Original Research Report

DISCOVERING PATTERNS OF CARDIOVASCULAR DISEASE AND DIABETES IN MYOCARDIAL INFARCTION PATIENTS USING ASSOCIATION RULE MINING

Anju Singh[1], Divakar Singh[2], Shikha Sharma[3], Kamal Upreti[4], Manish Maheshwari[5], Vimal Mehta[6], Jitender Sharma[7], Pratishtha Mehra[6], Pradeep Kumar Dabla[7]

[1]Sage University, Bhopal, India
[2]Barkatullah University Institute of Technology, Barkatullah University, Bhopal, India
[3]Department of Biochemistry, Chacha Nehru Bal Chikitsalaya, New Delhi, India
[4]Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India
[5]Makhanlal Chaturvedi National University of Journalism and Communication, Bhopal, India
[6]Department of Cardiology, Govind Ballabh Pant Institute of Postgraduate Medical Education & Research/ Associated Maulana Azad Medical College, New Delhi, India
[7]Department of Biochemistry, Govind Ballabh Pant Institute of Postgraduate Medical Education & Research/ Associated Maulana Azad Medical College, New Delhi, India

ABSTRACT

Cardiovascular diseases (CVDs) are a major cause of mortality in diabetic patients. Hypertensive patients are more likely to develop diabetes and hypertension contributes to the high prevalence of CVDs, in addition to dyslipidemia and smoking. This study was to find the different patterns and overall rules among CVD patients, including rules broken down by age, sex, cholesterol and triglyceride levels, smoking habits, myocardial infarction (MI) type on ECG, diabetes, and hypertension. The cross-sectional study was performed on 240 subjects (135 patients of ST-elevation MI below 45 years and 105 age-matched controls). Association rule mining was used to detect new patterns for early-onset myocardial infarction. A hotspot algorithm was used to extract frequent patterns and various promising rules within real medical data. The experiment was carried out using "Weka", a tool for extracting rules to find out the association between different stored real parameters. In this study, we found out various rules of hypertension like "Rule 6" says that if levels of BP Systolic > 131 mmHg, LpA2 > 43.2 ng/ml, hsCRP > 3.71 mg/L, initial creatinine > 0.5 mg/dl, and initial Hb ≤15 g/dl (antecedent), then the patient will have 88% chance of developing hypertension (consequent). Similarly for diabetes mellitus with finding their lift and confidence for different support like "Rule 6", if MI type on ECG = 'Inferior Wall MI' with STATIN=No, and levels of Triglycerides ≤325 (antecedent), then the patient had a 67% chance of developing diabetes mellitus. We concluded that early-onset myocardial infarction is significantly associated with hypertension and diabetes mellitus. Using association rule mining, we can predict the development of hypertension and diabetes mellitus in MI patients.

Keywords: *Cardiovascular diseases; diabetes; hypertension; association rule mining; hotspot; myocardial infarction*

Correspondence: Pradeep Kumar Dabla, Department of Biochemistry, Govin Ballabh Pant Institute of Postgraduate Medical Education & Research/ Associated Maulana Azad Medical College, New Delhi, India. Email: pradeep_dabla@yahoo.com

How to cite: Singh, A., Singh, D., Sharma, S. ., Upreti, K., Maheshwari, M., Mehta, V., Sharma, J., Mehra, P. ., & Dabla, P. K. (2022). Discovering Patterns of Cardiovascular Disease and Diabetes in Myocardial Infarction Patients Using Association Rule Mining. Folia Medica Indonesiana, 58(3), 242–250. https://doi.org/10.20473/fmi.v58i3.34975

pISSN:2355-8393 ● eISSN: 2599-056x ● doi: 10.20473/fmi.v58i3.34975● Fol Med Indones. 2022;58:242-250
● Submitted 15 May 2022 ● Received 28 Jul 2022 ● Accepted 23 Aug 2022 ● Published 5 Sep 2022
● Open access under CC-BY-NC-SA license ● Available at https://e-journal.unair.ac.id/FMI/

Highlights:

1. Association Rule Mining tools predict the association of early-onset Myocardial Infarction with Hypertension and Diabetes Mellitus.
2. Association Rule Mining tools using clinical and biochemical attributes can predict the development of Hypertension and Diabetes Mellitus in Myocardial Infarction patients.

INTRODUCTION

Cardiovascular diseases (CVDs) have become the leading cause of mortality globally since the turn of the century (Srinath et al. 2005). CVD occurred a decade earlier in the Indian population as compared to the European population, wherein CVD caused approximately 23% mortality in the elderly (<70 years of age). The fatality estimate in India is estimated to be double in comparison with the same age group (Joshi et al. 2007). The various factors responsible for the higher rate of CVD in the Indian population are hypertension, diabetes, metabolic syndrome, smoking, physical inactivity, unhealthy diet, and other environmental events which lead to the formation of the atheromatous plaque. Acute myocardial infarction (MI) is one type of CVD that alone is responsible for the nearly 1.5 million MI incidents yearly where nearly 33% of them lead to death (McNamara et al. 2019, Virani et al. 2021).

CVD is on the rise in practically all age groups presently, but numerous studies have recently revealed an increase in instances in the young population (less than 40 years of age) and increased mortality with acute coronary syndrome (Mohan 2005). The National Inpatient Sample data revealed that young MI patients had a higher predominance of risk factors, ischemic stroke, and a higher rate of hospital admission (George et al. 2011, Gupta et al. 2014). There is a scarcity of information about juvenile MI patients. Young MI patients are characterized as those who are under 45 years of age, according to the Framingham Heart Study (Doughty et al. 2002).

Traditionally, physicians diagnose CVDs based on their knowledge from their previous experience with patients with similar clinical presentations. However, healthcare workers and hospitals generate a huge amount of data that are difficult to be analyzed by the traditional approach. Association rule mining supports physicians in attaining more accurate diagnosis and early treatment to reduce mortality in CVD patients.

Among the various methods available in data mining, association rule mining extracts useful associations or causal relationships between sets of patterns present in the databases (Han et al. 2011, Hastie et al. 2009). The rules show the dependency of one item on another item and map it to discover the interesting relations between variables in the database. Exploratory data analysis in the medical domain is an emerging field and association rules mining plays a key role in finding the causes and solutions to various diseases. With the rapid growth of medical data sets, there is a need to explore the hidden patterns in the clinical data extracted patterns provide vital knowledge to medical professionals for making appropriate strategies and enhancing the performance of patient management tasks.

To find out the infrequent (rare) items, we have chosen low support and high confidence thresholds. Support is an indication of how frequently the items appear in the data whereas confidence indicates the number of times the if-then statements are found true. In the domain of business, a small number of rules may or may not be beneficial for customer analytics. Although we agree that confining low support and strong confidence yields a small number of rules where clinicians may infer some interesting patterns where they can explain even lesser-known incidents from such a small number of rules (Szathmary et al. 2010). Most of the combination of the symptoms will only be seen in a few patient instances, which is typically true in medical diagnosis. As a result, a method for mining the related rules and patterns will enable a most concentrated symptom investigation. In another study, the role of the Atherogenic Index of Plasma (AIP) was in conjunction with lipid indices like Lipid Tetrad Index (LTI) and Lipid Pentad Index (LPI) in identifying people who were at a higher risk of premature CA (Dabla et al. 2021).

Patil et al. (2010) aimed to find meaningful facts from the medical database and produce logical and user-friendly descriptions of patterns. They presented a model for association rules over numerical data and used an equal width mining interval method to desensitize continuous-valued elements, using an appropriate algorithm for extracting the fields to know whether the patient is going to develop diabetes mellitus or not. In another study, survival association rule mining is proposed, which addresses the various shortcomings in extracting rules (Simon et al. 2013). In the study, real data sets were used to extract the association rule. In another paper, association rule algorithm is applied to analyze the risk for diabetic patients. This study was performed to extract search relationships in the real data set (Soni et al. 2016). Shehabi and Baba (2021) suggested a new method, called MARC, to extract the more important association rules of two important levels: Type I and Type II. The approach relies on a multi-topographic unsupervised neural network model as well as clustering quality measures that evaluate the success of a given numerical classification model to behave as a natural symbolic model. Another software tool was devised for extracting frequent patterns and association rules from invoices provided by healthcare centers to improvise the quality of a large number of medical e-invoices reducing the irrelevant data (Agapito et al. 2019). Earlier mining rules were used to establish an association between collections of items in enormous databases (Vasoya & Koli 2016). Association rule mining was used recently to discover symptom patterns in COVID-19 patients rule mining (Tandan et al. 2021). As a result, in this paper, we focused on simple pattern mining techniques known as ARM (Association Rule Mining) to give a descriptive strategy for symptom rule extraction. We intended to uncover hidden correlations between symptom patterns in diabetic and hypertensive patients, which could help clinical decision-making for patient treatment.

A hotspot algorithm was applied to find patterns or associations between different attributes such as a complete set of biochemical evaluation testing along with detailed patient history including physical examination and electrocardiogram (ECG). Biochemical markers measured were lipid profile including total cholesterol, triglyceride, LDL-C, HDL-C, Apo A1, Apo B, and Lp(a) levels.

MATERIALS AND METHODS

This cross-sectional study was performed on the clinical data of cardiovascular disease patients presented at G.B. Pant Institute of Postgraduate Medical Education and Research in New Delhi, India. This research had been approved by the ethics committee of the institute and written approved consent was obtained from participants. A total of 240 subjects were included in the study with proper informed consent. Out of the total, 135 patients were of ST-elevation MI who were below 45 years of age whereas age-matched controls enrolled were 105 in number. A complete set of biochemical evaluations was conducted along with detailed patient history, physical examination, and electrocardiogram (ECG). Biochemical markers measured were lipid profile including total cholesterol, triglyceride, LDL-C, HDL-C, Apo A1, Apo B, and Lp(a) levels.

The Acute MI (AMI) diagnosis was based on the established criteria as:1). Ischemic chest pain lasting 20 min or more, 2) ECG evidence of myocardial injury: (a) ≥0.1 mv ST elevation in two contiguous leads other than V2–V3 where the cut-off point of ≥0.2 mv in men ≥40 years, ≥0.25 mv in men <40 years, or ≥0.15 mv in women. (b) New horizontal or down sloping ST depression ≥0.05 mv in two contiguous leads and/or T-wave inversion ≥0.1 mv in two contiguous leads with prominent R-wave or R/S ratio >1, 3). Positive biomarkers were creatine kinase-MB fraction and cardiac troponins. Then, further analysis was performed on clinical data obtained to detect common patterns in the rules discovered utilizing association rule mining.

The algorithm of the proposed framework works in four major steps as given below, having three stages including pre-processing of the dataset to make a compatible dataset for the algorithm, frequent pattern generation by Hotspot algorithm, and generation of rules. Hotspot learns a bunch of rules (showed in a tree-like design) that augment/limit an objective variable/ worth of interest. With an ostensible objective, one should search for sections of the information where there is a high likelihood of minority esteem occurring (given the imperative of a base help)**.**

To mine association rules, the first step is to create a dataset compatible with the Hotspot algorithm. Initially, data are stored in a Microsoft Excel sheet, we had to change it into comma-separated values and then convert it into ARFF (attribute relation file format). In the pre-processing of mine association rules, we discarded a few irrelevant attributes from the dataset. We had applied the prepossessing technique to make data suitable for the hotspot algorithm. After the pre-processing, data were found to be compatible with the hotspot algorithm. The first step was pre-processing of the medical dataset. The step was the removal of irrelevant attributes continued with the removal of records with missing class values. The second step was applying Hotspot algorithms for targeted class values. The third step was to scan and find the support value for each item, then the removal of infrequent items based on support values. The last step was the generation of rules.

Association rule helps to find a meaningful pattern that can be obtained in those states when R occurs, S occurs with certain possibilities. Literature for Rules Measures had been taken from various available relevant research papers on this subject (An et al. 2005, Chen & Chen 2008, Lau et al. 2006, Ordonez 2006).

For finding out interesting rules, we can apply various measures to set constraints on extracted rules. The minimum constraints are support and confidence.

Support: It denotes the % of the rule within transactions containing R ∪ S.

Support (R≥S) = P (R ∪ S).

Confidences: for the rule hold of form R≥S, confidence score lies between 0 to 1.

Confidence (R≥ S) = P(S | R)

= Support (R ∪ S) / Support (R)

= P (R and S) / P (R)

Lift: for the rule of form R≥S, if the lift score is 1 then the antecedent R and Consequent S are independent. If a score of lift is greater than 1 means that R and S are positively correlated. If a score is less than 1 means R and S are negatively correlated.

Lift (R≥S) = Support (R ∪ S) / Support (R) x Support (S)

= P (R and S) / P (R) P (S)

We discovered many rules that apply to the data set with nominal and real values, which only included clinically attributed data and omitted irrelevant factors. The correlation between R and S is determined by the lift value; independent (=1), positively related (>1), and negative related (1). The "Confidence" measures have the disadvantage of possibly misrepresenting the importance of an association.

In an association of R & S, for example, the Confidence score only considers the importance of item R, but not S. If S is also required, then there is a greater chance that a pattern having R will also have S, increasing the confidence measure. By measuring the strength of the relationship between R and S, the metric lift overcomes this difficulty. The data in transaction format, as well as how ARM computed and built these measures and criteria, are depicted in Figure 2.

RESULTS

The clinic-epidemiological profile of the study population is presented in Table 1

Table 1. Demographic variable of the cases under study

Variable		N (%)
Total No. of cases		135
Sex	Male	126
	Female	09
Age (Years), Mean ± SD		36.0 ± 4.5
Number of participants (15-30 Years)		21 (15.6%)
Number of participants (31-45 Years)		114 (84.4%)
BMI (kg/m^2)		25.7 ± 4.31
BMI >30 (kg/m^2)		21 (15.5%)
Systolic blood pressure (mmHg)		118.4 ± 15.5
Diastolic blood pressure (mmHg)		75.6 ± 9.61
Heart rate (beats/min)		88.2 ± 11.3
Sedentary lifestyle		126 (93.3%)
History of drug intake of statin		18 (13.3%)
Diabetes mellitus		12 (8.89%)
Hypertension		17 (12.6%)
Dyslipidemia		12 (8.89%)
Personal history	Addiction (Smoking cigarettes and beedi)	95 (70.4%)
	Tobacco consumption	90 (66.7%)
	Alcoholic consumption	41 (30.4%)
Frequency of chest pain		130 (96.2%)
Sweating		110 (81.4%)
Breathlessness		85 (62.3%)
Syncope		14 (10.4%)
ECG Changes	Anterior wall MI	82 (60.7%)
	Inferior wall MI	48 (35.5%)
	Lateral wall MI	5 (3.7%)

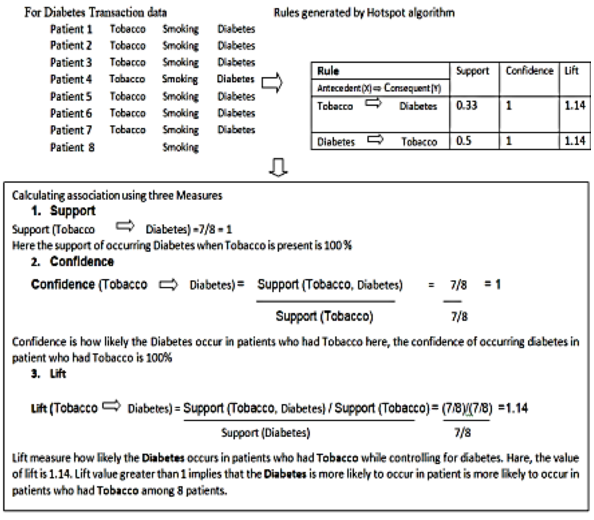

Figure 1. Example to show how ARM computed for diabetes transactional data

In Figure 1, data from eight patients were taken, and the rule mining method yielded two rules, with a support score of 0.33, confidence of 1, and lift of 1.14 for the antecedent (R) Tobacco and consequent (S) Diabetes in rule 1. According to Support 0.5, four out of eight patient data consist of the common attribute "Tobacco and Diabetes". With a confidence level of 0.80, 80 percent of Tobacco patients had Diabetes.

Similarly, lift 1.14 indicates that "Tobacco" and "Diabetes" are positively correlated. In Table 2 and Figure 2, significant rules for {Diabetes Mellitus = No} (N = 135 with 34 attributes in each) are given. Rule 2 says that if value of initial Hb = 7.2, Triglycerides >87, BP Diastolic ≤95} (antecedent), then the patient had a 100 % chance of {Diabetes Mellitus=No} (consequent). In Rule 3, if value of initial HbA1c ≤7.2, Hypertension = No and Heart Rate >70 (antecedent), then the patient had a 100% chance of “Diabetes Mellitus = No” (consequent).

Frequent Pattern for support value was 0.2 to 0.8 (min 20% to 80% of Diabetes Mellitus = No) and confidence (Chance of occurrence of both Antecedents and Consequents) for value was 0.96 to 1' with following attributes and their combinations: initial Hb value >13.7, Heart Rate >70, Triglycerides >87, BP Diastolic ≤95, Hypertension=No, Triglycerides ≤239, initial HbA1c ≤8.6, Heart Rate >70. {Diabetes Mellitus= Yes}.

Rule 6 says that if MI type on ECG=’Inferior Wall MI’, with STATIN=No, and value of Triglycerides ≤325 (antecedent), then the patient had a 67% chance of {Diabetes Mellitus}. Frequent Pattern For support value 0.3 to 0.6 (min 30-60% of Diabetes Mellitus) and confidence (Chance of occurrence of both Antecedents and Consequents) for valve 0.6 to 1 (60% to 100%) following and their combinations: initial Hb >4.6 to ≤13.7, Alcohol = Never, ApoB >0.61 to ≤0.84, Total cholesterol >114 to ≤192, BP Diastolic >62 to ≤142, ApoA1 >0.73 to ≤1.23, BMI >26.44 to ≤27.8159, LDL >47, Heart Rate >93 to ≤109, HDL >22.7 to ≤42.2, hsCRP >5.95 to ≤105, Age >18, STATIN=No, MI type on ECG=Inferior Wall MI.

In Table 3 and Figure 3, significant rules for {Hypertension = No} (N = 135 with 34 attributes in each) are given. Rule 1 says that if the value of MI type on ECG = ” Inferior Wall MI” and value of BP Systolic ≤118 (antecedent), then the patient had a 100 % chance of {Hypertension = No} (consequent). In the next Rule 4, if value of initial Hb value > 11.5, and value initial HbA1c ≤12.3,Diabetes Mellitus = No,BMI ≤32.8125, hsCRP ≤182.01 (antecedent), then the patient had a 100% chance of “Hypertension = No” (consequent). Frequent patterns for support value 0.2 to 0.9 (min 20% to 90% of Hypertension) and confidence (Chance of occurrence of both Antecedents and Consequents) for valve 0.93 to 1 were observed.

Table 2. Significant rules for diabetes mellitus in 135 patients with 34 clinical attributes

Rules	Antecedents (R)	Consequents (S) {diabetes mellitus}	Support	Confidence	Lift
R1	{initial_Hb_ value > 13.7, Heart_Rate >70}	{No}	0.2	1	1.16
R2	{initial_HbA1c ≤7.2, Triglycerides >87, BP_Diastolic ≤95}	{No}	0.6	1	1.16
R3	{initial_HbA1c ≤7.2, Hypertension=No, Heart_Rate >70}	{No}	0.7	1	1.16
R4	{Triglycerides ≤239, Heart_Rate >70, initial_HbA1c ≤8.6}	{No}	0.80	0.98	1.13
R5	{initial_HbA1c ≤8.6, Heart_Rate >70}	{No}	0.80	0.96	1.12
DM=Yes⇓ R6	{Triglycerides >239, MI type_on_ECG=Inferior Wall MI, STATIN=No, Triglycerides ≤325}	{Yes}	0.3	0.67	7.83
R7	{Heart_Rate >96, HDL ≤37.4, initial_Hb_value ≤13.7, Age >32, BMI ≤27.6398}	{Yes}	0.5	1	11.75
R8	{initial_Hb_value ≤12.9, Alcohol=Never, ApoB ≤0.84, Total_cholesterol >114, BP_Systolic >107}	{Yes}	0.6	0.78	9.14
R9	{Heart_Rate >93, HDL ≤37.4, ApoA1 ≤1.23, BMI ≤27.8159, initial_Hb_value ≤13.7}	{Yes}	0.6	0.7	8.22
R10	{Alcohol=Never, LDL >49, hsCRP ≤105, initial_Hb_value ≤15.2, ApoA1 >0.73}	{Yes}	0.9	0.23	2.69

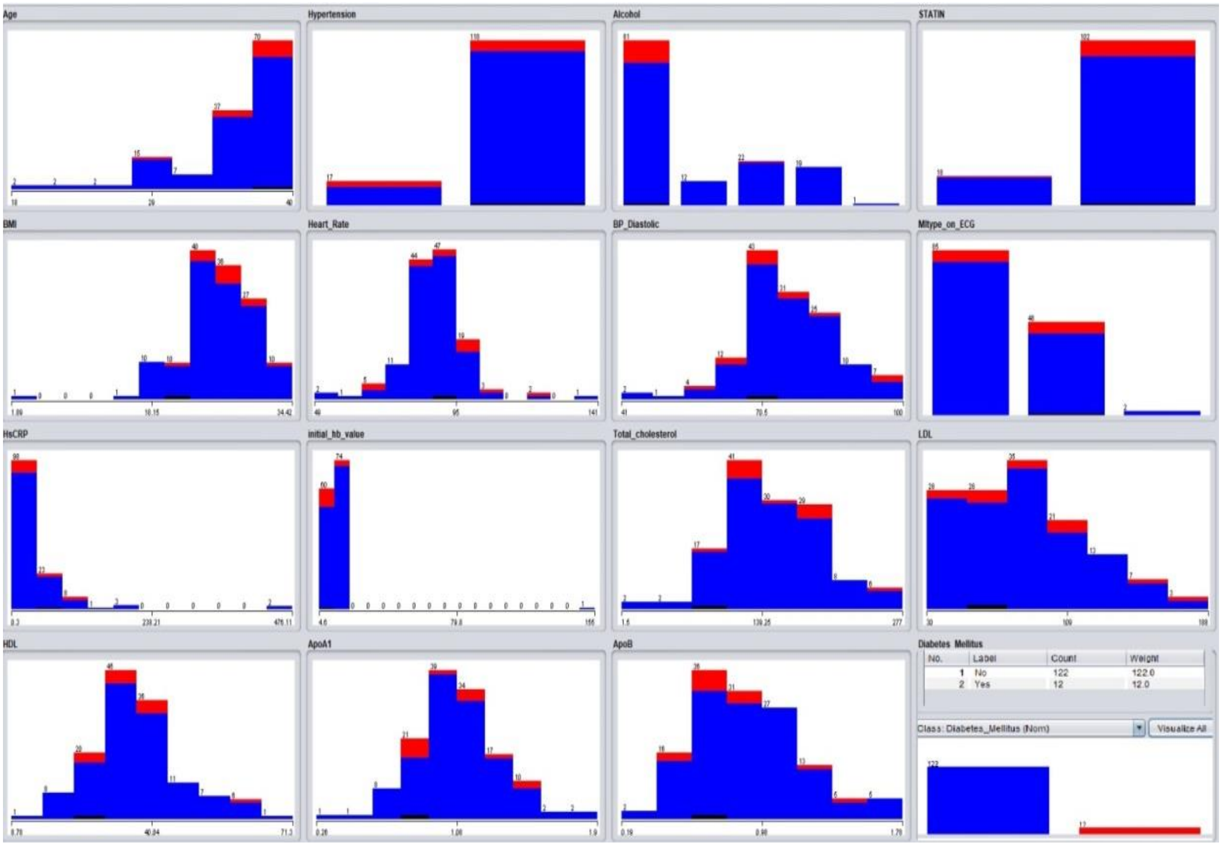

Figure 2. Association of different clinical parameters with diabetes mellitus

In Table 3 and Figure 3, significant rules for {Hypertension=Yes} Rule 6 says that if value of BP Systolic > 131, value of LpA2 >43.2, value of hsCRP >3.71, value of initial creatinine >0.5, and value of initial Hb ≤15 (antecedent), then the patient had an 88% chance of Hypertension (consequent). Frequent Pattern for support value 0.4 to 0.99 (min 40 % to 99% of Hypertension) and confidence (Chance of occurrence of both Antecedents and Consequents) for valve 0.19 to 0.88 following attributes and their combinations: BP Systolic >102, LpA2 >15.7, hsCRP >1.79, initial creatinine > 0.5, initial Hb ≤15.7, ApoB ≤1.25, ApoA1 ≤1.33, HDL >8.78, LDL ≤115 Total cholesterol >1.5 to ≤229, Physical Activity=No, Heart Rate >77 to ≤103.

Table 3. Significant rules for Hypertension in 135 patients with 34 clinical attributes

Rules	Antecedents (R)	Consequents (S) {hypertension}	Support	Confidence	Lift
R1	{MItype_on_ECG=Inferior Wall MI, BP_Systolic ≤118}	{No}	0.2	1	1.19
R2	{LpA2 ≤39.9, HDL >19.1, Triglycerides >52}	{No}	0.4	1	1.19
R3	{initial_Hb_value >12.6, ApoA1 ≤1.42, initial_creatinine_val ≤1.6, BP_Diastolic ≤92}	{No}	0.6	0.96	1.15
R4	{initial_Hb_value > 11.5, initial_HbA1c ≤12.3, Diabetes_Mellitus=No, BMI ≤32.8125, HsCRP ≤182.01}	{No}	0.8	0.96	1.15
R5	{BP_Systolic ≤144, Followup_status=Alive, LDL ≤148}	{No}	0.9	0.93	1.11
Hypertension = Yes ⇓ R6	{BP_Systolic >131, LpA2 >43.2, hsCRP > 3.71, initial_creatinine_val >0.5, initial_Hb_value ≤15}	{Yes}	0.4	0.88	7.26
R7	{BP_Systolic >122, LpA2 >39.9, hsCRP >2.78, Physical_Activity=No, Total_cholesterol ≤229}	{Yes}	0.6	0.71	5.92
R8	{LpA2 >39.9, BP_Systolic >115, Physical_Activity=No, Total_cholesterol >84, Heart_Rate >77}	{Yes}	0.8	0.4	3.32
R9	{hsCRP >1.79, BP_Systolic >102, LpA2 >15.7, Physical_Activity=No, initial_Hb_value ≤15.7}	{Yes}	0.99	0.2	1.64
R10	{LpA2 >15.7, BP_Systolic >102, Physical_Activity=No, initial_Hb_value ≤15.7, Heart_Rate ≤103}	[Yes}	0.99	0.19	1.57

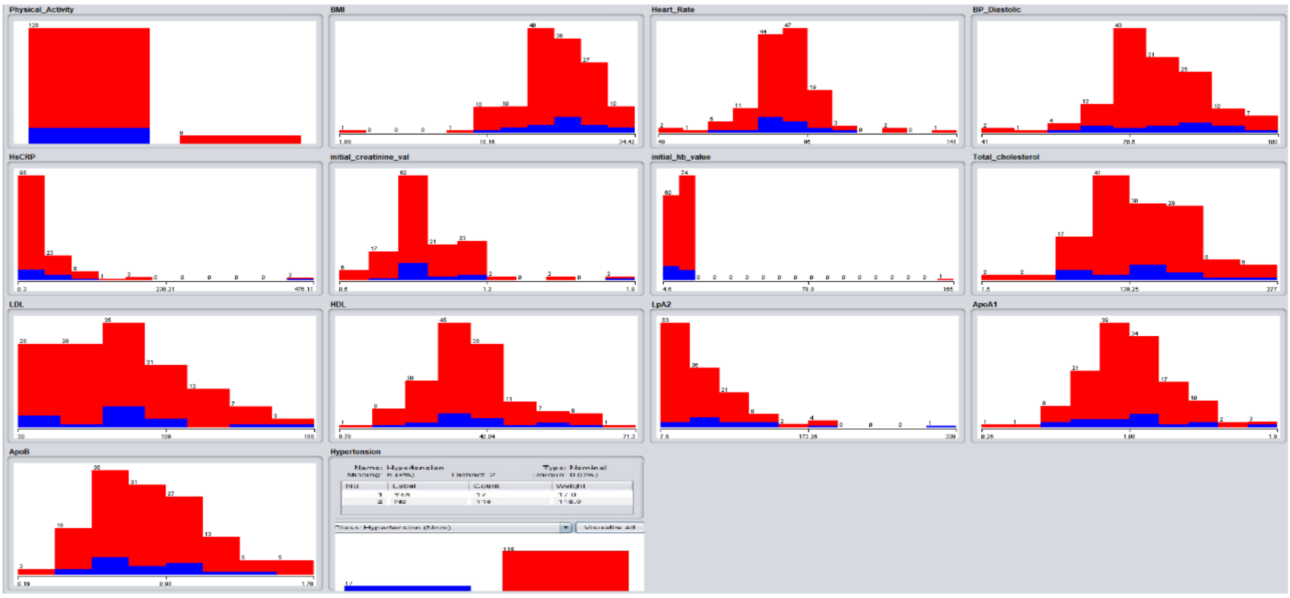

Figure 3 Association of different clinical parameters with hypertension

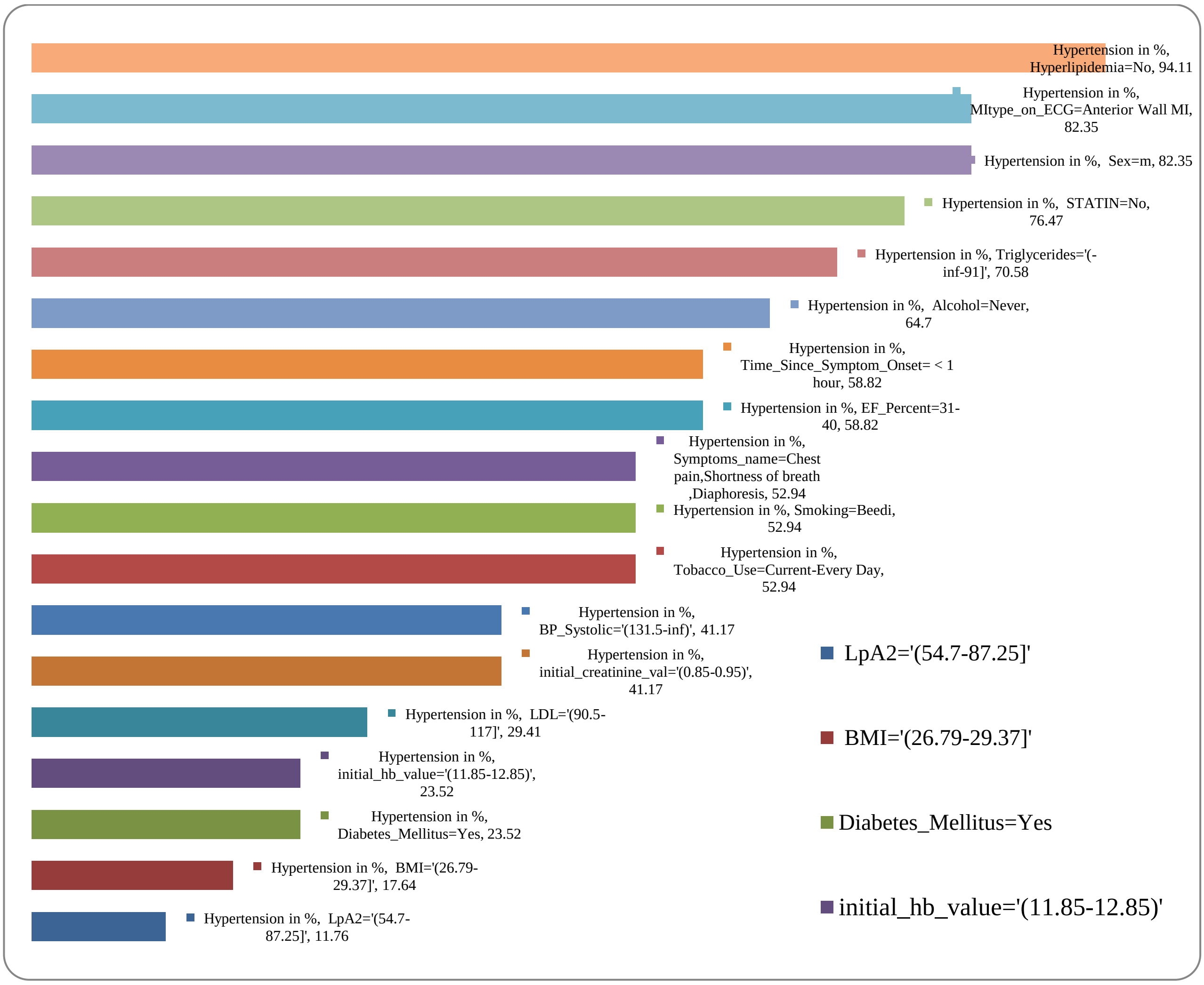

Figure 4. Relative frequency of different parameters for hypertension (N = 135)

DISCUSSION

In this study, 8.9% of the patients were diabetic, 12.6% were hypertensive, and 8.9% had a history of dyslipidemia, which was slightly low in comparison to previous studies where diabetes and hypertension were seen in 14.7 % and 38.1 % of young MI patients, respectively (Chan et al. 2012, McManus et al. 2011). Hypertension is under-diagnosed in young individuals, and if left untreated for a lengthy period, it might lead to a higher risk of MI in this age group (Chan et al. 2006). Although diabetes is less common in young people, it is still a hazard factor for MI in this age group. The adjusted OR of MI among 45-year-old men with diabetes was found to be 8.34% (95 percent CI, 1.67–41.6) as compared to non-diabetic young people (Oliveira et al. 2009). When plasma Lp(a) levels rise, in the absence of other risk factors, the risk of coronary artery disease increases with a high TC/HDL ratio (Dai et al. 2018). In the current study, the average BMI was found to be 25.7±4.31 kg/m^2, with 15.5% of the patients having a BMI of more than 30 kg/m^2. Available data supported our finding that young MI patients had a higher BMI than older MI patients (Hipp et al. 2000).

In our study, we found out various rules of cardiovascular diseases and classified them by finding their lift and confidence. In this paper, Hotspot was used to extract frequent patterns and various promising rules within real medical data. Thirty-four attributes of 135 individual patients were considered for the experiment. The experiment was carried out using the "Weka" tool. Weka is one of the tools used for extracting rules to find out the association between different stored real parameters.

Table 2 lists the ten significant rules for diabetic patients in order of highest support scores. There was a 100 percent certainty that a patient had an HDL value was 28.25 to 33.65 if he or she had a value of “BP Systolic” was 106.5 to 113.5 and Age was more than 39.5. As a result, the patients died. ARM came up with 127 regulations for men and no rules for women. Between the sexes, there was a variance in symptom rules. Figure 2 shows the relative frequency of different parameters for diabetic patients. We can see in the figure that if BMI is ranging between 26.79 to 29.37, then the chances of Diabetes is 33 percent. The association rule in Table 2 and Figure 2 clearly shows that the combination and values of these attributes were responsible for diabetes. This work will be beneficial for understanding patterns of different medical parameters for diabetic patients.

We can see rules in Table 3 and Figure 3 that different medical parameters were responsible for hypertension. For Hypertensive patients, 17% had a BMI value which was greater than 26.7. Similarly, 94% have had Hyperlipidemia, 82% sex was male, 82% had MI type on ECG was Anterior Wall MI, 70% had Triglycerides value below 91, 64% were not consuming alcohol, and 59% had ejection fraction (EF) percent between 31-40. The association rule in Table 3 and Figure 3 shows that the combination of these attributes was responsible for hypertension. Figure 4 shows the Relative Frequency of different parameters in Hypertensive patients. We can see in the Figure that if BMI is ranging between 26.79 to 29.37 then the chance of hypertension is 17.64%.

Strength and limitation

This work contributes to existing studies by further prove the validity of association rule mining in predicting the development of hypertension and diabetes mellitus, especially among myocardial infarction patients. However, as this was a single-center study, the study population might not be diverse and therefore resulted in a lack of universal findings.

CONCLUSION

We conclude that the Hotspot algorithm handles data with real value and generates rules that hold the highest confidence value, lift (>1) assists us in the advanced detection of hypertension and diabetes. In the present study, Diabetes Mellitus has been directly related to the levels of initial hemoglobin, alcohol, ApoB, ApoA1, total cholesterol, BMI, LDL, HDL, hsCRP, heart rate, diastolic blood pressure, age, statin use, and MI type on ECG. ARM techniques identify different patterns and rules to predict the development of diabetes mellitus in myocardial infarction patients. Also, our study finds patterns for hypertension and overall rules among patients, utilizing rules combination of clinical parameters like systolic blood pressure, heart rate, physical inactivity, and serum levels of LpA2, hsCRP, creatinine, hemoglobin, ApoB, ApoA1, HDL, LDL, and total cholesterol in myocardial infarction patients.

Thus, the association rule mining technique could be used to identify patterns utilizing clinical attributes and biochemical laboratory tests to predict the development of hypertension and diabetes mellitus in early-onset myocardial infarction patients.

Acknowledgement

We thank to the clinical data of cardiovascular disease patients presented at G.B. Pant Institute of Postgraduate Medical Education and Research in New Delhi, India.

Conflict of interest

None.

Funding disclosure

None.

Author contribution

AS, DS, KU,VM, and JS contributed to the conceptualization study, design and methodology. SS,MM and PM were data collection, data analysis. PKD were contributed to the final check the manuscript and grammar.

REFERENCES

Agapito G, Calabrese B, Guzzi P, et al (2019). Association rule mining from large datasets of clinical invoices document. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE.

An J, Chen Y, Chen H (2005). DDR: An index method for large time-series datasets. Inf. Syst. 30, 333–348.

Chan C, Chen W, Kuo H (2012). Circadian variation of acute myocardial infarction in young people. Am. J. Emerg. Med. 30, 1461–1465.

Chan M, Woo K, Wong H, et al (2006). Antecedent risk factors and their control in young patients with a first myocardial infarction. Singapore Med. J. 47, 27–30.

Chen Y, Chen F (2008). Identifying targets for drug discovery using bioinformatics. Expert Opin. Ther. Targets 12, 383–389.

Dabla P, Sharma S, Saurabh K, et al (2021). Atherogenic index of plasma: A novel biomarker and lipid indices in young myocardial infarction patients. Biomed. Biotechnol. Res. J. 5, 184–190.

Dai W, Long J, Cheng Y, et al (2018). Elevated plasma lipoprotein(a) levels were associated with increased risk of cardiovascular events in Chinese patients with stable coronary artery disease. Sci. Rep. 8, 1–9.

Doughty M, Mehta R, Bruckman D, et al (2002). Acute myocardial infarction in the young - The University of Michigan experience. Am. Heart J. 143, 56–62.

George M, Tong X, Kuklina E, et al (2011). Trends in stroke Hospitalizations and associated risk factors among children and young adults, 1995-2008. Ann. Neurol. 70, 713–721.

Gupta A, Wang Y, Spertus J, et al (2014). Trends in acute myocardial infarction in young patients and differences by sex and race, 2001 to 2010. J. Am. Coll. Cardiol. 64, 337–345.

Han J, Kamber M, Pei J (2011). Data mining: Concepts and techniques, third edition. Morgan Kaufmann, Massachusetts.

Hastie T, Tibshirani R, Friedman J (2009). The elements of statistical learning: Data mining, inference, and prediction, second edition. Springer, New York.

Hipp J, Güntze U, Nakhaeizadeh G (2000). Algorithms for association rule mining - A general survey and comparison. ACM SIGKDD Explor. Newsl. 2, 58–64.

Joshi P, Islam S, Pais P, et al (2007). Risk factors for early myocardial infarction in South Asians compared with individuals in other countries. JAMA 297, 286–294.

Lau R, Tang M, Wong O, et al (2006). An evolutionary learning approach for adaptive negotiation agents. Int. J. Intell. Syst. 21, 41–72.

McManus D, Piacentine S, Lessard D, et al (2011). Thirty-year (1975 to 2005) trends in the incidence rates, clinical features, treatment practices, and short-term outcomes of patients <55 years of age hospitalized with an initial acute myocardial infarction. Am. J. Cardiol. 108, 477–482.

McNamara K, Alzubaidi H, Jackson J (2019). Cardiovascular disease is a leading cause of death: How are pharmacists getting involved? Integr. Pharm. Res. Pract. 8, 1–11.

Mohan H (2005). Textbook of pathology. Anshan Publishers, Tunbridge Wells.

Oliveira A, Barros H, Azevedo A, et al (2009). Impact of risk factors for non-fatal acute myocardial infarction. Eur. J. Epidemiol. 24, 425–432.

Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. In: Transactions on Information Technology in Biomedicine. IEEE, pp. 334–343.

Patil B, Joshi R, Toshniwal D (2010). Association rule for classification of type-2 diabetic patients. In: Second International Conference on Machine Learning and Computing. pp. 330–334.

Shehabi S, Baba A (2021). MARC: Mining association rules from datasets by using clustering models. Int. J. Multidiscip. Stud. Innov. Technol. 5, 89–93.

Simon G, Schrom J, Castro M, et al (2013). Survival association rule mining towards type 2 diabetes risk assessment. AMIA Annu. Symp. Proc. Arch. 16, 1293–1302.

Soni U, Behara S, Krishnan K, et al (2016). Application of association rule mining in risk analysis for diabetes mellitus. Int. J. Adv. Res. Comput. Commun. Eng. 5, 548–551.

Srinath, R., Shah, B., Varghese, C., Al, E., 2005. Responding to the threat of chronic diseases in India. Lancet 366, 1744–1749.

Szathmary L, Valtchey P, Napoli A (2010). Generating rare association rules using the minimal rare itemsets family. Int. J. Softw. Informatics 4, 219–238.

Tandan M, Acharya Y, Pokharel S, et al (2021). Discovering symptom patterns of COVID-19 patients using association rule mining. Comput. Biol. Med. 131, 1–12.

Vasoya A, Koli N (2016). Mining of association rules on large database using distributed and parallel computing. In: 7th International Conference on Communication, Computing and Virtualization. pp. 221–230.

Virani S, Alonso A, Aparicio H, et al (2021). Heart disease and stroke statistics - 2021 update: A report from the American Heart Association. Circulation 143, 254–743.