Machine Learning and Data Mining in Diabetes Diagnosis and Treatment

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042049 doi:10.1088/1757-899X/490/4/042049

Machine Learning and Data Mining in Diabetes Diagnosis and Treatment

Bo He[1], Kuang-i Shu[2], Heng Zhang[3,*]

[1]School of computer and information science, Southwestern University, China
[2]School of computer and information science, Southwestern University, China
[3]School of computer and information science, Southwestern University, China

*Corresponding author e-mail: dahaizhangheng@163.com

Abstract. The remarkable progress of biotechnology and medical science has created the considerable amount of biomedical data. Diabetes mellitus (DM), a common chronic disease, has also been generated a large number of medical data in the process of diagnosis and treatment. So the exploration of medical data has become a hotpot. Nowadays, researchers are using machine learning to discover potentially valuable knowledge in medical data more than ever before. The purpose of this study is to systematically collate and review the application of machine learning, data mining method and supplement tools in the diabetes research field. Through sorting out, it was found that: a) the clinical data-sets were mainly used, b) about 82% of the articles based on diverse supervised machine learning method, c) deep learning method was widely used by researchers and the good experimental result have been achieved.

1. Introduction

Diabetes mellitus is a chronic disease that threatens human health severely. How to discover the resources of diabetic diagnosis and treatment data through machine learning and data mining is of great value for the treatment and recovery of the disease. Therefore, the application of machine learning and data mining method in DM research has attracted much attention. This paper will review these application in the research field of diabetes mellitus. The particular arrangements are as below:

The second, third, and fourth parts will introduce the related contents. The fifth part elaborates the research methods, the sixth part discusses, and the seventh part summarizes.

2. Diabetes Mellitus

Diabetes mellitus is a group of metabolic disorders. Symptoms are long-term high blood sugar levels in patients[1, 2].

Diabetes is usually divided into these types:

Type 1 DM (b-Cell Destruction), which accounts for only 5-10% of those with diabetes, results from the pancreas' failure to produce enough insulin due to loss of beta cells[3].

Type2 DM, which accounts for 90-95% of those with diabetes. The main clinical manifestation is insufficient insulin secretion or insulin resistance[4]. The cells concerned could not respond to insulin.

Several other types of diabetes can be attributed to many reasons.

All sorts of diabetes are at risk of complications due to long-term inability to cure completely[5, 6]. Adult diabetes patients will increase to 439 million by 2030, accounting for 7.7% of the

population.[7]. Nowadays, researchers have used machine learning and data mining methods to mine the medical data generated in the process of diseases, especially in diabetes ' diagnosis and treatment[8].

3. Machine Learning

ML(Machine Learning) link problems learned from data samples to general concepts of reasoning[8, 9]. The process of ML can be divided into: a) Discovering data dependencies from a given data set; b) Using the established dependencies to generate corresponding outputs for new inputs. According to the training methods of machine learning, it can be divided into the following three kinds[10]: a) supervised learning [11], b) unsupervised learning, c) semi-Supervised learning[12].

4. Data Mining and Knowledge Discovery in Database

Data mining is the process of discovering the value of information in large data sets involving multiple disciplines.[13]. The purpose is to derive strong association rules in the data set:

$$A_1 \wedge ... \wedge A_m \Rightarrow B_1 \wedge ... \wedge B_n \quad (1)$$

Where A_i (for $i \in \{1...m\}$) and B_j (for $j \in \{1...n\}$) are sets of attribute-values, from the relevant data sets in the database.

Knowledge discovery is mainly popular in the field of artificial intelligence and machine learning.The Knowledge Discovery (KDD) process is usually performed as follows: selection; data pre-processing; transformation; data mining; method evaluation[14]. which is depicted in Figure 1.

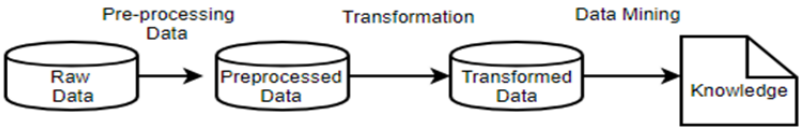

Figure 1. The step of the KDD process.

With the growth of biomedical data and the development of machine learning, the application of advanced technology to the information discovery of biomedical data has brought great help to the diagnosis and treatment of diabetes.

5. Research Method

This article searched for articles in the DBLP Computer Science Bibliography as of September 2018.[15]. Because machine learning is closely related to discovery, keywords “Diabetes ”are used to retrieve articles in the database in the process of searching articles.

After searching, we found 1398 articles with the keyword "Diabetes ", our research focuses on papers published in nearly four years ,and there are 476 articles. Figure 2 depicts the number of published articles that meet the search requirements each year.. After selection, we selected 70 articles. The classification and number of these articles are as follows: Prediction and Diagnosis in DM (27 articles); Complications (13 articles); Treatment methods (8 articles); Diabetes health management system (18 articles).

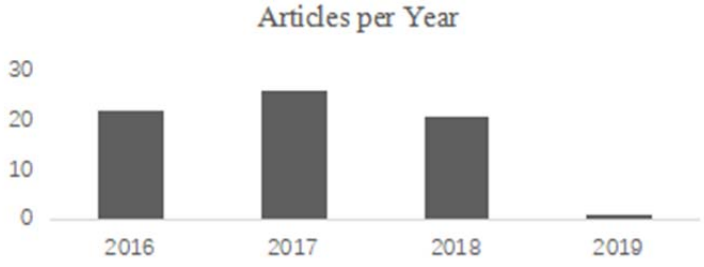

Figure 2. Numbers of articles collected per year that meet the search requirements.

5.1 Prediction and Diagnosis in DM

Diabetes mellitus is mainly detected by the following indicators: alpha-glycosylated hemoglobin (A1C) test, random blood sugar test, fasting blood sugar. Self-examination is mainly through blood collection to obtain blood sugar, mostly invasive detection. Researchers used the multi-modal similar and specific learning (MMSSL) approach to predict the occurrence of diabetes. The approach is more accurate and non-invasive[17]. Researchers propose to capture the facial and sublingual features of patients and extract the color, texture and geometric features of these three types of images, to achieve non-invasive diagnosis of patients. [16-20]Some researchers use electronic medical records and machine learning algorithm to predict the pathology of patients. Some of the biomarkers of patients can reflect the development of patients' condition. The commonly used biomarkers are metabolomics, proteomics, and genomics, which help to develop a large number of new biomarkers[21]. Researchers in the study used membrane fluidity of type 1 diabetes and the decision system that they have proposed to monitor the pathology of diabetes mellitus[22]. Applying data analysis to diabetes medical record analysis has improved the traditional recognition rate of type 2 diabetes [23].

At present, the prediction of diabetes mellitus is mainly through the use of machine learning methods to discover the case data of diabetic patients and to carry out biomarkers.

5.2Diabetic Complications

Hyperglycemia is the main pathophysiological feature of diabetes mellitus. Long-term hyperglycemia can cause tremendous damage to the patient's body. Researchers also predict the risk of complications by analyzing the patient's condition.

Because a large number of medical data show that patients with diabetes have a high risk of breast cancer[24].Researchers have proposed a hybrid machine learning solution to solve the problem of uneven data distribution. They put forward the value of machine learning in health care and epidemiological research.. This paper proposes a system that uses the support vector machine (SVM) combined with the PCA algorithm to extract features from dominant coronary morphology.[25]. A model for predicting hepatocellular carcinoma within six years after diagnosis of T2D was established[26]. Researchers proposed a new recursive graph method, Symbolic Recursive Map (SRP), and quantified the certainty of EEG by SRP_DET (deterministic SRP_DET). The recursive graph was used to determine the indicators of cognitive impairment in patients with type 2 diabetes [27].

5.3Drugs and Therapies

Patients with diabetes need drugs to adjust the high blood sugar level to the normal range. However, according to the current research situation, although these anti-diabetic drugs can effectively reduce the blood sugar level of patients and prolong the life span of patients, there are also some shortcomings: 1) there are side effects in many current treatment programs; 2) there are still challenges in exploring the pathological mechanism of diabetes mellitus, drug design and so on. However, with the accumulation of diabetic case data and the development of machine learning, advanced technology can be used to recommend more refined and personalized treatment plans for patients' specific conditions. Pathology can be further explored, and more effective anti-diabetic drugs can be designed. Researchers have overcome the inconvenience of the traditional patient treatment process requiring expert guidance for each insulin injection measurement[28]. A novel research suggested that consider the temporal sequences when recommending the bolus insulin doses instead of observing events in isolation. This method can effectively reduce the risk of hyperglycemia. Researchers have developed a non-invasive treatment for type 2 diabetes using ultrasound[29]. By establishing a new model describing the relationship between insulin and glucose, the researchers have explored the relationship between glucose and insulin, which provides a good reference value for other researchers[30].

5.4 Diabetic Health Management System

There is no cure for diabetes mellitus due to the increasing number of diabetic patients worldwide. So this disease will accompany the patient's life perennially.Every country and corresponding clinics

attach great importance to the health management of diabetic patients, which is used to help patients recover and enhance their sense of joy in life.

Researchers developed controllers based on patient's blood sugar level to control the patient's blood sugar level within the expected health scope; [31-33] Researchers mainly used external equipment to collect patient's physiological data information to establish models, so as to intervene in patient's clinical treatment and recovery. [34] The researchers designed a game where patients can record daily blood glucose measurements and nutrient intake into a digital diabetes diary. The researchers used the patient's electronic medical record to develop and train a Markov model for monitoring patients with myocardial infarction, stroke, retinopathy, kidney disease, diabetic foot syndrome, and hypoglycemia [35]. Researchers have established models to help doctors diagnose diabetes and have developed systems to monitor changes in the condition of people with diabetes [36,37].

6. Discussion

By sorting out the papers in recent years, we fund that most of the prediction of DM have high accuracy, generally more than 80%. About 82% of the studies used supervised learning, and the rest mainly used relevance analysis. Moreover, supervised learning methods such as decision tree algorithm, SVM algorithm and deep learning algorithm are the most common algorithms for DM prediction. Table 1 shows studies that compared more than four machine learning algorithms in various data-sets. SVM algorithm has achieved good prediction results in the diagnosis of diabetes mellitus. Researchers usually processed data sets into training sets and test sets for model training to obtain a general prediction model. Molecular biotechnology is used to study the pathology of diabetes mellitus, which has a guiding role in the treatment and recovery of diabetes mellitus. Many researchers also use advanced wearing equipment for remote monitoring, and use data mining method to obtain the influence of environment, diet and life rules that affect the development of diabetes mellitus.

Table 1. Comparison of various kind of algorithms.

Authors	Type of Diabetes	No. Of subject	Compared algorithms	Best Accuracy
Jinxing Li et al.[16]	Nonspecific	Healthy sample:548 DM samples:356	K-nearest neighbors(KNN); Support Vector Machine(SVM); Collaborative Representation Classifier(CRC)	SVM ACC=0.94
Malik et al.[23]	Type-2	175	SVM; Logistic Regression (LR); Artificial neural network (ANN)	SVM ACC=0.84
Jianfeng Zhang et al.[21]	Nonspecific	827	KNN; SVM; Naive Bayes	SVM ACC=0.90
Ermanno Cordelli et al.[22]	Type-1	25	SVM; Kernel KNN	SVM ACC=0.96

7. Conclusion

This paper summarized the application of machine learning and data mining in the diagnosis and treatment of diabetes mellitus. Researchers have made fruitful achievements in the field of diabetes diagnosis and treatment. Abundant diagnostic and therapeutic data is expected to give further in-depth research by using machine learning and data mining algorithms in the future.

Acknowledgments

This work was financially supported by SWU41015718 and SWU20710953.

References

[1] "About diabetes". World Health Organization. Archived from the original on 31 March 2014. Retrieved 4 April 2014.
[2] Shoback DG, Gardner D, eds. Greenspan's basic & clinical endocrinology (9th ed). New York: McGraw-Hill Medical.17. (2011)
[3] Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN. "Hyperglycemic crises in adult patients with diabetes". Diabetes Care. 32 (7),(2009): pp.1335–43.
[4] Amer Diabet Assoc,"Diagnosis and Classification of Diabetes Mellitus". DIABETES, (2010), CARE.33: pp.62-S69.
[5] "Diabetes Programme". World Health Organization. Archived from the original on 26 April 2014.
[6] "Diabetes Fact sheet N°312". WHO. October 2013. Archived from the original on 26 August 2013. Retrieved 25 March 2014.
[7] J.E.Shaw,R.A.SicreeP.Z.Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030", Diabetes Research and Clinical Practice, Volume 87,(2010) , 4-14.
[8] Bishop CM. Pattern recognition and machine learning. New York: Springer. (2006)
[9] Mitchell TM. The discipline of machine learning: Carnegie Mellon University. Carnegie Mellon University, School of Computer Science, Machine Learning Department. (2006)
[10] Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd Ed). Prentice Hall. ISBN 978-0137903955.
[11] Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.
[12] "Build with AI | DeepAI". DeepAI. Retrieved. (2018)
[13] "Data Mining Curriculum". ACM SIGKDD.(2014)
[14] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic"From Data Mining to Knowledge Discovery in Databases", International Conference on Scientific, 2002.
[15] "Records in DBLP ". Statistics. DBLP. (2018).
[16] Jinxing Li, Bob Zhang, Guangming Lu, Jane You:Body surface feature-based multi-modal Learning for Diabetes Mellitus detection. Inf. Sci. 2019，472: 1-14
[17] Jinxing Li, David Zhang, Yongcheng Li, Jian Wu:Multi-modal Fusion for Diabetes Mellitus and Impaired Glucose Regulation Detection. CoRR abs/1604.03443. (2016)
[18] Ting Shu, Bob Zhang, Yuan Yan Tang:An improved noninvasive method to detect Diabetes Mellitus using the Probabilistic Collaborative Representation based Classifier. Inf. Sci. 467: 477-488. (2018)
[19] Jinxing Li, David Zhang, Yongcheng Li:Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection. Inf. Sci. 384: 191-204. (2017)
[20] Ting Shu, Bob Zhang, Yuan Yan Tang: An extensive analysis of various texture feature extractors to detect Diabetes Mellitus using facial specific regions. Comp. in Bio. and Med. 83: 69-83. (2017)
[21] Jianfeng Z, Jiatuo X, Xiaojuan H, et al. Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images[J]. BioMed Research International, 2017, 2017:1-9.
[22] Ermanno Cordelli, Giuseppe Maulucci, A decision support system for type 1 diabetes mellitus diagnostics based on dual channel analysis of red blood cell membrane fluidity, Computer Methods and Programs in Biomedicine 162: 263-271. (2018)
[23] Malik S, Khadgawat R, Anand S, et al. Non-invasive detection of fasting blood glucose level via electrochemical measurement of saliva[J]. Springerplus, 2016, 5(1):701.
[24] Linglong Ye, Tian-Shyug Lee, Robert Chi: A Hybrid Machine Learning Scheme to Analyze the Risk Factors of Breast Cancer Outcome in Patients with Diabetes Mellitus. J. UCS ,2018,24(6): 665-681 .
[25] T.Araki, N. Ikeda, D. Shukla, P.K. Jain, N.D. SuriPCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology.

[26] Comput. Methods Programs Biomed., 128 (2016), pp. 137-158.
[27] Rau HH, Hsu CY, Lin YA, Atique S, Fuad A, Wei LM, et al. Development of a webbased liver cancer prediction model for type II diabetes patients by using an artificial neural network. Comput Methods Programs Biomed Mar 2016;125:58 – 65. http://dx.doi.org/10.1016/j.cmpb.2015.11.009 [Epub 2015 Nov 27].
[28] Dong Cui, Jinhuan Wang, Lei Wang, Shimin Yin, Zhijie Bian, Guanghua Gu: Symbol Recurrence Plots based resting-state eyes-closed EEG deterministic analysis onamnestic mild cognitive impairment in type 2 diabetes mellitus. Neurocomputing 203:102-110. (2016)
[29] Daniel Brown, Arantza Aldea, Rachel Harrison, Clare E. Martin, Ian Bayley: Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support. Artificial Intelligence in Medicine85: 28-42. (2018)
[30] Payam Sadeghi Shabestari, A Novel Approach to Numerical Modeling of Metabolic System: Investigation of Chaotic Behavior in Diabetes Mellitus. Complexity2018: 6815190:1-6815190:11. (2018)
[31] Mirza Mansoor Baig, Farhaan Mirza, Hamid Gholamhosseini, Jairo A. Gutiérrez, Ehsan Ullah:Clinical Decision Support for Early Detection of Prediabetes and Type 2 Diabetes Mellitus Using Wearable Technology. EMBC2018: 4456-4459.
[32] Hayat Al Mushcab, William George Kernohan, Jonathan G. Wallace, Roy Harper, Suzanne Martin:Self-Management of Diabetes Mellitus with Remote Monitoring: A Retrospective Review of 214 Cases. IJEHMC 8(1): 52-61. (2017)
[33] Özge Kart, Vildan Mevsim, Alp Kut: A mobile and web-based clinical decision support and monitoring system for diabetes mellitus patients in primary care: a study protocol for a randomized controlled trial. BMC Med. Inf. & Decision Making 17(1): 154:1-154:10. (2017)
[34] René Baranyi, Rainer Willinger, Nadja Lederer: DiaBeaThis - A gamified self-tracking portal to support people suffering from diabetes mellitus to control their blood glucose level. SeGAH 2018: 1-8.
[35] Wendelin Schramm, Fabian Sailer, Monika Pobiruchin, Christian Weiss: PROSIT Open Source Disease Models for Diabetes Mellitus.ICIMTH 2016: 115-118.
[36] Lucas Felipe Klein, Sandro José Rigo, Sílvio César Cazella, Ângela Jornada Ben: An Application for Mobile Devices Focused on Clinical Decision Support: Diabetes Mellitus Case. ISAmI 2016: 57-65.
[37] Vaclav Burda, Daniel Novák, Jakub Schneider: Evaluation of diabetes mellitus compensation after one year of using Mobiab system. EMBC 2016: 6002-6005.