

## Research Article

### A Novel Approach for the Diagnosis of Diabetes and Liver Cancer using ANFIS and Improved KNN

<sup>1</sup>C. Kalaiselvi and <sup>2</sup>G.M. Nasira

<sup>1</sup>Department of Computer Applications, Tiruppur Kumaran College for Women, Karpagam University,

<sup>2</sup>Department of Computer Applications, Chickanna Govt Arts College for Men, Tirupur, Tamilnadu, India

**Abstract:** The multi-factorial, chronic, severe diseases are cancer and diabetes. As a result of abnormal level of glucose in body leads to heart attack, kidney disease, renal failure and cancer. Many studies have been proved that several types of cancer are possible in diabetes patients having a high blood sugar. Many approaches are proposed in the past to diagnose both cancer and diabetes. Even though the existing approaches are efficient one, the classification accuracy is poor. An Enhanced approach is proposed to achieve a higher efficiency and lower complexity. Adaptive neuro fuzzy inference system is used to classify the dataset with the help of adaptive group based KNN. The Pima Indian diabetes dataset are used as input dataset and classified based on the attribute information. The experimental result shows the classification accuracy is better than the existing approaches such FLANN, ANN with FUZZYKNN.

**Keywords:** Artificial neural network, K-nearest-neighbour, liver cancer, neuro fuzzy

## INTRODUCTION

When the production of insulin is affected which control the glucose level of the body leads to a disease called diabetes mellitus. The diabetes mellitus is the one of the main reason for nerve damage, kidney disease and also cancer. Many researches are going on this area and some of the statistics report shows that most of the people will die because of the high glucose level lead to either cancer or heart diseases. Diabetes mellitus cannot be healed fully but it can be controlled by using medicines such as insulin, food items. The diagnosis of diabetes is important classification problem (Vigneri *et al.*, 2009). Some types of cancer such as pancreatic cancer, liver, endometrium cancer are commonly seen in diabetes patients.

Many researches have been conducted to find the relationship between the cancer and the diabetes mellitus recently in diabetes patients. The diabetes is a group of metabolic disorders and the reinterpretation is required, so the possibilities of cancer in diabetes patients are high (Vigneri *et al.*, 2009). The metabolic and hormonal disorders are affecting the diabetes patients in a different way. Some of the risk factors of the cancer and diabetes are physical activity, diet, alcohol, smoking and obesity. The biological link mechanisms between cancer and diabetes are hyperinsulinemia, hyperglycemia and inflammation. The certain diabetic patients are affected by cancer because of the exposure of the treatment of diabetes.

Many new techniques are used for diagnosing both cancer and diabetes. The diagnosis using machine learning technique is one of the existing techniques which have a transparent diagnostic knowledge. Machine learning is classified into connectionist learning and symbolic learning. The user can easily understand the rules of symbolic learning techniques and it is considered as a comprehensible technique. The best example for the symbolic technique is rule induction which is extensively used for medical diagnosis (Richards *et al.*, 2000; Lisboa *et al.*, 2000). Artificial neural network is an incomprehensible technique and best example for connectionist learning techniques. The connections and information are hidden from the user in this technique. Artificial neural networks have been applied in medical field for various tasks (Kononenko, 2001; Andrews *et al.*, 1995).

Artificial neural network has been applied to various processes such as pattern recognition and data classification in the medical field. It attracts the many researchers and it becomes multi-objective solution to the various problems (Asada *et al.*, 1990). ANN working process is based on the neurons in human brain. It is used to find the connection between the given data as input from the user and output data based on the information merged from large number of cases (Grenier *et al.*, 1994). The computer aided diagnosis with Bayesian is used for diagnosing lung diseases. The chest radiography, chest CT, clinical data values are the main parameters for diagnosing the lung disease. Even

**Corresponding Author:** C. Kalaiselvi, Department of Computer Applications, Tiruppur Kumaran College for Women, Karpagam University, Tirupur, Tamilnadu, India

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

though computer aided diagnosis is consider as a supplementary technique for the diagnosis, performance of CAD on radiologists are not evaluated.

In recent days K-Nearest-Neighbour (KNN) technique is usually used in data mining for pattern recognition and classification problems (Moreno-Seco *et al.*, 2003). This study examines the diagnosis of cancer and diabetes on the benchmark dataset using ANFIS by training using adaptive group-based kNN. The proposed approach is compared with the other data mining techniques utilize on the same dataset. It also examines if ANFIS integrates with AgKNN can increase its accuracy in the diagnosis of cancer and diabetes in patients. The rest study is divided into related works, proposed approach and performance evaluation.

### LITERATURE REVIEW

**Diabetes:** The method for diagnosing a diabetes using back propagation neural network algorithm is proposed in this study (Siti and Dannawaty, 2005). The blood pressure, glucose concentration in blood, serum insulin, BMI, number of times a person pregnant and age are the main parameters for the diagnosing diabetes. The main drawback of this framework is missing values of the input. Jayalakshmi and Santhakumaran (2010) improved this framework to overcome the drawback of this system. The input dataset is reformed using the missing values and it improves the framework by enhancing the classification precision. Data preprocessing technique also proposed in this study which is used to improve the speed of the framework. Dey *et al.* (2008) proposed a technique using a backpropagation neural network along with binary classification. The inputs to this framework are blood sugar test, post plasma, age. The percentage of performance of this framework is 92.5% when it is compared with previous framework. The prediction of glucose level for future based on the current level is using artificial neural network. Ann is trained using the parameters to predict the glucose level.

The Pima Indian diabetes dataset is used for research work. The frameworks with 22 different classifiers are used for the accurate classification using input as Pima Indian diabetes dataset. The performance evaluation of framework based on k means and knn is compared with existing framework and accuracy range of framework is 66.6 to 77.7% (Michie *et al.*, 1994). The model using hybrid k-means and decision tree with the help of support vector machine, GDA, reaching the classification accuracy of 82.05%. The author achieved different classification accuracy for framework with various combinations of ANN, DT\_ANN and cascaded GA\_CFS\_ANN (Humar and Novruz, 2008; Patil *et al.*, 2010; Karegowda *et al.*, 2011).

The insulin usage prediction can be done by using neuro fuzzy systems with the help invasive blood tests (Dazzi *et al.*, 2001). The neuro fuzzy based on the back propagation network and fuzzy logic and it is used to

find the connection between variables and rules. The neuro fuzzy system is trained using 1000 BG values and tested with 400 BG values were used to build the nomogram. When it is compared with the conventional control systems neuro-fuzzy system shows the better results in insulin variation and maintains constant body glucose level.

**Cancer:** The nuclear imaging method attracts the some of the researchers in the medical field. Most of the researches based on the information provided by patients along with images and remaining researches based on the images only. Computer aided design is used for the classification of affected cells from normal one and it used to predict the cancer. The combinations of detection and segmentation task are used in the tumor localization problem. Even though the standard segmentation methods are used in this problem, some drawbacks are there such as lack of sensitivity and specificity. Some of the reports stated that the image and data are mainly used for feasibility studies. They are (Ying *et al.*, 2004) proposed a method which gives better visual performance but failed to present quantitative results when it is compared with the other methods. Even though Guan *et al.* (2006) gives good sensitivity and specificity compared with the other methods, CT/PET images and 2 D scale image are used to false detection. In Saradhi *et al.* (2009) this study comparing the previous image processing methods with the supervised classification schemes.

### METHODOLOGY

**Proposed system for diagnosis of diabetes and cancer:** The proposed approach for diagnosing both cancer and diabetes using ANFIS and Adaptive group based KNN.

**ANFIS architecture:** To enhance the learning and adaptations of the adaptive system neuro fuzzy inference system is used in this model. The first order fuzzy inference system based on if then rules is used in ANFIS architecture (Karlık *et al.*, 2003):

**Rule 1:** If ( $x$  is  $A_1$ ) and ( $y$  is  $B_1$ )

Then ( $f_1 = p_1x + q_1y + r_1$ )

**Rule 2:** If ( $x$  is  $A_2$ ) and ( $y$  is  $B_2$ )

Then ( $f_2 = p_2x + q_2y + r_2$ )

where,  $x$  and  $y$  are the inputs,  $A_i$  and  $B_i$  are the fuzzy sets,  $f_i$  are the outputs within the fuzzy region specified by the fuzzy rule,  $p_i, q_i$  and  $r_i$  are the design parameters that are determined during the training process. The ANFIS architecture to implement these two rules is shown in Fig. 1, in which a circle indicates a fixed node.

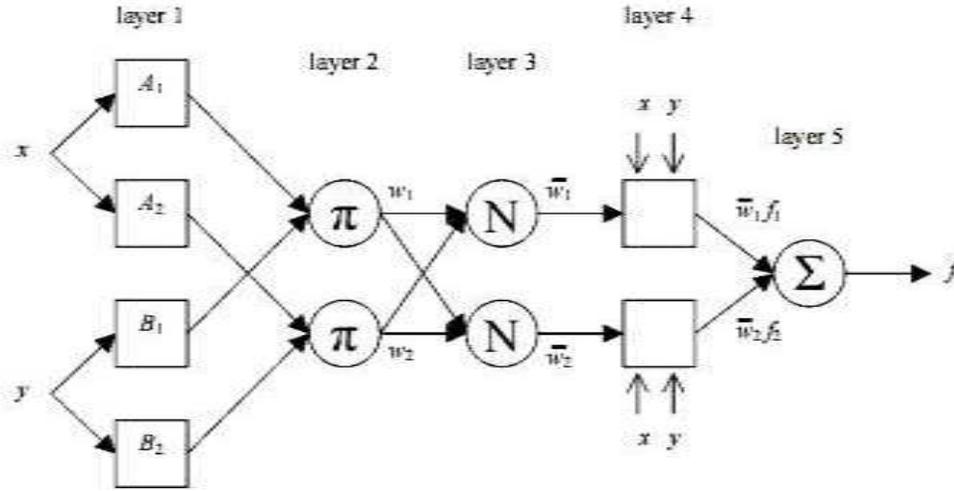


Fig. 1: ANFIS architecture

In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grade of the inputs, which are given by:

$$O_i^1 = \mu A_i(x), i = 1,2$$

$$O_i^1 = \mu B_{i-2}(y), i = 3,4$$

where  $\mu A_i(x)$ ,  $\mu B_{i-2}(y)$  can adopt any fuzzy membership function. For example, if the bell shaped membership function is employed,  $\mu A_i(x)$  is given by:

$$\mu A_i(x) = \frac{1}{1 + \left\{ \left( \frac{x - c_i}{a_i} \right)^2 \right\}^{b_i}}$$

where,  $A_i$ ,  $b_i$  and  $c_i$  are the parameters of the membership function, governing the bell shaped functions accordingly.

In the second layer, the nodes are fixed nodes. They are labeled with M, indicating that they perform as a simple multiplier.

The outputs of this layer can be represented as:

$$O_i^2 = w_i = \mu A_i(x) \mu B_i(y) \quad i = 1,2$$

which are the so-called firing strengths of the rules.

The above diagram depicts the typical adaptive neuro fuzzy inference system with 5 layers. In each layer circle depicts fixed node, square shows that adaptive node and each layer is used for different purpose. To obtain the preferred performance number, type, parameter and rules of fuzzy membership functions are used and it is selected based on the trial and error method. Although the better performances are achieved, these parameters are hard to use in some situations. To overcome this problem ANFIS is trained to get optimal premise and consequent parameters.

In the third layer, the nodes are also fixed nodes. They are labeled with N, indicating that they play a normalization role to the firing strengths from the previous layer.

The outputs of this layer can be represented as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1,2$$

which are the so-called normalized firing strengths.

In the fourth layer, the nodes are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and a first order polynomial (for a first order Sugeno model). Thus, the outputs of this layer are given by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad i = 1,2$$

In the fifth layer, there is only one single fixed node labeled with S. This node performs the summation of all incoming signals. Hence, the overall output of the model is given by.

It can be observed that there are two adaptive layers in this ANFIS architecture, namely the first layer and the fourth layer. In the first layer, there are three modifiable parameters  $\{a_i, b_i, c_i\}$ , which are related to the input membership functions. These parameters are the so-called premise parameters. In the fourth layer, there are also three modifiable parameters  $\{p_i, q_i, r_i\}$ , pertaining to the first order polynomial. These parameters are so-called consequent parameters.

**ANFIS classifier:** Adaptive neuro fuzzy inference system is the combination of fuzzy inference system and learning power of artificial neural network. It mainly aims to incorporate the best features of the fuzzy systems and neural network. The algorithms such as gradient descent and back propagation are used to

train the artificial neural network systems by regulating the membership functions and weights of the defuzzification.

The GUI parameters which is used to regulating the fuzzy inference system are mentioned below:

- Fuzzy Inference System (FIS) Editor (FIS editor handles issues such as the numbers of input and output variables; their names etc.)
- Membership Function Editor (Membership function editor defines the shapes of all the membership functions associated with each variable)

- Rule Editor (Rule editor is used for editing the list of rules that define the behavior of the system)
- Rule Viewer (Rule viewer is used for looking at the Fuzzy Inference System (FIS) to help diagnose the behavior of specific rules or study the effect of changing input variables)

The adaptive neuro fuzzy inference system process is explained. In ANFIS assume the model structure is used to relate the membership function, rules input and output in cyclic process. The collection of input and output data which is used to train the fuzzy inference model by regulating the membership function based on

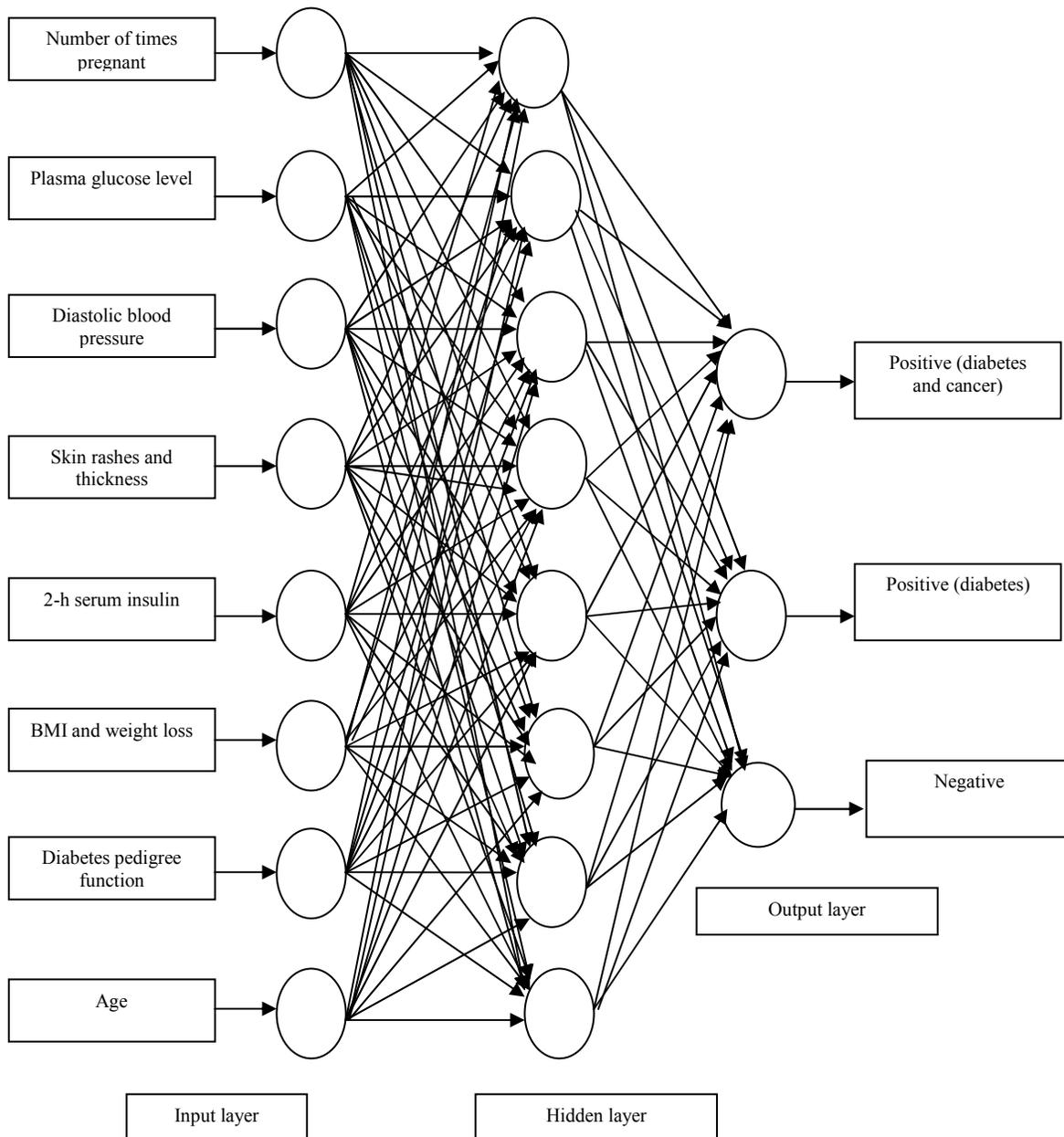


Fig. 2: Diagnosis of cancer and diabetes using ANFIS with KNN

the trial and error method. The parameters are selected based on the error criteria.

**Training ANFIS model:** ANFIS can be trained to learn from given data. As we can observe from the ANFIS architecture, in order to configure an ANFIS model for a specific problem, we need to specify the fuzzy rules and the activation functions (i.e., membership functions) of fuzzification neurons. The specific domain antecedent fuzzy sets can be specified for designing the fuzzy rules; while for the consequents of the fuzzy rules, the parameters (e.g.) are formed and adjusted by algorithm in the training process. Adaptive Group-Based kNN is used as training algorithm in ANFIS model (Fig. 2).

**Adaptive group-based KNN:** Traditional KNN have two main problems. They are experience dependent and sample category balance (Yuhas *et al.*, 1992). The empirical value should be preset and results are erroneous sometimes because of the large gap or distance between the documents or data.

To overcome the above mentioned problem an adaptive group based KNN is proposed. The training set or sample parameters are grouped into the multiple groups. The Adaptive Group-Based KNN algorithm is used for recording the set of observations or training set. i.e., examples of the k inputs and the desired outputs that should be computed in network and then the error (difference between actual and expected results) is calculated. The data are classified using ANFIS simultaneously in each group with random value of k and compare the results. The k value is changed according to the compared results. If the result is reliable then group size and k value are unchanged. If the result is similar emerge the group and reduce the k value. If the results are different then increase the k value. The groups should increase when the two opposing are offset.

$N_g(i)$ -Number of training group during  $i^{th}$  data processed.  $C_j(i)$ -Categorization result of  $i^{th}$  document by  $j^{th}$  group. The best example for the categorization calculation for groups is defined as diagnosing the liver cancer {number of times pregnant, plasma glucose level, blood pressure, skin rashes and age} as {1, 2, 3, 4, 5}.  $C_3(5)$  means third parameter is considered, i.e., blood pressures is going to be discussed.  $\bar{C}$  is the average value of different categories calculated by feature distance in groups. Sample data as:

$$\{\text{Number of times pregnant, plasma glucose level, blood pressure, skin rashes and age}\} \bar{C} = \frac{(3+2+3+1+1)}{5} = 2$$

Adaptive training group is determined as:

$$N_g(i+1) = \begin{cases} N_g + int \left( \frac{1}{N_g} \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 \right), \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 > \bar{C} \\ N_g \frac{1}{C} < \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 \leq \bar{C} \\ N_g - int \left( \frac{1}{\bar{C}} \right) \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 \leq \frac{1}{\bar{C}} \end{cases}$$

The samples data's are suggested to the variance of different groups by adjusting the grouping situation. If the variance of the grouping data is higher than the threshold then the categorization results are inaccurate. The reason for inaccurate results is more groups are required for final decision. If the variance is low means then the sample groups are merged without any disputes in classification results. Threshold value can be calculated as a by using lower and higher bound ( $1/C$  and  $C$ ).

The value of  $k$  can be calculated adaptively as:

$$k = \begin{cases} \gamma_i & \frac{1}{N_g} \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 < 1 \\ \gamma_i + int \left( \frac{1}{n} \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 \right), \frac{1}{N_g} \sum_{j=1}^{N_g} (C_j(i) - \bar{C})^2 \geq 1 \end{cases}$$

$\gamma_i$ -Random initial value of  $k$ . The random value can be tested by the system to check whether it is suitable or not. To obtain the exact categorization results the value of  $k$  should be adjusted.

The  $k$  value can be tested by the system to check whether it is suitable for group or not and can be set by algorithm with the help of training set. The groups are adjusted based on the variance of categorization results by different groups in real time (Chun-Hong and Wei, 2009; Coomans and Massart, 1982). Runtime complexity for  $n$  elements are  $3n - 1$  and computational complexity is calculated as a:

$$T = N_g \cdot k \cdot (3N_g - 1)$$

Training of network i.e., error correction is stopped when the value of the  $k$  has become sufficiently small and as desired in the required limits (Duch *et al.*, 2000). Total error for  $p^{th}$  observation of data set and  $j^{th}$  neuron in the output layer can be computed as:

$$E_k = t_k - y_k$$

where,  $t_k$  represents the desired target output,  $y_k$  represents the predicted from the system and  $E_k$  error correction

The problems of  $k$  nearest neighbor is reduced in adaptive group based KNN. When the AGKNN is compared with the traditional KNN the proposed algorithm shows the higher efficiency and robustness by solving the experience dependent problem and the

algorithm shows the accurate results by solving category balance.

The training data set used to train the neural network which is implemented in the client server architecture. The neural network contains 8 nodes in input layer based on the input attributes. The pima Indian diabetes dataset are used as input and applying the adaptive group based k nearest neighbor algorithm for training the network. The training set or sample parameters are grouped into the multiple groups. The data are classified simultaneously in each group with random value of k and compare the results. The same algorithm is applied to same network without propagation of errors and obtaining the efficient results. The proposed algorithm i.e., combination of ANFIS and AGKNN are compared with the previous methods and it outperforms the existing method in classification accuracy. The proposed algorithm accuracy is calculated by using 10 fold CV in weka classifier. The calculated accuracy is compared with the existing approaches such as FLANN, ANN with FUZZYKNN. In training data set the unwanted values or less significant values are removed by preprocessing the data to obtain the classification accuracy more than the existing methods.

## DATASET DESCRIPTION AND RESULTS

In our study we have used Pima Indian Diabetes data sets (Ngoc and Edward, year) for training and testing the neural network model.

### Attribute information:

- Number of times pregnant
- Plasma glucose level
- Blood pressure
- Skin rashes and thickness
- 2-h serum insulin
- BMI and weight loss
- Diabetes pedigree
- Age

The 8 input parameters are Number of times pregnant, Plasma glucose level, Diastolic blood pressure, Triceps skin fold thickness, 2-H serum insulin, body mass index, Diabetes pedigree function and Age. The output parameter name is class. Class value positive is interpreted as “tested positive for diabetes and cancer” or else “tested positive for diabetes” and class value negative is represented as “tested negative for diabetes and cancer”.

All the input parameters had numeric values. First parameter is total number of times the patient was pregnant. Second parameter is the value of oral glucose tolerance test which is used to find the amount of glucose level in the blood. Third parameter is the

diastolic blood pressure value which is measured in millimeter by Hilo gram. Fourth parameter is triceps skin fold thickness which is a measure in millimeter. Fifth parameter is 2-h serum insulin test values. It is conducted to find the amount of insulin creation in the patient body. Sixth parameter is the patents body mass index. It is calculated by the following formula:

$$\begin{aligned} \text{Body mass Index} \\ &= \text{Patient weight in kg} \\ &/ (\text{patient height in meter})^2 \end{aligned}$$

Seventh parameter the relationship function value of diabetes family hierarchy and eighth parameter is age of the person.

**Experimental results:** ANFIS is relatively fast to convergence due to its learning strategy and its easy interpretation. It is a more transparent model and its behavior can be explained in human understandable terms, such as linguistic terms and linguistic rules.

The ANFIS classifiers were trained with the Adaptive Group-Based KNN method in combination with the least squares method when attributes such as Number of times pregnant, Plasma glucose level, Diastolic blood pressure, Triceps skin fold thickness, 2-h serum insulin, body mass index, representing the patient details were used as inputs.

The fuzzy rule architecture of the ANFIS classifiers were designed by using a generalized bell shaped membership function. Each ANFIS classifier was implemented by using the MATLAB software package (MATLAB version 7.0 with fuzzy logic toolbox). The data sets were divided into two separate data sets-the training data set and the testing data set. The adequate functioning of the ANFIS depends on the sizes of the training set and test set.

The training data set was used to train the ANFIS model, whereas the testing data set was used to verify the accuracy and the effectiveness of the trained ANFIS model for classification.

The test performance of the classifiers can be determined by the computation of total classification accuracy and RMSE. The graphical representation for root mean square error is shown in the Fig. 3. Sensitivity vs. Specificity comparison:

TP : Number of True positives  
 FP : Number of False positives  
 TN: Number of True Negatives  
 FN: Number of False Negative

True Positive Rate (TPR) (also called sensitivity, hit rate and recall):

$$\text{sensitivity} = \frac{TP}{TP+FN}$$

False Positive Rate (FPR) (also called false alarm rate):

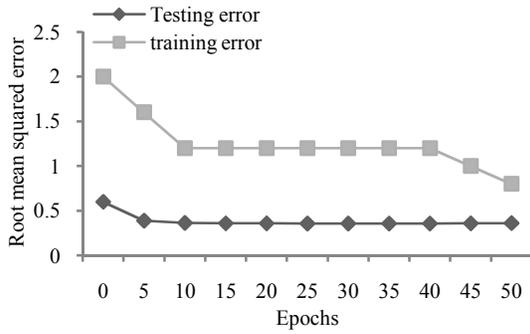


Fig. 3: The training and testing error of the developed ANFIS classifier

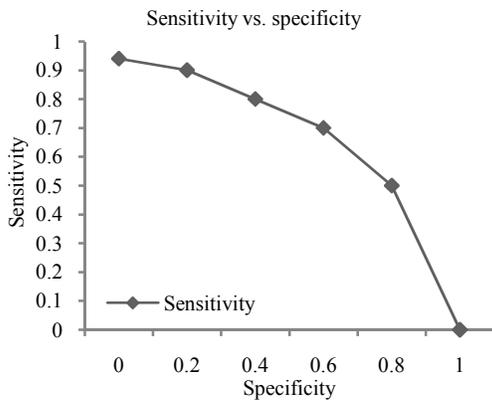


Fig. 4: Sensitivity vs. specificity graph for proposed ANFIS-AGKNN classifier

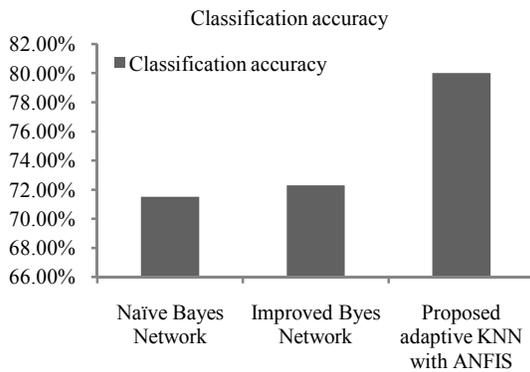


Fig. 5: Total classification accuracy of proposed ANFIS classifier

$$FalseAlarm = \frac{FP}{TN+FP}$$

$$specificity = \frac{TN}{TN+FP} = 1 - FalseAlarm$$

The graphical representation of the sensitivity and specificity comparison is shown in the Fig. 4. The large areas under ROC curves point out that the proposed ANFIS classifier has high specificity and sensitivity. As

Table 1: Performance of pima Indian diabetes model

Method	Accuracy (%)
Naïve bayes network	71.5
Improved byes network	72.3
Proposed adaptive KNN with ANFIS	80.0

a conclusion, these results show that the ANFIS based AGKNN approach provide the highest TCC accuracy in the detection of Cancer and Diabetes.

The total classification accuracy is defined as ratio of number of correct decisions and total number of cases (Table 1):

$$Accuracy = (TP + TN)/n$$

The graphical representation of the sensitivity and specificity comparison is shown in the Fig. 5. The performance of the proposed classifier is better than the other existing classifiers such as navie bayes network. As a conclusion, these results show that the ANFIS based AGKNN approach provide the highest TCC accuracy in the detection of Cancer and Diabetes.

### CONCLUSION

To predict both diabetes and cancer many researches has been conducted. The dataset used here is pima Indian diabetes dataset for predicting both cancer and diabetes. A novel approach by using Adaptive neuro fuzzy inference system is used and to train the neural network Adaptive group based k nearest neighbor algorithm is used. The input nodes in neural network are constructed based on the input attribute. The hidden nodes are used to classify given input based on the training dataset with the help AGKNN by grouping the training dataset. The experimental results of the proposed approach show that classification accuracy is better than the existing approaches. The proposed model has lower complexity, achieves higher efficiency and performs the pattern classification well than the traditional methods.

### REFERENCES

Andrews, R., J. Diederich and A.B. Tickle, 1995. Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowl-Based Syst., 8(6): 373-389.

Asada, N., K. Doi, H. MacMahon, S.M. Montner, M.L. Giger, C. Abe and Y. Wu, 1990. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: Pilot study. Radiology, 177: 857-860.

Chun-Hong, Z. and X. Wei, 2009. The approach to text automatic classification technology of characteristic databases on the SVM-KNN. Inform. Sci., Vol. 11.

- Coomans, D. and D.L. Massart, 1982. Alternative k-nearest neighbour rules in supervised pattern recognition. Part 2. Probabilistic classification on the basis of the kNN method modified for direct density estimation. *Anal. Chim. Acta*, 138: 153-165.
- Dazzi, D., F. Taddei, A. Gavarini, E. Uggeri, R. Negro and A. Pezzarossa, 2001. The control of blood glucose in the critical diabetic patient: A neuro-fuzzy method. *J. Diabetes Complicat.*, 15(2): 80-87.
- Dey, R., V. Bajpai, G. Gandhi and B. Dey, 2008. Application of artificial neural network technique for diagnosing diabetes mellitus. *Proceeding of 10 Colloquium and the 3rd IEEE Region International Conference on Industrial and Information Systems (ICIIS, 2008)*, pp: 1-4.
- Duch, W., R. Adamczak, K. Grabczewski, G. Zal and Y. Hayashi, 2000. Fuzzy and Crisp Logical Rule Extraction Methods in Application to Medical Data. In: Szczepaniak, P.S., P.J.G. Lisboa and J. Kacprzyk (Eds.), *Fuzzy System in Medicine*. Physica-Verlag, Springer, Berlin, Heidelberg, New York, pp: 593-616.
- Grenier, P., S. Chevret, C. Beigelman, M.W. Brauner, C. Chastang and D. Valeyre, 1994. Chronic diffuse infiltrative lung disease: Determination of the diagnostic value of clinical data, chest radiography, CT and Bayesian analysis. *Radiology*, 191: 383-390.
- Guan, H., T. Kubota, X. Huang, X.S.S. Zhou and M. Turk, 2006. Automatic hot spot detection and segmentation in whole-body FDG-PET images. *Proceeding of IEEE International Conference on Image Processing*, pp: 85-88.
- Humar, K. and A. Novruz, 2008. Design of a hybrid system for the diabetes and heart diseases. *Expert Syst. Appl.*, 35: 82-89.
- Jayalakshmi, T. and A. Santhakumaran, 2010. A novel classification method for classification of diabetes mellitus using artificial neural networks. *Proceeding of 2010 International Conference on Data Storage and Data Engineering*.
- Karegowda, A.G., A.S. Manjunath and M.A. Jayaram, 2011. Application of genetic algorithm optimized neural network connection weights for medical diagnosis of PIMA Indians diabetes. *Int. J. Soft Comput.*, 2(2).
- Karlık, B., M.O. Tokhi and M. Alci, 2003. A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. *IEEE T. Bio-Med. Eng.*, 50(11): 1255-1261.
- Kononenko, I., 2001. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.*, 23(1): 89-109.
- Lisboa, J.G., E.C. Ifeachor and P.S. Szczepaniak, Eds. 2000. *Artificial Neural Networks in Biomedicine*. Springer-Verlag, London, UK.
- Michie, D., D.J. Spiegelhalter and C.C. Taylor, 1994. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, New York.
- Moreno-Seco, F., L. Mico and J.A. Oncina, 2003. Modification of the LAESA algorithm for approximated k-NN classification. *Pattern Recogn. Lett.*, 24: 47-53.
- Ngoc, T.N. and S. Edward, year. UCI Machine Learning Repository and Archive. Retrieved form: [ics.uci.edu/ml/datasets.html](http://ics.uci.edu/ml/datasets.html).
- Patil, B.M., R.C. Joshi and D. Tosniwal, 2010. Hybrid prediction model for type-2 diabetic patients. *Expert Syst. Appl.*, 37: 8102-8108.
- Richards, G., V.J. Rayward-Smith, P.H. Sönksen, S. Carey and C. Weng, 2000. Data mining for indicators of early mortality in a database of clinical records. *Artif. Intell. Med.*, 22(3): 215-231.
- Saradhi, G.V., G. Gopalakrishnan, A.S. Roy, R. Mullick, R. Manjeshwar, K. Thielemans and U. Patil, 2009. A framework for automated tumor detection in thoracic FDG PET images using texture-based features. *Proceeding of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '09)*, pp: 97-100.
- Siti, F.B.J. and M.A. Dannawaty, 2005. Diabetes mellitus forecast using artificial neural networks. *Proceeding of Asian Conference of Paramedical Research*. Kuala Lumpur, Malaysia.
- Vigneri, P., F. Frasca, L. Sciacca, G. Pandini and R. Vigneri, 2009. Diabetes and cancer. *Endocr-Relat. Cancer*, 16: 1103-1123.
- Ying, H., F. Zhou, A. Shields, O. Muzik, D. Wu and E. Heath, 2004. A novel computerized approach to enhancing lung tumor detection in whole body PET images. *Proceeding of International Conference on IEEE Eng. Med. Biol. Soc. (EMBS, 2004)*. San Francisco, CA, pp: 1589-1592.
- Yuhas, R.H., A.F.H. Goetz and J.W. Boardman, 1992. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. *Proceedings of the Summaries of the 4th Annual JPL Airborne Geoscience Workshop*, 1: 147-150.