Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 9, September -2017

An Enhance Expert System for Diagnosis of Diabetes using Fuzzy Rules over PIMA Dataset

Hanslal Prajapati[1], Anurag Jain[2], Sanjay Kumar Pal[3]

[1,2,3]*Department of Computer Science and Engineering*
Radharaman Institute of Technology and Science, Bhopal, India

Abstract —*To improve the accuracy and to achieve better efficiency, a new Fuzzy Inference System (FIS) is proposed. This paper presents the design and implementation of a fuzzy logic based Expert System for Diagnosis of Diabetes. Pima Indian diabetic dataset is used as data set for designing the FIS. The proposed Expert System for diagnosis of diabetes is more accurate as compare to other existing approached present in the literature.*

Keywords-*diabetes; diagnosis; expert system; fuzzy logic; patient*

I. INTRODUCTION

Diabetes is a clinical syndrome distinguished by hyperglycemia due to absolute or relative deficiency of insulin. The lack of insulin affects the metabolism of the body. It causes to increase the blood sugar level. That is there is not enough insulin to reduce the percentage of glucose to its normal level [1, 2].

All the people with diabetes have one thing common. They have too much sugar or glucose in their blood. This is because a person who intake food is converted into glucose and the body is unable to remove glucose from the blood and deliver it to the cells as a source of energy in order to stay alive. A hormone called Insulin, secreted by the pancreas helps to convert glucose to energy [2].

An expert system is a computer program that incorporates knowledge to solve complex problems and can either replace or assist a human expert [3, 4]. Many expert systems have been developed to diagnose diabetes and heart diseases, where diagnosis is complex and involves up-to-date parameters. In this context, various researchers have proposed their methods.

Lee and Wang in [9] proposed a diabetes fuzzy ontology for diabetes diagnosis. Classical ontology cannot perform decision making using fuzzy data. The ontology needs to first identify the class membership of (depending on symptoms and concepts) the patient on par with his/her diagnosis inputs such as age, glucose concentration, etc., and after that a personal diagnosis is performed. The knowledge base is created using concepts and fuzzy relations using fuzzy variables and fuzzy numbers.

Kalpana and Kumar in [13] proposed a fuzzy determination procedure for improving accuracy, and for simplicity of diabetes diagnosis. Fuzzy based determination mechanism consists of fuzzy implication, fuzzy inference, and fuzzy aggregation. The system is deployed using a fuzzy rule base just before the defuzzification.

In [23], an attribute weighting method called fuzzy C-means clustering based attribute weighting (FCMAW) for classification of Diabetes disease dataset has been used. Pima Indian Diabetes dataset has been used to find out the accuracy of the method. Fuzzy C-means clustering is an improved version of K-means clustering method.

The rest of the paper is organised as follows. Section II gives a brief review of database used in this paper. Section III discusses proposed methodology. A detailed result analysis is mentioned in section IV. Section V presents conclusion and future work.

II. DATABASE USED FOR DIAGNOSIS OF DIABETES

Generally, Pima Indian Diabetes (PID) [24] is used in the diagnosis of diabetes. PID consists of 768 cases. It has only two decision classes, one is tested negative for diabetes 500 (65%) instances and the other is tested positive for diabetes (35%). Each subject has eight attributes, including: Number of times pregnant, Plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT), Diastolic blood pressure (mm Hg), Triceps skin fold thickness (mm), 2-hour serum insulin (mu U/ml), Body mass index (weight in kg/(height in m)^2), Diabetes pedigree function, Age (years).

III. PROPOSED METHODOLOGY

The proposed Fuzzy Logic Controller is designed using Matlab fuzzy logic tool for Diabetes Diagnosis which consists of 8 Linguistic Inputs & produces 1 output :

3.1. Input and output parameters

Input parameters

A1 Number of times Pregnant

A2 Plasma Glucose level
A3 Diastolic Blood pressure
A4 Skin rashes and thickness
A5 2 –Hrs Serum Insulin
A6 Body Mass Index
A7 Diabetes pedigree
A8 Age

Output parameters:

DM Diabetes Mellitus

Fuzzy Variables	Representation of Fuzzy	Fuzzy Numbers	Representation of Fuzzy Numbers	Fuzzy Triangular Numbers
Pregnant times	**A1**	**Low**	**A11**	**[0 0.475 7.213]**
		Medium	**A12**	**[0.475 7.213 12.1]**
		High	**A13**	**[7.213 12.1 17]**
Glucose Level	**A2**	**Low**	**A21**	**[0 88.84 152.6]**
		Medium	**A22**	**[88.84 152.6 175.8]**
		High	**A23**	**[152.6 175.8 199]**
Diastolic BP	**A3**	**Low**	**A31**	**[0 49.73 88.44]**
		Medium	**A32**	**[49.73 88.44 105.2]**
		High	**A33**	**[88.44 105.2 122]**
Skin Thickness	**A4**	**Good**	**A41**	**[0 4.584 36.49]**
		Average	**A42**	**[4.584 36.49 67.74]**
		Below	**A43**	**[36.49 67.74 99]**
Serum Insulin	**A5**	**Low**	**A51**	**[0 35.44 195]**
		Medium	**A52**	**[35.44 195 520.5]**
		High	**A53**	**[195 520.5 846]**
BMI	**A6**	**Low**	**A61**	**[0 24.11 39.87]**
		Medium	**A62**	**[24.11 39.87 53.49]**
		High	**A63**	**[39.87 53.49 67.1]**
Pedigree	**A7**	**Low**	**A71**	**[0.078 0.1405 0.8032]**
		Medium	**A72**	**[0.1405 0.8032 1.612]**
		High	**A73**	**[0.8032 1.612 2.42]**
Age	**A8**	**Low**	**A81**	**[21 21.48 45]**
		Medium	**A82**	**[21.48 45 63]**
		High	**A83**	**[45 63 81]**
Diabetes Mellitus	**DM**	**Low**	**DM1**	**[0 0.25 0.45]**
		Medium	**DM2**	**[0.35 0.55 0.75]**
		High	**DM3**	**[0.65 0.85 1]**

Table 1. Parameters of Triangular Membership Functions

3.2. Proposed Algorithm

INPUT

Input the fuzzy set for A1, A2, A3, A4, A5, A6, A7 and A8

OUTPUT
Output the fuzzy set for DM

METHOD
Begin
Step1: Input the crisp values for A1, A2, A3, A4, A5, A6, A7 and A8.
Step 2: Set the triangular membership function for the fuzzy number with equation.
Step 3: Built the fuzzy numbers for A1, A2, A3, A4, A5, A6, A7 and A8 for input set
Step 3.1: Built the fuzzy number for DM for the output set.
Step4: Fuzzy inference is executed by Mamdani's method.
Step 4.1: Input the rule as {Rule 1,2...k}
Step 4.2: Matching degree of rule with OR fuzzy disjunction are calculated for fuzzy input set (A11, A12, A13, A21, A22, A23, A31, A32, A33, A41, A42, A43, A51, A52, A53, A61, A62, A63, A71, A72, A73, A81, A82, A83, DM1, DM2 and DM3).
Step5: Defuzzify into the crisp values using centroid method
Step6: Present the knowledge in the form of human nature language.
End.

3.3. Proposed Fuzzy Rules

1. If (A1 is Low) or (A2 is Low) or (A3 is Low) or (A4 is Good) or (A5 is Low) or (A6 is Low) or (A7 is Low) or (A8 is Low) then (DM is Low) (1)
2. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
3. If (A1 is High) or (A2 is High) or (A3 is High) or (A4 is BelowAverage) or (A5 is High) or (A6 is High) or (A7 is High) or (A8 is High) then (DM is High) (1)
4. If (A1 is High) or (A2 is High) or (A3 is Medium) or (A4 is BelowAverage) or (A5 is High) or (A6 is High) or (A7 is High) or (A8 is High) then (DM is Medium) (1)
5. If (A1 is High) or (A2 is Medium) or (A3 is High) or (A4 is BelowAverage) or (A5 is High) or (A6 is High) or (A7 is High) or (A8 is High) then (DM is Medium) (1)
6. If (A1 is Low) or (A2 is High) or (A3 is High) or (A4 is Good) or (A5 is Low) or (A6 is Low) or (A7 is Low) or (A8 is Low) then (DM is Low) (1)
7. If (A1 is Low) or (A2 is High) or (A3 is Medium) or (A4 is Good) or (A5 is Low) or (A6 is Low) or (A7 is Low) or (A8 is Low) then (DM is Low) (1)
8. If (A1 is Low) or (A2 is Medium) or (A3 is High) or (A4 is Good) or (A5 is Low) or (A6 is Low) or (A7 is Low) or (A8 is Low) then (DM is Low) (1)
9. If (A1 is Medium) or (A2 is High) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
10. If (A1 is Medium) or (A2 is Medium) or (A3 is High) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
11. If (A1 is High) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
12. If (A1 is Medium) or (A2 is High) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
13. If (A1 is Medium) or (A2 is Medium) or (A3 is High) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
14. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is BelowAverage) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
15. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is High) or (A6 is Medium) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
16. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is High) or (A7 is Medium) or (A8 is Medium) then (DM is Medium) (1)
17. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is High) or (A8 is Medium) then (DM is Medium) (1)
18. If (A1 is Medium) or (A2 is Medium) or (A3 is Medium) or (A4 is Average) or (A5 is Medium) or (A6 is Medium) or (A7 is Medium) or (A8 is High) then (DM is Medium) (1)

IV. RESULT ANALYSIS AND DISCUSSION

The performance of the proposed fuzzy expert system has been developed in MATLAB fuzzy logic toolbox. Pima Indians Diabetes Database was chosen to evaluated data set.

Through Pima Indian Diabetes Database, Knowledge can analyzed based on the Fuzzification interface, Fuzzy assessment methodology and Defuzzification interface for the parameter very young.

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100\%$$

where

True Positive (TP): Sick people correctly identified as sick
False Positive (FP): Healthy people incorrectly identified as sick
True Negative (TN): Healthy people correctly identified as healthy
False Negative (FN): Sick people incorrectly identified as healthy

Authors	**Year**	**Method**	**Accuracy**
X. Chang and J. H. Lilly [5]	2004	Fuzzy	77
Goncalves, Vellasco, Pacheco and De Souza [6]	2006	Neuro-Fuzzy	78.26
Nor Ashidi Mat Isa, Wan MohdFahmi Wan Mamat [7]	2011	Clustered-HMLP	80.59
Aibinu A M, Salami M J E, Shafie A A [8]	2011	ARI+NN	81.28
Chang-Shing Lee [9]	2011	Fuzzy	77.3
M. Kalpana and A.V Senthil Kumar [10]	2011	Fuzzy	85.03
M.Kalpana and A.V Senthilkumar [11]	2011	Fuzzy	87.38
A.V Senthilkumar and M.Kalpana [12]	2011	Fuzzy	88.35
M. Kalpana and A.V Senthil Kumar [13]	2011	Fuzzy	89.32
Chikh M A, Saidi M, Settouti N. [14]	2012	AIRS2	82.69, 89.10
Ozcift A. [15]	2012	RBF+eACC	76.30
M. Kalpana and A.V Senthil Kumar [16]	2012	Fuzzy	89.52
Ahmad F, Isa NA, Hussain Z, Osman MK. [17]	2013	Improved GA	80.4
AnujaKumari V, Chitra R. [18]	2013	SVM	78
M. Kalpanaand A. V. Senthil Kumar [19]	2013	Fuzzy	83.08
M. Kalpana and A.V Senthil Kumar [20]	2013	Fuzzy	83.52
Ravi Kumar G, Ramachandra G A, Nagamani K [21]	2014	GA + SVM	77.3
Vaishali Jain and SupriyaRaheja [22]	2015	Fuzzy	87.2
Kemal Polat [23]	2016	FCMAW	91.41
Proposed Method	**2017**	**Fuzzy**	**91.61**

Table 2. Comparison of the proposed expert system with existing expert systems in terms of accuracy

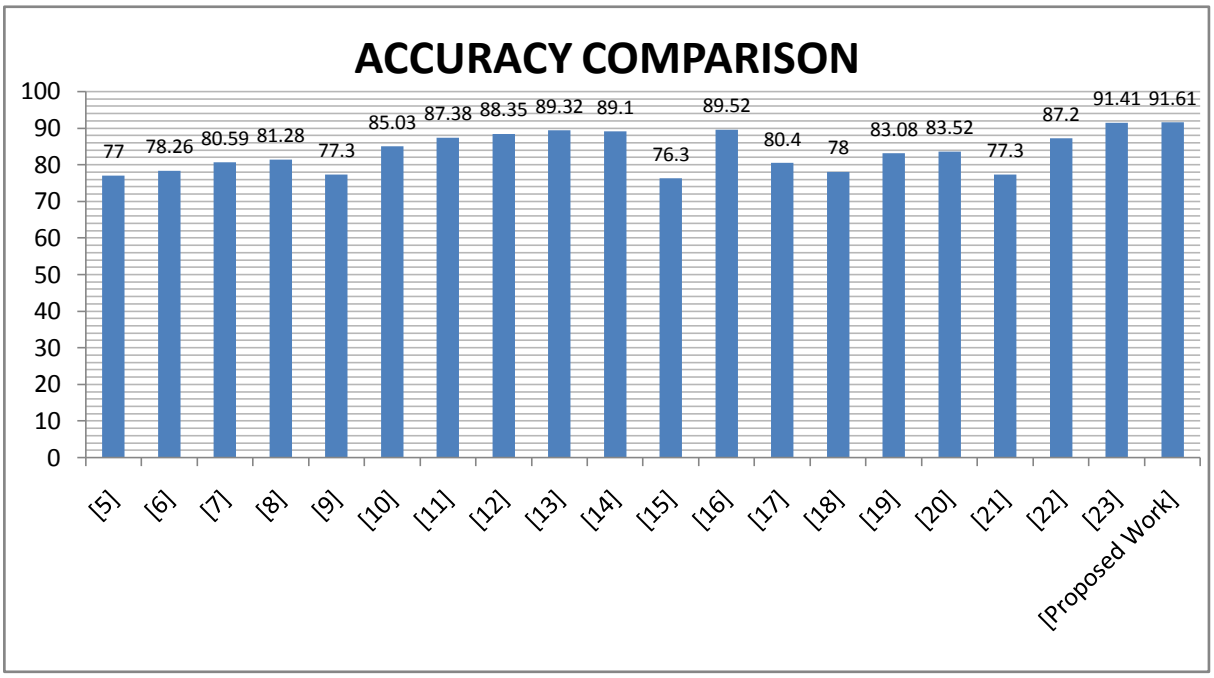

Figure 1. Comparison of the proposed expert system with existing expert systems in terms of accuracy

V. CONCLUSION

In order to diagnose any disease, the expert system developed by human may be a cheering way out to diminish cost, time, human efforts and medical error. In this paper, we have proposed Fuzzy Logic based Expert System for the diagnosis of Diabetes, most severe disease in India. It makes use of the well-known database for diabetes diagnosis. Through our implementation, 91.6 % accuracy is achieved.

Fuzzy controller can be used with various applications like neuro-fuzzy system, bioinformatics, expert system, air traffic signals, automatic power system and biological diagnostic. Our future work is to extend the system and then apply it to diagnose other severe diseases.

ACKNOWLEDGEMENT

The authors would like to thank Radharaman Institute of Technology and Science, Bhopal, India.

REFERENCES

[1] Nicki R. Colledge, Brian R. Walker and Stuart H. Ralston, Davidson's Principles and Practice of Medicine 21st Edition (Elsevier Limited,2010) 795-818.
[2] Smitha V, An Expert System for Diabetes Diagnosis, Master of Philosophy, Computer Science, Christ University Bangalore, Bangalore, India 2010.
[3] C. Yau and A. Sattar, "Developing Expert System with Soft Systems Concept," *Proceedings of International Conference on Expert Systemsfor Development*," pp. 79-84, 1994.
[4] P. Patra, D. Sahu, and I. Mandal, "An Expert System for Diagnosis of Human Diseases," *International Journal for Computer Applications(IJCA)*, Vol. 1, No. 13, 2010.
[5] X. Chang and J. H. Lilly, "Evolutionary design of a fuzzy classifier from data," IEEE transactions on systems, man, and cybernetics part b: cybernetics, vol. 34, no. 4, pp. 1894-1906, 2004.
[6] L. B. Goncalves, M. M. B. R. Vellasco, M. A. C.Pacheco and F. J. de Souza, "Inverted hierarchical neuro-fuzzy BSP system: A novel neuro-fuzzy model for pattern classification and rule extraction in databases," IEEE Transactions on Systems, Man, and Cybernetics-Part C: Application and Reviews, vol. 36, no. 2, pp. 236-248, 2006
[7] Nor Ashidi Mat Isa, Wan Mohd Fahmi Wan Mamat, "Clustered-Hybrid Multilayer Perceptron network for pattern recognition application, "Elsevier: Applied Soft Computing, January, 2011, 11(1), pp. 1457-1466.
[8] Aibinu A M, Salami M J E, Shafie A A, "A novel signal diagnosis technique using pseudo complex-valued autoregressive technique," *Elsevier: Expert Systems with Applications*, 2011, 38(8), pp. 9063-9069.
[9] Chang-Shing Lee and Mei-Hui Wang, "A Fuzzy System for Diabetes Decision Support Application", *IEEE Transaction on Systems, Man, andCybernetics*, Vol. 41, No.1, pp. 139-153, 2011.
[10] M. Kalpana and A.V Senthil Kumar, "Fuzzy Expert System for Diabetes using Fuzzy Verdict Mechanism," International Journal Advanced Networking and Applications, vol. 3, no. 2, pp. 1128-1134, 2011.
[11] M. Kalpana and A.V Senthil Kumar, "Enhanced Fuzzy Verdict Mechanism for Diabetes using Fuzzy Expert System," Journal of Computational Intelligence and Information Security, vol. 2, no. 8, 2011.
[12] A.V Senthil Kumar and M. Kalpana ,―Diagnosis of Diabetes using Intensified Fuzzy Verdict Mechanism‖, A. Abd Manaf et al. (Eds.): ICIEIS 2011, Part III, CCIS 253, pp. 123–135, Springer-Verlag Berlin Heidelberg 2011.
[13] M. Kalpana and A. Kumar, "Fuzzy Expert System for Diagnosis of Diabetes Using Fuzzy Determination Mechanism," *International Journal of Computer Science and Emerging Technology*, Vol. 2, No. 6,2011.
[14] Chikh M A, Saidi M, Settouti N., "Diagnosis of diabetes diseases using an Artificial Immune Recognition System2 (AIRS2) with fuzzy K-nearest neighbor," *Springer: J Med Syst.*, Oct 2012, 36(5), pp. 2721-29.
[15] Ozcift A., "SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease," *Springer: J Med Syst.*, Aug 2012, 6(4), pp. 2141-47.
[16] M. Kalpana and A.V Senthil Kumar, "Diagnosis of Diabetes using Correlation fuzzy logic in Fuzzy Expert System," International Journal of Advanced Research in Computer Science, vol. 3, no. 1, 2012.
[17] Ahmad F, Isa NA, Hussain Z, Osman MK., "Intelligent medical disease diagnosis using improved hybrid genetic algorithm--multilayer perceptron network," *Springer: J Med Syst*, April 2013, 37(2).
[18] Anuja Kumari V, Chitra R., "Classification of Diabetes Disease Using Support Vector Machine," *International Journal of Engineering Research and Applications*, April 2013, 3(2), pp. 1797-1801.
[19] M. Kalpanaand A. V. Senthil Kumar, "Design and Implementation of Fuzzy Assessment Methodology using K Ratio for Fuzzy Expert System," International Conference on Emerging Research in Computing, Information, Communication and Applications, pp. 62-67, 2013.
[20] M. Kalpana and A.V Senthil Kumar, "S Fuzzy Assessment Methodology using S Value to Diagnosis Diabetes," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 10, pp. 1350-1358, 2013.

[21] Ravi Kumar G, Ramachandra G A, Nagamani K, "An Efficient Feature Selection System to Integrating SVM with Genetic Algorithm for Large Medical Datasets," *International Journal of Advanced Research in Computer Science and Software Engineering*, Feb 2014, 4(2), pp. 272-277.

[22] Vaishali Jain and Supriya Raheja, "Improving the Prediction Rate of Diabetes using Fuzzy Expert System," International Journal of Information Technology and Computer Science, vol. 10, pp. 84-91, 2015.

[23] Kemal Polat, "Intelligent Recognition of Diabetes Disease via FCM Based Attribute Weighting," International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 10, no. 4, 2016, pp. 783-787.

[24] Pima Indians Diabetes Data Set, [Online], Available: http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

