Decision Tree Algorithm and Learning Algorithms for Blood Data Diagnosis

Zaki Saeed Tawfik[1]
zeki.saeed@hiuc.edu.iq

Alaulddin Adel Albla[2]
alladdin@hiuc.edu.iq

Mahmoud Abbas Mahmoud[3]
mahmoudalnimi@hiuc.edu.iq

Abstract: A systolic pressure has a top (systolic) and bottom (pulse pressure) number (diastolic) Normal. People who already have hypertension that is greater than average should ask their doctor about how to lower it. Previously, the level was set at 140/90 mm Hg for people below the age of 65and 150/80 mm Hg for those 65 and older. The purpose of this research is to establish the reference range of red blood cells hydroxylated (Hb A1c percent) within the (males and females) and to forecast Bp blood pressure diastolic. This means 70% to 79% of age 65- and 150/80-mm Hg for those ages 65 and older Diastolic blood pressure systolic. In this paper sampling is taken from 100andbased on data mining model, which is substitute for the decision tree calculative algorithm which one of one of the fields Artificial intelligent which analysis of appropriate decision-making can be used to visually represent decisions and processes operations prospecting.

Keywords: Data processing, Decision tree algorithm, Learning algorithms, Classification

[1] Prof. Dr.: Computer Techniques Engineering Department, Al-Hikma University College, Baghdad, Iraq
[2] Dr.: Computer Techniques Engineering Department, Al-Hikma University College, Baghdad, Iraq
[3] Assist. Prof. Dr.: Medical Instrumentation Techniques Engineering Department, Al-Hikma University College, Baghdad, Iraq

1. Introduction

Public health care advancement is still a driving force in the quest to walk the fine line between cost containment and health-care reliability. When the force of the blood flowing through your blood vessels is consistently too high, user have high blood pressure (usually regarded as hypertension) [1].. The systolic bp (the top number or lowest blood pressure between heart beats) is more important than diastolic blood pressure (the bottom number or lowest blood pressure between both heart beats) because it best predicts the risk of systolic blood pressure rises but your diastolic blood pressure remains normal, you get a condition called as isolated systolic hypertension. Isolated systolic hypertension is the lateral blood pressure has been the most popular type of vertical heart rate in people over the age of 65. This type of rapid heartbeat can also occur in young people. High blood pressure (also known as hypertension) occurs when the force of the blood exceeds normal levels.

In this paper we are describing a prototype measuring device based on data mining model, which is substitute for the classical calculative algorithm. The goal of this study is to find out what the normal range of glycated (HbA1c basis points) is in an Iraqi population Blood samples were collected from 100 healthy subjects (50 females and 50 males) ranging in age from 20 to 75 years. The reference value for HbA1c percent in females was (5.34-0.67) percent and (5.67-0.73) percent in males [2].

2. Application

As a dataset, data were obtained from 100 healthy individuals (50 females and 50 females) ranging in age from 20 to 75 years old. Microsoft SQL Server 2000 was the management information system used in the study. This scheme was used for 2 purposes: the software program was consistent with and useful with the dbms. The information to be investigated was previously recorded. The data set was compiled in 2010 from 100 healthy subjects as well as records from 100 patients.

2.1 The data process

Several steps were involved in the data exploration and presentation process. Information gathering, data classification and conversion, data mining, and data display are all aspects of data management were the steps involved [3].

2.2 Preparedness of data and conversation

During these steps, data from multiple tables was merged into a standard table. The system's dataset of females (F) and males (M) is gathered from people, and

greatest effect on Bp blood pressure systole females' ages range from 21 to 67, while males' ages range from 25 to 57. See Table 1.

For men, weight in various distances of kilos, height in cm in length partnership among them and Hb A1C put it in range after samples from female and male were obtained. The other qualities have less of an impact. Because some data in the selection was out of range in some attributes, the activity errors in the records were fixed and the data was made arrangements and placed in table 1.

There is a high correlation between the data collected from blood samples taken from 100 healthy subjects, both males and females. Sex, Age Yrs, WE. kgm, Height cm., Hb A1c, and Bp systolic diastolic are even included in the dataset see below.

Table 1: Data table training

Sex	No of Patients	Age Yrs	WE. kgm	Higfht Cm.	Hb A1c	Bp systolic diastolic (class)
F	6	21- 32	55-66	170-162	4.0-4.8%	110/70
F	20	33-36	50-70	152-171	4.8-5.1%	120/80
F	16	37-40	64-74	162-175	5.1-6.1%	130/80-130/90
F	8	38-67	70-85	152-171	5.8-6.6%	130/80-130/90
M	20	25-34	64-74	170-183	4.0-5.4%	130/80-130/90
M	14	35-37	70-109	170-180	4.6-6.1%	130/90
M	10	36-38	60-85	167-181	5.0-6.9%	140/90
M	6	39-57	90-101	174-190	5.6-5.9%	150/90

3. Decision Tree Algorithm

Figure 1 illustrates the decision tree algorithm. Member of the group of learning algorithms is the prediction model. Unlike other supervised learning methods, this method can solve regression and classification problems. Using a Classification Tree is to develop a training model able to accurately predict the class or value of an attribute value based on, earlier data by learning simple decision rules. To predict a class label for a record in Decision Trees, start at the trunk of the tree. The values of the root are calculated and the results of the record's attribute. Proceed to another node analysis and the comparison by following the branch that corresponds to that value3.

In machine learning, categorization process consists of a having to learn step and a step of prognostication. The fashion designer is developed in the learning stage based on the given data. The model is employed. Figure 1 shows decision tree algorithm. To response for information items in the prediction process, decision Tree is one of the most simple and widely used classification techniques [6].

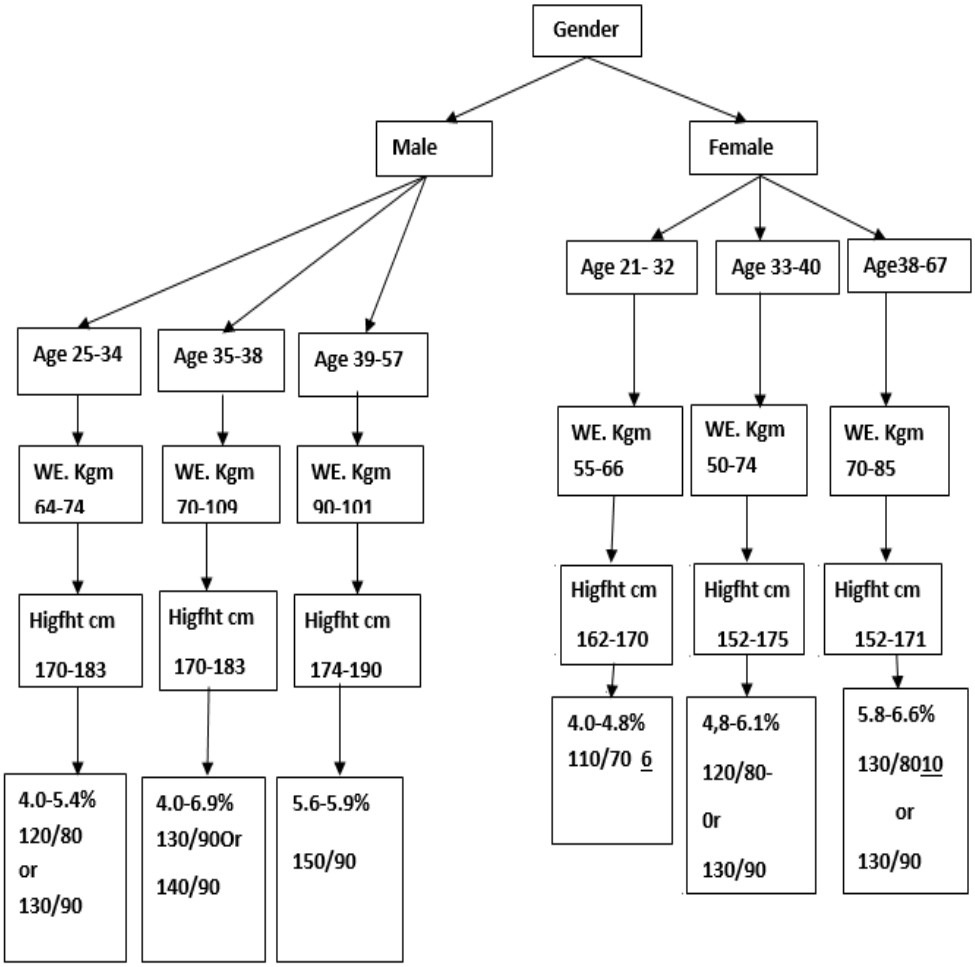

Figure 1: Decision tree algorithm

Rules

RULE1 IF Male and Age ≥25and ≤34and we.kgm ≥ 64 and≤74 and Higfht≥170 and≤183and Hb A1c=4.0%THEN Bp systolic diastolic = 120/80 ELLSE Bp systolic diastolic = 130/180.

RULE2 IF Male and Age ≥35and ≤38and we.kgm ≥70 and≤109and Higfht≥170 and≤183and Hb A1c=4.0% THEN Bp systolic diastolic = 130/90ELLSE Bp systolic diastolic = 140/90

RULE3 IF Male and Age ≥39and ≤57and we.kgm ≥90 and≤101and Higfht≥174 and≤190and Hb A1c=5.6.3% THEN Bp systolic diastolic = 150/90

RULE5 IF Female and Age ≥21and ≤32and we.kgm ≥55 and≤66 and higfht≥162 and≤170 and Hb A1c=4.0% THEN Bp systolic diastolic = 110/70 .

RULE6 IF Female and Age ≥33and ≤40and we.kgm ≥50 and≤74 and higfht≥152 and≤171and Hb A1c=4,81% THEN Bp systolic diastolic = 120/80 ELLSE Bp systolic diastolic = 130/90.

RULE7 IF Female and Age ≥38and ≤67and we.kgm ≥7o and≤85 and higfht≥153 and≤165and Hb A1c=5.8% THEN Bp systolic diastolic = 130/80 ELLSE Bp systolic diastolic = 130/90.

4. Classification

Classification techniques necessitate class definitions regarding data attribute values. They frequently describe these classes by examining the properties of data that is already known to correspond to the classes. Classification is a big data (pattern recognition) method used to predict data instance group membership, and that has risen to be one of the most crucial components of the information technology revolution affecting our lives. Artificial intelligence, the field from which machine learning arose, aims to make think the same way that humans do technique of artificial intelligence subfield that "provides computers without being explicitly programmed. For example, could [4] uses decision tree to predict and Classification techniques of paints include and neural networks paints are shown in Table 2.

Table 2: Classification of data attributes values

Male	Number	Hb A1c and Bp	Female	Number	Hb A1c and Bp
Age 25-34	20	4.0-5.4%120/80or 130/90	Age 21-32	6	4.0-4.8%110/70
Age 35-38	24	4.0-6.9%130/90 or 140/90	Age 33-40	36	4,8-6.1% 120/80 or 130/80
Age 39-57	6	5.6-5.9% 150/90	Age38-67	8	5.8-6.6% 130/80 or 130/90

5. Conclusions

The data gathering procedure was conducted out and explained clearly. For patients, the implementation expected the impact of age, year, and weight in Length in kg in cm in length, and the connection on Bp pulse rate systole and Hb A1C. Data mining techniques used in cluster analysis may produce more varied and significant results. This step presents the results of the data mining step. The impact of Hb A1c on Diastolic blood pressure for women is less than specified. Great connection among attributes values was found.

References

[1] Salwa H.N., Emad Abdul-Rehman (2010) "Determining the reference Rang Values of Glycosylated Hemoglobin (HbA1c) by Immunoiturbid Assayin Iraqi

[2] Chen L, Magliano DJ, Zimmet PZ (2011) The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol 8: 228-236

[3] Milley, A. (2000). Healthcare and data mining. Health Management Technology, 21(8), 44-47.

[4] Dr. Zeki S. Tywofik and Ali T. YASEEN "Application of Clustering as a Data Mining Tool in Bp systolic diastolic "Jan 1, 2016.

[5] Radhwan H. A. Alsagheer1 , Abbas F. H. Alharan2 , Ali S. A. Al-Haboobi3 "Popular Decision Tree Algorithms of Data Mining Techniques" June 2017, pg.133 – 142

[6] Yuqing, P., Xiangdan, H., Shang, L., "The KMeans Clustering Algorithm Based On Density and Ant Colony", IEEE Int. Conf. Neural Networks & Signal Processing Nanjing, China, 457-460, December 14-17, 2003.

[7] Thearling, K., "An Introduction to Data Mining",http://thearling.com/text/dmwhite/dm white.htm, 01 December 2003.

[8] Jelinek HF, Yatsko A, Stranieri A, Venkatraman S (2014) Novel Data Mining Techniques for Incomplete Clinical Data in Diabetes Management. British Journal of Applied Science & Technology 4: 4591-4606[7] Fayyad, U.M., Piatesky-Shapiro, G., Smyth, P., Uthurusamy, R., "Advances in data mining and knowledge discovery", MIT Pres, USA, 1994.

[9] Stamouli M, Pouliakis A, Mourtzikou A, Skliris A, Panagiotou I, et al. (2014) Evaluation of the Lipid Profile in Type 2 Diabetes Mellitus Patients in Greece. Clin Lab 60: 1593-1600.

خوارزمية شجرة القرار وخوارزميات التعلم لبيانات الدم

زكي سعيد توفيق[1]
zeki.saeed@hiuc.edu.iq

علاء الدين عادل المولى[2]
alladddin@hiuc.edu.iq

محمود عباس محمود[3]
mahmoudalnimi@hiuc.edu.iq

المستخلص: الضغط الانقباضي له رقم علوي (انقباضي) وأسفل (ضغط نبضي) (ضغط انبساطي) طبيعي . الناس يجب على الأشخاص الذين يعانون بالفعل من ارتفاع ضغط الدم أكثر من المتوسط أن يسألوا طبيبهم عن كيفية خفضه هو - هي .سابقًا ، تم ضبط المستوى على 90/140 ملم زئبق للأشخاص الذين تقل أعمارهم عن 65 عامًا و 80/150 ملم زئبق لمن هم بعمر 65 وما فوق .الغرض من هذا البحث هو تحديد النطاق المرجعي لخلايا الدم الحمراء هيدروكسيل (نسبة الهيموغلوبين A1c) داخل (الذكور والإناث) وللتنبؤ بضغط الدم Bp الانبساطي .هذا يعني 70٪ إلى 79٪ من العمر 65 و 80/150 ملم زئبق لمن هم في سن 65 وما فوق .الانبساطي ضغط الدم الانقباضي .في هذه الورقة ، تم أخذ العينات من 100 واستنادًا إلى نموذج التنقيب عن البيانات ، والذي يعد بديلاً عن الخوارزمية الحسابية لشجرة القرار والتي هي أحد الحقول الاصطناعية ذكي أي تحليل لصنع القرار المناسب يمكن استخدامه لتمثيل القرارات بصريًا وعمليات التنقيب عن العمليات.

الكلمات المفتاحية: معالجة البيانات، خوارزمية شجرة القرار، خوارزميات التعلم ، تصنيف

[1] استاذ دكتور: قسم هندسة تقنيات الحاسوب - كلية الحكمة الجامعة – بغداد – العراق
[2] مدرس دكتور: قسم هندسة تقنيات الحاسوب - كلية الحكمة الجامعة – بغداد - العراق
[3] استاذ مساعد دكتور: قسم تقنيات الأجهزة الطبية - كلية الحكمة الجامعة – بغداد - العراق