JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 1107-1124 (2020)
DOI: 10.6688/JISE.202009_36(5).0012

A Risk Assessment Approach of Hypertension Based on Mobile Crowd Sensing

HUAN-HUAN ZHAO[1,2,3], ZU-CHANG MA[1] AND YI-NING SUN[1]
[1]*Institute of Intelligent Machines*
Hefei Institutes of Physical Science, Chinese Academy of Sciences
Hefei, 230031 P.R. China
[2]*University of Science and Technology of China*
Hefei, 230026 P.R. China
[3]*School of Computer and Information Engineering*
ChuZhou University
ChuZhou, 239000 P.R. China
E-mail: {zhhh206; zuddrj; suddrj}@126.com

Mobile crowd sensing (MCS) makes full use of the sensing and wireless communication capabilities of smart mobile devices to collect real-time information effectively. It makes it possible to monitor people's health condition in real time. Our health information collected through MCS can be used to improve healthcare service. Hypertension is a widespread chronic disease, and preventing hypertension can effectively reduce the incidence of cardiovascular disease. In this paper, we propose a hypertension risk assessment approach based on mobile crowd sensing, which allows for real time health monitoring and warning. In order to stimulate the enthusiasm of MCS volunteers, optimized communication model is used to reduce the communication cost of non-data-users. Additionally, the current hypertension risk status of patients will be feed back to them in real time. In our approach, binary logistic regression is used to select risk factors of hypertension, and then the risk factors are used as the inputs of BP neural network to construct the risk prediction model. Furthermore, the hypertension risk is further divided into low risk, medium risk and high risk through cumulative distribution function. 4498 samples from a community health service center in Hefei area were used to evaluate the performance of the proposed approach. The experimental results show that the proposed approach can provide real-time, effective monitoring and dynamic feedback of the hypertension risk, offering a novel clinical tool for the early warning of hypertension. The proposed approach also provides a general framework for risk assessment of other chronic diseases.

Keywords: mobile crowd sensing, hypertension, risk assessment, BP neural network, real time

1. INTRODUCTION

Nowadays, with the aging of population and the prevalence of sub-health, the traditional interactions between doctors and patients have been unable to meet the increasing demands for healthcare. With the popularity of smart phones and development of wireless technology and sensors, a new sensing structure named mobile crowd sensing (MCS) has emerged [1]. Mobile crowd sensing refers to people sensing information about people and surrounding environment using smart phones and other mobile terminals carried around, and then providing useful information and services to users based on the collected in-

Received August 31, 2019; revised October 25 & December 5, 2019; accepted December 27, 2019.
Communicated by Xiaohong Jiang.

formation. MCS provides a new way for people to sense the world. It enables anyone who performs sensing tasks through mobile terminals to participate in the sensing process and provides universal services. As an important application branch, mobile health crowd sensing is a major component of MCS. It combines mobile crowd sensing with telemedicine. Participants upload health data collected by mobile terminals to cloud server and enjoy various services provided by medical institutions. Among the applications of health service, developing and constructing an effective risk assessment model for chronic diseases is of great value in chronic disease management. At present, the disease risk assessment service is mainly implemented based on the experience of doctors, which consumes a lot of time and resources of doctors for one-to-one intervention and continuous assessment. Therefore, there is an increasing demand for an automated disease risk assessment system, which can make full use of the collected information and ultimately provide health care decision support. The system is of great value for disease prevention. In order to meet this requirement, we need to collect large amount of effective data in real time, at the same time, an effective disease risk assessment approach is also required.

Chronic diseases are responsible for 70% of all deaths worldwide. Hypertension is a widespread chronic disease. At present, the total number of hypertensive patients can reach 972 million worldwide and this number continues to grow [2]. Hypertension is also an important risk factor of cardiovascular and cerebrovascular diseases [3, 4]. Studies have shown that hypertension is the first risk factor leading to death and the third factor of economic burden of disease [5]. According to the latest investigation [6], from 2012 to 2015, the prevalence rate for hypertension of Chinese adults had reached 27.9%, and keeps increasing continuously. Especially for adolescents, the situation is also unoptimistic [7]. The awareness, treatment and control rates of hypertension for Chinese adults were 51.6%, 45.8% and 16.8% respectively, which indicates that prevention and control of hypertension is unsatisfactory and needs to be improved in China at present.

The best way to improve the situation of hypertension is to prevent the occurrence of hypertension. Researches have shown the occurrence of hypertension is affected by heredity, lifestyle *etc*. Improving unhealthy lifestyle before the onset of hypertension can reduce the risk of hypertension and prevent the occurrence of hypertension effectively [8]. Therefore, the risk assessment of hypertension plays a vital role in preventing hypertension. Currently, the risk assessment of hypertension for Chinese people is still a challenge. For one thing, the data used to build the model are mainly from hospitals or specialized medical institutions. Most of the data are non-real-time with a long update cycle, and it is difficult to feed the assessment result back to individuals in real time. For another, the selection of hypertension risk factors and modeling methods are of vital importance on the performance of risk assessment model. In order to provide a basis for the primary prevention of hypertension for Chinese population, a hypertension risk assessment approach based on mobile crowd sensing was proposed in this paper. First, hypertension relative information including medical data, lifestyle data, individual characteristic data were collected through mobile crowd sensing network, then binary logistic regression was used to analysis of risk factors of hypertension from the collected information, and then hypertension prediction model was constructed based on BP neural network training by selecting the optimal threshold to solve the unbalanced problem of training sample. Based on the risk value, cumulative distribution function was used to classify hypertension risk by controlling error rate. The remainder of this paper is organized as follows. Section 2 summarizes related research work. Section 3 illustrates our proposed approach. Section 4 reports the experimental results on real world medical data set and presents the performance study. Finally, Section 5 gives conclusions and future work of this study.

2. RELATED WORK

A large number of researches have constructed considerable work in investigating remote health monitoring based on wireless sensor network (WSN). For example, in a multi-source healthcare architecture called MSHA [9], multi-source data were from sensors in wireless body domain networks and text-based input anywhere. Kaur *et al.* [10] proposed a general multi-sensor fusion method to realize continuous monitoring of remote patients using wireless body sensor network. Health risk assessment and decision making algorithm were used to make reliable health management decisions. However, the location of WSN is fixed and cannot monitor health information of individuals in various situations in real time. Mobile crowd sensing provides new ideas to solve these problems. Mehdi *et al.* [11] proposed a general MCS reference architecture for healthcare. Collection, networking and transmission of healthcare information in MCS were further studied [12, 13]. Yang *et al.* [14] and Wen *et al.* [15] explored how to take effective measures to encourage individuals to participate in MCS actively. Pryss *et al.* [16] taking a specific disease as an example, MCS was used in combination with data anonymization to provide a large number of daily data sets of patients at a low cost, thereby making clinical trials possible anywhere. In order to create an experimental database through MCS, Jovanovic *et al.* [17] proposed a machine learning method to determine the current hypertension status through ECG signals and to collect data of hypertensive patients. It indicated that it is feasible to determine hypertension status according to information collected from MCS. However, it only considered ECG signals to determine hypertension or not. In fact, there are many other risk factors and indicators for hypertension.

Many researchers have studied the risk factors and prediction models of hypertension, which indicated that different populations have different characteristics of hypertension. Age, gender, BMI, diabetes and blood pressure were the most common predictors [18]. Logistic regression [19, 20] and neural network [21-23] were employed to establish corresponding hypertension prediction models for different populations. Ture *et al.* [24] compared the performance of hypertension prediction models established by decision tree, statistical method and neural network, and found that neural network performed the best. The Framingham heart research center calculated the risk of hypertension in the next 1, 2, 4 years based on cohort studies [25]. Muntner *et al.* [26] and Mika *et al.* [27] researched the validation of Framingham prediction model for local residents, and the results showed that the model is validate for European population. But a South Korean study [28] showed that the Framingham hypertension risk prediction model underestimated the risk of South Koreans. Researchers from Japan, India, Iran and Swedish have developed risk prediction models of hypertension for local residents [29-32]. The research of Zheng *et al.* [33] showed that the Framingham hypertension risk prediction model couldn't assess hypertension risk of Chinese rural population effectively. Thus, it is urgent to develop a risk prediction model of hypertension for Chinese people. Chien *et al.* [34] established a hypertension risk assessment model for Chinese-Taiwanese. Li *et al.* [35] established a 15-year hypertension risk prediction model based on a cohort study of 3,899 people in 11 provinces of China. But neither model included lifestyle risk factors, which limits their application in primary prevention of hypertension. In order to evaluate the hypertension risk of Chinese steel workers, Wu *et al.* [36] established a hypertension risk score model for steel workers using LVQ neural network, which has a high accuracy. However, this model is only applicable to steel workers.

Although some progress has been made in the research of hypertension risk assessment, there are still some difficulties and limitations in applying the current research to the prevention of hypertension in the Chinese population. Regarding the collection of modeling data and the feedback of assessment results, telephone follow-up or face-to-face questionnaire are the most common approaches, these approaches may cost lots of time and resources, which makes them difficult to apply to a large population. Mobile devices can monitor activity and behaviors effectively anytime and anywhere, and provide real time, dynamic feedback in response to collected information. Regarding the assessment of hypertension risk, it is challenging to predict the risk of hypertension using risk factors. Improper selection of risk factors and modeling methods may affect the performance of the model. The neural network has shown good classification performance, but now it is mostly used in predicting the occurrence of hypertension or not, but in fact, risks assessment of hypertension is what we need. In order to solve the problems above, this study proposes a risk assessment approach for hypertension based on mobile crowd sensing. Experimental results show that the risk assessment approach can identify the risk of hypertension effectively and give real time feedback to users, providing novel guidance for primary prevention of hypertension.

3. PROPOSED APPROACH

In this section, we first present the framework and the flow of hypertension risk assessment approach, and then introduce the components of the framework. In the approach, a large amount of hypertension relative data, including medical data, lifestyle data and individual characteristic data can be collected via mobile equipment, which establishes a strong foundation for hypertension risk assessment. This risk assessment system learns medical knowledge from the collected information and simulates a doctor's diagnosis using machine learning to provide reliable risk assessment. As illustrated in Fig. 1, the entire framework of the proposed approach can be divided into three modules: data collection module, communication network module and data processing module.

The specific flow of the approach is shown in Fig. 2. When starting the task of building the model, in first step, health-relative information is collected by participants through smart phone. In the second step, the collected information is transmitted to cloud server for data storage through communication network. Third, the collected information is processed in the cloud server. Data preprocessing is used to deal with missing values and finish data standardization. For data processing, first, logistic regression is used to analyze the risk factors of hypertension. Second, BP neural network is used to construct a model for outputting the risk value of hypertension. Third, the cumulative distribution function was used to stratify the risk of hypertension by controlling the error rate. In this way, we construct the hypertension risk assessment model. The model can provide real-time service for participants and some institutes.

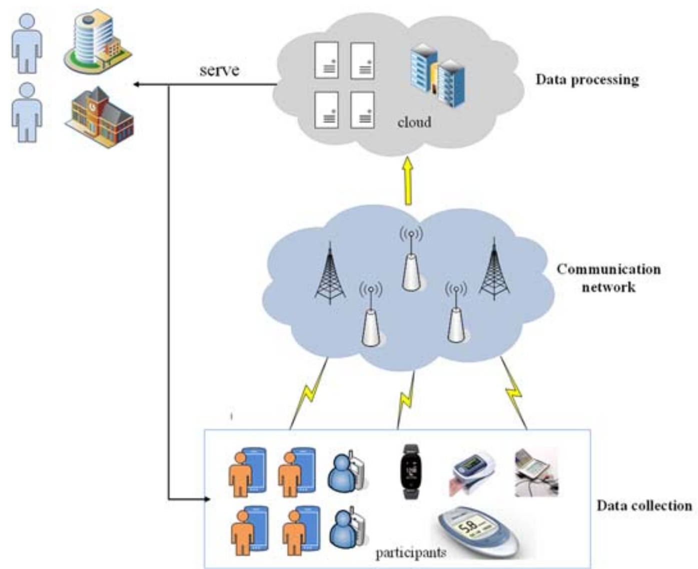


Fig. 1. Framework of hypertension risk assessment approach based on mobile crowd sensing.

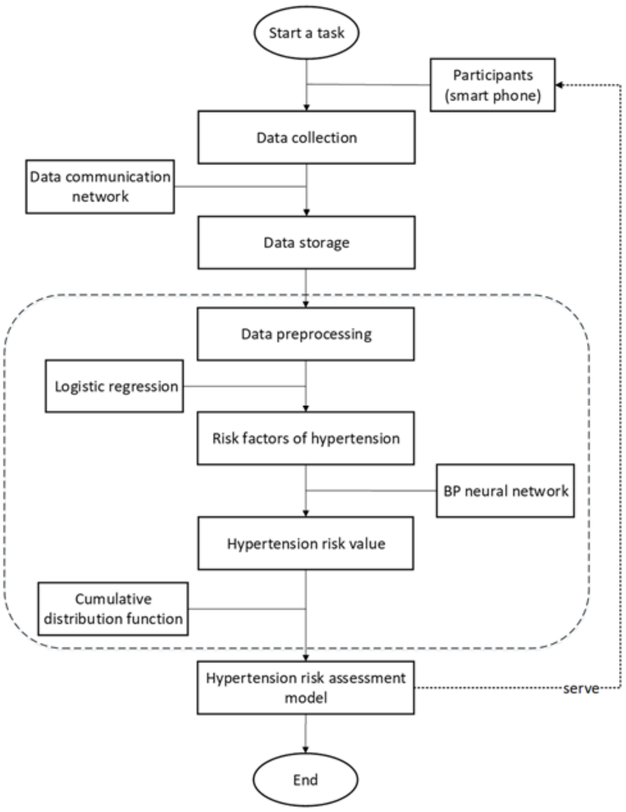


Fig. 2. Flowchart of hypertension risk assessment approach based on mobile crowd sensing.

3.1 Data Collection Module

Health-related information is multi-dimensional, and physiological indicators during work, exercise and even travel are an important part of health-related information. This kind of dynamic health information cannot be obtained from traditional medical system. Mobile crowd sensing provides a solution to this problem. Mobile crowd sensing is a kind of voluntary action, participants volunteered to submit their health-related information for the sake of common interests and benefits. Therefore, appropriate incentive mechanism should be taken to maintain the enthusiasm of participants. In this study, each time participants submit sensing information, the risk level of hypertension will be assessed and feedback to them as incentives. The hypertension relative information are gathered by smartphones, medical devices and then sent to the cloud server center. Data collection module is responsible for collecting medical data, lifestyle data, and individual characteristic data through mobile devices. For example, body fat scale can measure weight and body fat, blood pressure, blood lipids and blood glucose can be measured by blood pressure meter, blood lipid meter and blood glucose meter respectively; ECG can be measured by a single derivation electrocardiography sensor, temperature probe can measure temperature, and smart bracelet can collects information such as heart rate, sleep and physical activity; Lifestyle data can be collected through questionnaire on the APP in the smartphone. Therefore, a large number of hypertension relative information can be relayed from the device to the smartphone and then to the Internet.

3.2 Communication Network Module

The data collected from the information collection module is transmitted to the cloud server center through wireless access communication network and IP-based core network. Among them, the wireless access communication network is the network that connects the customer and service provider directly. It varies from a few hundred meters to a few miles in diameter. The core network adopts fiber structure with high transmission rate.

Wireless communication protocols for access communication networks include ZigBee (IEEE 802.15.4), Bluetooth (IEEE 802.15.1), Wi-Fi (IEEE 802.11) and 3G/4G/ 5G. In this study, ZigBee is excluded because it is not supported by smart phones at present. In order to meet the communication accuracy requirements, the distance between communication devices required for Bluetooth is between 10 meters and 100 meters. Therefore, in this system, the detection results of medical devices are transmitted to smart phones through Bluetooth. Wi-Fi is an important communication technology that can be used in mobile devices. Compared with Bluetooth, the transmission speed of Wi-Fi is faster and it can spread over a longer distance. Therefore, mobile users who have access to Wi-Fi can use Wi-Fi for data transmitting. For some scenarios where Wi-Fi is not available, mobile devices can transmit data over cellular network, which is the most widely used technology because of easy access to mobile phones despite their high cost. At present, full Wi-Fi coverage has not been achieved, and additional overhead may be incurred when using a cellular network for data transmission. Due to additional overhead, some users may not participate in mobile crowd sensing, especially for non-data-plan users. Therefore, inspired by researches [37, 38], the communication model we used in cellular network is shown in Fig. 3, the specific communication process is as follows:

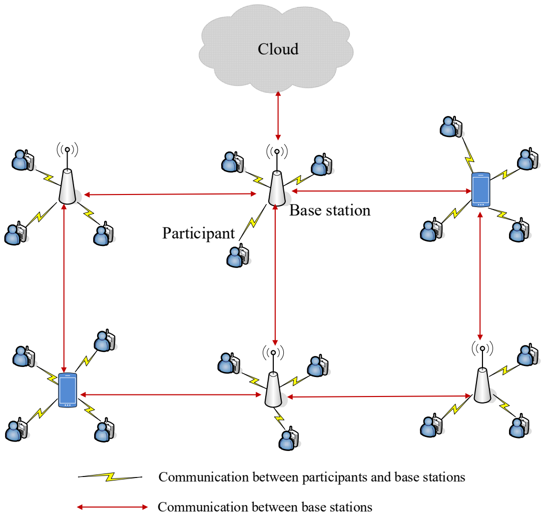


Fig. 3. Communication model in cellular network.

Step 1: Deploy and initialize base stations for data collection;
Step 2: Participants who are data-plan users can apply to become temporary base station nodes (in this way, they will be rewarded accordingly);
Step 3: Among the base stations within the allowed communication range, the participant chooses the base station with the shortest distance to submit the sensing data. In this case, Bluetooth and Wi-Fi are the most commonly used communication media;
Step 4: Base station nodes forward all collected data to the cloud service center through one or multiple hops.

This communication model can reduce the energy consumption of data transmission and balance the load of cellular network while encouraging users to participate in mobile crowd sensing.

3.3 Data Processing Module

In cloud service center, many kinds of analysis with health information can be implemented to provide medical and health management decision support for health service providers and also provide real-time health care services for individuals. For our approach of hypertension risk assessment, the data processing consists of three steps. Firstly, binary logistic regression was used to analyze the risk factors of hypertension. Secondly, BP neural network was used to construct hypertension prediction model which can output the risk value of hypertension. Third, the cumulative distribution function was used to stratify the risk of hypertension by controlling the error rate.

(A) Logical regression analysis

Logistic regression is a statistical method, which can explore the relationship between a categorical dependent variable and several independent variables. It has been widely used in epidemiological risk factor analysis. According to the number of possible values of dependent variables, logistic regression can be divided into binary logistic regression and multivariate logistic regression. Taking the risk factors for hypertension as an example, binary logistic regression analysis is required.

Dependent variable y stands for hypertension or not, independent variables $x_1, x_2, \ldots, x_n$ stand for n possible risk factors associated with hypertension. Logistic regression model formula is shown in Eq. (1), where β_i stands for the coefficient of x_i.

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{1}$$

We can express the predicted probability of y as Eq. (2). Logistic prediction model can be used to analyze $x_1, x_2, \ldots, x_n$ one by one, and finally obtain influential factors from the n independent variables.

$$p = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)]} \tag{2}$$

In this paper, binary logistic regression analysis is used to explore the correlation between independent variables (medical data, lifestyle data, individual characteristic data) and dependent variable (hypertension or not), so as to explore the risk factors of hypertension.

(B) BP neural network

Neural network is a supervised learning method, which can learn the law of the training data, thus can transform input data into the appropriate output, because the nerve can solve the problem of collinearity and interaction among variables and have strong tolerance of fault, the neural network has been widely applied to diseases prediction [39, 40].

BP neural network is a feed forward neural network model for classification and pattern recognition. The structure of typical neural network includes input layer, hidden layer and output layer, full connection is between one layer and another, that is, any neuron of each layer has connections to all neurons of the layer before, so that the relationship from input to output can be learned. The network is developed by updating weight iteratively using the back propagation algorithm and optimization algorithm, which contributes to reducing the error between the desired output and the actual output progressively. Algorithm 1 presents the pseudo-code of BP neural network.

Algorithm 1: Neural network learning with the back propagation algorithm

Input: N train samples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, with inputs x_i and corresponding outputs y_i, $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$, where is $x_{i1}, x_{i2}, \ldots, x_{ik}$ is k features of x_i.

Output: a BP neural network

1: Initializing weight $W(w_{ij})$ and threshold value $B(b_j)$ of network

2: Input a training sample (x_i, y_i), $i = 1, 2, \ldots, n$.

3: Calculate the output value of each node j of the hidden layer and output layer:

$$I_j = \sum_{i=1}^{m} w_{ij} * z_i + b_j$$

$$z_j = f(I_j)$$

Where m is the number of nodes in the previous layer

4: Calculate the error of each node j in the output layer:

$err_j = f'(I_j) * (y_{ij} - z_j)$

For (the last hidden layer: the first hidden layer)

Calculate the error of each cell j of the hidden layer:

$err_j = f'(I_j) * \Sigma_i(err_i * w_{ij})$

Where i is the neuron in the next layer of j

5: Adjust the weights and thresholds between two layers adjacent:

$w_{ij} = w_{ij} + \mu * err_j * z_i$

$b_j = b_j + \mu * err_j$

Where μ is the learning rate

6: If termination condition not satisfying, goto step 2; Else, stop

(C) Risk stratification method

Based on the risk value output by neural network, we used the accumulation distribution function to find the boundary between high risk and medium risk, medium risk and low risk in the training data [41], and divided the risk of hypertension into high, medium and low risk. The boundary between high risk and medium risk is chosen when 10% of non-hypertension was assessed as high risk, the boundary between medium risk and low risk is chosen when 10% of hypertension population assessed as low risk. Algorithm 2 presents the pseudo-code of risk stratification method.

Algorithm 2: Risk stratification method

Input: m risk values $r_{h1}, r_{h2}, \ldots, r_{hm}$ for hypertension, r_{hi} is the risk value of hypertension sample i,
k risk values $r_{l1}, r_{l2}, \ldots, r_{lk}$ for non-hypertension, r_{li} is the risk value of non-hypertension i

Output: boundary between risk levels $B = \{B_1, B_2\}$, B_1 is the boundary between medium risk and low risk, B_2 is the boundary between high risk and medium risk.

for (r = 0:0.001:1) **do**
    $h = 0$
    for ($r' = r_{h1} : r_{hm}$) **do**
        if ($r' \leq r$) **do**
            $h = +1$
        end if
    end for
    if ($h/m \geq 10\%$) **do**
            $B_1 = r$
            break
    end if
end for
for (r = 1:–0.001:0) **do**
    $l = 0$
    for ($r' = r_{l1} : r_{lk}$) **do**
        if ($r' \geq r$) **do**
            $l = +1$
        end if

end for
if ($l/k \geq 10\%$) **do**
$B_2 = r$
break
end if
end for

4. EXPERIMENTS

This section investigates the performance of the proposed approach against previous hypertension risk prediction model in hypertension detection, which is measured in terms of the sensitivity, specificity, AUC and Youden's index. In addition, the consistency of the training set and testing set is also investigated.

4.1 Datasets and Risk Factors of Hypertension

This subsection illustrates the preparation of the simulation data. The data analyzed in this paper is from a community health service center in Hefei area. There were a total of 4550 samples from April 2018 to July 2019, among which 52 samples were with missing values, and a total of 4498 cases were complete data. There were 1098 cases of hypertension and 3400 cases of non-hypertension in the sample. In addition to general physical examination and laboratory measurements, relevant lifestyle behaviors and habits, including history of chronic disease, lifestyle and physical activity habits were collected through face-to-face questionnaires by a professional worker in the community health service center. According to the diagnostic criteria of the Chinese guidelines for the prevention and treatment of hypertension, hypertension is defined as having been diagnosed as hypertension in hospital, or having an average systolic blood pressure greater than 140 or diastolic blood pressure greater than 90 during physical examination. According to previous researches, the risk factors of hypertension mainly include: gender, age, height, weight, Body Mass Index (BMI), family history of hypertension ("family history" was used for short in the rest of article), smoking, high-salt diet, diabetes, hyperlipidemia, physical activity, *etc.* [18]. Firstly, all these factors were included in the study. BMI is defined as weight (Kg) divided by square of height (m^2). Diabetes is defined as being told suffering diabetes by the doctor or fasting blood glucose is greater than 7.0mmol/L or 2 hours postprandial blood glucose is greater than 11.1mmol/L. Hyperlipidemia was defined as being told suffering hyperlipidemia by the doctor or total cholesterol is or greater than 5.2mmol/L or triglyceride is or greater than 1.7mmol/l or LDL-cholesterol is or greater than 3.4mmol/L or HDL-cholesterol is less than 1.0mmol/L. Family history is defined as parents or brothers and sisters having hypertension. Smoking is defined as total amount of smoking is more than 100 or smoking now. High-salt diet is defined as taking more than 6g of salt per day at ordinary times. Physical activity is divided into three categories, those with regular exercise habits (at least 3 times a week, 30 minutes each time) are defined as "regular exercise", those only with housework but without regular exercise were defined as "housework", and those without any regular physical activity are defined as "no exercise".

Logistic regression analysis has the capacity to determine significant factors. Therefore, the binary logistic regression was used to analyze the sample data, the dependent variable was hypertension or not, 1 stand for hypertension and 0 stand for non-hypertension. The independent variables including continuous variables: age, BMI, and categorical variables: gender, family history, smoking, high-salt diet, diabetes, hyperlipidemia, physical activity. When setting critical value of statistical significance p at 0.05, the correlation of dependent variable and each independent variable showed that age, BMI, family history, smoking, high-salt diet, diabetes, hyperlipidemia and physical activity was significantly correlated to hypertension ($p < 0.05$), which indicate they are predict indicators of hypertension that should be included into the model. While gender was not significantly correlated to hypertension ($p > 0.05$), but clinical studies have shown that gender is one of the influence factors of hypertension, so we would include gender as hypertension risk factors into the model.

4.2 Hypertension Prediction Model Based on BP Neural Network

Hypertension prediction model is established based on BP neural network. Where the input layer constitutes of nine risk factors, they are age, gender, BMI, family history, smoking, high-salt diet, diabetes, hyperlipidemia and physical activity. A node is used to represent the output; the output is a decimal number 0 or 1, which can show the risk of hypertension, and the greater the value means the higher the risk of hypertension. We set one hidden layer, thus can not only reduce the amount of computation, but also prevents overfitting. As regard to the number of nodes in the hidden layer, which is not only related to the number of nodes in the input and output layer, but also related to the complexity of the problem to be solved, the type of conversion function and the characteristics of sample data. In order to avoid the phenomenon of "over-fitting" and ensure the performance of the network, the most basic principle to determine the number of hidden layer nodes is minimizing the number of nodes in the hidden layer on the premise of meeting the requirements of precision. The number of nodes in the hidden layer in this model is determined to be 10, which contributes to the smallest correction error of the model. The Tanh was selected as activation function referring to previous research results [22]. Therefore, the predict model is shown in Fig. 4.

4.3 Experimental Results and Analysis

In order to implement the risk assessment approach of hypertension, sample data were randomly divided into training samples and testing samples according to the ratio 7 (a total of 3148 cases, in which 768 cases hypertension, 2380 cases non-hypertension): 3 (a total of 1350 cases, in which 330 cases hypertension, 1020 cases non-hypertension). The training samples were used to train the model, while the testing samples were used to verify the validity of the model. Referring to previous studies [42], to generate the binary hypertension status (Yes or No), the threshold is set at 0.5. But due to the unbalanced character of our sample data, which means hypertension population is far less than non-hypertension population, we consider all thresholds. The optimal threshold is the threshold which makes the sum of sensitivity and specificity of the training set reach the maximum value. A person is identified as hypertension when the output of neural network is larger than the threshold value, otherwise, he will be predicted as non-hypertension. We use the testing set to verify the performance of the model.

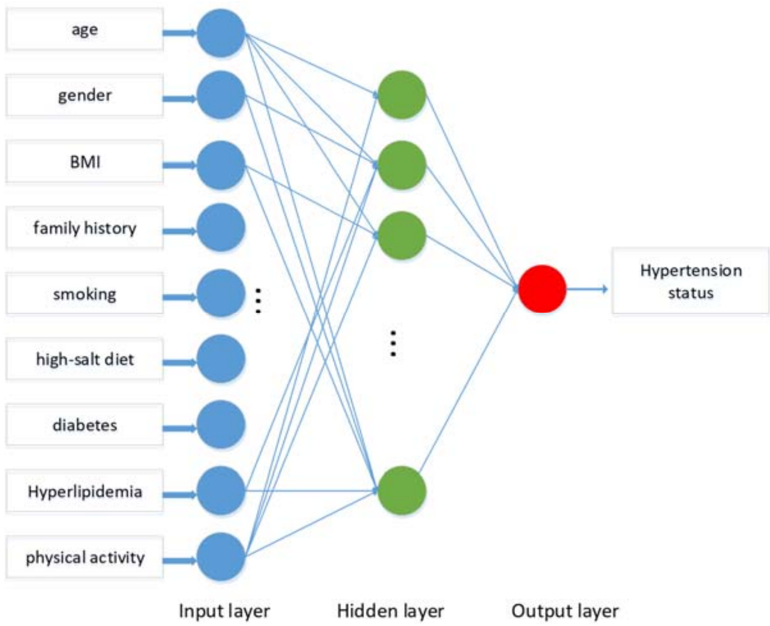


Fig. 4. Risk prediction model of hypertension based on neural network.

The accuracy, sensitivity, specificity, area under the curve (AUC) and Youden's index were used to evaluate the diagnostic performance of our model. Accuracy presents the probability of correctly identifying hypertension and non-hypertension. Sensitivity is defined as the probability of correctly identifying hypertension. Specificity refers to the probability of correctly identifying non-hypertension. AUC is defined as the area bounded by the coordinate axis under the receiver operating characteristic (ROC) curve, as a value, the classifier with larger AUC has better performance. Youden's index is equal to the sum of specificity and sensitivity minus 1. The sensitivity and specificity at different threshold for both training set and testing set are shown in Fig. 5, and the ROC curves of training set and testing set are shown in Fig. 6. Under the optimal threshold, for training set, the accur-

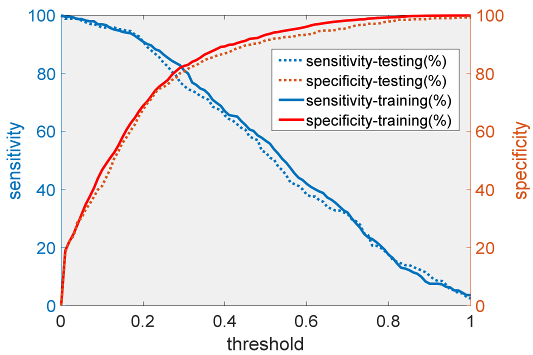


Fig. 5. Specificity and sensitivity of training set and testing set at different threshold value.

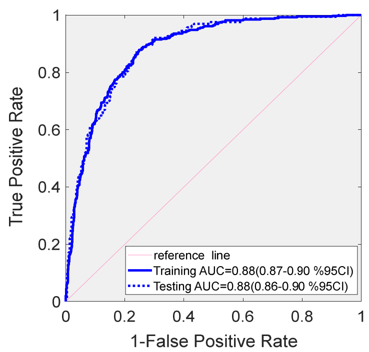


Fig. 6. ROC curves of training set and testing set.

acy was 77.78%(95%Confidence Interval(CI):77.73%-77.81%), the sensitivity was 87.73% (95%CI:87.69%-87.78%), the specificity was 74.61%(95%CI:74.57%-74.65%), and the AUC was 0.88(95%CI:0.87-0.90). The accuracy, sensitivity, specificity and AUC of the testing set were 77.95% (95%CI: 77.91%-77.99%), 87.82% (95%CI: 87.78%-87.86%), 74.75% (95%CI: 74.70%-74.80%), and 0.88(95%CI: 0.86-0.90) respectively. It is obvious that the training set have good consistency with the testing set in accuracy, sensitivity, specificity and AUC.

We compared the performance of the prediction model in this paper with the previous hypertension prediction models, and the results are shown in Table 1. This model outperforms other models in specificity, Youden's index and AUC, and also performs well in sensitivity.

Table 1. Comparison of hypertension prediction models.

Model	Method	Sensitivity (%)	Specificity (%)	Youden' index (%)	AUC
Fernando2018 [19]	logistic regression	77.00	68.00	45.00	0.73
Wang 2015 [22]	neural network	72.76	67.96	40.72	0.77
Mevlut Ture 2005 [24]	logistic regression	90.48	64.10	54.58	0.79
This approach	neural network	87.82	74.75	62.57	0.88

The above results show that our model has good performance on predicting hypertension. But in fact, our goal is to predict the risk of hypertension, and finally realize early warning of hypertension. Good performance of hypertension diagnostic, which can ensure the accuracy of risk degree, is the basis of risk classification. So, risk stratification of hypertension needs to be implemented further. Different from hypertension prediction, we let output of the neural network is not 0 or 1, but a continuous value between 0 and 1, which shows the risk of hypertension. We divided the risk of hypertension into high, medium and low, using the accumulation function to find boundaries between high risk and medium risk, medium risk and low risk in the training data. The boundary between high risk and medium risk is chosen when 10% of non-hypertension were predicted as high risk, the boundary between medium risk and low risk is chosen when 10% of hypertension population were predicted as low risk. The result of risk stratification is shown in Fig. 7.

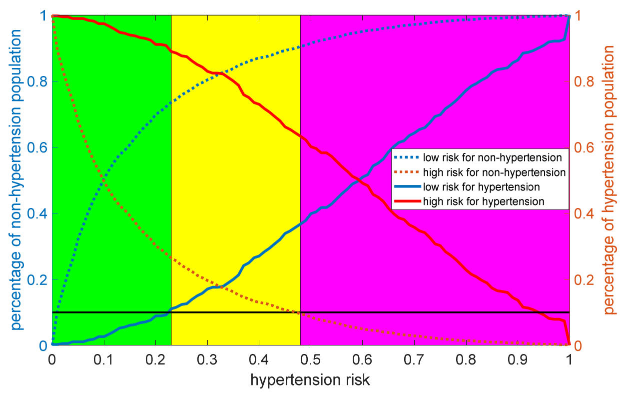


Fig. 7. Result of risk stratification.

We used the testing set to verify the performance of the risk stratification method. The results were shown in Table 2. Among 330 hypertension population, 22 (6.7%) were predicted as low risk, 88 (26.7%) were predicted as medium risk, 220 (66.7%) were predicted as high risk. Among 1020 non-hypertension population, 652 (63.9%) were predicted as low risk, 250 (24.5%) were predicted as medium risk, and 118 (11.6%) were predicted as high risk. which shows that our risk stratification method can effectively identify the risk of hypertension and provide a new idea for the early warning of hypertension.

Table 2. Performance of the risk stratification method in testing set.

	Total	#Low risk	%Low risk	#Medium risk	%Medium risk	#High risk	%High risk
hypertension	330	22	6.7%	88	26.7 %	220	66.7%
Non-hypertension	1020	652	63.9%	250	24.5%	118	11.6%

5. CONCLUSIONS

In the paper, we propose a hypertension risk assessment approach based on mobile crowd sensing. The participants can access the system through smart phones and upload their hypertension-related information anytime and anywhere. In order to encourage participants, the assessment results will be feed back to them in real time. Additionally, optimized communication model is used to reduce the communication cost of non-data-users. Then the collected information can be relayed to the cloud service center through the communication network. The data are processed and analyzed in the cloud service center to provide corresponding health care services. In our approach, binary logistic regression is used to identify risk factors of hypertension, and then BP neural network with risk factors as input is used to calculate the risk value of hypertension. Finally, further risk stratification of hypertension is realized by cumulative distribution function.

4498 samples in Hefei area were used to implement the simulation experiment. Results show that the average accuracy, sensitivity, specificity and AUC of this model are 77.95%, 87.82%, 74.75%, 0.88 respectively. In addition, the performance of training set is well consistent with the testing set, thus the model outperform the previous hypertension prediction model. Referring to risk level, the error is controlled within 10%, indicating the approach can assess hypertension risk effectively. Compared with the traditional health care service mode, this approach can collect a large number of hypertension related information effectively and provide real time, dynamic feedback in response to the collected information, which will effectively improve the interaction between individuals and doctors and the benefits of health care. This study provides a new idea for the early warning of hypertension. Although this paper only considers hypertension, this approach is also applicable to risk assessment of other chronic diseases. For future, we will explore effective approaches to alleviate network congestion which is caused by massive real-time data transmission.

ACKNOWLEDGMENT

This work was supported by the fund support of Major Science and Technology Project of Anhui Province (18030801133), the University Natural Science Foundation of Anhui Province (KJ2017B16), Natural Science Foundation of Anhui Province (1608085MF 127), and the University Natural Science Foundation of Anhui Province (KJ2017A419).

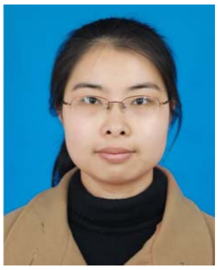
REFERENCES

1. N. D. Lane, E. Milezzo, H. Lu, *et al.*, "A survey of mobile phone sensing," *Communications Magazine*, Vol. 48, 2010, pp. 140-150.
2. S. Yusuf, S. Hawken, S. Ounpuu, T. Dans, A. Avezum, F. Lanas, *et al.*, "Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study," *LANCET*, Vol. 364, 2004, pp. 937-952.
3. J. K. Johansson, T. J. Niiranen, P. J. Puukka, and A. M. Jula, "Prognostic value of the variability in home-measured blood pressure and heart rate: the finn-home study," *Hypertension*, Vol. 59, 2012, pp. 212-218.
4. S. Lewington, R. Clarke, N. Qizilbash, R. Peto, R. Collins, and Prospective Studies, "Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies," *LANCET*, Vol. 360, 2002, pp. 1903-1913.
5. R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, *et al.*, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," *LANCET*, Vol. 380, 2012, pp. 2095-2128.
6. W. Chen, R. Gao, L. Liu, M. Zhu, W. Wang, Y. Wang, Z. Wu, H. Li, D. Gu, Y. Yang, *et al.*, "Summary of the report on cardiovascular diseases in China (2017)," *China Circulation Journal*, Vol. 33, 2018, pp. 1-8.
7. T. Midha, V. Krishna, B. Nath, R. Kumari, Y. K. Rao, U. Pandey, *et al.*, "Cut-off of body mass index and waist circumference to predict hypertension in Indian adults," *World Journal of Clinical Cases*, Vol. 2, 2014, pp. 272-278.
8. N. J. Aburto, S. Hanson, H. Gutierrez, L. Hooper, P. Elliott, and F. P. Cappuccio, "Effect of increased potassium intake on cardiovascular risk factors and disease: systematic review and meta-analyses," *Bmj-British Medical Journal*, Vol. 346, 2013, f1378.
9. O. H. Salman, M. F. A. Rasid, M. I. Saripan, and S. K. Subramaniam, "Multi-sources data fusion framework for remote triage prioritization in telehealth," *Journal of Medical Systems*, Vol. 38, 2014, No. 103.
10. P. Kaur, R. Kumar, and M. Kumar, "A healthcare monitoring system using random forest and internet of things (IoT)," *Multimedia Tools and Applications*, Vol. 78, 2019, pp. 19905-19916.
11. M. Mehdi, G. Muhlmeier, K. Agrawal, R. Pryss, M. Reichert, and F. J. Hauck, "Referenceable mobile crowdsensing architecture: a healthcare use case," *Procedia Computer Science*, Vol. 134, 2018, pp. 445-51.
12. P. Foremski, M. Gorawski, K. Grochla, and K. Polys, "Energy-efficient crowdsensing of human mobility and signal levels in cellular networks," *Sensors*, Vol. 15, 2015, pp. 22060-22088.

13. M. J. Moron, R. Luque, and E. Casilari, "On the capability of smartphones to perform as communication gateways in medical wireless personal area networks," *Sensors*, Vol. 14, 2014, pp. 575-594.
14. D. Yang, G. Xue, X. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Transactions on Networking*, Vol. 24, 2016, pp. 1732-1744.
15. Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, *et al.*, "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, Vol. 64, 2015, pp. 4203-4214.
16. R. Pryss, M. Reichert, B. Langguth, and W. Schlee, "Mobile crowd sensing services for tinnitus assessment, therapy and research," in *Proceedings of IEEE International Conference on Mobile Services*, 2015, pp. 352-359.
17. S. Jovanovic, M. Jovanovic, T. Skoric, S. Jokic, B. Milovanovic, K. Katzis, *et al.*, "A mobile crowd sensing application for hypertensive patients," *Sensors*, Vol. 19, 2019.
18. F. M. Sacks, L. P. Svetkey, W. M. Vollmer, L. J. Appel, G. A. Bray, D. Harsha, *et al.*, "Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet," *New England Journal of Medicine*, Vol. 344, 2001, pp. 3-10.
19. F. Lopez-Martinez, A. Schwarcz, E. Rolando Nunez-Valdez, and V. Garcia-Diaz, "Machine learning learning classification analysis for a hypertensive population as a function of several risk factors," *Expert Systems with Applications*, Vol. 110, 2018, pp. 206-215.
20. J. Y. Islam, M. M. Zaman, S. A. Haq, S. Ahmed, and Z. Al-Quadir, "Epidemiology of hypertension among Bangladeshi adults using the 2017 ACC/AHA hypertension clinical practice guidelines and joint national committee 7 guidelines," *Journal of Human Hypertension*, Vol. 32, 2018, pp. 668-680.
21. B. Sumathi and A. Santhakumaran, "Pre-diagnosis of hypertension using artificial neural network," *Global Journal of Computer Science & Technology*, Vol. 11, 2011, pp. 42-48.
22. A. G. Wang, N. An, G. L. Chen, L. Li, and G. Alterovitz, "Predicting hypertension without measurement: A non-invasive, questionnaire-based approach," *Expert Systems with Applications*, Vol. 42, 2015, pp. 7601-7609.
23. D. LaFreniere, F. Zulkernine, D. Barber, K. Martin, *et al.*, "Using machine learning to predict hypertension from a clinical dataset," in *Proceedings of IEEE Symposium Series on Computational Intelligence*, 2016, pp. 1-7.
24. M. Ture, I. Kurt, A. T. Kurum, and K. Ozdamar, "Comparing classification techniques for predicting essential hypertension," *Expert Systems with Applications*, Vol. 29, 2005, pp. 583-588.
25. N. I. Parikh, M. J. Pencina, T. J. Wang, E. J. Benjamin, K. J. Lanier, D. Levy, *et al.*, "A risk score for predicting near-term incidence of hypertension: The framingham heart study," *Annals of Internal Medicine*, Vol. 148, 2008, pp. 102-110.
26. P. Muntner, M. Woodward, D. M. Mann, D. Shimbo, E. D. Michos, R. S. Blumenthal, *et al.*, "Comparison of the framingham heart study hypertension model with blood pressure alone in the prediction of risk of hypertension the multi-ethnic study of atherosclerosis," *Hypertension*, Vol. 55, 2010, pp. 1339-1345.

27. K. K. Mika, B. G David, S. M. Archana, J. E. Ferrie, A. G. Tabak, J. Markus, *et al.*, "Validating the framingham hypertension risk score: results from the Whitehall II study," *Hypertension*, Vol. 54, 2009, pp. 496-501.
28. N. K. Lim, K. H. Son, K. S. Lee, H. Y. Park, and M. C. Cho, "Predicting the risk of incident hypertension in a Korean middle-aged population: Korean genome and epidemiology study," *Journal of Clinical Hypertension*, Vol. 15, 2013, pp. 344-349.
29. T. Otsuka, Y. Kachi, H. Takada, K. Kato, E. Kodani, C. Ibuki, *et al.*, "Development of a risk prediction model for incident hypertension in a working-age Japanese male population," *Hypertension Research*, Vol. 38, 2015, pp. 419-425.
30. S. Asgari, D. Khalili, Y. Mehrabi, S. Kazempour-Ardebili, F. Azizi, and F. Hadaegh, "Incidence and risk factors of isolated systolic and diastolic hypertension: a 10 year follow-up of the Tehran lipids and glucose study," *Blood Pressure*, Vol. 25, 2016, pp. 177-183.
31. T. Sathish, S. Kannan, P. S. Sarma, O. Razum, A. G. Thrift, and K. R. Thankappan, "A risk score to predict hypertension in primary care settings in rural India," *Asia-Pacific Journal of Public Health*, Vol. 28, 2016, pp. 26S-31S.
32. C. Fava, M. Sjogren, M. Montagnana, E. Danese, P. Almgren, G. Engstrom, *et al.*, "Prediction of blood pessure changes over time and incidence of hypertension by a genetic risk score in Swedes," *Hypertension*, Vol. 61, 2013, pp. 319-326.
33. L. Q. Zheng, Z. Q. Sun, X. G. Zhang, J. Li, D. Y. Hu, J. Chen, *et al.*, "Predictive value for the rural Chinese population of the framingham hypertension risk model: Results from Liaoning Province," *American Journal of Hypertension*, Vol. 27, 2014, pp. 409-414.
34. K. L. Chien, H. C. Hsu, T. C. Su, W. T. Chang, F. C. Sung, M. F. Chen, *et al.*, "Prediction models for the risk of new-onset hypertension in ethnic Chinese in Taiwan," *Journal of Human Hypertension*, Vol. 25, 2011, pp. 294-303.
35. G. Li, J. Liu, W. Wang, M. Wang, W. Xie, Y. Hao, *et al.*, "Prediction models for the 15 years risk of new-onset hypertension in Chinese people aged from 35 to 64 years old," Zhonghua Nei Ke Za Zhi, Vol. 53, 2014, pp. 265-268.
36. J. H. Wu, W. Wei, L. Zhang, J. Wang, R. Damasevicius, J. Li, *et al.*, "Risk assessment of hypertension in steel workers based on LVQ and fisher-SVM deep excavation," *IEEE Access*, Vol. 7, 2019, pp. 23109-23119.
37. J. Chen and J. Yang, "Maximizing coverage quality with budget constrained in mobile crowd-sensing network for environmental monitoring applications," *Sensors*, Vol. 19, 2019, pp. 2399-2415.
38. L. Wang, D. Zhang, Z. Yan, H. Xiong, and B. Xie, "effSense: A novel mobile crowd-sensing framework for energy-efficient and cost-effective data uploading," *IEEE Transactions on Systems Man Cybernetics-Systems*, Vol. 45, 2015, pp. 1549-1563.
39. I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Systems with Applications*, Vol. 34, 2008, pp. 366-374.
40. O. Niel, P. Bastard, C. Boussard, J. Hogan, T. Kwon, and G. Deschenes, "Artificial intelligence outperforms experienced nephrologists to assess dry weight in pediatric patients on chronic hemodialysis," *Pediatric Nephrology*, Vol. 33, 2018, pp. 1799-1803.
41. G. R. Hart, D. A. Roffman, R. Decker, and J. Deng, "A multi-parameterized artificial neural network for lung cancer risk prediction," *Plos One*, Vol. 13, 2018, e0205264.

42. Z. W. Wang, Z. Chen, L. F. Zhang, X. Wang, G. Hao, Z. G. Zhang, *et al.*, "Status of hypertension in China results from the China hypertension survey, 2012-2015," *Circulation*, Vol. 137, 2018, pp. 2344-2356.



Huan-Huan Zhao (赵欢欢) received her master degree in Computing Application Technology from Anhui Normal University in 2011. She is currently pursuing the Ph.D. degree at the Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences, China. She is also a Lecturer at School of Computer and Information Engineering, ChuZhou University, China. Her research interests include wireless communication networks and health information technology.

Zu-Chang Ma (马祖长) received his Ph.D. in Precision Instrument and Machinery from University of Science and Technology of China in 2007. He is currently a Professor at Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences. His research interests include health information technology and public health promotion service theory, technology, method and equipment.

Yi-Ning Sun (孙怡宁) received his Ph.D. in Electrical Engineering from the University of Rome in 1993. He is currently a Professor at Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences. His research interests include technology, system and mode of non-medical health promotion service and intelligent information processing technology.