JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CLASSIFICATION AND ANALYSIS OF BLOOD PRESSURE USING MACHINE LEARNING TECHNIQUES

[1]Nagireddy.Sowjanya, [2] Dr. Kunjam Nageswara Rao, [3]Dr. G. Sita Ratnam, [4]Dangety.Sowjanya
[1] M.Tech Student, [2] Professor, [3]Associate. Professor, [4]Research Scholar
[1, 2, 4] Department of Computer Science and Systems Engineering
[1, 2, 4] Andhra University College of Engineering, AP, India
[3]Lendi Institute of Engineering and Technology
[1]nagireddy0316@gmail.com, [2]kunjamnag@gmail.com, [3]sitagokuruboyina@gmail.com, [4]Sowjid91@gmail.com

Abstract

Blood pressure is a serious global public health problem that affects every individuals. Elevated blood pressure is a critical medical condition that affects the major organs like brain, kidneys, heart and sometimes causes death. It is one of the unresolved problem in health sectors. The main objective is to identify, classify and analyze blood pressure on real-life data related to blood pressure. Therefore, a multi supervised machine learning techniques like Logistic regression(LR),Naïve Bayes(NB) and Support Vector Machine(SVM) have been implemented to build a proposed model. Initially, data has been collected from individuals and is pre-processed. Then features are extracted and dataset is loaded into the model to identify, classify and analyze blood pressure. Finally, data visualization is done using bar graphs. Hence, the classification and analysis is done on dataset, SVM algorithm fulfilled highest accuracy of 96.8% when compared with the remaining LR,NB algorithms.

Keywords: Blood Pressure(BP), Systolic, Diastolic, Bioinformatics, Machine Learning, Bar graphs, Python, Anaconda Jupyter Notebook, Support Vector Machine(SVM), Naive Bayes(NB) and Logistic Regression(LR).

1. Introduction:

The field of informatics is a combination of software tools accompanied by physical devices which can be used in health sectors for easy disease diagnosis or cure. Informatics is enforced in the health care field especially for the management and use of patient health care information. This is defined as healthcare informatics. It is also termed as Bioinformatics, an evolving science with an increase in electronic health records associated with data analytic systems[1] which plays a major role acting as a multi-disciplinary field used in designing advanced devices or equipment which can be used in accessing biological data which reduces the work load for healthcare providers or experts. Efficient tools can be developed to interpret the results precisely and meaningfully. Bioinformatics will be helpful in disease diagnosis, severity detection, drug designing and treatment based on the data provided by centrally accessible global databases. Blood pressure variability is a major risk factor for heart attacks, strokes, heart failure, kidney disease, peripheral arterial disease and sometimes causes death. In some cases, low -level fluctuations in blood pressure will not affect an individual's health condition. By taking necessary home-remedies like limited amount of salt intake, diet, fruits/vegetables consumption and so on will help in managing blood pressure levels with the range. Much research is going on in this field for early detection of abnormal blood pressure ranges so as to reduce its impact.

Blood Pressure is the force of circulating blood on the walls of the arteries. Most of this pressure results from the heart pumping blood through the circulatory system[2]. Blood pressure uses two measurements: Systolic and Diastolic blood pressure. Systolic blood pressure is the force exerted when the heart is pumping. Diastolic blood pressure is the force exerted when the heart is at rest between beats. Generally, there are three types of blood pressure. They are low, normal and high blood pressure. The normal blood pressure is 120/80mmHg. When the blood pressure is less than 90/60mmHg person is having low BP also known as hypotension. When the blood pressure is greater than 130/90mmHg person is having high BP also known as hypertension[3]. The

major causes for change in blood pressure is stress, white-coat syndrome, medication, exercise, talking, laughter, food and drink, age, gender, marital status, alcohol take, hereditary, lack of sleep and so on. Most healthy individuals have variation in their blood pressure from minute to minute and hour to hour. These fluctuations happen within a normal range. If there is a variation observed in blood pressure levels with respect to either low bp or high bp, certain complications arise. Hence early identification will be helpful in prediction of an individual's health condition.

2. Literature Review:

Melin et al.[4], In this paper proposed the neural network and fuzzy inference system to classify the hypertension type based on the age, risk factors, and behavior of the blood pressure in a period of 24h.The author obtained classification performance as the maximum with proposed method. Singh et al.[5] focused on a new method called rule extraction from the support vector machine to diagnose hypertension in diabetes mellitus patients. And then, author achieved excellent results in the classification of hypertension types in people having diabetes mellitus. In another work, Abdullah et al.[6] modelled a fuzzy expert system (FES) to diagnose hypertension in male and female patients of age groups 10, 20, 30, and 40. The author modeled the hypertension cases for each age group based on the FES model. Das et al.[7] used different modeling techniques including Levenberg–Marquardt (LM), gradient descent (GD), and Bayesian resolution(BR)- based learning functions to model the hypertension types in people of age groups 20 and 40. In Shinde's work[8] author proposed two different approaches for the classification of hypertension types. These methods were the information gain-based feature selection and genetic algorithm-based feature selection for the classification of hypertension types. In this paper, real-life surveyed data related to blood pressure is collected from individuals and gathered data is preprocessed to classify blood pressure using machine learning algorithms like support vector machine(SVM), Naïve Bayes(NB) and Logistic regression(LR) techniques and analyze the comparison of accuracies when a model is built on blood pressure dataset.

3. Proposed Methodology:

Three supervised machine learning techniques are used to build a model for identification, classification and analysis of blood pressure. The techniques used in proposed model is Support Vector Machine(SVM), Naïve Bayes(NB) and Logistic Regression(LR). The gathered data is preprocessed and features are extracted manually. These features are fed as input to proposed model and efficiencies are compared.Finally SVM achieved the best accuracy in predicting the blood pressure.

Dataset:

In this study, real life surveyed database has been used to test the proposed models in the classification of blood pressure types . In the dataset, 14 attributes define the blood pressure types: age, gender, marital status , living area, systolic blood pressure (mmHg), diastolic blood pressure (mmHg), heart rate (bpm),smoking/chewing, diet, hereditary, education, occupational history , fast food addiction. Also, there are three classes: normal (healthy), low, high. Dataset has been collected for 1100 adult subjects aged more than 18.

The increase in blood pressure seen with aging is most likely related to arterial changes. Blood pressure change is more in men than women according to 24-hour ambulatory blood pressure monitoring system[9]. Unmarried individuals have high risk of blood pressure compared with married people. Simultaneously,people who live in polluted areas have risk of blood pressure change than rural living people. Nicotine in cigarette smoke is a big part of problem that raises blood pressure and heart rate[10]. Unhealthy diet and exercise leads to change. Members in the family share their genes, behaviors, lifestyles and environment that can influence their health and their risk for disease. Individuals who are educated have more knowledge on risks of blood pressure when compared to illiterate's[11]. High stress resulted in significant increase in systolic blood pressure[12]. Fast food meal can narrow the arteries and causes raise in blood pressure. Thus, these are some of the causes for fluctuations in blood pressure.

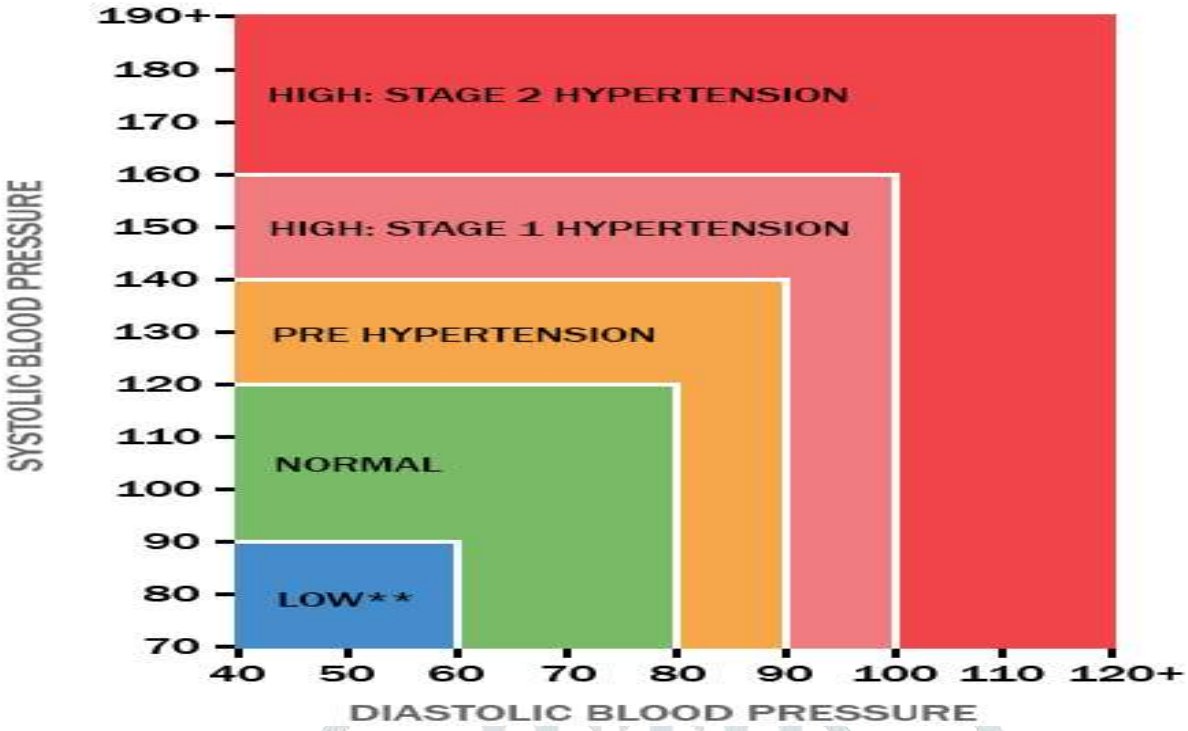

Fig. – 1 : Ranges of blood pressure categories

Next step is to extract features manually and the attributes are described for developing the model. If the number of input variables are reduced then it improves the performance of the model. variable selection, attribute selection etc. are other names for feature selection. Extracted features are given as input to model then data is classified after that visual representation of data using charts, graphs etc. is done for better understand.

Data is visualized using uni variate and multi variate visualization plots. Uni variate plots helps in understanding the position of observations in data variable. Whisker plots are used for Uni variate plotting of data. Whisker plot is also called as box plot is a convenient way of visually displaying the data distribution through their quartiles. Quartiles divide the dataset into quarters where one fourth of the dataset is stored in each quarter. Multi variate plotting is done using histograms. Histogram is an approximate representation for distribution of numerical data.

Data validation is done in which data is partitioned into training and testing data using K-fold cross validation technique. Cross validation is a re-sampling technique used to evaluate machine learning models on limited dataset. 80% of the data is used to train the model and 20% is used to validate the model. Model is validated to obtain accuracy. The complete methodology is described in Fig. 2.

Developing a model is the main step used to predict the accuracies of the algorithms implemented on the Blood Pressure Dataset. The algorithms implemented on blood pressure dataset are discussed below.

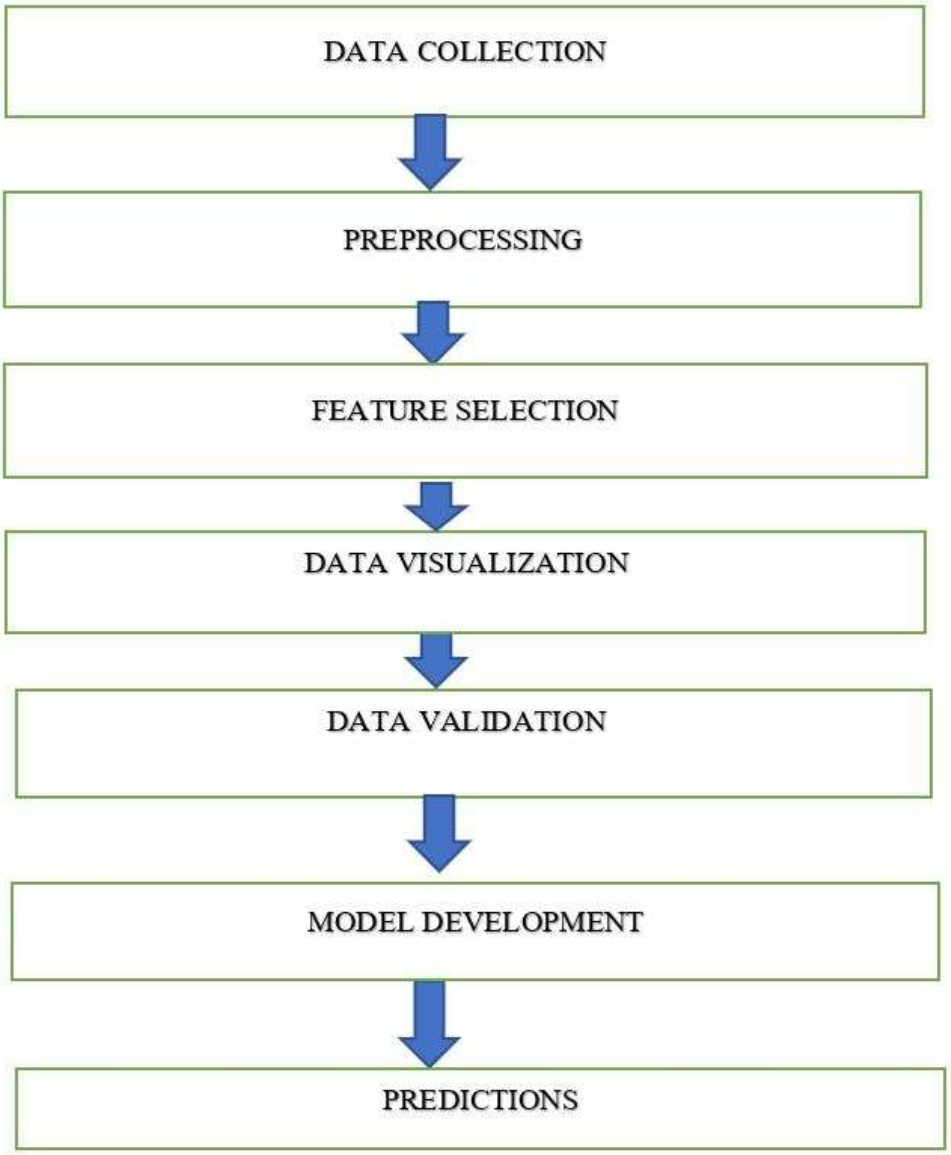

Fig. – 2 : Block diagram of proposed model

Support vector machine is a supervised machine learning algorithm which analyses the data for classification and regression analysis. This performs classification of data by constructing an N-dimensional hyper plane that optimally separates data points into two categories[13]. Based on hyperplanes, separation of data points are classified known as marginal distance. Using kernel functions like linear, sigmoid, rbf allows the SVM model to perform separations in complex boundary scenarios.

Probabilistic classifiers have a sub module called Naïve Bayes classifiers. Therefore, bayes theorem can be applied along with their assumptions with their respective features. These are highly scalable and require more parameters and variables in a learning problem. Algorithm works efficiently for large datasets.

Logistic regression predicts the output of a categorical dependent variable based on given set of independent variables used in classification and predictive analysis using the concept of probability[14]. It defines the relationship between one dependent binary variable and one or more nominal, ordinal ,interval level independent variables which produces discreet output.

4. Results:

The blood pressure dataset is loaded using python library pandas and proposed model is compared with one another to classify the types of blood pressure.

Table – 1 : Comparison of Accuracies of the proposed algorithms

Algorithms implemented	Accuracy obtained
Support Vector Machine(SVM)	96.8 %
Naïve Bayes(NB)	94.7%
Logistic Regression(LR)	83.6%

The accurate results are analyzed by the SVM algorithm. SVM is the popular supervised machine learning algorithm which predicts results efficiently when it is implemented on the BP dataset. It shows the highest accuracy of 96.8 percent as shown in Table 1. SVM algorithm comes under Linear Classifier and is also one amongst the Regression Techniques. Within this SVM algorithm, k-fold cross validation technique is used to achieve the accuracy of 96.8 percent. Similarly for NB algorithm is used giving moderate accuracy Simultaneously, LR Algorithm comes under Probabilistic Classifiers and it is the least used algorithm for predicting the model as it gives least accuracy among taken algorithms. The complete result analysis is featured. The Cost Function trains the SVM. Minimizing J(theta) value provides accurate SVM. In the below function $cost_1$ refers to cost of an example when y=1 and $cost_0$ refers to cost of an example when y=0. In SVM, Kernel functions like linear, sigmoid, rbf and poly are used to determine the cost value.

$$(\theta) = C\sum_{i=1}^{m}\left[y^{(i)}cost_1(\theta^T x^{(i)}) + (1-y^{(i)})cost_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{i=1}^{n}\theta_i^2$$

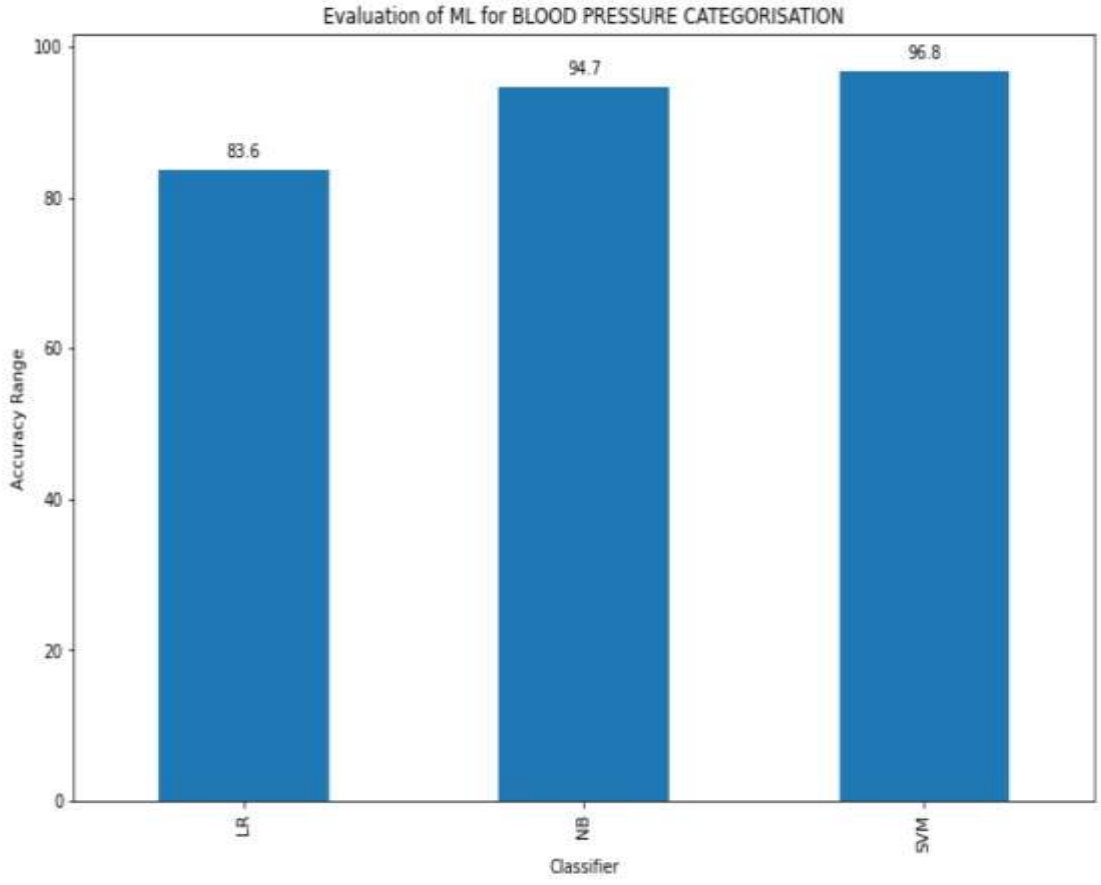

Fig. – 3 : Histogram representation of proposed model.

Visualizing the data can be easily understood to individuals. Histogram is a graphical representation that arranges group of data points in a user-specified intervals. SVM gets the highest accuracy of 96.8% when compared with LR and NB algorithms.

5. Conclusion:

Blood pressure changes time to time which is not detected without proper medical equipment. Now a days machine learning plays a major role in diagnosing diseases in healthcare industry. The main aim is to identify, classify and compare blood pressure using machine learning techniques. Three supervised machine learning techniques SVM, LR and NB are used to build a model. Data has been collected from individuals and is preprocessed. Features are extracted and fed as input to model. On comparing the results of all the three techniques it is identified that SVM attains highest accuracy i.e., 96.8% than others . Finally, data is visualized by using histograms and model efficiency is performed by Kernel functions like linear, sigmoid, rbf and poly. In future, this model can be furtherly implemented with more personal features on large datasets for better accuracy.

6. References:

[1]. A. A. Abdullah, Z. Zakaria, and N. F. Mohammad, "Design and development of fuzzy expert systems for diagnosis of hypertension," in Proceedings of the 2nd International Conference on Intelligent Systems, Modelling and Simulation (ISMS '11), vol. 10, pp. 131–141, IEEE, Kuala Lumpur, Malaysia, January 2011.

[2]. R. Nohria and P. S. Mann, "Diagnosis of Hypertension using Adaptive Neuro-Fuzzy Inference System," IJCST, vol. 6, pp. 2229–4333, 2015.

[3]. R. Yafawi, M. E. Knauft, K. Stokem, J. M. Palminteri, and M. WirthJ, "Pulmonary arterial hypertension," Encyclopedia of Cardiovascular Research and Medecine, pp. 181–194, 2018.

[4]. J. Mantas, J. Abellan, "Analysis and extension of the decision trees based on the imprecise probabilities application on noisy data, " Expert Systems with Applications, Vol. 41, no. 5, pp. 2514–2525, 2014.

[5]. H. Zhang and F. C. Lin, " Medical Diagnosis by the Virtual Physician, " IEEE Xplore Computer Based Medical System, pp. 296–302, 1999.

[6]. Chow CK, Teo KK, Rangarajan S, et al."Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high , middle , and low income",JAMA. 2013; 310(9):959 –968.

[7]. V. G. Vilkov, R. G. Oganov, and S. A. Shal'nova, "Comparative accuracy of neural network models for diagnosing latent arterial hypertension on the basis of data on daily blood pressure monitoring," Human Physiology, vol. 32, no. 6, pp. 657–661, 2006.

[8]. C. Cortes and V. Vapnik, "Support vector network," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[9]. J. A. Whitworth, "World Health Organization (WHO)/International Society of Hypertension (ISH) stated on management of hypertension," Journal of Hypertension, vol. 21, no. 11, pp. 1983–1992, 2003.

[10]. G. Beevers, G. Y. H. Lip, and E. O'brien, "Blood pressure measurement: Part I—Sphygmomanometry: Factors common to all techniques," BMJ, vol. 322, no. 7292, p. 981, 2001.

[11]. Ashok kumar , Vijaysinh Solanki, " Machine Learning Techniques Using WEKA classification for blood pressure risks " , (IJCSIT) International Journal of Computer Science and Information Technologies, Vol.34(12) ,2014, pp. 5857-5860.

[12]. P. Melin, I. Miramontes, and G. Prado-Arechiga, "A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis," Expert Systems with Applications, vol. 107, pp. 146–164, 2018.

[13]. Cheong A, Sazlina S-G, Tong S, Azah AS, Salmiah S. Poor blood pressure control and their associated factors among older people with hypertension: a cross-sectional study in six public primary care clinics in Malaysia. Malays Fam Physician. 2015;10(1):19.

[14]. Czernichow S, Zanchetti A , Turnbull F, Barzi F, Ninomiya T, Kengne AP et al. The effects of blood pressure reduction and of different blood pressure-lowering regimens on major cardiovascular events according to baseline blood pressure:meta-analysis of randomize trials. Journal of Hypertension 2011 29(1):4–16