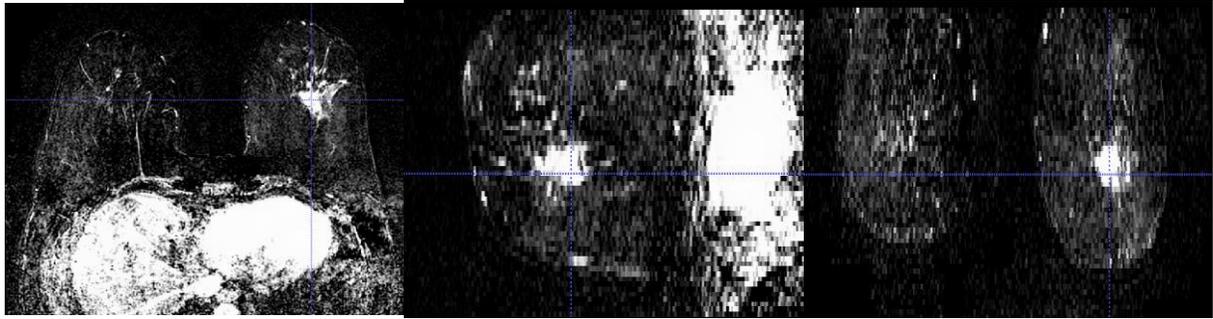
For improved diagnosis capability, MRI images were further enhanced with the Contrast Limited Adaptive Histogram Equalisation (CLAHE) technique.

#### 4.3.3.3 *Image Segmentation*

We extracted 193 (110 abnormal and 93 normal) regions of interest (ROI) from the TCGA-BRCA MRI images manually using the ITK-SNAP tool version 3.2.0 [85], according to the information provided about the location of the abnormality in the pathological reports for each patient. In the pathology report, cancer location and its stages are defined by the different tests performed by the physicians. These stages are ways to describe the cancer location, its growth and where it has spread. Once a stage is clear, it helps the doctor to decide which treatment is best to predict patient prognosis and recovery. The TCGA-BRCA data set contains MRI images samples for stages I–III. Figures 4.13, 4.14 and 4.15 show sample MRI images for stages I, II and III, respectively. With regard to the abnormal cases, most belong to stage II. The statistics of the abnormal ROI consist of 22 stage I, 70 stage II and 18 stage III cases.



**Figure 4.13. Sample T2-weighted MRI displaying Stage I from the TCGA-BRCA data set.**

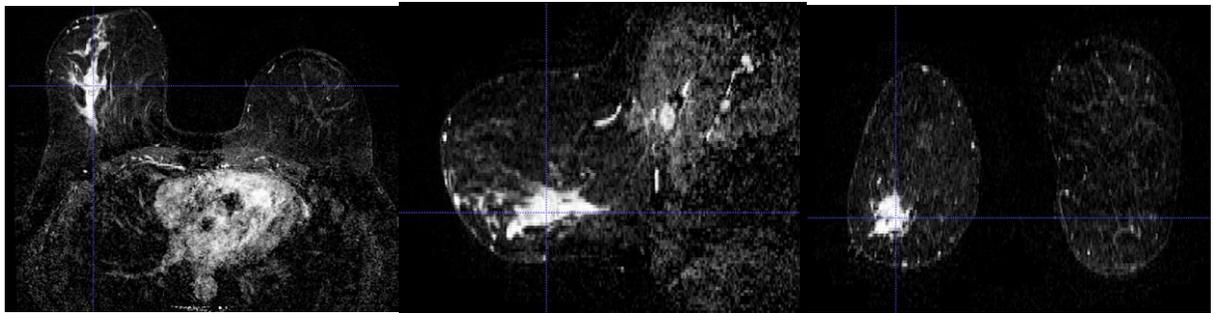


**Axial View**

**Sagittal View**

**Coronal View**

**Figure 4.14. Sample T2-weighted MRI displaying Stage II from the TCGA-BRCA data set.**



**Axial View**

**Sagittal View**

**Coronal View**

**Figure 4.15. Sample T2-weighted MRI displaying Stage III from the TCGA-BRCA data set.**

#### 4.3.3.4 *Feature Extraction*

We applied the LESH feature extraction technique to segmented ROI from MRI images and compared it with the benchmark wavelet-based feature extraction technique. The segmented ROI was divided into 16 sub-regions (each of size 32 X 32X16 voxels) further padded with zeroes to make it 32 x 32 x 32 size, thus calculating the Gaussian weight and hence the 3D-LESH (equation 4.19-4.20) for each sub-region. Furthermore, the LESH features were computed for each sub-region along 8 different orientations, accumulating into a 128-dimensional LESH

feature vector ( $16 \times 8 = 128$ ). This distribution of the voxels is depicted in Figure 4.16 below.

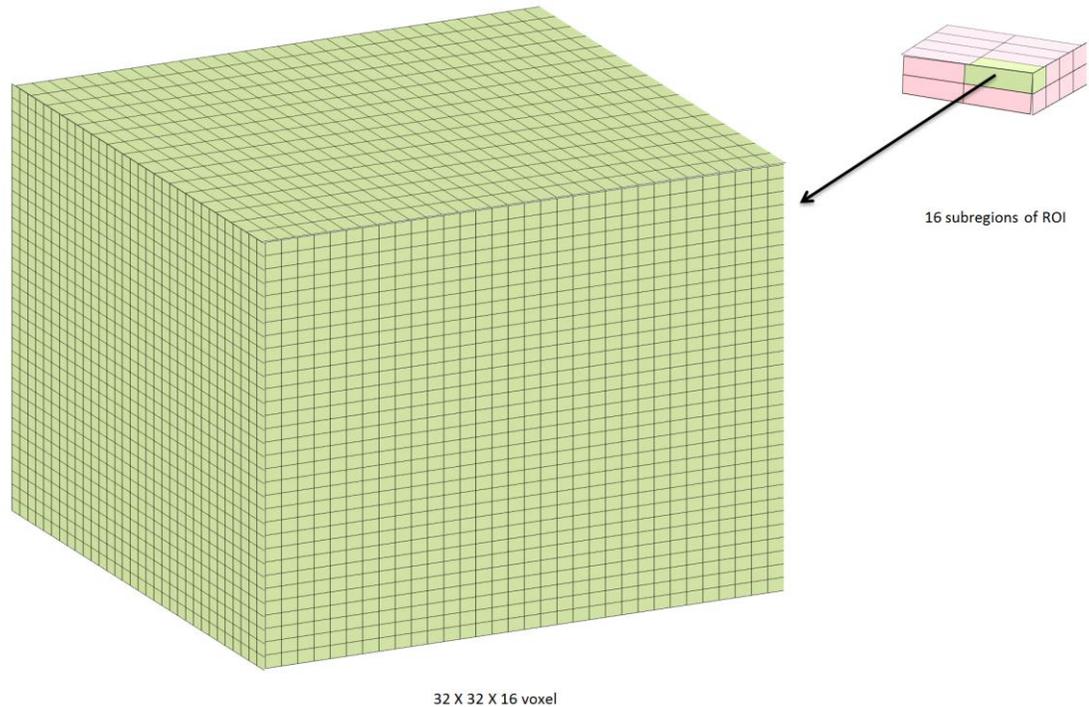


Figure 4.16. Graph depicting the breakdown of the ROI into sub-regions, for 3D LESH feature vector computation.

#### 4.3.3.5 *Feature Selection*

We studied the impact of selecting a subset of  $N$ -largest LESH coefficients, say  $h_N$ , from LESH feature vectors,  $h_{r,b}$ , for classification purposes. These selected features coefficients were fed to the classifier, to distinguish between different stages and between normal and abnormal cases.

#### 4.3.3.6 *Classification with Selected Machine Learning Classifiers*

Classification is an integral part of CDSSs for prediction and diagnosis of a specific disease. It assigns a class/label (from among the predefined classes) to new cases in a data set on the basis of the information incorporated into their respective feature set.

The selected LESH features,  $h_N$ , were fed to the classifiers SVM, ESN and ELM, and their performance was evaluated using classification accuracy and (AUC)  $A_z$  value.

#### 4.3.3.7 *Results and Discussion*

Experiments were conducted with TCGA-BRCA data sets to distinguish abnormal and normal cases and to classify abnormalities according to their stages in MRI images. We reported using classification accuracy and the ROC curve plots to measure the significance of the results generated.

We ran a grid search with 10-fold cross-validation to determine suitable parameters for SVM classifiers. In the case of SVM with the Radial Basis Function (RBF) kernel, the slope parameter  $\gamma$  and the penalty parameter  $C$  were tuned, while in the case of SVM with the linear kernel, only  $C$  was tuned. The polynomial kernel of degree two was employed for the experimentation. The results are shown in Tables 4.30–4.32. All reported results were averaged with 10-fold cross-validation.

The overall results indicate that LESH features combined with SVM classifiers provide efficient classification performance (Tables 4.30 & 4.31). The significant performance was achieved by the SVM linear classifier, which performed at the classification accuracy of 99.47% when 100 or all LESH features were selected (Table 4.30) and with an  $A_z$  value of  $0.9956 \pm 0.0046$  (Table 4.31) for classifying normal and abnormal cases. The SVM with RBF kernel also performed well by reaching the accuracy of  $98.95 \pm 0.52$  (Table 4.30) and  $A_z$  value of  $0.9909 \pm 0.066$  (Table 4.31). The ELM also resulted in the high classification accuracy of 99.47% (Table 4.30) and  $A_z$  value of  $0.9856 \pm 0.0136$  (Table 4.31) overall. Finally, SVM with the polynomial kernel reached the maximum performance of 98.48 % and  $A_z$  value of 0.9909, when 100 LESH feature were selected, and similarly ESN showed the maximum performance of 91.52 % and  $A_z$  value of 0.9226 with 100 selected LESH features.

The results in Table 4.32 verify that LESH is a good choice for detecting the cancer stage. In this respect, ELM outperformed any other classifier by distinguishing between different stages at the rate of 95.45% for stage I, 98.18% for stage II and 98.18% for stage III. The SVM linear classifier was capable of distinguishing between different stages of the breast cancer with maximum accuracy of 88.09% for stage I, 73.45% for stage II and 78.18 % for stage III (Table 4.32). The polynomial classifier was second in overall performance, followed by RBF, with its performance score; ESN again showed a low performance.

In all cases, multiclass classification was performed based on the one-for-all scheme to distinguish between different stages of abnormalities. Application of the proposed feature set in conjunction with machine learning classifiers affirmed the importance of LESH features.

In summary, the results reported in Tables 4.30–4.32 emphasise the superiority of LESH features for distinguishing between different stages of cancer, as well as detecting the existence of cancer in the medical images. Further, it is evident that selecting a subset from all available features does not deteriorate classification performance significantly. The ROC curve for different classifiers depicted in Figure 4.17 provides a comparison of the overall performance of these classifiers.

**Table 4.30. LESH-based classification accuracy % (abnormal/normal) for MRI data set.**

Features Selected	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
50	98.95	98.94	98.74	90.65	97.89
70	97.37	98.95	98.45	91.52	98.01
100	99.47	99.47	98.42	91.52	98.92
All	98.43	99.47	98.45	85.83	99.47

**Table 4.31. LESH-based classification Az value (abnormal/normal) for MRI data set.**

Features Selected	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
50	0.9909	0.9954	0.9859	0.8946	0.9758
70	0.9700	0.9650	0.9598	0.9222	0.9701
100	0.9954	0.9954	0.9909	0.9226	0.9894

## A Novel Application of LESH-based CDSS...

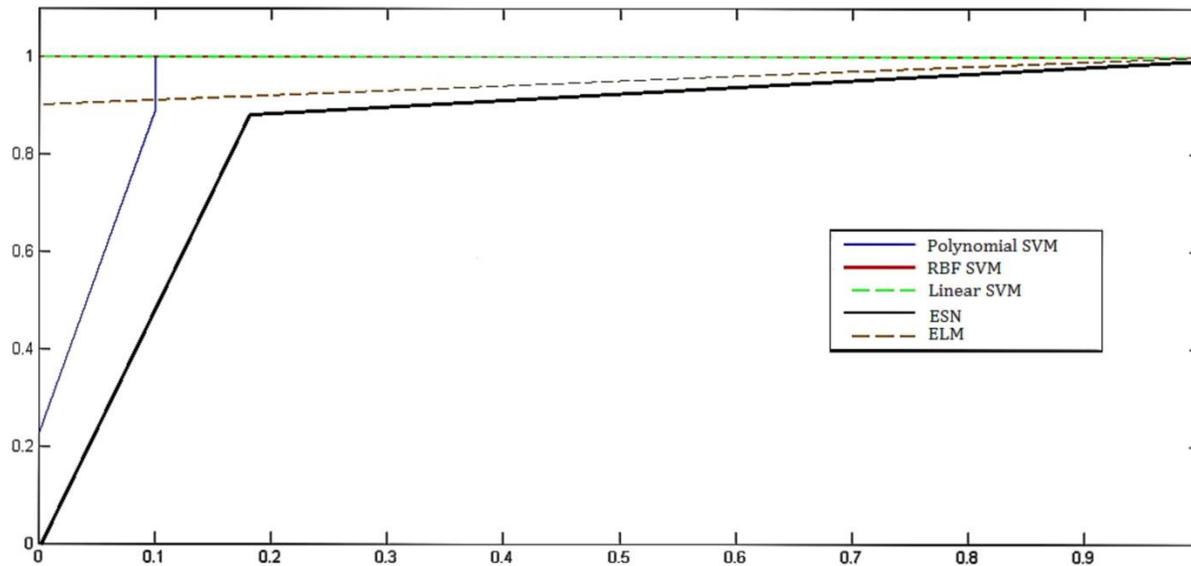
---

All	0.9843	1.0000	0.9864	0.8990	0.9861
-----	--------	--------	--------	--------	--------

---

**Table 4.32. Selected LESH feature (50,100 & all) with multiclass (one vs all) SVM classifier performance accuracy % for MRI data set.**

Stage	SVM with Linear			SVM with RBF			SVM with Polynomial			ESN			ELM		
	50	100	All	50	100	All	50	100	All	50	100	All	50	100	All
I	88.09	71.72	96.09	72.72	85.37	89.36	60.90	83.86	93.09	86.87	80.89	89.99	97.27	95.45	95.45
II	73.45	72.72	86.36	77.27	53.63	84.46	67.27	77.72	81.81	73.63	81.78	86.79	97.27	95.45	98.18
III	78.18	70.91	91.09	77.73	68.18	86.54	81.18	83.36	83.36	87.90	84.45	92.34	96.36	95.45	98.18



**Figure 4.17. ROC for classification between Abnormal/Normal cases for MRI data set (Linear SVM, and RBF SVM curves are overlapping, whereas Polynomial SVM curve is partially overlapping and thus some part of it is invisible.)**

4.3.3.8 **Statistical Analysis of classifiers performance for different feature subset selection using the t-test**

To observe the impact of different highest degree 3D-LESH coefficients upon the classification performance, we plotted these coefficients vs. classification performance accuracy (Figure 4.18). It can be seen that there is a certain range of coefficients where the 3D-LESH features perform consistently, while a certain set of features may assist in boosting the classification performance significantly.

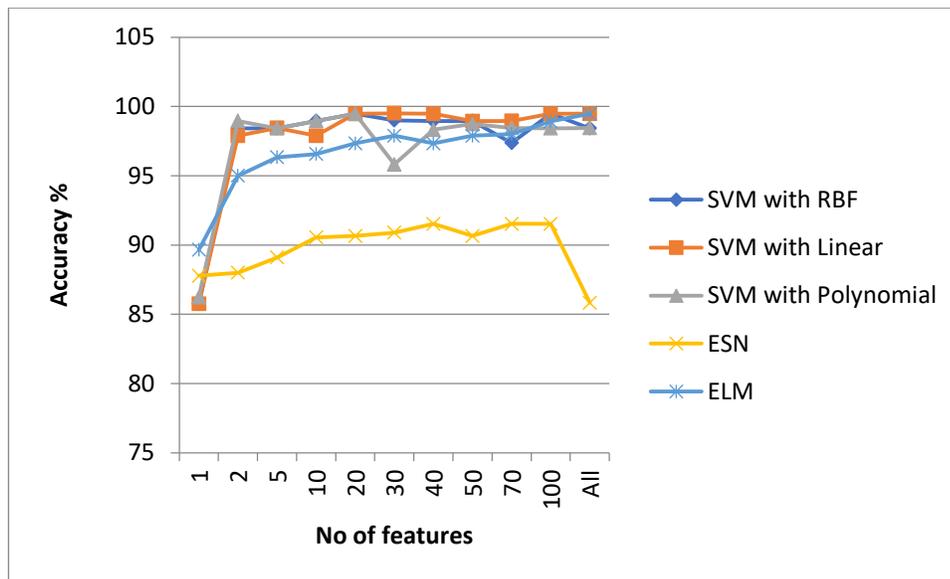


Figure 4.18. No. of LESH Features selected vs. Accuracy for the MRI dataset.

Furthermore, we compared the classification performance differences due to different feature subset selection (50, 70, 100, and all) using the t-test and the results are given in Table 4.33 below. Let  $\mu_{50}$ ,  $\mu_{70}$ ,  $\mu_{100}$  and  $\mu_{all}$  be the mean classification performance accuracies with different feature subsets for differentiating between malignant and benign cases. Then, the null hypothesis against the alternate hypothesis is provided in the Table 4.20 below:

**Table 4.33. The results of the t-test at level of significance = 0.05.**

<b>Null Hypothesis <math>H_a</math></b>	<b>Alternate Hypothesis <math>H_a</math></b>	<b>P-value</b>	<b>T-value</b>	<b>Conclusion</b>
$\mu_{70} - \mu_{50} = 0$	$\mu_{70} - \mu_{50} > 0$	0.34	2.13	Accepted
$\mu_{100} - \mu_{70} = 0$	$\mu_{100} - \mu_{70} > 0$	0.04	2.34	Rejected
$\mu_{all} - \mu_{100} = 0$	$\mu_{all} - \mu_{100} > 0$	0.17	2.15	Accepted

The above-mentioned t-test results suggest that N=100 selected features outperform the other selected feature set in enhancing the classification accuracy for diagnosing malignant cases. Further analysis of the feature subset selection and its effect on classification accuracy is part of our future plan.

**4.3.3.9 Comparative Analysis of the log-Gabor- and Gabor-based LESH Feature Extraction Technique for Classification Performance using t-test.**

We revised our experiments with the 3D-LESH feature extraction technique by replacing log-Gabor filter convolution with Fourier transform, with that of the Gabor- based convolution, where the Gabor filter is given as:

$$G(\omega, \theta) = \exp\left\{-\frac{\sigma^2}{2}(\omega - \omega_0)^2\right\} \times \exp\left(-\frac{\alpha(\theta)^2}{2\sigma_\alpha^2}\right) \quad (4.26)$$

Where parameters for the equation (4.26) are defined in Section 4.2.2.1, in detail. Tables 4.34-4.35 below summarise the results produced by the application of the Gabor-based LESH feature extraction technique, when used for classification.

**Table 4.34. Gabor-based LESH feature extraction-based classification results for binary classification between normal and abnormal cases for the MRI data set.**

Measure	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
ACCURACY %	99.47	99.47	97.92	90.00	95.99
$A_z$	0.9960	0.9856	0.9613	0.8623	0.9615

**Table 4.35. Gabor-based LESH feature extraction-based multiclass (one vs all) classification performance accuracy % for the MRI data set.**

Abnormality Type	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
STAGE I	82.50	98.75	82.50	87.87	94.89
STAGE II	81.25	80.00	77.50	83.72	96.72
STAGE III	97.50	97.56	81.25	95.46	97.54

Further application of the t-test for comparing the performance of the log-Gabor-based LESH feature extraction technique and the Gabor-based feature extraction technique is given below.

Let  $\mu_{lg}$  and  $\mu_g$  be the mean performance accuracies for the log-Gabor-based LESH and the Gabor-based LESH techniques, when classification is performed using SVM with Linear kernel. We tested the hypothesis:

$$\begin{aligned}
 H_0: \mu_{lg} - \mu_g = 0 & \text{ (log-Gabor-based LESH performs same as Gabor-based LESH)} \\
 & \text{against} \\
 H_a: \mu_{lg} - \mu_g > 0 & \text{ (log-Gabor-based LESH performs better than Gabor-based LESH)} \\
 & \text{(4.27)}
 \end{aligned}$$

**Table 4.36. The results of the t-test at level of significance = 0.05 for binary classification.**

Method	Alternate Hypothesis $H_a$	P-value	t-value	Null Hypothesis $H_0$
SVM (Linear)	$\mu_{lg} - \mu_g > 0$	0.0019	3.19	Reject
SVM (polynomial)	$\mu_{lg} - \mu_g > 0$	0.004	2.34	Reject
SVM (RBF)	$\mu_{lg} - \mu_g > 0$	0.003	1.89	Reject
ESN	$\mu_{lg} - \mu_g > 0$	0.0012	2.33	Reject
ELM	$\mu_{lg} - \mu_g > 0$	0.007	3.20	Reject

The results in Table 4.36 suggest that log-Gabor-based 3D-LESH features have higher capability to discriminate among malignant and benign cases, when compared to Gabor-based 3D-LESH features.

**4.3.3.10 Comparative Analysis of LESH and Wavelet-Based Feature Extraction for Classification Performance**

We compared the performance of the LESH feature extraction technique with the benchmarked wavelet feature extraction technique applied by Cristiane et al. [53] [30]. We extended this technique to three-dimensional MRI images with selection of the largest 100 wavelet features, as this seemed an appropriate choice for reducing the curse of dimensionality while keeping the performance intact. The results, presented in Tables 4.37 and 4.38 below, indicate obvious superior performance with LESH features use in both binary and multiclass classification.

**Table 4.37. Wavelet (Daubechies) feature extraction-based classification results for binary classification between normal and abnormal cases for MRI data set.**

Measure	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
---------	--------------	-----------------	---------------------	-----	-----

ACCURACY %	71.36	72.42	73.47	50.50	93.28
$A_z$	0.7460	0.7568	0.7659	0.5443	0.9249

**Table 4.38. Wavelet-based feature extraction for multiclass (one vs all) classification performance accuracy % for MRI data set.**

Abnormality Type	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
STAGE I	70.90	66.36	68.18	66.67	85.45
STAGE II	57.27	83.63	55.45	49.80	84.54
STAGE III	88.18	93.63	93.63	75.00	85.98

4.3.3.11 *Statistical Analysis of the Results with t-test*

The *t-test* was applied to calculate the significance of the improvement in accuracy by using LESH features compared with the wavelet texture features. We selected ELM-based classification accuracy for the test, since it has a higher performance ratio in both cases. The other classifiers' performances may also be tested, but we leave this for future research because of limited space. Let  $\mu_L$  and  $\mu_W$  be the mean performance accuracies for LESH and wavelet classifiers (mean is calculated from 10-fold cross-validation results), respectively. We tested the hypothesis:

$$\begin{aligned}
 &H_0: \mu_L - \mu_W = 0 \text{ (LESH performs same as wavelets)} \\
 &\quad \text{against} \\
 &H_a: \mu_L - \mu_W > 0 \text{ (LESH performs better than wavelets)} \quad (4.27)
 \end{aligned}$$

The results are given in Table 4.39 below.

A 0.05 level of confidence was used for experiments, which is standard for statistical testing. The results for binary classification (abnormal/normal) are shown in Table 4.39 below. The probability value (p-value) determined whether the null hypothesis should be rejected. In both binary and multistage cases, the p-value was higher than the 0.05 confidence level, which suggests that the difference between the classification accuracy was prominently significant. Thus, it can be stated that our proposed 3D LESH feature extraction technique performs significantly better than the wavelet-based technique. Other feature extraction methods may also be compared with our technique in future.

**Table 4.39. The results of t-test at level of significance = 0.05 for binary classification.**

Method	Alternate Hypothesis $H_a$	P-value	t-value	Null Hypothesis $H_0$
SVM (Linear)	$\mu_L - \mu_w > 0$	0.00004	9.39	Reject
SVM (polynomial)	$\mu_L - \mu_w > 0$	0.000019	7.45	Reject
SVM (RBF)	$\mu_L - \mu_w > 0$	0.000002	9.43	Reject
ESN	$\mu_L - \mu_w > 0$	0.000004	8.94	Reject
ELM	$\mu_L - \mu_w > 0$	0.038	2.002	Reject

#### 4.4 Conclusion & Discussion

The correct diagnosis is vital in designing an efficient CDSSs since FP as well as FN are both harmful. FP can cause unnecessary biopsies or other treatments that can be expensive and painful for the patients. Similarly, FN can cause an actual cancer to go undiagnosed, and thus may lead to mortality. A combination of

resilient methods at each stage of CDSSs help to enhance overall system performance, as evident from our proposed CDSSs described in this chapter.

LESH has been proved to be a superior feature descriptor. As mentioned earlier, LESH is a histogram of local energy, which is at a maximum level at the abrupt change of image intensity. Accordingly, it marks the texture variations in the local area. The highest degree coefficients correspond to the most significant and prominent set of features in the local area within an image. A subset of these features can be selected with some compromise on classification accuracy, which hence reduces the dimensionality. Experiments were performed with different numbers of the highest degree coefficients and determined that  $N = 100$  is the most appropriate number for improving classification accuracy. Overall, the results suggest that LESH generates an effective set of features that improves classification performance compared to a state-of-the-art wavelet-based feature selection approach, which is also evident from results above.

The reason for that is that although wavelets have efficient image representation they fail to represent discontinuities along curves and edges. Other competitive features are curvelets and contourlets. Curvelets are superior to wavelets by sustaining edges. Contourlets are closer to curvelets and are counted as a discrete form of curvelet transforms but have less clear directional features [177]. Other than wavelets, many feature extraction techniques have been mentioned in section 3.2.4. Most of these techniques focus their study on a specific type of abnormality and try to detect its malignancy, whereas LESH-based methodology has the

advantage of detecting the malignancy of any type of abnormality and can differentiate between different types of abnormalities with considerable efficiency.

When compared with a state-of-the-art wavelet-based feature selection approach, the overall results suggest that LESH generates an effective set of features that improves classification performance. This is due to the inability of wavelet features to represent discontinuities along curves and edges [177].

In lung cancer detection, CDSSs face multiple challenges due to inhomogeneous lung regions; similarity in density of arteries, ribs, bronchi, veins, and bronchioles in the lung region; and different types of nodule shapes like cavities and ground glass nodules [175]. All these make lung and nodule segmentation, feature extraction and classification cumbersome tasks [175]. Nodule characteristics such as shape, size, texture etc. can aid in the predicting their malignancy. Other factors which may add to system performance in detecting abnormalities in lungs include image acquisition and reconstruction parameters, the nodule location, data-set size, and optimisation of the system using cross-validation [176].

Most state-of-the-art methods employed by researchers to extract features from lung nodules ROI fail to represent discontinuities along curves and edges, and hence are incapable of presenting a resilient set of feature vectors that may help in the classification of different types of abnormalities. However, our proposed LESH feature extraction technique is capable of marking significant pattern variations in the medical images efficiently in both cases of classification, i.e. for distinguishing

lungs with/without nodules and for discriminating between malignant and benign nodules (Table 4.23 & 4.25).

Lastly, the chapter also presented a novel technique for detecting breast cancer in volumetric medical images based on a three-dimensional (3D) LESH model. It is a hybrid approach, which combines the 3D LESH feature extraction technique with machine learning classifiers to detect breast cancer from MRI images. The proposed system applies contrast-limited adaptive histogram equalisation (CLAHE) to the MRI images before extracting 3D LESH features. Further, a selected subset of features is fed to a machine learning classifier, namely support vector machine (SVM), extreme learning machine (ELM) or echo state network (ESN), to detect abnormalities and to distinguish between different stages of abnormalities. The results indicate the high performance of the proposed system. When compared with the wavelet-based feature extraction technique, statistical analysis testifies to the significance of our proposed algorithm.

The next chapter is a step further to enhancing the performance of each CDSSs described here through the novel application of the artificial immune network (AIN)-based hyper-parameter optimisation technique.

## **Chapter 5 - Nature-Inspired Algorithm for Machine Learning Classifier Optimisation**

The last chapter details our experiments and results related to the application of the CDSS framework (discussed in chapter 3) upon three different genres of datasets. We further extended these experiments to observe the implications of a novel AIN-based optimisation technique on enhancing the diagnostic capabilities of conventional CDSS. It was accomplished by the application of an Artificial Immune Network (AIN) for optimising machine learning classification algorithms incorporated as an integral part of CDSS. AIN accomplishes this task by searching for the best hyper-parameter set for a specific machine learning classification algorithm (also called model selection).

AIN belongs to the field of natural computing which has its roots in the workings of the immune system. Natural computing is the process of extracting ideas from nature to develop a computational model for problem-solving [192]. The research in this field aims to devise a theoretical model of the natural phenomenon that can be implemented in computers to reproduce the qualitative methods to solve complex problems, which cannot be satisfactorily solved by conventional techniques like linear, non-linear or dynamic programming. It works by simplifying the mechanisms presented in natural phenomena and so helps to make large numbers of entities traceable and highlight the minimal features necessary to enable some particular features of the system to be reproduced and observe some emergent approaches. Among the many approaches within computing inspired by

nature, the most well-known ones are the artificial networks, evolutionary algorithms, swarm intelligence and artificial immune systems [192].

In this chapter, we aimed to theoretically explore the immune system and apply AIN to solve the problem of machine learning classifiers optimisation. The rest of the chapter is organised as follows: section 5.1 describes some background knowledge of natural immune systems, its computational representation as an artificial immune system 5.2 describes the proposed AIN algorithm for optimising machine learning classifiers. 5.3 provides simulation results of the application of this optimisation technique using diverse types of datasets. Finally, the chapter ends with the discussion section, 5.4.

### **5.1 The Biological Immune System (IS)**

A natural immune system consists of cells and molecules that work together with our body to maintain a dynamic internal state of equilibrium. The immune system has the ability to resist disease-causing agents called pathogens and toxic substances called antigens. These pathogens can be viruses, bacteria, fungi etc. The primary role of the immune system is to detect these pathogens and eliminate them [193]. Cells (also called agents) in innate immune systems detect the pathogens by considering their specific molecular pattern, which can never be found in the host body cells. Once pathogens are detected, innate immune system cell signals (chemical signals) join with other immune cells to fight against pathogens [192].

The immune system has an adaptive property since they evolve through encounters with antigens. The immune system cells are also categorised as B-cells and T-cells (lymphocytes), which together help us to recognise and destroy specific substances/antigens produced and are capable of generating responses [192].

The immune system clones B-cells in an effort to remove infectious antigens. In cloning, B-cell goes through somatic hyper mutation to remove infection as well as prepare for any future attacks [192].

The system retains a memory of encounters with pathogens, which is recalled on further exposure to the same pathogens.

### **5.1.1 Immune System Concepts**

The immune system is comprised of B-cells (receptors which can recognise antigens) and T-cells (cells identifying antigens residing on the surface of other cells). Many theories have been proposed in literature to explain the immune system's responses to antigens. Most important of them are clone selection [194] and affinity maturation [196, 197] for developing an adaptive immune system for problem-solving [192].

#### **5.1.1.1 Clonal Selection**

This process produces a clone of B-cell or T-cell, which is capable of recognising an antigen-stimulus. This reproduction process may perform a mutation in proportion to cell affinity to the antigen parent cell [192].

#### 5.1.1.2 *Immune Memory*

Immune memory is acquired and maintained using clone expansion and selection and an immune network. In this network highly stimulated B-cells survive and less stimulated B-cells are removed [192].

#### 5.1.1.3 *Immune Network*

Jerne [195] suggested that B-cells exhibit a mutually reinforcing network, which stimulates as well as suppresses each other to avoid over population of B-cells to get a stable memory. In a self-regulatory fashion, a self-organising system is made up of immune cells and molecules, which interact with each other and the environment to produce dynamic behaviour [192].

#### 5.1.1.4 *Learning in Immune Networks*

This is a cognitive process which learns and responds as (1) recognises molecular shapes (2) preserves encounters with antigens (3) defines boundaries for networks (4) can and infer about future antigen patterns. In producing dynamic patterns of activities to regulate and maintain networks, these are metadynamics (which control the production and death of immune cells). The metadynamic has an oscillating pattern of production and death, which stabilises over time. It preserves the identity of IS overtime but is still allowed to adopt to a new situation [192]

#### ***5.1.1.1 Self/Non-Self Discrimination***

For IS to be in proper function it should be able to distinguish between its own cells and foreign cells, which are basically indistinguishable, otherwise it will kill its own antibodies and cause autoimmune diseases. So negative selection is vital.

We discuss the simulation of the immune system for computational modelling below.

### **5.2 Artificial Immune Systems (AIS)**

AIS are inspired by theoretical immunology and observe the immune function model. They encompass the algorithms that are developed using the ideas and metaphors from the biological immune system to solve problems [192].

The researchers, inspired by AIS, are developing algorithms to solve complex component problems. AIS have been applied in the area of fault diagnosis, computer security, virus detection, etc. Its main model comprises of the following features:

1. Recognition of malfunctioning self-cells and harmful non-self cells.
2. Feature extraction: Antigen Presenting Cells (APC) in IS extracting features from disease-causing cells (antigen) while removing noise from them and presenting them to other cells (lymphocytes).
3. Diversity: involves generation and maturation whereby the immune system adapts (somatic hyper mutation) to the foreign elements that confront it.

4. Learning: comprised of a process which helps the immune system to fine-tune its response over time.
5. Affinity maturation: B-cell produces antibodies (novel patterns) for increased affinity to antigens. Fighting with more antigens will increase affinity in antibodies. Affinity maturation consists of mutation and selection, and makes IS better at pattern recognition.

### 5.2.1 Immune Engineering Framework to design AIS

Castro and Timmis [192] proposed a layered framework for AIS, as below:

1. Population generation: representation of the components of the application domain
2. Affinity maturation: a mechanism to represent and evolve interaction among different system components and its environment: set of input, output and fitness function etc.
3. Algorithm to define dynamics (behaviour of the system): the examples are clonal selection, negative positive selection, immune network algorithms and bone marrow.

We will focus on immune networks here as our algorithm to optimise the machine learning classification is based on the immune network paradigm. These networks are described below.

### 5.2.2 Artificial Immune Networks (AIN)

This is a dynamic system, where the cells recognise each other and present some patterns of dynamic behaviour even without any foreign intervention. The network approach is helpful in developing computation tools due to its properties like learning, memory, self-tolerance, the diversity of cell population, and its interaction with the self and the environment. In general, the dynamics of continuous as well as discrete networks can be depicted as [192]:

$$RPV = NSt - NSu + INE - DUE \quad (5.1)$$

Where RPV is the rate of population variation, NSt is network simulation, NSu is the network suppression, INE is the influx of new elements and DUE is the death of unstimulated cells [192]. We further discuss a special algorithm based on the immune network paradigm below.

#### 5.2.2.1 *aiNet*

The aiNet model comprises a set of cells and molecules. The interconnection between them is represented as the degree of similarity between them: the closer they are, the more similar they are considered. Whenever the body is attacked by pathogens (antigens (Ag)), it responds by secreting antibodies (Ab), which help in recognising the antigens. The strength of the interaction between Ab and Ag is measured by the affinity of their match. This means that the cells (Ab) which can recognise antigen (Ag) better have a higher antigen affinity and hence are further increased in concentration and affinity, while the ones with lower affinity die out

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

over time. The *aiNet* algorithm [198] assumes a population of antigens to be recognised by the set of antibody cells in the network. A real-valued vector in a Euclidean shape space called network cells is used to represent the antibodies and antigens together. Their affinity is calculated using Euclidean distance. A set of cells with higher affinity are selected for cloning. These cloned cells undergo a somatic mutation that is inversely proportional to their antigen affinity rate. Finally, a certain number of affinity clones are maintained as network memory [192].

Affinity of the rest of the antibody cells is calculated and cells with affinity less than a certain threshold are eliminated from the network (clonal suppression).

To produce diversity, a set of antibody cells are generated randomly, and their affinity is computed with existing cells. Out of these new born cells, the one with affinity above a certain threshold is the only one added to the population and the rest are eliminated [192].

The *aiNet* learning algorithm is summarised here [198]:

1. Initialisation: generate a set of random population of network antibodies:
2. Antigenic presentation: repeat for each antigenic:
  - 2.1 Clonal selection and expansion: in this process, each cell is cloned according to its affinity rate. The higher the affinity, the more chances there are for the cell to be cloned.

2.2 Affinity maturation: the cloned cells are further mutated inversely proportional to its affinity rate while a set of the highest affinity clones are placed in clone memory.

2.3 Metadynamics: the memory clones with affinity of antigens less than a threshold are discarded.

2.4 Clonal Interaction: it is the computation of measure of affinity among clone cells.

2.4 Clonal suppression: clones with affinity less than a pre-specified suppression threshold are discarded.

3. Network interactions: it is the computation of measure of affinity among each pair of network antibodies (cells).

4. Network suppression: the network antibodies with an affinity rate of less than a certain threshold are discarded.

5. Diversity: a random set of new antibodies is generated and incorporated into the network population

6. Loop: repeat steps 2 to 4 until a certain condition is met.

### **5.3 Artificial Immune Network System-Based Optimisation**

Given a dataset  $U = (u_1(n), u_2(n), \dots, u_K(n))$  that consists of K feature vectors with their corresponding class labels  $Y' = (y'_1, y'_2, \dots, y'_k)$ , our goal is to optimise (using the

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

artificial immune network paradigm) the classification accuracy (fitness), which is given as follows:

$$Fitness = Accuracy (U) = \frac{\sum_{i=1}^K Assess(u_i)}{K} \quad u_i \in U \quad (5.2)$$

$$Assess(u_i) = \begin{cases} 1 & \text{if } classify(u_i) = \text{correct classification} \\ 0 & \text{if } classify(u_i) = \text{incorrect classification} \end{cases} \quad (5.3)$$

Where  $classify(u_i)$  is correct if  $y'_i(u_i)$  in the dataset is the same as the corresponding output label  $y_i(u_i)$ .

Let us consider the following example for understanding the calculation of accuracy measure. Suppose our sample dataset is given as in Table 5.1 below:

**Table 5.1. Example dataset**

$u_i$ for $i: 1$ – 10	Attribute1	Attribute2	Attribute3	Class Label $y_i(u_i)$ .	Predicted Class Classify( $u_i$ ) $= y'_i(u_i)$	$Assess(u_i)$
1	Yes	Large	128k	1	-1	0
2	No	Medium	120k	1	1	1
3	No	Small	100k	-1	1	0
4	Yes	Medium	234k	-1	-1	1
5	Yes	Small	456k	1	-1	0
6	No	Large	678k	-1	-1	1
7	No	Large	123k	1	1	1
8	Yes	Medium	455k	1	-1	0
9	Yes	Small	678k	-1	-1	1
10	Yes	Small	789k	1	-1	0

Given the above dataset, and the equation 5.2, Accuracy is given as  $=5/10=0.5$ .

As alternatives to using the accuracy, the receiver operating characteristic (ROC), the area under ROC (AUROC), sensitivity, specificity or other measures can be

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

taken as the fitness function. In the case of different classifiers, the output label  $y_i(u_i)$ .

The goal of obtaining the generalisation performance  $Fitness = Accuracy(U)$  of a specific classifier can be achieved via the application of cross-validation. This technique splits a given dataset  $(U, Y)$  into disjoint sets  $(U(train), Y(train))$  and  $(U(test), Y(test))$ . Thus, the proposed optimiser is trained over the training set; then, the classification performance is calculated as the mean of all of the performances applied to the testing set  $(U(test), Y(test))$ . The overall system structure to optimise a classifier through AIN can be depicted as follows:

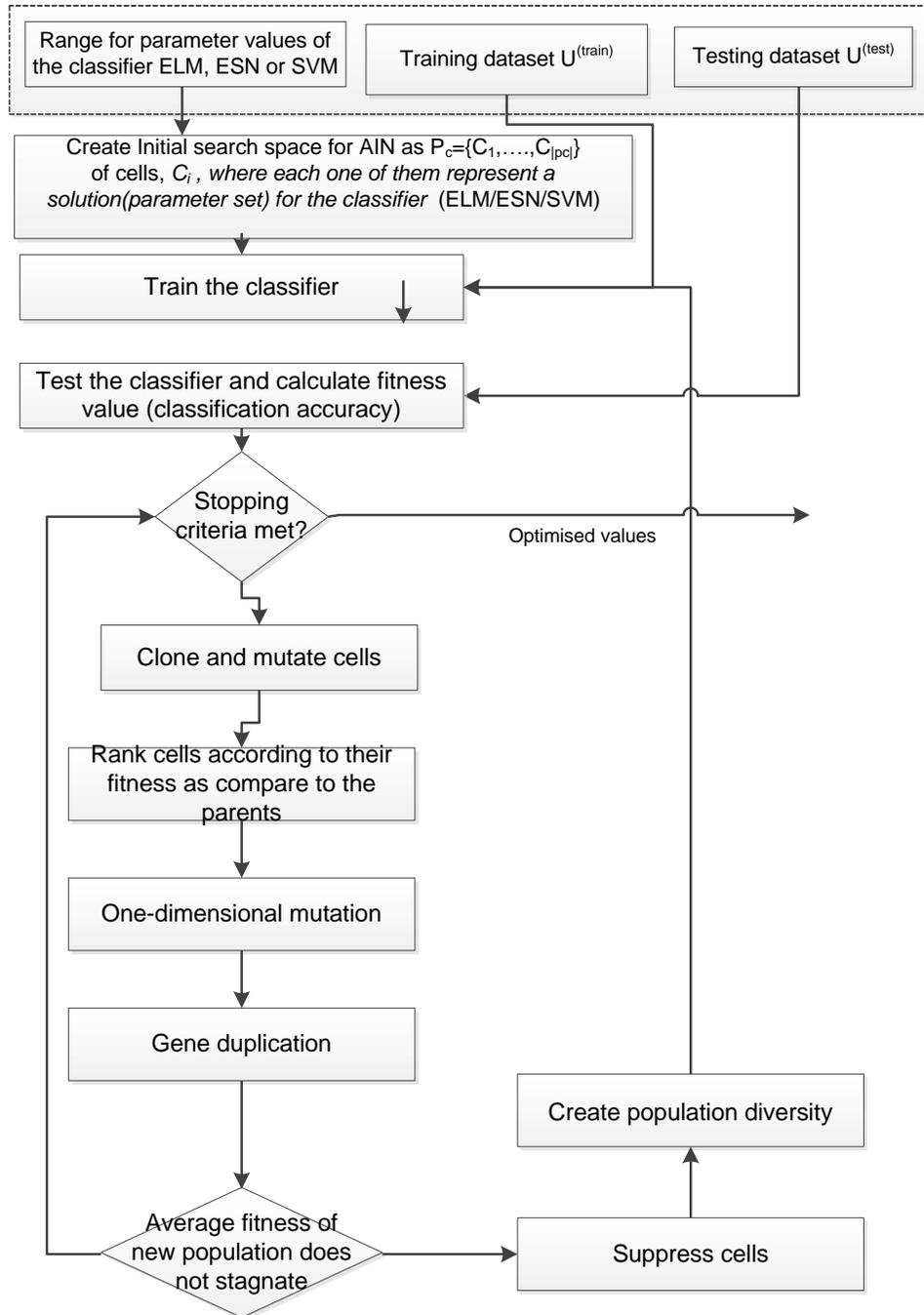


Figure 5.1. The overall system structure of the proposed algorithm.

### 5.3.1 Create the initial Search Space Population of the components of the application domain

The first step in solving the optimisation scheme is to create an abstract model of the application domain parameters as immune cells and molecules (AIN terminology) [192]. Perelson and Oster devised a shape-space approach to present them by using a data structure that is specific to an application. This data structure consists of string attributes that can be real-valued vectors, integer strings, binary strings or symbolic strings. The data structure should be sufficient to quantify the affinity between each pair of attributes [199]. Our data structure is a real-valued vector  $C$  (also called a network cell) that represents specific classifier parameters as a single feasible solution to the optimisation problem.

- 1) In the case of ELM, we optimised the input weights only; thus, the network cell is given as follows:

$$C = [a_{min}, a_{max}] \quad (5.4)$$

Where  $[a_{min}, a_{max}]$  are limited within the range of  $[-10, 10]$  for the input weight range.

- 2) In the case of the Echo State Network,

$$C = [(W_{min}^{in}, W_{max}^{in}), (W_{min}, W_{max}), (W_{min}^{out}, W_{max}^{out}), S_{in}, S, S_{out}, M] \quad (5.5)$$

Where  $(W_{min}^{in}, W_{max}^{in}), (W_{min}, W_{max}), (W_{min}^{out}, W_{max}^{out})$  are limited to the range of  $[-10, 10]$  for the input, output and reservoir weight range, respectively;  $S_{in}, S, S_{out}$  are the sparsity ratio for the input, reservoir and output weight matrices,

respectively, to be in the range of [0, 1], and M is the number of neurons or the reservoir size (network size), which is to be in the range of [1,300]. These values are generated randomly and are considered to be basic ESN parameters.

3) In the case of the SVM- RBF kernel

$$C = [\gamma, bc] \quad (5.6)$$

Where  $\gamma$  is a parameter within the range  $[2^{-10}, 2^{-9}, \dots, 2^4]$ , and  $bc$  is the box constraint within the range of  $[2^{-2}, 2^{-1}, \dots, 2^2]$ , respectively.

In all of the above cases,  $P_c = \{C_1, \dots, C_{|P_c|}\}$  is the set of the current network cell population that was generated, which represents a set of possible solutions.

### 5.3.2 Train the Classifier

For each cell  $C$ , a training dataset is used to train the specific classifier.

### 5.3.3 Calculate the Fitness

The testing dataset is used to calculate classification accuracy. For each cell  $C$  in the population  $P_c$ , the fitness is calculated using the performance accuracy measure given in “(5.3)”.

### 5.1.2 Clone and Mutate the Cells

For each cell in the population  $P_c$ , the algorithm generates a number  $N_c$  of clones [200]. Each of these clones is further mutated in such a way that the mutation  $C'$  of an original cell  $C$  is:

$$C' = C + \alpha N(0,1) \quad (5.7)$$

$$\text{Where } \alpha = (1/\beta) \exp(-f^*) \quad (5.8)$$

Here  $N(0,1)$  is a Gaussian random variable with a mean of zero and standard deviation of one, and  $f^*$  is the fitness function ( $Accuracy(U, C')$ ).  $\beta$  is the mutation parameter, which controls the function decay [200].

#### 5.3.4 Rank Cells

Each mutated cell  $C'$  is ranked according to its fitness compared to its parent. If its performance degrades, then it remains at 0 rank; otherwise, its rank is incremented by 1 [201]. Cells with ranks of above zero are selected for the subsequent operations.

#### 5.3.5 One-dimensional Mutation

The higher ranked cells are further mutated using golden section search [202]. In a traditional mutation, to obtain the best value for  $\beta$ , analysis of the function landscape is required but is not always possible. França et al. suggested a golden section search to find the value of  $\beta$  that will solve the problem of having slow or no convergence in the case of traditional mutation. Golden section search works by repeatedly subdividing the interval between two cells while looking for the section that has better performance. The process continues until a predefined criterion is met, i.e. the smallest length of the interval is met [203].

Given a cell  $C'$  and an Identity vector  $I$ , the golden section search will divide the area between them in search of the value that gives the best fitness of  $C'$ , which

will be selected to be  $\beta$ . Furthermore, equations (1, 2) are applied to generate the mutated cells.

### 5.3.6 Gene Duplication

Next, from the current population of the search space, a cell  $C_i$  is randomly chosen, and it replaces every other cell  $C_j$  if its rank is higher:

*For each cell  $C_i$  in  $P_c$  do*

*If  $C_i$  performs better than  $C_j$*

*Replace it in the original population.*

*End*

*End*

The process is used in chromosome transaction [204].

### 5.3.7 Suppress Cells

If the fitness value is not stabilised, then we suppress the cells that have the same fitness. The algorithm follows the principal of evolutionary strategies to select a mutated offspring such that a parent cell survives unless it is suppressed by one of the offspring [201]. It is a new method to suppress cells and avoid the chance of selecting more than one cell at the same peak of the fitness. The method considers the fitness of a cell while deciding whether to suppress it. Given the cells  $C_1$  and  $C_2$  from the search space population  $P_c$ , we let  $q_1 = (C_1, \text{fitness}(C_1))$ ,  $q_2 = (C_2, \text{fitness}(C_2))$  and  $q' = \text{projection of } q_{1/2} \text{ onto } q_2$ , where  $q_{1/2} = (q_1 + q_2)/2$ . The

interaction/affinity between the cells is computed based on the distance between  $q_{1/2}$  and a point  $q$ , which is computed as:

$$q' = \begin{cases} q + \left(\frac{v \cdot w}{\|v\|}\right) v & \text{if } q' \text{ falls inside segment } \overline{q_1 q_2} \\ q_1 & \text{if } q' \text{ falls outside segment } \overline{q_1 q_2} \text{ and close to } q_1 \\ q_2 & \text{if } q' \text{ falls outside segment } \overline{q_1 q_2} \text{ and close to } q_2 \end{cases} \quad (5.9)$$

Where  $v = q_2 - q_1$  and  $w = q_{1/2} - q_1$ . The above-mentioned work is based on the network suppression concept of AIN, where an affinity measure performs a mapping of the interaction between two cell vectors into a real-valued number, where recognition among the cells is considered to be proportional to the similarity: the smaller the distance is, the better the recognition [192]. Here, if  $dist(q_{\frac{1}{2}}, q')$  is less than threshold  $\sigma_s$ , then the cell that has the worst fitness value is removed, based on the consideration that they are close to each other in terms of their fitness.

### 5.3.8 Create Population Diversity

A random set of new cells is generated and added to the search space population  $P_G$ , to diversify it [200]. The above-mentioned process continues until the stopping condition is met, which can be a maximum number of iterations or stopping when the fitness stagnates, and we assume that we reach the optimum solution.

The resulting optimum values of the classifier hyper-parameters are reported in Table 5.1.

## 5.4 Simulation and Results

Given a dataset, we split it into training, validation, and test sets. Furthermore, the training dataset was fed to the machine-learning classification algorithm, with the known class labels, to build up a model from it [228].

It is a well-known fact that searching for the optimal set of values for the parameters using training set results causes overfitting, which means that we end up finding a relationship which does not hold for the dataset in general [228].

To avoid overfitting, a part of the training dataset is kept for learning (used to optimise the model) and validation (model performance is validated on this data for each run / for each hyper-parameter values). A validation set is used so that hyper-parameters with better performance on the validation set are selected (also called 'the hold-out method') [229]. The rest of the test set is kept aside and is used to find the general performance of the learned machine-learning classifier. Here, for each instance of the test set, the predicted value is compared again with the real class label [229]. In our case we split the dataset into 60% for training set, 20% for the validation set and 20% as test set for each of the experiments reported in this chapter.

We applied the proposed algorithm to two benchmarked machine learning datasets as well as on two-dimensional as well as three-dimensional medical images datasets.

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

We developed the algorithm in Matlab R2010a and executed it multiple times on a system that had an Intel Pentium i7- running at 2.3 GHz and 16 GB RAM.

In the case of ELM, we had the choice of tuning the input weights  $a_i$  as well as the bias  $b_j$ . We ran the AIN optimiser to optimise only one as well as both of them. We also experimented with a multiple range set to determine the best range for the search space for tuning the parameters, and it turned out that  $[-10, 10]$  is the appropriate range to search for the input weights. Similarly, in the case of ESN, we attempted to optimise the input, output and reservoir weights as well as the sparsity ratio and reservoir size. The appropriate range for the weights was determined to be  $[-10, 10]$  and for the sparsity ratio was  $[0, 1]$ . In the case of the network size, we selected the range  $[1-300]$  in our experiments. In the case of SVM ( $\gamma, bc$ ), AIN searched over  $[2^{-10}, 2^{-9}, \dots, 2^4]$  and  $[2^{-2}, 2^{-1}, \dots, 2^{12}]$ , ranges respectively. For each study, we executed the AIN optimiser for a different number of epochs  $[5, 10, 50, 100, \dots]$  with different population sizes  $[5, 10, 50, 100, \dots]$ . These diverse sets of case studies are discussed separately below.

### 5.4.1 Case Study I: Application to UCL Dataset

The first is the BUPA liver disorders dataset provided by BUPA Medical Research Ltd. This set contains 345 instances, each with seven attributes (the 7th attribute specifies the class). These attributes provide information about male patients' blood tests, drinking habits and disorder status (class) [204]. The second dataset, PIMA Indians Diabetes, contains 768 instances with eight attributes and a 9th attribute that specifies the class. The class can be one (1) if the patient tested

positive for diabetes and it is (0) otherwise [204]. To guarantee the results, we further partitioned the dataset into independent training and testing sets via 10-fold cross validation, and the algorithm was trained on the training set and tested on the testing set. The overall mean of the generated results is reported in Table 5.2 below.

Our experiments showed that the proposed algorithm has enhanced the classifier performance many times by tuning the hyper-parameters (Table 5.2). These classifiers have their merits and demerits, but overall, the results suggest that the SVM outperforms both the ESN and ELM on classification accuracy. On average, the AIN optimiser improved the ELM performance from 68.78% to 75.83% in  $2.13e+003$  CPU time (PIMA dataset), whereas in the case of the ESN, the improvement was from 69.85% to 74.08% in  $2.54e+003$  CPU time. Similarly, in the case of the SVM, it improved from 72.57% to 77.43% in  $2.79e+004$  CPU time. Similarly, for the BUPA dataset, the ELM performance improved from 61.38% to 73.03% in  $2.16e+003$  CPU time, whereas in the case of the ESN, the improvement was from 56.12% to 69.66% in  $2.27e+003$  CPU time, and in the case of the SVM, it upgraded from 62.00% to 74.23% in  $2.49e+004$  CPU time. In short, the SVM wins over the ELM and ESN because of providing a better performance but in more learning time, and the next best is the ELM because the optimisation takes less time to search the parameter space, compared to the SVM. For SVM, when compared with the conventional grid search algorithm, it can be seen that AIN provides the better model for classification enhancement in the case of both

datasets and in a lesser time. This is because the search space in the case of GRID search is limited.

**Table 5.2. Average of 10-fold cross validation results for classifier accuracy %/CPU Time for PIMA dataset.**

Classifier	Pima Indians Diabetes Dataset			
	Original	ES Optimised	AIN Optimised	Grid Search Optimized
ESN	69.85/ 0.3710	73.87/ 3.54+02	74.08/2.54e+003	-
ELM	68.78/0.3121	72.26/2.18e+02	75.83/ 2.15e+003	-
SVM	72.57/0.2990	76.12./4.36e+03	77.88/ 2.79e+004	77.43/7.35e+003

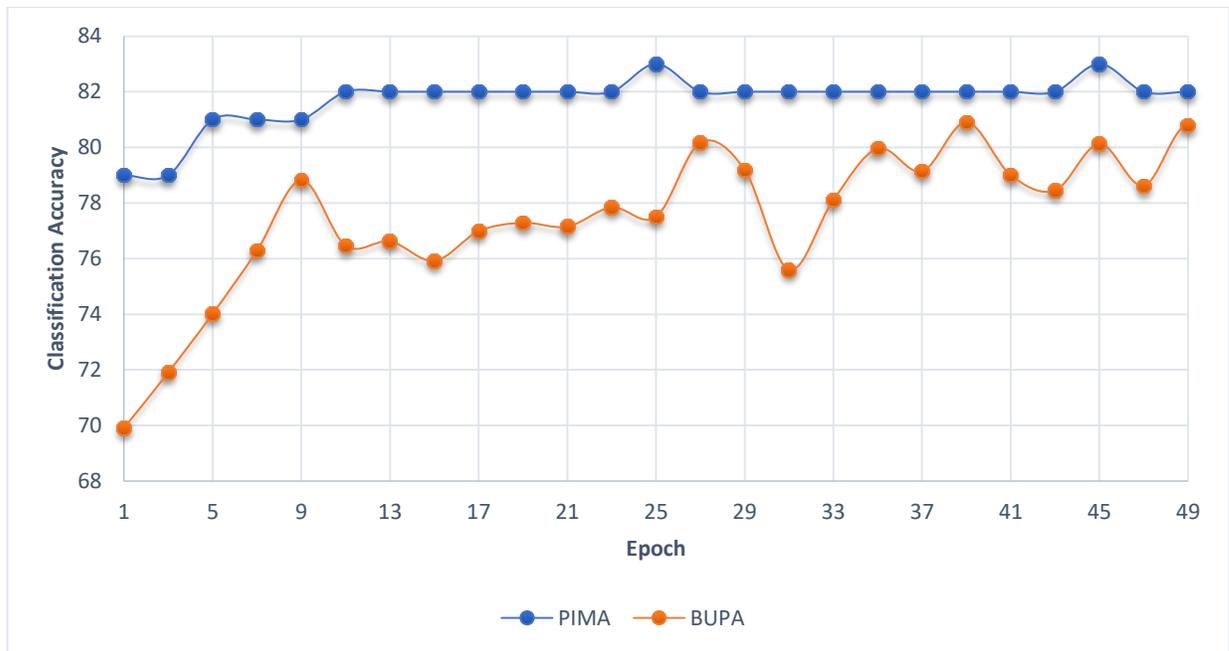
**Table 5.3. Average of 10-fold cross validation results for classifier accuracy %/CPU Time for BUPA dataset.**

Classifier	BUPA liver disorder dataset			
	Original	ES Optimised	AIN Optimised	Grid Search Optimized
ESN	56.12/0.1556	63.49/2.16+02	69.66/2.27e+003	-
ELM	61.38/0.1346	68.64/2.32+02	73.03/2.16e+003	-
SVM	62.00/ 0.1563	67.75/2.57e+03	74.23/ 2.49+004	73.79/8.68e+003

In the case of ESN, multiple runs of the algorithm conclude that the reservoir size contributes little to enhance the classifier performances because the same accuracy was achieved by a small network as well as by a large network; thus, we conclude that the range [1,300] is appropriate for experimentation purposes. Similarly, we experimented with the range of weights being [-5, 5], [-10, 10] and [-20, 20], for the ELM and ESN, and we found that having a larger range makes no difference in improving the performance and that the highest accuracy is achieved within the limited range of [-10, 10]. The sparsity ratio (ESN) must be within [0, 1], as suggested by [132], and our experiments suggest that higher classification accuracy is achieved with a high sparsity ratio for ESN. In the case of SVM, ( $\gamma$ ,  $bc$ ) were taken to be within the ranges  $[2^{-10}, 2^{-9}, \dots, 2^4]$  and  $[2^{-2}, 2^{-1}, \dots, 2^{12}]$ , respectively.

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

We experimented with varying the population size [10, 20, 50,100, 500...] and number of epochs and concluded that it is better to find a balance between the population size and number of epochs because, after a certain limit, the population size only makes the algorithm slow without contributing to the performance enhancement. Figs. 5.2, 5.3 and 5.4 depict the search pattern of the AIN optimiser for the ELM, ESN and SVM search space, respectively where Figs. 5.5-5.6 display the GRID search pattern for PIMA and BUPA dataset using SVM classifier.



**Figure 5.2. Learning curve epochs (time) for two datasets, using AIN-based ESN optimisation.**

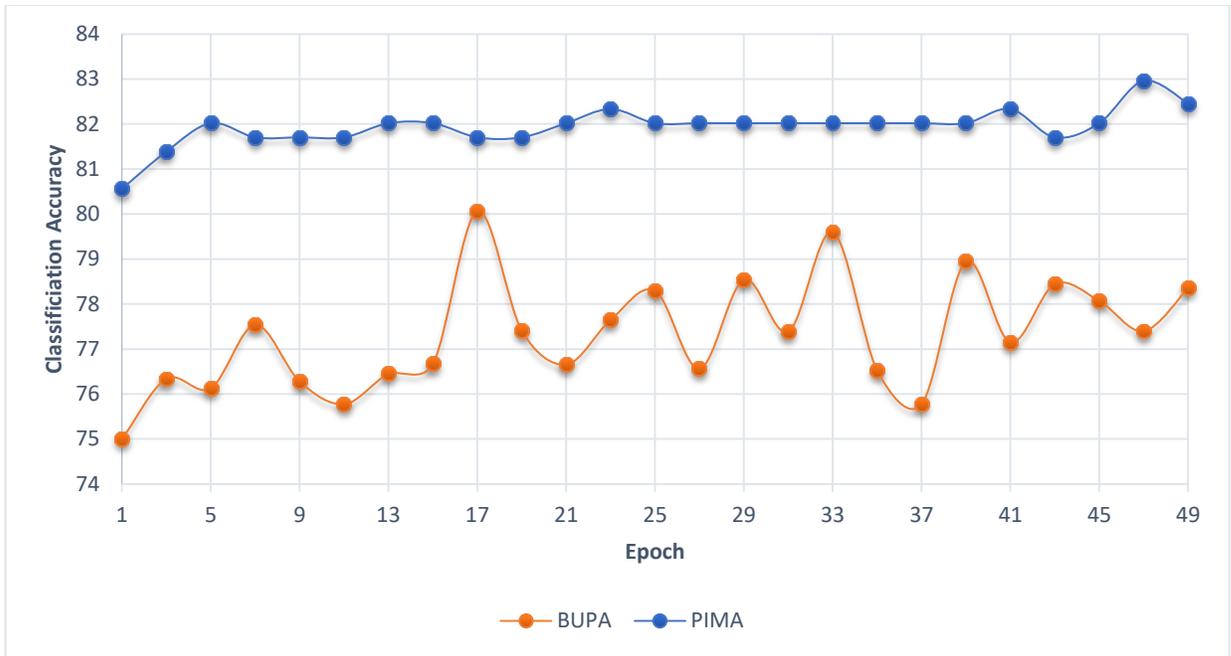


Figure 5.3. Learning curve over epochs (time) for two datasets, using AIN-based ELM optimisation.

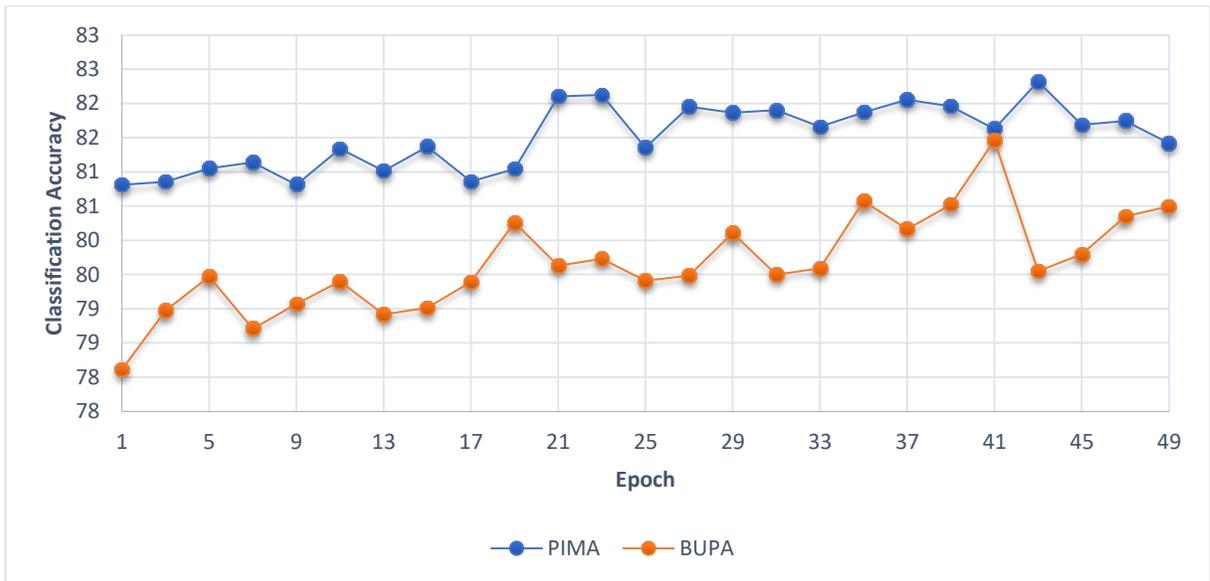
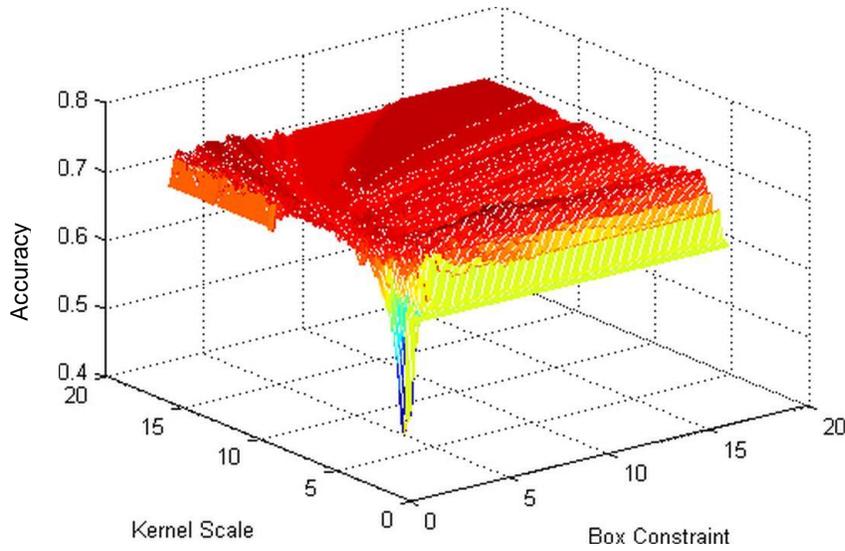
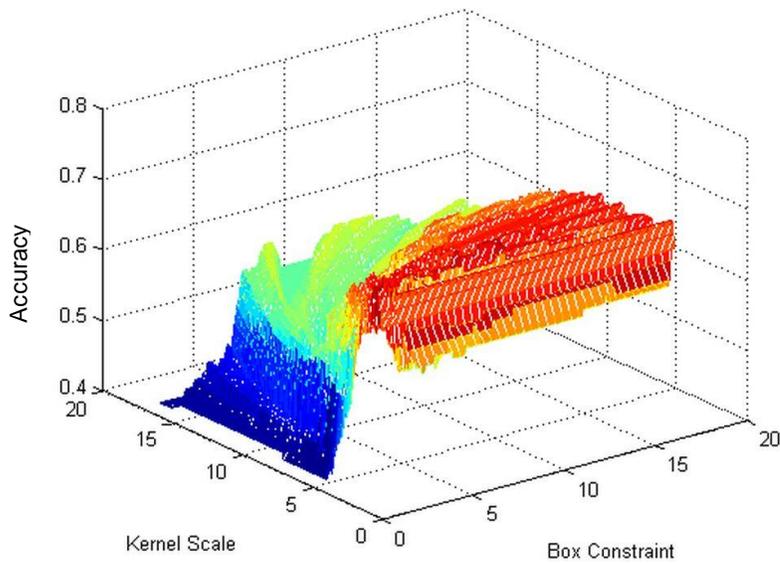


Figure 5.4. Learning curve over epochs (time) for two datasets, using AIN-based SVM optimisation.



**Figure 5.5** Grid search for SVM RBF for finding optimum values of (gamma and C) for the PIMA dataset.



**Figure 5.6** Grid search for SVM RBF for finding optimum values of (gamma and C) for the BUPA dataset.

#### 5.4.1.1 *Comparison with Evolutionary Strategy (ES) based Optimiser*

We compared our proposed technique with ES based hyper parameter optimisation. We set the weight range to be within the limits  $[-5, 5]$  and the network

reservoir size to be within the range [1-100], and we ran the algorithm with different population and generation sizes (100, 50 and 500) multiple times for the datasets. The ES was set to discrete recombination with the  $(\mu+\lambda)$ -selection (a set of parents and children both selected). The results of applying this algorithm to optimise the classifier are averaged in Table 5.2. The results suggest that ES also enhances the classification performance but the AIN optimiser outperforms the others in every case.

Furthermore, we applied statistical analysis with a t-test to highlight the significance of our method compared to the canonical and ES optimised methods.

### 5.4.1.1.1 t-test

Considering ELM to be representative of the classifiers, a t-test was applied to the performance accuracy that was computed from a 10-fold cross validation, for each dataset. The t-test was performed at a 0.05 level of significance.

Let  $\mu_E, \mu_M$  and  $\mu_{DE}$  be the mean performance accuracies from application of the ELM, ES-optimised ELM and AIN-optimised ELM classifier, respectively. We tested the null hypothesis:

$$H_0: \mu_{DE} - \mu_E = 0 \quad (5.10)$$

*(AIN optimisation does not make a difference in the classification performance)*

against  $H_a: \mu_{DE} - \mu_E > 0 \quad (5.11)$

*(AIN optimisation improves the classification performance)*

$$\text{and } H_0: \mu_{DE} - \mu_M = 0 \quad (5.12)$$

*(AIN optimisation generated performs as well as the ES optimiser)*

$$\text{against } H_a: \mu_{DE} - \mu_M > 0 \quad (5.13)$$

*(AIN optimisation performs better than the ES optimiser)*

**Table 5.4. The results of the t-test at the 0.05 level of significance.**

<b>Dataset</b>	<b>Alternate Hypothesis H<sub>a</sub></b>	<b>P-value</b>	<b>Null Hypothesis H<sub>0</sub></b>
Pima	$H_a: \mu_{DE} - \mu_E > 0$	0.0023	Rejected
Indians Diabetes Dataset	$H_a: \mu_{DE} - \mu_M > 0$	0.0015	Rejected
BUPA liver	$H_a: \mu_{DE} - \mu_E > 0$	0.0032	Rejected
disorder dataset	$H_a: \mu_{DE} - \mu_M > 0$	0.0028	Rejected

The t-test results (Table 5.4) clearly indicate the significance of the AIN optimiser because AIN-optimised ELM generates better results compared to the canonical ELM and compared to the ES-based optimised ELM. The same can be proven for ESN and SVM.

#### **5.4.2 Case Study II: Application to 2D Medical Images**

We further extended our experiments to optimise the performance of CDSS for breast cancer detection, as discussed in section 4.2.1, as well as lung cancer detection in section 4.2.2, by application of AIN for hyper parameter optimisation for SVM, ESN and ELM classifiers.

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

For the experiments, we selected MIAS, INbreast as well as JSRT datasets and the results are reported in the tables below.

**Table 5.5. Average of 10-fold classification results for classifier accuracy %/CPU Time for MIAS dataset.**

Classifier	Original	ES Optimised	AIN Optimised
ESN	94.35/ 0.0211	96.10/ 3.243e+02	97.20/ 2.06e+02
ELM	95.81/1.002	97.98/ 2.39e+02	98.65/ 1.09e+02
SVM	100.00/0.5156	-	-

**Table 5.6. Average of 10-fold classification results for classifier accuracy %/CPU Time for INbreast dataset.**

Classifier	Original	ES Optimised	AIN Optimised	GRID Search Optimised
ESN	95.12/0.101	97.10/3.07e+02	98.02/1.56e+03	-
ELM	98.29/1.260	99.45/2.39e+02	100.00/1.81 e+02	-
SVM	99.53/0.1252	100.00/945.10	100.00/978.00	100.00/698.70

**Table 5.7. Average of 10-fold classification results for classifier accuracy %/CPU Time for JSRT dataset.**

Classifier	Original	ES Optimised	AIN Optimised	GRID Search Optimised
ESN	97.66/ 0.1453	98.67/ 3.21e+02	99.31/3.61e+03	-
ELM	99.34/0.4260	100.00/ 1.09e+01	100.00/ 2.15e+01	-
SVM	97.98/0.2890	99.71/ 2.34e+02	100.00/ 922.00	100.00/897.90

From the results in the tables above (Table 5.5-5.7), it is evident that AIN has the capability to enhance machine-learning classifiers' performance significantly. The results have been averaged over a 10-fold cross validation applied on 100 highest degree LESH features selected from our previous study, reported in Chapter Two (sections 4.3.1 and 4.3.2). In the case of the MIAS dataset, the ESN performance has been enhanced from 94.13% to 97.20% in 2.06e+02 CPU time, by tuning the

network weights. In the case of ELM it has been enhanced from 95.81 to 98.65% in  $1.09e+02$  CPU time, while in the case of SVM the performance was already 100%. Similarly, in the case of the INbreast dataset, the performance of classifiers for distinguishing between malignant and benign cases has been enhanced as follows: for ESN the classification accuracy jumped from 95.12% to 98.02% in  $1.56e+03$  CPU time whereas in case of ELM it was enhanced from 98.29% to 100.00% in  $1.81e+01$  CPU time, and for SVM it was enhanced from 99.53% to 100% in 594.34 CPU time.

Similarly, to diagnose between malignant and benign chest nodules from the JSRT dataset, the enhancement of ESN improved from 97.66% to 99.31% in  $3.61e+03$  CPU time, whereas in the case of ELM, it was enhanced from 99.34% to 100% in  $2.15e+01$  CPU time, and for SVM it was from 97.98% to 100% in 922 CPU time.

### 5.4.2.1 *Comparison with Evolutionary Strategy (ES) Optimizer*

We compared our optimisation technique with ES. We set the range of weight to be within the limits  $[-5, 5]$  and the network reservoir size to be within the range  $[1-100]$ , and we ran the algorithm with different population and generation sizes (100, 50 and 500) multiple times for the datasets. The results applying this algorithm to optimise the classifier are averaged in Tables 5.5-5.7. The results suggest that ES and AIN are both capable of enhancing the classification performance, where ES takes less time, compared to AIN, in most of the cases but AIN performs better than ES. It is due to this fact that AIN has more operators, compared to ES, which

generates a wider search space where looking through a larger search space requires more time.

Furthermore, we applied statistical analysis with a t-test to highlight the significance of our method, compared to the canonical and ES-optimised methods.

#### 5.4.2.2 *t-test*

We considered ELM to be representative of the classifiers and an applied t-test to compare the performance using 10-fold cross validation for each dataset, at a 0.05 level of significance.

Considering again  $\mu_E, \mu_M$  and  $\mu_{DE}$  as the mean performance accuracies from application of the ELM and ES-optimised and AIN-optimised ELM classifier, respectively, according to the hypothesis described in eq. “(5.10, 5.11, 5.12, 5.13)”, our results are given in Table 5.8 below:

**Table 5.8. The results of the t-test at the 0.05 level of significance.**

<b>Dataset</b>	<b>Alternate Hypothesis <math>H_a</math></b>	<b>P-value</b>	<b>Null Hypothesis <math>H_0</math></b>
MIAS Dataset	$H_a: \mu_{DE} - \mu_E > 0$	0.006	Rejected
MIAS Dataset	$H_a: \mu_{DE} - \mu_M > 0$	0.01	Rejected
INbreast Dataset	$H_a: \mu_{DE} - \mu_E > 0$	0.003	Rejected
INbreast Dataset	$H_a: \mu_{DE} - \mu_M > 0$	0.0170	Rejected
JSRT Dataset	$H_a: \mu_{DE} - \mu_E > 0$	0.0014	Rejected

Dataset	Alternate Hypothesis $H_a$	P-value	Null Hypothesis $H_0$
JSRT Dataset	$H_a: \mu_{DE} - \mu_M > 0$	0.01	Rejected

The t-test results (Table 5.8) clearly indicate the significance of the AIN optimiser because AIN-optimised ELM generates better results compared to the canonical ELM as well as ES-optimised ELM. The same can be proven for ESN and SVM.

### 5.4.3 Case Study IV: Application to 3D breast MRI

Application of AIN-based optimisation technique, applied to the CDSS for detecting breast cancer through 3D LESH application to MRI images, is reported and discussed below.

We already applied an AIN-based hyper-parameter optimisation technique to ELM, ESN and SVM classifiers with experiments on UCL datasets as discussed above. Further, we extended these experiments to two-dimensional medical images, and hence applied the same technique to MIAS, INbreast (mammograms) and chest x-rays (JSRT dataset) datasets. Our experiments manifest that the proposed algorithm has enhanced the classifier performance many fold by tuning the hyper-parameters (Table 5.9).

We extend our experiments further via application of AIN-based optimised classifiers to diagnose abnormalities from 3D medical images. For this we selected a breast MRI dataset with enhanced 3D LESH features extracted from them, as described in section 4.3.3. Table 5.9 reports the rate of classification accuracy and

learning time for results obtained by using the MRI dataset, using the 100 largest features for discriminating between abnormal and normal MRI cases. The results have been averaged over 10-fold cross validation.

**Table 5.9. Average of 10-fold classification results for classifier accuracy %/ learning time for Breast MRI dataset.**

Classifier	Original	ES Optimised	AIN Optimised
ESN	83.45 / 0.1385	95.56 / 565.51	98.50 / 3.17e+03
ELM	97.82/0.0012	98.67/ 340.35	99.23/ 207.56+03
SVM	100.00/0.1936	-	-

We selected a range of values [1,300] for reservoir size estimation in ESN. Our experience demonstrated that increasing the network size does not contribute towards increasing the performance accuracy of the classifier, as the same accuracy was achieved by a small network as well as by a large network. For the weight range we selected [-10, 10] for ELM and ESN the sparsity ratio (ESN) must be within [0, 1], as suggested by [132], and our experiments suggest that a higher classification accuracy is achieved with a high sparsity ratio for ESN. In the case of SVM, ( $\gamma$ ,  $bc$ ) were taken to be within the ranges  $[2^{-10}, 2^{-9}, \dots, 2^4]$  and  $[2^{-2}, 2^{-1}, \dots, 2^{12}]$ , respectively. We experimented with varying the population size [10, 20, 50,100, 500...] and number of epochs and concluded that it is better to find a balance between the population size and number of epochs because, after a certain limit, the population size makes the algorithm slow without contributing to the performance enhancement.

#### 5.4.3.1 *Comparison with Evolutionary Strategy (ES) Optimiser*

The benchmark evolutionary algorithm ES was applied and compared with the AIN-based optimisation technique. The weight range was limited to [-10, 10] and the network reservoir size to be within the range [1-100]. The algorithm was run for different population sizes (50,70,100, ...500) for the MRI dataset. The results of applying this algorithm to optimise the classifier are averaged in Table 5.9 above. The results suggest that AIN has a performance comparable to the ES optimisation technique but takes more learning time. We aim to tune AIN, in future, by increasing its efficiency and hence decreasing its computational cost.

Furthermore, we applied statistical analysis with a t-test to highlight the significance of our method, compared to the canonical and ES optimised methods.

#### 5.4.3.2 *t-test*

We applied a t-test to authenticate the performance boost due to the application of AIN optimisation upon the classifier's accuracy. The t-test was applied on results generated using 10-fold cross validation with 0.05 level of significance.

Let  $\mu_{ME}$ ,  $\mu_{MM}$  and  $\mu_{MDE}$  be the mean performance accuracies from application of the ELM, ES-optimised ELM or the AIN-optimised ELM classifier, respectively, when applied on the breast MRI dataset. We tested the null hypothesis:

$$H_0: \mu_{MDE} - \mu_{ME} = 0 \quad (5.14)$$

*(AIN optimisation does not make a difference in the classification performance)*

$$\text{against } H_a: \mu_{MDE} - \mu_{ME} > 0 \quad (5.15)$$

*(AIN optimisation improves the classification performance)*

$$\text{and } H_0: \mu_{MDE} - \mu_{MM} = 0 \quad (5.16)$$

*(AIN optimisation performs as good as the ES optimisation)*

$$\text{against } H_a: \mu_{MDE} - \mu_{MM} > 0 \quad (5.17)$$

*(AIN optimisation performs better than the ES optimisation).*

Results for each classifier are reported in Table 5.10 below.

**Table 5.10. The results of the t-test at the 0.05 level of significance.**

Classifier	Alternate Hypothesis $H_a$	P-value	Null Hypothesis $H_0$
ELM	$H_a: \mu_{MDE} - \mu_{ME} > 0$	0.003	Rejected
	$H_a: \mu_{MDE} - \mu_{MM} > 0$	0.0042	Rejected
ESN	$H_a: \mu_{MDE} - \mu_{ME} > 0$	000014	Rejected
	$H_a: \mu_{MDE} - \mu_{MM} > 0$	0.0029	Reject
SVM	$H_a: \mu_{MDE} - \mu_{ME} > 0$	NIL	NIL
	$H_a: \mu_{MDE} - \mu_{MM} > 0$	NIL	NIL

The t-test results (Table 5.10) clearly indicate the significance of the AIN optimiser because AIN-optimised classifiers outperform the canonical classifiers (without hyper-parameter tuning). Compared with the ES-based optimisation technique, we conclude that the AIN optimisation technique performs better than ES optimisation while the learning time for ES is less.

## 5.5 Conclusions and Discussion

A self-configuring, fully automated strategy to optimise classifiers is formidable in the field of machine learning. This paper proposes such an automated optimisation technique, AIN, which is inspired by the biological immune system. We conducted experiments to evaluate the potential of AIN as an optimiser for three different machine learning classifiers using two UCL datasets, as well as two-dimensional and three-dimensional medical images.

The results suggest that the classification performance can be prominently increased using the AIN optimiser algorithm because it has the capability to reach local as well as global optima simultaneously.

We also observed that simple SVM, as well as optimised SVM, takes substantially less learning time, with better classification performance, compared to ESN and SVM. Most of the learning algorithms have the drawback that they can become stuck in a local minimum; additionally, their learning time can take days. In the case of ELM, the human intervention necessary is minimal to zero (in tuning the user-specific parameters), while the learning time is minimal and exhibits good universal approximation.

In any case, a larger network size does not contribute to enhancing the performance accuracy for ESN or ELM, whereas the range of weights as well as sparsity ratio play an important role in increasing their performance.

## Nature-Inspired Algorithm for Machine Learning Classifier Optimisation

In the case of ELM, we experimented with searching for most appropriate range of values to generate the hidden weights, as well as the appropriate number of neurons in the hidden layer. We also attempted optimisation of the hidden biases, but it did not turn out to add up to an ELM classification performance boost.

We compared AIN and ES optimisers and found that although ES takes less time it results in lower performance. In future, we aim to minimise the computational cost (the CPU time) of the AIN optimiser by fine-tuning the algorithm. Further, we will conduct experiments by optimising other machine learning algorithms with our proposed algorithm, as well as investigate the tendency of the proposed algorithm to accomplish predictive tasks. Intelligent fitness function (other than classification accuracy) may be introduced to lower generalisation error, and thus improve the performance of machine learning classifiers. The algorithm can be verified by using other machine learning datasets available online, which were kept pending due to space constraints here. Further, the algorithm's strength to enhance the performance of multiclass machine learning classifiers can be investigated in future.

## **Chapter 6 - Conclusion and Future Work**

The central idea behind our research is directed to contribute towards health care by suggesting an efficient CDSS, which will improve the diagnosis capability of medical practitioners. It was accomplished by developing non-knowledge-based CDSSs based on recent medical imaging, image processing, and machine-learning techniques, and further improving them by the application of an AIN based optimization algorithm. The suggested improvements will help to reduce the cost of health care, and enhance diagnosis outcomes. Further application of the above-mentioned framework using the diverse nature of medical datasets has been employed and results are reported in subsequent chapters.

Medical health research is generating an enormous amount of data, which cannot be handled with traditional method of information management, data analysis and patient diagnosis. Health professionals are constantly encountering information management activities, namely patient information management, sharing information with other researchers, planning diagnostic systems, keeping medical information updated, developing patient care strategies, laboratory result interpretation and radiologic studies. There comes a need for developing automated decision support systems, which can assist in individual patient care by removing the need for intensive, costly lab tests, and which can provide better diagnosis [4].

## Conclusion and Future Work

The following section 6.1 summarizes the overall research process outcome; limitations are discussed in section 6.2, and section 6.3 describes our future research endeavours.

### **6.1 Benefits of the Developed CDSSs**

The world of research in clinical decision support systems for health care is continuously changing and so is the world of computing. Thus we see the emergence of novel scientific issues that sit at the intersections of medical science, patient care and information technology. It is out of the question to practise modern medicine or medical research without the aid of state-of-the-art computational paradigms. Hence, our proposed new generation clinical decision support systems have the capability of offering better patient care by careful diagnosis of individual cases of the disease, and thus will assist in suggesting appropriate laboratory tests and drug dosage. Every stage of the framework has been discussed and relevant benchmark techniques (applied at each stage) are highlighted.

Clinical Decision Support Systems (CDSSs) play an important role in early diagnosis of fatal diseases such as cancer. These systems help in detecting or diagnosing abnormal conditions in medical datasets that may consist of images or other disease relevant information for multiple patients. We aim to enhance the performance of CDSSs by using adequate image processing and machine learning techniques, especially for medical imaging based cancer diagnosis. For this, we recently proposed a novel application of a local energy-based shape histogram (LESH) as the feature set for the recognition of abnormalities in x-ray images. We

## Conclusion and Future Work

investigated the implications of this technique on mammogram datasets of the Mammographic Image Analysis Society and INbreast dataset. In the evaluation, regions of interest were extracted from the mammograms, their LESH features were calculated, and they were fed to a support vector machine (SVM) classifier. In addition, the impact of selecting a subset of LESH features on classification performance was also observed and benchmarked against a wavelet based feature extraction method by Cristiane et al. [53] in chapter 4.

We further extended our original work to apply the LESH technique to detect lung cancer by analysing chest radiographs. The JSRT Digital Image Database of chest radiographs was selected for this experiment. Prior to LESH feature extraction, we enhanced the radiographic images using a contrast limited adaptive histogram equalization (CLAHE) approach. Selected cognitive machine learning classifiers, ELM, SVM and ESN were then applied for predicting the medical state (existence of benign or malignant cancer) present in the x-ray images. The simulation results, evaluated using the classification accuracy performance measure, and benchmarked against the wavelet based feature extraction technique, authenticated the distinct capability of our proposed framework for enhancing the diagnosis outcome.

Furthermore we developed a novel technique to detect breast cancer in volumetric medical images based on a three-dimensional (3D) LESH model. It is a hybrid approach, which combines a 3D LESH feature extraction technique with machine learning classifiers to detect breast cancer from MRI images. The proposed

## Conclusion and Future Work

system applied 3D contrast limited adaptive histogram equalization (CLAHE) to the MRI images before extracting 3D LESH features. Further, a selected subset of features is fed to a machine learning classifier namely SVM, ELM or ESN for detection of abnormalities, as well as to distinguish among different stages of abnormality. The results indicate high performance of the proposed system. When compared to the wavelet based feature extraction technique, statistical analysis testifies the significance of our proposed algorithm.

Furthermore, we also presented a novel application of Artificial Immune Network (AIN) for optimizing machine learning classification algorithms. We employed this technique in conjunction with selected machine learning classifiers namely ELM, SVM and ESN and validated it using the benchmark medical datasets of PIMA India diabetes and BUPA liver disorders. The results were investigated using the classification accuracy measure and the learning time of the algorithm. We also compared our methodology with a benchmarked multi-objective genetic algorithm (ES)-based optimization technique. The results authenticated the potential of AIN as an optimizer. The proposed algorithm was also verified on a two-dimensional dataset (mammograms and chest radiographs) and three-dimensional datasets (MRI dataset).

In future, large-scale studies are required to validate the results of our proposed systems and case studies clinically, and to assess the feasibility of deploying such models in real clinical practice.

## 6.2 Limitations of the Developed CDSSs

- The review and comparison of the state-of-the-art techniques in chapter 3 and chapter 4 is non-exhaustive due to limitations of time and space.
- The limited amount of dataset has been utilized to validate the proposed CDSSs, especially in case of the three-dimensional dataset, as the ratio of cases for multiple stages is biased towards stage II. In future, a large-scale dataset may be accessed for validation of these results and in addition the feasibility of these proposed systems in a real clinical environment may be assessed.
- The results reported as part of the research process demonstrate the capability of the proposed framework to address the diverse nature of medical diagnosis problems. Note that while the preliminary results reported in this dissertation should be taken with care, they do demonstrate the capability to address the diverse nature of medical diagnosis problems, and a range of contributions and potential impact is envisaged from this work, both for clinical practice and further research into employing such models in other clinical applications.
- AIN (chapter 5) is an efficient hyper-parameter optimizer, yet it has high computational cost.

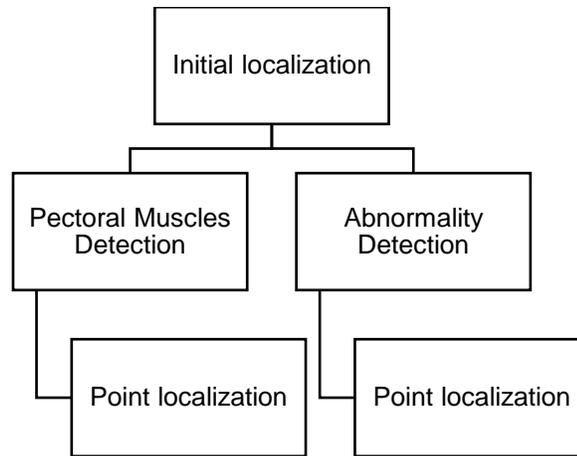
### **6.3 Future Expansions and Recommendations**

For future work, more extensive evaluation and clinical validation is required using additional clinical datasets benchmarked against other state-of-the-art feature selection and classification approaches. A hybrid approach to adaptively optimize an ensembled feature set based on a combination of LESH and other state-of-the-art feature extraction techniques while selecting significant features can be evaluated for enhancing classification performance. Furthermore, an intensive study of clinical significance of LESH features when applied to a real clinical dataset may lead to potential useful intimations. Other future endeavours are discussed below.

#### **6.3.1 Breast Cancer Detection with Gabor-filter-based cascading network**

We further plan to incorporate a Gabor Filter based cascading network technique as part of a CDSS for diagnosing breast cancer. These filters have an important quality in that their frequency and orientation representation matches the human visual system, which makes them a good representative/discriminator of the texture detail in the image [179]. Medical images contain abnormalities, which appear as variant textures within the image. Gabor features can be helpful in discriminating between different kinds of abnormalities inside the medical images. We aim to study and experiment with a Gabor Filter based cascading network method by Samir [180] for analysis of mammograms. This method works in layers and so performs faster than simple Gabor Filter techniques. The method consists of three layers given as:

## Conclusion and Future Work



**Figure 6.1. Three layer abnormality detection network..**

The first layer is intended to localize pectoral muscles and other texture variations inside the image. The second layer is to detect initial features in the pectoral area of the mammograms. It also intends to initialize features from the breast region for possible detection of abnormalities. The third layer is designed to landmark the pectoral muscles for removing it from the image. This layer also points the boundary of any abnormality that can be segmented from the breast region for further analysis. The overall can be represented as:



**Figure 6.2. Detection system with Gabor Filter.**

After segmentation of the region of interest, we will extract Gabor features from them and feed them to machine learning classifiers for diagnosis.

### **6.3.2 Towards a Novel Approach for Medical Image Segmentation**

As mentioned earlier, one of the integral parts of a CDSS is image segmentation and we have applied manual segmentation so far. In the case of medical image segmentation, this can be thought of as segmenting the organ or area of abnormality. In the literature, several segmentation methods are proposed. The main problem with these methods is the loss of information. Each method generates a segmentation image, which is different from the others. We aim to propose a segmentation technique that takes advantage of fusion theory to combine the outcome of several segmentation methods into a single segmented image [181-182].

### **6.3.3 Sensitivity Analysis (SA) as a performance booster for LESH features extraction**

We aim to enhance LESH feature performance by application of sensitivity analysis (SA) to find features that have less influence on classification results. SA allows ranking these features according to their effect on the model output. This will help to reduce imperfection related to the curse of dimensionality [183].

### **6.3.4 Minimization of computation cost of AIN based optimization technique**

We compared AIN and ES optimizers and found that although ES takes less time, it results in lower performance. In future, we aim to minimize the computational cost (the CPU time) of the AIN optimizer by fine-tuning the algorithm. Furthermore, we will conduct experiments by optimizing other machine learning algorithms with our proposed algorithm, and we will also investigate the tendency of the proposed

## Conclusion and Future Work

algorithm to accomplish predictive tasks. Intelligent fitness function (other than classification accuracy) may be introduced to reduce generalization error, and thus improve performance of machine learning classifiers. The algorithm can be verified upon other machine learning datasets available on line, which was kept pending due to space and time constraints here. In addition, the algorithm strength to enhance performance of multiclass machine learning classifiers can be investigated as well as application of this algorithm in the field of two-dimensional and three-dimensional datasets as part of our future work. Optimization of the novel multi-layer ESN proposed by Malik et al. [58] as well as a semi-supervised learning based 3C-SVM model developed by Yang et al. [63] is also part of our future endeavours.

### **6.3.5 Cognitively-inspired Iterative Learning System for an efficient clinical decision support system**

We aim to propose an iterative healthcare learning systems built on the core principle of learning from existing clinical practices through legacy clinical data, as well as utilizing existing clinical practice guidelines to facilitate more efficient clinical decision-making operations. To accomplish this, we aim to develop a cognition-inspired search algorithm based upon the Artificial Immune System (AIS) paradigm, which will enable knowledge discovery from medical datasets by extracting hidden patterns and relationships and form building blocks of the search heuristic.

## Conclusion and Future Work

AISs are inspired by theoretical immunology and observed immune function models. They encompass algorithms that are inspired by ideas and metaphors from the biological immune system to solve complex real-world problems [184].

The main AIS model comprises the following stages:

- Abstract representation: of key features/components of the application domain (as immune cells and molecules)
- Affinity maturation: a biologically-inspired mechanism to represent and evolve interaction among different system components and its environment (comprising a set of input, output and fitness function etc.)
- Meta-dynamics: algorithm to define dynamics (behaviour of the system) including negative positive selection, clonal selection, bone marrow and immune network algorithms.

As outlined above, the first step in AIS modelling is to create an abstract model of the application domain parameters in the form of immune cells and molecules [184]. For example, Perelson and G. Oster devised a new shape-space approach to represent these parameters using data structures specific to the application. This data structure consists of string attributes that can be real-valued vectors, integer strings, binary strings or symbolic strings. The data structure needs to be sufficient to quantify the affinity between each pair of attributes [185]. In our case, it will be patient information provided in the form of medical datasets (comprising medical diagnostic images, history, pathology reports and lab test reports).

## Conclusion and Future Work

In the context of clinical decision support systems, our proposed multi-disciplinary cognitively-inspired iterative learning system can benefit from exploiting state-of-the-art research advances in clinical pathways and national clinical guidelines, patient safety and quality enhancement practices, healthcare operations and evidence-based personalised care. All of these operations/components will form building blocks for the proposed next-generation healthcare systems based on our cognitively-inspired iterative learning framework.

### **6.3.6 State-of-the-art learning systems for radiologist training**

The training of radiologists must be as efficient as possible. To achieve this goal, Mazurowski et al. are developing adaptive computer-aided education systems for mammography [186-189]. The system aims to develop a technique, which can recognize error making patterns in trainee performance (false positives as well as false negatives) by application of LESH features [190-191].

### **6.3.7 Possible implications of a Cognition-Inspired Iterative Learning framework upon training radiologists.**

In our innovative approach, we plan to research and develop a novel cognitively-inspired iterative learning system based on the AIS paradigm for enhancing future clinical decision support, by better identifying (concealed) normal locations in medical images with more chance of causing false positive errors.

In summary, we expect to improve future radiology education efficiency by providing each trainee with a more natural, technology-driven interactive learning environment, which will enable them, and their mentors/consultants, to

## Conclusion and Future Work

appropriately tailor and personalize their iterative training on cases they find particularly difficult. To realize this goal, we will focus on modelling and analysing performance improvement in false positive errors by individual trainees for selected benchmark medical images (available online). We will investigate AIS as a novel iterative learning algorithm that can quickly identify patterns and extract association rules from complex medical images. Such hidden patterns and rules can then be explicitly used to reduce the complexity of the problem, and in the process provide human users/learners with new and relevant visual insights into the complex algorithm operation and underlying problem characteristics.

### **6.3.8 Towards eMobile Clinical Decision Support Systems**

Last but not least, we would advance our research to develop an mHealth application to assist practitioners in the medical diagnosis of fatal diseases, such as cancer, and to suggest further medical care. The application will also provide an environment for interactive learning by facilitating access to other healthcare practitioners in their daily practices.

## Bibliography

- [1] Ernesto Iadanza, Gabriele Guidi and Alessio Luschi, "Decision Support Systems in Healthcare," in *Clinical Engineering: From Devices to Systems*, Academic Press, Boston, 2016, pp. 75-93.
- [2] J. Michael Hardin, David C. Chhieng, "Data Mining and Clinical Decision Support Systems," in *Clinical Decision Support Systems: Theory and Practice*, Springer New York, 2007, pp. 44-63.
- [3] Pavel et al., "Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials," *BMJ* 2013, vol.346, February 2013.
- [4] Mark A. Musen, Blackford Middleton, Robert A. Greenes, "Clinical Decision-Support Systems," in *Biomedical Informatics*, Spring 2014, pp. 643–674.
- [5] Jerome A. Osheroff, Eric A. Pifer, Dean F. Sittig, Robert A. Jenders, *Clinical decision support implementers' workbook*, Chicago: HIMSS, 2004.
- [6] Eta S. Berner, Tonya J. La Lande, "Overview of Clinical Decision Support Systems," in *Clinical Decision Support Systems: Theory and Practice*, Springer New York, 2007, pp. 3-22.
- [7] Institute of Medicine (2007), *The learning healthcare system: Workshop summary*, Washington, DC: National Academies Press.
- [8] Lodwick GS, Haun CL, Smith WE et al. "Computer diagnosis of primary bone tumor," *Radiology* 80, pp. 273–275, 1963.
- [9] Myers PH, Nice CM, Becker HC et al., "Automated computer analysis of radiographic images," *Radiology* 83, pp. 1029–1033, 1964.
- [10] Winsbarg F, Elkin M, May J et al., "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology* 89, pp. 211–215, 1967.
- [11] Qiang Li, Robert M. Nishikawa, "Computer-Aided Detection and Diagnosis in Medical Imaging", CRC Press, 2015.
- [12] Fujita H, Doi K, Fencil LE, Chua KG., "Image feature analysis and computer-aided diagnosis in digital radiography. 2. Computerized determination of vessel sizes in digital sub- traction angiography," *Med Phys* 14, pp. 549–556, 1987.
- [13] Hoffmann KR, Doi K, Chan HP, Fencil L, Fujita H, Muraki A., "Automated tracking of the vascular tree in DSA images using a double-square-box region-of-search algorithm," *Proc SPIE* 626, pp. 326–333, 1986.
- [14] Giger ML, Doi K, MacMahon H., "Computerized detection of lung nodules in digital chest radiographs," *Proc SPIE* 767, pp. 384–386, 1987.
- [15] Giger ML, Doi K, MacMahon H. "Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields," *Med Phys* 15, pp. 158–166, 1988.
- [16] Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography," *Med Phys* 14, pp. 538–548, 1987.
- [17] Anke Meyer-Bäse, "I - Introduction," in *Pattern Recognition in Medical Imaging*, Academic Press, San Diego, 2004, pp. 1-13.
- [18] de Paredes E. S., *Atlas of Mammography*, 3rd ed., Lippincott Williams and Wilkins, Philadelphia, USA, 2007.

- [19] G.-B. Huang, "What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle," *Cognitive Computation*, vol. 7, pp. 263–278, 2015.
- [20] Fred A. Mettler, Walter Huda, Terry T. Yoshizumi, Mahadevappa Mahesh: "Effective Doses in Radiology and Diagnostic Nuclear Medicine: A Catalog," *Radiology 2008*, vol. 248, pp. 254–263.
- [21] Chikumaya, and Mikael Häggström (2002) - Image:Chest.png and Essential Clinical Anatomy, K.L. Moore & A.M. Agu., Lippincott, 2 ed., pp. 92, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=3415343>.
- [22] Stephen B. Edge, Carolyn C. Compton, "The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, 2010.
- [23] Romero, E. and González, F., "From Biomedical Image Analysis to Biomedical Image Understanding Using Machine Learning," in *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*, IGI Global, pp. 1-26, 2010.
- [24] Rafael C. Gonzalez , Richard E. Woods, Digital image processing, 2nd ed., Prentice Hall, 2002.
- [25] M. Sundaram, K. Ramar, N. Arumugam and G. Prabin, "Histogram based contrast enhancement for mammogram images," *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, 2011 International Conference on, Thuckafay, 2011, pp. 842-846.
- [26] M. Sundaram et al., "Histogram-Modified Local Contrast Enhancement for mammogram images," *International Journal of Biomedical Engineering and Technology*, vol. 9, no. 1, 2012.
- [27] Siddharth, Rohit Gupta, and Vikrant Bhateja, "A New Unsharp Masking Algorithm for Mammography Using Non-linear Enhancement Function," *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012)* , *Advances in Intelligent and Soft Computing*, vol. 132, pp. 779-786, 2012.
- [28] H. Cheng et al., "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Trans. Med. Imaging*, vol. 17, no. 3. pp. 442–450, 1998.
- [29] Gordon R, Rangayan R. M, "Feature enhancement Of Film mammograms using fixed and adaptive Neighborhoods," *Applied Optics* 23, pp.560-564, 1984.
- [30] A.Hassanien, and A.Badr, "A Comparative Study on Digital Mammography Enhancement Algorithms Based on Fuzzy Theory," *Studies in Informatics and Control*, vol. 12, pp.21-31, 2003.
- [31] Cheng, H. et al., "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Trans. Med. Imaging*, vol. 17, no. 3, pp. 442–450, 1998.
- [32] Gordon R, Rangayan R. M, "Feature enhancement Of Film mammograms using fixed and adaptive Neighborhoods," *Applied Optics* 23, pp. 560-564, 1984.
- [33] A.Hassanien, and A.Badr, "A Comparative Study on Digital Mammography Enhancement Algorithms Based on Fuzzy Theory," *Studies in Informatics and Control*, vol.12, pp. 21-31, 2003.
- [34] Alfonso Rojas Domínguez, Asoke K. Nandi, "Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 304–315, 2008.

- [35] Jianmin Jiang, Bin Yao, A.M. Wason, "Integration of fuzzy logic and structure tensor towards mammogram contrast enhancement," *Computerized Medical Imaging and Graphics*, vol. 29, pp. 83–90, 2005.
- [36] M. Sundaram, K. Ramar, N. Arumugam and G. Prabin, "Histogram based contrast enhancement for mammogram images," *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on*, Thuckafay, 2011, pp. 842-846.
- [37] I. Daubechies, "Ten Lectures on Wavelets," *PA: SIAM*, 1992.
- [38] Jinshan Tang, Xiaoming Liu and Qingling Sun, "A Direct Image Contrast Enhancement Algorithm in the Wavelet Domain for Screening Mammograms," *IEEE journal of selected topics in signal processing*, vol. 3, no. 1, February 2009.
- [39] Peter Heinlein, Johann Drexl, and Wilfried Schneider, "Integrated Wavelets for Enhancement of Micro calcifications in Digital Mammography," *IEEE transactions on medical imaging*, vol. 22, no. 3, march 2003.
- [40] Eri Matsuyama & Du-Yih Tsai & Yongbum Lee, "A Modified Undecimated Discrete Wavelet Transform Based Approach to Mammographic Image Denoising," *Digit Imaging*, vol. 26, pp. 48–758, 2013.
- [41] Jacob Scharcanski, Cláudio Rosito Jung, "Denoising and enhancing digital mammographic images for visual screening," *Comp. Med. Imag. and Graph*, vol. 30, no. 4, pp. 243-254, 2006.
- [42] Elizabeth Berry, *A Practical Approach to Medical Image Processing*, CRC Press, 2007.
- [43] B. Surendiran, A. Vadivel, "Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer," *IJMEI*, vol. 4, no. 1, pp. 36-54, 2012.
- [44] Yang Mingqiang et al., "A Survey of Shape Feature Extraction Techniques, Pattern Recognition Techniques," *book edited by Peng-Yeng Yin*, ISBN 978-953-7619-24-4, Published: November 1, 2008 under CC BY-NC-SA 3.0 license.
- [45] Pasquale Delogu, Maria Evelina Fantacci, Parnian Kasae, and Alessandra Retico , "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," *Computers in Biology and Medicine*, vol. 37, no. 10, pp.1479-1491, 2007.
- [46] Ramesh Jain et al., "Texture", *in Machine Vision*, McGraw-Hill, Inc., 1995.
- [47] Fritz Albrechtsen, "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices," *Image Processing Laboratory Department of Informatics*, University of Oslo, November 5, 2008.
- [48] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3 , no. 6, pp. 610–621.
- [49] K. Laws, "Textured Image Segmentation", Ph.D. Dissertation, University of Southern California, January 1980.
- [50] Anna N. Karahaliou, Ioannis S. Boniatis, Spyros G. Skiadopoulos, Filippos N. Sakellaropoulos, Nikolaos S. Arikidis, Eleni A. Likaki, et al., "Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications," *IEEE Trans Inf Technol Biomed*, vol. 12, no. 6, 2008, pp. 731–738.
- [51] <http://en.wikipedia.org/wiki/Wavelet>.

- [52] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 11, no. 7, pp.674–693, 1989.
- [53] Cristiane Bastos Rocha Ferreira, Díbio Leandro Borges, "Analysis of mammogram classification using a wavelet transform decomposition," *Pattern Recognition Letters*, vol. 24, pp.973-982, 2003.
- [54] E Candes and D Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," Available on <http://www.Curvelet.org/papers/Curve99.pdf> ,2000.
- [55] Mohamed Meselhy Eltoukhy, Ibrahima Faye, Brahim Belhaouari Samir, "Breast cancer diagnosis in digital mammogram using multi scale curvelet transform," *Computerized Medical Imaging and Graphics* 34, pp.269–276, 2010.
- [56] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2096, Dec 2005.
- [57] E. Romero and F. González, "From Biomedical Image Analysis to Biomedical Image Understanding Using Machine Learning," *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*, IGI Global, 2010, pp. 1-26, DOI:10.4018/978-1-60566-956-4.ch001 [accessed 15 November 2015].
- [58] Vladimir N. Vapnik, "The Nature of Statistical Learning Theory," *SpringerVerlag*, New York.
- [59] Herbert Jaeger , "The "echo state" approach to analysing and training recurrent neural networks," GMD Report 148, German National Research Center for Information Technology, pp. 43, 2001.
- [60] James Bergstra, and Yoshua Bengio, "Random Search for Hyper-Parameter Optimization," *J. Machine Learning Research* 13, 2012, pp. 281–305.
- [61] R. Bellman. Adaptive Control Processes: A Guided Tour. Princeton University Press, New Jersey, 1961.
- [62] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. of LION-5*, 2011.
- [63] S. Ramage and F. Hutter, SMAC source code, java version v2.06.01; URL: [www.cs.ubc.ca/labs/beta/Projects/SMAC](http://www.cs.ubc.ca/labs/beta/Projects/SMAC).
- [64] J. Snoek, H. Larochelle, and R.P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. of NIPS'12*, 2012.
- [65] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kegl, "Algorithms for Hyper-Parameter Optimization," in *Proc. of NIPS'11*, 2011.
- [66] X.C. Guo, J.H. Yang, C.G. Wu, C.Y. Wang, Y.C. Liang, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization," *Neurocomputing*, vol. 71, no. 16–18, October 2008, pp. 3211-3215.
- [67] , B.F. de Souza, A.C.P.L.F. de Carvalho, R. Calvo, R.P. Ishii, "Multiclass SVM Model Selection Using Particle Swarm Optimization," in *Hybrid Intelligent Systems*, 2006. HIS '06. Sixth International Conference on , vol., no., pp.31-31, Dec. 2006 doi: 10.1109/HIS.2006.264914.
- [68] J. Kennedy, R.C. Eberhart, et al., "Particle swarm optimization," in *proceedings of IEEE international conference on neural networks*, vol. 4, pp. 1942-1948, Perth, Australia, 1995.

- [69] Y. Shi, R.C. Eberhart, "A modified particle swarm optimizer," *in proceedings of the IEEE International Conference on Evolutionary Computation*, Piscataway, NJ, 1998, pp. 69–73.
- [70] X.C. Guo, J.H. Yang, C.G. Wu, C.Y. Wang, Y.C. Liang, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization," *Neurocomputing*, vol. 71, no. 16–18, October 2008, pp. 3211-3215, ISSN 0925-2312, <http://dx.doi.org/10.1016/j.neucom.2008.04.027>.
- [71] Cheng-Lung fsuck, Jian-Fan Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing*, vol. 8, no. 4, September 2008, pp. 1381-1391, ISSN 1568-4946, <http://dx.doi.org/10.1016/j.asoc.2007.10.007>.
- [72] An Introduction to Genetic Algorithms. Mitchell Melanie. A Bradford Book The MIT Press. Cambridge, Massachusetts • London, England. Fifth printing, 1999.
- [73] Melanie Mitchell, An Introduction to Genetic Algorithms, MIT Press Cambridge, MA, USA ©1998
- [74] L. Budin, M. Golub, and A. Budin., "Traditional Techniques of Genetic Algorithms Applied to Floating-Point Chromosome Representations," *proceedings of the 41st Annual Conference KoREMA*, pp.93-96, Opatija, 1996,
- [75] Coello Coello, Carlos, Lamont, Gary B., van Veldhuizen, David A., *Evolutionary Algorithms for Solving Multi-Objective Problems*, Second Edition, 2007, Springer US, Boston, MA.
- [76] Thomas Mazzocco, "Toward a novel predictive analysis framework for new-generation clinical decision support systems," Dept. of Computing Science and Mathematics, Univ. of Stirling, Stirling, UK, 2014.
- [77] M. Lichman. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [78] J. Suckling et al, "The Mammographic Image Analysis Society Digital Mammogram Database", *Exerpta Medical International Congress Series 1069*, pp.375-378, 1994.
- [79] IC Moreira et al., "INbreast: toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236-248, 2012.
- [80] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, pp. 71–74, 2000.
- [81] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, Maffitt D., Pringle M., Tarbox L., and Prior F., "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository." *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045 –1057, 2013.
- [82] K. Suzuki, *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis*, Hershey, PA: IGI Global. 2012, DOI:10.4018/978-1-4666-0059-1, pp. 1-524.
- [83] A. Laine, Jian Fan and Wuhai Yang, "Wavelets for contrast enhancement of digital mammography," *in IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 5, pp. 536-550, Sep/Oct 1995.
- [84] Shailaja Singh et al., "Performance Analysis of Mammographic Image Enhancement Techniques for Early Detection of Breast Cancer," *Advances in Parallel Distributed*

- Computing Communications in Computer and Information Science*, vol. 203, pp.439-448, 2011.
- [85] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, 2006, pp. 1116-28.
- [86] S. Pieper, M. Halle, R. Kikinis, 3D SLICER, *proc IEEE Int Symp Biomed Imaging*, 2004 Apr, pp. 632-5.
- [87] C. A. Schneider, W. S. Rasband, & K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis", *Nature methods*, vol. 9, no. 7, 2012, pp.671-675, PMID 22930834 (on Google Scholar).
- [88] S. Weeratunga and C. Kamath, "An investigation of implicit active contours for scientific image segmentation," *Video Communications and Image Processing. SPIE Electronic Imaging*, San Jose. 2004.
- [89] A. M. Khan, S. Ravi, "Image Segmentation Methods: A Comparative Study," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, 2013.
- [90] A. Taneja et al., "A performance study of image segmentation techniques," *Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 4th International Conference on. Noida. 1-6 (2015).
- [91] R.M. Haralick and L.G. Shapiro, "Image segmentation techniques," *Comput. Vis. Graph. Im. Proc.*, vol. 29, pp.100–132, 1985.
- [92] G. Kumar and P. K. Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems," *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on*, Rohtak, 2014, pp. 5-12.
- [93] D. F. Wang, L. Shi, P. A. Heng, "Automatic detection of breast cancers in mammograms using structured support vector machines," *Neurocomputing*, vol. 72, pp.3296–3302, 2009.
- [94] Rafayah Mousa, Qutaishat Munib, and Abdallah Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert Syst. Appl.*, vol. 28, no. 4, pp.713-723, 2005.
- [95] Fatemeh Moayedi, Zohreh Azimifar, Reza Boostani, Serajodin Katebi, "Contourlet-based mammography mass classification using the SVM family," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 373-383, April 2010.
- [96] F. Moayedi, Z. Azimifar, R. Boostani, S. Katebi, "Contourlet based mammography mass classification," in *Lecture Notes in Computer Science, Image Analysis and Recognition*, vol. 4633, Springer, Berlin, 2007, pp.923–934.
- [97] F. Moayedi, R. Boostani, Z. Azimifar and S. Katebi, "A Support Vector Based Fuzzy Neural Network Approach for Mass Classification in Mammography," in *proceedings 15th International Conference on Digital Signal Processing*, Jul. 2007, vol. 1-4, pp.240-243.
- [98] L. de Oliveira Martins, A. C. Silva, A.C. de Paiva, et al., "Detection of Breast Masses in Mammogram Images Using Growing Neural Gas Algorithm and Ripley's K Function," *J Sign Process Syst Sign Image Video Technol*, vol. 55, pp.77-91, 2009.
- [99] Geraldo Braz Junior, Anselmo Cardoso de Paiva, Aristófanés Corrêa Silva, Alexandre Cesar Muniz de Oliveira, "Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM," *Computers in Biology and Medicine*, vol. 39, no. 12, December 2009, pp. 1063-1072, ISSN 0010-4825, <http://dx.doi.org/10.1016/j.compbiomed.2009.08.009>.

- [100] Y. Wang, X. Gao and J. Li, "A Feature Analysis Approach to Mass Detection in Mammography Based on RF-SVM," *2007 IEEE International Conference on Image Processing*, San Antonio, TX, 2007, pp. V - 9-V - 12.
- [101] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, H.K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics*, vol. 29, no. 6, September 2005, pp. 419-429.
- [102] Defeng Wang, Lin Shi, Pheng Ann Heng, "Automatic detection of breast cancers in mammograms using structured support vector machines," *Neurocomputing*, vol. 72, no. 13–15, August 2009, pp. 3296-3302.
- [103] Alfonso Rojas Domínguez, Asoke K. Nandi, "Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection," *Computerized Medical Imaging and Graphics*, vol. 32, no. 4, June 2008, pp. 304-315.
- [104] Mohamed Meselhy Eltoukhy, Ibrahima Faye, Brahim Belhaouari Samir, "Breast cancer diagnosis in digital mammogram using multi scale curvelet transform" *Computerized Medical Imaging and Graphics*, vol. 34, pp.269–276, 2010.
- [105] Essam A Rashed, Ismail A Ismail, and Sherif I Zaki, "Multiresolution mammogram analysis in multilevel decomposition" *Pattern Recognition Letters*, vol. 28, no. 2, pp. 286-292, 2007.
- [106] Rafayah Mousa, Qutaishat Munib, and Abdallah Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert Syst. Appl.* vol. 28, no. 4, pp. 713-723, 2005.
- [107] Mohamed Meselhy Eltoukhy, Ibrahima Faye, Brahim Belhaouari Samir, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," *Computers in Biology and Medicine*, vol. 40, no. 4, April 2010, pp. 384-391.
- [108] E. Sakka, A. Prentza, I.E. Lamprinos, D. Koutsouris, "Microcalcification detection using multiresolution analysis based on wavelet transform," in *proceeding of the International Special Topic Conference on Information Technology in Biomedicine (IEEE-ITAB2006)*, Ioannina, Epirus, Greece, October 26–28, 2006.
- [109] Ju Cheng Yang, Jin Wook Shin, Dong Sun Park, "Comparing Study for Detecting Microcalcifications in Digital Mammogram Using Wavelets," *Intelligent Data Engineering and Automated Learning – IDEAL 2004, Lecture Notes in Computer Science*, vol. 3177, 2004, pp. 409-415.
- [110] R. Lippmann, "Pattern Classification using Neural Networks", *IEEE Communications Magazine*, pp. 48, November 1989.
- [111] Aijuan Dong; Baoying Wang, "Feature selection and analysis on mammogram classification," *Communications, Computers and Signal Processing*, *PacRim 2009. IEEE Pacific Rim Conference on*, vol., no., pp.731,735, 23-26 Aug. 2009.
- [112] Y. Sun, C.F. Babbs, E.J. Delp, "A Comparison of Feature Selection Methods for the Detection of Breast Cancers in Mammograms: Adaptive Sequential Floating Search vs. Genetic Algorithm," *IEEE-EMBS 2005. 27th Annual International Conference of the*, vol., no., pp.6532,6535, 17-18 Jan. 2006.
- [113] Gilad-Bachrach, Ran and Navot, Amir and Tishby, Naftali, "Margin Based Feature Selection - Theory and Algorithms," *Proceedings of the Twenty-first International Conference on Machine Learning, ICML '04*, 2004, Banff, Alberta, Canada.

- [114] Noelia Sánchez-Marroño, Amparo Alonso-Betanzos, María Tombilla-Sanromán, "Filter Methods for Feature Selection – A Comparative Study," *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, vol. 4881 of the series *Lecture Notes in Computer Science*, pp 178-187.
- [115] Jiliang Tang , Salem Alelyani , and Huan Liu, "Feature Selection for Classification: A Review," *Data Classification*. Jul 2014 , 37 -b64.
- [116] Alolfe, M.A; Mohamed, W.A; Youssef, A-B.M.; Kadah, Y.M.; Mohamed, AS., "Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms," *Radio Science Conference*, 2009. NRSC 2009. National , vol., no., pp.1,9, 17-19 March 2009.
- [117] "Student: The probable error of a mean," *Biometrika*, 1908, vol. 6, pp.1-25.
- [118] J. Zhang, J. Wang, H. Huacan, and S. Jiaguan, "A New Heuristic Reduct Algorithm Based on Rough Sets Theory," *4th International Conference, WAIM*, Chengdu, China, August 17-19, 2003, vol. 2762, 2003, pp 247-253.
- [119] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 11, no. 7, pp.674–693, 1989.
- [120] Q. Bai, "Analysis of Particle Swarm Optimization Algorithm," *Computer and Information Science*, vol. 3, no. 1, Pebruari 2010.
- [121] Jonas Mockus, "Application of Bayesian approach to numerical methods of global and stochastic optimization," *Journal of Global Optimization*, June 1994, vol. 4, no. 4, pp 347-365.
- [122] M. Tabassum and K. Mathew, "A Genetic Algorithm Analysis towards Optimization solutions," *Intern. Journal Digit. Info. Wireless Commun*, vol.4 no-1, pp. 124-142, 2014.
- [123] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga, "A review of feature selection techniques in bioinformatics, *Bioinformatics*," vol. 23, no. 19, August 24, 2007.
- [124] Yuan Ren and Guangchen Bai, "Determination of Optimal SVM Parameters by Using GA/PSO," *Journal of Computers*, vol. 5, no. 8, pp. 1160-1168, Aug., 2010.
- [125] X.C. Guo, J.H. Yang, C.G. Wu, C.Y. Wang and Y.C. Liang, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization," *Neurocomputing*, vol. 71, no. 16–18, Oct. , 2008, pp. 3211-3215.
- [126] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet and Sayan Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, pp. 131–159.
- [127] M. Dorigo, "Optimization, Learning and Natural Algorithms," Ph.D. thesis, Politecnico di Milano, Italy, 1992.
- [128] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [129] K. Polat, S. Sahan, H. Kodaz, S. Günes, "Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism," *Expert Systems with Applications*, vol. 32, no. 1, pp. 172-183, 2007.
- [130] J. Cho, M. Chun, and D. Lee, "Parameter Optimization of Extreme Learning Machine Using Bacterial Foraging Algorithm," in *International Symposium on Advanced Intelligent Systems*, 2007, pp. 742-747.

- [131] Fei Han, Hai-Fen Yao and Qing-Hua Ling, "An Improved Extreme Learning Machine Based on Particle Swarm Optimization," *Bio-Inspired Computing and Applications Lecture Notes in Computer Science*, vol. 6840, pp 699-704, 2012.
- [132] M. Lukoševičius, "A practical guide to applying echo state networks," in *Neural Networks: Tricks of the Trade*, 2nd ed., LNCS, vol. 7700, Springer, Heidelberg, 2012, pp. 659–686.
- [133] S. Basterrech, E. Alba and V. Snášel, "An experimental analysis of the Echo State Network initialization using the Particle Swarm Optimization," in *Nature and Biologically Inspired Computing (NaBIC)*, 2014 Sixth World Congress on, Porto, 2014, pp. 214-219.
- [134] A. Krause, V. Duerr, B. Blaesing and T. Schack, "Multiobjective optimization of echo state networks for multiple motor pattern learning," in *18th IEEE Workshop on Nonlinear Dynamics of Electronic Systems (NDES)*, 2010, pp. 190 – 193.
- [135] Fei Jiang, Hugues Berry, and Marc Schoenauer. "Supervised and Evolutionary Learning of Echo State Networks," in *Parallel Problem Solving from Nature – PPSN X Lecture Notes in Computer Science*, vol 5199, 2008, pp. 215-224.
- [136] Min Han and Meiling Xu, "Predicting Multivariate Time Series Using Subspace Echo State Network", *Neural Processing Letters*, vol. 41, no. 2, pp. 201-209, April, 2015.
- [137] Alexandre Devert, Nicolas Bredeche and Marc Schoenauer, "Unsupervised learning of Echo State Networks: A Case Study in Artificial Embryogeny," in *Artificial Evolution Lecture Notes in Computer Science*, vol. 4926, 2008, pp. 278-290.
- [138] Yuan Ren and Guangchen Bai, "Determination of Optimal SVM Parameters by Using GA/PSO," *Journal of Computers*, vol. 5, no. 8, pp. 1160-1168, Aug., 2010.
- [139] Kaizhu Huang, Danian Zheng, Jun Sun, Yoshinobu Hotta, Katsuhito Fujimoto and Satoshi Naoi, "Sparse learning for support vector classification," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1944-1951, Oct., 2010.
- [140] İlhan İlhan, Gülay Tezel, A genetic algorithm–support vector machine method with parameter optimization for selecting the tag SNPs," *Journal of Biomedical Informatics*, vol. 46, no. 2, April 2013, pp. 328-340.
- [141] Melody Y. Kiang, "A comparative assessment of classification methods," *Decision Support Systems*, vol. 35, no. 4, pp. 441-454, July, 2003.
- [142] Teuvo Kohonen, "Self-Organizing Maps," *Springer Series in Information Sciences*, vol. 30, 1995.
- [143] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational abilities," *Proc. Natl. Acad. Sci. USA*, 79, 2554-2558.
- [144] M. Dorigo, "Optimization, Learning and Natural Algorithms," Ph.D. thesis, Politecnico di Milano, Italy, 1992.
- [145] J.D. Farmer, N. Packard and A. Perelson, (1986) "The immune system, adaptation and machine learning", *Physica D*, vol. 2, pp. 187–204.
- [146] E. Romero, and F. González, "From Biomedical Image Analysis to Biomedical Image Understanding Using Machine Learning," in *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*, IGI Global. pp. 1-26, 2010.
- [147] Hetal Bhavsar, and Amit Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, 2012.

- [148] Sung-Nien Yu, Kuan-Yuei Li and Yu-Kun Huang, "Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model", *Computerized Medical Imaging and Graphics*, vol. 30, pp.163–173, 2006.
- [149] Muthu Rama Krishnan, Shuvo Banerjee, Chinmay Chakraborty, Chandan Chakraborty and Ajoy K. Ray, "Statistical analysis of mammographic features and its classification using support vector machine," *Expert Systems with Applications*, vol. 37, pp. 470-478, 2010.
- [150] Sung-Nien Yua, Kuan-Yuei Lib and Yu-Kun Huangac, "Detection of micro calcifications in digital mammograms using wavelet filter and Markov random field model," *Computerized Medical Imaging and Graphics*, vol. 30, no. 3, pp. 163-173, April, 2006.
- [151] Elif Derya Ubeyli, "Implementing automated diagnosis systems for breast cancer detection," *Expert Systems with Applications*, vol. 33, no. 4, pp.1054-1062, 2007.
- [152] Issam El-Naqa, Yongyi Yang, and Miles N. Wernick, "A Support Vector Machine Approach for Detection of Microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, Dec. 2002.
- [153] Cheng-Lung Huang, Hung-Chang Liao and Mu-Chen Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," *Expert Systems with Applications*, vol. 34, pp. 578–587, 2008.
- [154] Kemal Polat and Salih Güneş, "Breast cancer diagnosis using least square support vector machine," *Digital Signal Processing*, vol. 17, no. 4, pp. 694-701, July, 2007.
- [155] C. R. Huang, Y. T. Chen, W. Y. Chen, H. C. Cheng and B. S. Sheu, "Gastroesophageal Reflux Disease Diagnosis Using Hierarchical Heterogeneous Descriptor Fusion Support Vector Machine," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 588-599, March 2016.
- [156] Muhammad Hussain, Summrina Kanwal Wajid, Ali Elzaart and Mohamed A. Berbar, "A Comparison of SVM Kernel Functions for Breast Cancer Detection," in *CGIV*, 2011, pp. 145-150.
- [157] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-time epileptic seizure prediction using AR models and support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, May, 2010.
- [158] Ron Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2, pp. 1137-1143.
- [159] Tom Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, June 2006, pp. 861-874, ISSN 0167-8655, <http://dx.doi.org/10.1016/j.patrec.2005.10.010>.
- [160] A.P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recogn.*, 1997, vol. 30, no. 7, pp.1145–1159.
- [161] [https://en.wikibooks.org/wiki/Statistics/Testing\\_Data/t-tests](https://en.wikibooks.org/wiki/Statistics/Testing_Data/t-tests)
- [162] David M. Lane, [http://onlinestatbook.com/2/analysis\\_of\\_variance/intro.html](http://onlinestatbook.com/2/analysis_of_variance/intro.html)
- [163] M. S. Sarfraz and O. Hellwich, "Head pose estimation in face recognition across pose scenarios," *Int. Conference on Computer Vision Theory and Applications VISAPP*, vol. 1, 2009, pp. 235–242.
- [164] M. S. Sarfraz and O. Hellwich, "An efficient front-end facial pose estimation system for face recognition," *International Journal of Pattern Recognition and Image Analysis*, vol. 18, 2008, pp.434–441, 2008.

- [165] M. S. Sarfraz and O. Hellwich, "On head pose estimation in face recognition," *Computer Vision And Computer Graphics. Theory And Applications, Lecture Notes CCIS*, vol. 24, pp.162–175, 2009.
- [166] U. Zakir, I. Zafar, and E. A. Edirisingh, "Road sign detection and recognition by using Local Energy Based Shape Histogram (LESH)," *International Journal of Image Processing*, vol. 4, pp. 566–582, 2011.
- [167] M.C. Morrone and R.A. Owens, "A feature detection from local energy," *Pattern Recognition Letters* 6, pp.103–113, 1987.
- [168] P. D. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, pp. 1–26, 1999.
- [169] P. D. Kovesi, "Phase congruency: A low-level image invariant," *Psychological Research*, vol. 64, pp.136–148, 2000.
- [170] R.J. Ferrari, S. Allaire, A. Hope, J. Kim, D. Jaffray and V. Pekar, "Detection of point landmarks in 3D medical images via phase congruency model," *J. Braz. Comput. Soc.*, vol. 17, no. 2, pp. 117–132, 2011.
- [171] <http://reservoir-computing.org/node/129>
- [172] S. Creech, (2003).ANOVA, <http://www.statisticallysignificantconsulting.com/Anova.htm>. Accessed March 2014.
- [173] B. Verma, P. McLeod, and A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer," *Expert Syst Appl*, vol. 37, no. 4, pp. 3344–3351, 2010.
- [174] L.P Wong and H.T Ewe, "A Study of Lung Cancer Detection using Chest X-ray Images," 3rd APT Telemedicine Workshop, 27-28th Jan., 2005, Kuala Lumpur, Malaysia, pp. 210-214.
- [175] A. El-Baz, G. M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, A. Soliman, and B. Abdollahi, "Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies," *International Journal of Biomedical Imaging*, vol. 2013, Article ID 942353, 46 pages, 2013.
- [176] Heang-Ping Chan, Lubomir Hadjiiski, Chuan Zhou, Berkman Sahiner, "Computer-Aided Diagnosis of Lung Cancer and Pulmonary Embolism in Computed Tomography—A Review," *Academic Radiology*, vol. 15, no. 5, May 2008, pp. 535-555
- [177] Ma Jianwei Ma and G. Plonka, "The Curvelet Transform," *Signal Processing Magazine*, IEEE, vol. 27, no. 2, pp. 118-133, 2010.
- [178] R. Dosil, X.M. Pardo, X.R. Fdez-Vidal, "Decomposition of three-dimensional medical images into visual patterns," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 12, pp.2115–2118, 2005.
- [179] [http://en.wikipedia.org/wiki/Gabor\\_filter](http://en.wikipedia.org/wiki/Gabor_filter).
- [180] Samir Eljamel, "Intelligent Face Detection using a novel Gabor-filter based cascading network", BSc (Hons) Dissertation, University of Stirling, April 2014.
- [181] Imed Riadh Farah, Wadii Boulila, Karim Saheb Etabaâ, Basel Solaiman, Mohamed Ben Ahmed, "Interpretation of Multisensor Remote Sensing Images: Multiapproach Fusion of Uncertain Information," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 12, pp. 4142-4152, 2008.
- [182] Imed Riadh Farah, Wadii Boulila, Karim Saheb Etabaâ, Mohamed Ben Ahmed, "Multiapproach System Based on Fusion of Multispectral Images for Land-Cover Classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 12, pp. 4153-4161, 2008.

- [183] Wadii Boulila, Amine Bouatay, Imed Riadh Farah, "A Probabilistic Collocation Method for the Imperfection Propagation: Application to Land Cover Change Prediction," *JMPT*, vol. 5, no. 1, pp. 12-32, 2014.
- [184] Leandro Nunes de Castro, "Immunocomputing," in *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*.
- [185] Alan S. Perelson and George F. Oster, "Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination," *Journal of Theoretical Biology*, vol. 81, no. 4, pp. 645-670, 1979.
- [186] MA Mazurowski, JA Baker, HX Barnhart, & GD Tourassi, "Individualized computer-aided education in mammography based on user modeling: concept and preliminary experiments," *Medical Physics*, vol. 37, pp. 1152–1160. <http://doi.org/10.1118/1.3301575>, 2010.
- [187] MA Mazurowski, JA Baker, HX Barnhart, & GD Tourassi, "Identifying Error-making Patterns in Assessment of Mammographic BIRADS Descriptors among Radiology Residents Using Statistical Pattern Recognition," *Academic Radiology*, vol. 19, no. 7, pp. 865–871. <http://doi.org/10.1016/j.acra.2012.01.012>, 2012.
- [188] MA Mazurowski, JA Baker, HX Barnhart, & GD Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks*, vol. 21, no. 2-3, pp. 427–436. <http://doi.org/10.1016/j.neunet.2007.12.031>, 2008.
- [189] MA Mazurowski, J. Zhang, JY Lo, CM Kuzmiak, SV Ghate, & S. Yoon, "Modeling resident error-making patterns in detection of mammographic masses using computer-extracted image features: preliminary experiments," in *Proc. SPIE 9037, Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, 90370S (March 11, 2014); doi:10.1117/12.2044404
- [190] J. Zhang, JY Lo, CM Kuzmiak, SV Ghate, SC Yoon, & MA Mazurowski, "Using computer-extracted image features for modeling of error-making patterns in detection of mammographic masses among radiology residents," *Medical Physics*, vol. 41, no. 9, 91907, 2014.
- [191] J. Zhang, JI Silber, & MA Mazurowski, "Modeling false positive error making patterns in radiology trainees for improved mammography education," *Journal of Biomedical Informatics*, vol. 54, pp. 50–57, 2015.
- [192] Leandro Nunes de Castro, *Fundamentals of natural computing: an overview*, *Physics of Life Reviews*, Volume 4, Issue 1, March 2007, Pages 1-36, ISSN 1571-0645, <http://dx.doi.org/10.1016/j.plrev.2006.10.002>.
- [193] Jason Brownlee (2011), *Clever Algorithms: Nature-Inspired Programming Recipes* (1st ed.). Lulu.com.
- [194] FM., Burnet, "A modification of Jerne's theory of antibody production using the concept of clonal selection," *CA Cancer J Clin.*, 1976 Mar-Apr, vol. 26, no. 2, pp. 119-21.
- [195] NK Jerne, "Towards a Network Theory of the Immune System," *Ann. Immunol. (Inst. Pasteur)* 125C, 1974, pp. 373–389.
- [196] G. J. V. Nossal, "Life, Death and the Immune System," *Scientific American*, vol. 269, no. 3, pp. 21–30, 1993.
- [197] U. Storb, "Progress in Understanding the Mechanisms and Consequences of Somatic Hypermutation," *Immun. Rev.*, vol. 162, pp. 5–11, 1998.
- [198] L. N. de Castro, and F. J. Von Zuben, "aiNet: An Artificial Immune Network for Data Analysis", in *Data Mining: A Heuristic Approach 1*, pp.231-259.

- [199] Alan S. Perelson and George F. Oster, "Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination," *Journal of Theoretical Biology*, vol. 81, no. 4, pp. 645-670, 1979.
- [200] L. N. de Castro and J. Timmis, "An artificial immune network for multimodal function optimization," *Evolutionary Computation*, 2002. CEC '02. Proceedings of the 2002 Congress on, Honolulu, HI, 2002, pp. 699-704.
- [201] J. Timmis and C. Edmonds, "A Comment on opt-AINet: An Immune Network Algorithm for Optimisation," *In Lecture Notes in Computer Science*, vol. 3102, Springer, 2004, pp. 308-317.
- [202] Alexandre Devert, Nicolas Bredeche and Marc Schoenauer, "Unsupervised learning of Echo State Networks: A Case Study in Artificial Embryogeny," *in Artificial Evolution Lecture Notes in Computer Science*, vol. 4926, 2008, pp. 278-290.
- [203] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, "Nonlinear Programming Algorithms," in *Nonlinear Programming: Theory and Algorithms*, 2nd Ed., Wiley, 1993.
- [204] Fabricio Olivetti de França, Fernando J. Von Zuben, and Leandro Nunes de Castro, "An artificial immune network for multimodal function optimization on dynamic environments," in *Proc. of the 7th annual conference on Genetic and evolutionary computation (GECCO '05)*, ACM, New York, NY, USA, pp. 289-296.
- [205] A. Krause, V. Duerr, B. Blaesing and T. Schack, "Multiobjective optimization of echo state networks for multiple motor pattern learning," in *18th IEEE Workshop on Nonlinear Dynamics of Electronic Systems (NDES)*, 2010, pp. 190 – 193.
- [206] M. Lichman. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [207] Y. Wang, X. Gao and J. Li, "A Feature Analysis Approach to Mass Detection in Mammography Based on RF-SVM," *2007 IEEE International Conference on Image Processing*, San Antonio, TX, 2007, pp. V - 9-V - 12.
- [208] Mohamed Khalifa, "Clinical Decision Support: Strategies for Success," *Procedia Computer Science*, vol. 37, 2014, pp. 422-427, ISSN 1877-0509.
- [209] Shelda Mohan, M. Ravishankar, "Modified Contrast Limited Adaptive Histogram Equalization Based on Local Contrast Enhancement for Mammogram Images," *Mobile Communication and Power Engineering, Volume 296 of the series Communications in Computer and Information Science*. pp 397-403, 2013.
- [210] Summrina Kanwal Wajid and Amir Hussain, "Local energy-based shape histogram feature extraction technique for breast cancer diagnosis," *Expert Systems with Applications*, vol. 42, no. 20, Nov., 2015, pp. 6990-6999.
- [211] Summrina K. Wajid, A. Hussain, and B. Luo, "An efficient Computer Aided Decision Support System for breast cancer diagnosis using Echo State Network classifier," *Computational Intelligence in Healthcare and e-health (CICARE)*, 2014 IEEE Symposium on, Orlando, FL, 2014, pp. 17–24.
- [212] Summrina Kanwal Wajid and Amir Hussain, "Novel Artificial Immune Network-based Optimization of Machine Learning Classifiers," Submitted to *IEEE Trans. On Evolutionary Computation*, ISI Impact Factor (IF): 5.5.
- [213] Summrina Kanwal Wajid and Amir Hussain, "Local Energy-based Shape Histogram (LESH) Based Clinical Decision Support System for Breast Cancer Detection using Magnetic Resonance Imaging (MRI)," Submitted to *IEEE Trans. On Neural Networks and Learning Systems*, ISI Impact Factor (IF): 4.4.

- [214] Summrina Kanwal Wajid, Amir Hussain, Kaizhu Huang, and Wadii Boulila, "Lung Cancer detection using novel application of Local Energy-based Shape Histogram (LESH) Feature Extraction and Machine Learning Classification techniques," ICCI\*CC 2016, pp. 22-23 Aug, 2016, Stanford University, USA.
- [215] Summrina Kanwal Wajid, Amir Hussain, and Bin Luo, "An Investigation of Machine Learning and Neural Computation Paradigms in the Design of Clinical Decision Support Systems (CDSSs)," BICS, 2016.
- [216] H. Zhou, J Wu and J. Zhang. Digital Image Processing. Ventus Publishing, ISBN 978-87-7681-541-7, 2010.
- [217] Jianguo Zhang, Tieniu Tan, "Brief review of invariant texture analysis methods," *Pattern Recognition*, vol. 35/3, pp.735-747, 2002.
- [218] Bloch I, Heijmans H, Ronse C. Mathematical Morphology. In: Aiello M, Pratt-Hartmann I, Van Benthem J, editors. Handbook of Spatial Logics [Internet]. Dordrecht: Springer Netherlands; 2007. p. 857–944. Available from: [http://dx.doi.org/10.1007/978-1-4020-5587-4\\_14](http://dx.doi.org/10.1007/978-1-4020-5587-4_14).
- [219] Beyer H-G, Schwefel H-P. Evolution strategies – A comprehensive introduction. Nat Comput [Internet]. 2002 [cited 2017 Apr 12];1(1):3–52. Available from: <http://link.springer.com/10.1023/A:1015059928466>.
- [220] Rechenberg I (1965) Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Farnborough p. Library Translation 1122
- [221] Rechenberg I (1973) Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog Verlag, Stuttgart.
- [222] Schwefel H-P (1977) Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie, Interdisciplinary systems research; 26. Birkhäuser, Basel
- [223] Schwefel H-P (1981) Numerical Optimization of Computer Models. Wiley, Chichester.
- [224] Hartmann D (1974) Optimierung balkenartiger Zylinderschalen aus Stahlbeton mit elastischemund plastischem Werkstoffverhalten. Dr.-Ing. Dissertation, University of Dortmund, Dortmund.
- [225] Liu R, Liu E, Yang J, Li M, Wang F. Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search. In: Intelligent Control and Automation [Internet]. Springer Berlin Heidelberg; 2006 [cited 2017 Apr 12]. p. 712–21. Available from: [http://link.springer.com/10.1007/978-3-540-37256-1\\_87](http://link.springer.com/10.1007/978-3-540-37256-1_87)
- [226] Castellino RA. Computer aided detection (CAD): An overview. Cancer Imaging. 2005 Aug 23 [cited 2017 Apr 13];5(1):17–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16154813>.
- [227] M. L. Giger, N. Karssemeijer, and S. G. Armato, "Guest editorial computer-aided diagnosis in medical imaging," IEEE Trans. Med. Imaging, vol. 20, no. 12, pp. 1205–1208, Dec. 2001.
- [228] <https://en.wikipedia.org/wiki/Hyperparameter>
- [229] <https://www.oreilly.com/ideas/evaluating-machine-learning-models/page/5/hyperparameter-tuning>