Research Article

A Novel Machine Learning-Based Systolic Blood Pressure Predicting Model

Jiao Zheng 1 and Zhengyu Yu 2

1 Drug Clinical Trial Institution Department, Hunan Provincial People’s Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha, Hunan 410000, China
2 Faculty of Engineering and IT, University of Technology, Sydney, Sydney, NSW 2007, Australia

Correspondence should be addressed to Zhengyu Yu; zhengyu.yu@uts.edu.au

Received 26 March 2021; Accepted 22 May 2021; Published 8 June 2021

Academic Editor: Hassan Karimi-Maleh

Copyright © 2021 Jiao Zheng and Zhengyu Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Blood pressure (BP) is a vital biomedical feature for diagnosing hypertension and cardiovascular diseases. Traditionally, it is measured by cuff-based equipment, e.g., sphygmomanometer; the measurement is discontinuous and uncomfortable. A cuff-less method based on different signals, electrocardiogram (ECG) and photoplethysmography (PPG), is proposed recently. However, this method is costly and inconvenient due to the collections of multisensors. In this paper, a novel machine learning-based systolic blood pressure (SBP) predicting model is proposed. The model was evaluated by clinical and lifestyle features (gender, marital status, smoking status, age, weight, etc.). Different machine learning algorithms and different percentage of training, validation, and testing were evaluated to optimize the model accuracy. Results were validated to increase the accuracy and robustness of the model. The performance of our model met both the level of grade A (British Hypertension Society (BHS) standard) and the American National Standard from the Association for the Advancement of Medical Instrumentation (AAMI) for SBP estimation.

1. Introduction

Currently, hypertension or high blood pressure (BP) is one of the riskiest factors that affect cerebrovascular (CVDs) and cardiovascular diseases and causes around 31% of death in the world [1]. World Health Organization (WHO) reported 9.4 million death from hypertension in the world health statistics in 2014 [2]. After diabetes, hypertension is known as the second dangerous factor of cardiovascular disease [3]. Because many people unrealize the effect of hypertension and do not control it, it is also named silent killer. BP is one of the essential periodic features providing valuable medical information to diagnostic cardiovascular diseases. Diastolic blood pressure (DBP) is the lower bound of the BP. Systolic blood pressure (SBP) is the upper bound of the BP. The average BP is called mean arterial pressure (MAP) in a cardiac cycle [4]. Hypertension occurs and affects the internal body organs when the SBP is higher than 140 mmHg or the DSP is higher than 90 mmHg [5]. The typical value of MAP should stay between 70 mmHg and 110 mmHg [6]. Hypertension patients test BP occasionally. Incorrect measurements are recorded with high possibility due to the BP varies by eating habits, tobacco, stress, etc. Therefore, to achieve an accurate diagnosis, continuous BP monitoring is essential. Besides, the continuous BP record improves the prescriptions of appropriate medicine and treatment by doctors.

The most common and accurate BP measurement methods are cuff-based or invasive [7]. However, the equipment is limited to the health care centre and hospitals. The most common devices used for BP measurement are based on auscultatory and oscillometric methods, which determine the values of SBP and DBP with no risk and pain. However, the measurements from those devices are cuff-based and discontinuous. It is inconvenient to cuff inflation and deflation repeat for the patients. The development of cuff-less BP measurement has been proposed in the last decades [8].
transit time (PPT) has been introduced to estimate the cuffless continuous BP [9]. Also, continuous wave radar (CWR), bioimpedance (Blmp), and electrocardiogram (ECG) sensor are processed in BP estimation [10]. The most common estimation method is based on photoplethysmography (PPG) [7]. The pulse arrival time (PAT) can be calculated by PPG signal and ECG signal. Several challenges were applied in these methods, such as implantation of arterial wave propagation models, signal calibration, and various parameters from difference signals [11]. Recently, machine learning-based measurement methods are developed to reduce the calibration processes by using the signals of PPG and ECG [12]. BP is measured by a series of parameter of features extracted from the signals of PPG and ECG by machine learning. A single PPG signal-based measurement method is also proposed due to its simplicity. This approach can generate estimated BP continuously [13]. This paper presents a novel machine learning-based SBP predicted method by several features and SBP values.

2. Materials and Methods

A novel modelling method is proposed to predict the estimation of SBP in this paper. The proposed method is designed by the extraction of clinical measurement and lifestyle variables with machine learning techniques. Figure 1 illustrates the workflow of the SBP predicted method, which is summarised below.

1. Extract Data from Datasets. In this method, only 250 samples with low BP were implemented.

2. Extract Features. There were 501 features, including one target feature, SBP, 17 clinical features, and 483 genetic markers. This method initially covered 13 features. However, to increase the performance accuracy, part of the features was unselected when the model was evaluated.

3. Algorithm Comparison. Machine learning methods, such as linear regression (LR), support vector machine (SVM), decision tree regression (DTR), Gaussian process regression (GPR), and artificial neural network (ANN), had been computed to address the best approach for the model. The result indicated that ANN was the best accurate method.

4. In ANN, three different stages were designed to optimize the performance of the network

   (a) Three training algorithms, the Levenberg-Marquardt Algorithm (LMA), the Bayesian Regularization Algorithm (BRA), and the Scaled Conjugate Gradient Algorithm (SCGA), had been selected in the network

   (b) Various percentage of training, validation, and testing had been compared

   (c) The different numbers of hidden neurons had been adjusted

5. When the model was evaluated, kept the training algorithm, percentage of training, validation, and testing, and the number of hidden neurons. In order to validate the results by different features, the step of feature extraction was operated again by unselecting one of the features for each testing and validation. The results indicated that all 12 features were suitable for the model

6. After the final model was evaluated, the result was predictable with minimized errors

2.1. Datasets. The datasets we used in our research work are from Dr. Raymond Lam, GlaxoSmithKline, which includes eighteen feature variables and 500 subjects [14]. In this research, 250 subjects indicated as high blood pressure (hypertension) are higher than 140 mmHg, and 250 subjects are lower than 140 mmHg. Eighteen feature variables contain one response feature (SBP) and seventeen features (clinical covariates). Due to some variables of features were ambiguous classified and calculated, our research selected thirteen variables for training, testing, and estimating.

2.2. Systolic Blood Pressure. SBP represents BP value which is exerting against the walls of the artery when the heart is beating. SBP is an essential value for the BP measurement, and it is an important feature to detect hypertension [4]

2.3. Gender. In the datasets, the variable of gender is in binary. M denotes male, and F denotes female. The differences in gender indicate the difference in the regulation of BP [15]. A study of ambulatory BP monitoring for 24 hours has been presented recently [16]. The results have addressed that men have more possibility to have cardiovascular disease than age-matched women [17]. The BP in men is higher than the BP in women [18].

2.4. Marital Status. Marital status is one of the variables from the datasets. Y denotes married, and N denotes not married. According to the previous research, it is shown that never-married people have more risks of hypertension than married people in men. However, never-married women are associated with less risk of hypertension than married women [19].

2.5. Smoking Status. In the smoking status, Y denotes smoker, and N denotes nonsmoker. Research study indicates that the artery walls become sticky when the inhaled cigarette chemicals are absorbed into the bloodstream. The number of fatty plaques sticks to the artery walls, called atherosclerosis leading to cardiovascular disease. When the artery walls become narrower and narrower, the blood travels through the arteries difficulty [20].

2.6. Age. Age plays a vital role in relation with BP [21–23]. When the age increases, the blood vessels become stiffer, which can lead to a rise in BP as well as an increase in the risk of hypertension. People aged 50 or higher are the most prevalent hypertensions group, especially in isolated systolic hypertension. Due to the change of artery structure
associated with age increases, e.g., artery stiffness, the BP increases cardiovascular risk [24, 25].

2.7. Weight/Overweight. Weight is a continuous variable in the dataset; the unit is pound. The BP rises following by the bodyweight increase. Overweight will increase the possibility to develop hypertension [26]. In the datasets, three categories are classified. 1 denotes normal, 2 denotes overweight, and 3 denotes obese.

2.8. Height. A study proposed an inverse linear relationship between SBP and height. Lower SBP was associated with greater height [27]. In the datasets, height is a continuous variable with the unit (inches).

2.9. Body Mass Index (BMI). BMI is a measurement index indicating whether the body is obese or overweight [28]. The categories have been classified as follows:

(1) Underweight. Less than 20
(2) Normal Weight. 20-25
(3) Overweight. 26-30
(4) Obese. 30-above

It is calculated by weight and height.

\[
\text{Weight} \div \text{Height}^2 \times 703. 
\]  

2.10. Exercise Level. According to the research, exercise accelerates the heart pump. The faster the heart pumps, the higher SBP rises. During the exercise, the expected level of SBP is between 160 mmHg and 220 mmHg [29]. In this paper, the exercise level has been divided into three. 1, 2, and 3 denote low, medium, and high, respectively.

2.11. Alcohol Consumption. A recent clinical study has suggested that alcohol consumption can raise BP rapidly [30]. A single drink of alcohol affects an acute BP rise for 2 hours [31]. Moreover, a sustained BP rise can be caused by alcohol consumption for a few days. Long-time alcohol consumption links to risk factors, such as cardiovascular disease and high blood pressure [32]. In datasets, the alcohol consumption level is defined by 1, 2, and 3, which denote low, medium, and high, respectively.

2.12. Stress Level. The body generates a surge of hormones when they are in stressful situations. This action causes the heart to pump faster, and the blood vessels become narrow, which leads to spike in BP temporarily. The research finds that reducing stress can lower body BP [33]. We also have three levels of stress in datasets: 1 denotes low, 2 denotes medium, and 3 denotes high.

2.13. Salt (NaCl) Intake Level. High sodium or salt intake can contribute to high BP and speed up the risk of cardiovascular disease [34]. WHO reports that 5 grams or less salt intake for adults helps decrease BP and the risk of heart disease [35]. However, most people take 9-12 grams daily on average [36, 37]. This amount of salt consumption is twice the recommended maximum level. The member states of WHO agree to reduce world population’s salt consumption by 30% before 2025, which prevents 2.5 million deaths caused by high sodium consumption [38].
From the datasets, due to several variables of features are ambiguous, only 13 variables with 250 nonhypertension samples had been selected in the model. In order to predict the most accurate estimated SBP, an algorithm comparison had been designed. Algorithms such as LR, SVM, GPR, and ANN had been processed to generate the most optimized result. The outcomes from the algorithms had been recorded in Table 1.

After computing five machine learning methods, as the Table 1 shown, the ANN method had the best performance; the mean average error (MAE) was about 10.78 mmHg, which was lower than other methods. Therefore, the ANN was the primary method for our research to predict the SBP.

2.14. Training Algorithm

2.14.1. Levenberg-Marquardt. The LMA provides the solution to minimize the nonlinear least-squares problem [39].

It is one of the most popular algorithms for optimization [40]. In several kinds of problems, LMA generates better results than gradient descent and other conjugate gradient techniques [41]. It is a blended method of Gauss-Newton and vanilla gradient descent iteration [42]. If the solution is far from the true result, the algorithm acts like a gradient descent method: slow, ensures the converge. If the current solution is close to the true result, it acts like a Gauss-Newton method [43].

2.14.2. Bayesian Regularization. Bayesian regularization has also been named Bayesian regularized artiﬁcial neural networks (BRANNs), which is more reliable than the classical ANN backpropagation nets, with no need for prolix cross-validation. It is a machine learning algorithm that converts a nonlinear regression into a ridge regression [44].

2.14.3. Scaled Conjugate Gradient. As a supervised learning algorithm for feedforward neural network, SCG is one of the conjugate gradient algorithms. The operation of SCG is smoother and faster than the standard backpropagation nets. The training algorithm from SCG has benchmarked performance against the classical BP algorithm. BP utilizes the optimization theory of gradient descent with the selected variable from the user. The offline trained network uses a fixed variable, whereas the SCG algorithm uses these variables as the
second-order approximation. Less learning iterations occur to accelerate the learning process.

The ANN method in this paper had been separated into three different training methods, LMA, BRA, and the SCGA. We initially set the number of hidden neurons as 10, 80% data for training, 10% data for validation, and 10% data for testing to address the best training algorithm. The results had been illustrated in Table 2:

From Table 2, the results indicated that the BRA training method was the most accurate method.

In the next step, we adjusted the percentage of the training data and compared the result in Table 3.

According to Table 3, the most accurate result was generated by 90%, 5%, and 5% for training, validation, and testing, respectively.

Figure 2 presents the structure of ANN, the input layer contained 12 features, and the output layer contained 1 feature which was SBP. The training performance was evaluated by changing the number of hidden neurons on the hidden layer. In our model, as displayed in Figure 2, only one hidden layer applied. Initially, the hidden neurons had been set up as 10, and the different numbers of hidden neurons had been trained to determine the most accurate network structure.

As Table 4 presented, the best accurate model was with 15 hidden neurons. Therefore, the ANN applied the BRA training method with 90% training, 5% validating, and 5% testing, with 15 hidden neurons generated the most accurate results.

3. Results and Discussion

In Results and Discussion, as mentioned, several features were ambiguous. It is possible to generate more errors in the system. Several tests had been completed to compare the performance in order to avoid these ambiguous features. We validated that all the features used in the model were suitable to generate the result. The result validation process is presented as follows.

In Table 5, the above 12 features had been trained, validated, and tested in the ANN with the target feature, SBP. To validate the results, we unselected one of the features for each training. For instance, in the first training, the feature ID 1, gender, was excluded in the 12 features; therefore, only 11 features were operated in the first training. For each training, different error levels (from 5 mmHg to 30 mmHg) had been recorded in Table 6.

From Tables 6, 12 different features had been unselected one by one. The higher the error percentage was, the higher the accuracy was. When we unselected feature ID 3, smoking status, the results (26%, 58%, and 78%) were lower than any other results in the table, which meant the values of feature ID 3 were more significant than other values of features. However, if we unselected feature ID 2, marital status, the results (57%, 85%, and 95%) were higher than any other results in the table. Therefore, the values of feature ID 2 were less vital than any other values of features in the table. From Table 7, the error percentage after training 12 features were 67%, 89%, and 97% for error less than 5 mmHg, 10 mmHg, and 15 mmHg, respectively, whose accuracy was higher than any others in Table 6. Therefore, it is unnecessarily to unselect any feature in the training process, and all the 12 features should be trained.

Figure 3 presents the error histogram with 20 bins; the orange line indicates zero error, the blue bar represents training errors, and the test errors are red. The figure illustrated that the most significant number of errors are near the orange line, zero error; the number of errors decreased when the error became larger.

The results from Table 7 illustrated that our prediction model had 67%, 89%, and 97% accuracy when an error was less than 5 mmHg, 10 mmHg, and 15 mmHg, respectively. Comparing to the British Hypertension Society (BHS) standard [45], the performance of our model met the level of grade A (BHS standard) for SBP estimation.

Moreover, Table 8 illustrated the American National Standard from the Association for the Advancement of Medical Instrumentation (AAMI). According to the standard, the
BP estimation method for noninvasive should less than 5 mmHg and 8 mmHg for MAE and STD. Our results were 3.03 mmHg (MAE) and 6.11 mmHg (STD), which met the AAMI standard [46].

4. Conclusions

In this paper, we proposed a novel model based on machine learning algorithm to predict SBP. The model included three stages, which were input, calibration process, and output.

In the first stage, datasets associated with clinical and lifestyle features were selected and extracted. Values of 13 features, including SBP values, had been selected as training data. Five machine learning algorithms, such as LR, SVM, DTR, GPR, and ANN, had been compared in this stage. The result indicated that ANN had the best accuracy.

The calibration process was the second stage. Three different training algorithms, LMA, BRA, and SCGA, had been trained in ANN. After comparing with the value of MAE and STD, BRA was addressed as the best training algorithms. In the next step, the comparison of different percentage of data training, validation, and testing had been completed. The most accurate result was generated by 90%, 5%, and 5% for training, validation, and testing, respectively. The next step was to adjust the structure of ANN; we assumed that the ANN included one hidden layer, which initially contained 10 hidden neurons. Several different hidden neurons were applied to discover that the ANN generated the most accurate results in the condition of 15 hidden neurons.

In the third stage, the model was evaluated by amending inputted data of features. To validate the results, values of feature with significant uncertainties were unselected in the model. The accumulated error percentage of the evaluated model meets grade A in BHS standard and AAMI standard in estimating SBP. Therefore, our proposed predicted model was accurate and reliable. Nevertheless, the model would be optimized further if the data of the features were more robust.

Data Availability

The data used are from R. Lam, “Blood Pressure.” These are available at http://www.math.yorku.ca/Who/Faculty/Ng/ssc2003/BPMain.htm (accessed March 03, 2021).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] D. Mozaffarian, E. J. Benjamin, A. S. Go et al., “Heart disease and stroke statistics—2015 update: a report from the American Heart Association,” Circulation, vol. 131, no. 4, pp. e29–e322, 2015.

[2] W. H. Organization, World Health Statistics 2015, World Health Organization, 2015.

[3] C. Höchst, “Blood pressure variability: prognostic value and therapeutic implications,” ISRN Hypertension, vol. 2013, Article ID 398485, 16 pages, 2013.

[4] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa, “Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration,” Medical and Biological Engineering and Computing, vol. 38, no. 5, pp. 569–574, 2000.

[5] M. Korsager Larsen and V. V. Matchkov, “Hypertension and physical exercise: the role of oxidative stress,” Medicina, vol. 52, no. 1, pp. 19–27, 2016.
...
[38] W. H. Organization, *Reducing salt intake in populations: report of a WHO forum and technical meeting, 5-7 October 2006, Paris, France.*

[39] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[40] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, 1999.

[41] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for Non-linear Least Squares Problems*, 2004.

[42] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Science & Business Media, 2006.

[43] W. H. Press, H. William, S. A. Teukolsky, A. Saul, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge university press, 2007.

[44] F. Burden and D. Winkler, "Bayesian regularization of neural networks," *Artificial neural networks*, pp. 23–42, 2008.

[45] E. O’Brien, B. Waebier, G. Parati, J. Staessen, and M. G. Myers, "Blood pressure measuring devices: recommendations of the European Society of Hypertension," *BMJ*, vol. 322, no. 7285, pp. 531–536, 2001.

[46] A. f. t. A. o. M. Instrumentation, "American national standards for electronic or automated sphygmomanometers," 1987, ANSI/AAMI SP 10-1987.