Review on Diagnosis of Diabetes in Pima Indians

Anupriya K. Kamble
Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
MS (India)

Ramesh R. Manza
Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
MS (India)

Yogesh M. Rajput
Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
MS (India)

ABSTRACT
Diabetes is a disorder that most of the people suffer from and which also leads to death many of the times. Worldwide people suffer from the diabetes and the number is increasing day by day. Type 1 DM, Type 2 DM and Gestational diabetes are the types of diabetes. The main cause is due to prolong existence of high blood sugar level. There are many techniques and methods by which it can be diagnosed like image processing, pattern recognition, microwave tomography and so many and so forth. The present study mainly deals with the review for diagnosing diabetes in Pima Indians by using various pattern recognition techniques. The study done so far is by using a common database for each technique. While going through the review articles it was found that each different authors have applied different techniques. Some authors introduced new techniques and displayed their results by conducting new experiments and comparing with the old techniques. In which Comparative Disease Profile (CDP) and Separability of Split Value (SSV)gave accuracy of 76.4% and 74.8% respectively. It was also found that in some of the papers the technique was not used only for diabetes but for other diseases or disorder too. But as our aim was to study only on diabetes the results that are given have specifically mentioned in the result column itself with diabetes word specification. Invention of the new techniques gave satisfying result to authors. The above specified results will be much useful for the future study of the current review.

Keywords
Pima, diabetes, SSV, CDP

1. INTRODUCTION
A group of metabolic diseases in which there are high blood sugar levels over a prolonged period is known as diabetes mellitus(DM) commonly referred to as diabetes.[1] Diabetes occurs if either insulin production is inadequate or because the body cells do not respond properly to insulin, or both. It is explained that the Patients with high blood sugar will typically experience polyuria (frequent urination), they will become increasingly thirsty (polydipsia) and hungry (polyphagia). As of 2014, the estimated result shows that 387 million people worldwide have diabetes.[2] In which 8.3% of the adult population are present[3] with equal rates in both women and men.[4] Diabetes occurs when either the pancreas does not produce enough insulin or the cells of the body does not respond properly to the insulin produced.[5] Diabetes is not only the disease that gets suffer singly but after diabetes may invite other diseases with it. Diabetes is a metabolism disorder. Metabolism is referred as the way our body use digested food for energy and growth. A person with diabetes has a condition in which the quantity of glucose in the blood is too elevated (hyperglycemia). This happens because the body either does not produce enough insulin, produces no insulin, or has cells that do not respond properly to the insulin the pancreas produces. This results in too much glucose building up in the blood. This excess blood glucose eventually passes out of the body in urine. So, even though the blood has plenty of glucose, the cells are not getting it for their essential energy and growth requirements. The following table explains the diabetes properly:-

Table 1. Types of diabetes mellitus and its causes and effects:[5][6][7]

Sr. No.	Type of Diabetes	Begins from	Cause	% of diab-etes melli tus cases	Mostly affected areas
1.	Type 1 DM	Pancreas failed to produce enough insulin	Unkno-wn	10 %	North America and Europe
2.	Type 2 DM	Cells fails to respond to insulin properly (Insulin resistance)	Excessive body weight and not enough exercise	30 % 60 – 80 % 100 %	China and Japan Europe and Africa Pima Indians and Pacific Islanders
3.	Gestational diabetes	High blood sugar level of pregnant	High blood pressure without previous history of diabetes	2 – 10 %	Common

From the above table it can be clearly understood that the Type 2 DM affects most of the people from any area. Whereas Type 1 DM is observed less than the Type 2 DM. And Gestational diabetes is found leastly in any area. The Type 1 DM is found in majority of the immune-mediated nature, in which a T-cell-mediated auto immune attack leads to the loss of beta cells and thus insulin. Type 2 DM is the result of cells failing to respond to insulin that is it is occurred because of excessive body weight and not enough exercise. So to reduce

the number of increasing patients of this Type 2 DM just enough exercise is needed.Symptoms of high blood pressure are stated as frequent urination, increased thirst and increased hunger. It also states that if diabetes is untreated it may cause several complications, in which serious long-term complications includes cardiovascular disease, stroke, chronic kidney failure, foot ulcers, and damage to the eyes.[7] The first ever disease described by Egyptian manuscript from c.1500 BCE was diabetes[8] with mentioned as "too great emptying of urine". [9] At the same time the same disease was identified and classified as 'madhumeha' or 'honey urine' and specifying it as the urine that would attract ants.[9] It is said that the Greek Appollonius of Memphis in 230 BCE termed it as "diabetes" or "to pass through". [9] As major consequences and chances of dying persons are more there is a need to decrease the population affecting with the diabetes. The symptoms of diabetes can be seen in central brain, eyes, systemic, Breath, respiratory, Gastric and in urinary. So many different types of techniques have been used and invented for the detection of diabetes to reduce the number of diabetes patient and to prevent from it. Pattern recognition is one of the technique that is useful for the detection of diabetes. In this pattern recognition techniques many authors have used different algorithm, systems and methods. The present paper deals with the study of various pattern recognition techniques / algorithms / systems / methods used for detection of diabetes and the result obtained from it.

2. DATABASE[10]

The database referred for the present paper was from the National Institutes of Diabetes and Digestive and Kidney diseases; includes cost data which was donated by Peter Turney.

3. COMPARATIVE ANALYSIS

The present review has been studied taking into consideration the above database. For the detection of diabetes there were some existing techniques that have been used by authors and some of the authors invented the new techniques in which they have compared to the existing technique also by experimenting it and have mentioned it in their paper as well with full experiment. These technique used decision tree, neural network and tabu search algorithm. Only two techniques i.e. the Comparative disease profile (CDP) which is an automated system that compares the disease profile data and the seperability of split value (SSV) are useful to detect the diabetes. Rest all authors have either invented new technique or have not got accuracy but a different result from their research study. The techniques used and the result obtained from the respective author are described below :-

Table 2. Authors, techniques and result obtained in comparative analysis

Sr. No.	Authors	Technique / Algorithm	Result
1.	Peter Sykacek & Stephen Roberts	Adaptive Classification (Generalized nonlinear classifier & variational Kalman filter)	An approximate posterior i.e. a product of a multivariate Gaussian & a Gamma distribution.[11]
2.	Liping Wei & Russ B. Altman	Comparative disease profile (CDP)	Using only six attributes in the comparative disease profile gave a slightly higher cross validation accuracy of 76.4% of diabetes.[12]
3.	Kristin P. Bennett & Erin J. Bredensteiner	Parametirc Optimization Method for Machine Learning	No single error matric & combination of metrics used in MISMIN-P led to better results. [13]
4.	Andrew Watkins & Jon Timmis	Artificial Immune Recognition System (AIRS)	The revision to the algorithm do not sacrifice accuracy while increases the data reduction capabilities of AIRS[14]
5.	Stefan Rüping	Support Vector Machine (SVM)	Defined σ_{pp} by clipping f(x) at p_- and p+ instead of 0 & 1 is the new method & is compared with the existed approaches. [15]
6.	Marina Skurichina & Robert P.W. Duin	Boosting in Linear Discriminant Analysis	Usefulness of boosting does not depend on the instability of a classifier. [16]
7.	Lawrence O. Hall, Nitish Chawla & Kevin W. Bowyer	Combined decision tree learned in parallel.	The decision tree to rules with rule conflict resolution approach has promise. [17]
8.	Peter D. Turney	Hybrid Genetic Decision Tree Induction Algorithm	ICET a new algorithm is introduced and is compared here with three other algorithms for cost-sensitive classification.

			[18]
9.	Ilya Blayvas & Ron Kimmel	Interpolation & approximation algorithms for classification	For low dimensional training sets (D≤10) this gives an advantage in both runtime and memory usage, while maintaining classification performances comparable to the best reported results of existing technique. [19]
10.	Kristin P. Bennett, Ayhan Demiriz and Richard Maclin	Adaptive semi-supervised ensemble method	ASSEMBLE is introduced and results are obtained by using both decision both decision trees & neural networks which supports the method. [20]
11.	Jennifer A. Blue & Kristin P. Bennett	Hybrid tabu search	Hybrid tabu search method is introduced & computational results show that the hybrid algorithm outperforms the descent & tabu search algorithm used alone.[21]
12.	Altaf H. Khan	Novel feedforward network	The new network despite having strong constraints on its weights, has a generalization performance similar to that of its conventional counterpart. [22]
13.	Michael Lindenbabaum, Shaul Markovitch & Dmitry Rusakov	Lookahead algorithm	Proposed algorithm outperforms other methods.[23]
14.	Krzysztof Grabczewski & Wlodzislaw Duch	Separability of Split Value (SSV)	74.8% accuracy was obtained for diabetes. [24]

4. DISCUSSION

As stated above only Comparative Disease Profile (CDP) an automated system and Separability of Split Value (SSV) are useful for diagnosing diabetes of Pima Indians. So it is found that the need to detect diabetes through pattern recognition technique needs more attention as only two techniques give result of accuracy. Using the above techniques / algorithms we can further move on to detect diabetes instead of just inventing new methods. Good results have been obtained but the detection of diabetes is more important and hence it is a good point for the future work to be studied using the above techniques for the detection of diabetes. If any new algorithms / system in pattern recognition is useful in combination that it gets the accuracy of detection of diabetes that may also be very helpful for the further study. And the same techniques can be used for the other dataset as well.

5. CONCLUSION

Metabolic diseases in which there are high blood sugar levels which persist for longer period of time is nothing but diabetes. If diabetes is untreated then it can lead to many complications. These complications majorly can cause damage to the blood vessels, diabetes doubles the risk of cardiovascular disease and coronary arteries disease the major problem in diabetes for the reason of death accurately of about 75%. The detection of diabetes at its early stage can be helpful for the decrease in patients in the diabetes. Several methods are there to detect diabetes. Pattern recognition plays an important part in the detection of the diabetes disease. Some authors invented new pattern recognition techniques to detect diabetes and experimented it with the existing techniques and found satisfying data for the new ones. Each author have implemented separate techniques and no technique have been repeated in any paper. The Comparative disease profile (CDP) and Separability of Split Value (SSV) has obtained good accuracy for the detection of diabetes i.e. 76.4% and 74.8% respectively. From this also we can conclude that as Comparative disease profile (CDP) gives 76.4% accuracy it is better than Separability of Split Value (SSV).

6. FUTURE SCOPE

The techniques used in the present review can be used for the diabetes dataset. And results can be obtained. Going more elaborately in the topic in future it may be possible to detect the symptoms of diabetes at an early stage using the above mentioned database by using proper pattern recognition method. By doing this one can minimize the affecting persons from the other diseases that can be caused by diabetes. This study is very useful from the medical point of view.

7. REFERENCES

[1] "About diabetes". World Health Organization. Cited 2015 November 30.

[2] "Update 2014". IDF. International Diabetes Federation. Cited 2015 November 30.

[3] Shi, Yuankai; Hu, Frank B (2014 June 7). "The global implications of diabetes and cancer". *The Lancet 383*

(9933): 1947-8. doi:10.1016/S0140-6736(14)60886-2. PMID 24910221.

[4] Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V; et al. (2012 Dec 15). "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010.". Lancet 380 (9859): 2163-96 doi:10.1016/S01460-6736(12) 61729-2. PMID 23245607

[5] Shoback 2011. Greenspan's basic & clinical endocrinology. Edited by David G. Gardner, Dolores(9th ed.). New York:McGraw-Hill Medical.

[6] National Diabetes Clearinghouse (NDIC):National Diabetes Statistics 2011". U.S. Department of Health and Human Services. Cited 2015 Dec. 1.

[7] "Diabetes Fact Sheet No312". WHO. October 2013. Cited 2015 Dec. 1.

[8] [8] Ripoll, Brian C. Leutholz, Ignacio, 2011 April 25. Exercise and disease management [2nd edition]. Boca Raton : CRC Press. p. 25. ISBN 978-1–4398–2759–8.

[9] Editor, Leonid Poretsky, 2009. Principles of diabetes mellitus (2nd ed.). New York : Springer. p. 3. ISBN 978-0-387-09840-1.

[10] Murphy PM, Aha KW, 1994. UCI Repository of machine learning databases, http://www.ics.uci.edu/ ~mlearn/MLRepository.html], Irvine, CA: University of California, Department of Information and Computer Science.

[11] Peter Sykacek and Stephen J. Roberts, 2002. Adaptive Classification by Variational Kalman Filtering. NIPS.

[12] Liping Wei and Russ B. Altman. An Automated System for Generating Comparative Disease Profiles and Making Diagnoses. Section on Medical Informatics Stanford University School of Medicine, MSOB X215.

[13] Kristin P. Bennett and Erin J. Bredensteiner, 9. 1997. A Parametric Optimization Method for Machine Learning. INFORMS Journal on Computing.

[14] Andrew Watkins and Jon Timmis and Lois C. Boggess. Artificial Immune Recognition System (AIRS): An ImmuneInspired Supervised Learning Algorithm. (abw5,jt6@kent.ac.uk) Computing Laboratory, University of Kent.

[15] Stefan R uping. A Simple Method For Estimating Conditional Probabilities For SVMs. CS Department, AI Unit Dortmund University.

[16] Marina Skurichina and Robert P W Duin, 2000. Boosting in Linear Discriminant Analysis. Multiple Classifier Systems.

[17] Lawrence O. Hall and Nitesh V. Chawla and Kevin W. Bowyer. Combining Decision Trees Learned in Parallel. Department of Computer Science and Engineering, ENB 118 University of South Florida.

[18] Peter D. Turney, 1995. Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. CoRR, csAI/9503102.

[19] Ilya Blayvas and Ron Kimmel. Efficient Classification via Multiresolution Training Set Approximation. CS Dept. Technion.

[20] Kristin P. Bennett and Ayhan Demiriz and Richard Maclin, 2002. Exploiting unlabeled data in ensemble methods. KDD.

[21] Jennifer A. Blue and Kristin P. Bennett, 1996. Hybrid Extreme Point Tabu Search. Department of Mathematical Sciences Rensselaer Polytechnic Institute.

[22] Ahmed Hussain Khan and Intensive Care. Multiplier-Free Feedforward Networks. 174.

[23] Michael Lindenbaum and Shaul Markovitch and Dmitry Rusakov. Selective Sampling Using Random Field Modelling.

[24] Krzysztof Grabczewski and Wl/odzisl/aw Duch. THE SEPARABILITY OF SPLIT VALUE CRITERION. Department of Computer Methods, Nicolaus Copernicus University.

8. ACKNOWLEDGEMENT

I am thankful to University Grant Commission (UGC) for providing me a financial support in the form of Rajiv Gandhi National Fellowship.