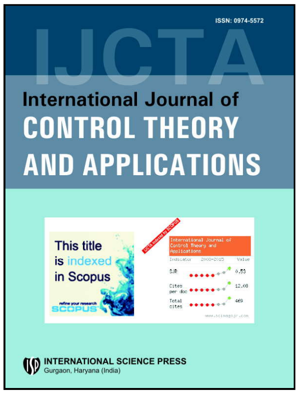

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 14 • 2017

On Improving Generalization of Classifiers using Feature Extraction

Prasenjit Dey[1] and Tandra Pal[1]

[1] *Department of Computer Science and Engineering, NIT Durgapur, India, Emails: prasenjitdey13@gmail.com, tandra.pal@gmail.com*

Abstract: In this work, we have improved the performance of self-organizing map-based probabilistic neural network (SOM-PNN) classifier and self-organizing map-based multilayer perceptron (SOM-MLP) classifier. The training time of the proposed classifiers is also much less than the SOM based classifiers existing in the literature.Firstly, a self-organizing map (SOM) network has been trained with the input data set. Then, a Gaussian function is used in each node of the SOM output grid. It takes the Euclidean distance between the input sample and the corresponding weight vector to produce the output of the node and provides a new representation of the input samples.However, the new representation is high in dimensionality, equal to the number of nodes in the SOM output grid.For data set with a larger number of training samples, high dimensionality may increase the training time of the classifiers. So, instead of considering all the samples of all classes, we have selected only half of the existing class samples from all classes using SOM clustering algorithm. It reduces the training time of the classifiers. PNN or MLP classifier is used over this subset of newly represented input samples. To show the superiority of the proposed SOM-PNN with reduced training set (RESOM-PNN), we have compared it with three different PNN classifiers: traditional PNN, PNN with reduced training set and SOM-PNN using ten standard classification data sets. Similarly,the proposed SOM-MLP with reduced training set (RESOM-MLP)has been compared with the traditional MLP, MLP with reduced training set and SOM-MLP classifiers using those ten data sets. Comparison results indicate the importance of the proposed SOM-based method for reduced extracted feature representation. Using Friedman test, we have shown that the proposed classifier is significantly better than the other existing classifiers.

Keywords: Feature extraction,Gaussian function, generalization, probabilistic neural network, multilayer perceptron, self-organizing map

1. INTRODUCTION

There are various types of artificial neural networks (ANNs) used as classifiers like probabilistic neural network (PNN), multilayer perceptron (MLP), radial basis function network (RBFN),etc. Generalizationis important for pattern classification task. However, the generalization of the classifiers is affected by overfitting problem. Foroverfitting problem, the classifiers show good accuracy with training patterns but their accuracy degrades with test patterns [1, 2]. Thus, we need to have a classifier with more generalization capability. Further, in many approachesexisting in the literature [2, 3], the generalization is achieved by compromising the time complexity of the algorithm. It demands a time efficient classifier with more generalization ability.

PNN is used as a classifier in many applications of different domains [4]. It is prone to overfitting if it fails to generate the proper class-wise probability density functions [1]. Figure 1 illustrates the basic architecture of PNN classifier. Like PNN, MLP may also fail to produce a satisfactory result due to overfitting [2]. It occurs when MLP fails to define class-wise nonlinear decision surfaces. Generally, it happens when the number of hidden nodes is more than the required. Figure 2 illustrates the basic architecture of MLP with the single hidden layer. Generalization ability depends on the representation of training data set which can be improved by feature extraction technique.

Self-organizing map (SOM) network has been used as a feature extraction tool in many neural network applications.Itmaps inputpattern from a high dimensional space to a low dimensional space [5]. It preserves inputspace topology using a neighborhood function. In SOM network, there are two layers: input layer and output layer. All the nodes of the output layer is connected with all the nodes of the input layer and each of the connections is associated with a weight. The weight vector corresponding to the connections of a node of the output layer to all the nodes of the input layer is called a codebook vector. The output layer represents m-dimensional (usually two-dimensional) feature map of input samples. The architecture of SOM network is shown in Figure 3.

SOM-based hybrid classifiers havemore generalization capability than conventional classifiers.However, the problem with this hybrid model is that the new feature space is high in dimensionality.For large data set, it increases the training time of the classifiers. The objective of the proposed approach is to improve the training time of the SOM-based hybrid classifiers.Here, we have considered SOM clustering algorithm to select only half of the input samples of each classes. It reduces the training time of the classifiers. Finally, the PNN or MLP classifier performs classification task over this reduced subset of extracted feature space. The

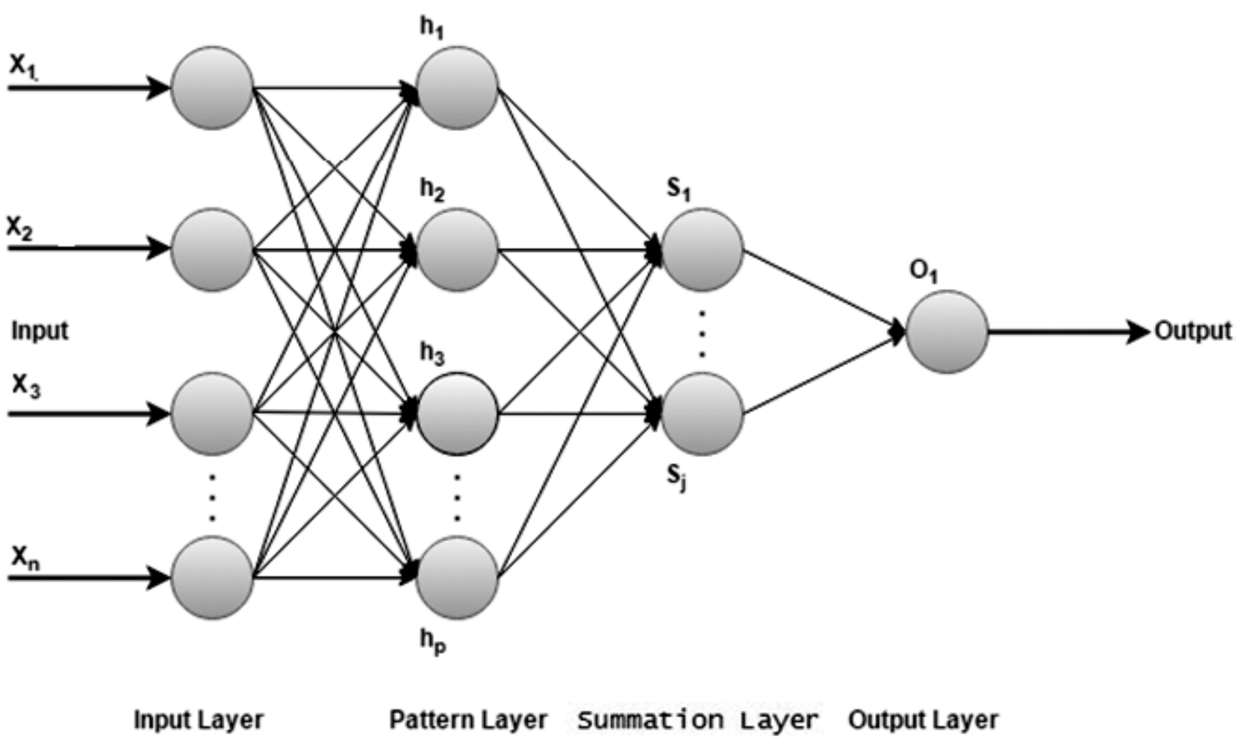

Figure 1: Architecture of ProbabilisticNeural Network

proposed model of classifier shows better classification accuracy than other existing classifiers even in reduced training time.

The organization of the paper is as follows. In Section II, we briefly review the literature of the existing workson generalization of classifiers. The proposed model along with the selection ofreduced subset from theextracted features is presented in Section III. The experimental results are shown in Section IV. Finally, conclusionis drawn in Section V.

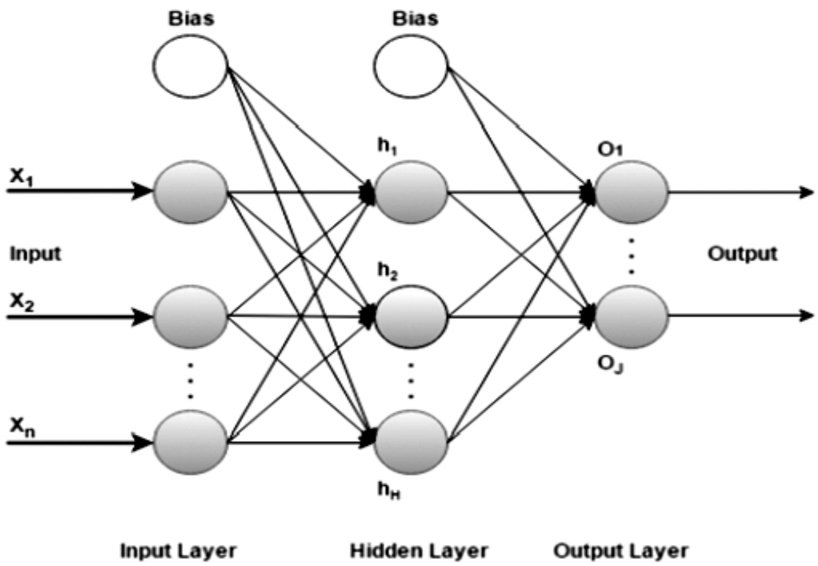

Figure 2:Architecture of Multilayer perception

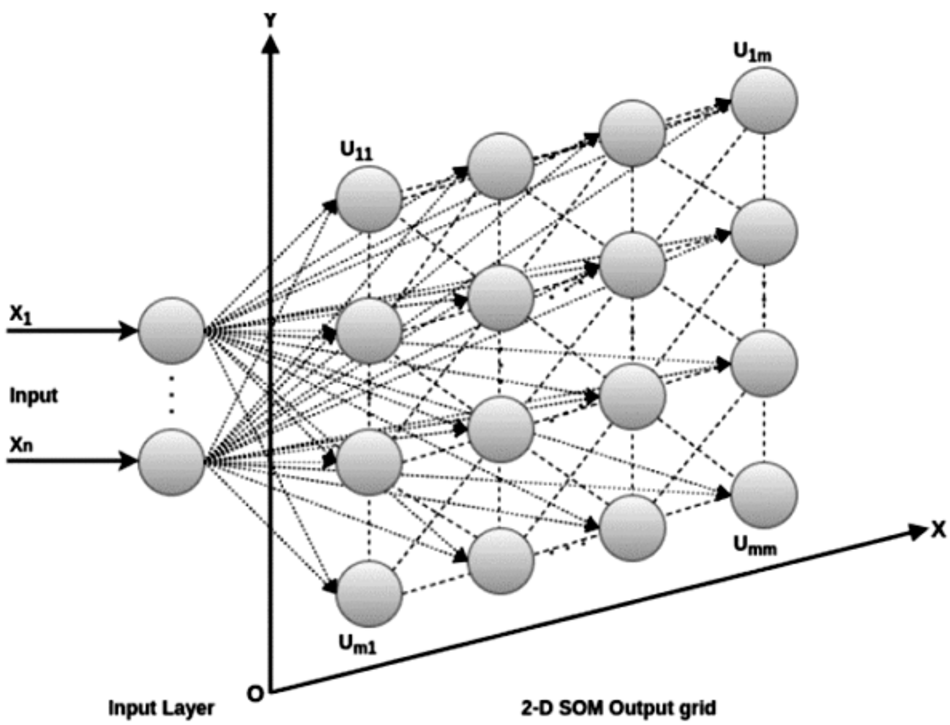

Figure 3: Architecture of Self-OrganizingMap

2. THE LITERATURE

In last few decades, PNN and MLP classifiers have been used in many neural network applications. The generalization capability of the PNN classifier was initially studied in [4]. The simulation results demonstrate that in most of the cases the PNN classifier has better generalization than back propagation neural network. In [6], the authors have used maximum likelihood learning approach to train the PNN classifier. Their algorithm generalized the class conditional probabilities even if the number of training samples for a particular class is less. Ganchev et al. [7] have studied generalized locally recurrent PNN. The authors, in [6], have used a recurrent layer between the pattern and output layers. It helps the model to identify the temporal and spatial correlation of the input data.Wang et al, in [8], for ECG arrhythmia classification, have used PNN classifier combined withfeature reduction methodbased onprincipal component analysis (PCA) and linear discriminant analysis(LDA). In [9], the author has shown that the generalization capability of the MLP can be improved by adding noise during the training of the network. In [10], the authors have used a population based incremental approach to improve the generalization of the multilayer perceptron. Bernier et al. [11] have used a regularization based approach to improve the generalization and noise immunity of the MLP. In [12], the authors have selected the training samples from the decision boundary without any information of the actual boundary using k-nearest neighbor to avoid overfitting in MLP classifiers.Below we have mentioned some approaches, existing in the literature, using PNN and MLP classifiers based on SOM.

2.1. SOM-based Approaches

SOM network has been used as a feature extraction tool in many neural network applications to improve the generalization [3, 13-16]. SOM is applied in a hybrid architecture with PNN in [13]. The model is used for volume data classification, where segment CT sloth data is used as volume data. In [14], SOM is used combined with PNN to reduce the training data set for the PNN classifier. Song et al., in [15], have used a hybrid structure of SOM and weighted PNN to estimate the PDF of the pixels of magnetic resonance images. In [3], the authors have proposed a new variant of SOM network with PNN classifier. Here, SOM network has been used as feature extraction tool and PNN as classifier over the newly extracted feature space. Empirical results demonstrate that their method improves generalization but increases the training time of the classifier.

In [17], the authors have proposed a neural network classifier combining SOM and MLP for isolated word recognition. They proposed that the two-dimensional feature map, which is first trained by SOM and then fine-tuned by K-means clustering algorithm, can beused as a sequential mapping function. It transforms acoustic vector sequences of speech signals into trajectories in a square binary matrix. An MLP is then used to classify the trajectories corresponds toeach word in the vocabulary. Another approach combining SOM and MLP for texture classification is presented in [18]. Here, the authors have extracted the texture features from the image with the help of Gabor filters. Thereafter, SOM is used over the outputof Gabor filter banks. Itmaps the inputs into*m*dimensional space, where *m* is the dimension of the SOM output layer. Then, the transformed feature vectors are fed into the MLP for classification. Huang et al., in [19], have proposed a hybrid model based onSOM and back-propagation neural network (BPNN) topredictthe remaining lifetimeofball bearing. After deriving minimum quantization error (MQE) indicator by SOM, BPNN is used over this to estimate residual life of ball bearing. In [20], the authors have proposed a cascaded architecture of neural networks with feature map using two-dimensional SOM and MLP. Here, after obtaining a two-dimensional map value corresponding to each input pattern, they fed the modified input pattern to the MLP for EMG classification.

2.2. ICA-based Approaches

Independent Component Analysis (ICA) is another feature extraction technique which transform the features in such a way that they become mutually independent. It has been used in combination with PNN classifier or back propagation neural network for EEG beat classification in [21]. Here, the authors empirically show that ICA combined with PNN performs better than ICA combined with back propagation neural network. In [22], the authors have demonstrated that regression by ICA and regression by MLP are similar to each other. They have also claimed that the output of each hidden layer neuron in an MLP is an estimate of one independent component.

3. PROPOSED METHOD

In the proposed model,the performance of PNN and MLP classifiers has been improved with the help of SOM. SOM is used for feature extraction of the input patterns. For this purpose, a Gaussian function is used in each node of the trained SOM output grid which provides a new representation of the input patterns. Then, to make it efficient for larger data sets, we have used SOM clustering algorithm which selects half of the existing samples from each of the classes. It gives a reduced extracted feature space from the original feature space. We have considered two neural network models,PNN and MLPto perform classification task over the newly formed reduced feature space. In case of PNN classifier, the proposed hybrid model is called as Reduced Extracted SOM-based PNN (RESOM-PNN) classifier. Similarly, the proposed hybrid model for MLP classifier is named as Reduced Extracted SOM-based MLP (RESOM-MLP) classifier. The proposed method is explained below in detail.

The output layer of SOM network is represented by two-dimensional grid of size $m \times m$, where the size of each dimension is m. There exists a weight vector w_j from each output node u_j to all input nodes. The weights between input and output layers are initialized randomly within the range [-0.5, 0.5].Let the p^{th} input pattern $x^p = (x_1, x_2 ... x_n)$ is fed into the SOM network, where n is the number of features available in the data set. For each node,a similarity measure d_j between the input x^p and the weight vector w_j is computed using Euclidean distance. The output node, which has the minimum value of d_j is the winning node for the corresponding input. Only the weight vector of the winning node along with the weight vectors of the nodes in its topological neighborhood are updated in the direction of the input. The neighborhood function used in our work is given below in (1).

$$g_i = \exp\left(-\frac{\left\|u_j - \hat{u}_j\right\|}{2\sigma^2}\right) \tag{1}$$

Here, g_j is the output of the j^{th} output node, σ is the spread of the neighborhood function and $\hat{u}_j$ is the winning node of SOM. The weight vector w_j^{t+1} is updated following (2).

$$\mathrm{w}_j^{t+1} = \mathrm{w}_j^t + \eta g_j \left\|\mathrm{x}^p - \mathrm{w}_j\right\| \tag{2}$$

where η is the learning rate, which has been initializedto $\eta_0 = 0.9$. During first *90%* of the training time, the learning rate has been reduced uniformly from 0.9 to 0.1 using (3).

$$\eta_i = \eta_0 \exp(-i/\tau) \tag{3}$$

Here, η_i is the learning rate at i^{th}iteration and τ is set to half of the training time. Weight updationfor each of the training patterns $X = [x^1, x^2, ..., x^P]$ available in the data set is done500times.After trainingthe network, we haveused a Gaussian function over the output node of SOM to get the new representation of the input pattern x^p. The final value $\hat{x}_j$ of each output node u_j in SOM is obtained using the Gaussian function over theEuclidean distance between the input patternx^p and the weight vector w_j. The corresponding mapping function is given below in (4).

$$\hat{x}_j = \exp\left(-\frac{\left\|x^p - w_j\right\|}{2\sigma^2}\right) \tag{4}$$

After achieving the new representation $\hat{x}p \in R^{m\times m}$ from the original space $x^p \in R^n$, SOM clustering algorithmhave been used to select the representative samples from each of the classes. Here, we have selected half of the training samples for each class. It reduces the number of hidden nodes in PNN classifier. As a consequence, training time of PNN classifier is also reduced. Similarly, in case of MLP, it reduces the training time by reducing the number of training samples to half of the original samples. Finally, PNN or MLP classifier is used over this new representation to perform the classification task.

In case of PNN classifier, we have fed the modified input pattern $\hat{x}p$ into the input layer which is directl**y** passed to the pattern layer. Each node i in the patternlayer is associated with a center and a spread associated with Gaussian activation function. The i^{th} input pattern acts as the center c_i of the i^{th} pattern layer node and σ_i determines the width of the corresponding Gaussian function. The output of the i^{th}pattern layer node is evaluated using (5) as given below.

$$\phi_i = \frac{1}{2\pi^{n/2}\sigma^n}\exp\left(-\frac{\left\|\hat{x}^p - c_i\right\|^2}{2\sigma^2}\right), i = 1,\ldots,P/2 \tag{5}$$

If the output class label of i^{th} input pattern is j, then the i^{th} node in the pattern layer is connected to the node j of the summation layer. The output of the j^{th} node in the summation layer is given below in (6).

$$s_i = \frac{1}{2P_j\pi^{n/2}\sigma^n}\sum_{p=1}^{P_j}\left(\exp\left(-\frac{\left\|\hat{x}^p - c_i\right\|^2}{2\sigma^2}\right)\right), j = 1,\ldots,C \tag{6}$$

Here P_j denotes the number of the patterns belonging to class j. Finally, each node of summation layer computes the sum of the pattern layer outputs connected to it and passes it to the node of output layer. The decision node outputs the class with the largest activation.

In case of an MLP, the feature vector x^p is fed in the input layer. Then, the hidden layer nodes compute the net output using the sigmoid as an activation function, given below in (7).

$$z_h = S\left(\sum_{i=1}^{n} w_{hi}^{Input}\,\hat{x}_i + b_h^{Input}\right) \; for 1 \le h \le H \tag{7}$$

Here, z_h is the output of the h^{th} hidden node, $S(.)$ is sigmoid activation function, w_{ih}^{Input} is weight between i^{th} input node and h^{th} hidden node, b_h^{Input} is the bias of h^{th} hidden node and H is the number of nodes in the hidden layer. The last layer acts as the network output layer. The output of the j^{th} node of the output layer is give below in (8).

$$o_j = S\left(\sum_{h=1}^{H} w_{jh}^{Hidden} z_h + b_j^{Hidden}\right) \quad for 1 \le j \le C \tag{8}$$

Here, w_{jh}^{Hidden} is the weight between h^{th} hidden node and j^{th} output node, b_j^{Hidden} is the bias of the j^{th} output node and C is the number of classes in data set. Finally, the output value computed from (8) is compared with the target output to obtain the network error, which is used to update the weights using gradient descent learning.

4. EXPERIMENTATION

4.1. Experimental Settings

The results, presented in this section are based on ten data sets, taken from the Keel website [23]. The detail of all the data sets are presented in Table 1. Neural network toolbox of MATLAB (Version 8.1) has been used for simulation purpose to assess the performance of the scheme. The spread value in the probability density function, for all four classifiers: traditional PNN, PNN with reduced training set, SOM-PNN, and the proposedSOM-PNN with reduced training set,i.e., (RESOM-PNN), is set to the same value. Similarly, for all of the four classifiers traditional MLP, MLP with reduced training set, SOM-MLP, and the proposed SOM-MLP with reduced training set(RESOM-MLP), we have considered gradient descent learning with the single hidden layer consisting of ten hidden nodes and the number of epochs was fixed at 10000 with learning rate η =0.1. We have first performed z-score normalization on the training data set and found the mean and standard deviation. The mean and standard deviation have been used for z-score normalizationof the test data set.

Table 1
Summary of Data Sets

Data set	*# Features*	*# Classes*	*Class Distribution*
Sonar	60	2	208 (97, 111)
Spectfheart	44	2	267(212,55)
Ionosphere	34	2	351(225, 126)
Vehicle	18	4	846 (212, 217, 218, 199)
Cleveland	13	5	297 (54, 35, 35, 13, 160)
Wine	13	3	178 (59, 17, 48)
Glass	9	6	214 (70, 76, 17, 13, 9, 29)
Pima	8	2	768 (500, 268)
Bupa	6	2	345 (145, 200)

4.2. Experimental Results: Comparison with other Models

To compare the proposed model (RESOM-PNN) with other classifiers, ten-fold cross-validation has been performed ten times for each of ten data sets. For all data sets, the mean and the standard deviation of classification accuracy, obtained from PNN, PNN with reduced training set, SOM-PNN and the proposed RESOM-PNN, is shown in Table 2. Similarly for comparison purpose, the mean and the standard deviation of classification accuracy of the four classifiers of MLP, i.e., traditional MLP, MLP with reduced training set, SOM-MLP and the proposed RESOM-MLPis shownin Table 3. Both the Tables show that the proposed SOM based classifiers (with reduced training set) have performed better in most of the cases among ten data sets. Missing values of the features of the Pima data set have been replaced by the mean values of the corresponding features.

Table 2
Classification accuracy (%) of traditionalPNN, PNN with reduced training set, SOM-PNNand the proposed SOM-PNN with reduced training set (RESOM-PNN)

Data set	PNN		SOM-PNN	
	With all samples	*With Reduced training set*	*With all samples*	*Proposed model with Reduced training set (RESOM-PNN)*
Sonar	72.46±4.13	68.22±2.26	73.13±1.36	73.52±2.15
Spectfheart	75.23±1.26	71.51±1.23	75.99±1.82	75.57±1.52
Ionosphere	74.47±2.87	74.03±1.32	76.02±3.08	75.94±3.87
Vehicle	69.17±1.43	68.54±0.92	71.96±1.49	72.17±1.21
Cleveland	61.02±0.98	62.33±0.74	63.58±1.95	64.41±1.85
Wine	94.27±1.62	95.87±2.06	95.28±1.52	98.17±0.13
Glass	66.80±2.37	65.99±1.62	68.24±1.25	70.24±1.61
Pima	69.71±1.23	69.85±0.52	70.51±2.54	71.52±2.85
Bupa	64.92±3.74	65.28±2.75	64.82±2.71	65.12±4.02
Tae	51.32±2.95	49.76±1.82	51.73±3.37	52.21±4.88

4.3. Statistical Significance: Friedman Test

To compare (i) the proposed method (RESOM-PNN) with SOM-PNN, PNN with reduced training set, and traditional PNN and (ii) the proposed method (RESOM-MLP) with SOM-MLP, MLP with reduced training set, and tradition MLP, we have performed statistical validation testing using Friedman test. In both the scenarios, the null hypothesis at á =0.05 level of significance has been rejected on the basis of the p-values obtained from the tests. It implies that the proposed method is significantly better than the other comparing classifiers.

Table 3
Classification accuracy (%) of traditional MLP, MLP with reduced training set, SOM-MLP and the proposedSOM-MLP with reduced training set (RESOM-MLP)

Data set	MLP		SOM-MLP	
	With all samples	*With Reduced training set*	*With all samples*	*Proposed model with Reduced training set (RESOM-MLP)*
Sonar	74.92±1.67	72:00±0.96	76.13±1.53	76.52±2.15
Spectfheart	77.23±2.36	75:18±1.63	78.78±0.56	78.36±1.98
Ionosphere	76.85±1.92	74:64±2.08	78.29±1.37	78.24±2.08
Vehicle	66.32±2.12	66:29±0.83	72.24±1.98	73.23±2.42
Cleveland	65.23±1.85	66.82±0.96	68.02±2.69	68.54±1.84
Wine	96.25±2.41	97:75±0.45	97.65±1.42	98.37±1.87
Glass	64.89±4.23	61:29±0.50	70.32±3.33	69.85±2.65
Pima	68.96±1.56	69:82±1.53	70.41±2.12	71.36±3.25
Bupa	63.29±4.01	64:73±1.96	65.31±2.46	65.98±3.88
Tae	50.98±2.26	48:79±1.61	52.52±1.99	53.02±2.63

5. CONCLUSION

Proper representation of data is required to improve the generalization of neural network classifiers. For this purpose, hybrid architecturesfor PNN and MLP classifiers have been developed in the article based on SOM. In the proposed model, after the training of the SOM network using the input data set, a Gaussian functionis used in the output layerto map the input space into a better feature space. But this representation of input data set increases the dimensionality of the original input data set. It will increase the training time of the classifiers, especially for the large data set.We have used SOM clustering algorithm to select half of the training class samples from all classes. The whole process improves the training time of the SOM-PNN and SOM-MLP based classifiers. Empirical results show that the proposed method improves the classification accuracy of the SOM-PNN and SOM-MLP based classifiers even in re-duced training time. Statistical tests show that our RESOM-PNN model has better generalization capability than traditional PNN, PNN with reduced training set,SOM-PNN model.The proposed RESOM-MLP model also shows better generalization capabil-ity than MLP, MLP with reduced training set,SOM-MLP.

REFERENCES

[1] Grim, J., Hora, J.: "Computational properties of probabilistic neural networks" In: International Conference on Artificial Neural Networks, Springer (2010) 31–40.

[2] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets" Neural computation 18, no. 7 (2006): 1527-1554.

[3] Dey, P., Pal, T.: "Extended self organizing map with probabilistic neural network for pattern classification problems" In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, Springer (2015) 167–174.

[4] Specht, D.F., Shapiro, P.D.: "Generalization accuracy of probabilistic neural net-works compared with backpropagation networks" In: Neural Networks, 1991., IJCNN-91-Seattle International Joint Conference on. Volume 1., IEEE (1991) 887–892

[5] Kangas, J.A., Kohonen, T.K., Laaksonen, J.T.: "Variants of self-organizing maps" IEEE transactions on neural networks 1(1) (1990) 93–99

[6] Streit, Roy L., and Tod E. Luginbuhl. "Maximum likelihood training of probabilistic neural networks"*IEEE Transactions on Neural Networks* 5, no. 5 (1994): 764-783.

[7] Ganchev, Todor D., Dimitris K. Tasoulis, Michael N. Vrahatis, and Nikos D. Fakotakis. "Generalized locally recurrent probabilistic neural networks with application to text-independent speaker verification."*Neurocomputing* 70, no. 7 (2007): 1424-1438.

[8] Wang, J.S., Chiang, W.C., Hsu, Y.L., Yang, Y.T.C.: "Ecg arrhythmia classification using a probabilistic neural network with a feature reduction method" Neurocom-puting 116 (2013) 38–45

[9] An, Guozhong. "The effects of adding noise during backpropagation training on a generalization performance"*Neural Computation* 8, no. 3 (1996): 643-674.

[10] Galiæ, Elvis, and Markus Höhfeld. "Improving the generalization performance of multi-layer-perceptrons with population-based incremental learning" In *International Conference on Parallel Problem Solving from Nature*, pp. 740-750. Springer Berlin Heidelberg, 1996.

[11] Bernier, Jose Luis, Julio Ortega, E. Ros, Ignacio Rojas, and Alberto Prieto. "A quantitative study of fault tolerance, noise immunity, and generalization ability of MLPs" *Neural Computation* 12, no. 12 (2000): 2941-2964.

[12] Chaudhuri, B., Bhattacharya, U.: "Efficient training and improved performance of multilayer perceptron in pattern classification" Neurocomputing 34(1) (2000) 11–27.

[13] Ma, F., Wang, W., Tsang, W.W., Tang, Z., Xia, S., Tong, X.: "Probabilistic segmen-tation of volume data for visualization using som-pnn classifier" In: Proceedings of the 1998 IEEE symposium on Volume visualization, ACM (1998) 71–78.

[14] Chen, Jie, Haiying Li, Shiwei Tang, and Jincai Sun. "A SOM-based probabilistic neural network for classification of ship noises" In *Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on*, vol. 2, pp. 1209-1212. IEEE, 2002.

[15] Song, T., Jamshidi, M.M., Lee, R.R., Huang, M.: "A modified probabilistic neural network for partial volume segmentation in brain mr image." IEEE Transactions on Neural Networks 18(5) (2007) 1424–1432

[16] Liu, Y., Weisberg, R.H., Mooers, C.N.: "Performance evaluation of the self-organizing map for feature extraction" Journal of Geophysical Research: Oceans 111(C5) (2006)

[17] Kuang, Z., Kuh, A.: "A combined self-organizing feature map and multilayer per-ceptron for isolated word recognition" IEEE Transactions on Signal processing 40(11) (1992) 2651–2657

[18] Raghu, P., Poongodi, R., Yegnanarayana, B.: "A combined neural network approach for texture classification" Neural Networks 8(6) (1995) 975–987

[19] Huang, R., Xi, L., Li, X., Liu, C.R., Qiu, H., Lee, J.: "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods" Mechanical Systems and Signal Processing 21(1) (2007) 193–207

[20] Huang, H.P., Liu, Y.H., Liu, L.W., Wong, C.S.: "Emg classification for prehensile postures using cascaded architecture of neural networks with self-organizing maps" In: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on. Volume 1., IEEE (2003) 1497–1502

[21] Yu, Sung-Nien, and Kuan-To Chou. "Integration of independent component analysis and neural networks for ECG beat classification" *Expert Systems with Applications* 34, no. 4 (2008): 2841-2846.

[22] Hyvärinen, A., Bingham, E.: "Connection between multilayer perceptrons and re-gression using independent component analysis" Neurocomputing 50 (2003) 211– 222.

[23] http://sci2s.ugr.es/keel/datasets.php.