Expert Systems with Applications 36 (2009) 8610–8615

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A comparative study on diabetes disease diagnosis using neural networks

Hasan Temurtas [a], Nejat Yumusak [b], Feyzullah Temurtas [c,d,*]

[a] Dumlupinar University, Department of Electrical and Electronics Engineering, Kutahya, Turkey
[b] Sakarya University, Department of Computer Engineering, 54187 Adapazari, Turkey
[c] Sakarya University, Department of Electrical and Electronics Engineering, 54187 Adapazari, Turkey
[d] Bozok University, Department of Electrical and Electronics Engineering, 66200 Yozgat, Turkey

ARTICLE INFO

Keywords:
Diabetes disease diagnosis
Multilayer neural network
Levenberg–Marquardt algorithm
Probabilistic neural network

ABSTRACT

Diabetes occurs when a body is unable to produce or respond properly to insulin which is needed to regulate glucose. Besides contributing to heart disease, diabetes also increases the risks of developing kidney disease, blindness, nerve damage, and blood vessel damage. Diabetes disease diagnosis via proper interpretation of the diabetes data is an important classification problem. In this study, a comparative pima-diabetes disease diagnosis was realized. For this purpose, a multilayer neural network structure which was trained by Levenberg–Marquardt (LM) algorithm and a probabilistic neural network structure were used. The results of the study were compared with the results of the pervious studies reported focusing on diabetes disease diagnosis and using the same UCI machine learning database.

1. Introduction

Diabetes is a major health problem in both industrial and developing countries, and its incidence is rising. It is a disease in which the body does not produce or properly use insulin, the hormone that "unlocks" the cells of the body, allowing glucose to enter and fuel them (Mohamed et al., 2002). Diabetes increases the risks of developing kidney disease, blindness, nerve damage, blood vessel damage and it contributes to heart disease. The cause of diabetes continues to be a mystery, although both genetics and environmental factors such as obesity and lack of exercise appear to play roles (Polat, Gunes, & Aslan, 2008). The most common form of diabetes is Type 2 diabetes (Acharya, Tan, & Subramaniam, 2008). This type diabetes results from insulin resistance (a condition in which the body fails to properly use insulin), combined with relative insulin deficiency. In Type 2 diabetes, either the body does not produce enough insulin or the cells ignore the insulin (Polat & Gunes, 2007). Although detection of diabetes is improving, about half of the patients with Type 2 diabetes are undiagnosed and the delay from disease onset to diagnosis may exceed 10 years. Thus, earlier detection of Type 2 diabetes and treatment of hyperglycaemia and related metabolic abnormalities is of vital importance. The Pima Indians of Arizona have the highest prevalence and incidence of Type 2 diabetes of any population in the world (Baier & Hanson, 2004; Knowler et al., 1991).

There are many factors to analyze to diagnose the diabetes of a patient, and this makes the physician's job difficult. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. But, this is not easy considering the number of factors she has to evaluate (Polat et al., 2008). To help the experts and helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, classification systems provide medical data to be examined in shorter time and more detailed. Expert systems and different artificial intelligence techniques for classification systems in medical diagnosis is increasing gradually. As for other clinical diagnosis problems, classification systems have been used for diabetes diagnosis problem (Polat & Gunes, 2007).

The multilayer neural networks (MLNNs) have been successfully used in replacing conventional pattern recognition methods for the disease diagnosis systems (Delen, Walker, & Kadam, 2005; Kayaer & Yıldırım, 2003; Temurtas, 2009). The back-propagation (BP) algorithm (Rumelhart, Hinton, & Williams, 1986) is widely recognized as a powerful tool for training of the MLNNs. But, since it applies the steepest descent method to update the weights, it suffers from a slow convergence rate and often yields suboptimal solutions (Brent, 1991; Gori & Tesi, 1992). A variety of related algorithms have been introduced to address that problem. A number of researchers have carried out comparative studies of MLNN training algorithms (Gulbag, 2006; Gulbag & Temurtas, 2006; Hagan, Demuth, & Beale, 1996; Hagan & Menhaj, 1994). Levenberg–Marquardt (LM) algorithm (Hagan & Menhaj, 1994) used in this study provides generally faster convergence and better estimation results than other training algorithms (Gulbag, 2006; Gulbag & Temurtas, 2006). On the other hand, the pervious

* Corresponding author. Tel.: +90 532 207 56 33.
E-mail address: temurtas@gmail.com (F. Temurtas).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.10.032

diagnosis result of LM algorithm reported on Pima Indian diabetes disease dataset (Kayaer & Yıldırım, 2003) was not better than the diagnosis results of other training algorithms. This can be because of that, LM algorithm converges very fast but it can cause the memorization effect when the overtraining occurs. If a neural network starts to memorize the training set, its generalization starts to decrease and its performance may not be improved for untrained test sets. Memorization of the training set can be because of the overtraining (Gulbag, 2006; Gulbag & Temurtas, 2006). So, before the starting of memorization, we must determine the optimum trained neural network using the maximum accuracy value of the test data. Using the optimum trained neural network, Pima Indian diabetes disease diagnosis can be made with better accuracy.

The probabilistic neural network (PNN) developed by Specht (1990) is a network formulation of 'probability density estimation'. It is a model based on competitive learning with a 'winner takes all attitude' and the core concept based on multivariate probability. The PNN provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. The network paradigm also uses Parzen Estimators which were developed to construct the probability density functions required by Bayes theory. The PNN uses a supervised training set to develop distribution functions within a pattern layer. The PNN structure is very useful for the classification problems and disease diagnosis systems (Gulbag, 2006; Gulbag, Temurtas, & Yusubov, 2007; Specht, 1990; Temurtas, 2009).

In this study, a comparative Pima Indian diabetes disease diagnosis was realized by using two different types of neural networks. The neural network classification models are multilayer neural network which was trained by LM algorithm and probabilistic neural network. In order to perform the study, the UCI machine learning database which is very commonly used among the other classification systems were used (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Hoshi et al., 2005; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008; Temurtas, 2009; ftp://ftp.ics.uci.edu/pub/machine-learningdatabases (accessed: 15.02.08)). The conventional (one training and one test) validation and 10-fold cross-validation techniques were performed to compare the accuracy of the neural network models. The results were also compared with the results of the pervious studies reported (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008) focusing on pima-diabetes disease diagnosis and using the same database.

2. Method

2.1. Data source

In order to perform the research reported in this manuscript, Pima Indian diabetes dataset taken from the UCI machine learning respiratory were used (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008; ftp://ftp.ics.uci.edu/pub/machine-learningdatabases (accessed: 15.02.2008)). The reason for using this dataset is that because it is very commonly used among the other classification systems that we have used to compare this study with for Pima Indian diabetes diagnosis problem. All patients in this database are Pima Indian women at least 21 years old and living near Phoenix, Arizona, USA.

The dataset which consists of Pima Indian diabetes disease measurements contains two classes and 768 samples. The class distribution is

- Class 1: normal (500)
- Class 2: Pima Indian diabetes (268)

All samples have eight features. These features are:

- Feature 1: Number of times pregnant.
- Feature 2: Plasma glucose concentration a 2 h in an oral glucose tolerance test.
- Feature 3: Diastolic blood pressure (mm Hg).
- Feature 4: Triceps skin fold thickness (mm).
- Feature 5: 2-h serum insulin (μU/ml).
- Feature 6: Body mass index (weight in kg/(height in m)^2)
- Feature 7: Diabetes pedigree function Feature 8: 2-h serum insulin (μU/ml).
- Feature 8: Age (years).

2.2. Previous studies

Classification systems have been used for Pima Indian diabetes disease diagnosis problem as for other clinical diagnosis problems. There have been a lot of studies reported focusing on Pima Indian diabetes disease diagnosis (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008). These studies applied different methods to the given problem and achieved high classification accuracies using the dataset taken from UCI machine learning repository. Polat et al. (2008) presented a cascade learning system based on Generalized Discriminant Analysis (GDA) and Least Square Support Vector Machine (LS-SVM) to diagnosis of Pima Indian diabetes disease. They have reported 78.21% classification accuracy using LS-SVM with 10-fold cross-validation (10x FC). They have also reported 79.16% classification accuracy using GDA–LS-SVM with 10x FC. Polat and Gunes (2007) have reported 89.47% classification accuracy (10x FC) on Pima Indian diabetes disease diagnosis using principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS). Kayaerand Yıldırım (2003) achieved 80.21% classification accuracy using general regression neural network (GRNN) for diagnosing Pima Indian diabetes. They have also reported 77.08% classification accuracy using multilayer neural network with LM algorithm. They have used conventional (one training and one test) validation method. Carpenter and Markuzon (1998) have presented an instance counting algorithm ARTMAP-IC and obtained 81% accuracy to test set. They have used conventional (one training and one test) validation method. Deng and Kasabov (2001) obtained 78.4% classification accuracy (10x FC) using ESOM. There has been several other studies reported focusing on the same dataset with accuracy between 59.5% and 77.7%. The accuracy values of these studies can be seen in Polat et al. (2008).

2.3. Diagnosis of the Pima Indian diabetes disease using multilayer neural network

In the first stage of this study, a multilayer neural network structure which was trained by Levenberg–Marquardt (LM) algorithm was used for the pima-diabetes disease diagnosis. The network structure used for this purpose is shown in Fig. 1. This network was the multilayer network (input layer, hidden layers, and output layer). The hidden layer neurons (50 neuron for each hidden layer) and the output layer neurons use nonlinear sigmoid activation functions. In this system, eight inputs were features, and two outputs are index of two classes. Equations used in the neural network model are shown in (1)–(3).

Outputs of the first hidden layer neurons are,

$$\vec{X}^{ih1}(n) = 1 \Big/ \left(1 + \exp\left(W^{ih1}(n) * \vec{f}(n) + \vec{b}^{ih1}(n)\right)\right) \quad (1)$$

Outputs of the second hidden layer neurons are,

$$\vec{X}^{ih2}(n) = 1 \Big/ \left(1 + \exp\left(W^{ih2}(n) * \vec{X}^{ih1}(n) + \vec{b}^{ih2}(n)\right)\right) \quad (2)$$

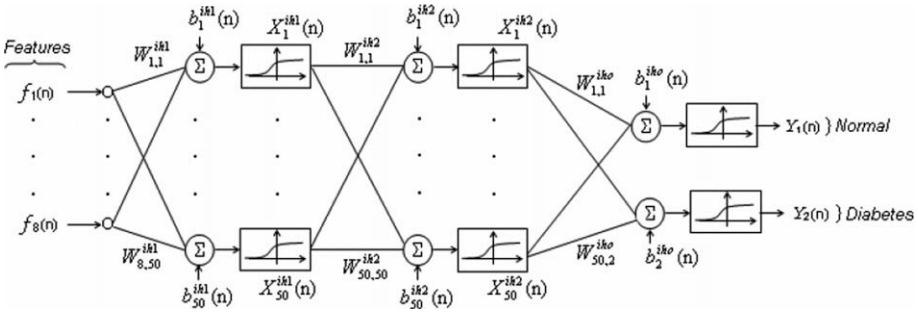

Fig. 1. Implementation of multilayer neural network for the pima-diabetes disease diagnosis.

Outputs of the network are,

$$\vec{Y}(n) = 1 \Big/ \left(1 + \exp\left(W^{ho}(n) * \vec{X}^{ih2}(n) + \vec{b}^{ho}(n)\right)\right) \tag{3}$$

where $W^{ih1}(n)$ are the weights from the input to the first hidden layer and $\vec{b}^{ih1}(n)$ are the biases of the first hidden layer, $W^{ih2}(n)$ are the weights from the first hidden layer to the second hidden layer and $\vec{b}^{ih2}(n)$ are the biases of the second hidden layer, $W^{ho}(n)$ the weights from the second hidden layer to the output layer and $\vec{b}^{ho}(n)$ are the biases of the output layer, $\vec{f}(n)$ values the features, $\vec{Y}(n)$ values the outputs for the class index, and n a training pattern index.

Levenberg–Marquardt algorithm (Hagan & Menhaj, 1994) was used in this study for the training of the multilayer neural network structure. Detailed computational issues about the application of the training algorithm to MLNN can be found in Gulbagand Temurtas (2006), Matlab documentation (2004).

2.4. Diagnosis of the Pima Indian diabetes disease using probabilistic neural network

At the second stage of this study, a probabilistic neural network was used for the pima-diabetes disease diagnosis. The network structure used for this purpose is shown in Fig. 2.

The PNN structure used in this study has a multilayer structures consisting of a single hidden layer (radial basis layer) of locally tuned units which are fully interconnected to an output layer (competitive layer) of two units, as shown in Fig. 2. In this system, real valued input vector is feature's vector, and two outputs are index of two classes. All hidden units simultaneously receive the eight-dimensional real valued input vector. The input vector to the network is passed to the hidden layer nodes via unit connection weights. The hidden layer consists of a set of radial basis functions. Associated with jth hidden unit is a parameter vector, called $\vec{c}_j$ a center. The hidden layer node calculates the Euclidean distance between the center and the network input vector and then passes the result to the radial basis function. All the radial basis functions are the same type (Gaussian). Equations which used in the neural network model are shown in (4)–(8).

$$X_j = \phi(\|\vec{f} - \vec{c}_j\| * b^{ih}) \tag{4}$$

$$\phi(x) = \exp(-x^2) \tag{5}$$

$$b^{ih} = 0.833/s \tag{6}$$

$$S_i = \sum_{j=1}^{h} W_{ji}^{ho} * X_j \tag{7}$$

$$Y_i = \begin{cases} 1, & \text{if } S_i \text{ is } \max of\{S_1, S_2\} \\ 0, & \text{else} \end{cases} \tag{8}$$

where $i = 1,2$, $j = 1,2,\ldots,h$, Y_i is the ith output (classification index), $\vec{f}$ the eight-dimensional real valued input vector, W_{ji}^{ho} the weight between the jth hidden node and the ith output node, $\vec{c}_j$ the center vector of the jth hidden node, s the real constant known as spread factor, b^{ih} the biasing term of radial basis layer, and $\phi(\cdot)$ the nonlinear radial basis function (Gaussian).

The PNN structures employed in the study utilized the newpnn function implemented in MATLAB. Detailed information about the realisation of the PNN structures can be found in the neural network toolbox part of MATLAB Documentation (Matlab documentation, 2004).

2.5. Measures for performance evaluation

2.5.1. Classification accuracy

Classification accuracy (Watkins, 2001) has been used for the comparison of studies reported in literature focusing on pima-diabetes disease diagnosis and using same database (Carpenter &

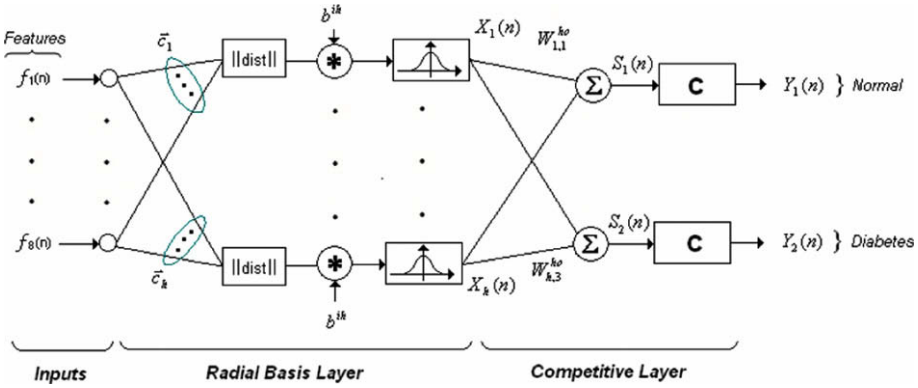

Fig. 2. Implementation of probabilistic neural network for the pima-diabetes disease diagnosis.

Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008). So, the results obtained in this study were compared with the results of the pervious reported studies by using classification accuracy.

Equations which used in the classification accuracies are shown in the following equations:

$$\textit{classification accuracy}(N) = \frac{\sum_{i=1}^{|N|} \textit{assess}(n_i)}{|N|}, \quad n_i \in N \tag{9}$$

$$\textit{assess}(n) = \begin{cases} 1 & \text{if } \textit{classify}(n) = nc \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where N is the set of data items to be classified (the test set), $n \in N$, nc is the class of the item n, and $classify(n)$ returns the classification of n by neural networks.

2.5.2. Validation of the estimation results

In this study, the conventional (one training and one test) validation and 10-fold cross-validation techniques were performed to compare the accuracy of the neural network models. For the conventional validation method, the first 576 cases were used as the training set and the remaining 192 cases were used as the test set (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003). In k-fold cross-validation (Delen et al., 2005; Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996; Polat & Gunes, 2007; Polat et al., 2008), whole data are randomly divided to k mutually exclusive and approximately equal size subsets. The classification algorithm trained and tested k times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus, k different test results exist for each training-test configuration (Delen et al., 2005). The average of these results gives the test accuracy of the algorithm.

3. Results

The classification accuracies obtained by this and best values of other studies for pima-diabetes disease dataset were presented in Table 1.

From Table 1, it can be seen easily that, 10-fold cross-validation and conventional validation approaches give a bit different classification results for the same methods used in this study. So, the comparison between methods used by this study and other studies can be made with respect to validation methods.

Table 1
Neural networks classification accuracies for Pima Indian diabetes disease diagnostic problem with classification accuracies obtained by other studies.

Study	Method	Classification accuracy (%)
Deng and Kasabov (2001)	ESOM (10x FC)	78.40
Polat and Gunes (2007)	PCA–ANFIS (10x FC)	89.47 (not reproducible)
Polat et al. (2008)	LS-SVM (10x FC)	78.21
	GDA–LS-SVM (10x FC)	79.16
This study	MLNN with LM (10x FC)	79.62
	PNN (10x FC)	78.05
Kayaer and Yıldırım (2003)	GRNN (conventional valid)	80.21
	MLNN with LM (conventional valid)	77.08
Carpenter and Markuzon (1998)	ARTMAP-IC (conventional valid)	81.00
This study	MLNN with LM (conventional Valid)	82.37
	PNN (conventional valid)	78.13
Other studies reported. Detailed list can be accessible in Polat and Gunes (2007)	Various methods (3x FC, 10x FC, conventional valid)	Between 59.5 and 77.7

As seen in this table pervious diagnosis result of the MLNN with LM algorithm reported on Pima Indian diabetes disease dataset (Kayaer & Yıldırım, 2003) is very far from the result of the MLNN with LM algorithm obtained by this study. Kayaer and Yıldırım, 2003 reported 77.08% classification accuracy using MLNN with LM algorithm. But we obtained 82.37% classification accuracy using the same method. This can be because of that, LM algorithm converges very fast but it can cause the memorization effect when the overtraining occurs. So, to prevent the memorization effect, we checked the accuracy values of the test during the training process as seen in Fig. 3 and we determined the optimum trained neural network using the maximum accuracy value of the test data. From this figure, it can be seen easily that, the MLNN with LM converged very fast and started to memorize the training set after ninth epoch. That is why; after ninth epoch, its generalization started to decrease and its performance wasn't improved for untrained test sets even its performance was decreased. So, a wrong trained MLNN with LM can cause worse accuracy results as in Kayaer and Yıldırım (2003). Additionally, the classification accuracy of MLNN with LM obtained by this study using correct training was better than those obtained by other studies for the conventional validation method as seen in Table 1.

Polat and Gunes (2007) have reported 89.47% classification accuracy (10x FC) on Pima Indian diabetes disease diagnosis using PCA and ANFIS as seen in Table 1. It was very promising with regard to the other classification applications in literature for this problem. On the other hand, the result for the same method must be reproducible. On the other hand, we obtained 66.78% classification accuracy (10x FC) on Pima Indian diabetes disease diagnosis using PCA and ANFIS. It is very far from 89.47% classification accuracy.

PCA is one of the most general purpose feature extraction methods (Chen & Chang, 1995). It aims to describe the data variance by constructing a set of new orthogonal features, called principal components (PCs). The PCs are a linear combination of the data variables and are mutually orthogonal. Every new PC describes a part of the data variance not explained by the previous ones. Due to this fact, usually a few first PCs are enough to represent well the data variance (Daszykowski, Kaczmarek, Vander, & Walczak, 2007). One drawback of typical PCA methods is that they are least squares estimation techniques and hence fail to account for outliers and atypical values which are common in realistic training and test sets (Torre & Black, 2001). A single bad outlier may cause that

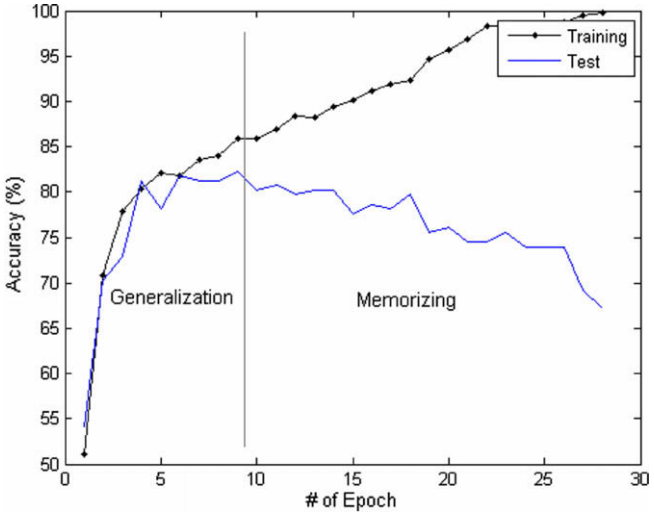

Fig. 3. Training and test performances of the MLNN with LM for the pima-diabetes disease diagnosis.

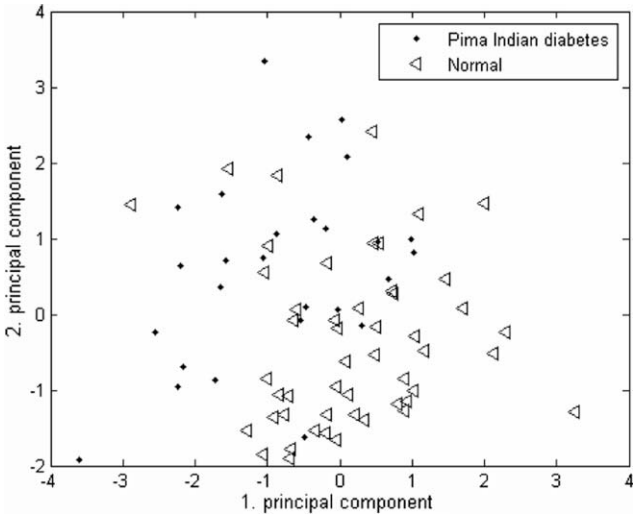

Fig. 4. The data distribution of a sample test set which used in 10-fold cross-validation techniques using the first two principal components.

principal components are distorted so as to fit the outlier well, leading to bad interpretation of the results. Outliers can also cause the so-called masking effect: due to their presence, the model is distorted in such a way that based on the principal components, no outliers can be detected. As soon as data are contaminated, classical PCA breaks down i.e. performance of PCA is seriously influenced by outliers and atypical values (Serneels & Verdonck, 2008; Campbell, 1980).

On the other hand, there a lot of outliers and atypical values in Pima Indian diabetes disease dataset. To show outliers and atypical values, we prepared Fig. 4. This figure illustrates the data distribution of a sample test set which used in 10-fold cross-validation techniques using the first two principal components. The outliers and atypical values can be easily seen from this figure.

ANFIS uses a hybrid learning algorithm to identify the membership function parameters of Sugeno type (Takagi & Sugeno, 1985) fuzzy inference system (FIS). A combination of least squares and back propagation methods are used for training FIS membership function parameters to model a given set of input/output data (Jang, Sun, & Mizutani, 1997). Application of ANFIS is become difficult when the numbers of inputs increase a bit (Gulbag, 2006). This is because of that the rule number increase very much. So, using of PCA is very suitable Pima Indian diabetes disease dataset because it decreases the numbers of inputs. However, improvement of the results is not expected because of that a lot of outliers and atypical values in Pima Indian diabetes disease data set. In fact, our result (66.78% classification accuracy) supports this idea.

On the other hand, higher classification accuracy values can be obtained if class indexes are used at PCA calculations mistakenly. We also obtained approximately 88.5% classification accuracy (10x FC) on Pima Indian diabetes disease diagnosis using PCA and ANFIS with including class indexes. This result is very close to the result (89.47% classification accuracy) obtained by Polat and Gunes(2007). But this type training and testing can be toughed as biased methods because the class indexes are given to classifier indirectly.

Furthermore, there are small differences for the application of the 10-fold validation at the studies presented by Polat and Gunes (2007) and Polat et al. (2008). They have reported 78 test data for one of the folds in their papers. But, they had to use 77 test data (768 samples/10-fold $\cong$ 77). This means that, they have used one extra common data in their test and training process. This common data increases real classification accuracy results as 1–2%. In addition to this, Polat et al. (2008) have reported 79.16% classification accuracy in their result table but they have put 82.05% classification accuracy in their comparison table. We illustrated their reported result in Table 1.

For the 10-fold cross-validation method, the classification accuracy of MLNN with LM obtained by this study using correct training was a bit better than those obtained by other studies except the classification accuracies by Polat and Gunes (2007) which is not reproducible as seen in Table 1.

In this study, we obtained similar classification accuracy values using PNN structures for both 10-fold cross-validation and conventional validation approaches. These results are quite good for Pima Indian diabetes disease diagnostic problem in comparison with the results obtained by the other studies especially for 10-fold cross-validation method as seen in Table 1. The performance of the PNN structure is very close to the best results of the literature for 10-fold cross-validation method as seen in the same table.

4. Conclusions

This paper presents a comparative study on Pima Indian diabetes disease diagnostic by using multilayer neural network which was trained by LM algorithm and probabilistic neural network. The results were also compared with the results of the pervious studies reported (Carpenter & Markuzon, 1998; Deng & Kasabov, 2001; Kayaer & Yıldırım, 2003; Polat & Gunes, 2007; Polat et al., 2008) focusing on pima-diabetes disease diagnosis and using the same database.

As the conclusion, the following results can be summarised:

- It was seen that neural network structures could be successfully used to help diagnosis of pima-diabetes disease.
- The classification accuracy of MLNN with LM obtained by this study using correct training was better than those obtained by other studies for the conventional validation method.
- For the 10-fold cross-validation method, the classification accuracy of MLNN with LM obtained by this study using correct training was a bit better than those obtained by other studies except the classification accuracies by Polat and Gunes (2007) which is not reproducible.
- The results obtained using PNN structures are also quite good for Pima Indian diabetes disease diagnostic problem in comparison with the results obtained by the other studies especially for 10-fold cross-validation method.

References

Acharya, U. R., Tan, P. H., Subramaniam, T., et al. (2008). Automated identification of diabetic type 2 subjects with and without neuropathy using wavelet transform on pedobarograph. *Journal of Medical Systems, 32*(1), 21–29.

Baier, L. J., & Hanson, R. L. (2004). Genetic studies of the etiology of type 2 diabetes in Pima Indians. *Diabetes, 53*, 1181–1186.

Brent, R. P. (1991). Fast training algorithms for multi-layer neural nets. *IEEE Transactions on Neural Networks, 2*, 346–354.

Campbell, N. A. (1980). Robust procedures in multivariate analysis 1: Robust covariance estimation. *Applied Statistics, 29*, 231–237.

Carpenter, G. A., & Markuzon, N. (1998). ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases. *Neural Networks, 11*, 323–336.

Chen, L. H., & Chang, S. (1995). An adaptive learning algorithm for principal component analysis. *IEEE Transactions on Neural Networks, 6*(5), 1255–1263.

Daszykowski, M., Kaczmarek, K., Vander, H. Y., & Walczak, B. (2007). Robust statistics in data analysis – A review basic concepts. *Chemometrics and Intelligent Laboratory Systems, 85*, 203–219.

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine Artificial Intelligence in Medicine, 34*(2), 113–127.

Deng, D., & Kasabov, N. (2001). On-line pattern analysis by evolving self-organizing maps. In *Proceedings of the fifth biannual conference on artificial neural networks and expert systems (ANNES)* (pp. 46–51).

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. G. R. (1996). *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI Press/ The MIT Press.

Gori, M., & Tesi, A. (1992). On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*, 76–85.

Gulbag, A. (2006). *Dagaalty work* ** . Ph.D. thesis, Sakarya University, Institute of Science & Technology.

Gulbag, A., & Temurtas, F. (2006). A study on quantitative classification of binary gas mixture using neural networks and adaptive neuro fuzzy inference systems. *Sensors and Actuators B, 115*, 252–262.

Gulbag, A., Temurtas, F., & Yusubov, I. (2007). Quantitative discrimination of the binary gas mixtures using a combinational structure of the probabilistic and multilayer neural networks. *Sensors and Actuators B: Chemical, 131*, 196–204. doi:10.1016/j.snb.2007.11.008.

Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. PWS Publishing.

Hagan, M. T., & Menhaj, M. (1994). Training feed forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks, 5*, 989–993.

Hoshi, K., Kawakami, J., Kumagai, M., Kasahara, S., Nisimura, N., Nakamura, H., et al. (2005). An analysis of thyroid function diagnosis using Bayesian-type and SOM-type neural networks. *Chemical & Pharmaceutical Bulletin, 53*, 1570–1574.

Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). *Neuro_Fuzzy and soft computing* (p. 07458). Upper Sllade River, NJ: Prentice Hall.

Kayaer, K., & Yıldırım, T. (2003). Medical diagnosis on Pima Indian diabetes using general regression neural networks. In *Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP)* (pp. 181–184).

Knowler, W. C., Pettitt, D. J., Saad, M. F., Charles, M. A., Nelson, R. G., Howard, B. V., et al. (1991). Obesity in the Pima Indians: Its magnitude and relationship with diabetes. *American Journal of Clinical Nutrition, 53*, 1543–1551.

Matlab® documentation. (2004). Version 7.0, release 14. The MathWorks Inc.

Mohamed, E. I., Linderm, R., Perriello, G., Di Daniele, N., Poppl, S. J., & De Lorenzo, A. (2002). Predicting type 2 diabetes using an electronic nose-base artificial neural network analysis. *Diabetes Nutrition & Metabolism, 15*(4), 215–221.

Polat, K., & Gunes, S. (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing, 17*(4), 702–710.

Polat, K., Gunes, S., & Aslan, A. (2008). A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. *Expert Systems with Applications, 34*(1), 214–221.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Serneels, S., & Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Computational Statistics & Data Analysis, 52*, 1712–1727.

Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks, 3*, 109–118.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 116–132.

Temurtas, F. (2009). A comparative study on thyroid disease diagnosis using neural networks. *Expert Systems with Applications, 36*, 944–949.

Torre, F. D., & Black, M. J. (2001). Robust principal component analysis for computer vision. In *Proceedings of the eighth international conference on computer vision (ICCV'01)* (Vol. 1, pp. 362–369).

Watkins, A. (2001). *AIRS: A resource limited artificial immune classifier*. Master thesis, Mississippi State University.