A New Approach for Diagnosis of Diabetes and Prediction of Cancer using ANFIS

C. Kalaiselvi
Research Scholar, Karpagam University
Assistant professor, Department of Computer Applications
Tiruppur Kumaran College for women
Tirupur, India
kalaic29@gmail.com

Dr. G .M. Nasira, Ph.D.,
Assistant Professor, Department of Computer Applications
Chikanna Govt Arts College for men
Tirupur, India
nasiragm99@yahoo.com

***Abstract* - The multi factorial, chronic, severe diseases like diabetes and cancer have complex relationship. When the glucose level of the body goes to abnormal level, it will lead to Blindness, Heart disease, Kidney failure and also Cancer. Epidemiological studies have proved that several cancer types are possible in patients having diabetes. Many researchers proposed methods to diagnose diabetes and cancer. To improve the classification accuracy and to achieve better efficiency a new approach like Adaptive Neuro Fuzzy Inference System (ANFIS) is proposed. The Pima Indian diabetic dataset is used as data set for classification.**

Keywords: Data Mining, Artificial neural networks, K–nearest neighbour, K-means clustering, back propagation

I. INTRODUCTION

Data Mining is one of the most innovative areas of computer science that uses various statistical techniques, classification, and clustering and pattern recognition for problems. The methodology lies in the ability to find patterns and relationships. It is also applied in forecasting tasks in medicine. In most of the areas of medicine, data mining proved that the results obtained with other methodologies give improved accuracy and performance.

Diabetes Mellitus is a disease occurs in which the amount of sugar in the blood cannot be regulated. This metabolic disease is very common in nowadays either the body does not produce enough insulin or the body does not respond to insulin produced. According to World Health Organization (WHO) tells that 37crores of people live with diabetes worldwide and it doubles before the year 2030. Because of diabetes 48 lakhs of people were died in the year 2012. 80% of people died were belonging to lower and middle class families. In India, 5crores and above are affected by diabetes and this becomes 7crores by some years. India holds number 2 place worldwide.

Diabetes cannot be fully cured and it can be controlled with the help of insulin and controlled diet. There are three main types of diabetes mellitus. Type-1 occurs when body failure to produce insulin completely. It requires injecting insulin or wearing an insulin pump. It affects mostly children usually thin. But it may strike at any age. This type is called as “Insulin Dependent Diabetes Mellitus” or “Juvenile diabetes”. Type-2 occurs when the body cannot effectively use the insulin produced. It requires diet, exercise and blood sugar level is lowered using drugs. It occurs above age 40. Mostly 90% of people are living with type-2 diabetes. This type is called as “Non Insulin Dependent Diabetes Mellitus” or “Adult onset diabetes”. The third type occurs when pregnant woman’s receptivity to insulin. 4% of all pregnant women are affected with this type. It can be controlled with insulin and diet. But 50-70% it may affect again. This type is called as “Gestational diabetes”. High blood sugar produces symptoms like polyuria, polydipsia and polyphagia. Diabetes mellitus causes serious complications such as heart disease, stroke, blindness, kidney failure and cancer. Some cancer types such as pancreatic cancer, liver cancer, and breast cancer are common in diabetic patients. When it affects cellular structure of body and may affect legs. The diagnosis of diabetes is one of the important classification problems [1].

The association between diabetes and cancer have been investigated that diabetes mellitus is not a single disease and diabetic patients are not considered to be homogeneous cohort. Therefore, if diabetes is associated with a small increase in the risk of cancer but requires important consequences at the population level. It depends on the factors that include diabetes duration, varying levels of metabolic control factors like obesity, alcohol, smoke and drugs used for diabetes treatment may influence the association between diabetes and cancer. Hyperinsulinemia, hyperglycemia, obesity and oxidative stress may cause cancer in diabetic patients and also drugs used to treat cancer may cause diabetes. While anti-diabetic drugs have minor influence on cancer risk, but insulin is a growth factor with pre-eminent metabolic, mitogenic effects and its action in malignant cells is favoured by mechanisms acting at both the receptor and post receptor level. In addition to well known diabetogenic effect of glucocorticoids and anti androgens, an increasing number of targeted anti cancer molecules may interfere with glucose metabolism acting at different levels on the signalling substrates shared by IGF-I and insulin receptors. Both diabetes and cancer needs clinical attention and better improved studies.

II. LITERATURE REVIEW

Numbers of techniques have been proposed to diagnose diabetes and cancer. Machine learning techniques are also proposed to diagnose diabetes and cancer [2-5]. Artificial Neural Networks (ANN) was applied in medical field for various tasks [6], [7]. ANN can be applied for pattern recognition and also for data classification [8]. The working process of ANN is based

978-1-4799-2877-4/14 $31.00 © 2014 IEEE
DOI 10.1109/WCCCT.2014.66

on the neurons in human brain. Back propagation neural network is used for binary classification. K Nearest Neighbor (KNN) is also used for pattern recognition and classification problems [9]. Nowadays HBALC test is suggested to get the average sugar level in blood. This paper examines the diagnosis of cancer and diabetes using ANFIS by training ANFIS using adaptive group based KNN. Comparison between various approaches that uses the same PIMA Indian data set was also discussed. The accuracy can also be improved by combining ANFIS with Adaptive group based KNN.

III. RELATED WORK

A. Diabetes

To diagnose diabetes, back propagation neural network algorithm is proposed [10]. Blood pressure, glucose concentration in blood, serum insulin, Body Mass Index (BMI) Age and other parameters are taken to diagnose diabetes. PIMA Indian diabetes as data set. The input dataset is reformed by assuming the missing values to improve the framework. It enhances the classification process [11]. ANN is also applied using back propagation neural network along with binary classification to predict diabetes [12]. The insulin usage prediction can be done using neuro fuzzy systems with the help of invasive blood tests. When compared with conventional control systems, neuro fuzzy systems show better results in insulin variation and maintain constant glucose level of the body [13-16].

B. Cancer

The nuclear imaging methods are proposed to diagnose cancer. Computer aided design (CAD) is used for classification of affected cells from normal one to predict cancer. The combinations of detection and segmentation task are used for tumour localization problem. These methods give better visual performance and failed to present quantitative results compared with other methods. CT/PET images and 2D scale images are used for false detection. Previous image processing methods compared with supervised classification schemes [17-19].

IV. PROPOSED SYSTEM FOR DIAGNOSIS OF DIABETES AND CANCER

The proposed approach for diagnosis of both diabetes and cancer using ANFIS with adaptive group based KNN.

A. ANFIS Classification

To enhance the learning process ANFIS is used. The first order fuzzy inference system based on if then rules is used in ANFIS architecture. The rules are

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
$f_2 = p_2x + q_2y + r_2$

In fuzzy region

x and y	Inputs
A_i and B_i	Fuzzy sets
f_i	Outputs
p_i, q_i and r_i	Design parameters

ANFIS incorporates the best features of fuzzy systems and neural networks. The algorithms such as gradient descent and back propagation are used to train the artificial neural network systems. Adaptive group based KNN is used with ANFIS to improve the efficiency.

$N_g(i)$ - Number of training group during i^{th} data is processed. $C_j(i)$ - Categorization result of ith document by jth group. C is the average value of different categories calculated by feature distance in groups.

Adaptive group is determined as

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

If the variance of the grouping data is higher than the threshold then the categorization results are inaccurate. If the variance is low means then the sample groups are merged without any disputes in classification results. Threshold value can be calculated using lower and higher bound ($1/C$ and C). The problems of KNN is reduced by AGKNN. 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.

PIMA Indian diabetes dataset is used as input and adaptive group based k nearest neighbour algorithm is used to train the neural network. The training set or sample parameters are divided into multiple groups. The unwanted values or less significant values are removed by pre-processing the data. The data are classified simultaneously in each group with random value of k. The results are compared with results in groups. If the results are similar then group value and k are unchanged. If the results differ then increase the k value. The training dataset used to train the neural network contains 8 nodes in Input layer based on the input attributes. The proposed algorithm, combination of ANFIS and AGKNN are compared with previous methods and it outperforms the existing methods in classification accuracy.

TABLE I. PERFORMANCE COMPARISON

Method	Accuracy
Naïve Bayes Algorithm	71.5 %
Improved Bayes Algorithm	72.3 %
k-means Algorithm	66-77%
ANFIS with Adaptive KNN	80 %

From the above table the accuracy of the previous approaches were compared with proposed approach.

V. DATASET DESCRIPTION AND RESULTS

The Pima Indian Diabetes data set [20] is used for training and testing the neural network model. Totally this dataset contains 768 Number of Instances.

A. Attribute information

- Number of times Pregnant
- Plasma Glucose level
- Diastolic Blood pressure(mm Hg)
- Skin rashes and thickness(mm)
- 2 –Hrs Serum Insulin
- BMI
- Diabetes pedigree
- Age

All the input parameters have numeric values. The first parameter is total number of times the patient pregnant. The second parameter is plasma glucose concentration 2 hours in an oral. The third parameter is the diastolic blood pressure value which is measured in mm by Hg. The fourth parameter is triceps skin fold thickness which is measured in mm. The fifth parameter is 2-hours serum insulin test which find the amount of insulin creation in the patient body. The sixth parameter is the patient's body mass index. It can be calculated as

Body Mass Index (BMI) = Patient's weight in kg / (Patient's height in meter) ^{2} (2)

The seventh parameter is the Diabetes pedigree which is the function value based on diabetes family hierarchy. The last parameter is age. Totally the Dataset contains 768 instances. The output parameter is classified into three categories. Class value positive-1 is interpreted as "Tested positive for Diabetes only", Class value positive-2 is interpreted as "Tested positive for both Diabetes and Cancer" and class value negative is interpreted as "Tested negative for Diabetes and Cancer".

VI. CONCLUSION

To predict both diabetes and cancer many researches are conducted. ANFIS is used to train the neural network. 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 of AGKNN. The experimental results show that the classification accuracy is better than existing approaches. The proposed approach gives higher efficiency and reduces complexity. The algorithm performs well and classifies the dataset well compared to traditional methods. The proposed work reduces the cost for different medical tests and helps the patients to take precautionary measures well in advance. In future the same method can also be applied in diagnosing other diseases like liver cancer etc.

REFERENCES

[1] Paolo Vigneri, Francesco Frasca, Laura Sciacca, Giuseppe Pandini and Riccardo Vigneri "Diabetes and cancer" Endocrine-Related Cancer (2009) 16 1103–1123.

[2] D. Cosic and S. Loncaric, "Rule-based labeling of CT head image," in *Lecture Notes in Artificial Intelligence 1211*, E. Keravnou, C. Garbay, R. Baud, and J. Wyatt, Eds. Berlin: Springer-Verlag, 1999, pp.453-456.

[3] W. Duch, R. Adamczak, K. Grabczewski, G. Zal, and Y. Hayashi, "Fuzzy and crisp logical rule extraction methods in application to medical data," in *Computational Intelligence and Applications 23*, P. S. Szczepaniak, Ed. Berlin: Springer-Verlag, 2000, pp.593-616.

[4] G. Richards, V. J. Rayward-Smith, P. H. Sönksen, S. Carey, and C. Weng, "Data mining for indicators of early mortality in a database of clinical records," *Artificial Intelligence in Medicine*, vol.22, no.3, pp.215-231,2000.

[5] P. J. G. Lisboa, E. C. Ifeachor, and P. S. Szczepaniak, Eds. *Artificial Neural Networks in Biomedicine*. London: Springer-Verlag, 2000.

[6] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," Artificial Intelligence in Medicine, vol.23, no.1, pp.89-109, 2001.

[7] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-Based Systems*, vol.8, no.6, pp.373-389, 1995.

[8] Asada N, Doi K, MacMahon H, et al. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study. *Radiology* 1990;177:857–860.

[9] Moreno-Seco, F., L. Mico, and J.A. Oncina, Modification of the LAESA Algorithm for Approximated k-NNClassification. . Pattern Recognition Letters, 2003. 24 p. pp. 47–53.

[10] Siti Farhanah Bt Jaafar and Dannawaty Mohd Ali, "Diabetes mellitus forecast using artificial neural networks", Asian conference of paramedical research proceedings, 5-7, September, 2005, Kuala Lumpur, Malaysia.

[11] T.Jayalakshmi and Dr.A.Santhakumaran, "A novel classification method for classification of diabetes mellitus using artificial neural networks". 2010 International Conference on Data Storage and Data engineering.

[12] Rajeeb Dey and Vaibhav Bajpai and Gagan Gandhi and Barnali Dey, "Application of artificial neural network technique for diagnosing diabetes mellitus", 2008 IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10.

[13] Humar, K., & Novruz, A. *Design of a hybrid system for the diabetes and heart diseases.* Expert Systems with Applications, 2008, 35, 82–89.

[14] B.M Patil, R.C Joshi, Durga Tosniwal, *Hybrid Prediction model for Type-2 Diabetic Patients,* Expert System with Applications, 37, 2010, 8102-8108.

[15] Asha Gowda Karegowda ,MA.Jayaram , ―Integrating Decision Tree and ANN for Categorization of Diabetics Data ―, International Conference on Computer Aided Engineering, December 13-15, 2007, IIT Madras, Chennai, India.

[16] Asha Gowda Karegowda , A.S. Manjunath , M.A. Jayaram *Application Of Genetic Algorithm Optimized Neural Network Connection Weights For Medical Diagnosis Of Pima Indians Diabetes,* International Journal on Soft Computing (IJSC), Vol.2, No.2, May 2011.

[17] I. E. Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad, and J. O. Deasy, "Exploring feature-based approaches in PET imaging for predicting cancer treatment outcomes," *Pattern Recog.*, vol. 42, pp. 1162–1171, 2009.

[18] G. V. Saradhi, G. Gopalakrishnan, A. S. Roy, R. Mullick, R. Manjeshwar, K. Thielemans, and U. Patil, "A framework for automated tumor detection in thoracic FDG PET images using texture-based features," in *Proc. Int. Symp. Biomed. Imag. (ISBI 2009)*, 2009, pp. 97–100.

[19] D. Coomans and D. L. Massart, "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," *Analytica Chimica Acta*, vol. 138, pp. 153–165, 1982.

[20] UCI machine learning repository and archive.ics.uci.edu/ml/datasets.html