# A Risk Factors Screening Method in the Context-aware System of Hypertension

Duoyi Xie[1,2], Guixia Kang[1,2] and Longfeng Chen[1,2]
[1]*Key Laboratory of Universal Wireless Communications, Ministry of Education,*
*Beijing University of Posts and Telecommunications, No.10 Xitucheng Road, Beijing, China*
[2]*Wuxi BUPT Sensory Technology and Industry Institute CO.LTD, Wuxi, China*

Keywords: Hypertension, Context-aware System, Feature Selection.

Abstract: Hypertension has become a health problem that seriously endangers human life and is the leading cause of cardiovascular disease. Many patients do not know exactly whether their blood pressure is well controlled or not, which makes their conditions worse. A context-aware intelligent system can help patients to analyse their control situation of blood pressure (BP) and provide feedback. It is especially important to determine whether the risk-factors input in the context-aware system of hypertension is appropriate. The choice of risk factors will affect the classification performance and accuracy of the system. The risk factors screening method for hypertension proposed in this paper combined the random forest algorithm and stability selection (RFSS). It can remove the redundant context information, and leave the key factors of BP control situation. Experimental results showed that the prediction accuracy achieved more than 77% prediction accuracy, and dimension of risk factors reduced by 59%. The results indicated that RFSS is an effective method in the screening of risk factors and the prediction of hypertension.

## 1 INTRODUCTION

Hypertension is a leading cause of death and disability-adjusted life-years worldwide (Lim et al., 2012). There are 270 million hypertensive patients in China. Annual patient increases by more than 10 million. More than 1 million people die from high blood pressure every year in China. Three-quarters of the survived patients were disabled due to hypertension. However, the awareness rate and control rate of hypertensive population are only 36% and 28%, respectively (Chen et al., 2014). Blood pressure of 120/80 mm Hg or higher is linearly related to risk for fatal and nonfatal stroke, ischemic heart disease, and noncardiac vascular disease, and each increase of 20/10 mm Hg doubles the risk for a fatal cardiovascular disease event (Carey and Whelton, 2012). Evidence shows that hypertension can be prevented and controlled through monitoring and treatment (Whitworth and Chalmers, 2004).

Wireless eHealth (WeHealth) has developed rapidly in China in recent years, and remote hypertension monitoring is one of its most important applications (Kang and Zhang, 2010). The hypertension monitoring system can increase the awareness rate and control rate of hypertensive population by feeding back their situation of blood pressure (Sandi et al., 2013). The management of hypertensive patients is mainly based on the blood pressure levels. However, the threshold of hypertension will be different due to the differences in patient function. Decisions on the management of hypertensive patients should not only take blood pressure levels into account, but also the other cardiovascular risk factors (Whitworth and Chalmers, 2004). Therefore, in addition to blood pressure, comprehensive analysis of other risk factors in the context information can help people better recognize their blood pressure condition. The formation of hypertension is related to a variety of factors, including age, lifestyle, family history, etc. (Hong-Tao et al., 2007; Kunes and Zicha, 2009). The context information obtained by the monitoring system contains risk factors and redundant information that affects the classification accuracy. Therefore, we need to screen the risk factors of hypertension in the context information to improve the classification effect of the monitoring system. In this way, the patients who take medicine can be properly guided to continue medication to prevent

Xie, D., Kang, G. and Chen, L.
A Risk Factors Screening Method in the Context-aware System of Hypertension.
DOI: 10.5220/0007672500490056
In *Proceedings of the 4th International Conference on Internet of Things, Big Data and Security (IoTBDS 2019)*, pages 49-56
ISBN: 978-989-758-369-8

the deterioration of the disease. At the same time, the study can assist doctors to diagnose patients' blood pressure condition individually.

In previous studies, feature selection algorithms were often used to screen for risk factors. Feature selection methods are divided into selection mechanisms: Filter, Wrapper, and Embedded. Filter relies on the data samples themselves, with a low time complexity. Ding and Peng (2005) proposed mRMR (minimum-redundancy maximum-relevancy) feature selection method. The algorithm considers not only the correlation of each feature attribute with respect to the class label, but also the redundancy of the internal relationship of the feature attributes set. Wrapper algorithms select features by using learning algorithm results, with a better result and a higher time complexity (Kohavi and John, 1997). Hu and Bao (2015) proposed a wrapper feature selection algorithm for short-term load forecasting (STLF) data, which has excellent selection effect. The embedded algorithm takes feature attribute selection as part of the training process and has balanced efficiency and accuracy. In recent years, it has become a research hotspot. Koenigstein and Paqued (2013) proposed an embedded algorithm based on Bayes matrix factorization. The classification model can automatically identify and use informational features, and useless features can be subtracted.

For a single method, the screening rate and accuracy are often not balanced. In this paper, we propose a context-aware system to assess the treatment effect in the intelligent monitoring of hypertensive patients. For the decision making module of the system, we propose a filter-embedded feature selection method to screen risk factors of hypertension. By using the support vector machine (SVM) classification algorithm to verify the screening effect, this method can improve the screening rate and ensure the accuracy of this system. The entire system therefore improves performance, which can service hypertensive patients preferably.

## 2 SYSTEM ARCHITECTURE

To make a comprehensive evaluation of hypertensive patients' treatment effect, blood pressure data should not be the only reference. The proposed monitoring system is context-aware and takes into account other risk factors. Risk factors screening method and data mining techniques are utilized in classifying the treatment effect. The overall architecture is shown in Figure 1. It consists of three modules: data acquisition (DA), decision making (DM), and diagnostic feedback (DF).

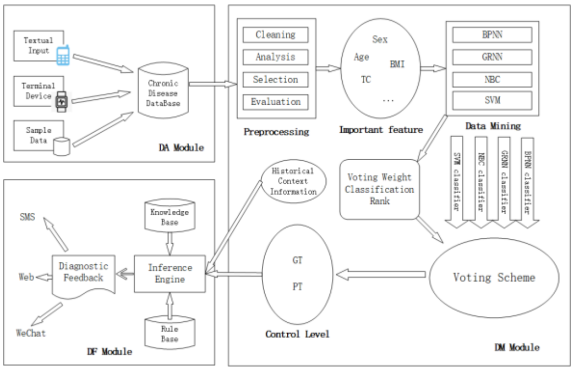

Figure 1: The proposed system architecture.

### 2.1 Data Acquisition Module

DA module gathers and stores patients' medical information. The way to obtain the context is divided into measurement data acquisition, text data acquisition and sample data acquisition according to different acquisition paths and uses. Measurement data is collected by hardware sensors. Text data is obtained by textual input. Sample data is stored in medical databases.

### 2.2 Decision Making Module

This is the core module of the entire system. It determines the level of patient treatment effectiveness. In the intelligent monitoring of hypertension, it is necessary to evaluate the level of blood pressure control status of hypertensive patients, and make corresponding diagnostic feedback according to the level. Data drawn from the DA module must be preprocessed first. The data preprocessing process is the core step and the focus of this paper. It includes data cleaning, analysis, feature selection, and evaluation. The specific method will be detailed in Section 3. Then the preprocessed data is put into the data mining algorithms for training and classification to judge the condition. By analyzing the patient's blood pressure, age, weight and other contextual information, the patients are divided into two groups by professional doctors: Good Treatment (GT) and Poor Treatment (PT). Finally the hypertensive patients are classified into these two groups by the system based on their risk factors. The accuracy of the DA module is verified by classification results.

### 2.3 Diagnostic Feedback Module

DF module is a small inference system to generate the feedback according to DM output and patient's relevant information. Knowledge Base contains all the knowledge needed for the proper treatment of hypertensive patients, varying from drug therapy to non-drug therapy (mainly lifestyle modifications). Rule Base is a set of rules regarding medical diagnosis. The inference engine gives an automatic feedback to the input treatment effect based on the Knowledge Base and Rule Base. The feedback is then returned to the patient, either by WeChat, Email, or on the web.

## 3 SCREENING METHODS

The DM module extracts the hypertension risk factors from the context information acquired from the DA module and inputs them into the classifier to obtain the patient's hypertension control condition. There is a large amount of multi-dimensional redundant information in the context information. Redundant information will decrease the speed of the system and interferes with normal decision making. Therefore, it is particularly important to filter risk factors from multi-dimensional context information correctly and effectively. Accurate extraction of features can increase the accuracy of the classifier and reduce data dimensions. After each step of screening, we use the same classification algorithm to evaluate the screening results to test the effect of the methods.

### 3.1 Data Preprocessing

This study screened 8619 people who met the criteria as the data set from Beijing We-Health Platform. The data included carcinoembryonic antigen (CEA), sex, age, height, weight, white blood cells, red blood cells, haemoglobin (HGB), red blood cell specific volume (HCT), erythrocyte mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), red cell distribution width coefficient variation (RDWCV), red cell distribution width standard deviation (RDWSD), platelet count (PLT), mean platelet volume (MPV), platelet distribution width (PDW), monocyte (MON), MON%, Granulocytes (GRA), GRA%, lymphocyte (LYM), LYM%, alanine aminotransferase (ALT), UREA, serum creatinine (Cr), uric acid (UA), serum total cholesterol (TC), triglyceride (TG), fasting plasma glucose (FPG), specific gravity (SG), PH, a total of 32 dimensions. Body mass index (BMI) can accurately reflect the combined indicators of height and weight, so we combined these two features into BMI. These data were labelled to indicate the type, GT and PT. According to statistics, there were about 47% of GT data and 53% of PT data. The data contained diastolic pressure and systolic pressure, but we did not consider them to ensure the objectivity of the risk factor screening process. We randomly divided the data set into 70% training set and 30% test set. In the following steps, the models were trained by 10-fold cross validation. A total of 2586 data were used for the tests. To guarantee the same ratio of the two types of data, there were 1372 cases of GT data and 1214 cases of PT data.

### 3.2 Filter by PCC

The continuous variable correlation measure based on Pearson's correlation coefficient (PCC) significance test can quickly test whether there is linear correlation between data. The source of context information used in this paper was the physical examination information. The evaluation index was mostly linear, so we used the PCC for preliminary screening. For the given sample points, we can calculate the PCC. This method can screen out features that have little correlation with blood pressure.

The calculation formula is as follows:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2} \qquad (1)$$

### 3.3 Random Forest Screening

Random forest algorithm has many advantages, including high accuracy, robustness and ease of use, making it one of the most popular machine learning algorithms (Madan et al., 2008). It randomly selects n samples from the sample set and K attributes from all attributes. After that, it selects the best segmentation attribute node to establish the CART decision tree. The above steps are repeated m times to establish m decision trees and obtain m classifiers (Breiman, 2001).

However, this kind of impurity-based screening method has a bias. Once a feature is selected, the importance of other features will drop sharply because the impurity has been lowered by the selected feature. So other features are difficult to reduce so much impurity. Therefore, the feature selected at first gets the high score while the score of other related features gets low score, which is easy to cause misunderstanding.

### 3.4 Random-forest Stability Selection

Stability selection is a method based on a combination of subsampling and selection algorithms. This method provides finite sample control for some error rates of false discoveries and hence a transparent principle to choose a proper amount of regularization for structure estimation (Nicolai and Peter, 2010).

We improved the screening method for the instability of feature scores in random forest algorithms. Firstly, the filter method was used. Then we combined the random forest algorithm and the subsampling to an improved screening method: random-forest stability selection (RFSS). Its main idea is to run random forest algorithm on different data subsets and feature subsets, repeat constantly, and finally aggregate feature selection results. We used the frequency (the number of times selected as an important feature divided by the number of times its subset was tested) at which a feature was considered to be an important feature as an evaluation indicator. Ideally, important features would score close to 100%. A slightly weaker feature score would be a non-zero number, while the most useless feature score would be close to zero.

### 3.5 Evaluation Algorithm Selection

For different screening methods, we used a unified classification algorithm to verify the screening effect. The support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000) has been introduced as an efficient technique for solving various function estimation problems, especially for the pattern classification problems (Vapnik et al., 1997). SVM is among the most robust and accurate methods in all well-known data mining algorithms. The final decision function of the SVM is determined by only a few support vectors, and the computational complexity depends on the number of support vectors. SVM can verify whether we have selected the most appropriate risk factors and avoids the "curse of dimensionality". Therefore, we used SVM to verify the screening effect.

## 4 EXPERIMENTAL RESULTS

We verified the accuracy of each screening method by SVM algorithm, and judged the screening effect of risk factors through multiple evaluation indicators including sensitivity, specificity, accuracy and dimension.

### 4.1 Performance Analysis

Table 1 shows the classification result of raw data without using a screening method through 10-fold cross-validation. 650 of the 1214 PT patients were predicted to be accurate and the sensitivity is 54%. The accuracy of this model is 66%. The risk factor dimension is 32.

Table 1: Classification result without using a screening method.

| | | Actual | | |
|---|---|---|---|---|
| | | GT | PT | Total |
| Predicted | GT | 1068 | 564 | |
| | PT | 304 | 650 | |
| | Total | 1372 | 1214 | 2586 |

Then we used the PCCs to filter features. We can see the result in figure 2. We sorted 32 features based on scores by PCCs. The feature with 0 value indicates that the probability of a two-tailed test is greater than 0.05, indicating that they cannot be used as a risk factor. Meanwhile, the correlation values of PLT, LYM, LYM%, MCV, MPV, and MON for blood pressure were less than 0.05. This means that these parameters, relative to other parameters, do not directly reflect or affect a patient's blood pressure condition. So we screened these parameters out because they were basically not helpful for the classification models.

Table 2 shows the result using the filter screening method through 10-fold cross-validation. The feature dimension reduced from 32 to 23, and the data indicated that the sensitivity and accuracy of the classification had increased significantly. 896 patients were predicted to be accurate and the sensitivity is 73%. The accuracy of this model raised to 75%. Dimension was 28% lower than before. However, the number of risk factors was still too large. Therefore, we used random forest screening methods to reduce feature dimension and further determine the risk factors on this basis.

Table 2: Classification result with using PCC.

| | | Actual | | |
|---|---|---|---|---|
| | | GT | PT | Total |
| Predicted | GT | 1073 | 318 | |
| | PT | 299 | 896 | |
| | Total | 1372 | 1214 | 2586 |

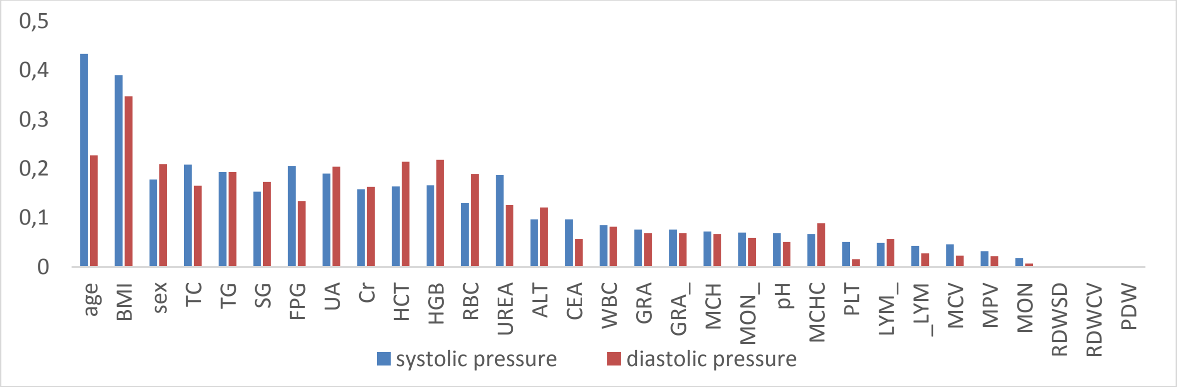

Figure 2: The scores by using PCC.

Figure 3 shows the scores of the features after using the random forest method. As mentioned in Section 3, once a feature was identified in the random forest screening process, the scores of other associated features may drop suddenly, leading to filter out some useful features. If we chose features with high score by the result, the sensitivity and accuracy would decrease in the classification process.

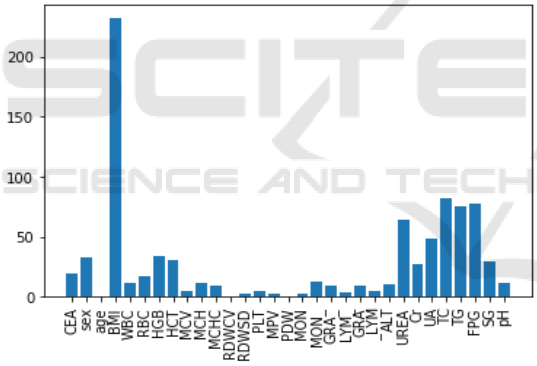

Figure 3: The scores by using random forest method.

Table 3 shows the classification result using random forest method after filter method with 10-fold cross-validation. The risk factors were further reduced to 13. The dimension was 59% lower than the original. But only 742 PT patients were predicted to be accurate. The sensitivity and the accuracy of this model decreased to 62% and 71%.

Table 3: Classification result with using random forest.

| | | Actual | | |
|---|---|---|---|---|
| | | GT | PT | Total |
| Predicted | GT | 1085 | 472 | |
| | PT | 287 | 742 | |
| | Total | 1372 | 1214 | 2586 |

In order to solve this problem, we chose RFSS method to calculate the key degree of each factor with more objective scores. Figure 4 shows the scores result by using the RFSS method.

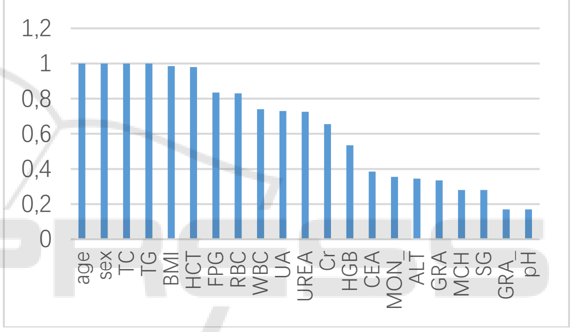

Figure 4: The scores result by using the RFSS method.

The highest four features' scores were 1.0, which means that they were selected as useful features (the score was affected by the regularization parameter alpha) every time. The next few features' scores began to decline, but the decline was not particularly sharp like the result by using the random forest method. It can be seen that RFSS method was helpful in overcoming overfitting and misunderstanding data. Good features did not have low scores due to feature correlation. So the method is better than the method only using random forest.

Table 4 shows the classification effect by RFSS method through 10-fold cross-validation. The risk factor dimension was still 13, but at the same time, we improved the classification sensitivity and accuracy. 909 of the 1214 PT patients were predicted to be accurate and the sensitivity raised to 75%. The accuracy of this model was 77%. This method achieved the desired effect, that is, less dimension and higher accuracy.

Table 4: Classification result with using RFSS.

| | | Actual | | |
|---|---|---|---|---|
| | | GT | PT | Total |
| Predicted | GT | 1082 | 305 | |
| | PT | 290 | 909 | |
| | Total | 1372 | 1214 | 2586 |

## 4.2 Discussions

Finally, we selected 13 features to be risk factors for our proposed context-aware system through the RFSS method. They are age, sex, TC, TG, BMI, HCT, FPG, RBC, WBC, UA, UREA, Cr and HGB. The accuracy of the results can be supported by the classification results and related literature.

In the last few years, there had been many studies on rick factors of hypertension. The results showed that age, sex (Virdis et al., 2002), BMI, genetic factors, waist circumference (Ashwell et al., 2012), TG (Wang, 2013), Urea (Pearson et al., 2001), were the most important risk factors affecting hypertension. This is consistent with the result of selecting important features by using a random forest algorithm.

Table 5 shows the comparison of classification effect by selecting different features as risk factors according to different screening methods. All methods had been tested by 10-fold cross validation. We can see that the sensitivity was low when we used all the features for classification, and the high dimension would affect the system performance. After the filter method, the indicators had improved, but the number of dimension still cannot make us satisfied. Therefore, in the method III and method IV, we adopted machine learning method, and the dimension was all controlled to the same number 13 to facilitate the comparison of experimental results. Compared with method I and III, the accuracy and sensitivity of method IV were greatly improved. Compared with method II, there was an advantage in the number of dimensions.

In the study of Maryam Tayefi et al., (2017), they used decision tree algorithm to study hypertension related factors, and the accuracy can reach 73%. Besides that, in the study of Wang et al., (2014), they used a logistic regression and artificial neural network-based approach to predict hypertension. The sensitivity and accuracy can reach 49% and 77%. For comparison, we used the same methods to train the data set of this study through 10-fold cross-validation. As shown in Table 6, The result for testing data set shows that the accuracy, sensitivity and specificity of using decision tree training model was 73%, 61%, 78%. And the neural network algorithm could reach 75%, 72% and 79%. As can be seen, our method performed better on these three indicators and this method is most suitable for feature screening.

Table 6: Results compared with other method.

| Measure | Decision Tree | Neural Network | Method RFSS |
|---|---|---|---|
| Sensitivity | 61% | 72% | 75% |
| Specificity | 78% | 79% | 78% |
| Accuracy | 73% | 75% | 77% |

Compared to other studies, the advantage of this study is that the sample size is large and 8619 people were used for modelling. Therefore, the method has strong applicability. Another advantage of this method is that the preselected physical indicators had 32 indicators, which was large enough to screen features. There are more iterations to increase the amount of experiment, which can improve the stability of the screening. However, there are some shortcomings in the evaluation. As a medical data study, we should also use statistical hypothesis testing to consider the random fluctuation in the results, so as to improve the accuracy and statistical significance of the results. We will improve this part in future studies.

The focus on this paper is to study how to screen features, it is necessary to analyse why the accuracy of method IV is not particularly high. Firstly, raw data set was acquired from physical examination, without specific data indicators like TOD, FHCD.

Table 5: The comparison of classification effect by using different screening methods.

| Method | Feature Selection | Dimension | Algorithm | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| I | None | 32 | SVM | 66% | 54% | 78% |
| II | PCC | 23 | SVM | 75% | 73% | 78% |
| III | Random Forest | 13 | SVM | 71% | 62% | 79% |
| IV | RFSS | 13 | SVM | 77% | 75% | 78% |

With the development of Internet of Things technology, patients will have more complete data, and the accuracy rate will increase. On the other hand, this SVM classifier was only used to judge the screening effect of each method. As mentioned in Chapter 2, the system classification module used weighted algorithm. Therefore, the classification effect of the whole system will be more accurate.

## 5 CONCLUSION

In this paper, we designed a risk factor screening module using different screening methods based on the proposed context-aware system for hypertension. After comparison and improvement, we selected the RFSS method combined by random forest and stability selection in four methods. We gradually filtered 32 parameters in context information obtained from DA module to 13 hypertension risk factors, and performed by SVM classification algorithm. Accuracy, sensitivity and specificity have been improved. This method can improve the screening rate and ensure the accuracy of this system. Therefore, the context-aware system of hypertension will improve performance by using this screening method. The output of the system can assist doctors and patients to have a comprehensive understanding of their blood pressure condition according to risk factors.

In the future research, we will increase the dataset and data parameters with the improvement of the performance of portable devices. On this basis, we will extend this work by applying DL technologies such as CNN, in order to see whether the accuracy can be increased. In addition, statistical hypothesis testing will be added to experimental verification and the results will be compared with clinical practice. Finally, the automated hypertension risk assessment approach will be improved based on future study. The diagnosis of doctors will be more accurate and comprehensive.

## ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (No.61471064), and National Science and Technology Major Project of China (No.2017ZX03001022).

## REFERENCES

Lim S. S., Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H. et al., 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859), pp.2224-2260.

Chen, W., R.Gao, L.Liu, 2014. China Cardiovascular Disease Report 2013. *Chinese Circulation Journal*, 29(07), pp.487-491.

Carey R. M, Whelton P K, 2018. Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Synopsis of the 2017 American College of Cardiology/American Heart Association Hypertension Guideline. *Annals of Internal Medicine*.

Whitworth J. A, Chalmers J., 2004. World health organisation-international society of hypertension (WHO/ISH) hypertension guidelines. *Clinical and Experimental Hypertension,* 26, pp. 747-752.

Guixia Kang, Meikui Zhang, 2010. Recent Advances of Wireless eHealth (WeHealth) in Terms of 'Sensing China'. *Proceedings of the 5th International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, pp.517-519.

Sandi, G., I. G. B. B. Nugraha and S. H. Supangkat, 2013. Mobile health monitoring and consultation to support hypertension treatment. *International Conference on ICT for Smart Society*, pp. 1-5.

Hong-Tao, Y. U., J. L. Wang and L. I. LiLua, 2007. Research on the risk factors of hypertension and OSAHS in the snore patients with different living habits. *Journal of Clinical Pulmonary Medicine*.

Kunes, J. and Zicha, J., 2009. The interaction of genetic and environmental factors in the ethology of hypertension. *Physiological Research*, 58, pp. S33.

Ding C., Peng H., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), pp. 1226-1238.

Kohavi R., John G. H., 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), pp.273-324.

Hu Z. Y, Bao Y. K., 2015. Hybrid filter-wrapper feature selection for short-term load forecasting. *Engineering Applications of Artificial Intelligence*, 40, pp.17-27.

Koenigstein N., Paqued U., 2013. Xbox movies recommendations: Variational Bayes matrix factorization with embedded feature selection. *Proceedings of the 7th ACM conference on Recommender Systems*.

Madan, A. K., H. Dureja S. Gupta et al., 2008. Topological models for prediction of pharmacokinetic parameters of cephalosporins using random forest, decision tree and moving average analysis. *Sci Pharm*, 76, pp.377–394.

Breiman, L., 2001. Random forest. *Machine Learning*, 45, pp. 5–32.

Nicolai M, Peter B, 2010. Stability selection. *Journal of the Royal Statistical Society*, 72(4), pp.417-473.

Cristianini N, Shawe-Taylor J, 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, UK, 1st edition.

V.Vapnik, S. Golowich, and A. Smola, 1997. *Support vector method for function approximation, regression estimation and signal processing, in Advances in Neural Information Processing Systems.* MIT Press, Cambridge.

Virdis, A., L. Ghiadoni, I. Sudano and et al, 2002. Endothelial function in hypertension: role of gender. *Journal of Hypertension Supplement Official Journal of the International Society of Hypertension*, 20(2), pp. 11–6.

Ashwell, M., Gunn, P. and Gibson, S., 2012. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. *Obesity Reviews*, 13, pp. 275–286.

Wang, X. X., 2013. Relationship of essential hypertension prognisis with serum TG, TC concentration. *Journal of Hainan Medical University*.

Pearson, D. L., Dawling, S., Walsh, W. F. et al, 2001. Neonatal pulmonary hypertension–urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. *N Engl J Med*, 344, pp. 1832–1838.

Tayefi, M.; H. Esmaeili, K. M. Saberi and et al, 2017. The application of a decision tree to establish the parameters associated with hypertension. *Computer Methods & Programs in Biomedicine*, 139, pp. 83–91.

Wang, A.; N. An, Y. Xia, L. Li and G. Chen, 2014. A Logistic Regression and Artificial Neural Network-Based Approach for Chronic Disease Prediction: A Case Study of Hypertension. *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*, pp. 45-52.