Scientific Research Publishing

Forecasting Alzheimer's Disease Using Combination Model Based on Machine Learning

He Li, Yuhang Wu, Yingnan Zhang, Tao Wei, Yufeng Gui*

College of Science, Wuhan University of Technology, Wuhan, China
Email: *guiyufeng@whut.edu.cn

Abstract

As the acceleration of aged population tendency, building models to forecast Alzheimer's Disease (AD) is essential. In this article, we surveyed 1157 interviewees. By analyzing the results using three machine learning methods—BP neural network, SVM and random forest, we can derive the accuracy of them in forecasting AD, so that we can compare the methods in solving AD prediction. Among them, random forest is the most accurate method. Moreover, to combine the advantages of the methods, we build a new combination forecasting model based on the three machine learning models, which is proved more accurate than the models singly. At last, we give the conclusion of the connection between life style and AD, and provide several suggestions for elderly people to help them prevent AD.

Keywords

Alzheimer's Disease, BP Neural Network, SVM, Random Forest, Combination Forecasting Model

1. Introduction

Alzheimer's Disease (AD) is a rapidly growing neurological issue all over the world. As a common form of dementia, it has neurological and behavioral repercussions. As the sixth leading cause of death, every 71 seconds, there is one person fall ill in AD. Terribly, the rate doubles roughly after age 65 [1].

Many machine learning methods have been developed to model, analyze and predict the relevant problems. For examples, Chung-Chou H. Chang proved smoking is an underlying triggering factor of AD [2]. C Reitz also indicated that smoking increases the risk of dementia by exploring the impact of the APOεE4 allele, sex and age on the association between smoking status and dementia [3]. As a consequence, doing research based on life style to forecast AD has great

significance nowadays.

To determine a more proper result, we use three machine methods to build prediction model. This paper shows three methods for the prediction of Alzheimer's disease—BP neural network, SVM and random forest. In the previous research, many papers show the methods of predicting diseases by machine learning. For example, in 2011, I.A. Illan introduced a computer aided diagnosis (CAD) system for the early detection of AD that combines mask-based feature reduction techniques and a pasting-votes method for aggregation of SVM classifiers ensemble applied to the relevant image components [4], which is defined as a hybrid scheme. Besides, as for the method of random forest, in 2017, Ashwani Kumar introduced models of classification for different AD genes were generated according to Mini-Mental State Examination scores and all other vital parameters to achieve the identification of the expression level of different proteins of disorder that may possibly determine the involvement of genes in various AD pathogenesis pathways [5], which is a new decision tree to solve the puzzle of AD pathogenesis through standard diagnosis scoring system. Like those, by machine learning, we can determine the inner relationship between life style and Alzheimer's disease. Moreover, by comparing the three models and considering the models synthetically, we can determine a new prediction model, which can maximum the advantages as well as having pretty accurate in predicting AD. All the work above is supported by Wuhan University of Technology.

2. Alzheimer's Disease Forecasting Model

2.1. Data Pretreatment

At first, we select elderly people aged more than 65 in Wuhan, China for example, and mainly surveyed via internet. Moreover, we visit hospitals and nursing homes to get enough papery questionnaires. To make the investigation believable, we eliminated the samples with liver, kidney and cardiopulmonary disease. Then we obtained 1157 questionnaires, including 1038 effective questionnaires in them. Next we code some of the factors as shown on Table 1.

2.2. Forecasting Based on Back Propagation Neural Network

Owing to the gradient descent optimization algorithm, back propagation (BP) neural networks can take only from the previous layer and send outputs only to the next layer. Moreover, it is characterized as self-learning and self-adaptability, as well as the significant improved performance in computation-intensive field. Therefore, it is commonly used to adjust the weight of neurons by calculating the gradient of the loss function in the context learning, which is widely used in such aspects as, classification, evaluation, identification and forecasting. As for disease forecasting, BP neural networks can constitute the input layers (symptoms), hidden layers and output (falling ill or not) layers, making it more objective to define the index weight. That is, the different impacts of different factors

Table 1. Assignment of the factors.

Factors	Assignment
Alzheimer's Disease	Patient = 1, Healthy Person = 0
Smoking History	Never = 1, <10 Years = 2, 10~20 Years = 3, >20 Years = 4
Drinking History	Never = 1, Inchoate = 2, <1 Time per Week = 3, 1~3 Times per Week = 4, >4 Times per Week = 5
Exercise	Never = 1, <1 Time per Week = 2, 1~3 Times per Week = 3, >4 Times per Week = 4
Getting Angry	Never = 1, Seldom = 1, Sometimes = 3, Usually = 4
Fruits & Vegetables	<1 Time per Week = 1, 2~5 Times per Week = 2, >5 Times per Week = 3

can be find out objectively. Above all, BP neural networks is a simulation of human nervous system, decreasing the subjectivity by finding the inner relations according to the selected training set, and it can be properly applied in the disease forecasting.

By training the BP neural networks, we can forecast the disease according the life style of the elderly people. In 1989, Hornik, Stinchcombe and White proved that a nonlinear neural network combining with three layers can approach any continuous function with any precision as long as it processes enough hidden nodes [6]. Now we assume that there are n neurons in the input layer, m neurons in the hidden layer, and one neuron in the output layer. As for this problem, the twenty factors can be treated as the input layers. Furthermore, we can determine the well-trained BP neural networks by the steps as follows [7].

- *Hidden layer stage:* The output of all neurons in the hidden layer can be described as:

$$net_j = \sum_{i=0}^{n} v_{ij} x_i, \quad j = 1, 2, \cdots, m \tag{1}$$

$$y_j = f_H\left(net_j\right), \quad j = 1, 2, \cdots, m \tag{2}$$

Here net_j is the activation value of the jth node, y_j is the output of the hidden layer, and f_H is called the activation function of a node, usually a sigmoid function as follow:

$$f_H(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

- *Output stage:* The *outputs* of all neurons in the output layer are given as follows:

$$O = f_O\left(\sum_{j=0}^{m} \omega_{jk} y_j\right) \tag{4}$$

where f_O is the activation function, usually a line function. All weights are assigned with random values initially, and they are modified by the delta rule according to the learning samples traditionally.

Then we set the output vector Y as follows:

$$Y \in \begin{cases} (0.5,1] & \text{when the interviewee is forecasted as AD patients} \\ [0,0.5) & \text{else} \end{cases} \tag{5}$$

Now we choose 730 specimens from all the 1038 interviewees randomly as the training set, select 154 specimens as the validation set, and let the 154 specimens remaining as the testing set. From the training set, we can adjust the weight and bias of each neuron to decrease the error until getting the satisfied minimum error, while the validation set can help us minimize the degree of over fitting. In the testing set, we can examine the accuracy of the BP neural network we build.

As for the number of neurons in hidden layer, we determine k, the number of neurons, by cutting and trying. After doing numbers of experiments, by using the empirical formula $k = \log_2 N$, we determine the primary value of k as 8. By the method of cut-and trial, we finally set the number of neurons in hidden layer as 65.

Before constructing the neural network, we normalize the data by using the max-min algorithm, so that we can eliminate the influence of dimension. And the algorithm can be described as follows:

$$X'_{ij} = \frac{X_{ij} - X_{i\min}}{X_{i\max} - X_{i\min}} \tag{6}$$

Consequently, the diagram of the BP neural networks structure is as shown in **Figure 1**.

At last, to judge whether an old man is going to fall in ill with the AD, we set a standard. If the output vector Y is more than 0.5, the old man will be an AD patient. Otherwise, he is not an AD patient. To solve this problem, we use the neural network toolbox in MATLAB to find the training regression, and the result is shown in **Figure 2**.
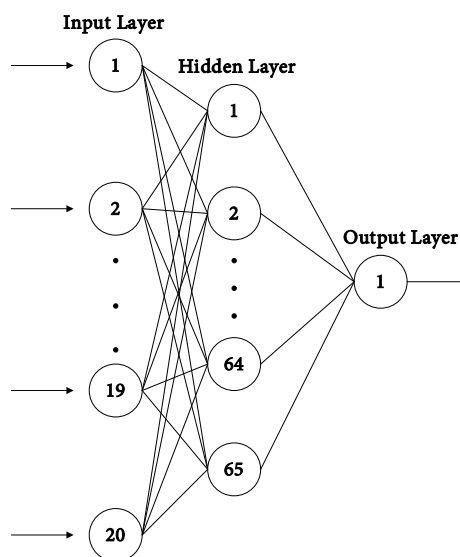

Figure 1. BP neural networks model structure.

From **Figure 2**, the linear transfer function of neurons is

$$Y = 0.86T + 0.066 \qquad (7)$$

And the correlation coefficient *R* between the predictive value and the real value is 0.96132, indicating that the prediction ability of the neural network is pretty well.

Then we check the neural network training performance, the mean squared error is shown in **Figure 3**.

As the increment of epoch during the training, the number of updates in the weights of neural network is becoming increasingly growing, and the fitted curve turns more and more overfitting. From **Figure 3**, we can determine that the best validation performance is 0.02968 at epoch 3.

In this way, we can determine the accuracy as shown on **Table 2**.

From **Table 2**, the accuracy is satisfying, so the neural network is well trained. By entering data of the elderly's early symbols, we can predict whether they will fall ill with the Alzheimer's Disease.
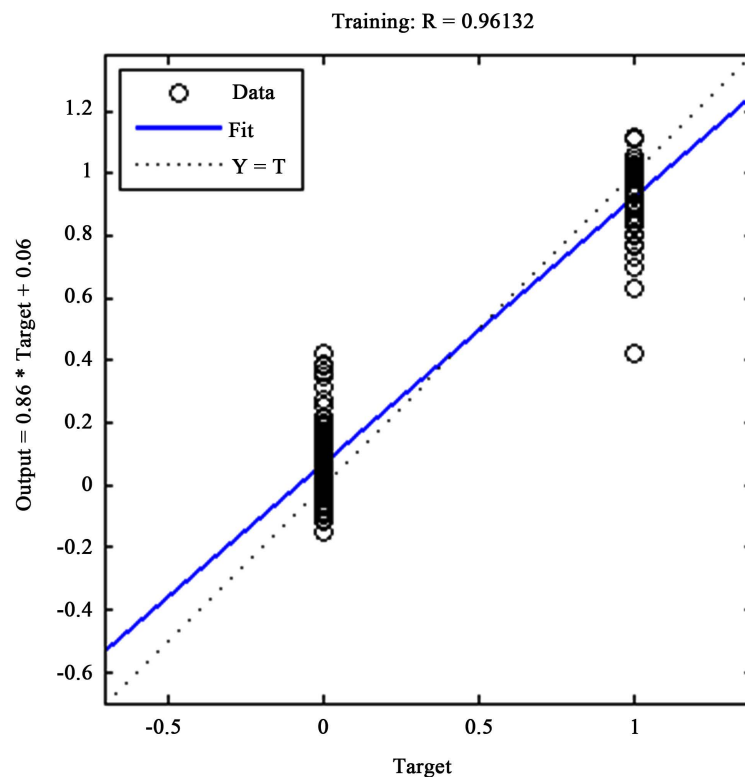

Figure 2. Neural network training regression.

Table 2. The results of BP neural networks.

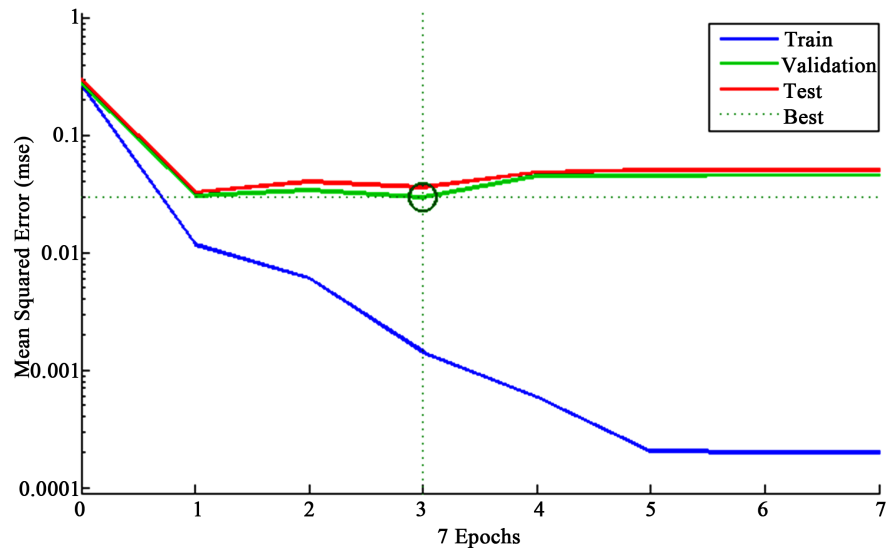| Groups	Training Set		Testing Set		Entire Sample	
	Total	Errors	Total	Errors	Total	Errors
Number of AD	161	11	48	8	209	19
Not AD	569	18	106	33	675	51
Accuracy	96%		83%		92%	

Figure 3. Neural network training performance.

2.3. Forecasting Based on Support Vector Machine

As a supervised learning model, support vector machine (SVM) can assign new examples to the two categories by a clear gap that is as wide as possible. As for this problem, to avoid over fitting, we structure the boundary conditions to relax the constraint conditions. By introducing the slack variable ξ_i, the constraint conditions can be described as follows [8]:

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i \tag{8}$$

$$s.t. \begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0, \quad i = 1, 2, \cdots, l \end{cases} \tag{9}$$

where $\xi = (\xi_1, \cdots, \xi_l)^{\mathrm{T}}$, and the parameter C refers to the penalty coefficient of the wrongly classified cases.

2.3.1. The Best Number of Training Set Samples

In the previous section, we use the convex hull theory to determine that the data is linear inseparable. Therefore, by setting the penalty factor parameter C as 1, we use radial basis function (RBF) to find the best number of the training samples. We optimize the number of training set via the grid-search method. And the accuracy is finally leveling out at around 87%. However, the tendency of the accuracy is rising up at the beginning and declining in late, it reaches its peak 6 when the number of the samples is 730, which is the optimum number of the training sample.

2.3.2. The Best Kernel Function

By using the kernel trick, SVMs can perform a non-linear classification efficiently, which can map the input factors into high-dimensional feature spaces. In the previous section, we determine the number of the training set is 730. Now

we choose the appropriate kernel function by testing the accuracy of training set and testing set via four different common kernel function: linear kernel, polynomial kernel and radial basis function. By doing 100 times experiments, we can derive the diagrams are as shown in Figures 4-6.

And the results are as Table 3.

From Table 3, RBF is the best kernel function. So we choose RBF as the kernel function, that is:
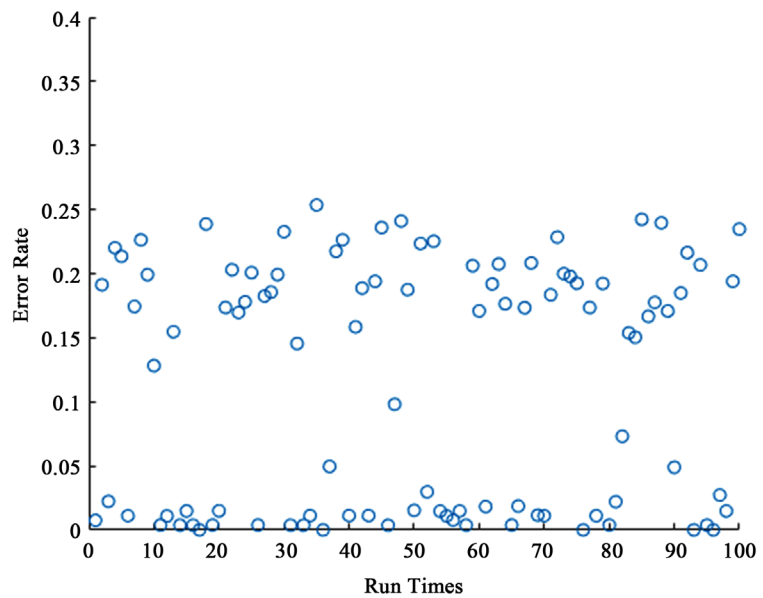
$$K = \exp\left(-\gamma \|x - x_i\|^2\right) \qquad (10)$$
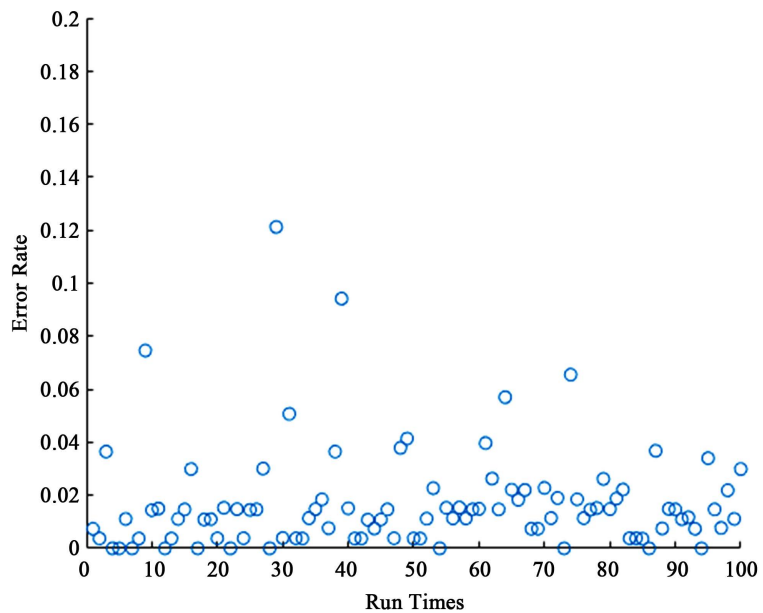

Figure 4. Error rate of linear kernel.
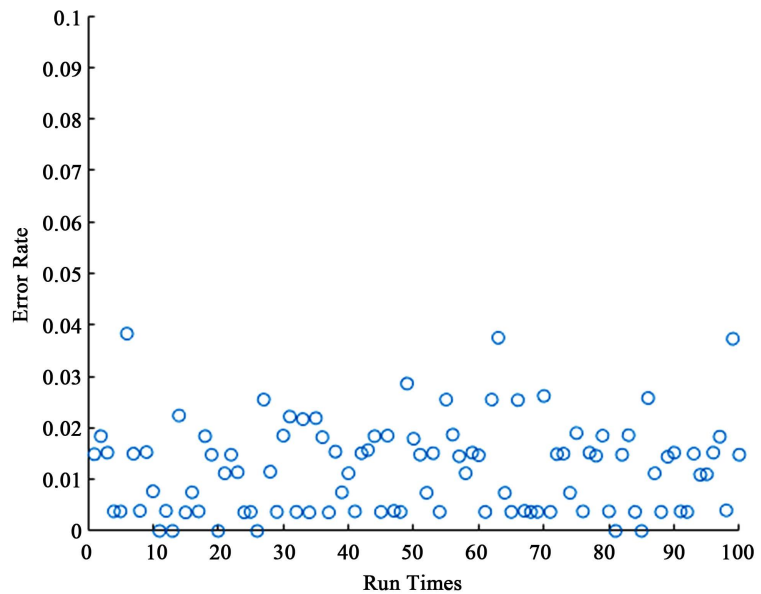

Figure 5. Error rate of cubic polynomial.

Figure 6. Error rate of RBF.

Table 3. Accuracy comparison table of four different kernel function.

Kernel Function	Linear Kernel	Polynomial Kernel	RBF
Train Accuracy	98.6%	98.2%	98.7%
Test Accuracy	98.1%	98.3%	98.9%

As for the function, the parameter γ in the RBF represents the bandwidth of the kernel function, which has great influence on the classification results. To make the problem easier, we define the value γ as 0.5 from the perspective of experience.

2.3.3. Results of SVM

From the previous section, we defined all the parameters that may be used in SVM, so that we can derive the accuracy of the prediction models applied in AD. And the prediction effect is as shown on Table 4.

From Table 4, we can derive that the accuracy of SVM is similar to that of BP neural network. So that we propose another method—random forest

2.4. Forecasting Based on Random Forest

Before introducing random forest model, let us discuss decision tree model at first. A decision tree, representing by the set of nodes, branches and leaves, can generate rules for classification. As for this problem, the endpoint nodes refer to whether an old man is an AD patient, and the other nodes refer to the living habits. To make the problem more impressive, we use random forest classification (RFC) to illustrate it.

RFC is a classification model contributed by several decision trees, and every decision tree has one classification result. Firstly, we select k samples randomly

Table 4. The Results of SVM.

Diagnosis	Forecast		Total	Accuracy
	AD	Not AD		
AD	241	26	267	90.3%
Not AD	42	729	771	94.6%
Total	283	755	1038	/

from the primary training set using *bootstrap*. Then we build k decision trees based on the k samples, so that we can derive the classification results. At last, we take votes to the classification results. The flow diagram of the algorithm is shown in **Figure 7**.

From **Figure 7**, random forest can improve the forecasting ability by constructing different training set and enlarging the difference between the classification models. After k epochs, we can get classification model series $\left\{ h_1(X), h_2(X), \cdots, h_k(X) \right\}$, by which we can construct a multi-classification model, and the criterion function is

$$H(x) = \arg \max_{Y} \sum_{i=1}^{k} I\left(h_i(x) = Y \right) \tag{11}$$

where $H(x)$ represents the multi-classification model, h_i refers to the decision tree model, Y refers to the target vector, $I(\cdot)$ is the indicative function. From the formula above, we can confirm the classification result by voting.

As for this problem, we select 692 samples from 1038 samples, then we adjust the parameter *mtry*—the numbers of the node, as well as *ntree*—the number of the trees in the forest. Now we do the control experiment to search the best value of the two parameters. Firstly, we adjust *mtry* by fixing *ntree*, the accuracy is shown in **Figure 8** after 100 times experiments.

From **Figure 8**, the error decreases when *mtry* turns from 1 to 2. Moreover, the error turns stable after 4, so we define $mtry = 4$.

Then we fix $mtry = 1$, and adjust the value of *ntree* using the method above, the result is shown in **Figure 9**.

From **Figure 9**, the error decreases when *ntree* turns from 10 to 30, but increases mildly later. Finally, we get the best result at $ntree = 80$.

In a word, we set the number of training set as 692, and we have $ntree = 80$, $mtry = 4$. The accuracy based on these parameters is 99.3%, then we plot the generalization error rate as in **Figure 10**.

From **Figure 10**, the error rate waves around 0.01, and the average error is 99.1%, proving the high classification accuracy and stability of the random forest model. Also, we can determine the importance of each factor based on contribution rate in decision tree classification, in which education has great effect on Alzheimer's Disease. Moreover, the factors such as exercise, chatting, smoking and eating habits also have pretty affection on AD.
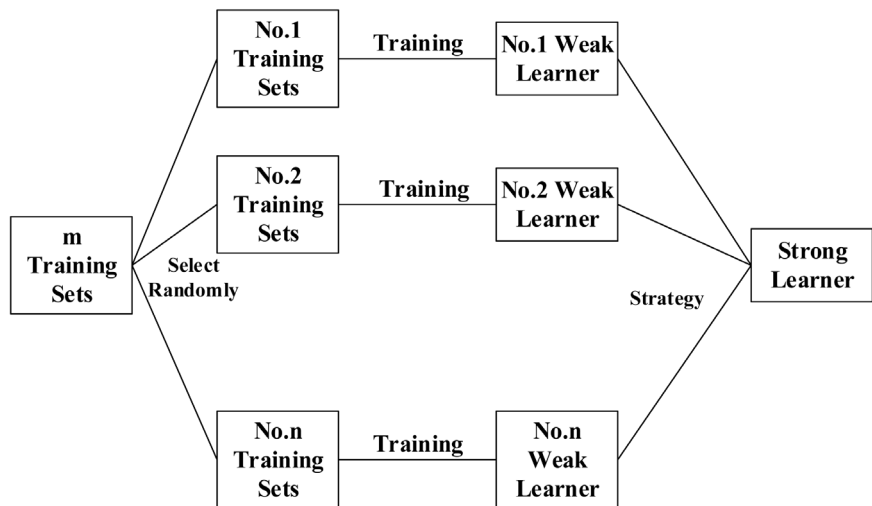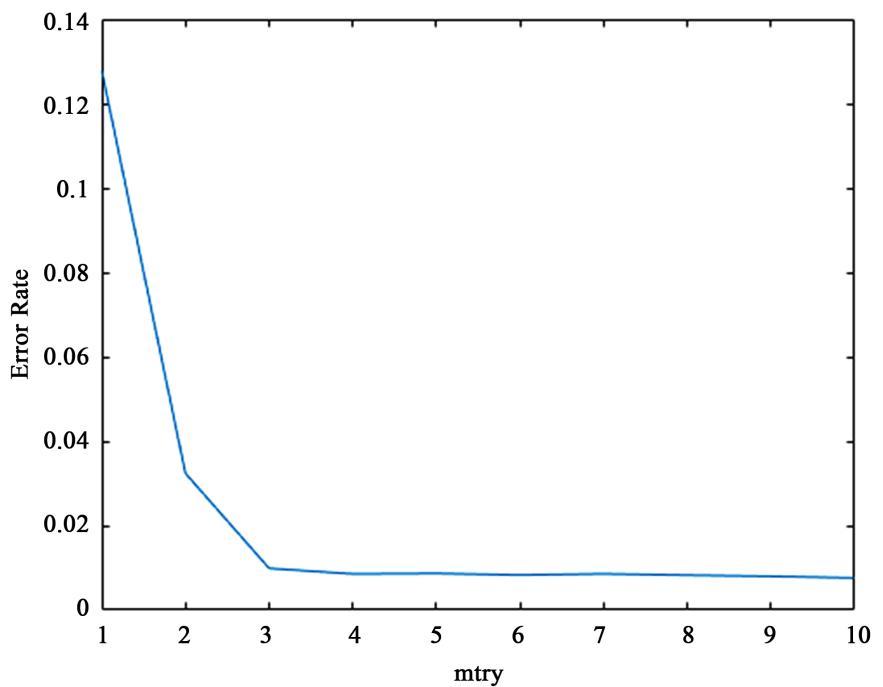
Figure 7. Algorithm flow diagram.

Figure 8. Accuracy caused by *mtry.*

3. Comparison between Three Models & Building Combination Forecasting Model

In the previous section, we predict Alzheimer's Disease by three methods: BP neural networks, SVM and random forest classification. Now we compare the three models and build a combination forecasting model based on the three machine learning models.

3.1. Data Pretreatment of the Three Models

The parameters needed by the three models are different. For BP neural network,
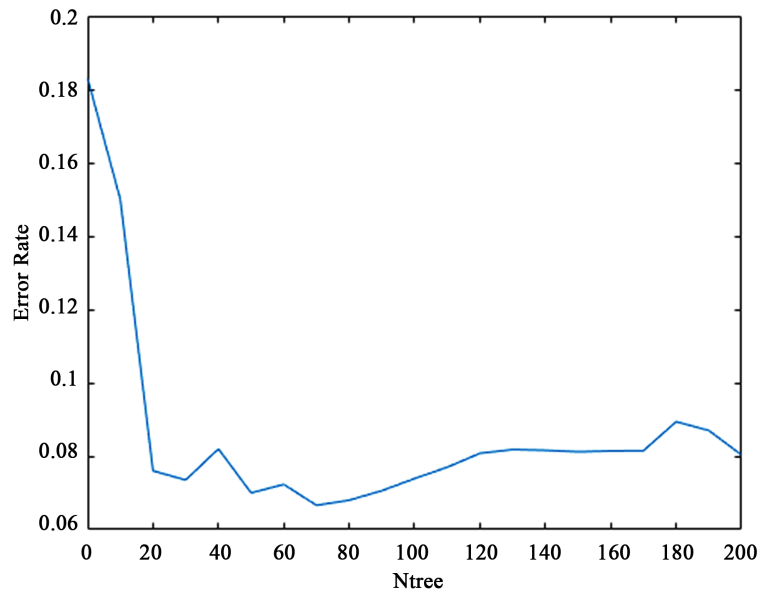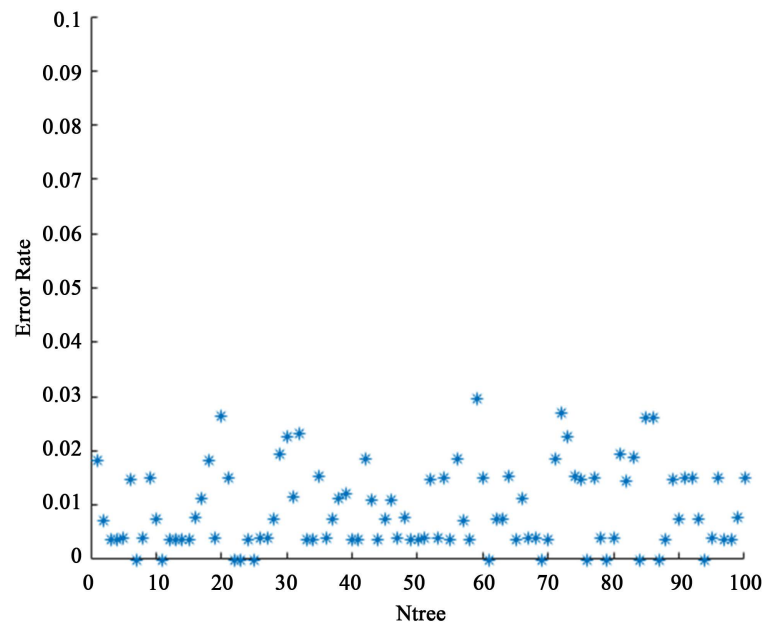
Figure 9. Accuracy Caused by *ntree.*

Figure 10. Generalization error rate.

the data should be normalized so that we can eliminate the influence of dimension. Similarly, data normalizing in SVM rely on the parameters of it. Particularly, random forest model does not have to do any data pretreatment. Thus, from the aspect of data pretreatment, random forest is the simplest method.

3.2. Parameters of the Three Models

By analyzing the parameters of the models, we can determine the sophistication of them, so that we can judge whether a model is easy to carry out.

From **Table 5**, SVM needs to define four parameters, indicating it needs more

Table 5. Parameters of the three models.

Model	How Many Parameters	Parameters
BP Neural Network	2	k, epoch
SVM	4	C, γ, degree, coef0
Random Forest	2	Mtry, ntree

extra work. BP neural network and random forest only need two parameters, so they are more easy to carry out.

3.3. Combination Forecasting Model Based on the Three Models

By comparing the three models, we can determine a final result. To improve classification accuracy, we consider the forecasting results synthetically to build a new model as follows.

At first, we define the probability formula of the three classification boxes. The probability formula of BP neural network is

$$
P = \begin{bmatrix}
e^{-\frac{E_{11}}{2\sigma^2}} & e^{-\frac{E_{12}}{2\sigma^2}} & \cdots & e^{-\frac{E_{1m}}{2\sigma^2}} \\
e^{-\frac{E_{21}}{2\sigma^2}} & e^{-\frac{E_{22}}{2\sigma^2}} & \cdots & e^{-\frac{E_{2m}}{2\sigma^2}} \\
\vdots & \vdots & \ddots & \vdots \\
e^{-\frac{E_{p1}}{2\sigma^2}} & e^{-\frac{E_{p2}}{2\sigma^2}} & \cdots & e^{-\frac{E_{pm}}{2\sigma^2}}
\end{bmatrix}
=
\begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1m} \\
P_{21} & P_{22} & \cdots & P_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
P_{p1} & P_{p2} & \cdots & P_{pm}
\end{bmatrix}
\tag{12}
$$

where E_{ij} refers to the Euclidean distance between the ith sample and the jth sample. And P refers to the initial probability matrix behind activation of neurons in the model layer Gaussian function.

$$
S = \begin{bmatrix}
\sum_{l=1}^{k} P_{1l} & \sum_{l=k+1}^{2k} P_{1l} & \cdots & \sum_{l=m-k+1}^{m} P_{1l} \\
\sum_{l=1}^{k} P_{2l} & \sum_{l=k+1}^{2k} P_{2l} & \cdots & \sum_{l=m-k+1}^{m} P_{2l} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{l=1}^{k} P_{pl} & \sum_{l=k+1}^{2k} P_{pl} & \cdots & \sum_{l=m-k+1}^{m} P_{pl}
\end{bmatrix}
=
\begin{bmatrix}
S_{11} & S_{12} & \cdots & S_{1c} \\
S_{21} & S_{22} & \cdots & S_{2c} \\
\vdots & \vdots & \ddots & \vdots \\
S_{p1} & S_{p2} & \cdots & S_{pc}
\end{bmatrix}
\tag{13}
$$

where S refers to the samples reached by the neural network summation layer belong to the initial probability of each class.

And we have

$$
p_{ij} = \frac{S_{ij}}{\sum_{l=1}^{c} S_{il}}
\tag{14}
$$

Similarly, probability formula of SVM is

$$
Pr(y=1\,|\,x) \approx P_{A,B}(f(x)) \equiv \frac{1}{1+\exp(Af(x)+B)}
\tag{15}
$$

As for random forest, the formula is

$$P_{i0} = \frac{n_{i0}}{ntree} \tag{16}$$

where n_{i0} refers to the number of judging the samples as 0. That is,

$$P_{i0} > P_{i1} \Rightarrow P_i = 0$$
$$P_{i1} > P_{i0} \Rightarrow P_i = 1 \tag{17}$$

And

$$\begin{cases} P_{i0} = \sum_{j=1}^{3} p_{ij0} \\ P_{i1} = \sum_{j=1}^{3} p_{ij1} \end{cases} \quad i = 1, 2, \cdots, n \tag{18}$$

where *i* refers to the number of samples, and *j* is the code of the three models (if $j = 1$, the model is BP neural network, else if $j = 2$, the model is SVM, else if $j = 3$, the model is random forest).

Moreover, P_{i0} is the total probability of the *i*th sample to fall ill, P_{i1} is the total probability of the *i*th sample to be healthy, and P_{ij0} is the total probability of the *i*th sample to fall ill under the *j*th algorithm.

By following the steps above, we use the advanced algorithm to find the accuracy of it, and the diagram is shown in **Figure 11**.

From **Figure 11**, the error rate based on the combination model is at around 0.7%, which is far lower than that of the other three machine learning methods solely.
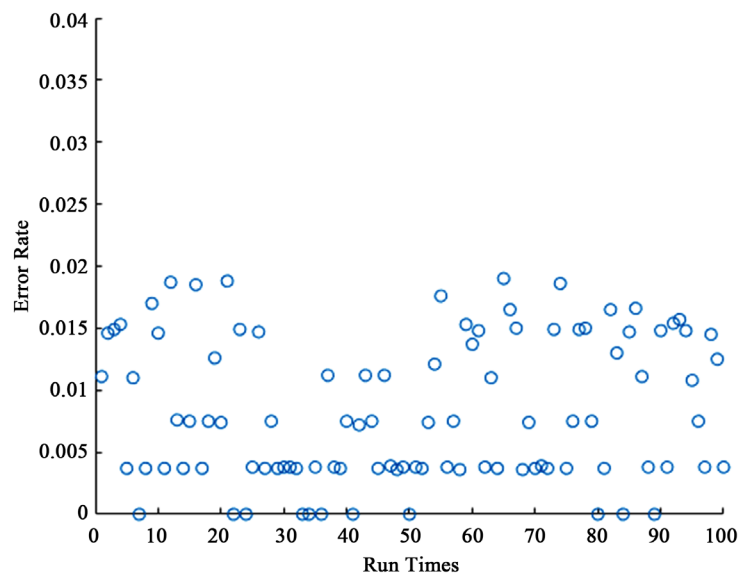

Figure 11. Error rate of combination model.

Table 6. Generalization errors of the models.

Model	BP Neural Network	SVM	Random Forest	Combination Model
G-Error	17%	1.1%	0.9%	0.7%

3.4. Accuracy of the Four Methods

In the previous section, we derived the error rate of every method. Then we compare the accuracy between the three machine learning models and the combination forecasting model as follows. By comparing the generalization error between the models, we can evaluate whether a model is accurate in forecasting AD. And the generalization errors are as shown on Table 6.

From Table 6, the combination forecasting model improves the accuracy significantly.

4. Conclusions and Suggestions

From the previous section, the three methods are all well applied in predicting Alzheimer's Disease. By adjusting the parameters, the classification results of the three methods are all improved. Among them, the random forest model is the best method, followed by SVM, and BP neural network has a more terrible effect compared with the other two models. Moreover, by building a new combination forecasting model based on the three machine learning, the error rate decreases accurately. However, we mainly discuss the accuracy in this paper without considering the complexity of the algorithm. In the future, we are going to research on this.

As for the life style, we suggest elderly people do more exercises to avoid AD. Moreover, taking group activities is also a pretty good idea. Above all, AD is an intractable problem and solving it is pretty difficult, trying our best to avoid it is all we should do.

Acknowledgements

References

[1] Alzheimer's Association (AA) (2012) Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, **8**, 131-168. https://doi.org/10.1016/j.jalz.2012.02.001

[2] Chang, C.-C., *et al.* (2012) Smoking, Death, and Alzheimer Disease: A Case of Competing Risks. *Alzheimer Disease and Associated Disorders*, **26**, 300-306. https://doi.org/10.1097/WAD.0b013e3182420b6e

[3] Reitz, C., *et al.* (2007) Relation between Smoking and Risk of Dementia and Alzheimer Disease: The Rotterdam Study. *Neurology*, **69**, 998-1005. https://doi.org/10.1212/01.wnl.0000271395.29695.9a

[4] Kumar, A. and Singh, T.R. (2017) A New Decision Tree to Solve the Puzzle of Alzheimer's Disease Pathogenesis through Standard Diagnosis Scoring System. *Interdisciplinary Sciences: Computational Life Sciences*, **9**, 107-115. https://doi.org/10.1007/s12539-016-0144-0

[5] Illan, I.A., Gorriz, J.M., *et al.* (2011) Computer Aided Diagnosis of Alzheimer's Disease Using Component Based SVM. *Applied Soft Computing*, **11**, 2376-2382. https://doi.org/10.1016/j.asoc.2010.08.019

[6] Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, **2**, 359-366.

[7] Zhang, Y.D. (2009) Stock Market Prediction of S&P 500 via Combination of Improved BCO Approach and BP Neural Network. *Expert Systems with Applications*, **36**, 8849-8854. https://doi.org/10.1016/j.eswa.2008.11.028

[8] Peker, M. (2016) A Decision Support System to Improve Medical Diagnosis Using a Combination of k-Medoids Clustering Based Attribute Weighting and SVM. *Journal of Medical Systems*, **40**, 116-132.