



Evaluating Performance of Some Classification Algorithms in Neural Networks

N.Srinivas¹, A.V.Dattatreya Rao² and K. Karteeka Pavan³

¹Andhra Loyola College, Vijayawada, Andhra Pradesh, India.

²Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

³RVR & JC College of Engineering, Guntur, Andhra Pradesh, India.

Abstract: Classification is a form of data analysis that extracts models describing important data classes, such models also called classifiers. In this paper an attempt is made for evaluating the performance of three neural network classifiers viz., Radial Basis Function, Multi Layer Perceptron and Probabilistic Neural Network. These classifiers are evaluated based on seven class wise (labels) and overall accuracy measures viz., Accuracy, Precision, Sensitivity, Specificity, Negative Predicted Value, F-measure and Mathews Correlation Coefficient. The results and conclusions obtained empirically are given in section 6.

I. Introduction

Classification is a form of data analysis that extracts models describing important data classes, such models also called classifiers. Big data allows us to unveil hidden patterns associated with a small subpopulation and weak commonalities across the whole population forcing us to concentrate on classification of data. In this paper prudent effort is made to evaluate performance of neural network classifiers. This paper is divided into six sections. Section 2 is devoted to describe the datasets used in the present study. Section 3 concisely describes the neural network classifiers. Accuracy measures are presented in Section 4. Methodology of the study is explained in Section 5. Observations followed by conclusions are stated in Section 6.

II. Description of Data Sets

Iris Data [Fisher R.A. (1936)] describes three varieties of Iris flower (Setosa, Versicolor, and Virginica) each having four features (Sepal length, Sepal width, Petal length and Petal width) for the purpose of Pattern recognition.

Glass Data [German B (1987)] has six classes (Building window float, Building window non-float, Vehicle windows float, Containers, Table ware and Headlamps) each having nine attributes (Refractive index, Sodium, Magnesium, Aluminium, Silicon, Potassium, Calcium, Barium and Iron)

Breast Cancer (BC) Data [Wolberg (1991)] The breast cancer data was obtained from the University of Wisconsin Hospital. The data set contains two classes Benign and Malignant having nine features - Clump Thickness, Uniformity of Cell size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nucleoli, Bland Chromatin, Normal Nucleoli and Mitoses.

Synthetic Datasets: These synthetic data sets are the simulated data sets of multivariate normal $N_p(\mu, \Sigma)$ for specified values of μ and Σ . Half moon, Path based and Spiral data sets are specific Bivariate Normal samples. Parameters of **Half moon (HM)** are

$$\mu_1 = \begin{pmatrix} 27.7150 \\ 9.0254 \end{pmatrix}; \mu_2 = \begin{pmatrix} 14.7010 \\ 21.0253 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 69.0757 & 2.5356 \\ 16.9593 & 12.8283 \end{pmatrix}; \Sigma_2 = \begin{pmatrix} 51.7205 & -1.9333 \\ & 12.8283 \end{pmatrix}$$

those of **Path based (PB)** are

$$\mu_1 = \begin{pmatrix} 19.315 \\ 19.0632 \end{pmatrix}; \mu_2 = \begin{pmatrix} 12.5119 \\ 15.9995 \end{pmatrix}; \mu_3 = \begin{pmatrix} 24.8941 \\ 16.2113 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 110.5167 & -3.1809 \\ 76.1151 & 6.5701 \end{pmatrix}; \Sigma_2 = \begin{pmatrix} 5.1379 & -0.1683 \\ & 6.5701 \end{pmatrix}; \Sigma_3 = \begin{pmatrix} 5.5303 & -0.2709 \\ & 7.7563 \end{pmatrix}$$

and the parameters of **Spiral (SP)** dataset are

$$\mu_1 = \begin{pmatrix} 20.4252 \\ 16.3257 \end{pmatrix}; \mu_2 = \begin{pmatrix} 17.3614 \\ 17.9324 \end{pmatrix}; \mu_3 = \begin{pmatrix} 17.5231 \\ 14.7901 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 50.2159 & 9.8161 \\ 49.0839 & 37.3353 \end{pmatrix}; \Sigma_2 = \begin{pmatrix} 60.8102 & -2.3806 \\ & 37.3353 \end{pmatrix}; \Sigma_3 = \begin{pmatrix} 43.9809 & -6.4972 \\ & 50.9929 \end{pmatrix}$$

Note: Datasets are downloaded from the UCI Machine Learning Repository.

III. Classification Techniques Implemented

In this paper three neural network classifiers are used to classify the data sets mentioned in the above sections. They are Radial Basis Function (RBF), Multi Layer Perceptron (MLP) and Probabilistic Neural Network (PNN). The brief presentation of three algorithms are given below

A. Radial Basis Function (RBF)

Radial Basis Function is basically an interpolating function which passes exactly through every data point. The exact interpolating function for noisy data is typically a highly oscillatory function. For noisy data the best generalization of interpolating function is the one which is much smoother and averages over the noise on the data. The limitation of exact interpolation is that the number of basis function is equal to the number of parameters in data set and for large data sets these functions will become very costly to evaluate and implement. By introducing number of modifications to the exact interpolation procedure, the Radial Basis Function Neural Network model is usually obtained. [Broomhead and Lowe (1988); Moody and Darken (1989)]. The modifications are as follows

1. The number M of basic functions need not equal the number H of data points and in typically much less than N .
2. The centres of the basic functions are no longer constrained to be given by input data vectors. Instead, the determination of suitable centres becomes part of training process.
3. Instead of having a common width parameter σ , each basis function is given its own width σ_j whose value is also determined during training.
4. Bias parameters are included in the linear sum. They compensate the difference the average value over the data set of the basis function activations and the corresponding average value of the targets.

When these changes are made we get the following form for RBF neural network mapping

$$y_k(X) = \sum_{j=1}^M \omega_{kj} \phi_j(X) + \omega_{k0} \quad (1)$$

If desired the basis ω_{k0} can be absorbed into the summation by including as an extra basis function ϕ_0 whose activation is set 1. In case of Gaussian basis function $\phi_j(X)$ can be taken as

$$\exp\left(-\|X - \mu_j\|^2 / 2\sigma_j^2\right) \quad (2)$$

where X is a d -dimensional input vector with elements x_i and μ_j is the vector determining the center of the basis function ϕ_j and has elements μ_{ji} . It may be noted that Gaussian basis functions in (2) are not normalized, since any overall factors can be absorbed into the weights ω_{kj} without loss of generality, where

$$\omega = [\omega_{kj}] = (\phi' \phi)^{-1} \phi' X \quad (3)$$

The optimal performance of the RBF is based on the spread parameter, and it is given as search code in the Matlab. A graph of Spread with correct classification and is shown in Appendix for various data sets.

B. Multi Layer Perceptron (MLP)

Multi Layer Perceptron used here is feedforward neural network model that maps set of input observations into a set of output observations. The MLP was trained by training set by the backpropagation algorithm. MLP consists of multiple layers each layer fully connected to next one. In MLP Backpropagation algorithm is adopted for training the data. The algorithms is as follows

$$\text{In a } (k>2) \text{ multi class classification there are } k \text{ outputs} \quad o_i^t = \sum_{h=1}^H v_{ih} z_h^t + v_{i0} \quad (4)$$

and we use softmax to indicate the dependency between classes, namely, they are mutually exclusive and exhaustive.

$$y_i^t = \exp o_i^t / \sum_k \exp o_k^t \quad (5)$$

$$\text{Where } y_i \text{ approximates } P(C_i | x^t). \text{ The error function is } E(W.V|X) = -\sum_t \sum_i r_i^t \log y_i^t \quad (6)$$

$$\text{And we get the updated equations using gradient descent } \Delta v_{ih} = \eta \sum_t (r_i^t - y_i^t) z_h^t \quad (7)$$

$$\Delta w_{hj} = \eta \sum_t [\sum_i (r_i^t - y_i^t) v_{ih}] z_h^t (1 - z_h^t) x_j^t \quad (8)$$

[Richard and Lippmann (1991)] have shown that given a network of enough complexity and sufficient training data, a suitably trained multilayer perceptron estimates posterior probabilities.

C. Probabilistic Neural Network (PNN)

The Probabilistic Neural Network (PNN) has two hidden layers and an output layer in addition to input layer. Hence this may be treated as multilayer feedforward network. The activation function of a neuron in PNN is statistically derived from the estimation of probability density function based on training patterns. Suppose that members of a set of exemplar patterns belong to k -class $S_1, S_2, \dots, S_k$. Assume that the observable patterns in $x \in \mathbb{R}^m$, whose posteriori probability

$$p_r(S_i | \mathbf{x}) = \frac{p(\mathbf{x}|S_i)p_r(S_i)}{p(\mathbf{x})} \quad (9)$$

where $p_i(\mathbf{x}|S_i)$, $i=1,2, \dots, k$ is the priori PDF of the pattern in the classes to be separated. $p_r(S_i)$, $i=1, 2, \dots, k$ are a priori probabilities of the classes (may be assumed to be equally likely); and $p(\mathbf{x})$ is assumed to be constant. The decision rule is to select class S_i for which $p_i(S_i|\mathbf{x})$ is maximum. This will happen if, for all $j \neq i$

$$p(\mathbf{x}|S_i)p_r(S_i) > p(\mathbf{x}|S_j)p_r(S_j) - (10)$$

Assuming that the priori probabilities are known and it is Gaussian for each class then the estimator for a priori

$$\text{PDF is } \hat{p}(\mathbf{x}|S_i) = \frac{1}{(2\pi)^{m/2} \sigma_i^m |S_i|} \sum_{j=1}^{n_i} \exp \left[\frac{-(\mathbf{x} - \mathbf{x}_j^{(i)})^T (\mathbf{x} - \mathbf{x}_j^{(i)})}{2\sigma_i^2} \right] \quad (11)$$

where $\mathbf{x}_j^{(i)}$ is the j^{th} exemplar pattern or training pattern from class S_i , $|S_i| = n_i$ denotes the cardinality of the set of patterns in class and σ_i is a smoothing parameter. The layers constituting the topology in the PNN are organized as follows. The input layer has m units, to which the m -dimensional input vector $\mathbf{x} \in \mathbb{R}^m$ is applied. The first layer has one pattern unit for each pattern exemplar. Therefore, each such pattern unit may be associated with a generic term depicted in the summation of equation (11) for the i^{th} class. The second hidden layer contains one summation unit for each class, each summation unit is used to realize a sum of the type equation (11) from the outputs in the previous layer. The output layer is the decision layer used for implementing the decision rule by selecting the maximum a posteriori probability, $p_r(S_i|\mathbf{x})$, from the outputs of the preceding summation layer for each i . The decision layer can be implemented by a winner-take-all competitive network. The network is constructed by setting weight vector to one of the pattern units equal to each distinct pattern vector in the training set from a certain class and then connecting the outputs of the pattern units to the appropriate summation unit or that class. Note that each new exemplar causes a new unit in the first hidden layer to be assigned to it, with the pattern vector equalling the input weight vector. The computation proceeds from the data applied to input layer. The negative of the argument of the exponential in the generic term for the summation of equation (11) as

$$\frac{\|\mathbf{x} - \mathbf{x}_j^{(i)}\|^2}{2\sigma_i^2} = \frac{1}{\sigma_i^2} \left[\mathbf{x}^T \mathbf{x}_j^{(i)} - \frac{1}{2} (\|\mathbf{x}\|^2 + \|\mathbf{x}_j^{(i)}\|^2) \right] \quad (12)$$

where the norm of the real-valued vector $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$ — (13)

IV. Deciding the Best Classifier

In order to decide the best classifier the below mentioned accuracy measures are used (i) Accuracy (ACC), (ii) Precision (PRE), (iii) Sensitivity (SEN), (iv) Specificity (SPE), (v) Negative Predicted Value (NPV), (vi) F-Measure (FM), (vii) Mathew's Correlation Coefficient (MCC) and (viii) Overall Accuracy (OVA). In order to compare classifiers, model accuracy is not only sufficient but also the accuracy of the class labels is necessary. So (i) to (vii) are accuracy measures calculated for each class label which varies from 0 to 1. The algorithm will be the best for which (i) to (vii) nearly equal to 1 and (viii) measures overall accuracy of the model. Along with the above accuracy measures Confusion Matrix was also used in discriminating the model.

V. Methodology for the Study of Relative Performance

Classification is also be defined as follows: Given a new multivariate observation, it is to determine to which group/class that it belongs to. In practice, as in this case, data sets are made available without arranging them into groups/classes. Hence clustering is a preprocessing step to classification. As we all know there are several classes of algorithms are existing posing an issue which of the resultant clusters are to be chosen for the purpose of classification. This issue was resolved partially by the authors in their earlier paper "*Relative Performance of K-Means, Single Linkage and Affinity Propagation in Cluster Analysis*" accepted in the next issue in TIJ's Research Journal of Science & IT Management (RJSITM) suggested the procedure of choosing the best clusters. Choosing those best clusters as the starting point, here the chosen classifiers are evaluated based on the accuracy measures described in the earlier section. In this section we briefly describe the methodology for deciding the best classifier. For this, the data sets are divided as Training Data and Test Data. Training Data is used to build / Train (estimating the parameters) a classifier. If the accuracy of the classifier is done with the same trained data, there is every possibility of bias entering in the decision making process. Therefore the test data is an independent sub data set used to know the accuracy of the classifier, which reflects how well the classifier correctly classifies new data points. The training and test data are divided in the ratio of 3:1 respectively. In RBF the optimal performance is based on spread parameter, the plot of spread values for training data is given in Fig. 1 through 6 and hence optimal spread value is selected graphically and such data will be trained based on that optimal value. Similarly for PNN the spread plot is given from Fig 7 through 12 to decide the optimal spread. For MLP the optimal performance depends on the choice of the parameters and its plot for various datasets are shown from Fig.13 through 18 to select optimal values of the parameters. The results are presented in tables T1 through T6 in Appendix. Confusion matrix for each algorithm and for various data sets are also presented in Appendix from tables T13 through T30. For a wide analysis experiments are conducted by choosing different sizes of datasets varying from 50% to 90% and results are reported in terms of overall accuracy in T7 through T12 in Appendix.

VI. Observations and Conclusions

The observations from various tables and consolidated suggestive conclusions are presented in this section.

A. Observations

- From Tables T1 through T6 the observations are as follows
- ✓ In case of Iris data 3 classifiers RBF, MLP and PNN performs equally better in providing classification with 100% accuracy.
- ✓ PNN recorded highest accuracy rate when compared to other two with 98% for Glass data
- ✓ For Breast Cancer data PNN classifier stood top with 98.85 % of accuracy when compared to other RBF and MLP
- ✓ For Half moon and Pathbased data three classification algorithms shows 100% accuracy.
- ✓ RBF and PNN classify the data with 100% accuracy, where MLP with 79.22% of accuracy only.
- From Tables T7 through T12 the observations are as follows
- ✓ In the case of Iris PNN is the only algorithm outperforms clean classification with zero error rate in all different sizes of train datasets varying from 50% to 90%.
- ✓ PNN recorded maximum accuracy for Glass data for different sizes of train datasets from 50% to 90%.
- ✓ For Breast Cancer on average RBF and PNN performs better.
- ✓ For Half moon data PNN scored highest percentage of accuracy when compared to others for different sizes of train data.
- ✓ PNN recorded highest accuracy rate for all different sizes of train data in the case of Path based and Spiral data.

B. Suggestive Conclusions

As the statistical consultants, ultimately we need to recommend the best classifier for a given data set with respect to accuracy for the use of practitioners as a thumb rule.

Notations used in various tables to present final conclusions are explained below

- R, M, and P respectively indicate the classifiers Radial Basis Function, MultiLayer Perceptron and Probabilistic Neural Network.
- BC, HM, PB and SP denote respectively the names of the data sets Breast Cancer, Half moon, Path based and Spiral.
- WAC denotes the Weighted Accuracy.
- ✓ The performance of order of choice of the algorithm for each dataset vs accuracy measures are given in Table CT1

Table CT1 Preference of order of choice of algorithm for each data type vs. accuracy.

DATASET	IRIS	GLASS	BC	HM	PB	SPIRAL
ACC	RMP	PMR	PMR	RMP	RMP	RPM

Interpretation: For example for Iris data performance of algorithms are Radial Basis Function, MultiLayer Perceptron and Probabilistic Neural Network with respect to maximum value of Accuracy and other entries in the table can be interpreted accordingly.

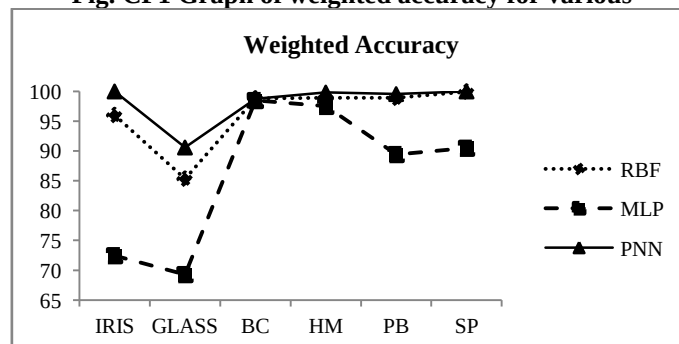
- ✓ The Weighted Accuracy of each data set with respect to the algorithm is presented in the following table CT2. Here weights are in the ratio 0.1, 0.2, 0.3, 0.4, 0.5 (The ratios are taken as the ratio of size of trained data set to that of test data set). Given a dataset and an algorithm, the highest weighted accuracy is indicated in bold figures in CT2 and is that algorithm is to be taken as the best.

Table CT2 Weighted Accuracy of Datasets vs. Algorithms

WAC	RBF	MLP	PNN
IRIS	95.99667	72.46	100
GLASS	85.31867	69.29067	90.60067
BC	98.78133	98.46667	98.74867
HM	98.91333	97.49267	99.82
PB	98.88667	89.41667	99.55333
SP	99.91333	90.47467	100
Average	96.30167	86.26689	98.12044

- ✓ The visual representation of table CT2 is presented in Fig CF1

Fig. CF1 Graph of weighted accuracy for various



✓ The performance of order of choice of the algorithm for each dataset vs weighted accuracy measures are given in Table CT2

Table CT2 Preference of order of choice of algorithm for each data type vs. weighted accuracy.

DATASET	IRIS	GLASS	BC	HM	PB	SPIRAL
WAC	PRM	PRM	RPM	RPM	PRM	PRM

Interpretation: For example for Iris data performance of algorithms are Probabilistic Neural Network, Radial Basis Function and MultiLayer Perceptron with respect to maximum value of Weighted Accuracy and other entries in the table can be interpreted accordingly.

✓ PNN classifier is the best classifier for all data sets with an average of **98.12044** and MLP classifier with least performance of **86.26689**

References

- Anil K. Jain (1989) Fundamentals of Digital Image Processing, PHI.
- Anderson, T.W. (1962). An Introduction to Multivariate Statistical Analysis. Wiley, New York.
- Andrews, H.C. (1972). Introduction to Mathematical Techniques in Pattern Recognition, John Wiley & Sons, New York.
- Alex Berson, Stephen Smith, Kurt Thearling (2000) Building Data Mining Applications for CRM, Tata McGraw Hill.
- Arun. K Pujari (2001) Data Mining Techniques, University Press (India) Ltd.
- Broomhead D. S, Lowe David (1988) Multivariate functional interpolation and adaptive networks, Complex Systems 2: 321-355.
- Bandyopadhyay. S et al (1995). Pattern classification with genetic algorithms. Patterns Recognition Letters, Vol. 16(8) pp. 801-808.
- Bart Kosko (1997) Neural Networks and Fuzzy Systems: A Dynamical Systems approach to machine Intelligence, PHI.
- Bose. N. K, Liang. P (1998) Neural Network Fundamentals with Graphs, Algorithms and Applications, Tata McGraw Hill.
- Bhuvan K. C (2005) Multivariate Analysis and its Applications, New Central Book Agency (P) Ltd.
- Christopher M. Bishop (1995) Neural Networks for Pattern Recognition, Oxford University Press.
- David Hand, Heikki Mannila, Padhraic Smyth (2004) Principles of Data Mining, PHI.
- Ethem Alpaydin (2010) Introduction to machine Learning, MIT Press.
- Fisher R. A (1936) The use of Multiple measurements in Taxonomic Problems, Annals of Eugencies
- Fuknaga Keinosuke (1972) Introduction to Statistical Pattern Recognition, Academic Press.
- German B (1987) Glass identification Database, UCI Machine Learning Repository.
- Han Jiawei, Kamber Micheline (2001) Data Mining: Concepts and Techniques. Morgan Kaufmann
- James A. Anderson (1999) An Introduction to Neural Networks, PHI.
- Johnson Richard A., and Wichern Dean. W(2001). Applied Multivariate Statistical Analysis, Prentice Hall International Inc., Englewood Cliffs, New Jersey.
- Jang-J-S.R, Sun.T-C, Mizutani. E (2003) Neuro-Fuzzy and Soft Computing: A computational approach to learning and Machine Intelligence.
- James A Freeman, David M. Skapura (2013) Neural Networks Algorithms, Applications and Programming Techniques.
- Mc Lachlan, G. J. (1992). Discriminant Analysis and Statistical Pattern. John Wiley, New York
- Mohamad H. Hassoun (1999) Fundamentals of Artificial Neural Networks, PHI.Margaret H. Dunham (2003) Data Mining, Introductory and Advanced Topics, Pearson Edition.
- Patrick, E. A. (1972) Fundamentals of Pattern Recognition, Prentice Hall, Englewood Cliffs, New Jersey.
- Prabhu C. S. R (2002) Data Warehousing, Concepts, Techniques, products and Applications, PHI.
- Richard, M. D and Lippmann, R. P (1991) Neural Network Classifiers Estimate Bayesian a posteriori Probabilities, Neural Computation 3, pp. 461-483
- Ripley B. D (1996) Pattern Recognition and Neural Networks, University of Oxford, Cambridge University Press.
- Robert S (2005) Pattern Recognition, Statistical Structural and Neural Approaches.
- Rajasekaran. S, Vijayalakshmi Pai. G. V (2006) Neural Networks, Fuzzy Logic and Genetic Algorithms Synthesis and Applications, PHI.
- William H. Wolberg and O.L. Mangasarian: Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990, pp 9193-9196
- Yegnanarayana. B (1999) Artificial Neural Networks, PHI.

APPENDIX

1. Tabular presentations of Accuracy measures of various datasets with respect to three classifiers.

Table T1: Accuracy measures for Iris data by applying three classifiers.

IRIS	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	
MLP	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	
PNN	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	

Table T2: Accuracy measures for Glass data by applying three classifiers.

GLASS	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	0.98	0.8	1	0.9783	1	0.8889	0.8847	92
	2	0.92	0.8	0.5714	0.9767	0.9333	0.6667	0.634	
	3	0.96	0.9394	1	0.8947	1	0.9688	0.9168	
	4	1	1	1	1	1	1	1	
	5	1	1	1	1	1	1	1	
	6	0.98	NaN	0	1	0.98	NaN	NaN	
MLP	1	0.96	1	0.5	1	0.9583	0.6667	0.6922	94
	2	0.94	0.75	0.8571	0.9535	0.9762	0.8	0.7673	
	3	0.98	0.9688	1	0.9474	1	0.9841	0.958	
	4	1	1	1	1	1	1	1	
	5	1	1	1	1	1	1	1	
	6	1	1	1	1	1	1	1	
PNN	1	1	1	1	1	1	1	1	98
	2	0.98	1	0.8571	1	0.9773	0.9231	0.9152	
	3	0.98	0.9688	1	0.9474	1	0.9841	0.958	
	4	1	1	1	1	1	1	1	
	5	1	1	1	1	1	1	1	
	6	1	1	1	1	1	1	1	

Table T3: Accuracy measures for Breast Cancer data by applying three classifiers.

BC	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	0.9598	0.913	0.9844	0.9455	0.9905	0.9474	0.9166	95.98
	2	0.9598	0.9905	0.9455	0.9844	0.913	0.9674	0.9166	
MLP	1	0.9713	0.9836	0.9375	0.9909	0.9646	0.96	0.9383	97.13
	2	0.9713	0.9646	0.9909	0.9375	0.9836	0.9776	0.9383	
PNN	1	0.9885	0.9844	0.9844	0.9909	0.9909	0.9844	0.9753	98.85
	2	0.9885	0.9909	0.9909	0.9844	0.9844	0.9909	0.9753	

Table T4: Accuracy measures for Half moon data by applying three classifiers.

HM	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
MLP	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
PNN	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	

Table T5: Accuracy measures for Pathbased data by applying three classifiers

PB	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	
MLP	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	
PNN	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	

Table T6: Accuracy measures for Spiral data by applying three classifiers

SP	CLASS	ACC	PRE	SEN	SPE	NPV	FM	MCC	OVA
RBF	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	
MLP	1	0.9221	0.8333	0.9615	0.902	0.9787	0.8929	0.8374	79.22
	2	0.8312	0.75	0.72	0.8846	0.8679	0.7347	0.6112	
	3	0.8312	0.7826	0.6923	0.902	0.8519	0.7347	0.614	
PNN	1	1	1	1	1	1	1	1	100
	2	1	1	1	1	1	1	1	
	3	1	1	1	1	1	1	1	

2. Tabular presentation six datasets with different Train data sizes by applying three classifiers

Table T7 Iris Accuracy for different Train data size

IRIS	0.1	0.2	0.3	0.4	0.5
RBF	100	93.3	100	95	94.67
MLP	66.7	93.3	66.7	75	66.7
PNN	100	100	100	100	100

Table T8 Glass Accuracy for different Train data size

GLASS	0.1	0.2	0.3	0.4	0.5
RBF	-	92.7	88.7	92.77	91.44
MLP	-	73.17	72.58	80.72	70.48
PNN	-	95.12	100	93.38	99.05

Table T9 BC Accuracy for different Train data size

BC	0.1	0.2	0.3	0.4	0.5
RBF	95.7	98.6	99.04	99.3	98.9
MLP	100	97.84	99.52	97.84	98.28
PNN	98.5	99.28	99.04	98.2	98.85

Table T10 HM Accuracy for different Train data size

HM	0.1	0.2	0.3	0.4	0.5
RBF	100	98.6	100	100	97.3
MLP	100	100	100	98.66	93.55
PNN	100	98.65	100	100	100

Table T11 PB Accuracy for different Train data size

PB	0.1	0.2	0.3	0.4	0.5
RBF	100	98.3	100	98.3	98.7
MLP	89.66	89.83	89.77	89.08	89.26
PNN	100	100	100	100	98.66

Table T12 SP Accuracy for different Train data size

SPIRAL	0.1	0.2	0.3	0.4	0.5
RBF	98.7	100	100	100	100
MLP	80	83.87	91.3	90.32	94.84
PNN	100	100	100	100	100

3. Visual representation of Spread Plots of various datasets by applying RBF classifier

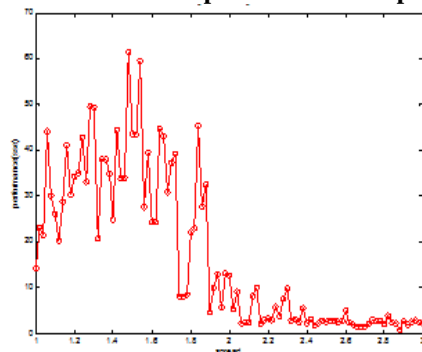


Fig.1 Iris Spread plot

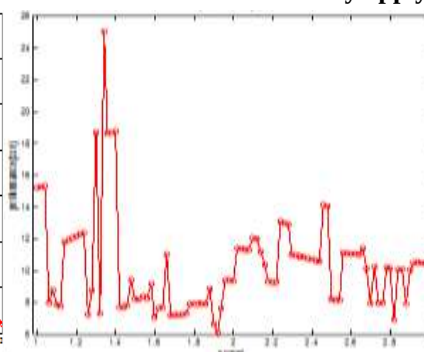


Fig.2 Glass Spread plot

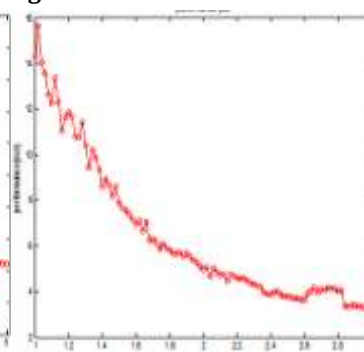


Fig.3 Breast Cancer Spread plot

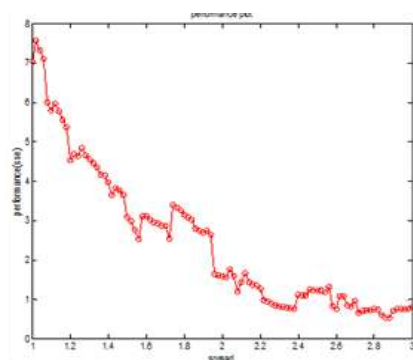


Fig.4 Half moon Spread plot

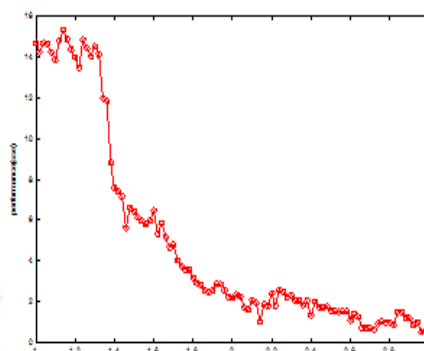


Fig.5 Pathbased Spread plot

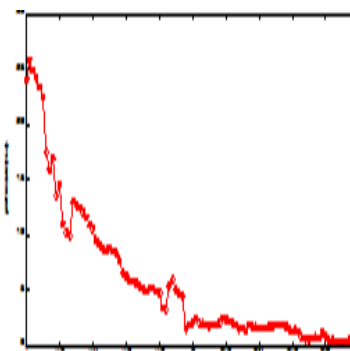


Fig.6 Spiral Spread plot

4. Visual representation of Spread Plots of various datasets by applying PNN classifier

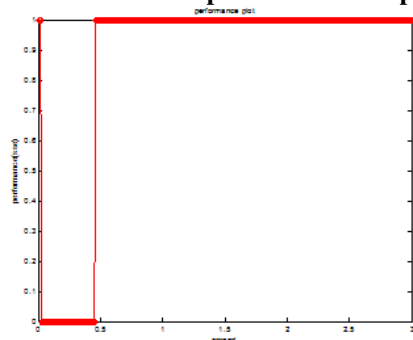


Fig.7 Iris Spread plot

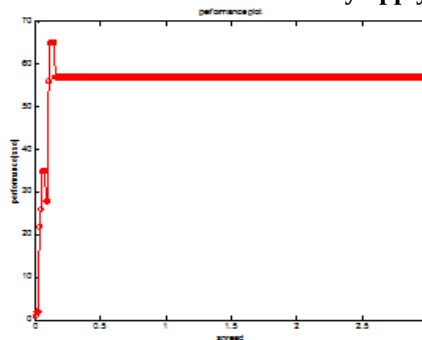


Fig.8 Glass Spread plot

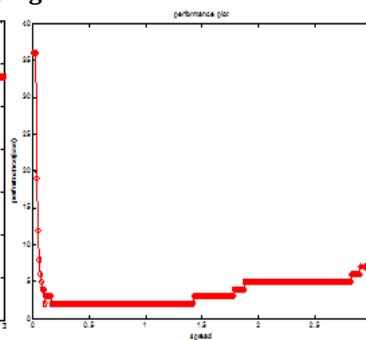


Fig.9 Breast Cancer Spread plot

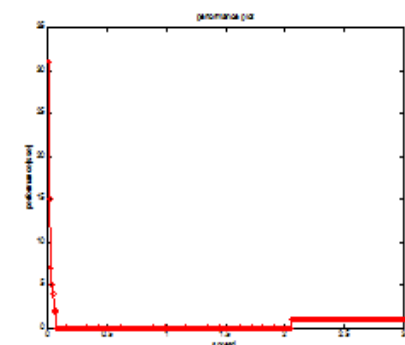


Fig.10 Half moon Spread plot

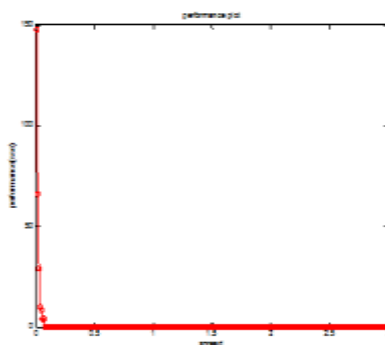


Fig.11 Pathbased Spread plot

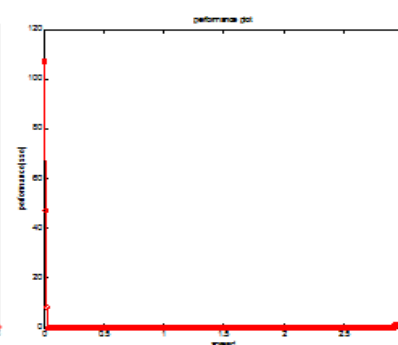


Fig.12 Spiral Spread plot

5. Visual representation of parameters in MLP of various datasets.

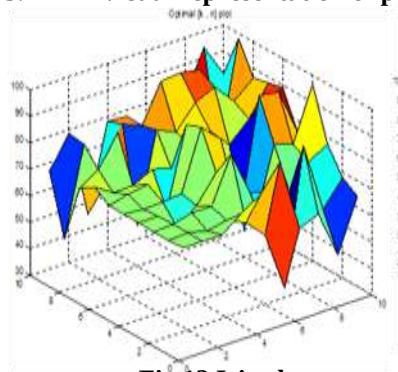


Fig.13 Iris plot

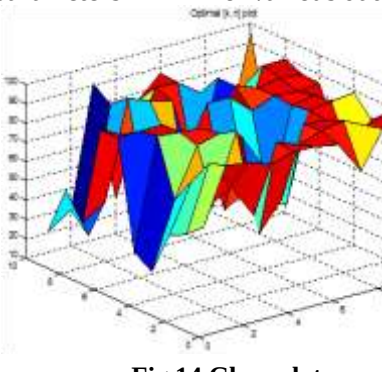


Fig.14 Glass plot

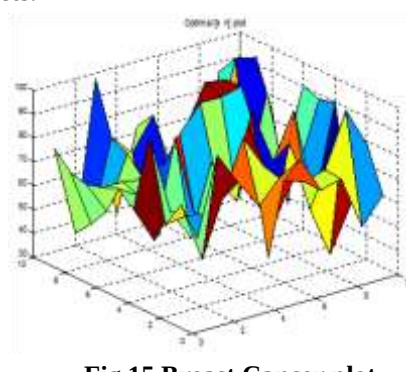


Fig.15 Breast Cancer plot

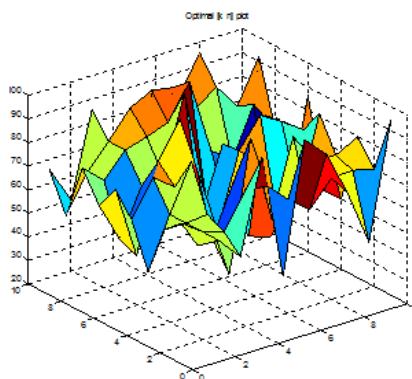


Fig.16 Half moon plot

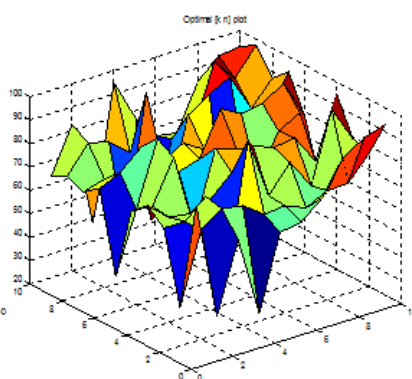


Fig.17 Pathbased plot

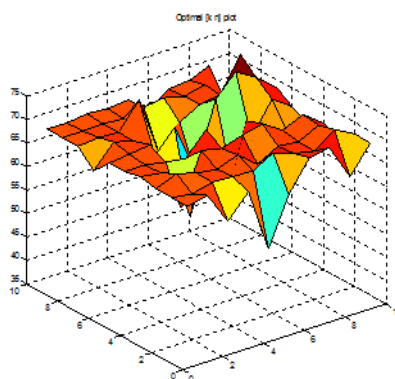


Fig.18 Spiral plot

6. Confusion Matrix (CM) each data set wise when specific classifier is performed.

Note: In the below tables values in the parenthesis indicates the percentage of accuracy of target class and output class.

Table T13 CM of Iris when RBF is performed

Output Class	1	2	3	Acc
	12 (33.33%)	0 (0.0%)	0 (0.0%)	100 0.0%
	0 (0.0%)	12 (33.33%)	0 (0.0%)	100 0.0%
	0 (0.0%)	0 (0.0%)	12 (33.33%)	100 0.0%
Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
Target Class				Acc
				1
				2
				3

Table T14 CM of Iris when MLP is performed

Output Class	1	2	3	Acc
	12 (33.33%)	0 (0.0%)	0 (0.0%)	100 0.0%
	0 (0.0%)	12 (33.33%)	0 (0.0%)	100 0.0%
	0 (0.0%)	0 (0.0%)	12 (33.33%)	100 0.0%
Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
Target Class				Acc
				1
				2
				3

Table T15 CM of Iris when PNN is performed

Output Class	1	2	3	Acc
	12 (33.33%)	0 (0.0%)	0 (0.0%)	100 0.0%
	0 (0.0%)	12 (33.33%)	0 (0.0%)	100 0.0%
	0 (0.0%)	0 (0.0%)	12 (33.33%)	100 0.0%
Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
Target Class				Acc
				1
				2
				3

Table T16 CM of Glass when RBF is performed

Output Class	1	2	3	4	5	6	Acc
	4 (8%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	80% 20%
	0 (0%)	4 (8%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	80% 20%
	0 (0%)	2 (4%)	31 (62%)	0 (0%)	0 (0%)	0 (0%)	93.9% 6.1%
	0 (0%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	100% 0%
	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (12%)	0 (0%)	100% 0%
	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 0%
Acc	100% 0%	57.1% 42.9%	100% 0%	100% 0%	100% 0%	0% 100%	92% 8%
Target Class							Acc
							1
							2
							3
							4
							5
							6

Table T17 CM of Glass when MLP is performed

Output Class	1	2 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	100% 0%
	2	2 (4%)	6 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	75% 25%
	3	0 (0%)	1 (2%)	31 (62%)	0 (0%)	0 (0%)	0 (0%)	96.9% 3.1%
	4	0 (0%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	100% 0%
	5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (12%)	0 (0%)	100% 0%
	6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	100% 0%
	Acc	50% 50%	85.7% 14.3%	100% 0%	100% 0%	100% 0%	100% 0%	94% 6%
	1	2	3	4	5	6	Acc	
	Target Class							

Table T18 CM of Glass when RBF is performed

Table 11: CM of Glass when RBF is performed								
Output Class	1	4 (8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	100% 0%
	2	0 (0%)	6 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	100% 0%
	3	0 (0%)	1 (2%)	31 (62%)	0 (0%)	0 (0%)	0 (0%)	93.9% 6.1%
	4	0 (0%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	100% 0%
	5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (12%)	0 (0%)	100% 0%
	6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	100% 0%
	Acc	100% 0%	85.7% 14.3%	100% 0%	100% 0%	100% 0%	0% 100%	98% 2%
	1	2	3	4	5	6	Acc	
	Target Class							

Table T22 CM of HM when RBF

Output Class	1	61 (65.6%)	0 (0%)	100% 0%
	2	0 (0%)	32 (34.4%)	100% 0%
Acc		100% 0%	100% 0%	100% 0%
		1	2	Acc
		Target Class		

Table T24 CM of HM when PNN is performed

Output Class	1	61 (65.6%)	0 (0%)	100% 0%
	2	0 (0%)	32 (34.4%)	100% 0%
Acc		100% 0%	100% 0%	100% 0%
		1	2	Acc
		Target Class		

Table T19 CM of BC when RBF

Output Class	1	63 (36.2%)	6 (3.4%)	91.3 8.7%
	2	1 (0.6%)	104 (59.8%)	99% 1%
Acc		98.4% 1.6%	94.5% 5.5%	96% 4%
		1	2	Acc
		Target Class		

Table T20 CM of BC when MLP

Output Class	1	60 (34.5%)	1 (0.6%)	98.4 1.6%
	2	4 (2.3%)	109 (62.6%)	96.5% 3.5%
Acc		93.8% 6.3%	99.1% 0.9%	97.1% 2.9%
		1	2	Acc
		Target Class		

Table T21 CM of BC when PNN

Output Class	1	63 (36.2%)	1 (0.6%)	98.4 1.6%
	2	1 (0.6%)	109 (62.6%)	99.1% 0.9%
Acc		98.4% 1.6%	99.1% 0.9%	98.9% 1.1%
		1	2	Acc
		Target Class		

Table T23 CM of HM when MLP

Output Class	1	61 (65.6%)	0 (0%)	100% 0%
	2	0 (0%)	32 (34.4%)	100% 0%
Acc		100% 0%	100% 0%	100% 0%
		1	2	Acc
		Target Class		

Table T25 CM of PB when RBF is performed

Output Class	1	8 (10.8%)	0 (0.0%)	0 (0.0%)	100 0.0%
	2	0 (0.0%)	33 (44.6%)	0 (0.0%)	100 0.0%
	3	0 (0.0%)	0 (0.0%)	33 (44.6%)	100 0.0%
	Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
		1	2	3	Acc
		Target Class			

Table T26 CM of PB when MLP is Performed

Output Class	1	8 (10.8%)	0 (0.0%)	0 (0.0%)	100 0.0%
	2	0 (0.0%)	33 (44.6%)	0 (0.0%)	100 0.0%
	3	0 (0.0%)	0 (0.0%)	33 (44.6%)	100 0.0%
	Acc	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
					Acc
Target Class					

Table T27 CM of PB when PNN is performed

Output Class	1	8 (10.8%)	0 (0.0%)	0 (0.0%)	0.0%
	2	0 (0.0%)	33 (44.6%)	0 (0.0%)	100 0.0%
	3	0 (0.0%)	0 (0.0%)	33 (44.6%)	100 0.0%
	Acc	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
Target Class					Acc

Table T28 CM of SP when RBF is performed

Output Class	1	26 (33.8%)	0 (0.0%)	0 (0.0%)	100 0.0%
	2	0 (0.0%)	25 (32.5%)	0 (0.0%)	100 0.0%
	3	0 (0.0%)	0 (0.0%)	26 (33.8%)	100 0.0%
	Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
					Acc
Target Class					

Table T29 CM of SP when MLP is performed

Output Class	1	25 (32.5%)	3 (3.9%)	2 (2.6%)	83.3% 16.7%
	2	0 (0.0%)	18 (23.4%)	6 (7.8%)	75% 25%
	3	1 (1.3%)	4 (5.2%)	18 (23.4%)	78.3% 21.7%
	Acc	96.2% 3.8%	72% 28%	69.2% 30.8%	79.2% 20.8%
	1	2	3	Acc	
Target Class					

Table T30 CM of SP when PNN is performed

Output Class	1	26 (33.8%)	0 (0.0%)	0 (0.0%)	100 0.0%
	2	0 (0.0%)	25 (32.5%)	0 (0.0%)	100 0.0%
	3	0 (0.0%)	0 (0.0%)	26 (33.8%)	100 0.0%
	Acc	100% 0.0%	100% 0.0%	100% 0.0%	100 0.0%
					Acc
Target Class					