An intelligent AAA++ approach to predict high blood pressure using PARP classifier

Satyanarayana Nimmala[a,*], Y. Ramadevi[b], B. Ashwin Kumar[a], R. Sahith[a]

[a] *Department of CSE, CVR College of Engineering, Vastunagar, Mangalpalli (V), Ibrahimpatnam (M), Rangareddy (D), Telangana, 501510, India*
[b] *Department of CSE, Chaitanya Bharathi Institute of Technology, Kokapet (V), Gandipet (M), Hyderabad (D), Telangana, 500075, India*

ABSTRACT

Objective: High Blood Pressure (HBP) is a major health challenge of many around the world. Existing research covers extensively how to treat HBP, but predicting HBP in advance based on biological and psychological parameters of a person is not covered in the literature. The objective of this paper is to predict HBP based on Bio-Psychological factors of a person.

Methods: We proposed an intelligent Rule-based classifier to predict HBP. The proposed model can be used to prevent HBP rather than using medication. In our approach, we considered AAA++ (Age, Anger level, Anxiety level, Obesity level (+), Cholesterol level (+)) of a person for experimental study. The proposed approach uses priority-based apriori rule pruning (PARP) classifier, which works in 3 stages. Stage 1: generate association rules using apriori. Stage 2: it uses the priority of an attribute to prune the association rules generated in stage 1. Step 3: Rules extracted in stage 2 are used to build a rule-based classifier to predict the class label of test instances. The Results of the proposed model are compared with JRip, PART, OneR and, ZeroR.

Results: Experimentation is done on real-time data set using 10 fold cross-validations. In each fold, 90% data is used to train the model and 10% is used to test the model. The proposed approach has shown improved accuracy (86.4%) and reduced mean length of a rule (1.7) compared to existing rule-based algorithms. Although JRip is good at accuracy (86.9%), but the proposed model has outperformed at the mean length of the rule (1.7).

Conclusion: The extracted rules after experimentation are understandable and informative to the technical and nontechnical community to predict HBP.

1. Introduction

The human body is made of trillions of cells, these cells needs oxygen, and energy for their livelihood. So the heart pumps the blood in order to provide oxygen and energy for all the tissues of the human body. While it pumps the blood, it creates some pressure inside the blood vessels; this pressure is called blood pressure. Blood pressure (BP) is normally represented as 120 over 80 or 120/80 mmHg or systolic blood pressure (SBP) over diastolic blood pressure (DBP). If SBP is greater than 140 mmHg or DBP is greater than 90 mmHg or both on repeated measurements, then the person is diagnosed as a victim of high blood pressure[15] or also called Hypertension. Nowadays HBP is one of the root causes of a brain stroke, heart attack, and kidney failure. BP of a person may be raised because of various reasons such as obesity, an unhealthy diet, excess bad cholesterol, excess sodium intake, smoking, consumption of alcohol, age, anger, anxiety, lack of physical exercise etc.[6] In this paper, we kept our eagle eye on the, impact of age, anger level, anxiety level, obesity level and cholesterol levels in raising the blood pressure.

Classification rule mining is one of the emerging concepts in data mining. The main aim of classification rule mining is to find the minimum set of rules in the form IF-THEN, based on training records.[1,2] Many studies suggested that, rule-based classification is highly expressive and also generates easily understandable, interpretable rules.[5] The main aim of the rule-based classification is to find a smaller set of overall rules using training records. If the number of attributes in training data set is large, all possible rules also large, this becomes computationally expensive.[7] But our proposed approach, we used modified apriori Algorithm to generate association rules, where many rules are pruned based on priority of the attribute appear in the antecedent part of the rule. Positive class rules are generated using attributes with high priority, and negative class rules are generated using attributes with low priority. The priority of each attribute is calculated using Pearson correlation coefficient value, as we considered priority,

---

the attribute space is reduced by half of the original attribute space.

2. Overview of association and rule-based classification

Association analysis is to find interesting relationship between data items in the large data sets and the relationship is represented in terms of association rules or set of frequent item sets.[8] The major area in which association analysis is used is market basket analysis. In this paper we, classified the test instances based on association rules.

2.1. Frequent item sets generation using Apriori Algorithm

Apriori generates frequent itemsets using support count, and then it extracts all possible association rules from frequent itemsets using confidence.[12] While generating frequent itemsets aprioris uses apriori property: if any set is not frequent, then all of its supersets also not frequent, based on this property we can skip many sets without finding their support and confidence.

2.2. Extracting association rules

Association rule represents the occurrence of a certain event together with occurrence of another event or set of events [4]. It is represented as given below.

Rule R: $A \Rightarrow B$ where A and B are disjoint sets and $A \cap B = \varphi$

The above rule R says appearance of B when A is appeared. The support count of an association rule represents the usage of the rule and the confidence of association rule represents the certainty of the rule. The lift of association rule represents the correlation coefficient value between the attributes participating in the antecedent and consequent part of the rule [9].

$$\text{Support count } (A \Rightarrow B) = \frac{\text{Occurrence of } (A, B)}{N} \quad (1)$$

$$\text{Confidence } (A \Rightarrow B) = \frac{\text{Occurrence of } (A, B)}{\text{Occurance } (A)} \quad (2)$$

$$\text{Lift } (A \Rightarrow B) = \frac{\text{Support } (A \Rightarrow B)}{\text{Support}(A) * \text{Support } (B)} \quad (3)$$

Here Occurrence of (A, B) is the number of times in which A, B together appeared, and N is total transactions or events. Association rule is said to be strong if it satisfies the threshold support count, and threshold confidence specified and having positive lift value.

2.3. Introduction to rule-based classification

The rule-based classifier classifies the records based on set of rules of the form "if condition is true then, it predicts the class label attribute value". Example rules are in the form of, Rule: IF (test condition) → Y. LHS of the rule is called antecedent or condition and RHS of the rule is called rule consequent.[8] In our proposed approach we used association rule-based classification. Our approach produces only rules such that the rule consequent is always either yes or no.

2.4. Existing work

In the last 5 decades, mathematical modeling of biological and physiological systems gained lot of interest and popularity, modeling and simulation of biophysiological systems is now emerging topic in the biomedical research (Gavaghan et al., 2006). These models can help us to improve the understanding of inter and intra relationship between biological and psychological elements of human body (Guyton et al., 1972). These models also can be used in various applications like disease prediction or diagnosis, disease treatment, decision-making and etc. (Weiss et al., 1978; Pedersen et al., 1996; Lewenstein, 2005; La

Rovere et al., 2008). These models can be part of software-based teaching and assessment (Coleman and Randall, 1983; Kofranek et al., 2005) of human biopsychological problems.

Existing research says age and blood pressure are positively correlated (Mikko Peltokangas at al. 2017, Elisabete Pinto 2007, Julie K.K. Vishram at al. 2012). As the person ages the pathways of the heart's pacemaker system starts collecting fat.[19] As we age, functionality of kidney to process sodium in the food, we consume decreases.[19] According to Richard N, Obesity and blood pressure are positively correlated (Richard N. Re 2009, Krzysztof Narkiewicz 2006). In the obese people, there is an increase in fatty tissue, which requires more nutrients and oxygen to live, so heart has to pump more blood to reach these new tissues, causing raise in the blood pressure.[20] Existing research work in the literature revels that the cholesterol and blood pressure are also positively correlated (Masaru Sakurai 2011, D. Sesso 2005, LA Ferrara 2002). If there is too much cholesterol in the blood, it can cause the arteries to narrow, and raises the blood pressure,[21] keeping this in view we took age, obesity, cholesterol levels as part of experimental study. Existing research also says, stress levels and blood pressure are positively correlated (Daniela Lucini 2007). If a person gets angry, then the fight or flight mode of Sympathetic Nervous System gets activated.[11] As a response nerves send more blood to muscles and, brain, which elevates the blood pressure.[13] In this paper we developed a model to predict the HBP based on biological and psychological factors of human body. Blood pressure is affected by many factors, such as excess sodium intake, lack of physical exercise, age, obesity, smoking, alcohol consumption, gender and etc. The existing work in the literature unfolds, the effect of age on BP, the effect of obesity on BP, the effect of blood cholesterol on BP, and the affect of stress level on the BP using simple linear regression,[11],[19–21] but the combined effect of all these on elevating the BP is not yet studied. In this paper, we proposed a new approach to find the combined effect of AAA + + on elevating the BP.

3. Proposed approach

We did our experiments on real-time data set; data is collected from 1000 people, where each person record consists of AAA + +, SBP and DBP.[16] Three attributes age, obesity, cholesterol levels are collected from medical reports obtained from a medical diagnostic center, Hyderabad, India. Two attributes anger level and anxiety level are collected through a personal interview. Blood pressure is measured repeatedly using the digital monitor. SBP and DBP are used to calculate mean arterial blood pressure, where mean arterial blood pressure is used to find the impact of each attribute. Class label attribute is made as YES, if SBP is more than 140 mmHg, or DBP is more than 90 mmHg or both, otherwise class label attribute is taken as NO.[18] The process used to calculate the anger level and anxiety level is discussed in Section 3.1 and Section 3.2.

3.1. Anger measurement

The anger level of a person is measured through personal interview; every person is posed to a set of predefined questions.[11] The answer obtained for each question is marked on the scale of 0–3, for example, if the question is, do you get angry if people do wrong to you, the answer could be 0 means (no, never), or it could be 1 means (yes, rarely), or it could be 2 means (yes, often) or it could be 3 means (yes, most of the time). We used 10 such predefined questions,[6] like do you get angry if any one hurts you, do you get angry if people are not on time, do you get angry if others won't respect you, do you get angry if things won't go on as you expected etc. And the answer is marked as 0 or 1 or 2 or 3. We considered the average value of all answers for experimental analysis.

3.2. Anxiety measurement

The anxiety level of a person is also measured through personal interview; every person is posed to a set of 20 predefined questions[6,11,13] such as do you have breathing difficulty often, do you compare always with others, do you feel that you are not doing worth noted, do you feel sense of dryness in the mouth, do you scare often without clear reason, the answer obtained for each question is marked on the scale of 0–3, for example if the question is, do you have constant fear about anything in life, the answer could be 0 (no never), or it could be 1 (yes, rarely), or it could be 2 (yes, often) or it could be 3 (yes, most of the time), and the answer is marked as 0 or 1 or 2 or 3. We considered the average value of all answers for experimental analysis.

3.3. Data preprocessing

It is the process in which a series of statistical and data mining techniques are used to convert the raw data collected into an understandable format of the Algorithm used. As apriori works well with binary and categorical attributes, in this stage, continuous attributes are converted into categorical attributes. To know the impact of each attribute on class label attribute, we considered correlation coefficient value. We considered the mean arterial blood pressure (MABP) as class label attribute while calculating PCC, Equation (4) is used to calculate PCC value. The calculated PCC value is used to tabulate priority of each attribute[14,17] as shown in Table 1.

$$PCC = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (4)$$

$$MAP = \frac{((2 * DBP) + SBP)}{3} \quad (5)$$

3.4. PARP algorithm to extract classification rules

In this paper, we proposed a new rule based classifier to extract classification rules named as priority-based Apriori Rule Pruning (PARP) classifier. It is an intelligent Algorithm that extracts the association rules generated from apriori using priority of an attribute. These rules are used as classification rules in the next stage to classify the input instances. Table 1 is used to set anxiety level and anger levels of a person as high priority attributes, the same table is used to set, age, obesity and cholesterol levels as low priority attributes. While extracting yes class rules the algorithm looks for rules with only high priority attributes with higher value or high priority attribute with medium value as part of conjunction. While extracting no class rules the algorithm looks for rules with only low priority attributes with low or medium value as part of conjunction. Extracted rules are ordered in descending order based on coverage of records. Later rules are swapped on the principle of super set and subset coverage.

3.5. PARP for YES class

Input: list of a rules from apriori, D is data set of training records with class label yes, TR is the Training records covered by rule R where SC is the thresh hold support count; MC is the minimum confidence of

the rule.
Output: Set of rules.

Algorithm. Papriori_yesclass_rulegeneration (D, SC, MC)
Step 1: Intialize LD = D where LD is left over data set after extracting the records covered by the selected rule.
Step 2: generate top 10 rules using apriori.
Step 3: select a rule R, that consist high priority attribute with high or medium value in its Antecedent part, and covers highest number of records.
Step 4: Write R;
Step 5: LD = LD-TR.
Step 6: if(LD > ((15/100)*D))

Papriori_yesclass_rulegeneration(LD, SC, MC)
end if
else
return;

The above Algorithm uses recursive approach to learn one rule in each iteration, and stop when left over records are less than 15% of the total records in the current iteration, after extracting the records covered by the rule.

3.6. PARP for NO class

Input: list of a rules from apriori, D is data set of training records with class label no, TR is the Training records covered by rule R where SC is the thresh hold support count; MC is the minimum confidence of the rule.
Output: Set of rules.

Algorithm. Papriori_noclass_rulegeneration (D, SC, MC)
Step 1: Intialize LD = D, LD is left over data set after extracting the records covered by The Selected rule.
Step 2: generate top 10 rules using apriori.
Step 3: select a rule R, that consist low priority attribute with low or medium value in its Antecedent part, and covers highest number of records.
Step 4: Write R;
Step 5: LD = LD-TR.
Step 6: if(LD > ((15/100)*D))

Papriori_yesclass_rulegeneration(LD, SC, MC)
end if
else
return;

The above Algorithm uses recursive approach to learn one rule in each iteration, and stop when left over records are less than 15% of the total records in the current iteration, after extracting the records covered by the rule.

4. Results and discussion

In this section, we briefly discussed about the process involved in generation of association rules, extracting these rules based on priority to build rule based classifier. We also compared the results obtained using proposed model with existing rule based classifiers.

4.1. Extracting rules using PARP generated from apriori

Apriori Algorithm with minimum support counts as 0.4 and thresh hold confidence as 1 is executed, using training data to generate the rules.[2] Minimum support count is set initially at 1 and subsequently, the value is reduced by 0.1 each time till at least one rule is generated. At minimum support count 0.4 we obtained the association rules for the given input, so we set minimum threshold support count as 0.4.

Table 1
Priority of risk factor based on PCC value.

Risk factor	Calculated value of PCC_i	Priority based on PCC_i
Age (RF_1)	0.3976	low
Obesity level (RF_2)	0.3266	low
Cholesterol level (RF_3)	0.2866	low
Anger level (RF_4)	0.5546	high
Anxiety level (RF_5)	0.7014	high

Table 2
Confusion matrix of proposed approach.

Actual class		Predicted class	
		YES	NO
Fold 1	YES	20	9
	NO	4	67
Fold 2	YES	17	9
	NO	3	71
Fold 3	YES	17	9
	NO	3	71
Fold 4	YES	21	7
	NO	9	63
Fold 5	YES	12	10
	NO	1	77
Fold 6	YES	14	10
	NO	4	72
Fold 7	YES	17	8
	NO	5	70
Fold 8	YES	14	12
	NO	3	71
Fold 9	YES	15	8
	NO	7	70
Fold 10	YES	20	7
	NO	9	64

Table 3
Class-wise accuracy details of existing rule based classifiers.

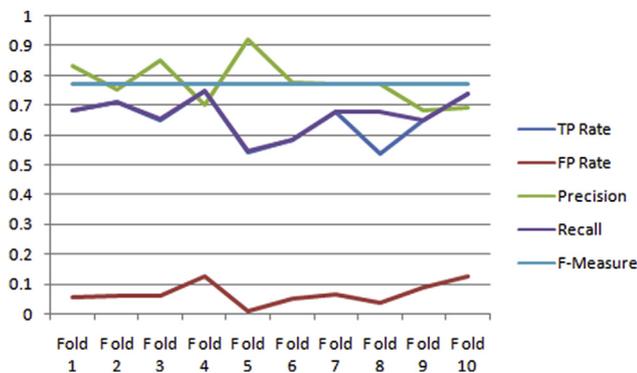| Algorithm used	Class	TP rate	FP rate	Precision	Recall	F-Measure
JRip	NO	0.936	0.331	0.894	0.936	0.915
	YES	0.669	0.064	0.778	0.669	0.719
PART	NO	0.936	0.359	0.886	0.935	0.910
	YES	0.641	0.065	0.767	0.641	0.698
OneR	NO	0.873	0.526	0.832	0.873	0.852
	YES	0.474	0.127	0.556	0.474	0.512
ZeroR	NO	1.000	1.000	0.749	1.000	0.856
	YES	0.000	0.000	0.000	0.000	0.000

Fig. 1. Performance of proposed approach with respect to yes class.

However at confidence 1 there exist many association rules, so confidence is set at 1. We used 10 fold cross-validations in generating association rules. In each fold, 90% data is used for training and 10% data is used for testing. Each time the training data set is split into yes class data and no class data. As apriori works well on binary valued attributes,[3] for each class while generating rules we considered only two values of the selected attribute. For yes class, attributes with values high, medium is taken and for no class, attributes with values low, medium is taken. Performances details of proposed classifier[6,14] and existing classifiers are tabulated in Tables 2–4. Figs. 1–3 represents performance of proposed approach against Yes class, and No class. The extracted rules in Fold 1 are given below.
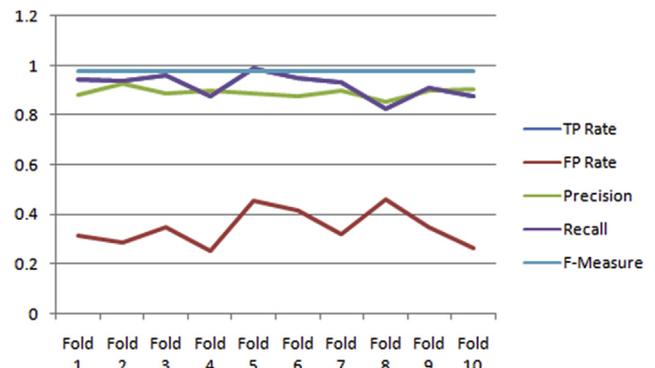

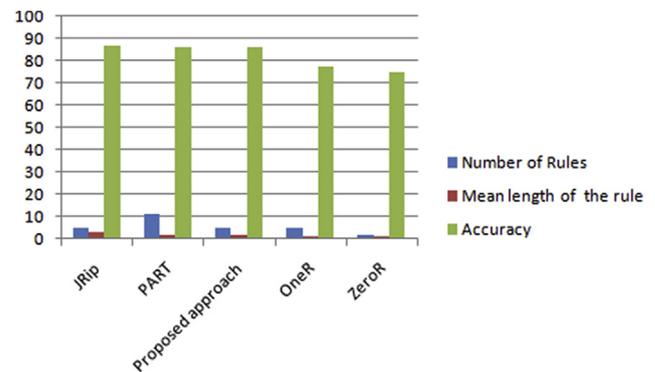Fig. 2. Performance of proposed approach with respect to no class.

Fig. 3. Performance analysis of existing and proposed Algorithm.

4.2. Rules obtained in Fold 1 using training data set 1

1 If (obesity ≤ 25) Then Class = No
2 If (25 < obesity ≤ 30) AND (anxiety ≤ 1) Then Class = No
3 If (anxiety > 2) Then class = Yes
4 If (150 < cholesterol ≤ 200) AND (1 < anxiety ≤ 2) Then Class = Yes
5 If (cholesterol > 200) AND (anger > 2) AND (1 < anxiety ≤ 2) Then class = Yes

4.3. Comparative study of proposed approach with existing rule-based classifiers

Table 4

4.4. Comparative analysis

In this paper, we have proposed a new model to classify the victims of HBP using real-time data set. The proposed model considers AAA++ to predict class label attribute, so far no research is found on AAA++ to predict high blood pressure in the literature. Existing research addresses only, the effect of age on blood pressure,[19,23] the effect of obesity on blood pressure,[20] the effect of cholesterol on blood

Table 4
Comparative analysis of existing rule-based classifiers and proposed classifier.

S. no	Algorithm used	Number of rules	Mean length of the rule	Accuracy
1	JRip	5	3.4	86.9
2	PART	11	1.6	86.1
3	Proposed approach	5	1.7	86.4
4	OneR	5	1	77.3
5	ZeroR	2	1	74.9

pressure[21,22] but their combined effect along with anger and anxiety levels is addressed in this article. As there are no existing algorithms in the past experimented on AAA + + to predict high blood pressure, we compared the experimental results with rule-based classifiers supported in WEKA. Though accuracy of proposed Algorithm is 86.4 but it is exceptional in reduced mean length comparatively with existing algorithms. Table 4 unfolds that, the proposed algorithm outperforms in terms of accuracy and mean length of a rule compared with JRip, PART, OneR, and ZeroR supported in WEKA.[10] In this paper performance details of fold 1 like TP rate, Fp rate, Precision, Recall and F-measure are mentioned, performance details of other folds of proposed model are not mentioned as we considered the aggregated values of the all the 10 folds for comparative analysis. The proposed model is experimented on real time data set, so far no research has conducted the experiments on AAA + +, so we compared our results with existing standard rule based classifiers.

5. Conclusion

This paper focus on the classification of a person whether he or she becomes victim of a HBP or not. Classification is done based on the age, anger level, obesity level, cholesterol level, and anxiety level of a person. We used PARP to extract the rules, later these rules are used as classification rules to build the rule-based classifier. Our proposed rule-based classifier has shown improved performance in terms number of rules, mean length of each rule in classifying the records, comparatively with existing rule-based classifiers like, JRip, PART, OneR and ZeroR. Our approach has shown 86.4% accuracy and 1.7 mean length of a rule in classifying the test instances using 10 fold cross-validations. In future, we would like to extract the fuzzy classification rules to improve the accuracy as well as certainty of the IF-THEN rules.

References

1. Chen Guoqing, Liu Hongyan, Yu Lan, Wei Qiang, Zhang Xing. A new approach to classification based on association rule mining. *Decis Support Syst.* 2006;42:674–689.
2. Nguyen Loan TT, Vo Bay, Hong Tzung-Pei, Thanh Hoang Chi. Classification based on association rules: A lattice-based approach. *Expert Syst Appl.* 2012;39:11357–11366.
3. Li W, Han J, Pei J. *CMAR, Accurate and Efficient Classification Based on Multiple Class-Association Rules, ICDM 2001.* San Jose, California: IEEE Computer Society; 2001:369–376.
4. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases.* San Francisco, CA, USA: VLDB '94, Morgan Kaufmann Publishers Inc.; 1994:487–499.
5. Ali K, Manganaris S, Srikant R. Partial classification using association rules. *Proceedings of KDD-97.* 1997; 1997:115–118.
6. Nimmala Satyanarayana, Ramadevi Y, Sahith R, Ramalingaswamy Cheruku. High blood pressure prediction based on AAA + + using machine-learning algorithms. *Cogent Engineering.* 2018;5:1497114.
7. Agrawal R, Imielinski T, Swami A. *Mining Association Rules between Sets of Items in Large Databases, Proceeding of 1993 ACM-SIGMOD International Conference on Management of Data.* Washington, D.C.: ACM Press; 1993:207–216.
8. Burdick D, Calimlim M, Gehrke J. MAFIA: A maximal frequent itemset algorithm for transactional databases. *International Conference on Data Engineering.* 2001; 2001.
9. Marinica Claudia, Guillet Fabrice. Knowledge-based interactive post mining of association rules using ontologies. *IEEE Trans Knowl Data Eng.* June 2010;22(6):784–797.
10. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tool and Technique with Java Implementation.* San Francisco: Morgan Kaufmann; 2000.
11. Hendricks L, Bore S, Aslinia D. The effects of anger on the brain and body. *National Forum Journal of Counseling and Addiction.* 2013;2:1–12.
12. Shendge Pratiksha, Gupta Tina. Comparative study of Apriori & FP growth algorithms. *Indian Journal of Research.* March 2013;2(3).
13. Pandey P, Kyung Lee E, Pompili D. A distributed computing framework for real-time detection of stress and its propagation in a team. *IEEE J Biomed Health Informs.* 2016;20:1502–1512.
14. Nimmala Satyanarayana, Ramadeve Y, Aswian Kumar B. An innovative data-driven computational model to predict high blood pressure based on AAA + +. *Int J Eng Technol.* 2018;7(3.3):114–118.
15. WHO. World Health Organization. Available: \url : www.who.int/gho/ncd/risk _factors/blood_pressure_prevalen ce_text/en/.
16. Millasseau SC, Kelly RP, Ritter JM, Chowieczyk PJ. The vascular impact of aging and vasoactive drugs: Comparison of two digital volume pulses measurements. *Am J Hypertension Sci.* 2003;16:467–472.
17. Satyanarayana N, Ramalinga swamy CH, Ramadevi Y. Survey of classification techniques in data mining. *Int J Innov Sci Eng Technol.* 2014;1:268–279.
18. Babbs CF. Oscillometric measurement of systolic and diastolic blood pressures validated in a physiologic mathematical model. *Biomed Eng Online.* 2012;11:1–22.
19. Peltokangas Mikko, Vehkaoja Antti, Verho Jarmo. Age dependence of arterial pulse wave parameters extracted from dynamic blood pressure and blood volume pulse waves. *IEEE J Biomed Health Inform.* 2017;21:142–149.
20. Mertens Ilse L, Van Gaal Luc F. Overweight, obesity, and blood pressure: The effects of modest weight reduction. *Obes Res.* 2000;8:270–278.
21. Bhende AM, Zade SB, Sitre SR. Influence of family hypertension on blood pressure, serum cholesterol, high-density lipoprotein cholesterol in general population. *Int J Pharma Bio Sci.* 2012;3:176–180.
22. Zhang Bing, Wei Zhiyao, Ren Jiadong, Cheng Yongqiang, Zheng Zhangqi. An empirical study on predicting blood pressure using classification and regression trees. *IEEE ACCESS, Special Section on Human-Centered Smart Systems and Technologies.* 2018;6:21758–21768.
23. Currie Gemma, Christian Delles. Blood pressure targets in the elderly. *J Hypertens.* 2018;36:234–236.