An Appraisal of Some Selected Works Reported in the Literature on the Classification Techniques Used in Diagnosis and Management of Diabetes Mellitus

Oladipo, I. D., Babatunde, A. O. & Jimoh , R. G.
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
odidowu@unilorin.edu.ng, babatunde.ao@unilorin.edu.ng, jimoh.rasheed08@gmail.com

ABSTRACT

This paper appraised some selected studies reported in the literature on classification techniques used in diagnosis and management of diabetes mellitus in clinical data repository. The studies identify some methods and techniques that are mostly implored in carrying out research in this domain. Finding from this appraisal would go a long way at improving and bridge the identified gaps for an enhancement of the present study and further studies in this subject domain.

Keywords – Classification, Diagnosis, Diabetes Mellitus, Repository, Management and Enhancement

1. INTRODUCTION

Several literatures dealing with method and techniques of classification, diagnosis and management of diabetes mellitus in clinical data repository are abundant. In the last decades, a variety of models were devoted to different aspect of diabetes diagnosis and management, which are published regularly on national and international journals. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely, efficiently and automation of the process would be highly beneficial. Clinical decisions are often based on doctors' intuition and experience rather than on the knowledge rich data hidden in database. This practice leads to biases, errors and excessive medical costs which affects the quality of service provided to patients. Integration of clinical decision support with computer based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

Decision support is a crucial function for decision makers in many industries. Typically, decision support systems help decision-makers to gather and interpret information and build a foundation for decision-making. However, in healthcare, decision support systems play an important role through the interpretative analysis of large-scale patient data, thus, allow doctors and nurses to quickly gather information and process it in various informed ways that facilitate diagnosis and treatment decision (Norm & Basudeb, 2012). Diabetes Mellitus is a major global health problem affecting about 194 million people across the world and is expected to rise to 300 million by 2025 (Yue, Paul, Norman & Roy 2004).

Diabetes occurs as result of a disorder in the metabolism where the amount of glucose in the blood is too high because the body cannot produce or properly use insulin. In order to achieve more effective diabetes clinical management, data mining technique can be applied to a patient database. Diabetes is a relevant disease for data mining technology study for a number of reasons, the huge amount of data is there and diabetes is a common disease that cost a great deal of money in diagnosis, treatment and management. Diabetes is a group of disease that is marked by high levels of blood glucose, resulting from defects in insulin production, action or both (Harleen & Siri, 2006).

Diabetes mellitus are of two types, Type 1 and Type 2. Type 1 result from a failure of the human body to produce insulin and is less common than type 2. People with Type 1 diabetes take insulin injections. Type 2 diabetes is a form of diabetes usually develops in adults age 40 and above and is most common in the age group of over 55years old. People with Type 2 are usually overweight with metabolic syndrome that includes obesity, elevated blood pressure and high level of blood lipids (Joseph, Goodall & Fos 2002; Yue, Paul, Norman & Roy 2004; Okoro, 2014).

This study, adopts an appraisal of some selected works reported in the literature on the classification techniques used in diagnosis and management of diabetes mellitus. The remaining part of this paper is organized as follows: Section 2 deals with the materials and methods while section 3 presents the results and discussion and section 4 the conclusion.

2. MATERIALS AND METHODS

The method used in this appraisal of work reported in the literature on the classification techniques used in diagnosis and management of diabetes mellitus was through the survey of journals and publications in the fields of computer science, engineering and healthcare, Artificial Intelligence in medicine, international journal of computer application, international journal of medical informatics, international journal of computer science & Engineering and Transactional on Engineering, computing and technology are some of the journals used in this study. The study also includes books chapters, dissertation, working papers and conference papers. Most recent publications are considered in this review with 2002 as the cutoff year. Findings form this appraisal would go a long way in coming up with a robust method to further the research in the development of clinical decision support system for the treatment of diabetes.

3. RESULTS AND DISCUSSION

The table below presents a succinct summary of results obtained from the research and the discussions of same.

S/N	Title of Paper	Author(s)	Year	Objective	Methodology	Dataset	Evaluation	Remarks/ Future work
1.	Data Mining a Diabetic Data Warehouse	Breault et al	2002	Examined diabetic data warehouse, showing a method of applying data mining techniques, and some of the data issues, analysis problems and result.	Use classification and regression Tree (CART) to deduce a series of association rules among different attributes.	An Integrated healthcare System, New Orleans with 30,383 diabetes patients dataset was used	The work revealed that data mining can discover novel association that is useful to clinicians and administrator.	CART was able to achieved classification accuracy of 59.5% which the authors found useful to derive rule set.
2.	Feature Selection and Classification Model Constriction on Type 2 Diabetic Patient's Data	Huang et al	2004	Use feature selection and classification model construction on Type 2 diabetic patient's data in an attempt to improve the efficiency of data mining algorithms. A feature selection technique Relief F is use with the data, which can rank the important attribute affecting Type 2 diabetes control	The work comparatively evaluated the performance of three machine learning methods Naïve Bayes, IB1 and C4.5 on the data set. C4.5 provided the best classification accuracy before and after feature selection mechanisms.	The dataset was extracted from commercial Diamond Clinical Information System which contains 2,017 Types 2 diabetic Patient's Clinical information was used.	The model provided a best predictive performance of 77%, which could assist decision making. The results are in accordance with diabetes guidelines and the Clinician's opinions. The performance of IB1 and Naïve Bayes were significantly improved after the application of Relief F.	Future concentrate on enhancing the performance of Relief F, as efficient and accurate feature selection is a prerequisite for generating predictive models from large medical dataset.

#	Title	Author	Year	Aim	Method	Population/Data	Result	Conclusion
3.	Profile of Nigerians with Diabetes Mellitus – Diabcare Nigeria study group (2008): Results of a multicenter study.	Chineneye, et al	2012	Review the profile of Nigerian with diabetes, their aim was to assess the clinical and laboratory profile, and evaluate the quality of care of Nigerian diabetics with a view of planning improved diabetes care.	They carry out a multicenter study of seven tertiary health institutions covering the six geopolitical regions of Nigeria. The clinical and laboratory parameters of diabetic out-patients were evaluated.	The study population included consenting adult diabetic patients attending the diabetes Clinics of Seven tertiary health institutions covering six geopolitical regions of Nigeria.	The result revealed that most Nigeria diabetics have sub optimal glycemic control, are hypertensive, and have chronic complication of DM.	Improved quality of care and treatment to target is recommended to reduce diabetes – related morbidity and mortality.
4.	Data Mining Technologies for Diabetes: A Systematic Review	Miroslav Marinov, Abu & Suzanne	2011	Review the application of data mining techniques in the field of diabetes research. The main goal of the work was to identify research goals, diabetes types, data sets, data-mining methods, data mining software and techniques and outcomes.	The work initially identified 31 article by search, and selected 17 articles representing various data-mining methods used for diabetes research.	MEDLINE Database Through PubMed	The work revealed that the application of data-mining techniques in the selected articles were useful for extracting valuable knowledge and generating new hypothesis for further scientific research /experimentation and improving health care for diabetes research and would be a valuable asset for diabetes patients.	Data mining has played an important role in diabetes research and would be a valuable asset for diabetes researchers because it can unearth hidden knowledge from a huge amount of diabetes related data. Thus ultimately improve the quality of healthcare for diabetes patients.
5.	Empirical Study of Applications of Data Mining Techniques in Healthcare	Kaur & Wasan	2006	Empirically study applications of Data mining Techniques in Healthcare, using classification techniques on medical data set of diabetic patients as a case study.	Examine the potential use of classification based data mining technique such as Rule based, decision tree and Artificial Neural Network to massive volume of healthcare data.	The dataset used for the study was created by George John and appears on the UCL ML Data Repository at http//icddics.uci.edu. It contains 8 continuous attributes and 768 instance and two classes that determine either a person is or not having the disease		

6.	Temporal Data Mining for the Assessment of Cost Related to Diabetes Mellitus Pharmacological Treatment.	Stefano et al	2009	Use temporal data mining for the assessment of costs related to diabetes mellitus pharmacological treatment.	The work show the application of a temporal data mining techniques to extract temporal association rules over an integrated repository including administrative and clinical data related to a sample of diabetic patients.	The dataset used was created by George John on UCL ML Data Repository at http//kddics.uci.edu. It contains 8 continuous attributes and 768 instance.	Given patient records with corresponding diagnosis, data mining methods are able diagnose new cases.	
7.	Software System for Diagnosis and Classification of Diabetes Mellitus	Ayeni & Adewale	2008	The objective is develop a numerical model equation scheme to develop software system.	Pascal and visual Basic.Net were used to develop the diabetes monitor software that engages model equations scheme to developed a software capable of diagnosing and classifying diabetes subjects.	Use data obtain from ten (10) patients to test the software	Using the model equations, a numerical scheme was used to develop a software capable of diagnosing and classifying diabetic subject	The software helps the medical practitioners to understand the extent of severity of this diseases in each patient.
8.	Intelligent and Effective Diabetes Risk Prediction System Using Data Mining	Kawsar et al	2012	An effective diabetes prediction system based on data mining and provided an efficient approach for the extraction of significant pattern from data warehouse for efficient prediction of diabetes was proposed.	The data were clustered using K-mean algorithms for identifying relevant and non-relevant data to diabetes. Next significant frequent patterns are discovered using AprioriTid and Decision Tree algorithm The method is implemented using Java.	Pre-processed 400 diabetes and non-diabetes patients from various diagnostic centres was used.	The method can efficiently and successfully predict the diabetes.	

9.	Application of Data Mining Methods and Techniques for Diabetes Diagnosis	Rajesh and Sangeetha	2012	Used data mining to find useful patterns to help in the treatment and diagnosis of diabetes datasets. The project aims for mining the relationship in Diabetes data for efficient Classification. The data mining methods and Techniques was explored to identify the suitable methods and techniques for efficient Classification of Diabetes datasets are mining useful patterns.	The data mining methods and Techniques was explored to identify the suitable methods and techniques for efficient Classification of Diabetes datasets are mining useful patterns. The researchers applied Data mining techniques to Classify Diabetes Clinical data and Predict the likelihood of a patient being affected with diabetes or not.	The training dataset used for data mining Classification was the Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Disease from UCI machine learning repository. The dataset contains 768 record samples, each having 8 attributes. This dataset was used for the Classification exercise, as the data is complete with no missing values.	Result was obtained after applying different Classification techniques to Pima India Diabetes database and the error rate results was obtained for the Ten(10) Classification Algorithm. A Classification rate of 91% was obtained for C4.5 algorithms.	Future enhancements of this work include improvisation of the C4.5 algorithms to improve the Classification rate to achieved greater accuracy in Classification.
10.	Medical data mining by fuzzy modeling with selected feature.	Ghazavi & Liao	2008	The paper presents a data mining study of medical data with fuzzy modeling methods that use feature subsets selected by some indices/methods.	Three fuzzy modeling method were employed which include the fuzzy k-nearest neighbor algorithm, a fuzzy clustering based modeling and adaptive network – based fuzzy inference system. For feature selection, a total of 11 indices /method are used.	Data set used are the Wiscosin breast cancer and the Pima India diabetes data set.	The classification accuracy and computational time are reported. Exhaustive testing of all possible combination of feature subsets with three features was also carried out. For the Wisconsin breast cancer dataset, the best accuracy of 97.17% was obtained, which is only 0.25% lower than that was obtained by	The paper shows that feature selection is important to mining medical data for reducing processing time and for increasing classification accuracy.

					exhaustive testing. For the Pima Indians diabetes dataset, the best accuracy of 77.65% was obtained, which is only 0.13% lower than that obtained by exhaustive testing.	
11.	A novel approach for the Diagnosis of Diabetes and Liver Cancer using ANFIS and improved KNN.	Kalaiselvi & Nasira	2014	The paper Examines the diagnosis of Cancer and diabetes on the bench mark dataset using ANFIS by training using adaptive group-based KNN. 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. ANFIS is relatively fast to converge due to its learning strategy and its easy interpretation. Each ANFIS classifier was implemented by using the MATLAB software package.		The experimental results accuracy is better than the existing approaches. The proposed model has lower complexity, achieves higher efficiency and performs the pattern classification better than traditional methods.	

12.	A Fast Computing Genetic – Neuro Fuzzy for Data Mining and its Application to Medical Diagnosis.	Kavita & Kavita	2013	The objective of the work is to simultaneously optimize the parameters and feature subset without degrading the ANFIS classification accuracy.	The research work involves the designing of frame work that incorporates genetic algorithm with neuro-fuzzy for feature selection and classification on the training dataset.	Ovarian cancer dataset.	The verification of the effectiveness of the proposed approach was tested on ovarian cancer dataset.	
13.	Neuro Fuzzy Classification and Detection Technique for Bioinformatics problems.	Othman & Yau	2007	The work identify new integration of classification or clustering algorithm especially in neuro fuzzy domain as compared to conventional or traditional method. The paper examine the suitability and performance of recurrent classification technique, fuzzy C means (FCM) as a classifier in neuro fuzzy system compared to sub clustering method.	After intensive literature reviews on various classification techniques, conventional or recurrent fuzzy C means is selected and this is integrated with ANFIS model. Six(6) stages was involved in the proposed model which started from the data input to output and was developed using MATLAB software.	In order to justify the selection of classification technique, the diabetes data which has total patterns of 768 with 8 inputs and 1 output is used to observe and investigate the rational of selected classification technique using ANFIS model.	The paper reveal that fuzzy c means techniques is better than sub-clustering technique and ANFIS model is a suitable and feasible method to be used as a supportive tools for bioinformatics problem.	

From all the literature review presented above, none of the past work discusses the management of Diabetic Mellitus. Our work therefore is to develop an efficient technique to improve management of Diabetes Mellitus. The future research methodology involves using fuzzy expert system for the management of diabetes mellitus to get efficient result.

4. CONCLUSION

In this study, several works related to classification, diagnosis and management of diabetes mellitus were reviewed to create background knowledge in the application of data mining in Healthcare domain. This study has shown that various methods and techniques have being used by researchers in the area of classification, management and diagnosis of diabetes mellitus. The appraisal has therefore helped us to come up with a robust and efficient technique for improving the management of diabetes mellitus using fuzzy expert system.

REFERENCE

1. Akbar, S., Rao, K. N. and Chandulal, J. A. (2011). Implementing Rule based Genetic Algorithm as a Solution for Intrusion Detection System. International Journal of Computer Science and Network Security, 11(8), pp. 138-144.
2. Andrea, L. H., Hsinchun, C., Susan, M. H., Bruce, R. S., Tobun, D. NG., Robin, R. S. & Kristin, M. T. (1999): Medical Data Mining on the Internet: Research on a Cancer Information System. Artificial Intelligence Review 13: pp 437 - 466
3. Ayeni, O. & Adewale, S. O. (2008): Software System for Diagnosis and Classification of Diabetes Mellitus Nigeria Computer Society (NCS) 22nd National Conference & 3oth Anniversary Proceedings 349 – 354.
4. Breault, J. L., Goodall, C. R. & Fos, P. J. (2002): Data Mining a Diabetic Data Warehouse. Artificial Intelligence in Medicine, 37 (1-2).
5. Chinenye, S., Uloko, A. E., Ogbera, A. O., Ofoegbu, E. N., Fasanmade, O. A.,
6. Fasanmade, A. A. & Ogbu, O. O. (2012): Profile of Nigerians with Diabetes Mellitus – Diabcare Nigeria study group (2008): Results of a multicenter study. Indian Journal of Endocrinology and Metabolism 16(4)
7. Divya, T. & Agarwal, S. (2013): A Survey on Data Mining Approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), pp 241 - 266
8. Ghazavi, S. N. & Liao, T. W. (2008): Medical data mining by fuzzy modeling with selected feature; Artificial Intelligence in Medicine 43, 195 - 206
9. Hacchesu, P. R., Ahmadi, M. & Sadoughi, F. (2013): Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients. Healthc Inform Res., 19(2): 121 – 129.
10. Han, J., & Kamber, M. (2001): Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, Inc. (Academic Press), San Francisco, CA.
11. Harleen, K & Siri, K. W. (2006): Empirical Study on Application of Data Mining Techniques in HealthCare. Journal of Computer Science, 2(2), pp 194 - 200
12. Jacob, S. G. & Ranani, R. G. (2011): Discovery of Knowledge Patterns in Clinical Data through Data Mining Algorithms: Multiclass Categorization of Breast Tissue Data. International Journal of Computer Applications, 32(7), pp 46 – 53.
13. Kalaiselvi & Nasira, G. M. (2014): A novel approach for the Diagnosis of Diabetes and Liver Cancer using ANFIS and improved KNN. Research Journal of Applied Sciences, Engineering & Technology 8(2); 243 - 250
14. Kavita, R. & Kavita, B. (2013): A Fast Computing Genetic – Neuro Fuzzy for Data Mining and its Application to Medical Diagnosis. International Journal of Engineering and Advanced Technology (IJEAT) 3(1); 2249 - 8958
15. Kawsar, A., Tasnuba, J., Ushin, F., Moniruzzaman, A. & Zamilur, R. (2012): Intelligent and Effective Diabetes Risk Prediction System Using Data Mining. Oriental Journal of Computer Science & Technology. 5(2) 215 – 221 Oriental Sci. Publishing Co., Indian.
16. Laetitia, J., Clarisse, D. and El-Ghazali, T. (2001). A Genetic Algorithm for Feature Selection in data-Mining for Genetics. 4th Metaheuristics International Conference. Portugal:Porto, pp. 29-33.