CHAPTER 6

Breast Cancer Disease Diagnosis §

6.1. Introduc on

Breast cancer is a global problem. Day by day, the problem is intensified except some developed countries. More causality is reported worldwide due to this cause. Intelligent automated decision support systems have been attempted with varying accuracies for early detection of breast cancer. One of the important tools in this context is neural network. Various feature selection techniques have been deployed as prerequisite. For taking final decision by the physicians, these features play vital roles. Different feature selection or extraction techniques, namely, correlation-based feature selection (CFS), rough set based feature selection and principal component analysis (PCA) have been considered in this work. For classification, incremental back propagation learning network (IBPLN), and Levenberg-Marquardt (LM) classification tested on UCI breast cancer data base have been used here. Classification results are presented in terms of classification accuracy, specificity, sensitivity, and receiver-operating characteristics curve area (AUC). The experimental results are found to be satisfactory after performing the average of 100 to 120 simulations.

As per reports published by American Cancer Society, it is observed that 229,060 patients in United States are newly diagnosed with breast cancer only in the year 2012 [1]. It was also estimated that the number of death cases due to this cause was – 39,920 for both sexes; 410 (male), and 39,510 (female). Breast cancer is the second leading cause of death in women. Clinical and medical research, diagnostic procedures and equipments, decision support systems (DSS) are some of the major areas for exploration to manage the menace.

Early diagnosis is one of the prime factors to avoid any fatal situation. This might be the one of the reasons of survival of 2.5 million breast cancer patients in United States. Presently, the three methods available for detecting breast cancer: surgical biopsy, mammography, and FNAC (Fine needle aspiration cytology). Though the accuracy of surgical biopsy is nearly 100% but it is costly, invasive, time consuming and painful [2] and it also requires expertise in the particular domain, which is a real

§ This chapter is based on the publication made by the author entitled “An Intelligent Hybrid System for Breast Cancer Classification”, Proc. Nat. Conf. on Research Trends in Computer Sc. & Application (NCRTCSA), pp. 21 – 27, ISBN 978-93-82338-95-6 © 2014 Bonfiring, February, 2014.

Table 6.1.. Some Classi ers and Their Classi ca on Accuracies on UCI Database

Year	Author	Method	Accuracy(%)
1990	Mangasarian et al.[6]	LP (10-fold CV)	97.50
1996	Quinlan [9]	C4.5 (10-fold CV)	94.74
1996	Hamilton et al. [10]	RIAC (10-fold CV)	95.00
1996	Ster and Dobnikar [11]	LDA (10-fold CV)	96.80
1998	Bennett and Blue [12]	LS-SVM (5-fold CV)	97.20
1999	Nauck and Kruse [13]	NEFCLASS (10-fold CV)	95.06
1999	Pena-Reyes and Sipper [14]	Fuzzy-GA1 (train: 75% - test: 25%)	97.36
2000	Setiono [15]	Neuro-rule 2a (train: 50% - test: 50%)	98.10
2002	Goodman et al. [16]	Optimized -LVQ (10-fold CV)	96.70
		Big-LVQ (10-fold CV)	96.80
		AIRS (10-fold CV)	97.20
2003	Abonyi and Szeifert [17]	Supervised fuzzy clustering (10-fold CV)	95.57
2005	Polat et al. [18]	FS-AIRS (10-fold CV)	98.51
2005	Ubeyli [19]	ME (train: 37% - test: 63%)	98.85
2007	Sahan et al. [20]	Fuzzy-AIS-knn (10-fold CV)	99.14
2007	Ubeyli [21]	SVM (train: 37% - test: 63%)	99.54
2007	Polat and Gunes [22]	LS-SVM (10-fold CV)	98.53
2009	Akay [23]	SVM-CFS (train: 80% - test: 20%)	99.51
2009	Karabatak and Cevdet-Ince [24]	AR + NN (3-fold CV)	97.40
2009	Ubeyli [25]	ANFIS (train: 37% - test: 63%)	99.08
2010	Huang et al. [26]	SBS-BPPSO (10-fold CV)	97.51
		SBS-BPLM (10-fold CV)	98.83
2011	Marcano-Cedeno et al. [27]	AMMLP (train: 60% - test: 40%)	99.26
2011	Marcano-Cedeno et al. [28]	AMMLP (train: 60% - test: 40%) average	99.58
		AMMLP (train: 60% - test: 40%) best	99.63
2011	Palanivel and Kumaravel [29]	Adaptive SVM (train: 80% - test: 20%)	99.87
2011	Fan et al. [30]	CBFDT (train: 50% - test: 50%)	98.90
2011a	Chen et al. [31]	RS-SVM (train: 80% - test: 20%) highest	100
		RS-SVM (train: 50% - test: 50%) average	96.87
2011	Xu et al. [32]	Kernel-based orthogonal transform (10-fold CV)	98.53
		Kernel-based orthogonal transform (train: 50% - test: 50%)	98.53
		Kernel-based orthogonal transform (train: 80% - test: 20%)	100
		Kernel-based orthogonal transform (train: 70% - test: 30%)	99.51
2011	Chen et al. [33]	PSO-SVM (10-fold CV)	99.3
2012	Our study	PCA-LM (train 68%-validation 16%-testing 16%) highest	100
		PCA-LM (train 68%-validation 16%-testing 16%) lowest	96.08
		PCA-LM (train 68%-validation 16%-testing 16%) average	98.55

problem. Accuracy of mammography, and FNAC varies from 68% to 79%, and 65% to 98% respectively [3][4]. Moreover, the visual interpretation requires experience of a doctor. So, a better automated diagnostic tool is solicited to diagnose breast cancer.

With the establishment of UCI machine learning repository [5], where one can find (i) Breast cancer Wisconsin (Original) data set, (ii) Breast cancer Wisconsin (Diagnostic) data set, (iii) Breast cancer Wisconsin (Prognostic) data set; various techniques and tools are being explored in the form of data mining for better diagnosis of breast cancer. Starting from 'Cancer diagnosis via linear programming' [6] in the year 1990, we find various data mining approaches in literature using 'machine learning / intelligent systems' tagged with various classification accuracies. Much of the studies used breast cancer Wisconsin (original) data set; an year wise development comparison of which is presented in Table 6.1. However, there are some studies [7] using breast cancer Wisconsin(diagnostic) data set as well as breast cancer Wisconsin (prognostic) data set. In addition to the UCI data sets, peoples use other cancer data sets as well [8]. This study considers breast cancer Wisconsin (original) data set; the details of which are discussed in section 2.

From Table 6.1, we find that various techniques have been attempted primarily to reach high classification accuracy. At the same time, easy implementation, low computational costs were the issues of consideration of studies. Linear programming (LP), decision tree, RIAC, linear discrete analysis (LDA), neuro-fuzzy, fuzzy-GA, learning vector quantization (LVQ), feed forward neural network rule extraction, combined logarithmic simulated anneling with perceptron, supervised fuzzy clustering, support vector machine (SVM), probabilistic neural network, recurrent neural network, combined neural network, multilayer perceptron neural network, least square support vector machine (LS - SVM), linear least squares, particle swarm optimization (PSO), association rules combined with neural networks (AR+NN), Kernel-based orthogonal transform, artificial metaplasticity neural network (AMMLP) are the different methods used in the studies listed in Table 6.1.

A few observations are noted here from Table 6.1. *First,* much of the works having high classification accuracy are based on hybrid approach. *Second*, neural network and SVM with various combinations with least square, f-score, rough set, fuzzy-artificial immune system lead to higher classification accuracy. *Third,* 10-fold cross-validation and train x% - test (100-x)% validation techniques have been deployed. Moreover, different combinations of train x% - test (100-x)% have been used; the rationale of which is not clear. *Fourth,* it is not clear from some of the papers that the accuracy they achieved are the results of the best simulation product or are an

average of several simulations [28]. *Fi h*, some studies used 699 instances with using replacements of missing values by mean value of the attribute [33]. Others use only 683 instances by excluding the instances with missing values. *Sixth,* different studies suggest different sets of reduced features which were combined with different classification algorithms producing good accuracy levels. When two or more different reduced sets of features produce same or comparable classification performance with the same classification algorithm; it is a question which set should be taken into consideration for further scrutiny by the physicians for final decision making. *Last,* but not the least, some studies pointed out the highest (100%) performance of a particular simulation but did not point out the lowest achievable performance of a simulation. In medical diagnosis, where there is a question of life and death, it is rather more important to see the worst performance of a system than a particular best simulation result.

In this study, we attempt four combinations (i) CFS + LM (ii) RS + LM (i) PCA + IBPLN, (ii) PCA + LM i.e. features are selected using correlation-based feature subset selection, rough set, and principle component analysis methods, and incremental backpropagation learning network and Levenberg – Marquardt algorithms are used as classifier. We focus on the worst simulation result, the best simulation result as well as an average of 100 simulations for combinations (i) and (ii) and 120 simulations for combinations (iii) and (iv).

The rest of the study is organized as follows. Section 2 discusses the Wisconsin breast cancer database. Section 3 discusses the applications. Section 4 discusses preliminaries of feature selection techniques such as CFS, RS, and PCA as data preprocessing steps. Section 5 discusses the preliminaries of ANN, IBPLN, and LM classification algorithms. Section 6 presents the modeling results. Section 7 specifies benchmarking parameters used to evaluate the performance of model. Experimental results are presented in section 8. Lastly, our conclusions are summarized.

6.2. Wisconsin Breast Cancer Database

This database was contributed by Dr. William H. Wolberg (1989-1991), University of Wisconsin Hospitals, Madison, USA [1]. The records came periodically as Dr. Wolberg reports his clinical cases. The total number of instances is 699; and the number of attributes is 10 plus the class attribute. As the class distribution, there are 458 (65.5%) Benign cases and 241 (34.5%) Malignant cases. There are 16 instances with missing values in the data set. The attribute information and statistics are shown in

Table 6.2. In our study, we discard these 16 instances with missing values and use 683 instances for analysis.

Table 6.2. Wisconsin Breast Cancer Data with Sta s cs

#	Attribute	Domain	Mean	Standard deviation
1.	Sample code number	id number	-	-
2.	Clump Thickness	1-10	4.44	2.83
3.	Uniformity of cell size	1-10	3.15	3.07
4.	Uniformity of cell shape	1-10	3.22	2.99
5.	Marginal adhesion	1-10	2.83	2.86
6.	Single epithelial cell size	1-10	2.23	2.22
7.	Bare nuclei	1-10	3.54	3.64
8.	Bland Chromatin	1-10	3.45	2.45
9.	Normal nuclei	1-10	2.87	3.05
10.	Mitosis	1-10	1.60	1.73
11.	Class	(2 for benign, 4 for malignant)		

6.3. Applica on

Basically, this study consists of two stages: the feature extraction and reduction phase by correlation based feature subset selection (CFS), rough set (RS), and principal component analysis (PCA); and classification phase by incremental back propagation learning networks (IBPLN), and Levenberg-Marquardt (LM) algorithms. The schematic view of our system is shown in Figure 6.1.

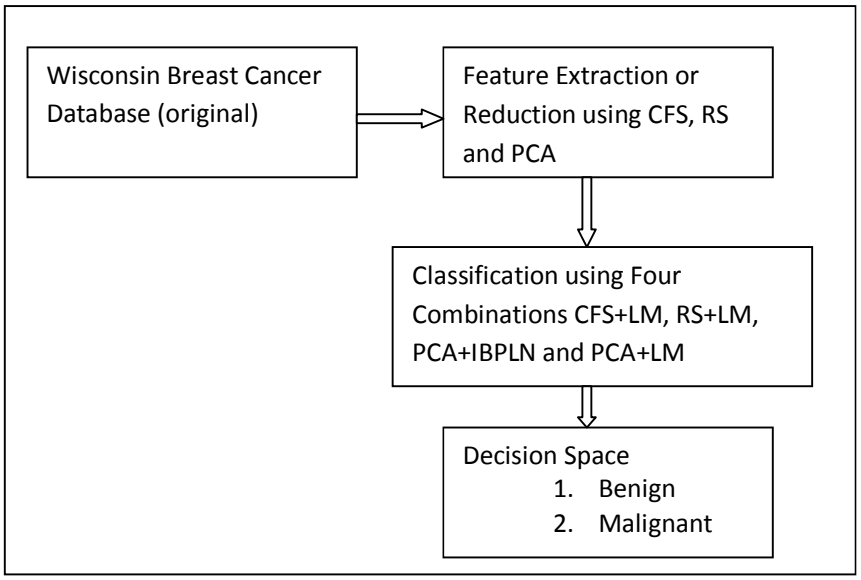

Figure 6.1. Block Diagram of a System for Breast Cancer Diagnosis.

6.4. Data Preprocessing

Though often neglected, data preprocessing is a vital step in data mining. Raw data may contain noise, missing value, outliers, and inconsistencies. During training phase, it is very much difficult to extract knowledge from data which are noisy and unreliable. The objective of data preprocessing is to improve the quality of data as well as the mining results. Different steps of data mining process are discussed in chapter 4.

There are 699 instances in Wisconsin breast cancer data set. 683 instances are taken into consideration after deleting 16 instances which contain missing value. We completely randomize the data sets after missing records deletion. There is no outlier in our data. The whole data set is partitioned into three: Training set (68%), Validation set (16%), and Test set (16%).

6.4.1. Feature Extrac on or Reduc on

Feature extraction or reduction or simply dimensional reduction is one of the strategies of data reduction where redundant, irrelevant and weakly relevant features are identified and removed. This step is needed to improve the performance of intelligent model. This work uses following three strategies to extract features.

- Correlation based feature subset selection (CFS)
- Rough Set (RS)
- Principal Component Analysis (PCA)

6.4.1.A. Correla on based Feature Subset Selec on (CFS)

Correlation based feature subset selection technique evaluates the relevancy of a subset of features in determining a class. In this process the relevancy of each individual features along with the inter-correlation between the features are taken into account. CFS algorithm identifies a subset of features which are relevant to a class and not redundant to any other features. It means that features, selected in the subset, are highly correlated with the class, but uncorrelated with other features. Based on test theory [18], the feature evaluation formula is defined as given below (Eq. 1.):

$$r_{fc} = \frac{k\overline{r_{fc}}}{\sqrt{\{k + k(k\_1)\,\overline{r_{ff}}\}}} \qquad (1)$$

where, k, $r_{fc,}$ $\overline{r_{fc}}$ and $\overline{r_{ff}}$ are number of attributes, correlation between the summed features, average of correlations between the features and the class variable and average of inter-correlation between features. As a preprocessing step supervised discretisation method is used to transform continuous features into categorical features. This is done to include nominal or categorical and also continuous or ordinal features in Eq. 1 [34]. The evaluation of the degree of relationship between nominal features has been done by using the theory of information gain [35]. Different heuristic search strategies like best-first and hill-climbing [36] are often implemented to identify feature subset in reasonable amount of time.

We implement CFS algorithm as feature evaluator. Out of 10 attributes of the said data set, 5 attributes are selected after applying CFS algorithm as shown in Table 6.3.

Table 6.3. Reduced feature sets applying CFS

Sr. No.	Reduced Attributes (CFS)
1.	Clump Thickness
2.	Uniformity of cell size
3.	Bare nuclei
4.	Bland Chromatin
5.	Normal nuclei

6.4.1.B. Feature Selec on using Rough Set

Rough set theory (RST) can be used as a tool to determine minimal feature subset from a data set that represents the original features. RST can be used to find out data dependencies and to reduce dimension of data set using only the data in the original data set without requiring any additional information (Pawlak, 1991; Polkowski, 2002). Recently, RST has been implemented in different areas of problem domain and is an important topic to researchers. RST can be used to determine a subset of original attributes (called reduct) with discretized values which represents the whole data set and thus other attributes can be removed without any loss of information.

Let I=(U,A) is an information system, where U is a nonempty finite set of objects (Universe of discourse) and A is a nonempty finite set of attributes. IND(P) is an associated equivalence relation for any $P \in A$, which is as follows:

$$IND(P) = \{(x,y) \in U^2 \mid \forall_a \in P, a(x) = a(y)\} \quad (2)$$

x and y are indiscernable by attribute from P if $(x,y) \in IND(P)$.

Let us consider $X \subseteq U$. By constructing P-lower ($\underline{P}X$) and P-upper (PX) approximations of X, X can be approximated using the information contained within P. Where:

$$\underline{P}X = \{x \mid [x]_P \subseteq X\} \quad (3)$$

$$PX = \{x \mid [x]_P \cap X \neq \oplus\} \quad (4)$$

If P and Q are in equivalence relation over U, the positive region can be defined as:

$$POS_P(Q) = \bigcup_{x \in U/Q} \underline{P}X \qquad (5)$$

If P, Q ∈ A, Q is said to be dependent on P and the rough set degree of dependency K, where $0 \leq K \leq 1$ are defined using the definition of positive region as give below:

$$K = Y_P(Q) = |POS_P(Q)| \; / \; |U| \qquad (6)$$

If C is the set of conditional attributes and D is the decision attribute, a feature $x \in C$ is indispensable in P, if $Y_P(D) = Y_{P-x}(D)$; otherwise x is indispensable with respect to D. A set $B \in C$ is independent if all of its attributes are indispensable. A set of features $R \in C$, is said to be reduct of C if R is independent and $POS_R(D) = POS_C(D)$.

We apply rough set as attribute evaluators. The reduced features sets by applying rough set are shown in Table 6.4. We adopted the methodology of Chen et al. [31] for rough set reduction.

Table 6.4. Reduced feature sets applying rough set

Sr. No.	Reduced Attributes (rough set)
1.	Clump Thickness
2.	Uniformity of cell shape
3.	Bare nuclei
4.	Marginal adhesion
5.	Mitosis

6.4.1.C. Feature Selec on using Principal Component Analysis (PCA)

It is very much difficult to identify any pattern in data for a data set of huge dimension. Principal component analysis (PCA) is a tool to find a pattern in such cases efficiently. After discovering the pattern, the data set can be compressed by reducing the dimension of it and retaining the information of the original data set. PCA is one of the most powerful methods to analyze data. The steps in PCA are as follows:

- For each attribute in the given data set subtract the mean value of the corresponding attribute from the original value of it.
- Evaluate the covariance matrix.
- Compute the eigenvectors and eigenvalues of the covariance matrix
- Eigenvectors are arranged by their eigenvalues in descending order.
- If the dimension of the data set be n, from n number of eigenvectors, one may choose m number of eigenvectors by neglecting the eigenvectors with smaller eigenvalues. It can be done as removing the eigenvectors with smaller eigenvalues do not cause much loss of information. Thus the final data set reduced to m dimension and that can be given as follows:

$$\text{Final Data Set} = \text{Row\_Feature\_Vector X Row\_Data\_Adjust} \quad (7)$$

where, Row_Feature_Vector is a matrix in which each row represents a column transposed eigenvector and eigenvalues of the eigen vector corresponding to the i^{th} row is greater than that of $(i+1)^{th}$ row. The Row_Data_Adjust is the mean adjusted data transposed.

We apply PCA as attribute evaluators. We got seven extracted features after applying PCA as follows : 0.381Unif_cell_size + 0.378Unif_cell_shape + 0.346Bland_Chromatin + 0.336Sing_Epi_cell_size+0.336Normal_Nucleoli..., 0.906Mitoses-0.261Bare_nuclei-.228Bland_Chromatin + 0.164Sing_Epi_cell_size - 0.141Clump_Thickness..., -0.866Clump_Thickness+0.413Margi_Adhesion + 0.213Bland_Chromatin+0.134Normal_Nucleoli + 0.088Sing_Epi_cell_size..., - 0.499Bare_nuclei-0.493Margi_Adhesion + 0.427 Sing_Epi_cell_size + 0.417Normal_Nucleoli-0.259Mitoses..., '0.69 Normal_Nucleoli - 0.637 Sing_Epi_cell_size + 0.228 Bland_Chromatin-0.146Unif_cell_size-0.125Bare_nuclei...', 0.655Margi_Adhesion - 0.609Bare_nuclei-0.299Bland_Chromatin+0.243Clump_Thickness - 0.148 Mitoses..., '-0.7Bland_Chromatin+0.46 Normal_Nucleoli + 0.403Bare_nuclei + 0.211Sing_Epi_cell_size-0.205Unif_cell_size...'.

6.5. Ar cial Neural Network Architecture

Recently, artificial neural network (ANN) has been used as a powerful tool to medical domain. The most useful application in this domain is to analyze complex clinical data for classification problems that is ANN is used to identify in which class a patient can be assigned. ANN is a computational model containing a number of interlinked processing elements, called nodes, which are able to operate in parallel like human brain. The models are designed by the inspiration of biological nervous system. The different tasks involved in designing ANN model are:

- deciding which type of the network is to be considered – feedback or feedforward.
- deciding the number of layers in the network.
- deciding the number of neurons in each layer.
- deciding the activation function of node.
- deciding the way in which the nodes are interlinked.

In general, balancing the trade-off between accuracy and generalizability is the prime characteristic of selecting a model. The ANN model selection includes choice of network architecture and feature selection. The hold-out data set called the *valida on set* would be useful helping all these decisions successful [37]. Validation set is a part of our data used to tune the network topology or network parameters other than weights. In our networks, we use logistic function of the form $F(x) = 1/ (1+e^{-x})$ in the hidden and output nodes. Theoretically, a network with one hidden layer and logistic function as the activation function at the hidden and output nodes is capable of approximating any function arbitrarily closely, provided that the number of hidden nodes are large enough [38]. So, we use one input layer, one hidden layer, and one output layer. The number of neurons in ANN is always a problem; too few hidden nodes as well as too many hidden nodes have certain problems. To overcome the problem, the formula proposed by Goa [39] was used in our study. The said formula is as follows:

$$s = \sqrt{(a_1 m^2 + a_2 mn + a_3 n^2 + a_4 m + a_5 n + a_6)} + a_7 \qquad (8)$$

where s is the number of neuron, m is the number of input, n is the number of output, a_1 ~ a_7 are undefined coefficients. Using least mean square (LMS), Huang et al. [26] derived the following formula:

$$s= \sqrt{} (0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35) + 0.51 \quad (9)$$

In the present study for the combinations CFS + LM and RS + LM, the value of m = 5, n = 2; and hence s = 5 after round off. So in this study we use 5 neurons at the hidden layer. The corresponding network architecture is shown in Figure 6.2. For another approach where PCA is used to extract features, the values of m and n are 7 and 2 respectively; and hence s= 6 after round off. So, in this study, we use six neurons at the hidden layer for all combinations, which is shown in Figure 6.3.

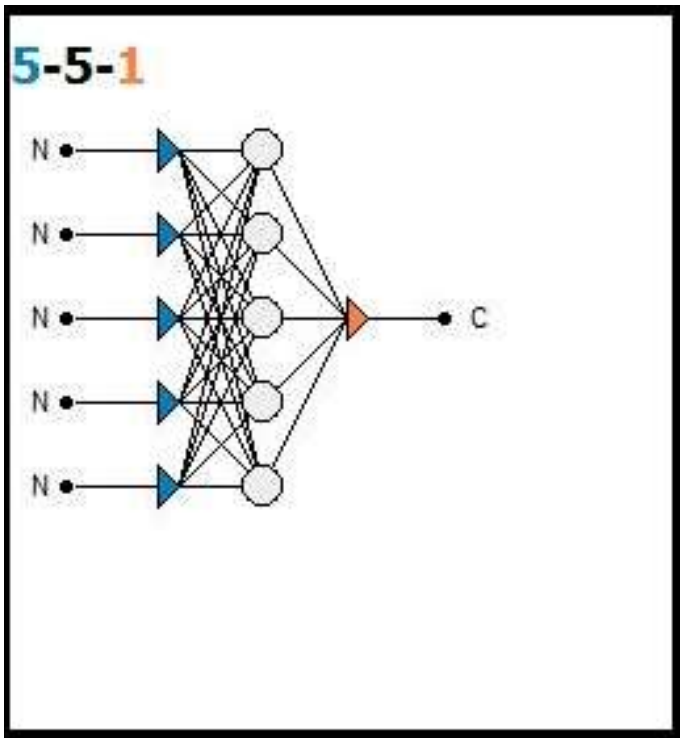

Figure 6.2. ANN Architecture for the Combina ons CFS + LM and RS + LM

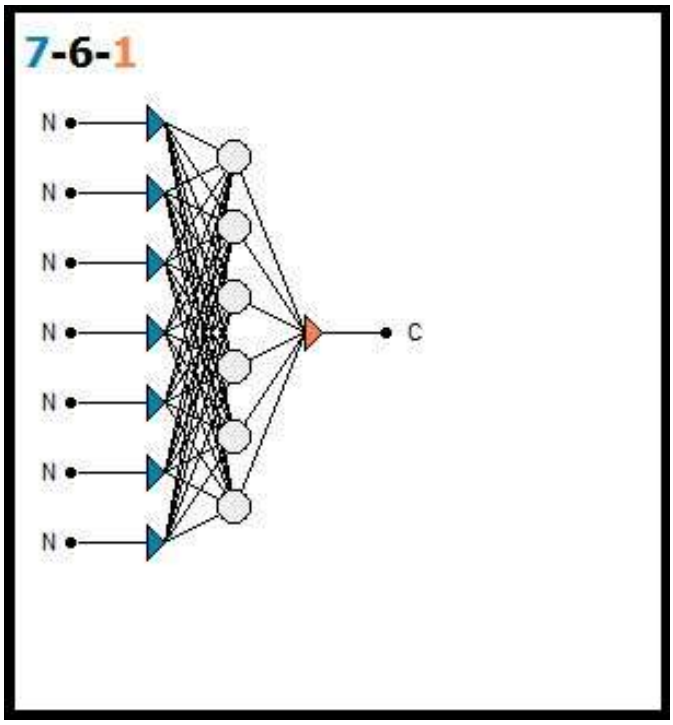

Figure 6.3. ANN Architecture for the Combina ons PCA + IBPLN and PCA + LM

The final step is to adjust the link-weights and threshold values based on training data set. Hundreds of training or learning algorithms are available. Two algorithms, namely, incremental back propagation learning network (IBPLN) and Levenberg-Marquardt (LM) algorithms are used in this study.

6.5.1. Incremental Back Propaga on Learning Network

The main issue of incremental learning network design is the way of learning new knowledge without forgetting the previous knowledge [40]. Incremental backpropagation learning network (IBPLN) is the modification of incremental learning method, which employs bounded weight modification, structural adaptation and applies initial knowledge to constrain the learning process. IBPLN is not like incremental by nature. The network is modified for the arrival of new instance and stored the previous one if it is not in conflict with the new. IBPLN algorithm is discussed in chapter 3.

6.5.2. Levenberg-Marquardt Algorithm

Levenberg-Marquardt (LM) algorithm is considered as a standard technique to solve non-linear least square problems. A parameterized function can be fitted into a set of data points by minimizing the sum of squares of errors between the data points and the function. The least square problem is non-linear if the function which has been fitted is not linear. LM algorithm is a repetitive method to minimize the sum of squares of errors between the data points and the function through a sequence of updation of values of parameters. LM algorithm is basically a curve fitting technique which combines two methods of minimization, namely, Gauss-Newton method and gradient descent method. In case of Gauss-Newton method the sum of squares of errors is minimized by finding the minimum of quadratic function, which is considered as the least square function. Updation of parameters in the steepest descent direction has been done in the case of gradient descent method to minimize the sum of squares of errors. These two minimization techniques are combined into LM method. LM algorithm has been discussed in chapter 3.

6.6. Modeling Results

WEKA [41] was used for feature set reduction using CFS and PCA. RSES [42] was used for feature reduction using rough set. The classification algorithms using four combinations were implemented in Alyuda NeuroIntelligence [43]. All the mentioned tasks were executed on Intel Core Solo T1350 CPU (1.86GHz, 533MHz FSB, 2MB L2 cache) with 512MB DDR2 RAM.

Table 6.5 and Table 6.6 show the network structure, epochs, numbers of retains, and numbers patterns used in training, validation and testing phases for different combinations. As overtraining control measure, we retain the copy of the network with the lowest validation error.

TABLE 6.5. Network Parameters Applying to WBCD

Types Classifiers	Network structure			Epochs (retrain)	Numbers patterns		
	I	HL	O		Training	Validation	Testing
CFS + LM	5	5	1	2000(10)	465	109	109
rough set + LM	5	5	1	2000(10)	465	109	109

Table 6.6. Network Parameters Applying to WBCD.

Types Classifiers	Network structure			Epochs(retrain)	Numbers patterns		
	I	HL	O		Training	Validation	Testing
PCA + IBPLN	7	6	1	2000(10)	465	109	109
PCA + LM	7	6	1	2000(10)	465	109	109

6.7. Performance Measure

The classification accuracy, sensitivity, specificity and area under ROC curve (AUC) have been computed as performance measure. The formulations are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \text{ x } 100\% \quad (10)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \text{ x } 100\% \quad (11)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \text{ x } 100\% \quad (12)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives respectively. The area under ROC curve (AUC) is an important measure of classification performance that is being used in biomedical research to assess the performance of diagnostic tests [44]. AUC close to one indicates more reliable diagnostic result.

6.8. Experimental Results

The compiled results from 120 simulations of our studies using the combinations CFS + LM and RS + LM are shown in Table 6.7 [45].

The following observations are noted below:

- Out of the two methods, CFS+LM shows better performance in terms of CCR, Specificity, Sensitivity, and AUC.
- The methods, used here, provide 100% classification accuracy as the highest performance using 68% training, 16% validation, and 16% testing data sets. This is comparable to other similar studies [31], [32].
- The lowest classification performance is 94.29% for the combination CFS + LM. We could not compare this result as there is no such data available from the literature of this kind.

The compiled results from 100 simulations of our studies using combinations PCA + IBPLN and PCA + LM are shown in Table 6.8.

The following observations are noted below:

- Out of the two methods, PCA+LM shows best performance in terms of CCR, Specificity, Sensitivity, and AUC.
- All of the methods provide 100% classification accuracy as the highest performance using 68% training, 16% validation, and 16% testing data sets. This is comparable to other similar studies [31], [32].
- The lowest classification performance is 96.08%. This result could not be compared due to unavailability of any such result obtained from previous studies.

- **TABLE 6.7. Results from 120 Simula ons [45]**

Methods	Test set (CCR%)			Specificity			Sensitivity			AUC		
	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg
CFS +LM	**100(6)**	**94.29(4)**	**97.45**	**100(19)**	**84.21(1)**	**95.28**	**100(38)**	**94.20(1)**	**98.53**	**100(10)**	**94(1)**	**99.27**
Rough set+LM	100(3)	93.33(1)	97.23	100(16)	83.87(1)	94.94	100(40)	93.42(1)	98.46	100(5)	93(1)	99.11

TABLE 6.8. Results from 100 Simula ons

Methods	Test set (CCR%)			Specificity			Sensitivity			AUC		
	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg
PCA +IBPLN	100(10)	95.1(1)	98.04	100(27)	85.71(1)	96.36	100(45)	95.65(1)	98.86	100(18)	97(1)	99.57
PCA+LM	**100(16)**	**96.08(1)**	**98.55**	**100(32)**	**89.29(1)**	**96.92**	**100(60)**	**97.06(1)**	**99.31**	**100(29)**	**98(3)**	**99.65**

6.9. Conclusion

Four combinations of intelligent diagnostic system have been tried for breast cancer diagnosis. It is noted that none of the previous works pointed out the lowest performance of their systems. We argue that the lowest performance of an intelligent system, especially in the field of medical diagnosis, should also be a judging parameter for the several simulations [28]. So, we present here the highest, lowest, and the average behavior of the methods used.

This work provides better result as compared to the results of Chen et al. [31]. In a significant number of cases, specificity, sensitivity, and AUC have reached 100% which are comparable to other similar studies [31], [32]. While this work does not claim the highest performance achiever, but at the same time reports that a combination of seven extracted features derived using PCA would have been worthwhile when the final decision is made by the doctors. Moreover, the lowest achievable performance with 96.08% classification accuracy is relatively better than classical techniques. The lowest, the highest, and the average performance of a DSS should be judged by a user of the system before using the same.

References

1. http://www.cancer.org/Cancer/BreastCancer/OverviewGuide.
2. http://cgm.cs.mcgill.ca/~beezer/cs644/cancer.html.
3. Fletcher S. W., Black W., Harris R., Rimer, B. K. & Shapiro S., "Report of the International Workshop on Screening for Breast cancer", *J. of the Na onal Cancer Ins tute,* Vol. 85, pp -1644-1656, 1993.
4. Giard R. W. M., & Hermans J., "The Value of Aspiration Cytologic Examination of the Breast", A statistical review of the medical literature, *Cancer,* Vol. 96, pp - 2104-2110, 1992.
5. http://archive.ics.uci.edu/ml/datasets/UCI Machine Learning Repository.
6. Mangasarian O. L., & Wolberg W. H., "Cancer Diagnosis via Linear Programming", *SIAM News*, Vol. 23(5), pp. 1 – 18, 1990.
7. Peng Y., Wu Z. & Jiang J., "A Novel Feature Selection Approach for Biomedical Data Classification", *Journal of Biomedical Informa cs,* Vol. 43, pp. 15-23, 2010.
8. Thongkam J., Xu G., Zhang Y., & Huang F., "Breast Cancer Survivability via AdaBoost Algorithms, *Australian Workshop on Health Data and Knowledge Management (HDKM 2008)"*, Wollongong, NSW, Australia, 2008.
9. Quinlan J. R., "Improved use of Continuous Attributes in C4.5", *Journal of Ar cial Intelligence Research,* Vol. 4, pp. 77 – 90, 1996.
10. Hamilton H. J., Shan N., & Cercone N., "RIAC: A Rule Induction Algorithm based on Approximate Classification", *Technical Report, CS 96-06, University of Regina*, 1996.
11. Ster, B. & Debonikar, A., "Neural Networks in Medical Diagnosis: Comparison with Other Methods", *Proc. of the interna onal conference on engineering applica ons of neural networks*, pp. 427-430, 1996.
12. Bennett K. P., & Blue J. A., "A Support Vector Machine Approach to Decision Trees", N*eural Networks Proceedings,* Vol. 3, pp. 2396-2401, 1998.
13. Nauck D., & Kruse R., "Obtaining Interpretable Fuzzy Classification Rules from Medical Data"*, Ar cial Intelligence in Medicine*, Vol. 16(2), pp. 149-169, 1999.
14. Penna-Reyes C. A. & Sipper M., "A Fuzzy-Genetic Approach to Breast Cancer Diagnosis", *Ar cial Intelligence in Medicine,* Vol. 17(2), pp. 131-155, 1999.

15. Setiono R., "Generating Concise and Accurate Classification Rules for Breast Cancer Diagnosis", *Ar cial Intelligence in Medicine,* Vol. 18(3), pp. 205-219, 2000.

16. Goodman D. E., Boggess L. C., & Watkins A. B., "Artificial Immune System Classification of Multiple-Class Problems", *Proc. of the ar cial neural networks in engineering* (ANNIE), pp. 179-183, 2002.

17. Abonyi, J., & Szeifert, F., "Supervised Fuzzy Clustering for the Identification of Fuzzy Classifiers", *Pa ern Recogni on Le ers,* Vol. 24(14), 2195-2207, 2003.

18. K. Polat, S. Sahan, H. Kodaz, and S. Gunes, "A New Classification Method for Breast Cancer Diagnosis: Feature Selection Artificial Immune Recognition System (FS-AIRS)", ICNC 2005, LNCS 3611, pp. 830-838, 2005.

19. Ubeyli E. D., "A Mixture of Experts Network Structure for Breast Cancer Diagnosis", *J. Med. Systems,* Vol. 29(5), pp. 569-579, 2005.

20. Sahan S., "A New Hybrid Method based on Fuzzy-Artificial Immune System and K-NN Algorithm for Breast Cancer Diagnosis", *Comput Biol Med,* Vol. 37(3), pp. 415-423, 2007.

21. Ubeyli E. D., "Implementing Automated Diagnostic Systems for Breast Cancer Detection", *Expert systems with applica ons,* Vol. 33(4), pp. 1054 – 1062, 2007.

22. Polat K. & Gunes S., "Breast Cancer Diagnosis using Least Square Support Vector Machine", *Digital Signal Processing,* Vol. 17(4), pp. 694-701, 2007.

23. Akay M. F., "Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis", *Expert Systems with Applica ons,* Vol. 36(2), pp. 3240-3247, 2009.

24. Karabatak M., & Ince M. C., "An Expert System for Detection of Breast Cancer based on Association Rules and Neural Network", *Expert Systems with Applica ons,* Vol. 36(2), pp. 3465-3469, 2009.

25. Ubeyli E. D., "Adaptive Neuro-Fuzzy Inference Systems for Automatic Detection of Breast Cancer", *J Med Syst,* Vol. 33(5), pp. 353-358, 2009.

26. Huang M. L., Hung Y. H., & Chen W. Y., "Neural Network Classifier with Entropy Based Feature Selection on Breast Cancer Diagnosis", *J Med Syst,* Vol. 34(5), pp. 865-873, 2010.

27. Marcano-Cedeno A., Quintanilla-Dominguez J., & Andina D., "WBCD Breast Cancer Database Classification Applying Artificial Metaplasticity Neural Network", *Expert Systems with Applica ons*, Vol. 38, pp. 9573-9579, 2011.

28. Marcano-Cedeno A., Quintanilla-Dominguez J., & Andina D., "Breast Cancer Classification Applying Artificial Metaplasticity Algorithm", *Neurocompu ng,* Vol. 74, pp. 1243-1250, 2011.

29. Palanivel J., & Kumaravel N., "An Efficient Breast Cancer Screening System Based on Adaptive Support Vector Machines with Fuzzy C-means Clustering", *European J of Scien c Research,* Vol. 51(1), pp. 115-123, 2011.

30. Fan C. Y., Chang P. C., Lin J. J., & Hsieh J. C., "A Hybrid Model Combining Case-based Reasoning and Fuzzy Decision Tree for Medical Data Classification", *Applied So Compu ng,* Vol. 11(1), pp. 632-644, 2011.

31. Chen H. L., Yang B., Liu J., & Liu D. Y., "A Support Vector Machine Classifier with Rough Set based Feature Selection for Breast Cancer Diagnosis", Expert Syst Appl, Vol. 38(7), pp. 9014-9022, 2011.

32. Xu Y., Zhu Q., & Wang J., "Breast Cancer Diagnosis Based on a Kernel Orthogonal Transform", *Neural Comput and Applic,* DOI 10.1007 /s00521-011-0547-0, 2011.

33. Chen H. L., Yang B., Wang G., Wang S. J., Liu J., & Liu D. Y., "Support Vector Machine based Diagnostic System for Breast Cancer using Swarm intelligence", *J Med Syst*, DOI 10.1007/s10916-011-9723-0, 2011.

34. Fayyad U. M., and Irani K. B., "*Mul -interval Discre sa on of Con nuous-valued A ributes for Classi ca on learning",* Proc. of the XIIIth Int. Joint Conf. on AI, Morgan Kaufmann, 1993.

35. Quinlan J. R., *C 4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.

36. Rich E., and Knight K., *Ar cial Intelligence*, McGraw-Hill, 1991.

37. Hung M. S., Shankar M., & Hu M. Y., "Estimating Breast Cancer Risks Using Neural Networks*", J. Opera onal Research Society,* Vol. 52, pp. 1-10, 2001.

38. Hornik K., Stinchcombe M., & White H., "Multilayer Feedforward Networks are Universal Approximator", *Neural Network*, Vol. 2, pp. 359-366, 1991.

39. Goa D., "On Structures of Supervised Linear basis Function Feedforward Three-layered Neural Networks", *Chin. J. Comput.*, Vol. 21(1), pp. 80-86, 1998.

40. LiMin F., et al., "Incremental Backpropagation Learning Networks", IEEE Transactions of Neural Networks, Vol. 7(3), pp. 757 – 761, 1996.

41. Hall E., Frank G., Holmes B., Pfahringer P., Reutemann I., & Witten H., "The WEKA Data Mining Software: An Update", *SIGKDD Explora ons,* Vol. 11(1), 2009.

42. http://logic.mimuw.edu.pl/~rses

43. Alyuda NeuroIntelligence 2.2, http://www.alyuda.com

44. Bradley, A. P., "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms", *Pa ern Recogni on*, Vol. 30 (7), pp. 1145-1159, 1997.

45. Samanta R. K., Mitra M., "A Neural Network Based Intelligent System for Breast Cancer Diagnosis", Proc. Int. Conference on Intelligent Infrastructure, CSI-2012, Kolkata, Published by Tata McGraw Hill Pvt. Ltd., pp. 20 – 25, 2012.