

# Combination of Neural Networks and Fuzzy Clustering Algorithm to Evaluation Training Simulation-Based Training

Lida Pourjafar

Department of Computer, Ahvaz  
Branch, Islamic Azad University,  
Ahvaz, Iran

Department of Computer, Khou  
Zestan Science and Research Branch,  
Islamic Azad University, Ahvaz Iran

Mehdi Sadeghzadeh

Department of Computer  
Engineering, Mahshahr Branch  
Islamic Azad University, Mahshahr,  
Iran

Department of Computer,  
Khouzestan Science and Research  
Branch,  
Islamic Azad University, Ahvaz, Iran

Marjan Abdeyazdan

Department of Computer Science,  
College of Electricity and Computer,  
Mahshahr Branch Islamic Azad  
University, Iran

Department of Computer,  
Khouzestan Science and Research  
Branch,  
Islamic Azad University, Ahvaz, Iran

**Abstract**—With the advancement of computer technology, computer simulation in the field of education are more realistic and more effective. The definition of simulation is to create a virtual environment that accurately and real experiences to improve the individual. So Simulation Based Training is the ability to improve, replace, create or manage a real experience and training in a virtual mode. Simulation Based Training also provides large amounts of information to learn, so use data mining techniques to process information in the case of education can be very useful. So here we used data mining to examine the impact of simulation-based training. The database created in cooperation with relevant institutions, including 17 features. To study the effect of selected features, LDA method and Pearson's correlation coefficient was used along with genetic algorithm. Then we use fuzzy clustering to produce fuzzy system and improved it using Neural Networks. The results showed that the proposed method with reduced dimensions have 3% better than other methods.

**Keywords**—Educational Data Mining; Simulation-Based Training; Dimensions Reduction; ANFIS

## I. INTRODUCTION

Data mining is a kind of computer-based information system (CBIS), which can be used for big data warehouse, peer review, production information, and knowledge discovery. The traditional term of mining is affected the foundations of data mining. But instead of searching for minerals, here discover the knowledge. The purpose of data mining is to identify data patterns, hidden links with organized information, communication rules are structured, the unknowns are estimated to be classified topics, create homogeneous clusters of issues and a wide variety of findings that do not come easily obtained by classical CBIS be uncovered. By the way, the results of data mining are invaluable the basis for decision-making.

Education, a new basin for the use of data mining to discover knowledge, decision-making and provide recommendations. The use of data mining in education is early

stages and created the field of "educational data mining". The first decade of this century marks the beginning of educational data mining.

Educational data mining can be an example for the design, assignments, methods and algorithms for data discovery learning environments. The purpose of data mining is to find patterns and make predictions of the behavior and development of trained people, content knowledge application environment, assessments, training and application functions to define. With the emergence development of computer technology, simulation systems closer to reality and is one of the most important and has become efficient tool in education. As mentioned, a virtual environment that simulates real conditions creates absolutely arises [1-2].

## II. LITERATURE REVIEW

Tian et al in an article work on the evaluation of simulation-based training for pilots. In this article, they have used cubic learn and krikpatrick model. In fact, to review and evaluate of results is used of simplified krikpatrick model. The results show that simulation-based training to 26% better than the education based on booklet [3].

Pamela et al to evaluate simulation-based training for teaching assistants midwife at the birth of babies. Simulation-based training was conducted on 111 persons and 14 persons were trained as usual. This research was conducted at medical centers in Ghana. The analysis was performed using 4 surfaces krikpatrick model. The results showed that better results are Simulation-based training [4].

Natassia et al in an article work on driver training base on simulation. In this paper, they do the impact of simulation training for drivers of vehicles [5].

Sophia et al in an article work on eye surgery Simulation-based training. This article is used a review of five different database. In this paper, expressed validity of the model and learned the ability to transfer to the operating room has been measured [6].

Shu-Hsien et al [7] Recent advances in data mining involves the collection of the works of the last decade have offered.

Collection approaches began Narli, Ozgan and Alcan [8], the theory of multiple intelligences unaccounted for identifying the relationship between people and their learning styles used. By using multiple intelligences the data collected to teach the learning style scales and tests for prospective teachers.

Gonzalez et al [9], the hidden conditional random field applied to complete the reading assignment predict. They are classified as self-paced dialogue as a matter of classification to classify sequences considered dialogue to assess.

Barbell et al [10], they seek to solve some issues: How can learning processes using process models and control rule based, the optimization? How can process models created based on the concept of learning styles? Therefore, they are a method for modeling a student using a combination of learning styles and approaches proposed mining process, and it gave way to model pupil.

Terry, Pardos, Sarkozy and Heffernan [11], a clustering strategy with common building envelope created to predict the results of students' self-learning function.

Yu et al [12] features expanded with redundancy and discrete optimization techniques. Series of features were learned by logistic regression L1. Then, features were dense to help statistical techniques and random trees. Finally, the results were combined together with adjusted linear regression.

Approaches complete set with Levy and Wilensky [13] began in the behavior of query students studied in complex models, while their goal is to create an equation linking the physical parameters of the system. They looked at how students adapt their systems with different behaviors.

### III. PROPOSED METHOD

In the framework presented in Figure 1 feature selection is shown. In this paper we proposed a method based on genetic algorithms with two different fitness function based on linear discriminant analysis and Pearson coefficient. A genetic algorithm is an effective method for solving optimization problems.

Here problem space is multi-dimensional, discrete and complex. Genetic algorithm represents each chromosome is a binary vector and each number represents a feature. If  $I$  show the collection features so that it is composed of  $N$  member, in this case a subset  $X$  of  $Y$  ( $X \subseteq Y$ ) represents the  $i$ -th chromosome  $I$  with  $N$  gene so that it is equal to one if the  $i$ -th feature was selected and the otherwise is zero (Figure 1).

In addition to, simply coding solutions genetic algorithm for solving such problems is very appropriate because the search space of exponentially in a very difficult, complex and non-linear services. algorithm starts with a  $P$  random populations. The fitness of each chromosome created using appropriate fitness function is calculated as described.

After calculating the fitness, 80% chromosomes are

selected using the roulette wheel. Combine the two point method is carried out. Then mutations in chromosomes or individuals carried a distinct possibility.

#### A. Fitness function based on linear discriminant analysis

We use a fitness function with two statements, a  $J$  index called the separability index. The second term represents the number of members of set and if it is desirable to be less. Separability index  $J$  is derived from linear discriminant analysis.

In this paper, it is assumed that each feature can be modeled as a random variable. Separability index can calculate uses the covariance matrix of the features.

Since the variance of the random variable dispersion around the mean to express a certain value the covariance matrix of  $n$  variables, probability distribution around the mean vector in  $n$ -dimensional space. If we have  $n$  random variables  $\{X_1, X_2, \dots, X_n\}$ , so that each variable is  $m$  samples (dimensions  $m \times n$  stored in the matrix  $D$ ), the covariance matrix  $\Sigma$ , an  $n \times n$  is the matrix of the elements in row  $i$  and column  $j$  indicates the covariance between the variables  $x_i$  and  $y_j$  is the (equation 1):

$$\begin{aligned} \Sigma[i, j] &= Cov(x_i, y_j) \\ Cov(x_i, y_j) &= \frac{1}{m} \sum_{l=1}^m (D(l, i) - \mu_i)(D(l, j) - \mu_j) \end{aligned} \quad (1)$$

Here  $\mu_i$  and  $\mu_j$  are mean's variables in matrix  $D$ .

Two classes of observations by a specified means and variances to consider. Fisher separation between the two distributions for the variance between two classes to the variance within the class definition:

$$S = \frac{\sigma_{between}^2}{\sigma_{within}^2} = \frac{(\bar{w}^T \bar{\mu}_{y=1} - \bar{w}^T \bar{\mu}_{y=0})^2}{\bar{w}^T \sum_{y=1} \bar{w} + \bar{w}^T \sum_{y=0} \bar{w}} = \frac{(\bar{w} \cdot (\bar{\mu}_{y=1} - \bar{\mu}_{y=0}))^2}{\bar{w}^T (\sum_{y=0} + \sum_{y=1}) \bar{w}} \quad (2)$$

$$S_W = w^T S_W w \quad (3)$$

$$S_B = w^T S_B w \quad (4)$$

According to the above definitions can define separability index using eq 5:

$$J(I) = tr\left(\frac{W^T \sum_B W}{W^T \sum_W W}\right) \quad (5)$$

Here,  $W$  is the transition matrix is defined as follows:

Available features four-dimensional space, so we assume  $N = 4$ . If the chromosome  $I = (0,1,0,1)$  is, in this case on chromosome 2 and 4 are traits  $I$  so:

$$W = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (6)$$

$tr(\cdot)$  is a trace matrix. High levels of separability index  $J(I)$  show that the subsidiary created by chromosomes  $I$  have been

at the center of the well separation On the basis of the proposed fitness function and genetic algorithms can be defined relationship 7:

$$F(I) = \frac{J(I)}{N} + K \frac{N - N_I}{N} \quad (7)$$

Here, N number of available features,  $N_I$  number of features on chromosomes  $I$  and K constant to determine the effect of the second sentence in eq 7.

### B. Pearson correlation-based fitness function

A feature largely is correlated with response variable and with other features in sub-features is at a lower correlation. Correlation is a measure of the strength of the relationship between the two variables in bi-directional.

$$Fitness_s = \frac{N\bar{r}_{cf}}{\sqrt{N + N(N-1)\bar{r}_{ff}}} \quad (8)$$

S subset of N feature,  $\bar{r}_{cf}$  is that the average correlation feature-class and  $\bar{r}_{ff}$  is average correlation features - features. Eq 8 is Pearson correlation relationship in which all variables have been homogeneous.

### C. Neural Networks

Back propagation method is the iterative process that runs on a set of training samples and each of the outputs obtained by comparing the output target and this process continues until a specified condition stems. Target values can be labels for training (for classification issues) and continuous values (numerical calculations). For all samples by repeating the correction process is weights modified to minimize the mean square error between the outputs of the network and target values.

This modification is done in the weights in the reverse direction, so that the beginning of the last layer (output layer) started in the hidden layer continues to be the first hidden layer, hence it is called back propagation. There is also no guarantee the convergence of weight. The algorithm are shown in Table 1.

### D. Fuzzy Inference System

At first, a fuzzy inference system based on the existing database defined using Takagi-Sugeno-Kang. Using FCM, fuzzy set of rules for modeling the behavior of the database is extracted. The number of fuzzy inputs used here is the number of features of the database and the number of outputs equal to the number of classes (clusters). Since the four clusters (scores), so we considered 4 levels of output. To define the initial fuzzy system of fuzzy membership functions are used. Figure 2 shows an example of fuzzy rules that used for clustering with 8 features in 4 clusters.

### E. Neuro-Fuzzy Systems

Figure 3 shows the structure of neuro-fuzzy system for 8 features. As shown any features connected to four membership functions and the fuzzy rules have been used here. Combined method is used for training.

### F. Datasets Used

Preparing the database in an educational institution in Mahshahr shooting took place. A total of 200 patients were used for this database. At first tab contains personal information was prepared and then people were trained based simulation. The following characteristics were obtained from each individual 17 are introduced.

The first feature points was shot from a distance of 175 meters. This feature was extracted from the training and expertise in the institute's rate.

The second feature shooting accuracy in the training. Shooting accuracy of the diversity of positions x and y on Targets for each beam i is obtained. Since each person was 6 shooting at this stage of the relationship 9 was used to calculate accuracy:

$$\alpha = \frac{\sum_{i=1, j=i+1}^6 \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{6} \quad (9)$$

Smaller numbers indicate greater accuracy in high regard. The third feature indicates the shooting accuracy in the training. Shooting accuracy of the calculated difference between each beam to the center of target position is calculated from equation 10.

$$\alpha = \frac{\sum_{i=1}^6 \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{6} \quad (10)$$

$x_i$  and  $Y_i$  the center of target location is here considered to be the origin.  $x_i$  and  $y_i$  the location of each shot is on target.

The fourth feature is in testing phase accuracy that can be obtained from the above equation.

The fifth feature of the profile of the individual units had been achieved.

The sixth feature of confidence in the firing position is that professionals in the Institute's rate.

The seventh Feature is grade in school was that of profiles of individuals.

The eighth Feature is Confidence in eight shooting rule. The ninth feature of confidence in the usefulness of the use of guns.

The tenth features is the accuracy of the test step. The eleventh feature is the body mass index was calculated from height and weight. Relationship between body mass and BMI of 11 is as follows.

$$BMI = \frac{mass_{kg}}{height_m^2} \quad (11)$$

The twelfth degree of confidence about the performance characteristics shooting.

The thirteenth features is the left eye and 14th feature is the right eye sight.

The 15<sup>th</sup> Feature is the location of weapons.

The 16<sup>th</sup> feature is the months of military service status. The seventeenth feature is belief in focus.

#### IV. EVALUATION OF THE PROPOSED METHOD

For this study we used a computer with certain properties, Including:

Processor: Intel Pentium(R) CPU G620, 2.60 GHz  
2.60GHz

Installed memory (RAM): 4.00 GB

For modeling and simulation software program MATLAB version R2014a (8.3.0) 64-bit was used.

##### A. Neural Networks Standard

For non-reducing features the network includes 17 input that used the same database features. The number of neurons in the hidden layer to all states is constant and equal to 20. Since each class is shown with binary code so the output layer of the Neural Networks includes four neuron.

As described in previous, two different methods have been used to reduce the dimensions by using a genetic algorithm Table 2 shows characteristics of each method with the code number. The purpose of coding here, references in the text to the new method is better.

Table 3 shows recognition rates for the standard Neural Networks for database without reducing the size of the feature. As shown in table recognition rate for data in LDA1 and LDA2 is 94.5%.

##### B. Fuzzy Inference System

Table 3 shows recognition rates for fuzzy inference system for database no decrease in size. Here confusion matrix for all the data is shown. Here is a detection rate of 94.5% for the entire data. Here's detection rate is slightly less than the Neural Networks. Recognition rate for fuzzy inference system for databases with size reduction for LDA2 is better than other. Here confusion matrix for all the data is shown. Here 95% detection rate for all data.

##### C. neuro-fuzzy system

Figure 4 confusion matrix for neuro-fuzzy systems for database shows no decrease in size. Here confusion matrix for all the data is shown. Here is a detection rate of 94.5% for the entire data. Detection rate here is like Neural Networks.

Figure 5 confusion matrix for neuro-fuzzy system for databases with size reduction method shows LDA3. Here confusion only for all data matrix is shown. Here 95% detection rate for all data. Such as Neural Networks and fuzzy detection rate here is better than PC.

In the case of the Neural Networks with 13 features we have the highest detection rate is 94.5 and the amount is lower than the other two methods. While this happened to fuzzy inference systems also feature 13 to achieve the highest detection rate. But the combination of neuro-fuzzy system achieved the highest rate of diagnosis for nine features with

this mode with less computing power to achieve better accuracy.

#### V. CONCLUSION

Here a genetic algorithm with binary encoded with two fitness function includes linear discriminant analysis and Pearson coefficient was used. Here the results of three simulations in various environments, including neural networks, fuzzy systems based on fuzzy clustering and neuro-fuzzy were compared. The highest detection rates in the neural networks, fuzzy systems and neuro-fuzzy inference system was shown. As has been stated that the Neural Networks is used in a state where the number is 13 features we have the highest detection rate is 94.5 and the amount is lower than the other two methods. While this happened to fuzzy inference systems also feature 13 is required to achieve the highest detection rates. But the combination of neuro-fuzzy system achieved the highest rate of diagnosis for 9 Features can achieve better accuracy.

#### REFERENCES

- [1] Bell, B.S. & Kozlowski, S.W.J. (2002). Adaptive guidance: enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, 55, 267-307.
- [2] Bell, B.S., Kanar, A.M. & Kozlowski, S.W.J. (2008). Current issues and future directions in simulation-based training in North America. *International Journal of Human Resource Management*, 19(8), 1416-1434.
- [3] Yongliang Tian, Hu Liu, Jiao Yin, Mingqiang Luo, Guanghui Wu, (2015), Evaluation of simulation-based training for aircraft carrier marshalling with learning cubic and Kirkpatrick's models, *Chinese Journal of Aeronautics*, Volume 28, Issue 1, 152-163
- [4] Pamela Andreatta, Florence Gans-Larty, Domitilla Debpuur, Anthony Ofosu, Joseph Perosky, (2011), Evaluation of simulation-based training on the ability of birth attendants to correctly perform bimanual compression as obstetric first aid, *International Journal of Nursing Studies*, 48, 10, 1275-1280
- [5] Natassia Goode, Paul M. Salmon, Michael G. Lenné, (2013), Simulation-based driver and vehicle crew training: Applications, efficacy and future directions, *Applied Ergonomics*, 44, 3, 435-444
- [6] Ann Sofia S. Thomsen, Yousif Subhi, Jens Folke Kiilgaard, Morten la Cour, Lars Konge, (2015), Update on Simulation-Based Surgical Training and Assessment in Ophthalmology : A Systematic Review, *Ophthalmology*, Available online.
- [7] Shu-Hsien, L., Pei-Hui, C., & Pei-Yuan, H. (2012). Data mining techniques and applications – a decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303-11311.
- [8] Narli, S., Ozgen, K., & Alkan, H. (2011). In the context of multiple intelligences theory, intelligent data analysis of learning styles was based on rough set theory. *Learning and Individual Differences*, 21(5), 613-618.
- [9] Gonzalez-Brenes, J. P., Duan, W., & Mostow, J. (2011). How to classify tutorial dialogue? comparing feature vectors vs. Sequences. In *Proceedings of the 4th international conference on educational data mining*, 169-178
- [10] Holzhtuter, M., Frosch-Wilke, D., & Klein, U. (2012). Exploiting learner models using data mining for e-learning: a rule based approach. In A. Pena-Ayala (Ed.), *Intelligent and adaptive educational- learning systems: achievements and trends, smart innovation, systems and technologies*, 77-105.
- [11] Trivedi, S., Pardos, Z. A., Sarkozy, G. N., & Heffernan, N. T. (2012). Co-clustering by bipartite spectral graph partitioning for out-of-tutor prediction. In *Proceedings of the 5th international conference on educational data mining*, 33-40.

- [12] Yu, H. F., Lo, H. Y., Hsieh, H. P., Lou, J. K., G.McKenzie, T., Chou, J.W., Chung, P. H., Ho, C. H., Chang, C. F., Wei, Y. H., Weng, J. Y., Yan, E. S., Chang, C. W., Kuo, T. T., Lo, Y. C., Chang, P. T., Po, C., Wang, C. Y., Huang, Y. H., Hung, C.W., Ruan, Y. X., Lin, Y. S., Lin, S. Lin, H. T., & Lin, C. J. (2010). Feature engineering and classifier ensemble for KDD cup 2010. In Proceedings of the KDD 2010 cup 2010 workshop knowledge discovery in educational data, 1–12.
- [13] Levy, S. T., & Wilensky, U. (2011). Mining students' inquiry actions for understanding of complex systems. *Computers & Education*, 56(3), 556–573.
- [14] Jiawei Han, (2012), *Data Mining: Concepts and Techniques*, third edition, elsevire, USA, 398

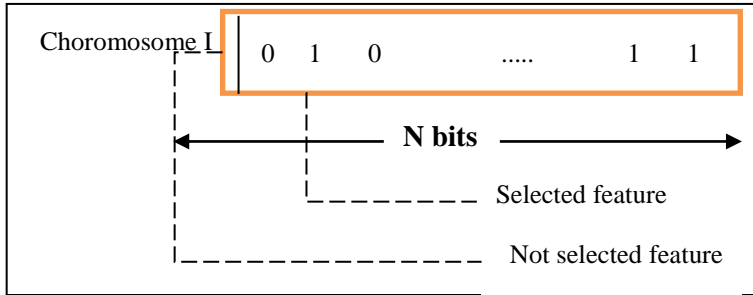


Fig. 1. Is an example of coding set by a series of two-bit Features

TABLE I. BACK PROPAGATION ERROR METHOD BASED [14]

<p><b>Input:</b></p> <ul style="list-style-type: none"><li>■ <math>D</math>, a data set consisting of the training tuples and their associated target values;</li><li>■ <math>l</math>, the learning rate;</li><li>■ <math>network</math>, a multilayer feed-forward network.</li></ul> <p><b>Output:</b> A trained neural network.</p> <p><b>Method:</b></p> <ol style="list-style-type: none"><li>(1) Initialize all weights and biases in <math>network</math>;</li><li>(2) <b>while</b> terminating condition is not satisfied {</li><li>(3)     <b>for</b> each training tuple <math>X</math> in <math>D</math> {</li><li>(4)         // Propagate the inputs forward:</li><li>(5)         <b>for</b> each input layer unit <math>j</math> {</li><li>(6)             <math>O_j = I_j</math>; // output of an input unit is its actual input value</li><li>(7)         <b>for</b> each hidden or output layer unit <math>j</math> {</li><li>(8)             <math>I_j = \sum_i w_{ij} O_i + \theta_j</math>; // compute the net input of unit <math>j</math> with respect to the previous layer, <math>i</math></li><li>(9)             <math>O_j = \frac{1}{1+e^{-I_j}}</math>; } // compute the output of each unit <math>j</math></li><li>(10)         // Backpropagate the errors:</li><li>(11)         <b>for</b> each unit <math>j</math> in the output layer</li><li>(12)             <math>Err_j = O_j(1 - O_j)(T_j - O_j)</math>; // compute the error</li><li>(13)         <b>for</b> each unit <math>j</math> in the hidden layers, from the last to the first hidden layer</li><li>(14)             <math>Err_j = O_j(1 - O_j) \sum_k Err_k w_{jk}</math>; // compute the error with respect to the next higher layer, <math>k</math></li><li>(15)         <b>for</b> each weight <math>w_{ij}</math> in <math>network</math> {</li><li>(16)             <math>\Delta w_{ij} = (l)Err_j O_i</math>; // weight increment</li><li>(17)             <math>w_{ij} = w_{ij} + \Delta w_{ij}</math>; } // weight update</li><li>(18)         <b>for</b> each bias <math>\theta_j</math> in <math>network</math> {</li><li>(19)             <math>\Delta \theta_j = (l)Err_j</math>; // bias increment</li><li>(20)             <math>\theta_j = \theta_j + \Delta \theta_j</math>; } // bias update</li><li>(21)         } }</li></ol>
--

1. If (in1 is in1cluster1) and (in2 is in2cluster1) and (in3 is in3cluster1) and (in4 is in4cluster1) and (in5 is in5cluster1) and (in6 is in6cluster1) and (in7 is in7cluster1) and (in8 is in8cluster1) then (out1 is out1cluster1) (1)
2. If (in1 is in1cluster2) and (in2 is in2cluster2) and (in3 is in3cluster2) and (in4 is in4cluster2) and (in5 is in5cluster2) and (in6 is in6cluster2) and (in7 is in7cluster2) and (in8 is in8cluster2) then (out1 is out1cluster2) (1)
3. If (in1 is in1cluster3) and (in2 is in2cluster3) and (in3 is in3cluster3) and (in4 is in4cluster3) and (in5 is in5cluster3) and (in6 is in6cluster3) and (in7 is in7cluster3) and (in8 is in8cluster3) then (out1 is out1cluster3) (1)
4. If (in1 is in1cluster4) and (in2 is in2cluster4) and (in3 is in3cluster4) and (in4 is in4cluster4) and (in5 is in5cluster4) and (in6 is in6cluster4) and (in7 is in7cluster4) and (in8 is in8cluster4) then (out1 is out1cluster4) (1)

Fig. 2. Proposed four Fuzzy rules

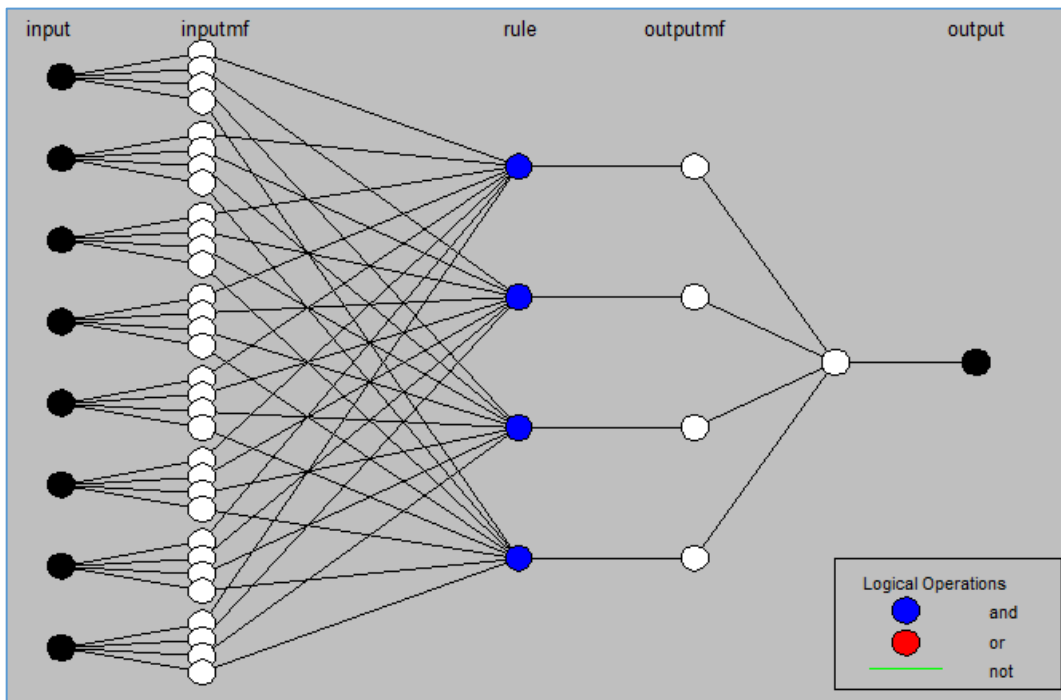


Fig. 3. Neuro-fuzzy inference system for 8 Features

TABLE II. NUMBER OF PROPERTIES WITH EACH PROCEDURE CODE

Dimension reduction method	Pierson coefficient	LDA K=0.8	LDA K=4	LDA K=8	LDA K=10
Number of feature	5	15	13	9	8
Method code	PC	LDA1	LDA2	LDA3	LDA4

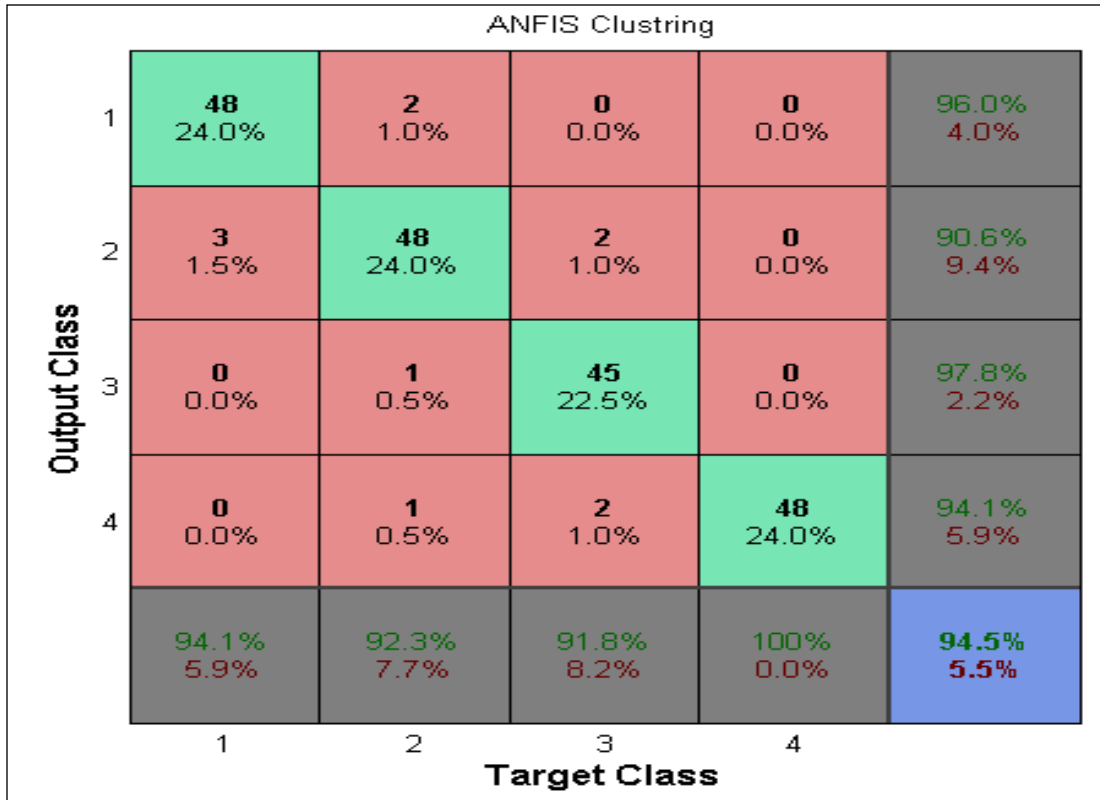


Fig. 4. Shows the confusion matrix for neuro-fuzzy system with no loss of features

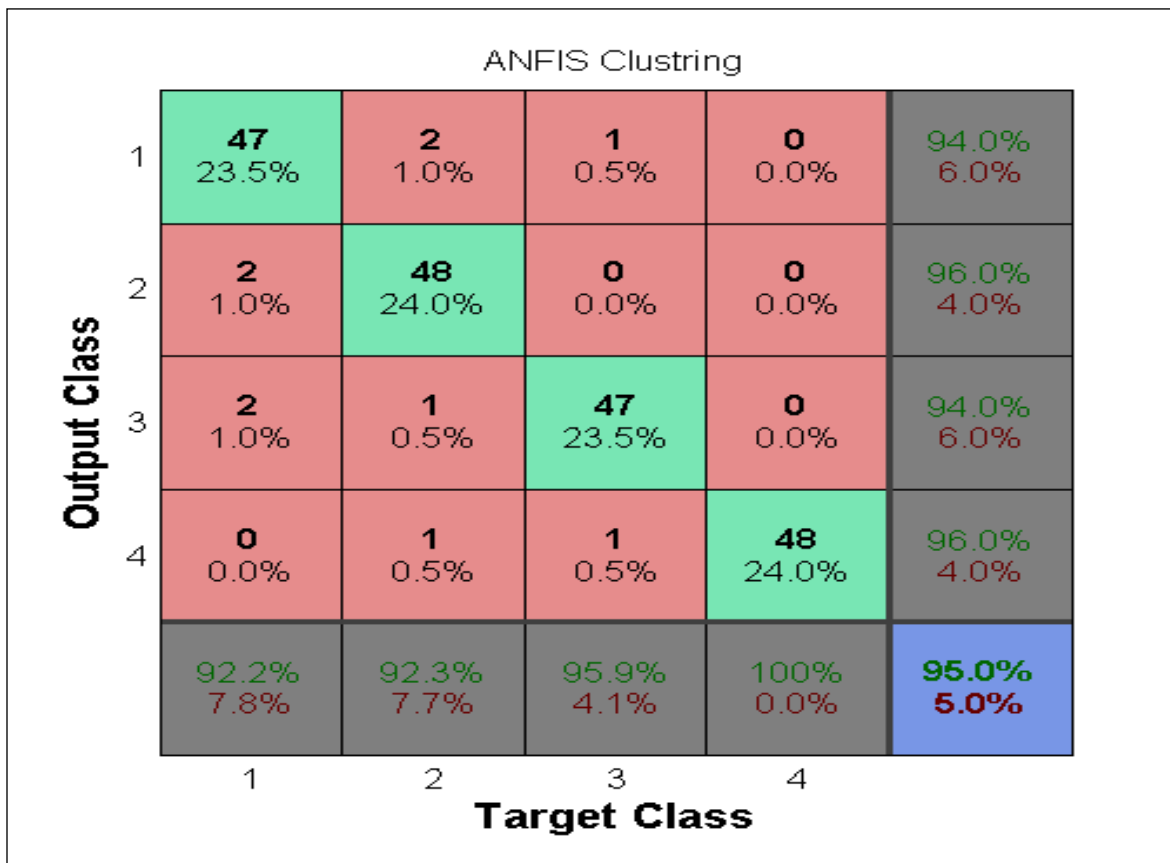


Fig. 5. Confusion matrix for neuro-fuzzy system with reduced features using LDA3

TABLE III. RECOGNITION RATE IN ALL CLASSIFIERS

Method	Dimension reduction	Recognition rate
Neural networks	17 Features	93.50
	PC	78.50
	LDA1	94.50
	LDA2	94.50
	LDA3	93.00
	LDA4	92.00
Fuzzy Inference System	17 Features	94.50
	PC	74.00
	LDA1	95.00
	LDA2	95.00
	LDA3	94.50
	LDA4	94.50
Neuro fuzzy	17 Features	94.50
	PC	83.00
	LDA1	94.50
	LDA2	94.00
	LDA3	95.00
	LDA4	94.50