Predicting High Blood Pressure Using Decision Tree-Based Algorithm

Satyanarayana Nimmala, Y. Ramadevi, Srinivas Naik Nenavath and Ramalingaswamy Cheruku

Abstract High blood pressure, also called as hypertension, is a state developed in biological system of human beings by knowingly or unknowingly. It may occur due to varied biological and psychological reasons. If high blood pressure state is sustained for a longer cycle, then the person may be the victim of heart attack or brain stroke or kidney disease. This paper uses a decision tree-based J48 algorithm, to predict whether a person is prone to high blood pressure (HBP). In our experimental analysis, we have taken certain biological parameters such as age, obesity level, and total blood cholesterol level. We have taken the real-time data set of 1045 diagnostic records of patients in the age between 18 and 65. These are collected from a medical diagnosis center Doctor C, Hyderabad. Records (66%) are used to train the model, and remaining 34% records are used to test the model. Our results showed 88.45% accuracy.

Keywords Classification · Decision tree · Blood pressure monitoring

S. Nimmala (✉)
Department of Computer Science and Engineering, Osmania University,
Hyderabad 500007, India
e-mail: satyauce234@gmail.com

Y. Ramadevi
Department of Computer Science and Engineering, CBIT, Hyderabad 500075, India
e-mail: yrdcse.cbit@gmail.com

S. N. Nenavath
School of Computer and Information Sciences, University of Hyderabad,
Hyderabad 500046, India
e-mail: srinuphdcs@gmail.com

R. Cheruku
Department of Computer Science and Engineering, National Institute of Technology Goa,
Ponda 403401, Goa, India
e-mail: rmlswamygoud@gmail.com

D. Reddy Edla et al. (eds.), *Advances in Machine Learning and Data Science*,
Advances in Intelligent Systems and Computing 705,
https://doi.org/10.1007/978-981-10-8569-7_6

1 Introduction

When the heart beats, it pushes the blood against arteries with some force, which creates some pressure called as systolic blood pressure. The heart takes rest during the beats, while the pressure inside the arteries is called diastolic blood pressure. Most of the literature says that hypertension (HBP) or heart rate change leads to heart failure or stroke [1]. Recent health statistics show that men and women of age above 25 are prone to hypertension [3]. We consider blood pressure level, obesity condition, total cholesterol level as reported by laboratories and scientific studies as shown in Table 1.

In this research work, we address the classification methodology to predict whether a person is a victim of HBP or not. We have taken parameters like age, obesity level, and complete blood cholesterol level of a person.

The rest of the paper is organized as follows. A background of the factors influencing HBP is given in Sect. 2. Procedure for identifying a specific classifier based on the performance measures is given in Sect. 2.1, and Sect. 4 evaluates the performance of a classifier. Section 5 concludes and gives future outline of the paper.

2 Background Work

This section describes overview of the factors influencing HBP.

2.1 Factors Effecting BP

Blood pressure (BP) is mainly effected by the cardiac output (CO) and total peripheral resistance (TPR), which is calculated using (1).

$$BP = CO * TPR \tag{1}$$

Table 1 Ranges for BP, obesity, and cholesterol

BP	Low range	Normal range	Borderline	High range
Systolic	<90	90–130	131–140	>140
Diastolic	<60	60–80	81–90	>90
Total cholesterol level	<125	125–200	200–239	>240
Obesity	<18 (underweight)	18–24.9 (normal)	25–29.9 (overweight)	>30 (obese)

Here, CO is effected by increased venous return or stroke volume or heart rate and sympathetic activity. TPR is effected by the resistance that acts against the blood flow in the arteries. It may be due to a blood clot or fat in the blood vessels. CO effects the systolic blood pressure, whereas TPR effects the diastolic blood pressure [4]. So, CO is proportional to a number of heart beats per minute (HBM) and volume of blood (BV) pushed in each beat.

$$CO = HBM * BV \tag{2}$$

$$BP = Systolic/Diastolic \tag{3}$$

$$Systolic/Diastolic = CO/TPR \tag{4}$$

Sometimes, we may consider mean arterial blood pressure (MABP) which can be calculated as

$$MABP = (2 * (Diastolic + Systolic))/3 \tag{5}$$

If MABP [2] is within the range, then all organs and tissues will get enough blood, oxygen, and nutrients. Our paper focuses on age, obesity, and cholesterol levels of a person in elevating the blood pressure.

2.2 Impact of Age, Obesity, and Cholesterol on Hypertension

Aging is inevitable although a person has a healthy diet and exercise regularly. As we age, arteries may become narrow and harden and the ability of body to process sodium in the diet decreases [6]. The person is said to be obese if his body mass index (BMI) value is more than 30. In obese people, there is increased fatty tissue which needs more blood to live. High blood cholesterol is one of the main reasons for fat deposits in arteries which may harden the arteries also. However, age, obesity, and high blood cholesterol are playing their role to elevate the blood pressure of a person.

2.3 Nervous System

It will monitor changes in the state of a human body and transmits signals to and from different parts of the body. It is of two types [5]: voluntary system (VS) and automatic nervous system (ANS).

2.3.1 Voluntary System

It deals with movement and sense. The reaction of VS happens, when we move our hand in any direction or close our eye. It is under the control of human sense.

2.3.2 Automatic Nervous System

It controls the heartbeat, digestive system, BP, etc. It is not under the control of human sense. It is of two types, namely sympathetic nervous system (SNS) and parasympathetic nervous system (PNS).

Sympathetic Nervous System:

When we experience a sudden fear or high anxiety, SNS helps to prepare the body to defend itself by sending more blood to brain, muscles. This may increase BP. If SNS is active for a longer period, it is not good for human body (Fig. 1).

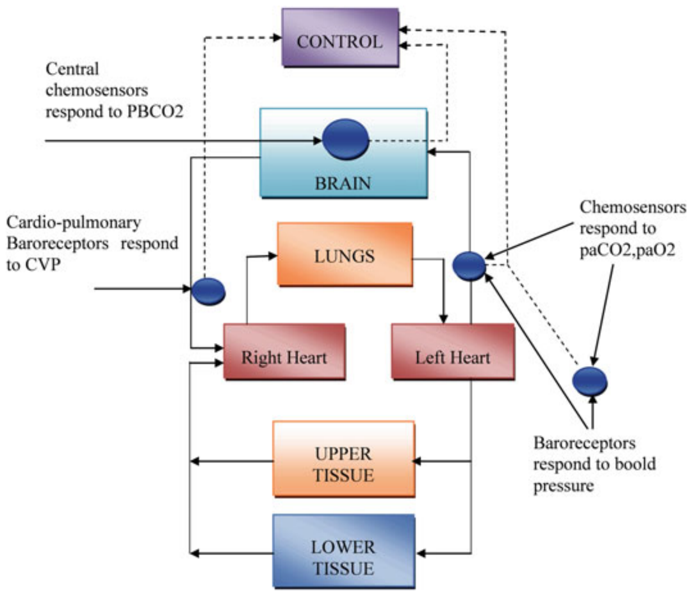


Fig. 1 Sympathetic action and its effects on circulation

Parasympathetic Nervous System:

PNS looks after the immune system to work effectively. PNS is active in the states like being calm, relaxed, and positive. It mostly works against the HBP.

2.4 How Brain Monitors BP

The BP management refers to heartbeat, arterial blood pressure, total blood volume, O_2, and CO_2 levels in the blood [7]. It uses the sensory system to know the state of the human system and update the same to the brain. Then brain converts this information to response signals via the peripheral nervous system, where the control listeners are waiting to take action on the received information. They may act to increase or decrease blood flow to various tissues by a change in systemic resistance. The response of the control is designed to maintain overall stable blood flow in the human body.

3 Proposed Methodology

In our work, we use a data mining classification technique. It borrows concepts from probability, statistics, fuzzy logic, neural networks to predict class labels of an object. The classification model will have two stages [8]: 1. training stage, where the model is trained with a set of records whose class labels are already known and 2. testing stage, where the model is going to predict class labels of a set of records whose class labels are unknown. There are various classifiers; we used a decision tree-based classifier named J48.

The classifier evaluation is most often based on prediction accuracy (the percentage of correct prediction divided by the total number of predictions). If accuracy is unsatisfactory, then to find the reason different factors must be observed like: 1. Relevant attributes of the problem are considered, 2. it may need more training records, 3. the dimensionality of the problem is too high, 4. algorithm that we select may not be suitable, and 5. parameter tuning is needed.

4 Experimental Result and Analysis

In our experiments, we use a classifier called J48-C 0.25-M2 (weka.classifiers.trees) to classify the tuples. The input data set is doctorcwekacsv-weka.filters.unsupervised.attribute.Remove-R2-weka.filters.unsupervised.attribute.Remove-R4. We have used 1045 number of records for our analysis and considered four attributes like age, obesity level, cholesterol, and BP for evaluation (Table 2).

Table 2 Performance measures of a classifier

Measure	Value
Classified instances (correctly)	314
Kappa statistic	0.7586
Mean absolute error	0.1506
Root mean squared error	0.3026
Relative absolute error	31.9906%
Root relative squared error	62.1366%
Total number of instances	355
Accuracy	88.4507%
Error rate	11.5493%

TP rate, FP rate, precision, recall, F-measure, Matthews correlation coefficient (MCC), receiver operating characteristic curve (ROC), precision–recall curve (PRC) are calculated for each class for the test data as presented in Table 3.

We trained the algorithm using 66% of input records and used remaining 34% of records to test the algorithm. The time taken to build the model is 0.05 s and time taken to test the model is 0.05 s. Some instances predicted correctly, some are predicted incorrectly, and the accuracy of the classifier, values of different errors and error rate of a classifier are as presented in Table 2.

The matrix in Table 4 represents 120 records, whose class labels are YES and predicted as YES, but 17 records whose class labels are YES and predicted as NO. For class label NO, 24 records are predicted as YES, and 194 records are predicted as NO.

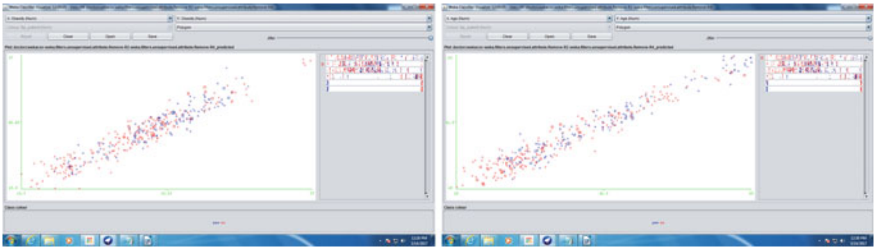
The total number of right and wrong predictions is drawn using decision tree-based J48 algorithm. Figure 2a represents the distribution of test samples predicted right, wrong for selected attribute obesity. Figure 2b represents the distribution of test samples predicted right, wrong for selected attribute age. Figure 3a represents the

Table 3 Detailed accuracy by class

TP rate	FP rate	Precision	Recall	F-measure	MCC	ROC area	PRC area	Class
0.876	0.110	0.833	0.876	0.854	0.759	0.932	0.871	Yes
0.890	0.124	0.919	0.890	0.904	0.759	0.931	0.942	No

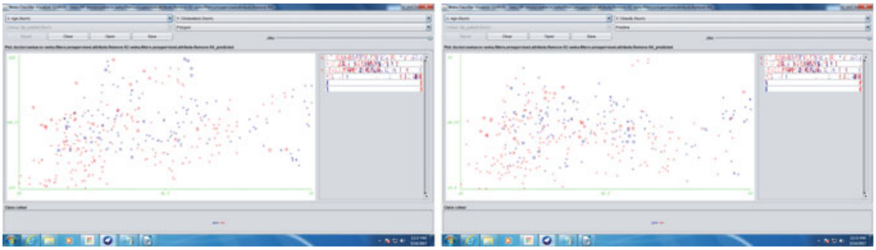
Table 4 Confusion matrix

Actual class versus predicted class	a = YES	b = NO
a = YES	120	17
b = NO	24	194



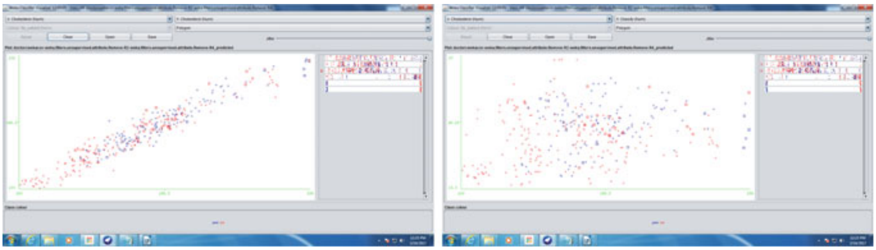
(a) obesity vs obesity (b) age vs age

Fig. 2 Comparison of error rate



(a) age vs cholesterol (b) age vs obesity

Fig. 3 Comparison of error rate



(a) cholesterol vs cholesterol (b) cholesterol vs obesity

Fig. 4 Comparison of error rate

distribution of test samples predicted right, wrong for selected attributes age, cholesterol. Figure 3b represents the distribution of test samples predicted right, wrong for selected attributes age, obesity. Figure 4a represents the distribution of test samples predicted right, wrong for selected attribute cholesterol. Figure 4b represents the distribution of test samples predicted right, predicted wrong for selected attributes cholesterol, obesity. Accuracy and error comparison of different classifiers are shown in Table 5.

Table 5 Accuracy and error comparison of different classifiers

Classifier	Accuracy	MAE	RMSE	RAE
Naive Bayes	70.7042	0.3714	0.4483	78.9052
Linear regression	70.9859	0.3859	0.4481	81.9946
REP tree	87.3239	0.1659	0.3093	35.2512
Multilayer perceptron	75.2113	0.3009	0.3993	63.9322
J48	88.4507	0.1506	0.3026	31.9906

MAE Mean absolute error; *RMSE* Root mean squared error; *RAE* Relative absolute error

5 Conclusion and Future Work

This paper used a decision tree-based J48 algorithm to predict whether a person is prone to HBP or not. By analyzing the experimental results, we conclude that if the age is more than 30 and cholesterol is more than 205, then the possibility of a person prone to the HBP is high. For certain records, if age is between 35 and 50, cholesterol is less than 150, and even obesity is more than 30 then as per the experimental analysis, the person may not be prone to HBP. If the age of a person is more than 55 irrespective of cholesterol and obesity levels, then that person is prone to HBP. As part of the future work, we will consider parameters like anxiety, depression along with age, obesity, and cholesterol, to predict HBP of a person using multiple regression-based classifiers.

References

1. AbuDagga, A., Resnick, H.E., Alwan, M.: Impact of blood pressure telemonitoring on hypertension outcomes: a literature review. Telemed. e-Health **16**(7), 830–838 (2010)
2. Cheng, H.M.: Measurement accuracy of a stand-alone oscillometric central blood pressure monitor: a validation report for MicrolifeWatchBP Office Central. Amer. J. Hypertens. **26**(1), 42–50 (2013)
3. Ding, X.R.: Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio. IEEE Trans. Biomed. Eng. **63**(5), 964–972 (2015)
4. Forouzanfar, M.: Oscillometric blood pressure estimation: past, present, and future. IEEE Rev. Biomed. Eng. **8**, 44–63 (2015)
5. Forouzanfar, M.: Ratio-independent blood pressure estimation by modeling the oscillometric waveform envelope. IEEE Trans. Instrum. Measure. **63**, 2501–2503 (2014)
6. Liu, J.: Model-based oscillometric blood pressure measurement: preliminary validation in humans. In: Proceedings of IEEE Conference on Engineering, Medical Biological Soc, pp. 1961–1964 (2014)
7. Mukkamala, R.: Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. IEEE Trans. Biomed. Eng. **62**, 1879–1901 (2015)
8. Satyanarayana, N., Ramalingaswamy, C.H., Ramadevi, Y.: Survey of classification techniques in data mining. Int. J. Innov. Sci. Eng. Technol. **1**, 268–278 (2014)