Journal of Diabetes Science and Technology
Volume 5, Issue 6, November 2011

REVIEW ARTICLE

Data-Mining Technologies for Diabetes: A Systematic Review

Miroslav Marinov, M.S.,[1] Abu Saleh Mohammad Mosa, M.S.,[1] Illhoi Yoo, Ph.D.,[1,2] and Suzanne Austin Boren, Ph.D., MHA[1,2]

Abstract

Background:

The objective of this study is to conduct a systematic review of applications of data-mining techniques in the field of diabetes research.

Method:

We searched the MEDLINE database through PubMed. We initially identified 31 articles by the search, and selected 17 articles representing various data-mining methods used for diabetes research. Our main interest was to identify research goals, diabetes types, data sets, data-mining methods, data-mining software and technologies, and outcomes.

Results:

The applications 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 patients. The results could be used for both scientific research and real-life practice to improve the quality of health care diabetes patients.

Conclusions:

Data mining has played an important role in diabetes research. Data mining would be a valuable asset for diabetes researchers because it can unearth hidden knowledge from a huge amount of diabetes-related data. We believe that data mining can significantly help diabetes research and ultimately improve the quality of health care for diabetes patients.

J Diabetes Sci Technol 2011;5(6):1549-1556

Author Affiliations: [1]Informatics Institute, University of Missouri, Columbia, Missouri; and [2]Department of Health Management and Informatics, University of Missouri School of Medicine, Columbia, Missouri

Abbreviations: (CART) classification and regression tree, (BGL) blood glucose level, (FBG) fasting blood glucose, (HbA1c) glycated hemoglobin, (IB1) instance-based learning version 1, (ICU) intensive care unit, (MeSH) medical subject headings, (STS) structural time series, (T1DM) type 1 diabetes mellitus, (T2DM) type 2 diabetes mellitus, (TA) temporal abstraction

Keywords: blood glucose level, classification, data mining, diabetes mellitus, feature selection, systematic review

Corresponding Author: Illhoi Yoo, Ph.D., CE718 Clinical Support and Education Building, Five Hospital Dr., Columbia, MO 65212; email address *yooil@health.missouri.edu*

Introduction

Diabetes is often called a modern-society disease because widespread lack of regular exercise and rising obesity rates are some of the main contributing factors for it. According to data from the 2011 National Diabetes Fact Sheet,[1] 25.8 million people, or 8.3% of the U.S. population, have diabetes**.** The estimated total cost of diabetes in the United States for 2007 was $174 billion. Worldwide, the picture is very similar, with an estimated 285 million people affected by diabetes in 2010, representing 6.6% of the world's adult population. Health care expenditures for diabetes are expected to be $490 billion for 2030, accounting for 11.6% of the total health care expenditure in the world.[2]

As we can see from these facts, problems related to diabetes are many and quite costly. Diabetes is a very serious disease that, if not treated properly and on time, can lead to very serious complications, including death. This makes diabetes one of the main priorities in medical science research, which in turn generates huge amounts of data. Constantly increasing volumes of data are very well suited to be processed using data mining that can readily handle them. Using data-mining methods in diabetes research is one of the best ways to utilize large volumes of available diabetes-related data for extracting knowledge. Both descriptive (association and clustering) and predictive (classification) data-mining methods are used in the process. These data-mining methods are different from traditional statistic approaches in many ways (for details, refer to reference 3). One major difference between them is that the process of data mining is much more complex than that of statistical approaches.

Methods

Data Sources

MEDLINE (1947–2010) was searched using PubMed for eligible articles using the combination of the following terms: "diabetes mellitus" [medical subject headings (MeSH) major topic] and "data mining" (MeSH terms), "data" (all fields) and "mining" (all fields), or "data mining" (all fields). No limitation of publication date was used in the search; however, the earliest eligible article was published in 1999.

Inclusion and Exclusion Criteria

Inclusion criteria of the articles for this study were the use of various data-mining methods in the research and studies of diabetes. This study excluded the articles that do not use data mining for diabetes research, mentioned data mining as a side fact but never actually discussed it in detail, or were not published in English.

Study Selection and Data Extraction

Titles and abstracts of identified citations were screened based on inclusion and exclusion criteria described earlier. Potentially eligible articles were then reviewed in full text. Two of the investigators independently abstracted information from each article. Any discrepancies between the two investigators were resolved through discussion and reference to the original articles. Information extracted from the articles and presented in the table includes: (1) purpose of the study; (2) group/topic of research; (3) diabetes type; (4) data set used; (5) data-mining methods applied; (6) data-mining software and technology utilized; and (7) outcome of the data-mining application.

Results

A total of 31 articles were identified based on the searches conducted in MEDLINE from PubMed (**Figure 1**). Initially, articles were screened based on the title and abstract, and two articles were excluded, as both of the articles discussed diabetes mellitus affecting a mining town population as opposed to an application of data mining. The remaining 29 articles were reviewed in full text, and 12 additional articles were excluded based on the fact that there was no data mining used, no data-mining method was specified and discussed, or they were written in a language different than English. At the end, 17 articles complied with our eligibility criteria and were included in this study[4–20] (**Table 1**). The articles were grouped by the topic of research: interpretation and prediction of blood glucose level (BGL), features selection, genomic data analysis, and other.

Interpretation and Prediction of Blood Glucose Level

This group of studies aims to analyze and interpret the diabetes data and, based on the results, try to find ways to predict BGL. Four studies that analyzed BGL in diabetes patients were identified.[4–6,20] Two of the studies were on prediction of BGL,[6,20] one was on interpretation of BGL,[5] and another one was on both interpretation and prediction of BGL.[4]

Breault and colleagues[6] applied a classification and regression tree (CART) using the CART data-mining software (Salford Systems, San Diego, CA) on data of 15,902 diabetes patients and observed that the most important variable associated with bad glycemic control (HbA1c >9.5) is age. Patients below the threshold of 65.6 years old have worse glycemic control than older people, which was very surprising to the clinicians. Using this knowledge, they can target the specific age groups that are more likely to have poor glycemic control.

In order to predict next-morning fasting blood glucose (FBG), Yamaguchi and colleagues[20] used dataForest software (Yamatake Co., Japan)[21] to apply classification on a data set collected from four type 1 diabetes mellitus (T1DM) patients over a period of 150 days. The authors constructed a model for predicting next-morning FBG based on FBG, metabolic rate, food intake, and physical condition and concluded that the physical condition is highly correlated with FBG and its best variable to predict FBG.

Bellazzi and colleagues[5] used a combination of structural time series (STS) analysis, based on Bayesian network, and temporal abstraction (TA) to interpret past BGL data in order to extract and visualize the trends and daily cycles of BGL. First, data was analyzed with STS, with the results in the form of time varying series over a specific time period. Then, the second step was to apply TA on the results from the first step for further interpretation. At the end of the process, the final results were a trend diagram and a daily cycle diagram that visually represent the BGL.

Bellazzi and Abu-Hanna[4] applied two data-mining algorithms—TA and subgroup discovery—to predict the BGL on two types of patients—patients who self-monitor their BGL at home and intensive care unit (ICU) patients. This classification technique allows identification of the group of ICU patients with risk of diabetes and preventing treatments that can harm them.

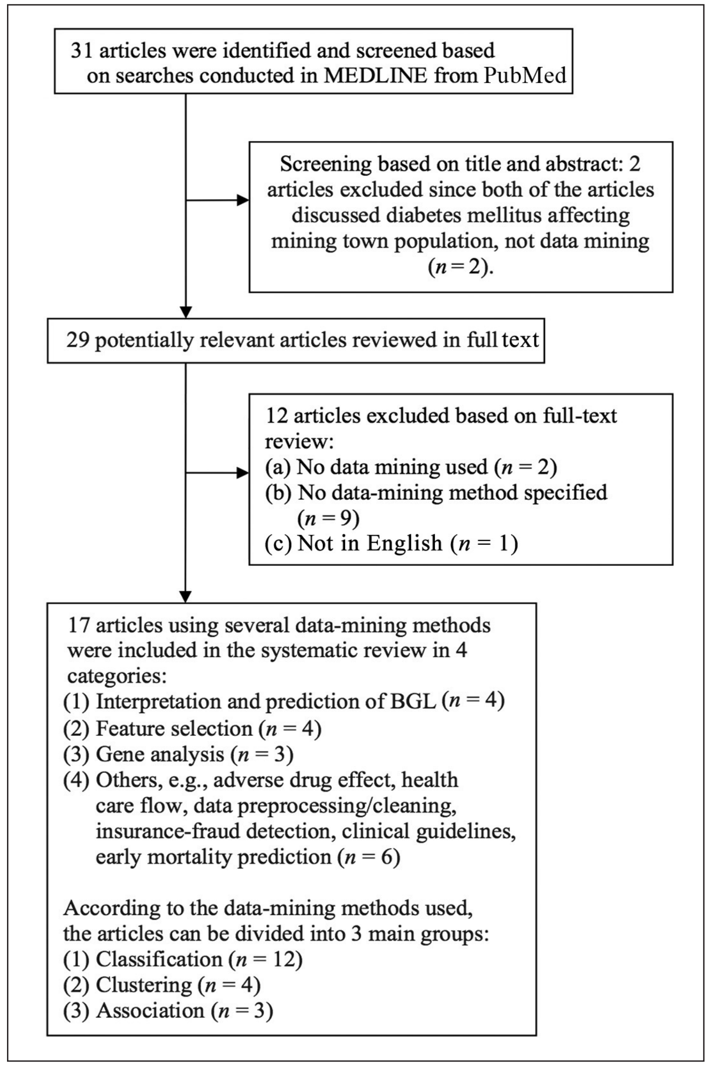

Figure 1. Trial flow diagram.

Table 1.
Data-Mining Technologies for Diabetes[a]

Author (Year)	Study Purpose	Group/Topic of Research	Diabetes Type	Data set	Data-Mining Methods	Software	Outcome
Bellazzi & Abu-Hanna, 2009[4]	Patient need	Interpretation and prediction of BGL	N/A	Blood glucose home-monitoring data, ICU blood glucose data	Association/ Temporal abstraction, Classification/ Subgroup discovery	N/A	Trends and daily cycles of BGL, predict high levels of BGL
Bellazzi *et al.*, 1998[5]	Patient need	Interpretation of BGL	N/A	Blood glucose home-monitoring data	Association/Temporal abstraction	N/A	Trends and daily cycles of BGL
Breault *et al.*, 2002[6]	Science research	Prediction of BGL	N/A	15,902 patients with diabetes	Classification/CART	CART software by Salford Systems	Best predictor and rules to predict glycemic control

Continued →

Table 1. Continued

Author (Year)	Study Purpose	Group/Topic of Research	Diabetes Type	Data set	Data-Mining Methods	Software	Outcome
Brown *et al.*, 2005[7]	Science research	Genomic data analysis	T2DM	LocusLink database	Clustering	ExQuest	Candidate genes that contribute to diabesity
Concaro *et al.*, 2009[8]	Science research	Healthcare flow	N/A	101,339 health care events	Association/Temporal abstraction	N/A	Temporal association rules on sequences of hybrid events
Covani *et al.*, 2009[9]	Science research	Genomic data analysis	T2DM	Gene list associated with T2DM, periodontitis, and sinusitis	Clustering/ Hierarchical, k-means	STRING 7.0	Leader genes and interactions among them
Duhamel *et al.*, 2003[10]	Science research	Data preprocessing/ cleaning	T2DM	23,601 records of T2DM patients	Clustering/ k-means, Classification/Decision tree	Spina	Cleaned data
DuMouchel *et al.*, 2008[11]	Science research	Adverse drug effect	N/A	2.4 million reports from FDA AERS database	Classification/ Proportional Reporting Ratio, Bayes Multi-Item Gamma Poison Shrinker, Logistic regression	N/A	Diabetes-related adverse effect associated with antipsychotic drugs
Gerling *et al*, 2006[12]	Science research	Genomic data analysis	T1DM	2D gel proteome data	Clustering/k-means, principal component analysis	GeneSpring version 6.2	37 differentially expressed spots
Huang *et al*, 2007[13]	Science research	Feature selection	T2DM	2064 patient information: 1148 male, 916 female	Classification/Naïve Bayes, IB1, Decision tree—C4.5	N/A	Best predictor for each algorithm
Liou *et al*, 2008[14]	Science research	Insurance-fraud detection	N/A	Taiwan's national health insurance database	Classification/Neural Network, Classification Tree	SPSS Clementine 7	Fraudulent health care provider
Miyaki *et al*, 2002[15]	Science research	Feature selection	T2DM	165 patient's records	Classification/CART	SPSS Answer Tree 2.1 (IBM, Chicago, IL)	Best predictor
Richards *et al*, 2001[16]	Science research	Prediction of early mortality	N/A	21,000 patient's clinical records	Classification/ Simulated annealing	Lanner data-mining package	Association between patient's observation and early mortality
Sigurdardottir *et al*, 2007[17]	Science research	Feature selection	T2DM	21 articles from Medline, Scopus, and CINAHL	Classification/Decision tree—C4.5	WEKA	Factors predict changes in HbA1c level
Toussi *et al*, 2009[18]	Science research	Clinical guideline	T2DM	Patient records with missing or incomplete rules in guideline	Classification/Decision tree—C5.0	SPSS Clementine 10.1	New rules to enrich guideline
Wright *et al*, 2005[19]	Science research	Feature selection	T1DM, T2DM	MQIC data warehouse: 50,428 records extracted for this study from 3.6 million records	Classification/ Reconstructibility analysis (RA)	OCCAM with RA package	Risk factor for diabetes mellitus
Yamaguchi *et al*, 2006[20]	Patient need	Predict BGL	T1DM	FBG, metabolic rate, food intake, and physical condition	Classification	dataFOREST	Next-morning FBG

[a] N/A, not applicable

Features Selection

There are many factors that influence diabetes and the different aspects of the disease, and quite often it is unfeasible to take all in consideration. The feature selection method addresses this problem by selecting a few factors that are the most influential for each particular case. Four of the studies were conducted in order to identify a set of the best predictors from the data set of diabetes patients.[13,15,17,19]

Huang and colleagues[13] applied Naïve Bayes, IB1classifier, and CART C4.5 on information collected from 2,064 patients (1,148 males and 916 females), and identified the five most important factors that influence blood glucose control: (1) age; (2) diagnosis duration; (3) need for insulin treatment; (4) random blood glucose measurements; and (5) diet treatment. Using these five factors, 95% predictive accuracy and 98% sensitivity was achieved.

Myiaki and colleagues[15] conducted a study to find the best predictors for diabetes vascular complications using CART on data from 165 type 2 diabetes mellitus (T2DM) patients. The authors found that age (cutoff: 65.4 years) was the best predictor, and depending on the age, the second best predictor was body weight (cutoff: 53.9kg) for the group above 65.4 or systolic blood pressure for the group below 65.4.

Sigurdardottir and colleagues[17] reviewed outcomes of educational interventions in T2DM reported in 21 articles. The results were analyzed with CART C4.5 using WEKA (Waikato Environment for Knowledge Analysis) data-mining software in order to identify the best factors that predict changes in glycated hemoglobin (HbA1c) level. This study found that the effect of the factor "diabetic education intervention" in diabetes treatment is significant and achieved a notable drop (.8–2.5%) in HbA1c levels. This study also suggested that other factors such as duration, education content, and intensity of education have no impact on changes in HbA1c.

Wright and colleagues[19] applied reconstructibility analysis on de-identified data from 50,428 T1DM or T2DM patients to identify risk factors for various complications of diabetes. This study discovered that elevated urine microalbumin level is the best predictor of microalbuminuria (serious nephrologic complication of diabetes).

Genomic Data Analysis

Three studies applied data-mining methods for genomic data analysis related to diabetes.[7,9,12]

A data-mining method called leader-gene approach was applied by Covani and colleagues[9] in order to study the association between periodontitis (a disease related to inflammation and infection of the ligaments and bones that support the teeth) and T2DM, and found that periodontitis and T2DM share four leader genes. The leader-gene method clusters genes from a list according to their weighted number of links using hierarchical or k-means clustering, and genes belonging to the highest ranked cluster are identified as leader genes. They also found that sinusitis (a local inflammatory infectious disease like periodontitis) and periodontitis do not share any leader genes. The outcome of the study lacks experimental validation, as it is purely theoretical. However, it provides new hypothesis for further experimentation.

A data-mining study has been carried out by Brown and colleagues[7] in order to identify candidate genes that contribute to T2DM using the gene-expression mining tool ExQuest[22]. The genes selected by ExQuest were assessed by searching PubMed for articles associated with the gene names. It was found that a number of candidate genes, revealed by previous study, have little or no prior association with diabetes. This study provided a theoretical method of identifying genes related to diabetes that can also be applied for identifying candidate genes related to other genetic diseases. However, no experimental validation was provided.

Another data-mining study by Gerling and colleagues[12] demonstrated a new approach for analyzing two-dimensional gel-proteome data using GeneSpring (version 6.2, Sillcon Genetics, Redwood, CA) software. Clustering algorithm (k-means, self-organizing maps, and hierarchical) is applied on the data to form groups of similar gene expression and reveal correlations among the genes that may be unnoticed otherwise.

Other Studies

Six studies analyzed diabetes health care and clinical data for various purposes such as health care flow analysis, data preprocessing/cleaning, adverse drug effect analysis, insurance fraud detection, clinical guidelines enrichment, and prediction of early mortality.[8,10,11,14,16,18]

Concaro and colleagues[7] used TAs in a large clinical diabetes database containing 101,339 events to generate temporal association rules based on sequences of hybrid events. An example of such a rule is, "There is 56% probability of increase in glycemia during the follow-up visit of an overweight subject who is under antihypertensive therapy and showing a normal glycemic value and high HbA1c ." In extracting diagnostic and therapeutic pattern from health care data, the application of TA is very useful because it takes into account the time-element factor. However, this study was very specific to the data set used and did not provide a general method appropriate for any health care data set.

Data preprocessing and cleaning is a very important step in the process of knowledge discovery in data. Duhamel and colleagues[10] demonstrated a five-step preprocessing method for improving data mining in a large clinical data set containing information on 23,601 T2DM patients. The poorly filled fields (i.e., containing numerous missing values) are identified by applying k-means clustering algorithm. In order to study and handle the missing data, the decision tree was implemented and rules were produced to impute missing values.

Diabetes-related adverse drug effect was analyzed by DuMouchel and colleagues.[11] This study measured the association with various glycemic events from Food and Drug Administration AERS database by applying three methods: (1) proportional reporting ratio; (2) Bayes multi-item gamma poison shrinker; and (3) logistic regression. The authors found that the association differed for different drugs (i.e., low association—ziprasidolle, aripiprazole, haloperidol, and risperidone; medium association—quetiapine; and strong association—clozapine and olanzapine). Even though they could not generate a "class effect" hypothesis, the authors provided enough results to support future studies on the subject.

The detection of fraudulent insurance claims was conducted by Liou and colleagues[14] on a data set containing random sample of data from diabetes patients and health care providers from Taiwan's national health insurance database. This study identified fraudulent health care providers by applying logistic regression, neural networks, and classification trees C5.0 using Clementine 7 data-mining software (IBM, Chicago, IL). Average accuracy of the classification-tree algorithm was highest (99.37%) among 3 data-mining techniques applied, and presented the optimal solution.

Richards and colleagues[16] applied simulated annealing for predicting early mortality on a data set containing 21,000 diabetes patients' clinical records using Lanner data-mining package. The study generated a final rule set that includes six rules as indicators of early mortality. Among them, the most significant sign of early mortality was diabetic neuropathy, which was unknown to clinicians and was confirmed in a later study. However, the findings of the study were very specific to the data set used, and future studies are required to determine generalized rules.

Toussi and colleagues[18] found that clinical guidelines always stay behind actual practice. The authors analyzed records of patients corresponding to the clinical condition not covered in the clinical guidelines by applying the C5.0 decision-tree algorithm using SPPS Clementine (version 10.1, IBM, Chicago, IL) data-mining software, and generated 27 new rules based on the physicians' prescriptions. These new rules were compared with the original guidelines, which showed similarities with those added to the newer version. Although the presented method may be useful for generating clinical guidelines, it is based only on the decisions of physicians, and not on the evidence from scientific study. As such, the proposed method can be applied only as the opinion of clinical experts when there is lack in clinical guidelines.

Discussion

Diabetes affects a considerable number of people in the United States as well as worldwide. This, in turn, could generate a large amount of patient-related information. The use of data-mining methods and technologies to analyze these large quantities of data is well suited to discovering new knowledge in the area of diabetes research and providing better health care in everyday practice.

Implication for Practice

Diabetes is a lifelong disease, and one of the main goals in its treatment is to keep the BGL within normal limits, thus making controlling the BGL a part of patients' everyday life. Four of the 17 articles applied data-mining techniques in order to achieve better control of the BGL. Data-mining methods were found to be very useful for interpretation and prediction of BGL. However, the outcome of current research is usually limited to the data set used and the lack of ability to produce universal prediction rules applicable on other data sets related to diabetes.

The diabetes data set may contain a large amount of information for every patient including personal, clinical, and social information. These factors may not have a significant effect on the BGL, and thus in predicting BGL. Four articles discussed application of data-mining methods for feature selection, choosing the best factors in predicting the BGL, and the effects each of them have on the level of blood glucose. However, the results of these articles only contain the age of patients as the common best predictor. Further research is required to produce a universal set of best predictors of BGL. Such a set could also be used to generate hypotheses for future research in order to identify the best factors for controlling BGL.

The use of data-mining methods were also found to be very useful in analysis of diabetes-related health care flow and adverse drug effects, preprocessing/cleaning diabetes-related data for mining, detecting diabetes-related fraudulent insurance claims, enriching existing diabetes-related clinical guidelines by incorporating new guidelines, or detecting signs of early mortality. However, all studies were very specific to the data set used, and similar studies using different data sets may reveal a gold standard for analyzing diabetes-related data sets for extracting important knowledge.

Implication for Research

Genomic data analysis is also important in the area of diabetes research because diabetes is a genetic disease. Several data-mining software for gene expression analysis are available, such as ExQuest, GeneSpring, and STRING. Three studies applied data-mining methods for diabetes-related genomic data analysis. The results of these research studies had no experimental validation, but provided important hypothesis for future experimentation.

Limitations

The articles for this systematic literature review were selected exclusively from MEDLINE based on the search query described. We wanted to find and analyze all diabetes-related articles that used the term "data mining" in the biggest biomedical data set in the world. If authors did not use these terms, we considered the article irrelevant to our case.

Conclusions

Using data mining to deal with the avalanche of clinical data collected from patients and generated from the research and management of diabetes is a valuable asset that can help researchers and clinicians provide better health care for the patients affected by this modern-society disease. Data-mining techniques are becoming more widely used in the field of diabetes—12 of the articles were published after 2005, and 4 were published in 2009 alone. This is one more confirmation that data mining in biomedicine has a good future and will be used more and more in the area of diabetes in particular.

References:

1. Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Available from: http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf. Accessed April, 1 2011.
2. IDF Diabetes Atlas. The Economic Impacts of Diabetes. Available from: http://www.diabetesatlas.org/content/economic-impacts-diabetes. Accessed April 1, 2011.
3. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L. Data Mining in Healthcare and Biomedicine: A Survey of the Literature. J Med Syst. 2011 May 3. Epub ahead of print.
4. Bellazzi R, Abu-Hanna A. Data mining technologies for blood glucose and diabetes management. J Diabetes Sci Technol. 2009;3(3):603–12.
5. Bellazzi R, Magni P, Larizza C, De Nicolao G, Riva A, Stefanelli M. Mining biomedical time series by combining structural analysis and temporal abstractions. Proc AMIA Symp. 1998:160–4.
6. Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. Artif Intell Med. 2002;26(1–2):37–54.
7. Brown AC, Olver WI, Donnelly CJ, May ME, Naggert JK, Shaffer DJ, Roopenian DC. Searching QTL by gene expression: analysis of diabesity. BMC Genet. 2005;6:12.
8. Concaro S, Sacchi L, Cerra C, Bellazzi R. Mining administrative and clinical diabetes data with temporal association rules. Stud Health Technol Inform. 2009;150:574–8.
9. Covani U, Marconcini S, Derchi G, Barone A, Giacomelli L. Relationship between human periodontitis and type 2 diabetes at a genomic level: a data-mining study. J Periodontol. 2009;80(8):1265–73.
10. Duhamel A, Nuttens MC, Devos P, Picavet M, Beuscart R. A preprocessing method for improving data mining techniques. Application to a large medical diabetes database. Stud Health Technol Inform. 2003;95:269–74.
11. DuMouchel W, Fram D, Yang X, Mahmoud RA, Grogg AL, Engelhart L, Ramaswamy K. Antipsychotics, glycemic disorders, and life-threatening diabetic events: a Bayesian data-mining analysis of the FDA adverse event reporting system (1968-2004). Ann Clin Psychiatry. 2008;20(1):21–31.
12. Gerling IC, Singh S, Lenchik NI, Marshall DR, Wu J. New data analysis and mining approaches identify unique proteome and transcriptome markers of susceptibility to autoimmune diabetes. Mol Cell Proteomics. 2006;5(2):293–305.
13. Huang Y, McCullagh P, Black N, Harper R. Feature selection and classification model construction on type 2 diabetic patients' data. Artif Intell Med. 2007;41(3):251–62.
14. Liou FM, Tang YC, Chen JY. Detecting hospital fraud and claim abuse through diabetic outpatient services. Health Care Manag Sci. 2008;11(4):353–8.
15. Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. J Epidemiol. 2002;12(3):243–8.
16. Richards G, Rayward-Smith VJ, Sönksen PH, Carey S, Weng C. Data mining for indicators of early mortality in a database of clinical records. Artif Intell Med. 2001;22(3):215–31.
17. Sigurdardottir AK, Jonsdottir H, Benediktsson R. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. Patient Educ Couns. 2007;67(1–2):21–31.
18. Toussi M, Lamy JB, Le Toumelin P, Venot A. Using data mining techniques to explore physicians' therapeutic decisions when clinical guidelines do not provide recommendations: methods and example for type 2 diabetes. BMC Med Inform Decis Mak. 2009;9:28.

19. Wright A, Ricciardi TN, Zwick M. Application of information-theoretic data mining techniques in a national ambulatory practice outcomes research network. AMIA Annu Symp Proc; 2005;829–33.

20. Yamaguchi M, Kaseda C, Yamazaki K, Kobayashi M. Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining. Med Biol Eng Comput. 2006;44(6):451–7.

21. Tsutsui H, Kurosaki A, Sato T. Nonlinear Modeling Technique Using Historical Data for Case TCBM: Topological Case Based Modeling. Transactions-Society of Instrument and Control Engineers. 1997;33(9):947–54.

22. Brown AC, Kai K, May ME, Brown DC, Roopenian DC. ExQuest, a novel method for displaying quantitative gene expression from ESTs. Genomics. 2004;83(3):528–39.