An Evolutionary Feature Selection Technique Using Polynomial Neural Network

Amit Saxena[1], Dovendra Patre[2], Abhishek Dubey[3]

[1] **Prof. & Head, Dept. of CSIT, G.G.Vishwavidyalaya, Bilaspur (C.G.),India.**

[2] **Project Fellow, Dept. of CSIT, G.G.Vishwavidyalaya, Bilaspur (C.G.),India.**

[3] **Lecturer, Dept. of IT, Salalah College of Technology, Salalah, Oman.**

Abstract

In this paper we propose a novel approach for feature subset selection by the Polynomial Neural Network (PNN) using Genetic Algorithm (GA). A randomly selected subset of features of a dataset is passed to the PNN as input. The classification accuracy of PNN is taken as the fitness function of GA. In the conventional PNN approaches, published in literature so far, the processing by PNN takes large computation time due to the expansion of the whole network at different levels. In the proposed scheme, less number of features selected stochastically using the GA, prevents PNN to grow at very early stages which reduces the computation cost as well as time. The proposed scheme is simulated on six benchmark databases and classification accuracies obtained using proposed PNN classifiers are compared with those obtained using three other existing approaches. It is observed that the classification accuracies using proposed scheme are quite satisfactory compared to existing three schemes. The strength of proposed scheme is justified in two ways: (i) its high classification accuracy with much less computational cost in the presence of reduced number of features and (ii) much less execution time taken by it as compared to other schemes.

Keywords: *Polynomial Neural Net, Genetic Algorithm, Feature Selection, Pattern Classification.*

1. Introduction

It is known that a certain set of features in a dataset that characterize a data point (class in most cases) for a classification problem may not be equally important; some features can be derogatory and may even have harmful effect on the task. Feature selection techniques aim to discard the bad and irrelevant features from the available set of features. This reduction of features may improve the performance of classification, function approximation, and other pattern recognition systems in terms of speed, accuracy, and simplicity [13]. Another importance of feature selection is in the task of mining large databases which is also known as dimensionality reduction [52]. Dimensionality reduction can be done by selecting a small but important subset of features and generating (extracting) a lower dimensional data preserving the distinguishing characteristics of the original higher dimensional data [42]. Feature Selection leads to savings measurement cost and time because some of the features get discarded. This concept can be utilized in pruning those networks which otherwise would have taken a large time to compute in presence of all features. In practice it is found that additional features actually degrade the performance of a classifier designed using class-conditional density estimates when the training set is small with respect to the dimensionality [46], [47]. This usually happens because classifiers estimate the class-conditional densities from the available training data. Thus, if the dimensionality is increased keeping the size of the training set fixed, the number of unknown parameters automatically increases and the reliability of the estimate decreases. As a consequence, the performance of the classifier, constructed from a fixed number of training instances, may degrade with the increase in dimensionality as was illustrated by Trunk [57]. When feature selection methods use class information, we call it supervised feature selection otherwise it is an unsupervised feature selection [35], [11]. Polynomial Neural Networks (PNN) [24] have emerged recently as an extension of Artificial Neural Networks (ANN) [51], to take care of some of the limitations of the latter. The PNN are also used as classifiers like ANN. The measurement cost of classification accuracy in terms of complexity and time depends on the number of features and the size of PNN, i.e. its architecture. The latter has been addressed in literature [30], [31]. The PNN based classification using feature selection approach has not been applied so far to our perception. In this paper we investigate a scheme to evaluate the performance of PNN classifier using reduced number of features by applying genetic algorithm (GA).

This paper is organized as follows: Section 2 presents review of work done. Section 3 presents basics of PNN. Section 4 outlines GA. Section 5 presents proposed scheme. Simulation study is presented in Section 6. Results are discussed in Section 7 followed by Conclusions in Section 8.

2. Review of the work done

The problem of feature selection has been well addressed in literature [13] and it has been tried out in various paradigms. Previous studies on feature subset selection have focused mainly around statistical approaches like principal component analysis (PCA) [21], linear discriminant analysis (LDA) [17], etc. These methods attempt to reduce the dimensionality of the feature space by creating new features which are combination of the original ones; known as feature extraction techniques. The main drawback of these methods is that the newly created features lose their original identity. Blum and Langley [5] have given an excellent survey on feature selection in machine learning. These approaches are different in evaluation of the feature subsets. There are many feature selection algorithms that use soft computing/computational intelligence tools. GA is used in [8] and [48] to select the relevant feature subsets. Neural Networks are used in [3], [16], [50], [53],[55] and [61] for feature selection. Feature selection has also been attempted using fuzzy and neurofuzzy techniques [12], [49]. There are also specialized methods to deal with feature selection for very large dimensional data sets that are typical in application areas such as bioinformatics [1], [25]. In [26], [27] and [28] MacKay has considered neural network learning in a Bayesian framework. MacKay and Neal proposed a feature selection mechanism in the Bayesian learning framework called automatic relevance detection (ARD) [37]. In the ARD model, each input variable is associated with a hyper parameter that controls the magnitude of the weights of connections out of that input unit. The significance of an input variable is determined according to the posterior distributions of these hyper parameters. In [41], Pal and Chintalpudi developed an integrated feature selection and classification scheme based on the multilayer perceptron (MLP) architecture. The feature selection phase in their method was integrated with the main learning task, and the MLP learned certain feature modulators along with the conventional weights and biases of a neural network. In [9], a neurofuzzy system was developed for simultaneous feature selection and system identification. The methodology developed in [9] was modified for a classifier in [10]. The feature selection methods described in [9], [10], and [41] are termed as online methods. In an online method, the feature selection phase is integrated with the task of learning other parameters of the system. The ARD [37] and its variant [56] also learn hyper parameters associated with the input features. Some evolutionary techniques [59] also learn the importance of input features along with other parameters of the classifier. For the classification problem, LIKNON [4] uses a linear programming formulation to learn feature weights along with other parameters of the separating hyperplane.

Sivakumar [54], report on the plausible solution for ascertaining the composition of gases, during complex boiler flue gas data analysis, by taking a number classification problem as a model. For this purpose an indigenously developed arithmetic residue (AR) scheme has been devised as a feature selection technique. For the purpose of classification of data (number of classes of gases), a probabilistic neural network has been implemented and its classification capability has been analyzed first for the data acquired from ORSAT analyzer and then for the data from KANE® analyzer.

Yumin et.al. [60] recently proposes an Ant Colony Optimization based approach for feature selection using rough sets. Saxena et al. have used GA algorithms with Sammon's function for unsupervised feature selection [52].

It will also be worth addressing PNN and some of its applications in feature selection for classification. Classification is one of the core challenging tasks [30] in data mining [29], [22], pattern recognition [15], web mining [23], bioinformatics [2], and financial forecasting [7], [14]. The goal of classification [34] is to assign a new entity into a class from a pre-specified set of classes. One of the popular and widely used techniques is the method of feed-forward neural network (FNN) [58]. Although such FNN can classify a wide range of dataset properly/accurately yet the classification model cannot be comprehensible due to a large number of synaptic connections including weights and biases. In order to achieve high classification accuracy in FNN framework, one has to provide a well defined structure of FNN such as the number of input nodes, hidden and output neurons, and assumption of a proper set of relevant features. In this regard the trial and error method has been used to arrive at such kind of structures which are computationally very expensive. Similarly there are other methods like rule extraction and decision tree [44], [45], which provides comprehensible rules that are based on the trade-off between the complexity and the classification accuracy of the rules. Misra et al. [30] presents a Particle Swarm Optimization (PSO) approach to reduce optimized architecture of PNN for classification task.

3. Polynomial Neural Networks (PNN)

PNN is a flexible neural architecture whose topology is not predetermined or fixed like a conventional ANN but is grown through learning layer by layer. The design is based on Group Method of Data Handling (GMDH) which was invented by Prof. A. G. Ivankhnenko in the late 1960s [19], [38], [20] and [39]. He developed GMDH as a means for identifying nonlinear relations between input and output variables. As described in [40], the GMDH generates successive layers with complex links that are individual terms of a polynomial equation. The individual terms generated in the layers are partial descriptions (PDs) of data being the quadratic regression polynomials with two inputs. The first layer is created by computing regressions of the input variables and choosing the best ones for survival. The second layer is created by computing regressions of the values in the previous layer along with the input variables and retaining the best candidates. More layers are built until the network stops getting better based on termination criteria. The selection criterion used in this study penalizes the model that become too complex to prevent overtraining.

In a feed-forward neural network (FNN) [43], to achieve high classification accuracy, one has to provide in advance, a well defined structure of FNN, such as, the number of input nodes, hidden and output neurons, and assume a proper set of relevant features. To alleviate this drawback of FNN; PNN can be used for classification purposes. Evolutionary approach based PNN generates populations or layers of neurons/simulated units/partial descriptions (PDs) and then trains and selects those neurons, which provide the best classification. Using this approach during learning, the PNN model generates the new population of neurons and the number of layers and the complexity of the network increases [32], [33] until a predefined criterion is met. Such models can be comprehensively described by a set of short-term polynomials thereby developing a PNN classifier. Coefficients of PNN can be estimated by least square fitting. The network architecture grows depending on the number of input features, PNN model selected, number of layer required, and the number of PD's preserved in each layer. Fig. 1 shows a basic PNN model with all inputs. Fig. 2 describes how a PD is computed at a node of a PNN's layer with reduced features using proposed scheme.

The GMDH belongs to a kind of inductive self-organization data driven approach. It requires small data samples which are able to optimize the structure of the models objectively and this relationship between input–output variables can be well approximated by Volterra functional series, the discrete form of which is Kolmogorov–Gabor polynomial [20], i.e.

$$y = C_0 + \sum_{k_1} C_{k_1} x_{k_1} + \sum_{k_1 k_2} C_{k_1 k_2} x_{k_1} x_{k_2} + \sum_{k_1 k_2 k_3} C_{k_1 k_2 k_3} x_{k_1} x_{k_2} x_{k_3} + \ldots\ldots\ldots \tag{1}$$

Where C_k denotes the coefficients or weights of the Kolmogorov–Gabor polynomial and x vector is the input variable. Further a new GMDH algorithm has also been developed by Ivakhnenko [20], [36] which is a form of Kolmogorov–Gabor polynomial. He proved that a second order polynomial, i.e.

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{2}$$

This takes only two input variables at a time and can reconstruct the complete Kolmogorov–Gabor polynomial through an iterative procedure. The GMDH-type polynomial neural networks are multi-layered models consisting of the neurons/active units/partial descriptions (PDs), whose transfer function is a short-term polynomial described in Eq. (2). At the first layer L = 1, an algorithm, using all possible combinations of two inputs out of m variables, generates the first population of PDs. Hence, the total number of PDs in first layer is n = m (m-1)/2 and the output of each PD in layer L = 1 is computed by applying the Eq. (2). The coefficient vector of the PDs is determined by the least square estimation approach. The further details of the PNN model and least square estimation technique can be found in [30].

4. Genetic Algorithm (GA)

First pioneered by John Holland in the 1960s, GA have been widely studied with interest, experimented and applied in many fields in science and engineering worlds. GA is an evolutionary algorithm, which optimizes a fitness function to find the solution of a problem. Different evolutionary algorithms have been used for feature selection. In a typical GA, each chromosome represents a prospective solution of the problem. The problem is associated with a fitness function – higher fitness refers to a better solution. The set of chromosomes is called a population. The population goes through a repeated set of iterations (or generations) with crossover and mutation operations to find better solutions. At a certain fitness level or after a certain number of generations, the procedure is stopped and the chromosome giving the best solution is preserved as the best solution of the problem. A detailed description of GA can be found in [18].

5. Proposed Scheme

Proposed scheme is depicted in Figure-3. Feature set represents the set of all features in the dataset. A subset of this feature set is selected randomly which becomes a part of population (one chromosome) to be used in GA. The existence of a feature in the subset is represented by a 1 and its absence by a 0 in every chromosome of the population. This subset of features is given as input to PNN. The classification accuracy of PNN is calculated using training and testing patterns from the dataset. This accuracy serves as the fitness value of the GA. Similarly the fitness of other subsets of features in the initial population is also calculated. The simple one point crossover and mutation operations are applied on initial population to produce a modified population. The fitness' of the chromosomes of the modified population are compared with those of the initial population. The better chromosomes (Subsets) are retained in the next population. This completes one generation and the population with chromosomes of higher fitness values becomes the initial population for the next generation. This process continuous for a number of generations and at a satisfactory level the process is stopped. The feature subset with best classification accuracy of PNN is noted for comparison.

6. Simulation Studies

The performance of proposed scheme is evaluated using the benchmark databases. A summary of these databases is given in Table 1 which is also available in the UCI machine repository [6]. Proposed scheme was simulated on a Pentium-III machine. For computing classification accuracy of PNN, cross validation was used. Each dataset was divided into two folds, one for training and other for testing. We have taken 50% patterns in fold 1 and remaining 50% in fold 2. The number of generations and sizes of populations used in the proposed scheme for different datasets are shown in Table 2. Table 3 presents the Times of execution, classification accuracies for different datasets using PNN, RCPNN with Gradient Descent, RCPNN with PSO and Proposed Scheme. Figure 4 shows a co mparative bar chart for proposed scheme versus PNN, RCPNN with Gradient Descent, RCPNN with PSO schemes with respect to classification accuracies. Figure 5 shows a comparative line graph for proposed scheme versus RCPNN with Gradient Descent, RCPNN with PSO schemes with respect to time. As time consumed by P1 is very high compared to P2,P3 and P4, we did not show P1 in the Figure 5 just to show in details the variations among P2,P3 and P4.

7. Results and Discussions

Table 3 presents a comparison between the performances evaluated using proposed scheme (P4) and those evaluated using other three schemes viz. conventional PNN (P1), RCPNN with Gradient Descent (P2) and RCPNN with PSO (P3) published in literature [30], [31]. The classification accuracy and the execution cost have been taken as the performance indexes. In P2 and P3, efforts are made to reduce the architecture of the PNN. However number of features has not been reduced in any of the three schemes. By observing Table 3, it is noted that in proposed scheme, a very less number of features are capable to produce a higher classification accuracy e.g. in iris data set only two features are required to produce 99.11 % accuracy against other schemes which take all (4) features but produce a lesser accuracy (86.22,95.56,98.67). Similar results are observed for Bupa, WBC, and Thyroid datasets. For Wine and Pima datasets, the classification accuracies using proposed scheme are obtained fractionally smaller than that with other schemes but important to note that our investigation takes less number of features to acquire that much accuracy. This small difference in accuracy can be accepted looking to overall figures in comparison to number of features. In addition to classification accuracy; execution time is another parameter which has been compared with other three schemes. In proposed scheme, time of execution for iris dataset is 0.193 sec versus 254, 0.428, 0.830 sec respectively in other three schemes which is a noticeable difference. In each dataset, proposed scheme takes much less time compared to that in published schemes.

8. Conclusions

In this paper, we have proposed a n ovel scheme of feature selection using Polynomial Neural Networks (PNN). A number of recent publications have used PNN for different diversified applications including classification. There have been research publications to reduce the size of a conventional PNN. It is known that PNN grows from the first layer on the basis of number of features (inputs) and partial derivatives produced due to these inputs in the subsequent layers. If number of features is reduced, then the growth of partial derivatives will also stop at immediate next layer and there will be no more partial derivatives onward due to this dead node. To select subsets of features in order to reduce irrelevant / derogatory features, we applied GA. The fitness of GA is measured by computing the classification accuracy obtained by PNN for a selected subset of features. The scheme is tested for six datasets and in each case, proposed scheme outperforms other three known schemes published for the same purpose in terms of time of execution as well as classification accuracy. It is observed that time and classification accuracy, both is greatly improved by applying proposed scheme. It justifies our investigation. A further extension of proposed scheme is to test it on very large datasets like Microarrays, spatial datasets which will be scope for future research in this direction.

9. Acknowledgement

The authors acknowledge University Grants Commission (UGC), New Delhi, India to support this investigation through the Major Research Project.

10. References

[1] Al-Mubaid, H., Ghaffari, N., 2006. Identifying the most significant genes from gene expression profiles for sample classification. In Proc. IEEE Conf. Granural Comput.655–658.

[2] Baldi, P., Brunak, S., 1998. Bioinformatics: The Machine Learning Approach. MIT Press, Cambridge, MA.

[3] Beleue, L.M., Bauer, K.W., 1995. Determining input features for multilayered perceptron, Neurocomputing. 7(2), 111–121.

[4] Bhattacharyya, C., Grate, L.R., Rizki, A., Radisky, D., Molina, F.J., Jordan, M.I., Bissell, M.J., Mian, I.S., 2003. Simultaneous classification and relevant feature identification in high-dimensional spaces: Application to molecular profiling data. Signal Process.83, 729–743.

[5] Blum, L., Langley, P., 1997. Selection of relevant features and examples in machine learning, Artif. Intell. 97(1), 245–271.

[6] Blake, L., Merz, C.J., 2001. "UCI Repository of machine learning databases, "http://www.ics.uci.edu/ ~mlearn/MLRepository.html.junio".

[7] Boero, G., Cavalli, E., 1996. Forecasting the exchange range: A comparison between econometric and neural network models. AFIR II, 981.

[8] Brill, F.Z., Brown, D.E., Martin, W.N., 1992. Fast genetic selection of features for neural network classifiers, IEEE Trans. Neural Netw. 3(2),324–328.

[9] Chakraborty, D., Pal, N.R., 2001. "Integrated feature analysis and fuzzy rule-based system identification in a neuro-fuzzy paradigm," IEEE Trans. Syst. Man Cybern. B, Cybern. 31(3), 391–400.

[10] Chakraborty, D., Pal, N.R., 2004. A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification. IEEE Trans. Neural Netw. 15(1),110–123.

[11] Dash, M., L iu, H., Yao, J., 1997. Dimensionality reduction for unsupervised data. In Proceedings of 19th IEEE International Conference on Tools with AI. ICTAI.

[12] De, R., Pal, N.R., Pal, S.K., 1997. Feature analysis: Neural network and fuzzy set theoretic approaches, Pattern Recognit. 30(10), 1579–1590.

[13] Debrup, Chakraborty, Pal, N.R., 2008. Selecting Useful Groups of Features in a Connectionist Framework. IEEE Transactions On Neural Networks. 19(3), 381-396.

[14] Derrig, R.A., Ostaszewski K., 1995. Fuzzy techniques of pattern recognition in risk and claim classification. J. Risk Insurance. 62, 447–482.

[15] Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. John Wiley and Sons (Asia) Pte. Ltd.

[16] Engelbrecht, P, 2001. A new pruning heuristic based on variance analysis of sensitivity information. IEEE Trans. Neural Netw. 12(6), 1386–1399.

[17] Fukunaga, K., 1989. Statistical Pattern Recognition. New York: Academic.

[18] Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.

[19] Ivakhnenko, A.G., 1971. Polynomial theory of complex systems, IEEE Trans. Syst., Man Cybern-I. 364–378.

[20] Ivakhnenko, A.G., Madala H.R., 1994. "Inductive learning algorithm for complex systems modeling". Boca raton:CRC Inc.

[21] Jollife, T., 1986. Principal Component Analysis. New York: Springer-Verlag.

[22] Kamber,M., Han, J., 2006. Data mining: Concepts and techniques, 2nd ed. CA: Morgan Kaufmann Publisher. San Francisco.

[23] Kosala, R., Blockeel, H., 2000. Mining Research: A Survey. ACM SIGKDD Explorations. 2 (1), 1–15.

[24] Ladislav, Zjavka, 2010. Generalization Of Patterns By Identification With Polynomial Neural Network, Journal of ELECTRICAL ENGINEERING. 61(2), 120–124.

[25] Liu, H., Yu L., 2005. Towards integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17(4), 491–502.

[26] MacKay, D.J.C., 1992a. Bayesian interpolation. Neural Comput. 4(3), 415–447.

[27] MacKay, D.J.C., 1992b. A practical Bayesian framework for backprop networks, Neural Comput. 4(3), 448–472.

[28] MacKay, D.J.C., 1992c. The evidence framework applied to classification networks, Neural Comput.4(5), 698–714.

[29] Michalski, R.S., Karbonell J.G., Kubat, M., 1 998. Machine Learning and Data Mining: Methods and Applications. John Wiley and Sons, New York.

[30] Misra, B.B., Dehuri, S., Dash, P.K., Panda G., 2008a. A reduced and comprehensible polynomial neural network for classification, Pattern Recognition Letters of Elsevier Journal. 29(2008), 1705–1712.

[31] Misra, B.B., Dehuri, S., Dash, P.K., Panda, G., 2008b. Reduced Polynomial Neural Swarm Net for Classification Task in Data Mining, IEEE Congress on Evolutionary Computation (CEC 2008).

[32] Misra, B.B., Satapathy, S.C., Biswal, B.N., Dash, P.K., Panda, G., 2006a. Pattern classification using polynomial neural networks, IEEE Int. Conf. on Cybernetics & Intelligent Systems (CIS).

[33] Misra, B.B., Satapathy, S.C., Hanoon, N., Dash, P.K., 2006b. Particle swarm optimized polynomials for data classification, Proc. of the IEEE Int. Conf. on Intelligent Systems Design and Application.

[34] Mitchel, T.M., 1997. Machine Learning. McGraw Hill.

[35] Mitra P., Murthy, C.A., Pal, S.K., 2002. Unsupervised feature selection using feature similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence. 24(3), 301-312.

[36] Muller ,J., Lemke, F., Ivakhnenko, A.G.,1998. GMDH algorithms for complex systems modeling. Math and Computer Modeling of Dynamical Systems. 4, 275-315.

[37] Neal, R.M., 1996. Bayesian learning for neural networks, in Lecture Notes in Statistics. Springer-Verlag(118), Berlin, Germany.

[38] Nikolaev, N. L., Iba, H., 1999. Automated discovery of polynomials by inductive genetic programming, in: Zutkow J, Ranch J (eds.), Principles of Data Mining and Knowledge Discovery (PKDD'99), Springer, Berlin, pp. 456-462.

[39] Oh, S.K., Pedrycz, W., Park, B.J., 2003, "Polynomial neural networks architecture: analysis and design," Computers and Electrical Engineering. 29, 703-725.

[40] Oh ,S.K., Pedrycz W., 2002. The design of self-organizing Polynomial Neural Networks. Information Sciences. 2002(141), 237-258.

[41] Pal, N.R., Chintalapudi, K.K., 1997. A connectionist system for feature selection, Neural Parallel Sci. Compute. 5(3), 359–381.

[42] Pal, N.R., 2002. Fuzzy logic approaches to structure preserving dimensionality reduction. IEEE Transactions on Fuzzy Systems. 10(3), 277-286.

[43] Pao, Y.H., 1989. Adaptive pattern recognition neural networks. Addison Wesley, MA.

[44] Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco, CA

[45] Quinlan, J.R., 1987. Generating production rules from decision trees. in: Proc. Internat. Joint Conf. on Artificial Intelligence, San Francisco, CA: MorganKaufmann, pp. 304–307.

[46] Raudys, S.J., Jain A.K., 1991. Small sample size effects in statistical pattern recognition: Recommendations for Practioners. IEEE Transactions on Pattern Analysis and Machine Intelligence. 13(2), 252-264.

[47] Raudys, S.J., Pikelis, V., 1980. On dimensionality sample size classification error and complexity of classification algorithms in pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2, 243-251.

[48] Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A., Jain, A.K., 2000. Dimensionality reduction using genetic algorithms. IEEE Trans. Evol. Comput. 4(2), 164–171.

[49] Rezaee, M.R., Goedhart, B., Lelieveldt, B.P.F., Reiber, J.H.C., 1999. Fuzzy feature selection, Pattern Recognition. 32, 2011–2019.

[50] Ruck, D.W., Rogers, S.K., and Kabrisky M., 1990. Feature selection using a multilayered perceptron. J. Neural Netw. Comput. 40–48.

[51] Saxena, A., Dubey, A., 2010. An Analytical Review of Classification Methodologies. APEEJAY Journal of Management & Technology. ISSN: 0974-3294.5(1), 27-37.

[52] Saxena, A., Pal, N.R., Vora, M., 2010. Evolutionary methods for unsupervised feature selection using Sammon's stress function, Springer Journal on Fuzzy Information and Engineering. 2(3), 229-247.

[53] Setino, R., 1997. Neural network feature selector. IEEE Trans. Neural Netw. 8(3), 654–662.

[54] Sivakumar, A., Kannan, K., 2009. A novel feature selection technique for number classification problem using PNN-A plausible scheme for boiler flue gas analysis Original Research. Article Sensors and Actuators B: Chemical. 139(2), 280-286.

[55] Steppe, J.M., 1996. Integrated feature and architecture selection. IEEE Trans. Neural Netw. 7(4), 1007–1014.

[56] Sykacek, P., 2000. On input selection with reversible jump Markov chain Monte Carlo sampling in Advances in Neural Information Processing Systems. S.A. Solla, T.K. Leen and K.-R. M¨uller (Eds.), MIT Press.12, 638–644.

[57] Trunk, G.V., 1979. A problem of dimensionality: A simple example. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1(3), 306-307.

[58] Tsoi, A.C., Pearson, A.R., 1991. Comparison of three classification techniques, CART,C4.5, and multiplayer perceptrons. Advances in Neural Information Processing Systems 3, 963969.

[59] Yao, X., 1999. Evolving artificial neural networks. Proc. IEEE. 87(9), 1423–1447.

[60] Yumin, Chen, Duoqian, Miao, Ruizhi, Wang, 2010. A rough set approach to feature selection based on ant colony optimization. Pattern Recognition Letters. 31(2010), 226–233.

[61] Zurada, J.M., Malinowski, A., Usui, S., 1997. Perturbation method for detecting r edundant inputs of perceptron networks, Neurocomputing. 14, 177–193.

Figure-1 Basic PNN Model [31]

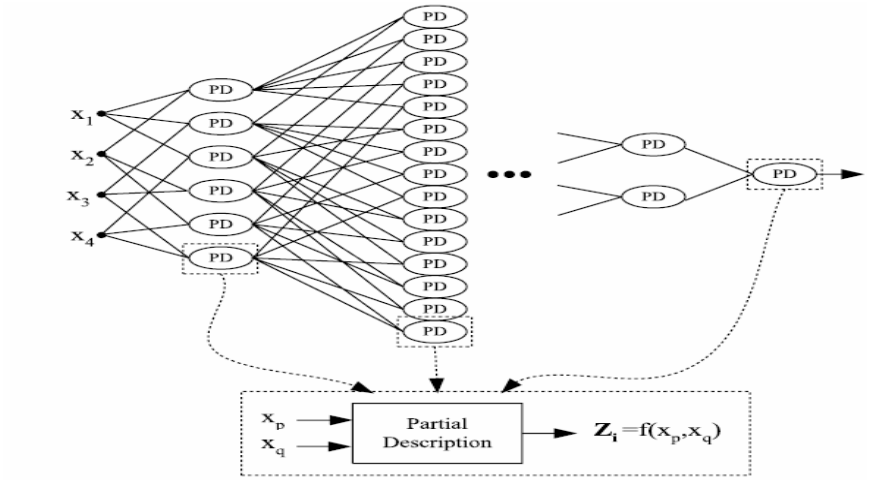


Figure-2 Computation of PD at a node of a PNN’s layer with some inputs treated absent and indicated by 0 using proposed scheme.

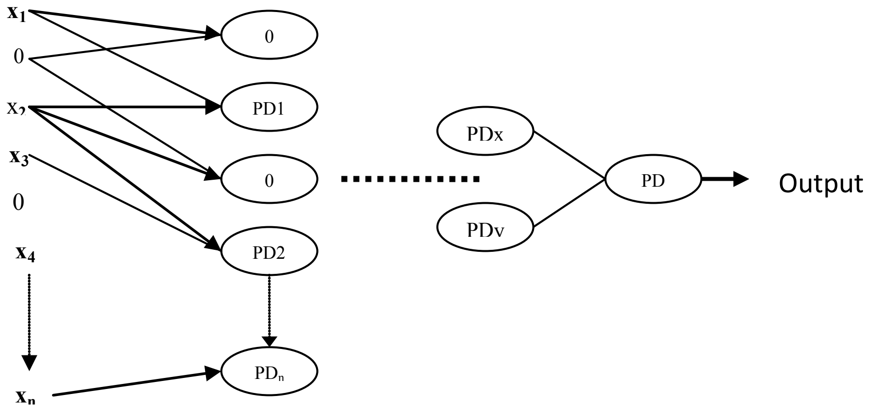


Figure-3 Proposed Technique

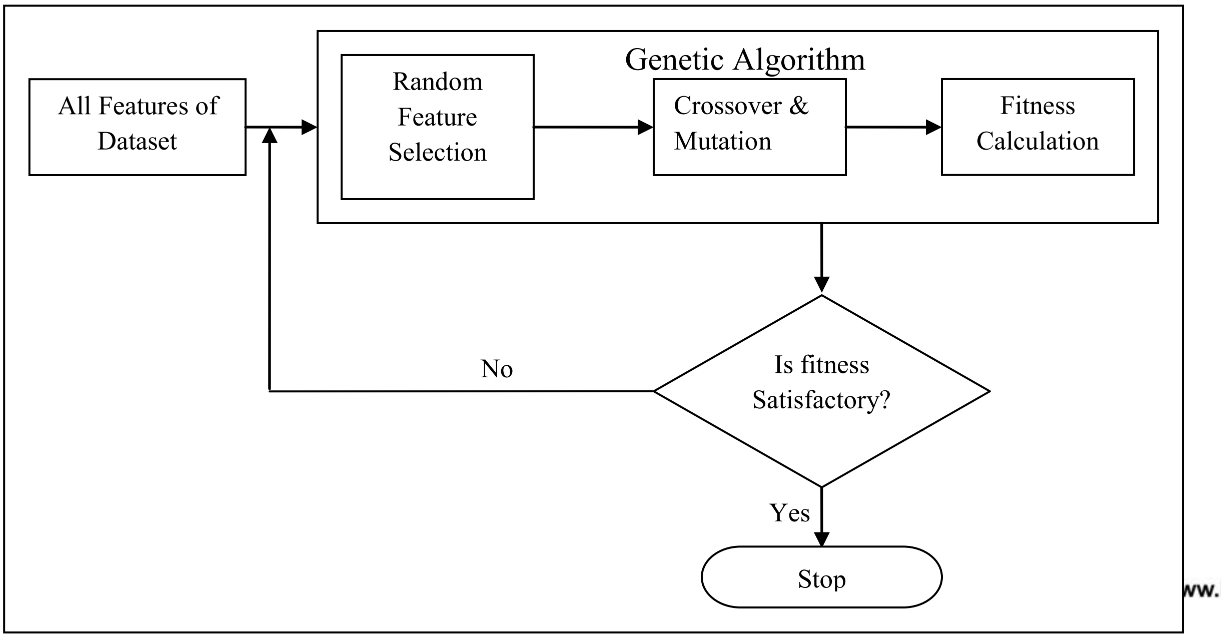


Figure-4
Comparison of Classification Accuracies between proposed scheme and other schemes

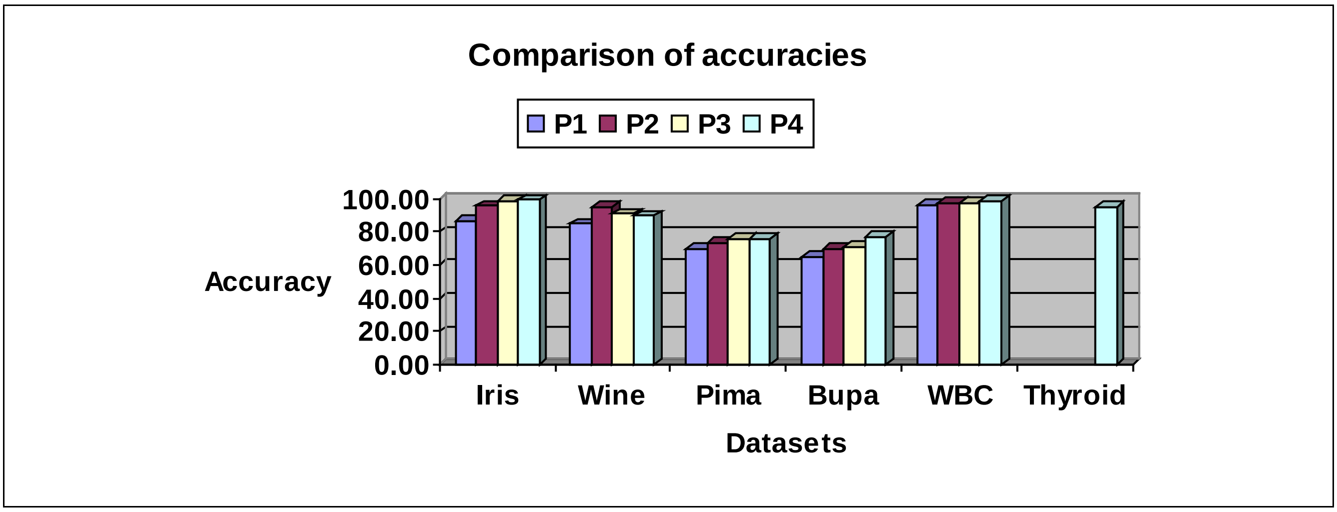
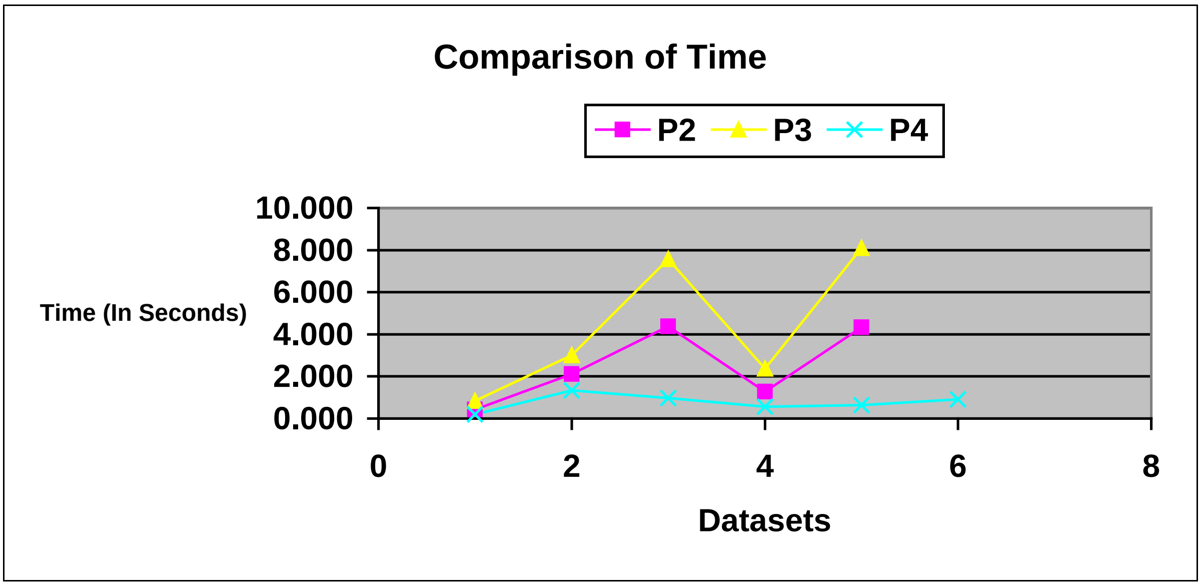


Figure-5
Comparison of execution time between proposed and other schemes



***Results due to P1 being very large have not been shown in this figure.**

Table-1
Description of the Data Set Used

Data Set	Total Patterns	Attributes	Classes	Patterns in Class1	Patterns in Class2	Patterns in Class3
Iris	150	4	3	50	50	50
Wine	178	13	3	59	71	48
Pima	768	8	2	268	500	-
Bupa	345	6	2	145	200	-
WBC	699	9	2	458	241	-
Thyroid	215	5	3	150	35	30

Table-2
Population size and number of generations used in GA

Dataset	Population	Generation
Iris	20	15
Wine	50	20
Pima	60	16
Bupa	25	25
WBC	60	18
Thyroid	20	15

Probability of Crossover =0.5 and Probability of Mutation = 0.3 for all data sets

Table-3
A Comparison of Times of execution and respective Classification Accuracies obtained through proposed scheme with PNN, RCPNN with Gradient Descent and RCPNN with PSO.

Data Set	Feature				Time (In Seconds)				Classification Accuracy			
	P1*	P2*	P3*	P4*	P1*	P2*	P3*	P4*	P1*	P2*	P3*	P4*
Iris	4	4	4	2	254	0.49	0.83	0.193	86.22	95.56	98.67	99.11
Wine	13	13	13	10	449	2.12	3.01	1.34	84.83	95.13	90.95	89.65
Pima	8	8	8	4	1576	4.38	7.57	0.973	69.45	73.35	76.04	75.64
Bupa	6	6	6	4	705	1.28	2.37	0.56	65.29	69.57	70.87	77.34
WBC	9	9	9	5	1425	4.33	8.1	0.64	95.9	97.14	97.64	98.35
Thyroid	-	-	-	4	-	-	-	0.913	-	-	-	95.00

***P1- PNN**
***P2-RCPNN with Gradient Descent**
***P3- RCPNN with PSO**
***P4- Proposed Scheme**