Journal of King Saud University – Computer and Information Sciences (2013) **25**, 127–136

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Application of data mining: Diabetes health care in young and old patients

Abdullah A. Aljumah, Mohammed Gulam Ahamad, Mohammad Khubeb Siddiqui *

College of Computer Engineering and Sciences, Salman bin Abdulaziz University, Saudi Arabia

Received 21 May 2012; revised 2 September 2012; accepted 1 October 2012
Available online 23 October 2012

KEYWORDS

Data mining;
Oracle data mining tool;
Prediction;
Regression;
Support vector machine;
Diabetes treatment

Abstract This research concentrates upon predictive analysis of diabetic treatment using a regression-based data mining technique. The Oracle Data Miner (ODM) was employed as a software mining tool for predicting modes of treating diabetes. The support vector machine algorithm was used for experimental analysis. Datasets of Non Communicable Diseases (NCD) risk factors in Saudi Arabia were obtained from the World Health Organization (WHO) and used for analysis. The dataset was studied and analyzed to identify effectiveness of different treatment types for different age groups. The five age groups are consolidated into two age groups, denoted as $p(y)$ and $p(o)$ for the young and old age groups, respectively. Preferential orders of treatment were investigated. We conclude that drug treatment for patients in the young age group can be delayed to avoid side effects. In contrast, patients in the old age group should be prescribed drug treatment immediately, along with other treatments, because there are no other alternatives available.

1. Introduction

There was a time when data were not readily available. As data became more abundant, however, limitations in computational capabilities prevented the practical application of mathematical models. At present, not only are data available for analysis but computational resources are capable of supporting a variety of sophisticated methods. Consequently, data mining tools are now being used for clinical data. The bottleneck in data analysis is now raising the most appropriate clinical questions and using proper data and analysis techniques to obtain clinically relevant answers.

In this paper, we have generated reports through the use of data mining tools on a pre-compiled dataset for non-communicable diseases in Saudi Arabia.

Data mining is the process of selecting, exploring and modeling large amounts of data. This process has become an increasingly pervasive activity in all areas of medical science research. Data mining has resulted in the discovery of useful hidden patterns from massive databases. Data mining problems are often solved using different approaches from both computer sciences, such as multi-dimensional databases, machine learning, soft computing and data visualization; and statistics,

* Corresponding author. Tel.: +966 547006246.
E-mail addresses: aljumah@sau.edu.sa (A.A. Aljumah), gulamahamad@hotmail.com (M.G. Ahamad), m.khubeb@sau.edu.sa (M.K. Siddiqui).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578 © 2013 Production and hosting by Elsevier B.V. on behalf of King Saud University.
http://dx.doi.org/10.1016/j.jksuci.2012.10.003

including hypothesis testing, clustering, classification, and regression techniques.

Diabetes is a major health problem in Saudi Arabia. Diabetes is the most common endocrine disease across all population and age groups. This disease has become the fourth leading cause of death in developed countries and there is substantial evidence that it is reaching epidemic proportions in many developing and newly industrialized nations (Gan, 2003). Recent research in Saudi Arabia shows that the number of patients with Diabetes Mellitus is increasing drastically.

The following six types of treatments were identified in the 2005 World Health Organization's NCD report of Ministry of Health, Saudi Arabia (http://www.emro.who.int/ncd/pdf/stepwise_saa_05.pdf, 2005) and are discussed below:

(a) Drug
(b) Diet
(c) Weight reduction
(d) Smoke cessation
(e) Exercise
(f) Insulin

A. Drug: Oral medications, in the form of tablets help to control blood sugar levels in patients whose bodies still produce some insulin (the majority of people with type 2 diabetes). Drugs are usually prescribed to patients with diabetes (type 2) along with recommendations for making specific dietary changes and getting regular exercise. Several drugs are often used in combination to achieve optimal blood sugar control.

B. Diet: Patients with diabetes should maintain consistency in both food intake timings and the types of food they choose. Dietary consistency helps patients to prevent blood sugar levels from extreme highs and lows. Meal planning includes choosing nutritious foods and eating the right amount of food at the right time. Patients should consult regularly with their doctors and registered dieticians to learn how much fat, protein, and carbohydrates are needed. Meal plans should be selected to fit daily lifestyles and habits.

C. Weight reduction: One of the most important remedies for diabetes is weight reduction. Weight reduction increases the body's sensitivity to insulin and helps to control blood sugar levels.

D. Smoke cessation: Smoking is one of the causes for uncontrolled diabetes (http://medweb.bham.ac.uk/easdec/prevention/smoking_and_diabetes.htm#help, 2011). Smoking doubles the damage that diabetes causes to the body by hardening the arteries. Smoking augments the risk of diabetes.

E. Exercise: Exercise is immensely important for managing diabetes. Combining diet, exercise, and drugs (when prescribed) will help to control weight and blood sugar levels. Exercise helps control diabetes by improving the body's use of insulin. Exercise also helps to burning excess body fat and control weight.

F. Insulin: Many people with diabetes must take insulin to manage their disease.

Diabetes is a particularly opportune disease for data mining for a number of reasons (Breault et al., 2002). First, many diabetic databases with historic patient information are available. Second, new knowledge about treatment patterns of diabetes can help save money. Diabetes can also produce terrible afflictions, such as blindness, kidney failure, and heart failure. Finally, physicians need to know how to quickly identify and diagnose potential cases.

In the present investigation we apply regression based data mining techniques to diabetes data from Saudi Arabia in 2005. The goal of this investigation is to use data mining techniques to discover patterns that identify the best mode of treatment for diabetes across different age groups in Saudi Arabia.

2. Related work

A literature survey reveals many results on diabetes, and specifically the diabetes problem in United States. The diabetic data warehouse was formed by a large integrated health care system in the New Orleans area with 30,383 diabetic patients. The classification and regression tree approach was used to analyze these data sets (Breault et al., 2002). The diabetes in Saudi Arabia has also been investigated. Researchers found that the overall prevalence of diabetes in adults in KSA was 23.7%. A longitudinal study was recommended to demonstrate the importance of modifying risk factors for the development of diabetes and reducing the prevalence of diabetes in KSA (Al-Nozha et al., 2004). Data mining indicated that education level did not predict changes in HbA1c levels (Sigurdardottir et al., 2007). The status of insulin therapy for preschool-age children with type 1 diabetes has also been studied. In this study, conducted with data acquired over a 10 year period (1993–2002), the daily insulin therapy and episodes of severe hypoglycemia were identified in a population of 142 patients diagnosed with diabetes at less than 5 years of age (Yokotaa et al., 2005). A study of diabetes was conducted in France in 2001 to characterize the therapeutic management and control of diabetes and modifiable cardiovascular risk factors in patients with type 2 diabetes receiving specialist care. The study was proposed to 575 diabetologists across France. This is due in part to the severity of diabetes in the patients seen by specialists in diabetes care; however, both awareness and application of published recommendations need to be reinforced (Charpentier et al., 2003). The diabetic control study in Asia and Western Pacific Region indicates that the incidence of type 1 diabetes is increasing in many parts of Asia, where resources may not enable targets for glycemic control to be achieved. The aims of this study were to describe glycemic control, diabetes care, and complications in youth with type 1 diabetes from the Western Pacific Region and to identify factors associated with glycemic control and hypoglycemia. A cross-sectional clinic-based study on 2312 children and adolescents (aged 18 years; 45% males) from 96 pediatric diabetes centers in Australia, China, Hong Kong, Indonesia, Japan, Malaysia, Philippines, Singapore, South Korea, Taiwan, and Thailand was also conducted. Clinical and management details were recorded, and finger-pricked blood samples were obtained for central glycated hemoglobin (HbA1c) (Craiga et al., 2007). The prevalence of diabetes mellitus, islet auto antibodies, type 2 diabetes, and LADA were all found to be elevated in the south of Spain (Soriguer-Escofet et al., 2002). Diabetes is divided into two subgroups based on the specific mechanisms causing the disease: diabetes in which genetic susceptibility is clarified at the DNA level and diabetes associated with other diseases or conditions (Richards et al., 2001). Smoking habits were reported to the Swedish National Diabetes Register (NDR) to study associations between smoking,

glycemic control, and micro albuminuria and identify trends associated with smoking habits. Studies concluded that smoking habits in patients with diabetes were widespread, especially in young females with type 1 diabetes, and in middle-aged type 1 and type 2 diabetes patients. The study recommended that these groups be targeted for smoking cessation campaigns. Smoking habits were also associated with both poor glycemic control and microalbuminuria, which were found to be independent of other study characteristics (Nilsson et al., 2004). Weight loss goals among participants enrolled in an adapted Diabetes Prevention Program (DPP) were also investigated. In a real-world translation of the DPP, lifestyle intervention participants who achieved their weight loss goal were more likely to have monitored their dietary intake and frequency and were also more likely to have increased their physical activity markedly. These findings highlight the importance of supporting participants in lifestyle interventions to initiate and maintain dietary self-monitoring and to increase levels of physical activity (Harwell et al., 2010). A study has also been conducted to carry out the prediction analysis for treating hypertension using regression-based data mining techniques (Almazyad et al., 2010).

3. Material and methods

3.1. Data collection

The 2005 dataset is a standard NCD risk factor report from the Ministry of Health, Saudi Arabia. It is freely available from the World Health Organization's web link titled 'Data and Statistics' (http://www.emro.who.int/ncd/pdf/stepwise_saa_05.pdf, 2005). Six tables are available in the Oracle 10g database named 'drug', 'diet', 'weight', 'smoke_cessation', 'exercise' and 'insulin' (Tables 1–6). Each table corresponds to a type of diabetes treatment in Saudi Arabia diabetes patients and includes six columns (Sr_No, Age, N, Small_n, Percentage, SE): the 'Sr_No' column lists serial numbers that serve as a primary keys for the table, 'Age' indicates the age range of patients, 'N' provides the total number of patients in each age group, 'Small_n' shows the number of patients whose treatment was effective in controlling the disease, 'Percentage' indicates the percentage of patients whose diabetes was effectively controlled, and 'SE' indicates the standard error.

3.2. Tools and techniques

Various types of data mining tools are currently available and each has its own merits and demerits. For the analysis of WHO's NCD report on Saudi Arabia, we have concentrated on diabetic data and used predictive regression analysis to investigate which mode of treatment is more effective for each age group. We used the Oracle Data Miner version 10.2.0.3.0.1; build 2007 for data mining, which acts as a client to the Oracle 10g database server, release 10.2.0.3.0.

Oracle Data Miner release 10.1 is a graphical user interface that helps data analysts mine data in Oracle databases to find valuable hidden information, patterns, and insights. Data analysts can mine data with Oracle Data Miner's easy-to-use wizards that guide them through data preparation, data mining, model evaluation, and model scoring. As data analysts transform the data, build models, and interpret results, Oracle Data Miner can automatically generate the code needed to

Table 1 Drug.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	3	28.6	11.7
2	25–34	13	7	52.4	13.2
3	35–44	70	46	65.2	7.2
4	45–54	130	96	73.7	4.7
5	55–64	142	102	71.8	3.9

Table 2 Diet.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	3	27.2	12.7
2	25–34	13	9	69.8	13.3
3	35–44	70	40	56.6	7.3
4	45–54	130	88	67.8	4.8
5	55–64	142	88	62.2	6

Table 3 Weight.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	3	27.3	12.7
2	25–34	13	5	40.7	12.9
3	35–44	70	30	43.2	6.1
4	45–54	130	54	41.4	5.8
5	55–64	142	42	29.4	3.8

Table 4 Smoke cessation.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	2	20.1	11.9
2	25–34	13	2	15.6	9.4
3	35–44	70	13	18.7	4.5
4	45–54	130	20	15	3.6
5	55–64	142	14	9.8	2.6

Table 5 Exercise.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	6	52.8	15.3
2	25–34	13	5	22.9	11.9
3	35–44	70	32	13.6	4.3
4	45–54	130	55	17.2	4.2
5	55–64	142	44	20.7	3.3

Table 6 Insulin.

Sr_No	Age	*N*	Small_n	Percentage	SE
1	15–24	11	6	20.1	11.9
2	25–34	13	3	15.6	9.4
3	35–44	70	10	18.7	4.5
4	45–54	130	22	15	3.6
5	55–64	142	29	9.8	2.6

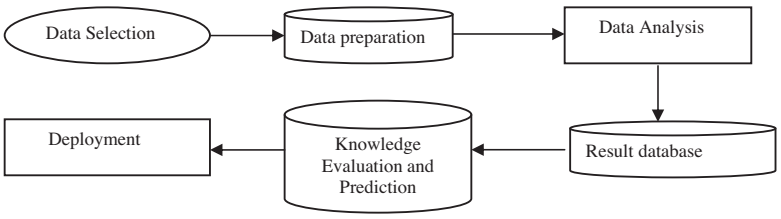


Figure 1 Data mining architecture of diabetes.

transform data mining steps into an integrated data mining/BI application (Oracle Technology Network tool: http://www.oracle.com/technology/products/bi/odm/odminer.html).

The data mining process for identifying the most effective mode of treatment for each age group, particularly for younger and older age patients, was divided into six steps. The processing blocks are shown in Fig. 1: '**Data Mining Architecture of Diabetes**'.

A. Data selection: The first stage of the mining process is data selection from the WHO's NCD report of Saudi Arabia. In this step, the data are prepared and errors such as missing values, data inconsistencies, and wrong information are corrected.

B. Data preparation: The data preparation stage is crucial for data analysis. Databases stored in .xls or .mdb format were found to be insufficient. The Oracle Data Miner software requires input to be provided in a particular format. Consequently, it was deemed necessary to convert the database to Oracle Database 10g format to facilitate use with the Oracle Data Miner.

C. Data analysis: In the data analysis stage, data are analyzed to achieve the desired research objectives, for example by selecting the appropriate target values from the master table. In a data mining engine, the data mining techniques comprise a suite of algorithms such as SVM, Naïve Bayesian, etc. In this study, we used a regression technique that employed a support vector machine algorithm.

D. Result database: At this stage, the desired algorithm and associated parameters have been chosen. The Oracle Data Miner software has a specific option, such 'publish', that processes the raw data and creates a result database.

E. Knowledge evaluation and pattern prediction: This stage extracts new knowledge or patterns from the result database. An informative knowledge database is generated that facilitates pattern forecasting on the basis of prediction, probabilities, and visualization.

F. Deployment: The final stage of this process applies a previously selected model to new data to generate predictions.

3.2.1. Predictive data mining

Predictive data mining is a term that is generally applied to data mining projects that seek to identify one or more predictive statistical or neural network models. In the present case, we would like to predict the effective treatment for diabetes mellitus.

Because the data are numeric, a regression technique was employed through the Oracle Data Miner software. The regression technique is described below.

3.2.2. Regression

Regression is a data mining technique that is used to predict a value. For example, the rate of success for patient treatment can be predicted using regression techniques. A regression takes a numeric dataset and develops a mathematical formula to fit the data. A regression task begins with a dataset of known target values. In the present investigation, the 'Percentage' values in each table are used as a target value.

The straight line regression analysis involves a responsible variable, 'y', and a single predictor variable, 'x'. This technique is the simplest form of regression and models 'y' as a linear function of 'x':

$$y = b + wx,$$

where the variance of y is assumed to be constant and b and w are regression coefficients that specify the y-intercept and slope of the line, respectively. The 'w' and 'b' regression coefficients can also be considered weights, so that we can write $y = w_o + w_1 x$ (Almazyad et al., 2010).

3.2.3. Algorithm: support vector machine (SVM)

The support vector machine is a training algorithm for learning classification and regression rules from data. The SVM is based on statistical learning theory. The SVM solves the problem of interest indirectly, without solving the more difficult problem. The support vector machine presents a partial solution to the bias variance trade-off dilemma. There are two ways of implementing SVM. The first technique involves mathematical programming and the second technique employs kernel functions. When kernel functions are used, SVM focuses on dividing the data into two classes, P and N, corresponding to the case when $y_i = +1$ and $y_i = -1$, respectively. The support vector classification searches for an optimal separating surface, called a hyperplane, which is equidistant from each of the classes. This hyperplane has many important statistical properties and kernel functions are non-linear decision surfaces (Burbidge and Buxton, 2001).

If training data are linearly separable, then a pair ($\boldsymbol{w}$, b) exists such that

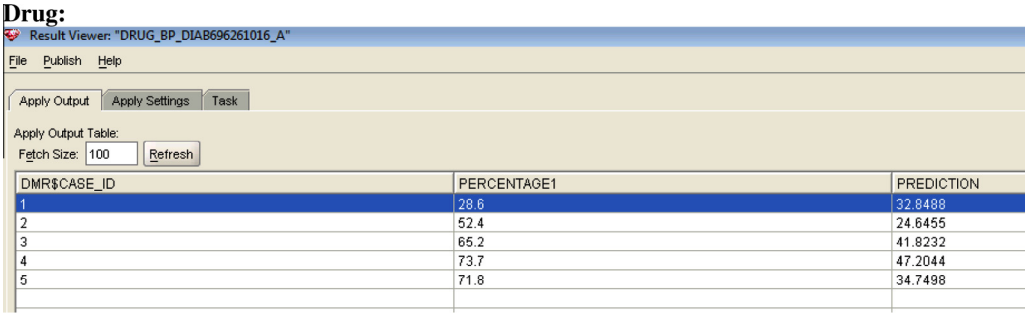
$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \text{ for all } \mathbf{x}_i \in P, \text{ and}$$

$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \text{ for all } \mathbf{x}_i \in N$$

where w is a weight vector and b is a bias. The prediction rule is given by:

$$f = \text{sign}(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b).$$

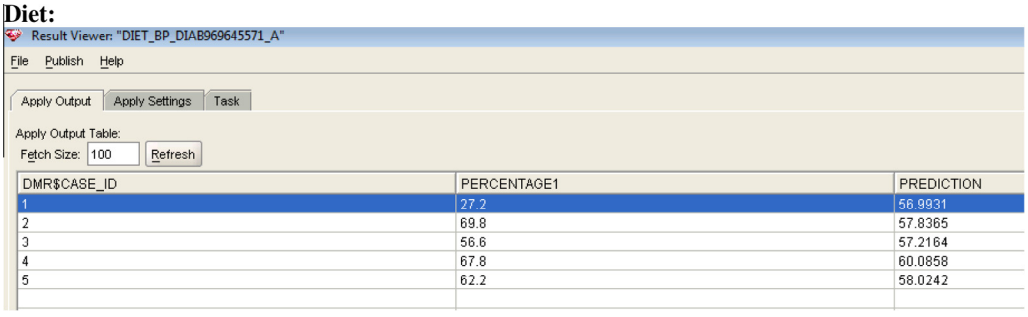
Drug:



DMR$CASE_ID	PERCENTAGE1	PREDICTION
1	28.6	32.8488
2	52.4	24.6455
3	65.2	41.8232
4	73.7	47.2044
5	71.8	34.7498

Figure 2 Prediction of age group with drug percentage.

Diet:



DMR$CASE_ID	PERCENTAGE1	PREDICTION
1	27.2	56.9931
2	69.8	57.8365
3	56.6	57.2164
4	67.8	60.0858
5	62.2	58.0242

Figure 3 Prediction of age group with diet percentage.

4. Experimental analysis

We performed data mining analysis on the Saudi Arabia NCD data using the Oracle Data Miner tool. The five age groups were re-classified into two age groups: Young and Old. Predictions based on the young group and old age groups were denoted as $p(y)$ and $p(o)$, respectively. The$p(y)$ group included the 15–24, 25–34 and 35–44 age groups, while the $p(o)$ group included the 35–44, 45–54 and 55–64 age groups. It should be noted that the '35–44' group is common to both of the two age groups. The mathematical expressions for the predictions of $p(y)$ and $p(o)$ are stated below:

$$p(y) = \sum_{y=1}^{3} p(y), \tag{1}$$

$$p(o) = \sum_{o=3}^{5} p(o), \tag{2}$$

$$p(y) \cap p(o) = \text{'3'}. \tag{3}$$

The database for diabetic treatment is shown in Tables 1–6. The 'Sr_No' contains the primary keys for each database, holding values 1, 2, 3, 4 and 5. The serial number 1 indicates an age of '15–24'; 2 is associated with patients with ages '25–34'; 3 indicates an age of '35–44'; 4 corresponds to an age group of '45–54'; and 5 is related to patients aged '55–64'.

The prediction results for each treatment type are provided below.

Fig. 2 shows the prediction of drug treatment effectiveness in each age group. As per Eqs. (1) and (2) the prediction had been indicated as below:

$$p(y) = 101.31, \tag{3}$$

$$p(o) = 164.77. \tag{4}$$

Eqs. (3) and (4) $p(o) > p(y)$. This implies that drug treatment is more effective for patients in the old age group than patients in the young group. Apart from medication, exercise, weight reduction and smoke cessation are also important aspects of effective diabetes treatment. Drugs are also important for controlling blood sugar levels. Sometimes a single medication is effective for the treatment of diabetes, while in other cases a combination of medications is more effective.

Each class contains one or more specific drugs. Diabetes drugs work in various ways to lower blood sugar:

- Stimulating the pancreas to produce and release more insulin,
- Inhibiting the production and release of glucose from the liver, which reduces the patient's need for insulin,
- Blocking the ability of stomach enzymes to break down carbohydrates, or
- Make tissues more sensitive to insulin (http://www.mayoclinic.com/health/730diabetes-treatment/DA0008, 2011).

Fig. 3 shows the prediction of effectiveness for dietary treatment. The prediction results are provided in Eqs. (5) and (6) for young and old age groups, respectively. These results indicate that dietary treatment is more effective for patients in the old age group than patients in the young group.

Predictions are summed as indicated in Eqs. (1) and (2):

$$p(y) = 172.04, \tag{5}$$

$$p(o) = 175.32. \tag{6}$$

Eqs. (5) and (6) indicate that $p(o) > p(y)$.

Proper diet is a treatment option for diabetes patient. A doctor

Weight:

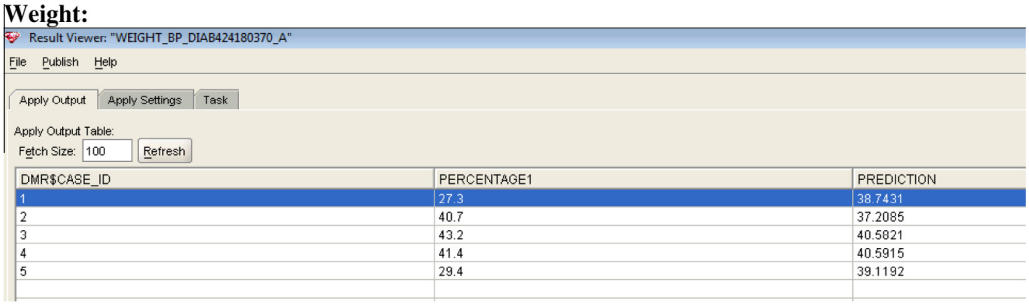


Figure 4 Prediction of age group with weight percentage.

Smoke Cessation:

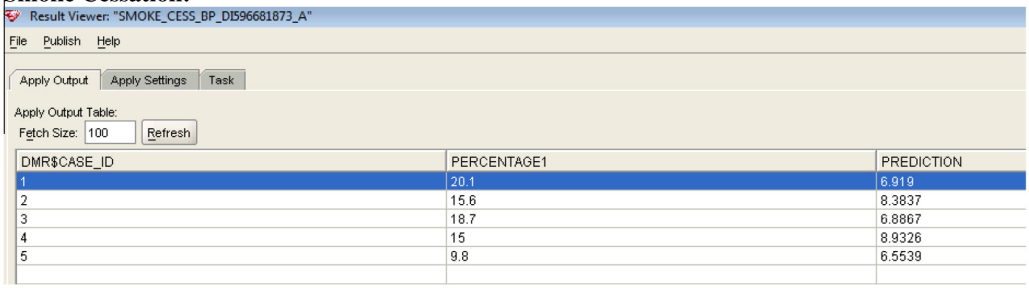


Figure 5 Prediction of age group with smoking cessation percentage.

Exercise:

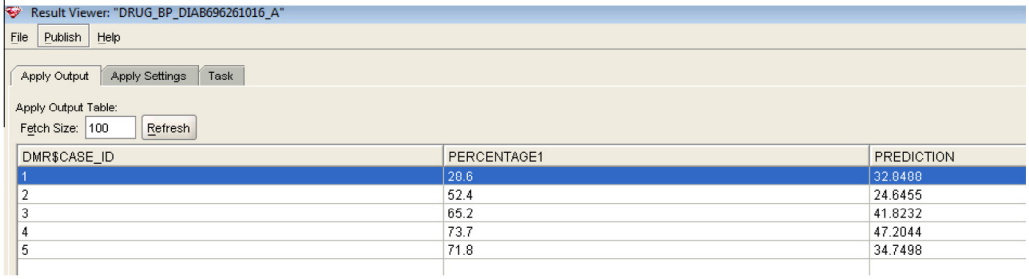


Figure 6 Prediction of age group with exercise percentage.

will usually prescribe diet as part of diabetes treatment. A dietician or nutritionist can recommend a diet that is not only healthy but also palatable and easy to follow. Pre-printed, standard diets are no longer the norm. Many experts, including the 'American Diabetes Association, recommend that 50–60% of daily calories come from carbohydrates, 12–20% from protein, and no more than 30% from fat.

Fig. 4 shows the prediction of effectiveness for weight reduction treatment. The results are provided in Eqs. (7) and (8) for the young and old age groups, respectively.

The predictions have been summed according to Eqs. (1) and (2):

$$p(y) = 116.53, \tag{7}$$

$$p(o) = 120.29. \tag{8}$$

The above equations indicate that $p(o) > p(y)$. When compared to younger patients, weight reduction treatment is more effective for patients in the old age group vs. patients, in the young group. If a person is overweight and has type 2 diabetes, weight loss lowers blood sugar levels, improves overall health, and aids general well-being. A previous study indicated that weight loss improved both insulin sensitivity and β-cell capacity (78–119%), as determined by the homeostasis model assessment method (HOMA) (Dixon and O'Brien, 2002). Weight loss alone is not a cure for type 2 diabetes. However, a large amount of weight loss (100 lbs) was found to reduce the prevalence of type 2 diabetes from 27% to 9% after 6 years of follow-up (Tayek, 2002)

$$p(y) = 22.18, \tag{9}$$

$$p(o) = 22.37. \tag{10}$$

Eqs. (9) and (10) indicate that $p(y) \approx p(o)$.

Fig. 6 shows the prediction of effectiveness for exercise treatment. Result are provided in Eqs. (11) and (12) for the young and old age groups, respectively. Exercise treatment is more effective for patients in the old age group than patients in the

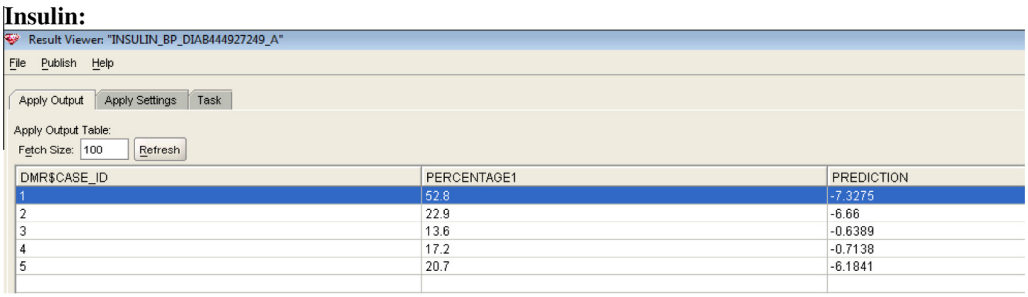


Figure 7 Prediction of age group with insulin percentage.

Table 7 Comparison of prediction.

Treatment	Prediction of effectiveness in the young group $p(y)$	Prediction of effectiveness in the old age group $p(o)$	Comparison of $p(y)$ with $p(o)$
Drug	101.31	164.77	$p(o) > p(y)$
Diet	172.04	175.32	$p(o) > p(y)$
Weight	116.53	120.29	$p(o) > p(y)$
Smoke cessation	22.18	22.37	$p(o) \approx p(y)$
Exercise	99.31	123.77	$p(o) > p(y)$
Insulin	−14.62	−15.07	$p(o) \approx p(y)$

young age group.Predictions are computed as described in Eqs. (1) and (2):

$$p(y) = 99.31, \tag{11}$$

$$p(o) = 123.77. \tag{12}$$

Eqs. (11) and (12) indicate that $p(o) > p(y)$, which implies that exercise treatment is more useful to patients in the old age group versus patients in the young group.

Fig. 7 shows the prediction of effectiveness for insulin treatment. The results are provided in Eqs. (13) and (14) for patients in the young group and old age group, respectively. Insulin treatment is effective for the age group of patients.Predictions are provided below based on Eqs. (1) and (2):

$$p(y) = -14.62 \tag{13}$$

$$p(o) = -15.07 \tag{14}$$

Eqs. (13) and (14) indicate that $p(y) \approx p(o)$.Diabetes mellitus, often simply referred to as diabetes, is a disorder caused by decreased production of insulin or by a decreased ability to use insulin. This decrease in insulin production or insulin utilization results in an increase in blood glucose levels. Insulin therapy is now the standard treatment method for type 1 diabetes. Treatments vary depending on the patient's blood glucose level, age, body mass index, genetic, food consumption and activity level (Gulcin Yildirim et al., 2011). When insulin dosages are based on blood glucose levels, doctors can tailor the medication to the patient's need, such as rapid-acting, short-acting, intermediate-acting, long-acting etc. (http://diabetes.niddk.nih.gov/dm/pubs/medicines_ez/insert_C.aspx).

5. Results and discussion

A comparison of the predictions, $p(y)$ and $p(o)$, are given in Table 7.

The NCD data of Saudi Arabia was used for analysis. The data mining tool has been correctly applied. To a domain expert (physician), some of the results are logical and make clinical sense, but not all. The results obtained here underscore the need to match available data with queries that produce clinically meaningful information. A mismatch may result in statistically appropriate, but clinically inappropriate, data analysis.

Data mining techniques are becoming increasingly popular and have been employed to clinically process medical data. In the present context, we would like to explore the effective treatment of diabetes and study the effectiveness of different diabetes treatments in Saudi Arabia.

Diabetes is a serious health problem all over the world and is significantly more prevalent among Saudi male patients. The prevalence of diabetes in Saudi Arabian male patients is directly proportional to the age of the patient, leading to an increased morbidity rate. Using a regression technique that was applied to diabetes data from WHO, predictions were made as to the effectiveness of each treatment type. The predictions are recorded in Table 7, which compares predictions regarding treatment effectiveness among young and old age groups in response to all six modes of treatments.

The predictions regarding drug treatment effectiveness indicate a wide difference between the $p(y)$ and $p(o)$ values. Based on the fact that $p(o) > p(y)$, drug treatment appears to be more effective for patients in the old age group versus those in the young group. This pattern indicates that drug treatment is effective for both groups of patients but more effective for patients in the old age group. Diabetes drugs are usually prescribed, along with recommendations to make specific dietary changes and exercise regularly, to people with type 2 diabetes. Young patients are advised to concentrate more on other modes of treatment, like diet control, weight reduction, exercise, and smoke cessation. Therefore, this mode of treatment predicts a positive treatment for patients in the old age group as compared to patients in the young group.

Dietary treatment for diabetic patients is strongly recommended for both young and old age groups. The prediction results strongly indicate that regular dietary control is necessary for both old age diabetic patients and young age diabetic patients. Metabolic control minimizes increased blood sugar levels, which in turn helps to minimizing chronic end complications. A special focus on narrow dietary goals was found to have large effects on metabolic control (Sigurdardottir et al., 2007). This prediction indicates that dietary control is more useful for patients in the old age group compared to the young age group, $p(o) > p(y)$. This finding clearly indicates that metabolic activity is slower in old age diabetic patients when compared to young diabetic patients. Therefore, the food management plays a vital role in the control of diabetic treatment for both groups but more so for old age patients.

The comparative view shown in Fig. 4 indicates the effectiveness of weight reduction treatment for diabetes for the young and old age groups. Because $p(o) > p(y)$, weight reduction is strongly recommended for old age diabetic patients. On average, weight loss also resulted in a reduction of HbA1c, which is associated with a decrease in fasting blood glucose (Tayek, 2002). Weight reduction is also, therefore, expected to be an effective treatment for both the age groups, particularly the old age group. Hence, the prediction $p(o) > p(y)$ stands as a better treatment for diabetes.

The comparative view of prediction for smoke cessation treatment for diabetes in Fig. 2 indicates that quitting smoking is an effective treatment for both group of patients $p(y)$ and $p(o)$. Table 7 and Fig. 5 indicates that the prediction value of smoke cessation treatment is both the groups of patients that simply indicates the treatment is effective for both the groups of patients. Smoking is especially harmful for people with diabetes. The main ingredient of cigarettes is nicotine, which causes large and small blood vessels to harden and narrow, resulting in reduced blood flow to the rest of the body. Because diabetes already presents a risk of developing health problems such as heart disease, stroke, kidney disease, nerve damage, foot problems, etc., smoking exacerbates these risks (http://diabetes.webmd.com/diabetes-smoking-cessation-tips, 2011). Therefore, the present study the prediction $p(y) = p(o)$ strongly recommends the total eradication of smoking for both groups of patients.

Exercise also plays a vital role in the treatment of diabetes. Table 7 and Fig. 6 present predictions of effectiveness for exercise in the treatment of diabetes for patients in both age groups. Exercise lowers patient's blood sugar levels, resulting in fewer medications for the patients, and improves insulin sensitivity. Combining diet, exercise, and medicine (when prescribed) will help to control the patient's weight and blood sugar level by improving the body's use of insulin. Exercise helps to burn excess body fat and control weight, improving insulin sensitivity, muscle strength, and increasing bone density and strength. Daily exercise maintains lower blood pressure, which protects against heart and blood vessel diseases by lowering 'bad' LDL cholesterol and increasing 'good' HDL cholesterol, improving blood circulation and reducing patient's risk of heart disease (http://diabetes.webmd.com/guide/exercise-guidelines, 2011). The results obtained here, $p(o) > p(y)$, indicates that exercise treatment is more effective for patients in the old age group than patients in the young group. Based on the present investigation, exercise treatment is more strongly recommended for older patients versus younger patients. Recent research shows that blood glucose levels normally rise with age (http://www.oldagesolutions.org/Health/Diabetes.aspx, 2012). Therefore, the old age group of diabetic patients need more exercise treatment compared to the young age group.

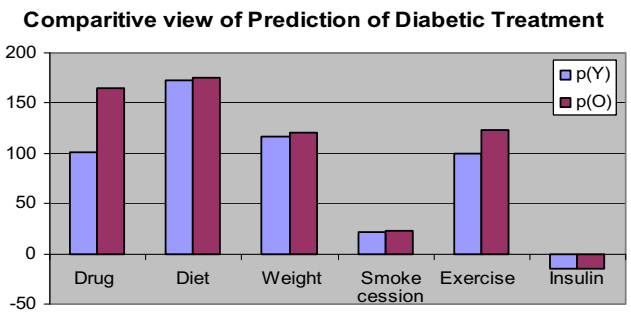


Figure 8 Comparison of $p(y)$ and $p(o)$.

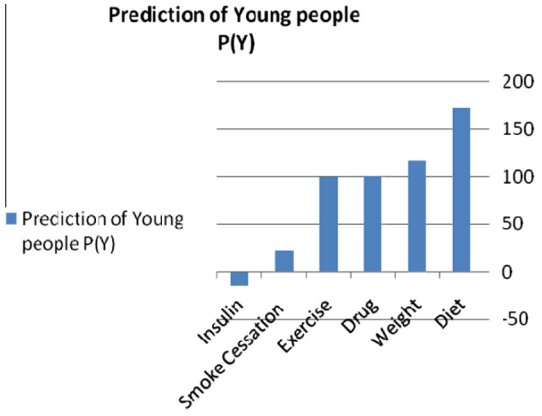


Figure 9 Prediction order of control of diabetes of $p(y)$.

Table 7 and Fig. 7 present prediction of effectiveness for the treatment of diabetes with insulin. The prediction values for both age groups is negative insulin is not administered orally. If we assume oral treatments are positive, then insulin treatment is negative because it is injected into the blood. Prior to 2005, treatment of diabetes did not include simultaneous administration of both drugs and insulin. Therefore, the prediction values are negative due to the drug administration. There is a clear demarcation between treatment types: insulin dependent (type 1), non insulin dependent (type 2). In the present investigation, as shown in Fig. 7 and Table 7, predictions for the effectiveness of insulin treatment are negative for both groups of diabetes patients. In type 2 diabetes, the problem is not a lack of insulin output but an increasing resistance to the effects of insulin by the body. (See Fig. 8).

In the present study, insulin administration was investigated using data from the WHO's NCD report. The study was limited to the treatment of type 1 and type 2 diabetics.

5.1. Preferential order of treatments based on prediction

The predicted effectiveness of diabetic treatments patterns of prediction for young and old patients $p(y)$ and $p(o)$ are depicted separately in Figs. 9 and 10. It is quite interesting to note that the orders of treatment predictions are different for each group of patients.

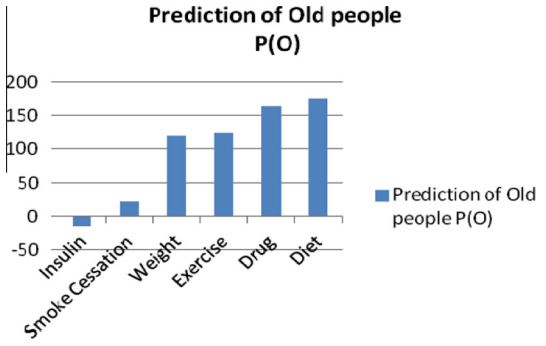


Figure 10 Prediction order of control of diabetes of $p(o)$.

In the literature, one can find an exhaustive description of different treatment methods for diabetic patients. There is no specific demarcation, however, over the treatment modes for patients of young and old age groups. In the present investigation, data mining predictions are based on the age group. Moreover, these predictions provide insight as to the order for preferential modes of treatment in each group.

The young age group, $p(y)$, is predicted to have a preferential order of diet control, weight control, drug treatment, exercise treatment, smoke cessation, and finally, insulin.

The preferential order of modes of treatment for old age group patients, $p(o)$, differs from $p(y)$. The predictions indicated here are diet control, drug treatment, exercise, weight control, smoking cessation, and finally, insulin.

It is evident from these investigations that while patients in the old age group should be immediately prescribed drug and diet control treatments, drug control treatments can be delayed for patients in the young group because drugs have other side effects. As there is no other alternative for old age patients, the drugs are prescribed well in advance. Other modes of treatment are comparable for both age groups.

6. Conclusion

The prevalence of diabetes is increasing among Saudi Arabian patients. The present study concludes that elderly diabetes patients should be given an assessment and a treatment plan that is suited to their needs and lifestyles. Public health awareness of simple measures such as low sugar diet, exercise, and avoiding obesity should be promoted by health care providers. In this study, predictions on the effectiveness of different treatment methods for young and old age groups were elucidated. The preferential orders of treatment were found to be different for the young and old age groups. Diet control, weight reduction, exercise and smoking cessation are mutually beneficial to each other for the treatment of diabetes. The collective and collaborative modes of these treatment, along with drug (type 2) and insulin (type 1) treatment, are found effective control the effects of diabetes.

Acknowledgments

We would like to first give thanks to Almighty Allah. We are also grateful to all those who assisted in conducting this research project, especially Salman bin Abdulaziz University, Ministry of Higher Education, and Kingdom of Saudi Arabia for the generous Grant extended for this project. Finally, we would like to extend our thanks to the Reporting Committee for NCD Risk Factors Standard Report-2005, Ministry of Health, including those with World Health Organization (WHO), EMRO, Kingdom of Saudi Arabia.

References

Almazyad, A.S., Ahamad, M.G., Siddiqui, M.K., Almazyad, A.S., 2010. Effective hypertensive treatment using data mining in Saudi Arabia. Journal of Clinical Monitoring and Computing 24 (6), 391–401. http://dx.doi.org/10.1007/s10877-010-9260-2, Springers.

Al-Nozha, M.M., Al-Maatouq, M.A., Al-Mazrou, Y.Y., Al-Harthi, S.S., Arafah, M.R., Khalil, M.Z., Khan, N.B., Al-Khadra, A., Al-Marzouki, K., Nouh, M.S., Abdullah, M., Attas, O., Al-Shahid, M.S., Al-Mobeireek, A., 2004. Diabetes mellitus in Saudi Arabia. Saudi Medical Journal 25 (11), 1603–1610.

Breault, Joseph L., Goodall, Colin R., Fose, Peter J., 2002. Data mining a diabetic data warehouse. Artificial Intelligence in Medicine 26, 37–54.

Burbidge, Robert, Buxton, Bernard, 2001. An Introduction to Support Vector Machines for Data Mining. Young OR12, University of Nottingham, Operational Research Society, Operational Research Society. doi: 10.1.1.103.7639.

Charpentier, G., Genès, N., Vaur, L., Amar, J., Clerson, P., Cambou, J.P., Guéret, P., 2003. Control of diabetes and cardiovascular risk factors in patients with type 2 diabetes:a nationwide French survey. Diabetes & Metabolism 29, 152–158, www.e2med.com/dm.

Craiga, Maria E., Jonese, Timothy W., Silinka, Martin, Pingg, Yeo Jing, 2007. Diabetes care, glycemic control, and complications in children with type 1 diabetes from Asia and the Western Pacific Region. Journal of Diabetes and its Complications 21, 280–287.

Diabetes Group of the Andalusian Society of Endocrinology and Nutrition, Soriguer-Escofet, Federico, et al, 2002. Prevalence of latent autoimmune diabetes of adults (LADA) in Southern Spain. Diabetes Research and Clinical Practice 56, 213–220.

Dixon, J.B., O'Brien, P.E., 2002. Health outcomes of severely obese type 2 diabetic subjects 1 year after laparoscopic adjustable gastric banding. Diabetes Care 25, 358–363.

Gan, D. (Ed.), 2003. Diabetes Atlas, 2nd ed. Brussels: International Diabetes Federation. <http://www.eatlas.idf.org/webdata/docs/Atlas%202003-Summary.pdf> (accessed 19.06.11).

Gulcin Yildirim, Emirhan, Karocha, Adem, Ucar, Tamer, 2011. Dosage planning for diabetes patients using data mining methods. Procedia Computer Science. http://dx.doi.org/10.1016/j.procs.2011.01.018 1374-1380, Published by Elsevier Ltd..

Harwell, Todd S. et al, 2010. Factors associated with achieving a weight loss goal among participants in an adapted Diabetes Prevention Program. Primary Care Diabetes.

Nilsson, P.M. et al, 2004. Smoking is associated with increased HbA1c values and microalbuminuria in patients with diabetes—data from the National Diabetes Register in Sweden. Diabetes & Metabolism 30, 261–2688, http//www.e2med.com/dm.

Obtained through the Internet. <http://www.mayoclinic.com/health/diabetes-treatment/DA00089> (accessed 31.05.11).

Obtained through the Internet. <http://medweb.bham.ac.uk/easdec/prevention/smoking_and_diabetes.htm#help> (accessed 31.05.11).

Obtained through the Internet. <http://diabetes.webmd.com/diabetes-smoking-cessation-tips> (accessed 31.05.11).

Obtained through the Internet. <http://diabetes.webmd.com/guide/exercise-guidelines> (accessed 31.05.11).

Richards, G., Rayward-Smith, V.J., Sonksen, P.H., Carey, S., Weng, C., 2001. Data mining for indicators of early mortality in a

database of clinical records. Artificial Intelligence in Medicine 22, 215–231.

Sigurdardottir, Arun K., Jonsdottir, Helga, Benediktsson, Rafn, 2007. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. Patient Education and Counseling 67, 21–31.

Tayek MD, John A., 2002. Is weight loss a cure for type 2 diabetes? American Diabetes Association. Diabetes Care 25 (2), 397–398. http://dx.doi.org/10.2337/diacare.25.2.397.

World Health Organisation, NCD risk factor, standard report of Ministry of Health, Saudi Arabia. 2005. Obtained through the Internet: <http://www.emro.who.int/ncd/pdf/stepwise_saa_05.pdf> (accessed 15.12.09).

Yokotaa, Ichiro, Amemiya, Shin, Kida, Kaichi, Sasaki, Nozomu, Matsuura, Nobuo, 2005. Past 10-year status of insulin therapy for preschool-age Japanese children with type 1 diabetes. Diabetes Research and Clinical Practice 67, 227–233.