ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 11(42), DOI: 10.17485/ijst/2018/v11i42/132665, November 2018* ISSN (Online) : 0974-5645

Data Mining Driven Models for Diagnosis of Diabetes Mellitus: A Survey

F. S. Ishaq, L. J. Muhammad, B. Z. Yahaya and Y. Atomsa

Department of Mathematics and Computer Science, Federal University of Kashere, Nigeria; faisalsishaq@gmail.com, mljtech@gmail.com, baladada57@yahoo.com, au.nlaro@gmail.com

Abstract

Objective: In this study, a systematic effort was employed to identify and review data mining concept, tasks and model evaluation techniques, Knowledge Discovery and Data mining process Model (KDDM) model process and research articles published with reputable journal publishers that employed data mining techniques for diagnosis of Diabetes Mellitus. **Method/Analysis:** The findings from this work have been drawn from the published articles reviewed and the frequency analysis was used for the analysis of the reviewed works. **Finding:** The result of the study showed that, classification data mining task has been the most successfully and most frequently used data mining tasks for diagnosis of DM and the mostly commonly used classification data mining algorithms are Support Vector Machine and decision tree algorithms. **Novelty/Improvement:** In the study Support Vector Machine was realized to be most efficient data mining algorithm for diagnosis of Diabetes Mellitus using either clinical or biological and clinical dataset of Diabetes Mellitus. Despite its popularity, SVM algorithm should be further improved in the future work so as to further improve its efficiency.

Keywords: Algorithm, Data Mining, Diabetes; Diagnosis, Knowledge, Pattern

1. Introduction

Data mining is a process of uncovering hidden patterns or useful knowledge from large dataset or database. Therefore, it is process of discovering novel, valid, useful and understandable hidden patterns or knowledge in a database or dataset[1]. Since 1960s, data mining techniques have become a subdivision of applied Artificial Intelligence (AI)[2]. As a result of its explosive rapid growth, data mining have turn out to be an increasingly vital research area[3]. Yet it is becoming widely accepted in medical field, because of its efficient analytical techniques and algorithms, as it uncovers useful and valuable knowledge in medical datasets or databases.

In medical filed, data mining has proved to provide several benefits such as prediction or detection of cause of diseases, availability of medical solutions to patients at low cost and identification of medical treatment methods among others[4]. To open a window of comparatively better resources data mining techniques are employed to enhance the sensitivity and/or specificity of disease detection and diagnosis. Hence, this substantially reduces the accompanied cost as unwanted and expensive medical tests are bypassed[5].

Many data mining techniques had been used for diagnosis of Diabetes Mellitus which include classification, clustering, neural network, association rule, prediction and summarization. However, the most widely used data mining techniques are clustering and classification techniques[6]. Those data mining techniques are increasingly becoming popular because they have been used for mining medical dataset for uncovering hidden patterns and useful knowledge[7].

Diabetes Mellitus (DM) is the one of the chronic diseases in the world today especially in most of the developed countries and now it is rapidly spreading in many of the developing countries. It is evident that, DM is also posing a threat on developing countries more than their counter parts thus developed countries. At beginning of the twentieth century, Dr. Cook described DM as the uncommon disorder in Africa. Yet, some compelling data available proved it's prevalence in the continent. An

*Author for correspondence

estimate of 1% of the population living in rural areas of Africa has DM, while 5% to 7% were recorded in urban areas and sub-Saharan Africa respectively[8]. Diagnosis of disease such as diabetes is one of the applications where DM algorithms are proving successful results in the recent years[9]. Therefore, Leveraging data mining techniques is key approach to utilizing large volume of available dataset of diabetes for uncovering and extracting useful knowledge and hidden patterns for diagnosis of this chronic diagnosis thus DM. and it would increasingly reduce the prevalence of DM especially in developing countries particularly in Africa. Nonetheless, this study aims at reviewing data mining concept, process model, classification tasks, model evaluation techniques and knowledge discovery. Hence, publish an article that employs data mining models for diagnosis of DM.

2. Knowledge Discovery and Data Mining Process Model

Knowledge Discovery and Data mining process Model (KDDM) defined the set of activities that involve extraction of hidden patterns or useful knowledge from database which include how the data is stored, accessed and retrieved, how to efficient and scalable algorithms that can be used to analyzed large dataset, how to interpret and visualize the results and how to model and support the interaction between machine and human. KDDM also involves learning and analysis support for analyzing application domain. However, DM is one of the series of steps involve in KDDM process[10]. KDDM is nontrivial process for identifying novel, valid, untimely understandable and potentially useful knowledge or patterns in a large dataset[11,12]. The reason behind defining and implementing KDDM process model is to make sure that end product is useful to its target users, therefore only using well defined and formal methods such desirable goals can be successfully achieved. In a note shell, set of steps involve in KDDM processes model include selection, pre-process, transformation, data mining, interpretation and evaluation. Figure 1 shows a typical KDDM process model.

But, the most important step in the entire KDDM process model is a data mining because it is exemplifying

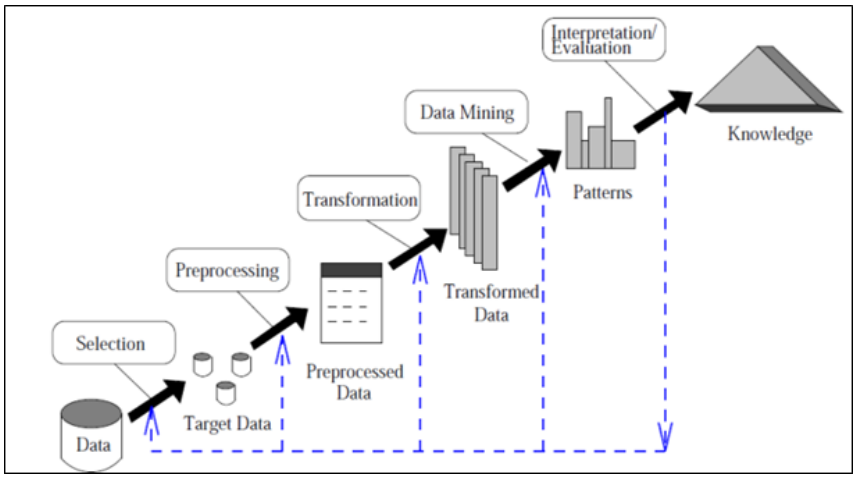

Figure 1. Typical KDDM process model[38].

the application of data mining algorithms in analyzing data for uncovering useful patterns or knowledge[13].

There are many types KDDM process models but the mostly widely used KDDM process models are Academia and CRoss-Industry Standard Process for Data Mining (CRISP-DM) process models. Academia KDDM process model has eight steps which include understanding application domain, creating target dataset, data cleaning and pre-processing, data reduction and projection, choosing data mining algorithm, data mining, interpretation of mined patterns, and consolidating discovered knowledge[11]. Industrial CRISP-DM process model had been developed with support from IBM. The model has six steps which include business understanding, data understanding, data preparation, data mining or modeling, evaluation and deployment[14]. Neither of the KDDM process models is the best; however the best model is how best the it has been used for uncovering useful, valid and novel patterns or knowledge from the dataset.

3. Data Mining Tasks

The tasks of data mining are very distinct and diverse because many knowledge or patterns exist in a dataset or database, therefore different techniques are needed to find the different kind of knowledge or patterns[15]. Medical researchers are motivated to use data mining techniques for knowledge discovery due to large growth of medical databases especially in developed countries. When the volume of database increases, data mining techniques play important role in uncovering hidden patterns and extracting useful knowledge to provide better diagnostic capabilities and patient care for diabetes and other diseases[16]. Based on the knowledge or pattern, the data miner looking for, tasks in data mining can be classified into classification, clustering, association rule, neutral network and summarization[15].

3.1 Classification

This is the most studied tasks in data mining for many decades by statistics and machine learning communities among others. In this task, the aim is to predict the value or class of user specified goal attributes based on the value of other input attributes called predicting attributes[17]. Classification data mining algorithms employ a set of pre-classified examples to develop a model that can classify the population of the data at large[18]. To predict certain outcome based on a given input, classification algorithms process a training dataset that containing a set of attributes and respective outcomes which are usually called prediction or goal attributes[4]. The accuracy of the data mining classification algorithm defined how good the algorithm is. Example of data mining classification algorithms includes decision tree induction, Support Vector Machine, Bayesian classification among others. Many works utilized classification data mining algorithms for diagnosis of DM[19].

3.2 Clustering

Clustering data mining task are used to identify the similar class of objects in the dataset called clusters or group for a set of objects whose classes are unknown. Clustering data mining algorithms are further used to identify sparse and dense regions in object space and overall distribution pattern and correlation among attributes of the dataset. Inter classes similarities are minimized and maximized if the objects are so clustered and it is done based on some conditions or criteria defined on the attributes of the objects. When the clustered are defined and decided, the objects are labeled with their corresponding cluster. The common characteristics of objects in a cluster are summarized to form the class description[15]. However classification data mining algorithms can be used for effective means for differentiating classes or groups of objects in the dataset but it is costly, so clustering can be used for pre-processing for attributes subset selection and classification. Example of the algorithms of clustering data mining technique include partitioning methods, density based methods, model based method, grid base method, hierarchical agglomerative method among others[20]. Like classification data mining algorithms, clustering data mining algorithm had been used for diagnosis of DM in the works[21-23]. But however, most of works used clustering data mining algorithm, together with some classification data mining algorithms.

3.3 Association Rule

Association is the discovery of connection or togetherness of objects. Such kind of connection or togetherness is called association rule. Association data mining task usually used to find a frequent item set findings among large data sets. In the standard form of this each data instance (or "record") consists of a set of binary attributes called items. Each instance usually corresponds to a customer transaction, where a given item has a true or false value depending on whether or not the corresponding customer bought that item in that transaction. Association data mining task algorithms are able to generate rule with confidence values less than one. But however, the number of possible association rules for a given dataset is usually very large and have a high proportion of association rules are often little value[17]. Types of association data mining algorithms include Apriority Algorithm, multilevel, qualitative, multidimensional association rules among others. Like other data mining techniques, association rule had been used for diagnosis of DM[24,25].

3.4 Summarization

Summarization is the abstraction or generalization of data. The summarized result is in smaller set which gives a general and summarized overview of data with usually aggregate information. For instance, the long distance calls of customer can be summarized in to total minutes, total calls, and total spending instead of detailed call is

presented to sale manager for customer analysis[26]. It is a key data mining concept for finding a compact description of a dataset. Some of the simple summarization data mining techniques include mean and standard deviation that are often applied for data analysis, generation automated report and data visualization[27]. However, summarization data mining techniques are not commonly used for diagnosis of diseases including DM.

3.5 Neutral Network

Neutral network methods are rarely used for data mining task as they often produce incomplete models and also demands extensive training period[28]. In this case, a set of connected input are linked to the output set via a weighted h connections. In order to accurately predict the class labels of the input tuples, the network is allowed to learn by continuously altering the weights. Hence, grant it remarkable ability to drive meaning from imprecise or complicated data. Consequently, extract patterns or knowledge to detect trends that are hard to detect by either computer techniques or human. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs. Types of neural networks include back Propagation and forward Propagation among others[29]. Neural network data mining techniques are employed to diagnose DM as in the work of[30].

4. Data Mining Model Evaluation Techniques

The data mining model evaluation techniques are used for evaluating the performance of the various data mining algorithms include the following:

1. *Accuracy*: measure the proportion of the total number of predictions that were correct.

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \quad (1)$$

2. *The recall or True Positive (TP) rate*: measure the proportion of positive cases that were correctly identified, as calculated using the equation.

$$TP = \frac{d}{c+d} \quad (2)$$

3. *The False Positive (FP) rate*: measure the proportion of negative cases that were incorrectly identified, as calculated using the equation.

$$FP = \frac{b}{a+b} \quad (3)$$

4. *The True Negative (TN) rate*: measure the proportion of negative cases that were classified correctly, as calculated using the equation.

$$TN = \frac{a}{a+b} \quad (4)$$

5. *The False Negative (FN) rate*: measure the proportion of positive cases that were unclassified correctly, as calculated using the equation.

$$FN = \frac{c}{c+d} \quad (5)$$

6. *Precision:* measure the proportion of predicated positive cases that were correct, as calculated using the equation.

$$\text{Precision} = \frac{d}{b+d} \quad (6)$$

7. *Receiver Operator Characteristic test (ROC)*: is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test.

While performance evaluation metrics of the association rule include the following:

1. Support determines how often a rule is applicable to a given data, as calculated using the equation:

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (7)$$

2. Confidence determine how frequently item in Y appear in transaction that X, as calculated using the equation:

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma N} \quad (8)$$

5. Diabetes Mellitus

DM is also called Diabetes. It is a kind of metabolic disease which occurs as a result of the abnormal increase in

the blood glucose level. This happens because of either the defects in insulin secretion, insulin reaction or both. The chronic hyperglycemia of diabetes is often associated with dysfunction, long term damage and failure of various organs such as kidneys, eyes, hearts, nerves and blood vessels[31]. DM is usually resulting in high disease burden in developing countries such as Nigeria. According to prediction of World Health Organization (WHO) by year 2030, approximately 350 million people will be affected by diabetes in the world[9].

There are three types of DMs. The absolute lack of insulin is the main cause of Type I DM (sometimes identified as Insulin Dependent DM or Juvenile Onset DM). People at risk of affecting this type of DM are usually identified by serological evidence of an autoimmune pathological process occurring in the pancreatic islets and genetic markers[2].

However, people that are either deficient in relative insulin secretory response or resistive to insulin action or both are diagnosed with DM type II that is, non-insulin dependent DM or Adult onset DM. This amount to 90% cases of DM[8].

Finally, any abnormal rise in the blood glucose noticed during pregnancy is recognized as Gestational DM (GDM)[32]. Moreover, GDM is often accomplished by cardiac diseases, renal complications and peripheral vascular diseases[33].

Early recognition or identification of DM to people or people with high risk of developing DM is very important challenge in the medical field, in order to reduce the prevalence of the disease. The availability of huge amount of data especially DM dataset lead to the need for powerful data analysis techniques to mine or extract novel and useful patterns or Knowledge in those available huge data. However, scholar and researcher had long been using or applying data mining algorithms and techniques to improve data analysis on those available huge data sets. Diagnosis of disease such as DM is one of the applications where data mining techniques are proving successful results in the recent years.

6. Methods

Only research articles published by reputable journal publishers and employed data mining techniques for diagnosis of DM, were searched, identified and reviewed in this work. These papers include conference papers and journal articles.

Many scholars employed classification, neutral network, association rules and clustering data mining techniques and algorithm for diagnosis of diabetes in their research. In[22] diabetes disease classified using SOM, PCA and NN for clustering, noise removal and classification task for development expert system for diagnosis of diabetes. The experimental results of the study showed that the effectiveness of incorporating the clustering and PCA techniques for classification of accuracy of diabetes disease and it to help to improve the classification accuracy of diabetes disease by more than 12%, 13%, 2% and 0.6%. According to the study review in[16] for leveraging data mining and machine learning for diagnosis of diabetes mellitus, support vector machine arise as the most successful and widely used data mining algorithm. A hybrid data mining and case base reasoning model for monitoring and predicting diabetics was presented in the study of[10]. Support Vector Machine as the classifier and case based reasoning cycle for determining and predicting sugar level of diabetics. In[19] intelligent modeling system for diagnosis of diabetes type II based on quantum particle swam optimization and algorithm and weighted least squares support vector machine was presented, however it overcame the disadvantage of large sample data and slow model building. The accuracy of the model is high due to mixed kennel function; add self-adapting weights and linear system of equation in the training model of WLS-SVM with QPSO algorithm. A hybrid prediction model for diagnosis of diabetes type II using K means clustering and C4.5 algorithm has been proposed in[22], the model was built on extracted patterns by leveraging C4.5 algorithm with 92% accuracy and it was further improved using neural network. The improved model separated the dataset into two groups and the model with 97% accuracy. A hybrid prediction model for diagnosis of diabetes type II with F-score feature selection approach towards improvement of the performance Support Vector Machine data mining classifier was proposed in[9]. The improved performance of the SVM classifier measured in terms of accuracy, sensitivity and area under curve and indeed F-score feature selection approach improved the performance of classifier. The proposed model achieved 98%, which was highest accuracy for diabetes dataset compared to other models in literatures reviewed in the work. Support Vector Machine algorithm was used to identified the effectiveness of different types of treatment of diabetic patients for different age groups using dataset of non-communicable disease

risk factors in Saudi Arabia which obtained from WHO in[6]. Age group for young diabetic patients were denoted as p(y) and while old age group denoted as p(o) respectively. Support Vector Machine algorithm was effectively used to identified preferential orders of treatment of diabetics' patients of each age group. In[5] a trained neutral network model for diagnosis of gestational diabetes mellitus was developed with four layers feed forward network, back propagation and regulation algorithm. The model has eight neurons, two hidden layers have ten neutrons each and the output layer has one neutron which served as the diagnosis result. The model was incorporated into a web based application so as to enhance its usage. The experimental result showed that, the model has over 95% accuracy. A data mining hybrid model based on K-means and logistic regression algorithms for predicting diabetes mellitus type 2 was presented in[34]. The model was proposed to improve the accuracy of the prediction of diabetes mellitus type II and result of the study showed that, the model attained a 3.04% higher accuracy of prediction compared to other researches in the work. In[35] performance analysis of data mining algorithms which include C4.4 Tree, K-Nearest Neighbors, and Random Forest, Support Vector Machines for predicting diabetes mellitus. Results of the study showed that, C4.5 algorithm achieved higher accuracy of 73.82% success rate. Hence, it has higher accuracy than other three algorithms before pre-processing of the diabetic dataset while after pre-processing both KNN and Random forest have much better accuracy of 100% than other algorithms. In[36] comparison the performance of Artificial Neural Networks (ANNs), logistic regress and decision tree algorithms for predicting diabetes using risk factors was presented. The result of the work showed that, logistic regression algorithm achieved a higher accuracy of 76.13% with a sensitivity of 79.59% and a specificity of 72.74%, while ANN algorithm reached a classification accuracy of 73.23% with a sensitivity of 82.18% and a specificity of 64.49%; and the decision tree (C5.0) achieved a classification accuracy of 77.87% with a sensitivity of 80.68% and specificity of 75.13%. Therefore, C5.0 decision tree algorithm has best classification accuracy, followed by the logistic regression model, and the ANN gave the lowest accuracy. Support Vector Machine for modeling diabetes mellitus was used in[37]. The dataset of diabetes data were classified into classification schemes in which classification Scheme I is diagnosed or vs. pre-diabetes and Classification Scheme II is undiagnosed diabetes or pre-diabetes. The dataset was sourced from National Health Nutrition Examination Survey from 1999 -2004. The result of the studies showed that, SVM data mining algorithm is a promising data mining modeling algorithm for detecting persons with common diseases such as diabetes and pre-diabetes in the population and abilities of the SVM models for Classification Schemes I and II, according to the area under the Receiver Operating Characteristic (ROC) curve, were 83.5% and 73.2%, respectively. A technique of discovering association rules of DM with complication was presented. Apriority Algorithm which is association rule of DM techniques was leveraged as a means to discover a relationship among DM with complication by focusing on diagnostic information[23]. Five best rules were formed, four out of the rules have conf: (1) while one rule has conf: (0.59). In[7], Apriority Algorithm of association rule data mining technique was used for classification of DM type II patients. Two set of rules were generated, the rules for patients who have no DM and those with DM. Ten rules were generated for patients with no DM while four rules were generated for patients with DM. the best rule for patient with no DM has conf: (1) and with DM has conf: (0.84).

7. Results and Discussion

The findings from this work have been drawn from the published articles reviewed in the literature review of section of the paper. Classification data mining task has emerged the most frequently used and successful data mining tasks for diagnosis of DM. from the review, the mostly commonly used classification algorithms is SVM and Decision tree algorithms. But however, SVM rises as the most successful data mining algorithm for in biological and clinical diagnosis of DM. It should be noted that, 43% of the articles reviewed used classification data mining algorithm only, with 83% of these paper used SVM algorithm. 29% of the papers reviewed used a hybrid of clustering and classification data mining techniques for diagnosis of DM. 14% of the articles reviewed used association rule data mining task and 100% the papers used Apriority Algorithm. Only 7% of the papers used neutral network data mining task while only 7% of the paper used a hybrid of neutral network and classification data mining techniques for diagnosis of DM. 64% of the works reviewed were for diagnosis DM Type II, 14% were for all type of DM and only 7% of the papers were for diagnosis

of generational DM. Table 1 compared the reviewed articles based on the type data mining tasks and algorithms used, and type of DM detected. All the paper were evaluated with one or more of the of data evaluation techniques discussed in section 4.

Table 1. Comparing the reviewed articles according to data mining tasks, algorithms and DM

Publication	Type of DM	DM Task	DM Algorithm
22	Type II	Clustering, classification	SOM, PCA, NN
16	Type II	Classification	SVM
10	All types	Classification	SVM
38	Type II	Classification	WLS-SVM, QPSO
29	Type II	Clustering, classification, neural network	K means, C4.5, back propagation
9	Type II	Classification	SVM
6	All types	Classification	SVM
5	Gestational	neutral network	Forward network, back propagation
15	Type II	Clustering, classification	K-means and logistic regression
17	Type II	Clustering, classification	C4.4, KNN, Random Forest, SVM
34	All types	Neural network, classification	(ANNs), logistic regress and decision tree
37	All types	Classification	SVM
23	Type II	Association rule	Apriority
7	Type II	Association rule	Apriority

8. Conclusion

In this work, a systematic effort was employed to identify and review data mining concept, tasks and model evaluation techniques, and KDDM model process. Likewise, research articles that employed data mining techniques for diagnosis of DM published with reputable journal publishers were equally reviewed. There are many significant research works done in almost all aspect of identification and diagnosis of DM. With increase of huge amount of electronic health data give rise towards further diagnosis and treatments of DM by leveraging data mining techniques especially SVM classification data mining algorithm in enriched datasets which may include both clinical and biological information. However, despite all its popularity, SVM algorithm should be further improved in the future work so as to further address its weaknesses and improve its efficiency.

9. Funding

This work was supported by the Tertiary Education Trust FUND (TETFUND), Nigeria, as an Institution Based Research Fund (IBR) for Federal University, Kashere, Gombe State, Nigeria.

10. References

1. Muhammad LJ,Sani S, Yakubu A, Yusuf MM, Elrufai TA, Mohammed IA, Nuhu AM. Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano –Wudil Highway, International Journal of Database Theory and Application. 2016; 10(2):197–206.
2. Ha S,BaeS, Park S. Web Mining for Distance Education, Proceeding of IEEE International Conference on Management of Innovation and Technology; 2000. p. 715–19.
3. Liao S, Chu P, Hsiao P. Data mining techniques and applications – A decade review from 2000 to 2011, Expert Systems with Applications, Elsevier. 2012; 39:11303–11. https://doi.org/10.7312/li--16274-040.
4. Tomar D, Agarwal S. Survey on data mining approaches for healthcare, SERSC, International Journal of Bio-Science and Bio-Technology. 2013; 5(5):241–66. https://doi.org/10.14257/ijbsbt.2013.5.5.25.
5. Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes, Procedia Computer Science. 2016; 82:115– 21. https://doi.org/10.1016/j.procs.2016.04.016.
6. Padawale SN, Jadhav BD. Survey on the various techniques used for the diagnosis of diabetes mellitus, IOSR, Journal of Electronics and Communication Engineering. 2015; 25–29.

7. Aljumah AA, Ahamad MG, Siddiqui MK. Application of data mining: Diabetes health care in young and old patients, Journal of King Saud University – Computer and Information Sciences, Elsevier. 2013; 25:127–36.
8. Ogbera AO, Ekpebegh C. Diabetes mellitus in Nigeria: The past, present and future, World Journal of Diabetes. 2014; 5(6):905–11. https://doi.org/10.4239/wjd.v5.i6.905. PMid: 25512795, PMCid: PMC4265879.
9. Shivakumar BL, Alby S. A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes. International Conference on Intelligent Computing Applications; 2014. p. 163–73. https://doi.org/10.1109/ICICA.2014.44.
10. Kurgan LA, Musilek PA. Survey of knowledge discovery and data mining process model, Cambridge University Press, the Knowledge Engineering Review. 2006; 21(1):1–24.
11. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases, Artificial Intelligence Magazine. 1996; 17:37–54.
12. Fayyad G, Piatesky-Shapiro G, Smyth P, Uthurusamy R. Advances in Knowledge Discovery and Data Mining, Cambridge, AAAI Press; 1996.
13. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, ChouvardaI. Machine learning and data mining methods in diabetes research, Computer Structure and Biotechnology Journal. 2017; 15(1):104–16. https://doi.org/10.1016/j.csbj.2016.12.005. PMid: 28138367, PMCid: PMC5257026.
14. Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A. Discovering Data Mining: From Concepts to Implementation. Prentice Hall Saddle River, New Jersey; 1998.
15. Luo Q. Advancing Knowledge Discovery and Data Mining. IEEE, Proceeding of Knowledge Discovery and Data Mining, Australia; 2008. p. 3–5.
16. Wasan SK, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics, Data Science Journal. 2006; 5:119–26. https://doi.org/10.2481/dsj.5.119.
17. Freitas AA. Data Mining and Knowledge Discovery with Evolutionary Algorithms. Berlin: Springer-Verlag; 2002. https://doi.org/10.1007/978-3-662-04923-5.
18. Ramageri RM. Data mining techniques and applications, Indian Journal of Computer Science and Engineering. 2016; 4:301–05.
19. Chi Y, Liu X, Xia K, Su C. An Intelligent diagnosis to type-2 diabetes based on QPSO algorithm and WLS-SVM, Intelligent Information Technology Application Workshops. 2018; 117–21. PMid: 29469163, PMCid: PMC5865475.
20. Sagar P, Prinima, Indu. Analysis of prediction techniques based on classification and regression, International Journal of Computer Applications. 2017; 163(7):47–51.
21. Nilashi M, Ibrahim O, Dalvi M, Ahmadi H, Shahmoradi L. Accuracy improvement for diabetes disease classification: A case on a public medical dataset, Fuzzy Information and Engineering. 2017; 9(3):345–57. https://doi.org/10.1016/j.fiae.2017.09.006.
22. Priya S, Rajalaxmi RR. An improved data mining model to predict the occurrence of type- 2 diabetes using Neural Network, International Journal of Computer Applications® (IJCA). 2012; 1–4.
23. Yang H, Huang S, Wang JX. Type-2 diabetes mellitus prediction model based on data mining, Informatics in Medicine Unlocked, Springer. 2018; 10(1):100–07.
24. Kasemthaweesab P, Kurutach W. Association analysis of Diabetes Mellitus (DM) with Complication States Based on Association Rules. Proceeding of the 7th IEEE Conference on Industrial Electronics and Applications; 2012. p. 1453–57. https://doi.org/10.1109/ICIEA.2012.6360952.
25. Patil BM, Joshi RC, Toshniwal D. Association Rule for Classification of Type-2 Diabetic Patients. Proceeding of Machine Learning and Computing (ICMLC) Second International Conference on Machine Learning and Computing; 2010. p. 330–34. https://doi.org/10.1109/ICMLC.2010.67. PMid: 20216911, PMCid: PMC2831780.
26. Yongjian F. Data Mining: Tasks, Techniques and Applications, IEEE Potentials. 1997; 16(4):1–12. https://doi.org/10.1109/45.624335.
27. Van MV, Vreeken J, Siebes A. Compression picks the item sets that matter. In: Proceedings of the ECML PKDD'06; 2006. p. 585–92.
28. Zhang GP. Neural Networks for Data Mining. In: Maimon O., Rokach L. (Eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA; 2009. p. 1–47. https://doi.org/10.1007/978-0-387-09823-4_21.
29. Craven M. Shavlik J. Using neural networks for data mining, Future Generation Computer Systems. 1997; 13:211–29. https://doi.org/10.1016/S0167-739X(97)00022-8.
30. Alade AO, Sowunmi OY, Misra S, Maskeliūnas R, Damaševičius R. A Neural Network Based Expert System for the Diagnosis of Diabetes Mellitus. International Conference on Information Technology Science; 2017. p. 14–22.
31. American Diabetes Association. Diagnosis and classification of diabetes mellitus, Diabetes Care. 2008; 31(1):55–60.
32. American Diabetes Association. Standards of medical care in diabetes, Diabetes Care. 2008 January; 31(1):12–54.
33. Georga EI, Protopappas VC, Fotiadis DI. Glucose Prediction in Type-1 and Type-2 Diabetic Patients Using

Data Driven Techniques. In: Funatsu, Ed., Knowledge-Oriented Applications in Data Mining; 2011. p. 1–22.
34. Yuan CZ, Isa D, Blanchfield P. A Hybrid data mining and case-based reasoning user modeling system (HDCU) for monitoring and predicting of blood sugar level. Proceeding in International Conference of Computer Science and Software Engineering; 2008 1. p. 653–56. https://doi.org/10.1109/CSSE.2008.1095.
35. Ilango BS, Ramaraj NA. Hybrid prediction model with F-score feature selection for type II Diabetes databases. A2CWiC '10 Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India; 2010.
36. Kandhasamy JP, Balamurali S. Performance analysis of classifier models to predict diabetes mellitus, Elsevier, Procedia Computer Science. 2015; 47:45–51. https://doi.org/10.1016/j.procs.2015.03.182.
37. Meng X, Huang Y, Rao D, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors, Kaohsiung Journal of Medical Sciences, Elsevier Taiwan. 2013; 29(2):93–99.
38. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes, BMC Medical Informatics and Decision Making. 2010; 10(1):16–10 https://doi.org/10.1186/1472-6947-10-16. PMid: 20307319, PMCid: PMC2850872.