

Deep Vein Thrombosis Predisposing Factors Analysis Using Association Rules Mining

*Muthanna D.R. AL-Assal

ABSTRACT

Background: DVT is a very common problem with a very serious complications like pulmonary embolism (PE) which carries a high mortality, and many other chronic and annoying complications (like chronic DVT, post-phlebotic syndrome, and chronic venous insufficiency) ,and it has many risk factors that affect its course, severity ,and response to treatment.

Objectives: Most of those risk factors are modifiable, and a better understanding of the relationships between them can be beneficial for better assessment for liable patients , prevention of disease, and the effectiveness of our treatment modalities. Male to female ratio was nearly equal , so we didn't discuss the gender among other risk factors.

Type of the study: A cross- sectional study.

Methods: Data taken from 114 patients with DVT were analyzed by association rules mining. Immobility was the most important risk factor.

Results: Smoking add more risk to immobile, post operative patient. Age per se has no effect. 100% of patients with long

bone fracture, were immobile. Fever occurred in one third of post operative patients who develop DVT. **Conclusions:** Association rules mining allow better and faster analysis of more data with an interactive powerful system, which saves time and effort, and discovers the relations among many factors to one or more than one factors. So, we use this method for analysis in this study, and we get the above mentioned relations, which are important for the future management of DVT.

Keywords: DVT risk factors, analysis of DVT risk factors, Data mining and DVT.

*Al-Kindy College Medical Journal 2016: Vol. 12 No.2
Page: 104-109*

*Lecturer/Dept. of Surgery , college of Al-Kindy medical college-Baghdad University.

Received 5th June 2016, accepted in final 3th Oct 2016

Corresponding to : Muthanna D.R. AL-assal

Knowledge discovery and data mining: With the enormous amount of data stored in files, databases, and other repositories, is increasingly important, if not necessary, to develop powerful means for analysis and perhaps interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process.

The Knowledge Discovery in Databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps:

- Data cleaning: also known as data cleansing, it is a phase in which noise data and irrelevant data are removed from the collection.
- Data integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.

- Data selection: at this step, the data relevant to the analysis is decided on and retrieved from the data collection.
- Data transformation: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- Data mining: it is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- Pattern evaluation: in this step, strictly interesting patterns representing knowledge are identified based on given measures.
- Knowledge representation: is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

It is common to combine some of these steps together. For instance, data cleaning and data integration can be performed together as a pre-processing phase to generate a data warehouse. Data selection and data transformation can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data.

The KDD is an iterative process. Once the discovered knowledge is presented to the user, the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed, or new data sources can be integrated, in order to get different, more appropriate results. Data mining derives its name from the similarities between searching for valuable information in a large database and mining rocks for a vein of valuable ore. Both imply either sifting through a large amount of material or ingeniously probing the material to exactly pinpoint where the values reside. It is, however, a misnomer, since mining for gold in rocks is usually called "gold mining" and not "rock mining", thus by analogy, data mining should have been called "knowledge mining" instead. Nevertheless, data mining became the accepted customary term, and very rapidly a trend that even overshadowed more general terms such as knowledge discovery in databases (KDD) that describe a more complete process. Other similar terms referring to data mining are: data dredging, knowledge extraction and pattern discovery [1, 2].

Deep Vein Thrombosis: Deep vein thrombosis, or DVT, is a blood clot that forms in a vein deep in the body. Blood clots occur when blood thickens and clumps together,[3,4]. Most deep vein blood clots occur in the lower leg or thigh. They also can occur in other parts of the body. A blood clot in a deep vein can break off and travel through the bloodstream. The loose clot is called an embolus. It can travel to an artery in the lungs and block blood flow. This condition is called pulmonary embolism, or PE. PE is a very serious condition. It can damage the lungs and other organs in the body and cause death [3,4,5]. Blood clots in the thighs are more likely to break off and cause PE than blood clots in the lower legs or other parts of the body. Blood clots also can form in veins closer to the skin's surface. However, these clots won't break off and cause PE,[3,4,5].

Causes Of Deep Vein Thrombosis: Blood clots can form in the body deep veins if:

- A vein's inner lining is damaged. Injuries caused by physical, chemical, or biological factors can damage the veins. Such factors include surgery, serious injuries, inflammation, and immune responses.
- Blood flow is sluggish or slow. Lack of motion can cause sluggish or slow blood flow. This may occur after surgery, if the patient is ill and in bed for a long time, or if he were traveling for a long time.
- The blood is thicker or more likely to clot than normal. Some inherited conditions (such as factor V Leiden) increase the risk of blood clotting. Hormone therapy or birth control pills also can increase the risk of clotting, i.e. Virchow's triad.

- The risk factors for deep vein thrombosis (DVT) include:
 - A history of DVT.
 - Conditions or factors that make the blood thicker or more likely to clot than normal. Some inherited blood disorders (such as factor V Leiden) will do this. Hormone therapy or birth control pills also increase the risk of clotting.
 - Injury to a deep vein from surgery, a broken bone, or other trauma.
 - Slow blood flow in a deep vein due to lack of movement. This may occur after surgery, if the patient is ill and in bed for a long time, or if he were traveling for a long time.
 - Pregnancy and the first 6 weeks after giving birth.
 - Recent or ongoing treatment for cancer.
 - A central venous catheter. This is a tube placed in a vein to allow easy access to the bloodstream for medical treatment.
 - Older age. Being older than 60 is a risk factor for DVT, although DVT can occur at any age.
 - Overweight or obesity.
 - Smoking.
- The risk for DVT increases if the patient has more than one of the risk factors listed above, [3,4,5].

Signs and Symptoms of Deep Vein Thrombosis:

- Only about half of the people who have DVT have signs and symptoms. These signs and symptoms occur in the leg affected by the deep vein clot. They include:
 - Swelling of the leg or along a vein in the leg.
 - Pain or tenderness in the leg, when standing or walking
 - Increased warmth in the area of the leg that's swollen or painful
 - Red or discolored skin on the leg
- Some people aren't aware of a deep vein clot until they have signs and symptoms of PE. Signs and symptoms of PE include:
 - Unexplained shortness of breath
 - Pain with deep breathing
 - Coughing up blood
 - Rapid breathing and a fast heart rate also may be signs of PE,[3,4,5].

Deep Vein Thrombosis Diagnosis :

- The doctor will diagnose deep vein thrombosis (DVT) based on the medical history, a physical exam, and test results. He or she will identify the risk factors and rule out other causes of the symptoms.

- For some people, DVT might not be diagnosed until after they receive emergency treatment for pulmonary embolism (PE).

Medical History: To learn about the medical history, the doctor may ask about:

- The overall health.
- Any prescription medicines in taking.
- Any recent surgeries or injuries the patient has.
- Whether the patient had been treated for cancer.

Physical Examination: The doctor will check the legs for signs of DVT, such as swelling or redness. He or she also will check the blood pressure and the heart and lungs.

Diagnostic Tests: The doctor may recommend tests to find out whether the patient has DVT

Common Test: The most common test for diagnosing deep vein blood clots is ultrasound. This test uses sound waves to create pictures of blood flowing through the arteries and veins in the affected leg .The doctor also may recommend a D-dimer test . A D-dimer test measures a substance in the blood that's released when a blood clot dissolves. If the test shows high levels of the substance, the patient might has a deep vein blood clot. If the test results are normal and the patient had few risk factors, DVT isn't likely,[5,6].

Other Tests : The patient may need blood tests to check whether he has an inherited blood clotting disorder that can cause DVT. This may be the case if the patient has repeated blood clots that are not related to another cause. Blood clots in an unusual location (such as the liver, kidney, or brain) also may suggest an inherited clotting disorder.

Other tests used to diagnose DVT include magnetic resonance imaging (MRI) and computed tomography ,or CT, scanning. These tests create pictures of the organs and tissues If the doctor thinks that the patient has PE, he or she may recommend more tests, such as a lung ventilation perfusion scan (VQ scan). A lung VQ scan shows how well oxygen and blood are flowing to all areas of the lungs,[5,6,7].

Methods: Study design: A prospective observational study was conducted from August the 1st 2011, to September the 1st 2013.

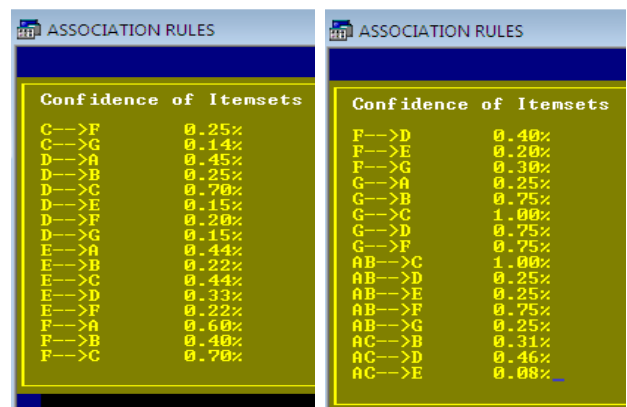
Study sample:The sample was collected from outpatient clinics in two general hospitals, and a private vascular clinic at Baghdad. All patients who was diagnosed to have DVT , was included in this study .Data that was collected include the name , age, gender, history of surgical intervention ,history of Contraceptive pills, history of immobility history of febrile illness before the DVT, smoking habits, alcohol intake, family history of DVT, , doplex scan results.....

Seven predisposing factors (age<=40, Contraceptive pills, immobility, febrile illness, history of surgery before DVT, smoking, pregnancy) was chosen to be analyzed by the association rules mining method.The total number of patients studied was 114.

Association Rules Mining: An Association Rule is a rule, which implies certain association relationships among a set of objects in a database. In this process people discover a set of association rules at multiple levels of abstraction from the relevant set(s) of data in a database. For example, one may discover a set of symptoms often occurring together with certain kinds of diseases and further study the reasons behind them. Since finding interesting association rules in databases may disclose some useful patterns for decision support, selective marketing, financial forecast, medical diagnosis, and many other applications, it has attracted a lot of attention in recent data mining research. In this section we introduce the standard definition of A-priori Algorithm that is used to discover the association rules,The *support* of an itemset I $sup(I)$, is defined as the number of transactions in the database containing I . The *minimum support* (min_sup), is a user predefined threshold. An itemset is *frequent* if its support is not less than the min_sup . An itemset with k items in called a k -itemset.

Let D be a set of transactions and $I = \{i_1, i_2, \dots, i_m\}$, an itemset is a subset of I . Given X and Y are itemsets, an *association rule* is of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \Phi$, where the $sup(X \cup Y) \geq min_sup$, and the *confidence* of $X \Rightarrow Y$ is not less than a predefined threshold, min_conf , the *confidence* of $X \Rightarrow Y$ is

$$\frac{sup(X \cup Y)}{sup(X)}$$



Figure(1) : shows some of The Association rules

Mining association rules may require iterative scanning of large transactions or relational databases, which is quite costly in terms of processing. Therefore, there is a need to study efficient mining of association rules in long transaction and /or relational databases. Or in other

word, it is necessary to evaluate efficiency of association rule mining among different algorithms [8]. To accomplish our aim which is to reduce the number of scans to one, let us discuss the following idea, when choosing set of items contains A, B, C the A-priori scans the database three times one for the sets which contains one item like A, B, C and the second scan is for the sets which contains two items like {A,B} , {A,C} , {B,C}, the third scan is for the sets with three items like {A, B, C}. Since the number of itemsets can be obtained as follows:

No. of itemsets = $2^n - 1$ where n is the number of items .

and by using the permutation of the giving items which obtaining by concatenations programming technique we can determine firstly the itemsets and we can simply compute the frequent of each itemset in the database by just one main scan[9]. In this research we applied the Interactive KDD (*Knowledge Discovery in Databases*) system for fast mining association rules[Figure2] [10].

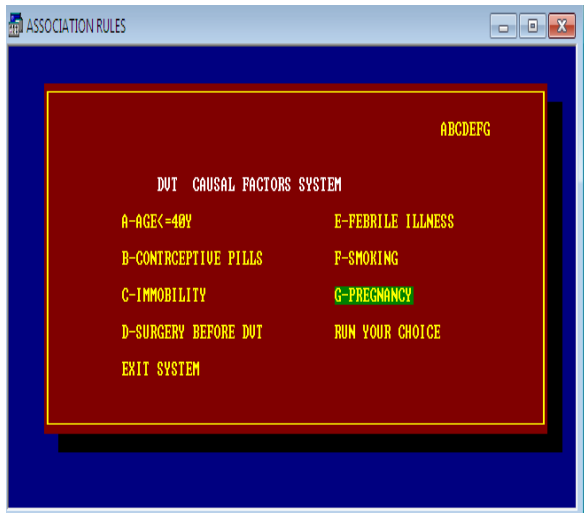


Figure (2) shows The KDD System Interface for Vein Thrombosis predisposing causal factors.

And for the first time we used the Mean of the items confidence (MOC) which can be defined as follows:

$$MOC(A) = \{conf(A \rightarrow B) + conf(A \rightarrow C) + conf(A \rightarrow D)\} / 3.$$

The formula gives the mean of the occurrence of item A among the items B, C and D which means the influence of item A on the others.

Results : By analyzing our collected data regarding risk factors of vein thrombosis by association rules mining, we got so many relations among those factors, from which we can extract the significant relations that could have an effect on the pathogenesis, severity, management and the course of the disease. Regarding the age 49% of patients were age<=40 year,23% had history of contraceptive pills,49% gave

history of immobility, 35% had surgery before DVT, 16% had febrile illness, 18 % were smokers and 19% had pregnancy[figure3].

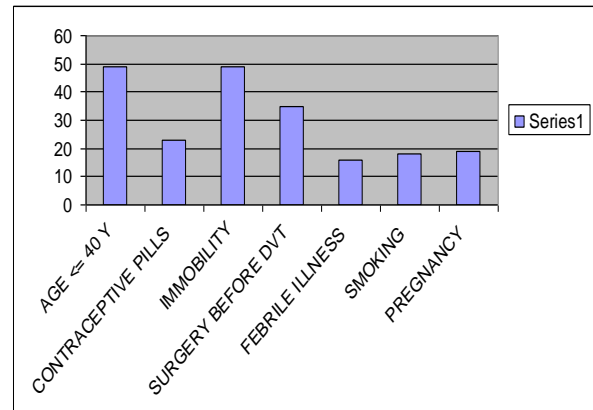


Figure (3) factors percentage from the given data

Association rules results: From those so many relations that have been extracted we chose the most significant relations (rules) with high confidence.

1. All patients (100%) with long bone fractures were immobile.
2. All alcoholic patients (100%) were immobile.
3. All smoker patients <=40 years, with surgery before DVT, were immobile.
4. 70% of patients with surgery before DVT were immobile, while 50% of immobile patients had surgery before DVT.
5. 60 % of smoker patients were <=40 years.
6. 75% of alcoholics had surgery before DVT, and had a long bone fracture.
7. 75% of patients<=40 years, with long bone fractures, were smokers.
8. All immobile pregnant women with age under 40 years had contraceptive bills.

Mean of confidence and the effects of factors:

1. Occurrence of age<=40 year with the other factors, [Figure4]. Contraceptive pills ,29%, immobility 46%, surgery before DVT 32%, febrile illness 14%, smoking 21%and pregnancy 39%. After discovering all frequent item sets, the algorithm of generating association rules uses the subsets of a frequent item set as antecedents to generate the rules (Figure 1)

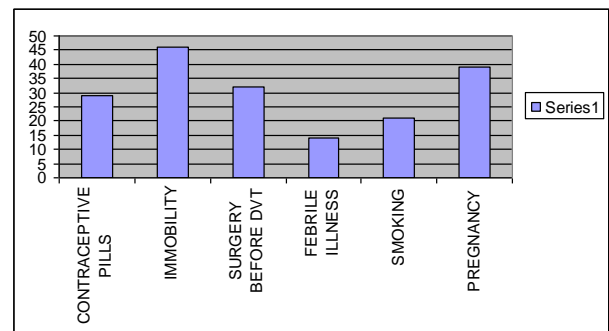


Figure (4) shows the Occurrence of age the with other factors

2. Occurrence of Contraceptive pills with the other factors, [Figure 5]. Age 62%, immobility 54%, surgery before DVT 46% , febrile illness 8%, and pregnancy 38%.

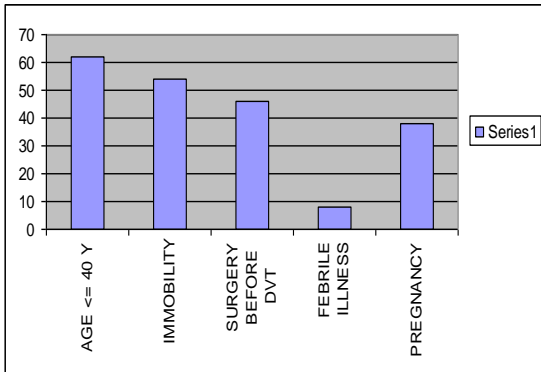


Figure (5) shows the Occurrence of contraceptive pills with the other factors

2-Occurrence of immobility with the other factors, [Figure 6]. Age 46%, Contraceptive pills 25%, surgery before DVT 50% , febrile illness 14%, smoking 25% , and pregnancy 11%.

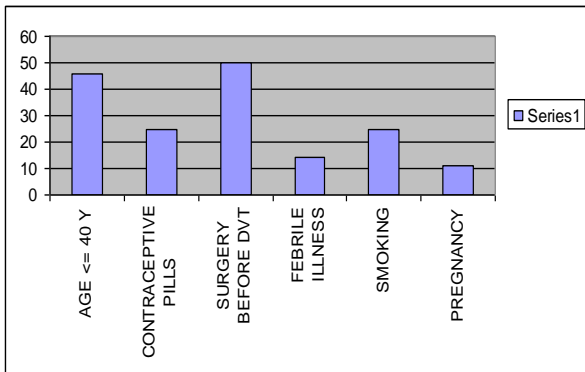


Figure (6) shows the Occurrence of immobility with the other factors

3-Occurrence of surgery before DVT with the other factors, [Figure 7]. Age 45%, Contraceptive pills 30%, immobility 70% , febrile illness 15%, smoking 20% and pregnancy 20%.

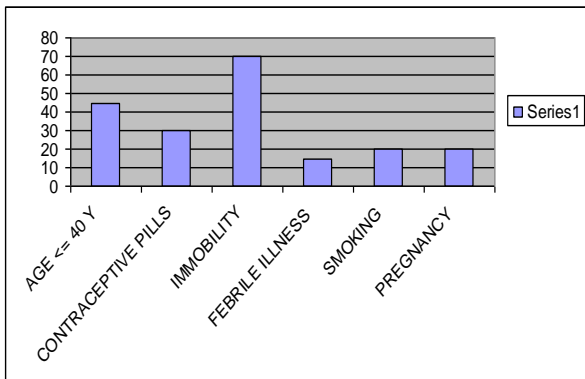


Figure (7) shows the Occurrence of surgery before DVT with the other factors

4-Occurrence of febrile illness with the other factors, [Figure 8].Age 44%, Contraceptive pills 11%, immobility 44%, surgery before DVT 33% , and smoking 22% , and pregnancy 33%.

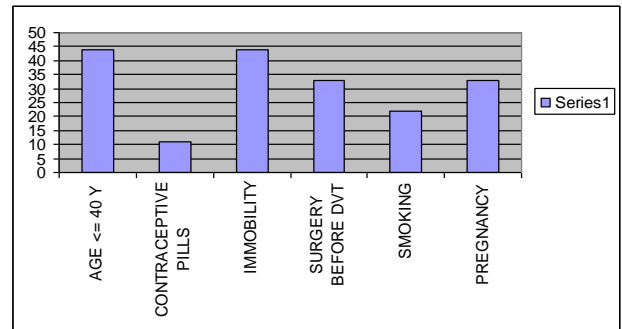


Figure (8) shows the Occurrence of febrile illness with the other factors

5-Occurrence of smoking with the other factors, [Figure 9].Age 60%, immobility 70%, surgery before DVT 40%and Febrile illness 20% .

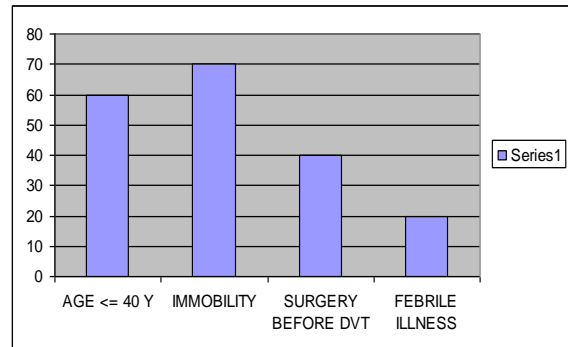


Figure (9) shows the Occurrence of smoking with the other factors.

6- Occurrence of pregnancy with the other factors, [Figure10].Age 100%, Contraceptive pills 45%, immobility 27%, surgery before DVT 36%, and febrile illness 27%, there is no relationship between smoking and pregnancy.

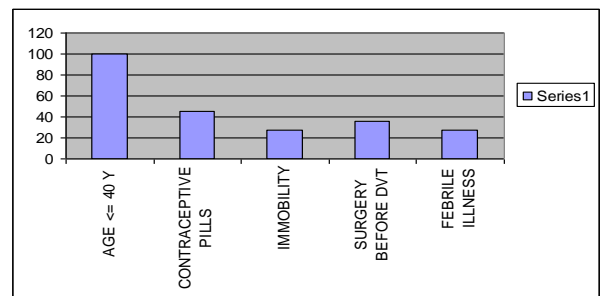


Figure (10) shows the Occurrence of pregnancy with the other factors

7- Mean of confidence (occurrence of the other factors with each single factor), [Figure 11]. Age 30%, Contraceptive pills 42%, immobility 29%, surgery before

DVT 33%, febrile illness 31%, smoking 48% and pregnancy 47%.

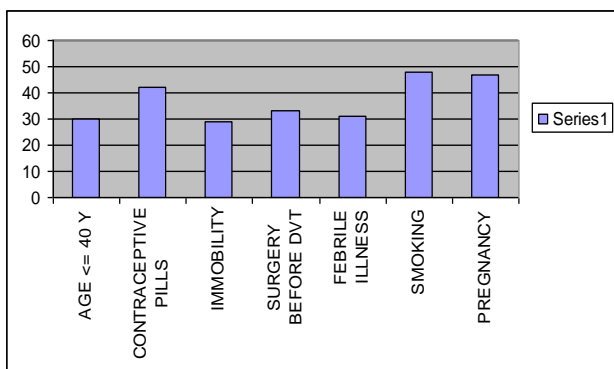


Figure (11) shows the Mean of confidence .

Discussion: Male to female ratio was nearly equal , so we didn't discuss the gender among other risk factors. Immobility , smoking , contraceptive pills , postoperative fever, where important factors in DVT occurrence. While age wasn't a significant factor.

Conclusions and Recommendations: 1. Immobility is a very important risk factor for DVT after surgery, so, early mobilization of patients after surgery will reduce the incidence of DVT. 2. Smoking , add more risk for DVT, when combined with immobility after surgery. So, cessation of smoking perioperatively will reduce the risk of DVT. 3. Occurrence of age with other factors gives no important relations because DVT could occur at any age. 4. Patients taking contraceptive pills, have higher risk of DVT post operatively So those patients should be considered as high risk patients, and given peri - operative anticoagulation. 5. About third of patients with surgery before DVT had history of fever. This means that prevention and control of post operative fever will reduce the incidence of DVT after surgery by about 33%. 6. Association rules mining allow better and faster analysis of more data with an interactive powerful system, which saves time and effort, and discovers the relations among many factors to one or more than one factors. So, we use this method for analysis in this study, and we get the above mentioned relations, which are important for the future management of DVT.

References :

1. RosamondW, Flegal K, Furie K, et al. Heart disease and stroke statistics–2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2008;117:e25-e146.
2. Allender S, Scarborough P, Peto V, et al. European cardiovascular disease statistics. *Eur Heart Netw* 2008:1-112.
3. Hirsch AT, Criqui MH, Treat-Jacobson D, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. *JAMA* 2001;286:1317-24
4. Schwartz's Principles of surgery f.Charles Brunicardi ,8th Edition, chap 22,2004.
5. Rutherford's Vascular surgery 7th edition vol1 Jack.I.Cronenwett and Johnson,K.wayne Johnston. Saunder Elsevier,Section3, chap 25, chap 26, chap 27, chap 28, chap 29, 2010.
6. Sabiston DC:Disorder of the arterial system.In Textbook of surgery edited by David C. Sabiston .14 ed ,vol.1 philadelphia W,.B.Saunders Company,1991 1:1618-1722.
7. Norgren L, Hiatt WR, Dormandy JA, Nehler MR, Harris KA, Fowkes FG. Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). *Eur J Vasc Endovasc Surg* 2007;33:S1-S75.
8. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", In Proc. of Int. Conf. on Very Large Data Bases (VLDB'94), Santiago, Chile, September 1994, pp. 487-499.
9. Johnny G, Developing A-priori Algorithm for Fast Mining Association Rules , Al-Taqani, Refereed Scientific Journal, Foundation of Technical Education, Vol. 22 , No.5, 2009, Baghdad-Iraq.
10. Johnny G, Interactive KDD system for fast mining association rules, Al-Taqani, Refereed Scientific Journal, Foundation of Technical Education, Vol. 22 , No.5, 2009, Baghdad-Iraq.