

## Research Article

### Hypertension Interventions using Classification Based Data Mining

Abdullah A. Aljumah and Mohammad Khubeb Siddiqui

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

**Abstract:** In the present study, we would like to gain the insight of the medical data through classification based data mining technique. The data sets of NCD (Non Communicable Diseases) risk factors, a standard report of Saudi Arabia 2005, in collaboration with WHO (World Health Organisation, 2005) were employed on Saudi hypertension patients. Computing the probability and prediction of hypertension disease intervention are evaluated through ROC (Receiver Operating Characteristics). The Area under Curve (AUC) of depicted ROC plots are calculated, the AUC of ROC is the indicative of prediction of the intervention. The AUC of ROC of five hypertension interventions is obtained based on which we distinguish which mode of intervention is more appropriate. Present analysis predicts that smoking cessation is the best intervention followed by exercise, diet, weight and drug for the hypertension intervention in Saudi Arabia.

**Keywords:** Classification, data mining, hypertension, Naïve Bayesian algorithm, Receiver Operating Characteristics (ROC)

#### INTRODUCTION

Data mining provides a way to get at the information buried in the data. Data mining creates models to find hidden patterns in large, complex collections of data, patterns that sometimes elude traditional statistical approaches to analysis because of the large number of attributes, the complexity of patterns, or the difficulty in performing the analysis. Data mining is the analysis of observational data set to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Data mining typically deals with data that have already been collected for some purpose other than data mining analysis. The ROC is a technique under data mining classification. ROC curves are very helpful for classification and prediction analysis of the performance it is common practice that ROC are used in medical decision. In medical diagnosis and intervention good ranking classification is highly desirable for the better performance of the intervention and diagnosis. This can be done through ROC curves and computing Area under Curve (AUC). In the present study the five modes of intervention of hypertension are drug, diet, weight, smoke cessation and exercise. The ROC curves have been plotted for these different five modes of intervention and AUC's are computed using the Oracle Data Miner 10 g tool. The results are compared on the basis computed AUC.

Hypertension may lead to serious conditions, such as heart disease and stroke. High blood pressure

symptoms may be insidious; signs of seriously elevated high blood pressure include severe headache, vision problems and chest pain. The exact causes of high blood pressure are unknown but may be related to hereditary and environmental factors. Interventions of hypertension includes drug, diet, weight, smoke cessation and exercise. The goals of intervention are to reduce blood pressure and control the heart disease. Intravenous (given directly into the vein) medications may be prescribed if hypertension is extremely severe and intensive care is necessary. Blood pressure should be checked at regular intervals (as recommended by your health care provider) to monitor the condition. Frequent blood pressure measurements taken at home are often recommended for people with difficult-to-control high blood pressure (Bernard *et al.*, 2007) states that hypertension is a common disease that contributes importantly to the high cardiovascular morbidity and mortality observed in industrialized countries. The proper diagnosis and management of this disorder affords considerable reduction of the risk of developing cardiac, cerebral and renal complications. Approximately 95% of patients with high blood pressure exhibit the so-called essential or primary form of hypertension. Various mechanisms are involved in the pathogenesis of this type of hypertension. This heterogeneity accounts for the diverse therapeutic approaches that have been utilized and for the rationale for individualizing intervention programs. In a small fraction of patients, the elevation of blood pressure is due to a specific cause (secondary hypertension). The

**Corresponding Author:** Mohammad Khubeb Siddiqui, College of Computer Engineering and Sciences, Salman bin Abdulaziz University, Saudi Arabia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

recognition of such patients has improved markedly in recent years. This is relevant since secondary hypertension can often be cured by appropriate interventions.

**Modes of intervention:** As mentioned in World Health Organization's NCD report of Saudi Arabia following five types of intervention are discussed below.

**Drug:** Drug intervention to lower a patient's blood pressure is worthwhile even if the blood pressure does not fall to the target of 140/90 mmHg or below on intervention with several drugs (or if the addition of more drugs is inappropriate or declined). Diuretics are usually recommended as the first line of therapy for most people who have high blood pressure. If one drug doesn't work or is disagreeable, other types of diuretics are available (<http://www.mayoclinic.com/health/high-blood-pressure/DS00100/DSECTION=treatments-and-drugs>).

**Diet:** A variety of dietary modifications are beneficial in the intervention of hypertension. This includes:

- Reduction of sodium intake
- Moderation of alcohol
- Weight loss in the obese patients
- Possibly increasing potassium and calcium intake
- Ingestion of a vegetarian diet or fish oil supplements (Chobanaian *et al.*, 2003)

**Weight:** There is a direct association between blood pressure and body weight and/or abdominal adiposity. Weight reduction confers a range of other cardiovascular health benefits including reduced insulin resistance and hyperlipidaemia and reduced risk of left ventricular hypertrophy and obstructive sleep apnoea. Mancia *et al.* (2007) every 1% reduction in body weight lowers systolic blood pressure by an average of 1 mmHg. National Heart Foundation of Australia (2008) states that losing 4.5 kg reduces blood pressure or prevents hypertension in a large proportion of overweight people, while losing 10 kg can reduce systolic blood pressure by 6-10 mmHg. In overweight patients with hypertension, the weight-reducing diets can achieve a 3-9% decrease in body weight and may reduce systolic and diastolic blood pressure by approximately 3 mmHg. The weight reduction confers a range of other cardiovascular health benefits including reduced insulin resistance and hyperlipidaemia and reduced risk of left ventricular hypertrophy and obstructive sleep apnoea (Mancia *et al.*, 2007).

**Smoke cessation:** Smoking is a strong independent risk factor for cardiovascular diseases. Quitting is acknowledged to be one of the most effective lifestyle interventions for preventing cardiovascular disease and

premature deaths. Smoking causes an immediate increase in blood pressure and heart rate that persists for more than 15 min after one cigarette. People who smoke show higher ambulatory blood pressure levels than non-smokers (Mancia *et al.*, 2007). Elevated blood pressure and smoking are the two most important risk factors for subarachnoid haemorrhage in the Asia-Pacific region. The risk of myocardial infarction is 2-6 times higher and the risk of stroke is three times higher in people who smoke, compared with non-smokers (National Heart Foundation of Australia, 2008). The smoking cessation markedly reduces overall cardiovascular risk, including the risk of coronary heart disease and stroke, compared with continued smoking. In patients with coronary heart disease, smoking cessation is associated with a 36% reduction in the risk of all-cause mortality (Critchley and Capewell, 2004). Although smoking is known to increase the risk of developing hypertension, there is currently no evidence that smoking cessation directly reduces blood pressure in people with hypertension (Mancia *et al.*, 2007).

**Exercise:** It is clear that exercise (physical activity) lowers resting and daytime ambulatory blood pressure. In clinical trials of people with hypertension, regular aerobic activity reduced systolic blood pressure by an average of 6.9 mmHg and diastolic blood pressure by 4.9 mmHg (Fagard and Cornelissen, 2007). Regular exercise has an independent cardio protective effect. Regular exercise is associated with an increase in high-density lipoprotein cholesterol and with reductions in body weight, waist circumference, percentage body fat, insulin resistance, systemic vascular resistance, plasma noradrenaline and plasma renin activity (Fagard and Cornelissen, 2007).

## LITERATURE REVIEW

The literature survey reveals many results on hypertension intervention. Abdullah *et al.* (2013) proposed work shows various models for each type of diabetic intervention and analysis is carried out using classification based data mining technique. Almazayad *et al.* (2010) have investigated on the data sets of NCD (Non Communicable Diseases) risk factors, a standard report of Saudi Arabia 2005, in collaboration with WHO (World Health Organisation, 2005) 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 to 64 years. The five age group of NCD data had been put into two age groups of young and old denoted as 'Y' and 'O' respectively. Data mining showed that the all the five modes of interventions were effective for older people due to hypertension at older ages

(Williams, 2009). Hypertension is one of the most important causes of cardiovascular morbidity and mortality and its intervention is a major focus of primary and secondary disease prevention strategies. The intervention of hypertension continues to evolve and the need for guidance on the use of newer screening tools, techniques for blood pressure measurement and different classes of drug therapies led to the first European guidelines for the management of arterial hypertension being issued in 2003 by the European Society of Hypertension (ESH) and the European Society of Cardiology (ESC). The first update of these guidelines in 2007 crystallized much of the modern thinking about the evaluation and intervention of patients with hypertension with a sharp focus on detailed assessment of subclinical organ damage and cardiovascular disease risk, as well as differential blood pressure intervention targets and thresholds for those at different levels of risk. This review focuses on the 2007 ESH/ESC Guidelines, highlighting the evolution of intervention strategies in order to meet the challenge of improving blood pressure control in Europe. In particular, development of patient-centred intervention strategies, the benefits of blood pressure lowering, drug-specific influences over clinical outcomes, recommendations for the pharmacological intervention of hypertension and the role of combination therapies are discussed.

Al-Nozha *et al.* (2007) have determined the prevalence of hypertension among Saudis in both genders, between the ages of 30-70 years in rural as well as urban communities. They have carried out the study on 17,230 subjects. This study was part of a major national study on Coronary Artery Disease in Saudis Study (CADISS). This was a community-based study conducted by examining subjects in the age group of 30-70 years of selected households during a 5-year period between 1995 and 2000 in Saudi Arabia. The data has been obtained from history using a validated questionnaire and examination including measurement of blood pressure. The data were analyzed to provide prevalence of hypertension. This study used the Logistic regression technique to develop a risk assessment model for prevalence of hypertension. The prevalence of hypertension was 26.1% in crude terms. For males, the prevalence of hypertension was 28.6%, while for females; the prevalence was significantly lower at 23.9% ( $p < 0.001$ ). The urban population showed significantly higher prevalence of hypertension of 27.9%, compared to rural population's which showed lower prevalence of 22.4% ( $p < 0.001$ ). The prevalence of CAD among hypertensive patients was 8.2 and 4.5% among normotensive subjects ( $p < 0.001$ ). The increasing weight was directly proportional to the increased prevalence of hypertension. El-Hazmi and Warsy (1999) have presented the hypertension studies in Saudi Arabia. As per the study they have compared a

few isolated and three comprehensive studies covering the whole Kingdom of Saudi Arabia. The investigators found different prevalence of hypertension in different areas of the kingdom. They have unified the diagnostic procedures and determined the factors behind such significant differences. Al-Nozha *et al.* (1997) screened 13,700 individuals of both sexes in all age groups. As per the WHO standards the blood pressure  $>160/95$  mmHg is considered as hypertension. Nazim-Uddin (1994) examined individuals with ages ranging between 18-50 years in the northern province of Saudi Arabia. He reported that if diastolic blood pressure of  $>90$  mmHg was considered as the cut-off point for definition of hypertension, the prevalence of hypertension would be 15.2%, while if  $>95$  mmHg was used, the prevalence of hypertension would be 5.25%. Al-Nozha and Osman (1998) reported the results of a National Nutrition Survey, in which blood pressure was checked in 17,892 individuals from age 12 years and above. There were 6260 adults over 18 years of age. The prevalence of systolic and diastolic hypertension was determined by using two cut-off values i.e.,  $>160/95$  and  $>140/90$ . The prevalence of systolic and diastolic hypertension in the adult population was 5.3 and 7.3%, respectively, using the cut-off value of 160/95. There were significant geographical variations. The prevalence of systolic blood pressure was the highest in Taif, Farasan and Hail and lowest in Asir, Jizan and AlMadina. The prevalence of diastolic blood pressure was highest in Al-Qassim, Jeddah, Tabouk and Hail and lowest in Makkah. Interestingly, females generally showed a higher prevalence compared to the male counterparts in all geographical areas. As per the university of York report (Effective Matters, 1999), the prevalence of hypertension peaked in the 40-70 years age group and increased significantly compared to the prevalence in the 18-29 and 30-39 years age group. Hypertension is widely considered to be one of the most important risk factors for these diseases and is strongly associated with death from stroke, congestive heart failure and coronary heart disease. Using a conservative definition of essential hypertension ( $>160$  mmHg systolic or 95 mmHg diastolic), it is estimated that in the UK more than half of the 10 million people over the age of 65 are hypertensive. The high prevalence of the disease and rapidly growing number of older people suggest that hypertension is an enormous public health problem and yet detection and intervention remain relatively low. Ira and Ozlem (2009) developed the predictive analysis to classify sixteen diseases based on discharge summary in this study, the authors have predicted whether each disease is present, absent or questionable. This has been a multiclass classification task, they have evaluated these sixteen diseases and showed that it improves significantly over voting and stacking, when they had used for multiple class classification.

**MATERIALS AND METHODS**

**Data collection:** The searched database was collected from World Health Organisation’s web link titled ‘Data and Statistics’, which was NCD risk factor, a standard report of Ministry of Health, Saudi Arabia, 2005 ([http://www.who.int/chp/steps/2005\\_Saudi Arabia\\_STEPS\\_Report\\_EN.pdf](http://www.who.int/chp/steps/2005_Saudi_Arabia_STEPS_Report_EN.pdf)).

The table ‘hyp\_intervention’ have been designed in oracle 10g database.

The table includes 7 columns (sr\_no, age, N, small\_n, Percentage, se, typ\_intervention). The ‘sr\_no’ indicates the serial number, which acts as a 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 who have been cured with the particular type of intervention, the percentage indicates the percent of cured patients by specific mode of intervention and ‘se’ indicates the standard error, the ‘typ\_intervention’ column is externally added in the table just to easily classify the data on the basis of patient intervention the column indicates the mode of intervention (Table 1).

**Tools and techniques:** Different types of data mining tools are available and each has its own merits and demerits. For the analysis of data of WHO’s NCD report of Saudi Arabia, we have concentrated on data of hypertension patients and studied predictive analysis to investigate which mode of intervention is more

Table 1: Dataset of hypertension

SR_No	Age	N	Small_N	(%)	S.E.	TYP_intervention
1	15-24	10	3	32.2	16.2	Drug
2	25-34	19	6	29.9	11.5	Drug
3	35-44	35	23	64.3	8.8	Drug
4	45-54	77	62	81.3	4.8	Drug
5	55-64	99	90	90.8	2.6	Drug
6	15-24	10	2	19.7	13.0	Diet
7	25-34	19	2	10.8	6.8	Diet
8	35-44	35	21	60.4	9.5	Diet
9	45-54	77	45	58.7	6.6	Diet
10	55-64	99	52	53.0	8.5	Diet
11	15-24	10	2	19.7	13.0	Weight
12	25-34	19	5	27.0	10.4	Weight
13	35-44	35	17	48.5	9.1	Weight
14	45-54	77	26	33.4	5.3	Weight
15	55-64	99	39	39.7	7.8	Weight
16	15-24	10	3	32.3	16.2	Smoke_cessation
17	25-34	19	2	10.8	7.3	Smoke_cessation
18	35-44	35	4	11.7	5.5	Smoke_cessation
19	45-54	77	13	16.7	4.4	Smoke_cessation
20	55-64	99	32	13.3	2.5	Smoke_cessation
21	15-24	10	3	32.3	16.2	Exercise
22	25-34	19	6	33.3	10.2	Exercise
23	35-44	35	13	36.9	7.9	Exercise
24	45-54	77	23	29.9	5.6	Exercise
25	55-64	99	28	27.9	5.4	Exercise

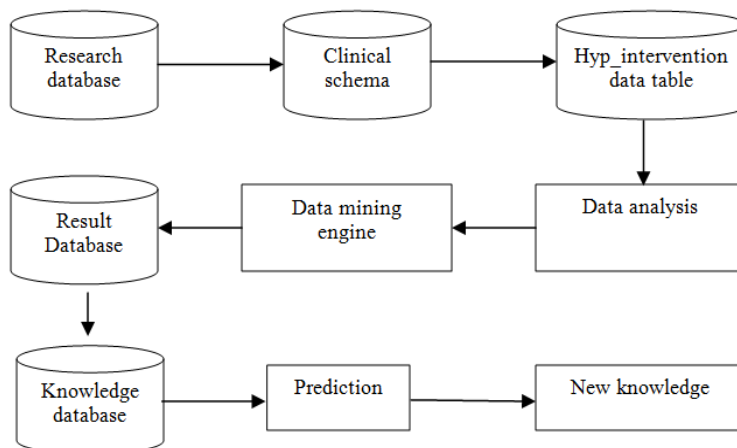


Fig. 1: Architecture of medical data mining

effective using classification technique of data mining. We have used the oracle data miner version 10.2.0.3.0.1; build 2007 for the mining activity that act as a client and oracle 10 g database release 10.2.0.3.0 as a server.

The data mining activity as shown in Fig. 1 is visualized as three tier process. Three-tier architecture are medical data mining, data integration and visualization system. The first tier is database server tier, which consists of:

- Source medical data
- Schema mappings
- Ontologies

Where data and metadata are prepared and stored. The database for research is termed as research database and the schema is termed as clinical schema that is authenticated by password. In clinical schema all database objects are designed and incorporate to interact with ODM. The master table named as 'hyp\_intervention' is designed which includes the attributes 's\_no', 'age', number of patient 'N', patient cured 'small\_n', 'percent' and 'typ\_intervention'. The middle tier is called 'data mining application' which also holds three items:

- Data analysis
- Data mining engine
- Result database

Data analysis is the step where data is analysed as to maintain the research objective like selecting target value from master table. In data mining engine, the data mining technique is a suite of algorithms such as SVM, Naïve Bayesian etc., for business logic events. The desired algorithm is to be chosen, settings the desired parameters of algorithms to process the raw data and store the results in the result database. The third tier holds the knowledge database. This predicts the new knowledge as result data. This is shown as knowledge database that is informative data which facilitates the forecast the problem on the basis of prediction, probabilities and visualization of the results in form of new knowledge or mined data.

Oracle 10 g Data Miner (ODM) is the mining tool that we selected in presented research. ODM version 10.1 is data mining software embedded in the Oracle 10 g Database Enterprise Edition (EE) that enables one to discover new insights hidden in data (Berger, 2004). The Oracle Data Mining suite is made up of two components, the data mining Java API and the Data Mining Server (DMS) (Oracle Data Mining, 2004). The DMS is a server-side, in-database component that performs data mining that is easily available and scalable. The DMS also provides a repository of

metadata of the input and result objects of data mining. The Oracle Data Miner is a component that resides within the DMS and is responsible for the actual data mining.

All data mining components have been added to the Oracle Data Miner, this also includes the ODM Browser interface which provides a full set of mining wizards.

Berger (2004) states that Oracle Data Miner supports supervised learning techniques (classification, regression and prediction problems), unsupervised learning techniques (clustering, associations and feature selection problems), attribute importance techniques (for finding key variables), text mining and has a special algorithm for life sciences sequence searching and alignment problems. The availability of these algorithms provides all the necessary tools required in gathering information from the dataset.

The main advantage of using Oracle Data Miner is that all data mining processing occurs within the Oracle database. We took this into consideration when selecting the mining tool to use. Other mining tools force you to extract the data out of the database before the actual mining process which may result in a number of flaws in these mining tools. The aspect of ODM where everything occurs in the database results in a more secure and stable data management which enhances productivity as the data does not have to be extracted from the database before mining it.

All the functionalities and advantages mentioned here have resulted in Oracle Data Miner being the most preferred mining tool to use when performing the algorithm evaluation and gathering of information from the dataset.

Oracle Data Miner helps data analyst mine their oracle data to find valuable hidden information, patterns and new insights. Data analysts can mine data with Oracle Data Miner's easy-to-use wizards that guide them through the data preparation, data mining, model evaluation and model scoring process. As the data analyst transforms the data, builds models and interprets results, Oracle Data Miner can automatically generate code needed to transform the data mining steps into an integrated data mining/BI application (<http://www.oracle.com/technology/products/bi/odm/odminer.html>).

ODM ([www.oracle.com/technology/documentation/datamining.html](http://www.oracle.com/technology/documentation/datamining.html)) is applicable in a variety of business, public sector, health care and other environments. The common thread running through all data mining projects is the goal of analyzing individual behavior.

The term "behavior" can be interpreted, to include:

- The vulnerability of an individual to a certain disease
- The purchasing habits of a customer

- The likelihood that an item passing through an assembly line will be flawed
- The characteristics observed in an individual indicating membership in a particular segment of the population

We adopt Oracle Data Miner (ODM) for ROC analysis on modes of hypertension intervention from the database of World Health Organisation (WHO). We have designed the database in Oracle 10 g that acts as server and ODM act as a client.

For running the Oracle Data Miner following privileges are required to the clinical schema which would connect to oracle database. For this purpose the SQL (Structured Query Language) query language has been used to access the database tables.

The table in the schema speaks about the mode of hypertension intervention i.e., drug, diet, weight, smoke cessation and exercise.

We had applied the classification technique, which is well known supervised mining technique that the data mining commonly utilizes. In our case it takes 'percentage' as numerical attribute which acts as its target for prediction of the results and compare the mode of hypertension intervention with the help of ROC.

**Classification:** Classification is one of the most important task in data mining, Classification of a collection consists of dividing the items that make up the collection into categories or classes. In the context of data mining, classification is done using a model that is built on historical data. The goal of predictive classification is to accurately predict the target class for each record in new data, that is, data that is not in the historical data.

A classification task begins with build data (also known as training data) for which the target values (or class assignments) are known. Different classification algorithms use different techniques for finding relations between the predictor attributes' values and the target attribute's values in the build data. These relations are summarized in a model; the model can then be applied to new cases with unknown target values to predict target values. A classification model can also be applied to data that was held aside from the training data to compare the predictions to the known target values; such data is also known as test data or evaluation data. The comparison technique is called testing a model, which measures the model's predictive accuracy. The application of a classification model to new data is called applying the model and the data is called apply data or scoring data. Applying a model to data is often called scoring the data.

While applying the Classification mining techniques on ODM needs the target attribute the target should be discrete value e.g., 0 or 1, male or female etc.

In the present study the target attribute is 'treat\_hyp' attribute that includes the value of mode of intervention that acts as discrete value, means the distinct value of this column shows the all five mode of hypertension intervention.

**Algorithms for classification:** Oracle Data Miner provides the following algorithms for classification:

- Decision Tree Algorithm
- Naive Bayes Algorithm
- Adaptive Bayes Network Algorithm
- Support Vector Machine Algorithm

We adopted the Naïve Bayes Algorithm for applying classification technique on the data.

**Naïve Bayesian:** The Naïve Bayes algorithm can be used for both binary and multiclass classification problems. NB builds and scores models extremely rapidly; it scales linearly in the number of predictors and rows. NB makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence. Bayes' Theorem states that the probability of event A occurring given that event B has occurred ( $P(A|B)$ ) is proportional to the probability of event B occurring given that event A has occurred multiplied by the probability of event A occurring ( $(P(B|A)P(A))$ ).

Naive Bayes makes the assumption that each attribute is conditionally independent of the others, that is, given a particular value of the target, the distribution of each predictor is independent of the other predictors. In practice, this assumption of independence, even when violated, does not degrade the model's predictive accuracy significantly and makes the difference between a fast, computationally feasible algorithm and an intractable one.

**ROC:** ROC 'Receiver Operating Characteristics' is the powerful statistical technique used in signal detection in radio frequency band. Presently this has been a marvellous technique in medical data mining under classification analysis to classify the medical data. ROC is a type of curve that shows the trade-off between the true positive rate or sensitivity (proportion of positive tuples that correctly identified) and the false positive rate (proportion of negative tuples that are incorrectly identified as positive). It allows us to visualize the trade-off between the rate at which the model can accurately recognize 'yes' cases versus the rate at which it mistakenly identifies 'no' cases as 'yes' for different portions of the test set. Any increase in the true positive rate occurs at the cost of an increase in the

false positive rate. The area under the ROC curve is measure of accuracy of the model.

The ROC curve allows us to explore the relationship between the sensitivity and specificity of a clinical test for a variety of different thresholds points, thus allowing the determination of an optimal value. In order to determine the presence or absence of a disease. We often have to carry out a test which provides a result on a continuous measure.

The Oracle Data Mining ROC computation calculates the following statistics:

- **Probability threshold:** The minimum predicted positive class probability resulting in a positive class prediction. Different threshold values result in different hit rates and false alarm rates
- **True negatives:** Negative cases in the test data with predicted probabilities strictly less than the probability threshold (correctly predicted)
- **True positives:** Positive cases in the test data with predicted probabilities greater than or equal to the probability threshold (correctly predicted)
- **False negatives:** Positive cases in the test data with predicted probabilities strictly less than the probability threshold (incorrectly predicted)
- **False positives:** Negative cases in the test data with predicted probabilities greater than or equal to the probability threshold (incorrectly predicted)
- **Hit rate ("true positives" in oracle data miner):** (true positives/ (true positives + false negatives))
- **False alarm rate ("false positives" in oracle data miner):** (false positives/ (false positives + true negatives))

$$\text{True Positive Rate} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

$$\text{False Positive Rate} = \text{False Positive} / (\text{True Negative} + \text{False Positive})$$

The vertical and horizontal axis of ROC curve represents the true positive rate and false positive rate respectively.

The ROC has been used in the machine learning and data mining for model evaluation particularly for medical decision making. ROC graph are drawn the true positive rate as a function of false positive rate. The curves are generated using the classification threshold from positive value to the most negative value. In a random classification the ROC curve is straight line with 45° of phase angle with the x-axis or whose slope is  $m = 1$ . The area under the curve is a measure of accuracy of the intervention or the diagnostics etc.

**Area under curve:** The area under curve measures the discriminating ability of a binary classification model. The larger the AUC, the higher the likelihood that an

actual positive case would be assigned a higher probability of being positive than actual negative case. AUC is a portion of the area of the unit square; its value always lies between 0 and 1.0.

The Area under curve is defined as the area or space under the ROC curve.

Let us mathematically defined Area Under Curve,  $n$  a binary classification with 'm' positive values and 'n' negative values, let 'C' be a fixed classifier that outputs strictly ordered list for these values. Let 'x1.....xm' be the output of 'C' on the positive values and 'y1.....yn' its output on the negative values and denoted by  $1x$  the indicator function of a set  $x$ . AUC, denoted as 'A' which is associated to 'C' is given computed by:

$$A = (\sum_{i=1}^m \sum_{j=1}^n 1_{x_i > y_j}) / mn$$

which is the value of the Wilcoxon Mann Whitney Statistics, (Hanaley and McNeil, 1982).

Thus, the AUC is closely related to the ranking quality of the classification. It takes up the comparison between two classifications and estimates the probability  $P_{xy}$ . If all positive values are ranked higher than the negative values then the area under curve 'A' becomes 1. Any deviation from this ranking decreases the AUC and the expected AUC value for random ranking is 0.5.

## EXPERIMENTAL ANALYSIS

We have carried out the ROC analysis on the NCD data of Saudi Arabia using Oracle Data Miner tool. The AUC of ROC analysis is obtained for the prediction of hypertension intervention.

**Drug:** The Fig. 2 is the ROC of drug intervention for the hypertension. The AUC of the ROC for this intervention is obtained as 0.61. As per the ROC the more the area the better is the intervention mode but curve is gracing the diagonal of ROC, or the curve is far away from the upper left corner which indicates the poorer intervention as compared to other modes of intervention.

**Diet:** The Fig. 3 is the ROC of diet intervention. The estimated AUC value is 0.72 of this ROC and is relatively large and it is covering the upper left corner of ROC curve which is the indicative of better intervention as compared to drug.

**Weight:** The Fig. 4 of ROC curve shows for weight reduction intervention. The AUC of this ROC is estimated at 0.60. The AUC of drug and weight are almost same 0.61 and 0.60, respectively and they are far away from upper left corner and the AUC is closed to diagonal of ROC which indicating relatively less

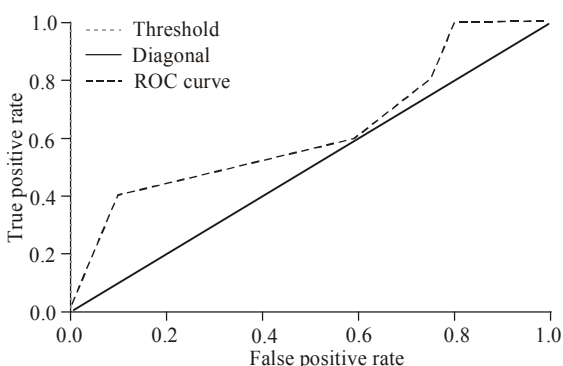


Fig. 2: ROC curve of intervention drug

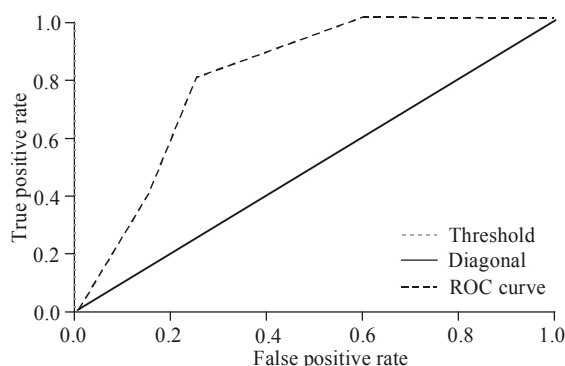


Fig. 5: ROC curve of intervention smoking cessation

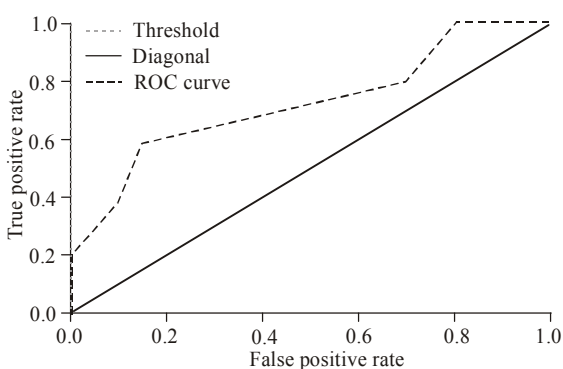


Fig. 3: ROC curve of intervention diet control

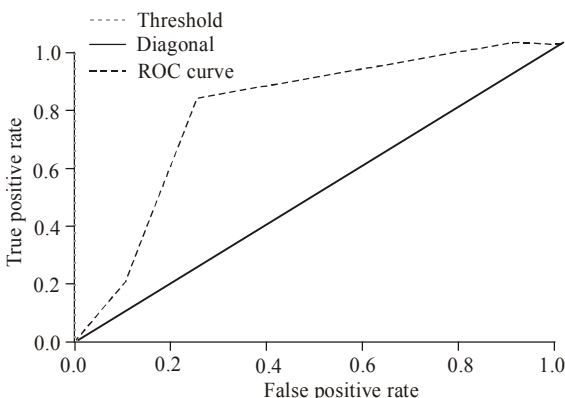


Fig. 6: ROC curve of intervention exercise

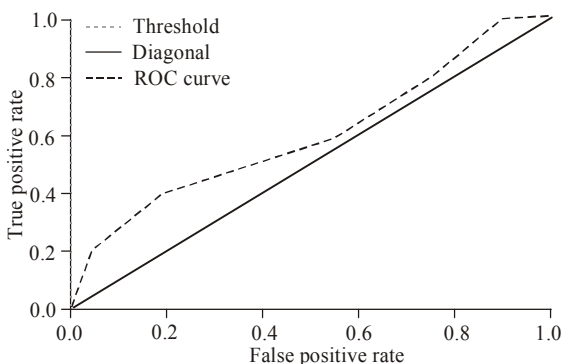


Fig. 4: ROC curve of intervention weight reduction

effective intervention for the hypertension intervention. Weight reduction confers a range of other cardiovascular health benefits including reduced insulin resistance and hyperlipidaemia and reduced risk of left ventricular hypertrophy and obstructive sleep apnoea.

**Smoke cessation:** The Fig. 5 is the ROC of smoking cessation intervention. The AUC is of this ROC is obtained as 0.79. The AUC value is highest of all the interventions. The AUC is large and deeply closer to upper left corner of the ROC indicating an excellent intervention mode for the hypertension intervention.

**Exercise:** The above Fig. 6 is the ROC graph of exercise intervention. The AUC of the ROC is estimated as 0.76 which is second highest as compared to other intervention. The AUC is closer to upper left corner of ROC. This indicates the second best intervention of hypertension.

## DISCUSSION

The values of area under ROC curve of all five modes of intervention are shown in Table 2.

The statistical view of ROC plot of types of hypertension intervention is illustrated Fig. 7.

The calculated area under ROC curve is stated in Table 2 and the statistical chart that is made on the basis of AUC values has been shown in Fig. 7.

We can say that the AUC of smoking cessation and exercise are approximately equal. This has been reported by Almazayad *et al.* (2010).

The data mining technique of classification has employed using the Oracle Data miner. The ROC curves have been depicted in figures. The area under curve has been tabulated in Table 2. From the table AUC of smoking cessation is highest 0.79. Avoiding the smoking is the vital role in reducing the hypertension. This is followed by exercise 0.76 which indicates exercise intervention is better in contribution



Table 2: AUC value corresponding to interventions

Intervention	Area under curve
Drug	0.61
Diet	0.72
Weight	0.60
Smoke cessation	0.79
Exercise	0.76

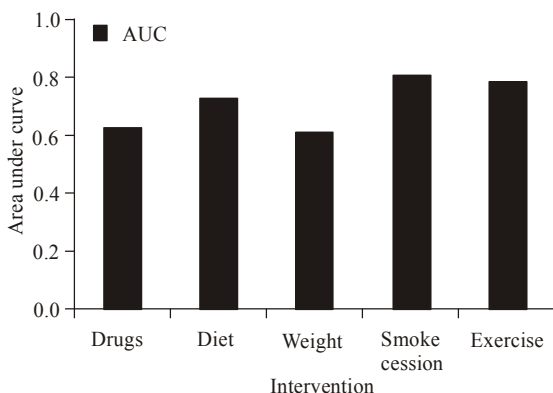


Fig. 7: Hypertension intervention

to the hypertension intervention. Then comes to the diet intervention, whose AUC is 0.72 it means that diet control has its significant effect on hypertension intervention. Lastly, the AUC of drug and weight are 0.61 and 0.60, respectively. However, these factors contribute relatively small when compared to other intervention. If we classify then through grading system for hypertension intervention, we can justify in saying, smoking cessation and exercise are best intervention, followed by diet control which is the better intervention and drug and weight intervention are very good intervention to have an overall grip over the hypertension. Smoking cessation markedly reduces overall cardiovascular risk, including the risk of coronary heart disease and stroke, compared with continued smoking. In patients with coronary heart disease, smoking cessation is associated with a 36% reduction in the risk of all-cause mortality. Although smoking is known to increase the risk of developing hypertension, there is currently no evidence that smoking cessation directly reduces blood pressure in people with hypertension (Mancia *et al.*, 2007). The hypertension is, in fact, significantly related to an increased risk of developing myocardial infarction, stroke, congestive heart failure, chronic kidney failure and mortality (Staessen *et al.*, 2003).

## CONCLUSION

In the present investigation, on the data of NCD report of WHO of Saudi Arabia, we have conducted the comparative study of all five types of hypertension intervention by using ROC curve analysis of classification technique using data mining. In the ROC analysis, the AUC is the indicator of prediction, more the calculated AUC value better the prediction. In the present study the AUC of ROC curve for smoking

cessation has the highest value. Hence, we can predict that smoking cessation is the best intervention followed by exercise, diet, weight and drug for the hypertension intervention in Saudi Arabia. Hence, all hypertension patients are unambiguously advised to stop smoking.

## REFERENCES

- Abdullah, A.A., K.S. Mohammad and G.A. Mohammad, 2013. Application of classification based data mining technique in diabetes care. *J. Appl. Sci.*, 13: 416-422.
- Almazyad, A.S., M.G. Ahamad, M.K. Siddiqui and A.S. Almazyad, 2010. Effective hypertensive treatment using data mining in Saudi Arabia. *J. Clin. Monitor. Comp.*, 24(6): 391-401.
- Al-Nozha, M.M. and A.K. Osman, 1998. The prevalence of hypertension in different geographical regions of Saudi Arabia. *Ann. Saudi Med.*, 18: 401-7.
- Al-Nozha, M.M., M.S. Ali and A.K. Osman, 1997. Arterial hypertension in Saudi Arabia. *Ann. Saudi Med.*, 17: 170-4.
- Al-Nozha, M., M. Mansour, M. Abdullah, M.R. Arafah, M.Z. Khalil, N.B. Khan, Y.Y. Al-Mazrou, M.A. Al-Maatouq, K. Al-Marzouki, A. Al-Khadra, M.S. Nouh, S.S. Al-Harhi, M.S. Al-Shahid and M.A. Al Mobeireek, 2007. Hypertension in Saudi Arabia. *Saudi Med. J.*, 28(1): 77- 84.
- Berger, C., 2004. Oracle Data Mining (Know More, Do More, Spend Less). An Oracle White Paper, Retrieved from: [http://www.oracle.com/technology/products/bi/odm/pdf/bwp\\_db\\_odm\\_10gr1\\_1004.pdf](http://www.oracle.com/technology/products/bi/odm/pdf/bwp_db_odm_10gr1_1004.pdf). (Accessed on: Jan 3, 2011).
- Bernard, W., B. Hans-Rudolph, B. Michel and N.C. Jay, 2007. Hypertension. *Cardiovascular Medicine*. Springer, London.
- Chobanaian, A.V., G.L. Bakris, H.R. Black and W.C.ushman, 2003. The seventh report of the joint national committee on prevention, detection, evaluation and intervention of high blood pressure: The JNC 7 Report. *JAMA*, 289: 2560-72.
- Critchley, J. and S. Capewell, 2004. Smoking cessation for the secondary prevention of coronary heart disease. *Cochrane Database Syst. Rev.*, 2004(1): CD003041.
- Effective Matters, 1999. The University of York. 4(2), Retrieved from: <http://www.york.ac.uk/inst/crd/EM/em42.pdf>. (Accessed on: 26/3/2010).
- El-Hazmi, M.A. and A.S. Warsy, 1999. Hypertension in Saudi Arabia. *Saudi J. Kidney Dis. Transpl.*, 10: 365-371.
- Fagard, R.H. and V.A. Cornelissen, 2007. Effect of exercise on blood pressure control in hypertensive patients. *Eur. J. Cardiovasc. Prev. Rehabil.*, 14: 12-7.
- Hanaley, J.A. and B.J. McNeil, 1982. The meaning and use of the area under a Receiver Operating Characteristics (ROC) curve. *Radiology*, 143(1): 29-36.

- Ira, G. and U. Ozlem, 2009. Specializing for predicting obesity and its co-morbidities. *J. Biomed. Inform.*, 42(2009): 873-886.
- Mancia, G., G. De Backer, A. Dominiczak, R. Cifkova, R. Fagard, G. Germano *et al.*, 2007. Guidelines for the management of arterial hypertension. The task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J. Hypertens.*, 25: 1105-87.
- National Heart Foundation of Australia, 2008. Guide to Management of Hypertension. Assessing and Managing Raised Blood Pressure in Adults. Retrieved from: [http://www.heartfoundation.org.au/Professional\\_Information/Clinical\\_Practice/Hypertension.htm](http://www.heartfoundation.org.au/Professional_Information/Clinical_Practice/Hypertension.htm). (Accessed on: 25/3/2010).
- Nazim-Uddin, K., 1994. Prevalence of hypertension in Saudi Arabia. *Practitioner*, 5(11): 805-6.
- Staessen, J.A., J. Wang, G. Bianchi and W.H. Birkenhager, 2003. Essential hypertension. *Lancet*, 361: 1629-1641.
- Williams, B., 2009. The changing face of hypertension intervention: intervention strategies from the 2007 ESH/ESC hypertension guidelines. *J. Hypertens.*, 27: S19-S26.
- World Health Organisation, 2005. NCD Risk Factor, Standard Report of Ministry of Health, Saudi Arabia. Retrieved from: [http://www.who.int/chp/steps/2005\\_SaudiArabia\\_STEPS\\_Report\\_EN.pdf](http://www.who.int/chp/steps/2005_SaudiArabia_STEPS_Report_EN.pdf). (Accessed on: December 15, 2009).