Data Mining Perspective: Prognosis of Life Style on Hypertension and Diabetes

Abdullah Aljumah and Mohammad Siddiqui
College of Computer Engineering and Sciences, Salman bin Abdulaziz University, KSA

***Abstract**: In the present era, the data mining techniques are widely and deeply useful as decision support systems in the fields of health care systems. The proposed research is an interdisciplinary work of informatics and health care, with the help of data mining techniques to predict the relationship among interventions of hypertension and diabetes. As the study shows persons who have diabetes can have chances of hypertension and vice versa. In the present work we would like to approach the life style intervention of hypertension and diabetes and their effects using data mining. Life style intervention plays a vital role to control these diseases. The intervention includes the risk factor like diet, weight, smoking cessation and exercise. The regression technique is used in which dependent and Independent Variables (IV) are defined. The four interventions are treated as (IV) and two diseases hypertension and diabetes are Dependent Variables (DV). We have established the relationship between hypertension and diabetes, using the data set of Non Communicable Disease (NCD) report of Saudi Arabia, World Health Organisation's (WHO). The Oracle Data Miner (ODM) tool is used to analyse the data set. Predictive data analysis gives the result that interventions weight control and exercise have the direct relationship between them in both the diseases.*

Keywords: *Oracle data mining tool, prediction, regression, support vector machine, hypertension, diabetes.*

Received April 10, 2014; accepted June 23, 2014; published online March 8, 2015

1. Introduction

The statistics state that Saudi Arabia is one of the highest numbers of patients of diabetes and hypertensions, the Non Communicable Disease (NCD) data indicates an alarming increase in the incidence and prevalence of diabetes and hypertension [1]. Researchers have to think to control these diseases and to give their possible solutions. Diabetes mellitus and hypertension are interrelated diseases that strongly predispose an individual to atherosclerotic cardiovascular disease [11]. Many researchers have defined the term data mining. It is an analysis of observational data sets to find the hidden relationships and to summarize the data in novel ways that are worth to data owner. Data mining gain novel and deep insights of datasets that can support to decision making. Data mining applications is in various sectors of real world. One of the renowned applications is health care knowledge management which is widely used in clinical system. The obtained knowledge from data mining can be applied to sort out health care issues and to serve the patients in better way.

Proposed research work provides an overview of this emerging field, with the aim of clarifying how data mining techniques are applicable on healthcare analysis to predict the relevance of modes of diabetic and hypertension interventions. For the past few years, the 'data mining' technique steadily has been increased in the medical and clinical diagnostics.

In this paper, we give a methodological review of data mining, focusing the analysis process using regression technique to establish the interrelationship between hypertension and diabetic interventions control.

The latest survey reveals that the patients of hypertension and diabetes have been dramatically increased in Saudi Arabia and reports show that they have gained epidemic form [22].

1.1. Interventions of Hypertension and Diabetes

The important aspect is prediction of interrelationship between hypertension and diabetes. As mentioned in World Health Organisation's (WHO) NCD report of Ministry of Health, Saudi Arabia [23] following are the four types of interventions selected in common in both the diseases which are causing explosive and alarming disease level, which are pictorially presented below in Figure 1.

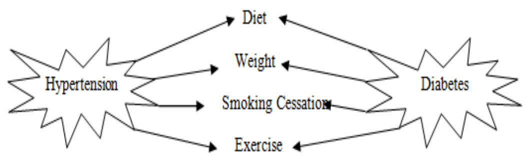


Figure 1. Hypertension and diabetes with common interventions.

2. Related Works

The literature survey is rich in hypertension and diabetes interventions. This proposed work shows various models for each type of diabetic intervention and analysis is carried out using classification based data mining technique [2]. The dataset was studied and analyzed to identify effectiveness of different treatment types for different age groups, five age groups are consolidated into two age groups, that are denoted as $p(y)$ and $p(o)$ for the young and old age groups, respectively [1]. In this study, analysis is divided into three different focuses based on the patient's healthcare costs and examine whether more complex analytical models sing several data mining techniques in SAS® Enterprise Miner™ 7.1 can better predict and explain the causes of increasing diabetes in adult patients in each cost category. The preliminary analysis shows that high blood pressure, age, cholesterol, adult BMI, total income, sex, heart attack, marital status, dental checkups and asthma diagnosis are among the key risk factors [17]. The investigation was carried out on the data sets of NCD risk factors, a standard report of Saudi Arabia 2005, in collaboration with WHO with the help of the Oracle Data Miner (ODM), the data sets for different age groups in case of blood pressure intervention for hypertension for male using different modes had been studied. The age group was in between of 15 years to 64 years [3]. In another investigation recommendations were given for aggressive management of hypertension early to prevent its damaging consequences if left untreated. Public health awareness of simple measures, such as diet, exercise and avoiding obesity, to maintain normal arterial blood pressure need to be implemented by health care providers [4].

The diabetes in Saudi Arabia had been investigated and found that the overall prevalence of diabetes adult patients in KSA is 23.7%, the study further recommended a longitudinal study to demonstrate the importance of modifying risk factors for the development of diabetes and reducing its prevalence in KSA [5]. In another work, concentrated on dietary changes of patient and served it as an initial treatment before drug therapy. In those hypertensive patients already on drug therapy, lifestyle modifications, particularly a reduced salt intake, can further lower BP. The current challenge to healthcare providers, researchers, government officials and the general public is developing and implementing effective clinical and public health strategies that lead to sustained dietary changes among individuals and more broadly among whole populations [6]. They illustrate a two-phase analysis procedure to simultaneously predict hypertension and hyperlipidemia. Common risk factors of these diseases picked up by data mining and majority vote. This study uses common risk factors to build MARS predictive models for hypertension and hyperlipidemia [8]. In another, study the recommends the intake of weight reducing diet, regular exercise and restrict alcohol and salt intake for effectiveness of lifestyle interventions for hypertension management [9]. Cigarette smoking, obesity and sedentary life style are known to increase risk of coronary and other vascular disease. Yet eliminating, or reducing, these risk factors through lifestyle modifications are a significant challenge to patients and their physicians.

When applying this paradigm to exercise, physicians can motivate patients by making them aware of the benefit of even moderate levels of activity, outlining a specific exercise program and setting appropriate goals and following up on their patient's progress. Studies show that physicians can have a major positive impact on smoking cessation merely by asking patients are advised to quit smoking. Dietary intervention should be tailored to individual patients, their food preferences and ethnic backgrounds [20]. Diabetes mellitus and hypertension are interrelated diseases that strongly predispose an individual to atherosclerotic cardiovascular disease [10]. Georga *et al.* [11] have used machine learning approach to the prediction of daily glucose time series as well as of hypoglycemic event, which relies on support vector and Gaussian process for regression.

This paper describes the analysis of a database of diabetic patient's clinical records and death certificates. The objective of the study was to find rules that describe associations between observations made of patients at their first visit to the hospital and early mortality. Pre-processing was carried out and a Knowledge Discovery in Databases (KDD) package, developed by the Lanner Group and the University of East Anglia, was used for rule induction using simulated annealing. The most significant discovered rules describe an association that was not generally known or accepted by the medical community; however, recent independent studies confirm their validity [16].

In this paper, the rapid miner tool had used to analyze a Pima Indians diabetes data set, which collects the information of patient's with and without developing diabetes. The discussion follows the data mining process. The focus will be on the data pre-processing, including attribute identification and selection, outlier removal, data normalization and numerical discretization, visual data analysis, hidden relationships discovery and a diabetes prediction model construction [12].

In this study, in case of diabetic patients the classification and regression tree approach is used to analyze the data sets [7]. Data mining had been done between the relationships in diabetes data for efficient classification. The data mining methods and techniques will be explored to identify the suitable methods and techniques for efficient classification of Diabetes dataset and in mining useful patterns. A classification rate of 91% was obtained for C4.5 algorithm [15].

In this paper, a wrapper based approach that integrates multi objective genetic algorithms and the target learning algorithm is presented in order to evolve optimal subsets of discriminatory features for robust pattern classification. In the training phase the multi objective genetic algorithms is used to find a subset of relevant attributes that minimizes both classification error rate and size of the classifier discovered by the classification algorithm, using the Pareto dominance approach. Two different classifiers were compared in this study using genetic feature subset selection: Decision tree, a feed forward neural network. Moreover, two different MOGAs namely elitism-based MOGA and Non-dominated sorting genetic algorithm have been employed separately in the training phase. Experimental results reveal that both of these proposed variants of MOGA combined with classifiers namely decision trees/FFNN yield improved classification performance and reduced classification time as compared to standard classifiers namely decision trees decision tree or standard feed-forward networks. Moreover, NSGA performs better than the elitism based approach in terms of classification time [13]. In work the guidelines emphasize the importance of lifestyle interventions like diet, weight control, smoking cessation and regular exercise, which are simultaneously effectively reducing blood pressure and the associated cardiovascular risk [14]. This study compared two traditional classification methods (logistic regression and fisher linear Discriminant analysis) and four machine-learning classifiers (neural networks, support vector machines, fuzzy c-mean and random forests) to classify persons with and without diabetes [18].

3. Material and Methods

3.1. Data Collection

The data repository for this study was sourced from the WHO, NCD risk factor, a standard report of ministry of health, Saudi Arabia, 2005 WHO, NCD risk factor, standard report of ministry of health, Saudi Arabia [23]. This is publicly available dataset on the internet.

The report broadly includes the medical history of high blood pressure, diabetes, hypertension, obesity, etc., the present investigation focussed on the exploring the hidden pattern and trends of risk factors of hypertension and diabetes dataset for male patients. The dataset of these two diseases are taken because of their rapidly increasing rate.

3.2. Schema Design

As shown in Figure 2. Eight tables have been designed four tables for hypertension and four tables for diabetes in Oracle database named as ‘hyp_diet’, ‘hyp_weight’, ‘hyp_smoke_cessation’, ‘hyp_exercise’ for hypertension and ‘diab_diet’, ‘diab_weight’, ‘diab_smoke_cessation’, ‘diab_exercise’ for diabetes respectively.

Each table indicates the mode of hypertension and diabetes disease intervention among patients in Kingdom of Saudi Arabia. Six attributes are in each table (sr_no, age, N, small_n, Percentage, SE). The ‘sr_no’ indicates the serial number, which represent the unique identifier in the table (primary key), ‘age’ indicates the age of patients, ‘N’ indicates the total number of patient of each age group, ‘small_n’ column shows the number of patients whose interventions are normal, the percentage indicates the percent of hypertension and diabetic controlled patients by specific mode of interventions and ‘SE’ indicates the standard error. After schema design of eight tables, the diet interventions of both hypertension and diabetes disease are linked using primary key ‘sr_no’ in ODM followed same for weight, smoking cessation and exercise interventions.

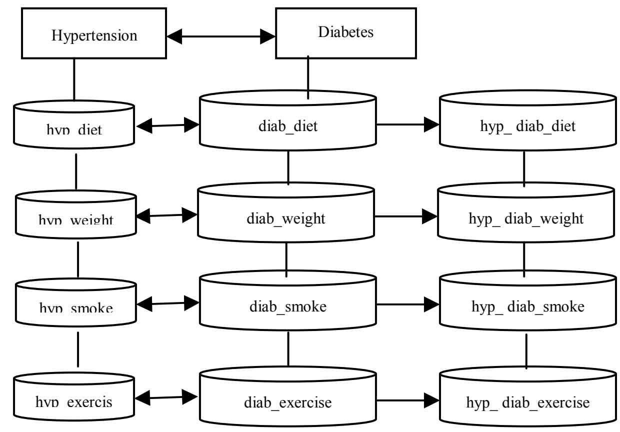


Figure 2. Schema design of dataset.

3.3. Tools and Techniques

Numbers of data mining tools are available today with each having its own pros and corns. For the analysis of data set of NCD report of Saudi Arabia, we have chosen the tool ODM version 10.2.0.3.0.1; build 2007 for the mining activity that acts as a client and oracle 10g database release 10.2.0.3.0 as a database server. This software tool is highly appreciated for building the predictive model to find the hidden patterns. Henceforth, this research is based on regression techniques of data mining to investigate the intervention of these NCD diseases and established the relationship between the modes of intervention for both the diseases and their mutual effect.

3.3.1. Architecture of Model Designing

The processing blocks have been shown in the Figure 3.

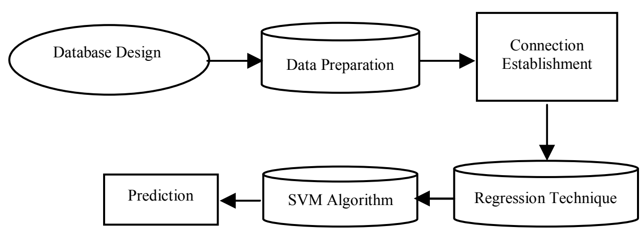


Figure 3. Architecture of model designing.

- Data Selection: Choosing the desired dataset from the schema is one of the primary and crucial tasks, here the selection process is performed four times because we have four interventions.
- Data Preparation: At each selection process, it is to be performed to join the tables like diet of hypertension (hyp_diet) to diet of diabetes

(diab_diabetes) through the common primary key based attribute sr_no and gives hyp_diab_diet as shown in Figure 2.

- Connection Establishment: In the present scenario, Oracle 10g database is server and ODM 10g is client, to connect the schema, ODM requires a set of privileges from the schema (SYSDBA) of server side.
- Regression Technique: It is a type of data mining technique that predicts a numeric value. The analysis used to model the relationship between one or more independent/predictor values and dependent/ response variable. It explains and predicts a Dependent Variable (DV) based on linear relationship with an Independent Variable (IV). The response variable is target value to be predicted and the regression task begins with a data set in which the target variable is known. In the present work the data set values of 'percentage' of patients in each mode of intervention. The IV is the interventions like diet, weight reduction, smoking cessation and exercise. The DV are the diseases hypertension and diabetes.
- Applied Algorithm. Support Vector Machine (SVM): SVM proposed by [19] is a novel approach for addressing classification problems. In its simplest linear form, an SVM is a hyperplane that separates a set of positive elements from a set of negative elements with the maximum interclass distance, of the so-called margin. In Figure 4 shows such a hyper plane with the associated margin. In Figure 4 the formula of SVM is stated as $u=w^T x_i + b$, where w is a normal vector (weight coefficient vector), x_i is the input vector and b is the bias/intercept term. Based on that, we can get the class u where u is 1 or -1. The distance between a training vector x_i and the boundary is called the margin.

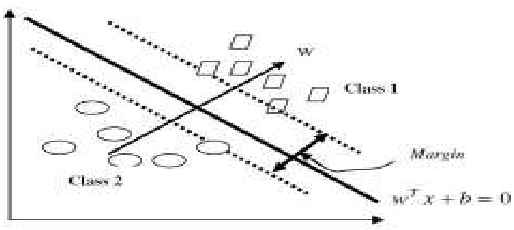


Figure 4. Hyperplane with the maximal margin by a linear SVM.

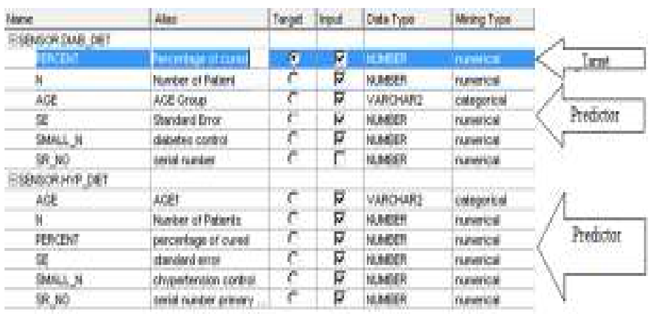
In Figure 4 the formula of SVM is stated as $u=w^T x_i +b$, where, w is a normal vector (weight coefficient vector), x_i is the input vector and b is the bias/intercept term. Based on that, we can get the class u where u is 1 or -1. The distance between a training vector x_i and the boundary is called the margin. According to the original theory by [19], we want to find the margin m where $w^T x_i+b\geq 1$ for all $x_i \in P$ $w^T x_i+b\leq -1$ for all $x_i \in N$ and in order to separate the elements which are in a positive or a negative class. SVM algorithm finds the weight vector and its implementation in various fields, like biomedical and prognosis problems. SVM uses an epsilon-insensitive loss function to solve regression problems. SVM regression tries to find a continuous function such that the maximum number of data points lie within the epsilon-wide insensitivity tube. Predictions falling within epsilon distance of the true target value are not interpreted as errors. The epsilon factor is a regularization setting for SVM regression. SVM algorithm is used to predict outcomes based on text data. The regression model is build by SVM by using one of two available kernels, linear or Gaussian. The linear kernel omits the nonlinear transformation altogether so that, the resulting model is essentially a regression model.

- Prediction: The last episode of whole process is what will be the outcome, to be precise predictions are. It is obtained through the regression technique**.** Oracle data mining builds and applies data mining models inside the Oracle Database, results are immediately available and this process is termed as knowledge deployment.

4. Data Analysis

In the analysis of dataset, ODM version 10.2.0.3.0.1; build 2007 had been used for the data mining activity. The ODM tool acts as a client and oracle 10g database release 10.2.0.3.0 acts as a server. The main advantage of using ODM is all data mining processing occurs within the Oracle database. Since, both ODM and Oracle database are the products of Oracle Corporation it's easy to establish the connection between the ODM clients and database server. Another important aspect of ODM is that, we get secured and stable data management which enhance the accuracy of result.

In the present analysis shown in Figure 5 the target attribute is 'percent' that explains number of patients get cure, based on the target attribute it gives the predictive results.



Name	Alias	Target	Input	Data Type	Mining Type
SENSOR.DIAB_DIET					
PERCENT	Percentage of cured			NUMBER	numerical
N	Number of Patient			NUMBER	numerical
AGE	AGE Group			VARCHAR2	categorical
SE	Standard Error			NUMBER	numerical
SMALL_N	diabetes control			NUMBER	numerical
SR_NO	serial number			NUMBER	numerical
SENSOR.HYP_DIET					
AGE	AGE1			VARCHAR2	categorical
N	Number of Patients			NUMBER	numerical
PERCENT	percentage of cured			NUMBER	numerical
SE	standard error			NUMBER	numerical
SMALL_N	chypertension control			NUMBER	numerical
SR_NO	serial number primary			NUMBER	numerical

Figure 5. Building data using regression.

The below Tables 1, 2, 3 and 4 show the output prediction for each risk factor i.e., diet, weight, smoke cessation and exercise.

The column of Tables 1, 2, 3 and 4 illustrates:

- *Sr*: Serial number a primary key attribute.
- *Age*: Age of patients.

- *N*1: Number of diabetic patients.
- *S*1: Number of intervene diabetic patients.
- *SE*1: Standard error for diabetic patients intervenes.
- *%*1: Percentage of intervene diabetic patients.
- *N*2: Number of hypertension patients.
- *S*2: Number of intervene hypertension patients.
- *SE*2: Standard error for hypertension patients intervene.
- *%*2: Percentage of intervene hypertension patients.
- *P*: Prediction.

Table 1. Predictive value of diet in hypertension and diabetes.

Sr	Age	N1	S1	SE1	%1	N2	S2	SE2	%2	P
1	15-24	10	2	13	19.7	11	3	12.7	27.2	**48.2**
2	25-34	19	2	6.8	10.8	13	9	13.3	69.8	**52**
3	35-44	35	21	9.5	60.4	70	40	7.3	56.6	**56.9**
4	45-54	77	45	6.6	58.7	130	88	4.8	67.8	**67.4**
5	55-64	99	52	8.5	53	142	88	6	62.2	**69**

Table 2. Predictive value of weight in hypertension and diabetes.

Sr	Age	N1	S1	SE1	%1	N2	S2	SE2	%2	P
1	15-24	10	3	12.7	27.3	11	2	13	19.7	**40.07**
2	25-34	19	5	12.9	40.7	13	5	10.4	27	**40.84**
3	35-44	35	30	6.1	43.2	70	17	9.1	48.5	**43.05**
4	45-54	77	54	5.8	41.4	130	26	5.3	33.4	**39.14**
5	55-64	99	42	3.8	29.4	142	39	7.8	39.7	**40.52**

Table 3. Predictive value of smoke cessation in hypertension and diabetes.

Sr	Age	N1	S1	SE1	%1	N2	S2	SE2	%2	P
1	15-24	10	2	11.9	20.1	11	3	16.2	32.3	**13.60**
2	25-34	19	2	9.4	15.6	13	2	7.3	10.8	**16.09**
3	35-44	35	13	4.5	18.7	70	4	5.5	11.7	**18.20**
4	45-54	77	20	3.6	15	130	13	4.4	16.7	**15.49**
5	55-64	99	14	2.6	9.8	142	10	3.2	13.3	**15.29**

Table 4. Predictive value of exercise in hypertension and diabetes.

Sr	Age	N1	S1	SE1	%1	N2	S2	SE2	%2	P
1	15-24	10	6	15.3	52.8	11	3	16.2	32.3	**40.11**
2	25-34	19	5	11.9	22.9	13	6	10.2	33.3	**41.57**
3	35-44	35	32	4.3	13.6	70	13	7.9	36.9	**45.42**
4	45-54	77	55	4.2	17.2	130	23	5.6	29.9	**41.82**
5	55-64	99	44	3.3	20.7	142	28	5.4	27.9	**40.66**

The ODM tool has been accurately imposed over the analysis of dataset. The predicting results are shown in Tables 1, 2, 3 and 4 as an attribute name 'P', The value of 'P' of all four risk factors are logical that makes medical perspective that gives clinically meaningful information. The usage of data mining in medical data analysis is increasing. The present data analysis explores the hidden pattern of interrelationship between hypertension and diabetes.

5. Discussion

The current research work relates data mining to medical informatics in the terms of life style intervention on hypertension and diabetes and what are their relationships among them. Four models have been designed to establish the relationship among these two disease using regression techniques**.**

This is a statistical analysis for predicting the accuracy of the life style intervention of diabetes and hypertension. In the present investigation, SVM algorithm has been adopted to predict the risk factors. Separately analysing the prediction of risk factors has shown in Figure 6 from Tables 1, 2, 3 and 4 all four common risk factors of hypertension and diabetes. The different life style interventions are compared for each age group. Hypertension and diabetes have a strong relationship; diabetes mellitus and hypertension are common diseases that coexist at a greater frequency than that of individual prediction [11].

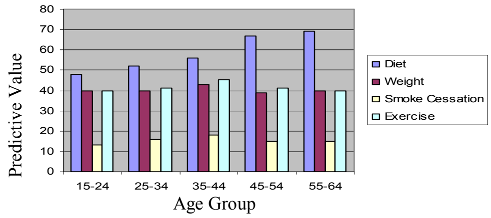


Figure 6. Statistical view of prediction of risk factors.

Risk factor Diet has the highest prediction value as compare to other risk factors especially it has the better outcome in the age group of more than 55 years. Diet intervention in both, to be kept under control, otherwise this may lead to Cardiovascular Diseases (CVD) and other arthroscleroses.

Over weight is the complex problem in health that cause increases the rating of hypertension and diabetes both. In the data analysis section 4 results shows the prediction of risk factor weight is same for age group 15-24, 25-34 and 55-64 is 40. For age group 45-54 the prediction value is 39, at last for age group 35-44 it predictive value is 43.

Fortunately, lifestyle changes including weight management often improve the risk factors associated with hypertension and diabetes.

The result about smoking cessation has different from other risk factors; smoking cessation has a role in the hypertension and diabetes. However, smoking cessation has vital role in pulmonary diseases like Tuberculosis. The smoking intervention modification does not have linear relation with the hypertension and diabetes. However, smoking cessation adds to the interventions to control the diabetes and hypertension it is always recommended to quit the smoking habit in the daily routine [21]. The risk factor exercise, predictive value is rated as 40 for all five age group and it is highly recommended to all type of age groups. Again, the life style modifications recommended for diabetes are very similar to those for hypertension and diabetes.

6. Conclusions

The aim of this study is to investigate the interrelationship between hypertension and diabetes risk factors using SVM based data mining technique. As both have the alarming prevalence in country Saudi Arabia. Present paper focus on the lifestyle modification and to adopt common intervention of both diseases in their daily routine, so that the diseases can be control. As per our predictive analysis, in section 4 we found that diet management have good prediction in all the age group and it is highly recommended that patients have to consult dietician for their diet modification or management. One more remarkable point in Table 5 and Figure 6 is that prediction of diet is directly proportional to the age group, as the age increases the prediction also increases. However, interventions weight and exercise have approximate prediction value and the result shows that both interventions are equal importance in both the diseases.

Table 5. Predictive values of risk factor hypertension and diabetes.

Age	Prediction of Interventions			
	Diet	Weight	Smoke Cessation	Exercise
15-24	48	15-24	48	15-24
25-34	52	25-34	52	25-34
35-44	56	35-44	56	35-44
45-54	67	45-54	67	45-54
55-64	69	40	15	40

References

[1] Aljumah A., Ahamad M., and Siddiqui M., "Application of Data Mining: Diabetes Health Care in Young and Old Patients," *Journal of King Saud University-Computer and Information Sciences*, vol. 25, no. 2, pp. 127-136, 2013.

[2] Aljumah A., Siddiqui M., and Ahamad M., "Application of Classification based Data Mining Technique in Diabetes Care," *Journal of Applied Sciences*, vol. 13, no. 3, pp. 416-422, 2013.

[3] Almazyad A., Ahamad M., Siddiqui M., and Almazyad A., "Effective Hypertensive Intervention using Data Mining in Saudi Arabia," *Journal of Clinical Monitoring and Computing*, vol. 24, no. 6, pp. 391-401, 2010.

[4] Al-Nozha M., Al-Maatouq M., Al-Mazrou Y., Al-Harthi S., Arafah M., Khalil M., Khan N., Al-Khadra A., Al-Marzouki K., Nouh M., Abdullah M., Attas O., Al-Shahid M., and Al-Mobeireek A., "Diabetes Mellitus in Saudi Arabia," *Saudi Medical Journal*, vol. 25, no. 11, pp. 1603-1613, 2004.

[5] Al-Nozha M., Abdullah M., Arafah M., Khalil M., Khan N., Al-Mazrou Y., Al-Maatouq M., Al-Marzouki K., Al-Khadra A., Nouh M., Al-Harthi S., Al-Shahid M., and Al-Mobeireek A., "Hypertension in Saudi Arabia," *Saudi Medical Journal*, vol. 28, no. 1, pp. 77-84, 2007.

[6] Appel L., Brands M., Daniels S., Karanja N., Elmer P., and Sacks F., "Dietary Approaches to Prevent and Treat Hypertension: A Scientific Statement from the American Heart Association," *Hypertension*, vol. 47, no. 2, pp. 296-308, 2006.

[7] Breault J., Goodall C., and Fose P., "Data Mining a Diabetic Data Warehouse," *Artificial Intelligence in Medicine*, vol. 26, no. 1-2, pp. 37-54, 2002.

[8] Chang C., Wang C., and Jiang B., "Using Data Mining Techniques for Multidiseases Prediction Modeling of Hypertension and Hyperlipidemia by Common Risk Factors," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5507-5513, 2011.

[9] Diabetes and High Blood Pressure, available at: http://www.patient.co.uk/health/diabetes-and-high-blood-pressure, last visited 2014.

[10] Dickinson H., Mason J., Nicolson D., Campbell F., Beyer F., Cook S., Williams B., and Ford G., "Lifestyle Interventions to Reduce Raised Blood Pressure: A Systematic Review of Randomised Controlled Trials," *Journal of Hypertens*, vol. 24, no. 2, pp. 215-233, 2006.

[11] Epstein M. and Sowers J., "Diabetes Mellitus and Hypertension," *Hypertension*, vol. 19, pp. 403-418, 1992.

[12] Georga E., Protopappas V., Bellos C., Makriyiannis D., and Fotiadis D., "Interpretation of Long-Term Clinical Diabetes Data and Prediction of Glycemic Control based on Data Mining Technique," *in Proceedings of the International Conference on Health Informatics*, Vilamoura, pp. 315-318 2014.

[13] Han J., Rodriguez J., and Beheshti M., "Diabetes Data Analysis and Prediction Model Discovery using RapidMiner," *in Proceedings of the 2nd International Conference on Future Generation Communication and Networking*, Hainan Island, pp. 96-99, 2008.

[14] Krishna B. and Kaliaperumal B., "Efficient Genetic-Wrapper Algorithm based Data Mining for Feature Subset Selection in a Power Quality Pattern Recognition Application," *the International Arab Journal of Information Technology*, vol. 8, no. 4, pp. 397-405, 2011.

[15] Mancia G., DeBacker G., Dominiczak A., Cifkova R., Fagard R., and Germano G., "ESH-ESC Practice Guidelines for the Management of Arterial Hypertension: ESH-ESC Task Force on the Management of Arterial Hypertension," *Journal Hypertens*, vol. 25, no. 9, pp. 1751-1762, 2007.

[16] Rajesh K. and Sangeetha V., "Application of Data Mining Methods and Techniques for Diabetes Diagnosis," *the International Journal of Engineering and Innovative Technology*, vol. 2, no. 3, pp. 224-229, 2012.

[17] Richards G., Rayward-Smith V., Sonksen P., Carey S., and Weng C., "Data Mining For Indicators of Early Mortality in A Database of Clinical Records," *Artificial Intelligence in Medicine*, vol. 22, no. 3, pp. 215-231, 2001.

[18] Sa-Ngasoongsong A. and Chongwatpol J., "An Analysis of Diabetes Risk Factors using Data Mining Approach," available at:

http://www.lexjansen.com/mwsug/2012/PH/MWSUG-2012-PH10.pdf, last visited 2012.

[19] Tapak L., Mahjub H., Hamidi O., and Poorolajal J., "Real-Data Comparison of Data Mining Methods in Prediction of Diabetes in Iran," *Healthc Inform Res*, vol. 19, no. 3, pp. 177-85, 2013.

[20] Vapnik V., *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, USA, 1995.

[21] Winslow E., Bohannon N., Brunton S., and Mayhew H., "Lifestyle Modification: Weight Control, Exercise and Smoking Cessation," *Am J Med*, vol. 101, no. 4, pp. 25-31, 1996.

[22] Who Country Cooperation Strategy for Saudi Arabia., available at: http://www.who.int/countryfocus/cooperation_strategy/ccs_sau_en.pdf, last visited 2011.

[23] World Health Organisation, NCD risk factor, standard report of Ministry of Health, Saudi Arabia., available at: http://www.who.int/chp/steps/2005_SaudiArabia_STEPS_Report_EN.pdf, last visited 2005

Abdullah Aljumah received PhD in electronic engineering from University of Wales, UK. His main area of research is artificial intelligence, digital design and data mining. Currently, he is working as an associate professor as well as Dean of the College of Computer Engineering and Sciences and Vice Rector of Salman Bin Abdulaziz University, Saudi Arabia. He is also, a consultant for several Government Organizations and a member of councils of various boards and commissions. Aljumah has published a number of research papers in reputed Conferences and Journals.

Mohammad Siddiqui received his BTech in computer science engineering from Uttar Pradesh Technical University, India and MS from BITS, India. His research field is databases, data warehousing and application of data mining. Currently, he is working as a Researcher in College of Computer Engineering and Sciences, Salman bin Abdulaziz University, Saudi Arabia. He worked as an Oracle DBA for various telecom based project. he had done funded projects by Deanship of Scientific Research, Salman bin Abdulaiz University, Ministry of Higher Education, Saudi Arabia. He has published a number of research papers in reputed Conferences and Journals. He is reviewer of various reputed Journal.