Artificial Neural Network-Based Predicting the Risk of Complicating Ventricular Tachyarrhythmia after Acute Myocardial Infarction During Hospitalization

Xifeng Zheng
1.Guangxi Medical University2.Affiliated Hospital of Guangdong Medical University
https://orcid.org/0000-0002-1745-9251
Weidong Nong
Guangxi Medical University
Dehui Feng
Affiliated Hospital of Guangdong Medical University
Junxian Wang
Affiliated Hospital of Guangdong Medical University
Yan He (hygxmu@163.com)
https://orcid.org/0000-0003-3020-2992

Research article

Keywords: artificial neural network, ventricular tachyarrhythmia, acute myocardial infarction

DOI: https://doi.org/10.21203/rs.3.rs-39562/v1

Abstract

An artificial neural network (ANN) model was developed to predict the risks of complicating ventricular tachyarrhythmia (VTA) in patients with acute myocardial infarction (AMI). We enrolled information of 503 patients with 13 risk factors from the affiliated hospital of Guangdong medical university from January 2017 to December 2019. Risk factors were dimensionally reduced and simplified as new variables by principal component analysis (PCA). The cohort were randomly divided into a training set and a testing set at the ratio of 70%:30%. Training set was used to develop a model for the prediction of VTA while testing set was used to evaluate the performance of the model. Three new comprehensive variables by PCA are able to reflect all information of the original data. We determined the prediction model with optimizing parameters by cyclic searching which includes an input layer of three comprehensive variables, a single hidden layer composed of two neurons and a output layer. The area under curve (AUC) is 0.812 in training set and confusion matrix with accuracy 94.60%, sensitivity 63.04%, specificity 99.35%, positive predicative value 93.55%, negative predictive value 94.70%. The model displayed a decreased but medium discrimination with an AUC of 0.688 in the independent testing cohort, confusion matrix with accuracy 87.42%, Sensitivity 39.26%, specificity 98.37%, positive predicative value 84.62%, negative predictive value 87.68%. The research suggests that ANN model could be used to predict the risk of complicating ventricular tachyarrhythmia after acute myocardial infarction while should be further improved.

1. Introduction

Ventricular tachyarrhythmia (VTA) is a common but fatal complication of acute myocardial infarction (AMI), especially in elderly patients[1]. It is also the main cause of sudden cardiac death(SCD) and seizes for 4–5 million lives per year globally[2]. According to reports of the American Heart Association in 2015, the survival to hospital discharge after suffering SCD was estimated to be 23.8%-35.9%[3]. Considering the extremely poor prognosis, it is of great significance to identify the high-risk patients in early stage.

In the booming era of big data and artificial intelligence, machine learning has become a useful and popular scientific research method in data analysis and prediction[4]. Artificial neural network (ANN) is a kind of machine learning methods, which copies the biologic neural system sample and consists of an input layer, signal or multiple hidden layers and an output layer. The algorithm has abilities to analysis data in nonlinear relationship and not necessitate to notice the distributional assumptions (such as normality).These advantages have caused considerable interests in medical research and used to establish disease prediction models in cardiology, sleep medicine and oncology[5–8].The purpose of our study is to construct a prediction model based on cardiovascular risk factors by ANN to identify the high-risk patients that may complicate of VTA after AMI.

2. Methods

2.1 Study design and data collection

The ethical review was approved by the ethics committee of the affiliated hospital of Guangdong medical university and patient informed consent was waived for this retrospective analysis.

We enrolled the information of the patients from the electronic medical record database from January 2017 to December 2019 who were clearly diagnosed as AMI, fulfilling the Fourth Universal Definition of Myocardial Infarction (2018)[9]. Arrhythmic events were recorded by reviewing electrocardiogram or Holter monitor. VTA were defined as sustained ventricular tachycardia, ventricular fibrillation that resulting in defibrillator shocks and non-sustained ventricular tachycardia. Features and subclassification about risk factors are listed in Table-1:

Table-1: Risk factors considered in the study

Risk factors	Features and subclassification
General information	
Age(years)	<60y, between 60y-75y,>75y
Gender	Female/Male
Diabetes history	Yes/No
Hypertension history	Yes/No
Information about AMI	
Type of AMI	STEMI/NSTEMI
hsTnT	Not more than 5 times threshold
	More than 5 times threshold
	More than 10 times threshold
PCI timing	Hospitalized in 24h
	Hospitalized more than 24h
	Hospitalized without PCI
NT-proBNP	Normal
	Not more than 5 times threshold
	More than 5 times threshold
EF	>50%, between 40%-50%, <40%
Hypokalemia	Yes/No
Relevant organs information	
Infection (pneumonia or catheter-related infection)	Yes/No
eGFR(kidney)	>60ml/min, between 30-60 ml/min, <30% ml/min

2.2 Statistical analysis and Model established

Statistical analysis and model construct were performed by R3.6.1 software. R packages “psych”, “ggplot2”, and “pheatmap” were used to execute and visualize the principal component analysis (PCA). The package “neuralnet”, "NeuralNetTools", "dplyr" and " pROC " were used to develop and validate artificial neural network model. The process of analysis is as follows steps:

2.2.1. Extracting the features from the original variables by principal component analysis (PCA) which is a multivariate statistical method with a long history and a widely used range. The principal components can reflect mostly or all information of the original data while each variable is independent from others, avoiding multiple collinearity and helpful to develop a model[10].

2.2.2. The cohort were randomly divided into a training set and a testing set at the ratio of 70%:30%. A standard feed-forward, back-propagation neural network is the simplest form of ANNs that consisting of an input layer, a hidden layer, and an output layer was applied in the study due to its relative simplicity and stability[11]. The operation process of the model is as follows: the new comprehensive principal components were introduced from input layer to the hidden layer, which consists of several neurons as information receiver. All the neuron connections have a different weights and bias parameter. The former one represents the importance of the corresponding input compared with other inputs, the latter one is used to correct the calculation results of the weight and input. The information is transformed to nonlinearly by the sigmoid activation function and passed into output layer that calculates results whether complicating of VTA. It should be noted that the optimal number of neurons in hidden layer was determined through trial and error, since no accepted theory currently exists for predetermining the optimal number. We use cycle searching to determine the optimal number of neurons of the model in the study. The mathematical operations in the model can be generalized as follows[12]:

$$y=\text{Activation}\left(\sum_{i}^{N} wi \cdot xi + b\right)$$

Note: y = output result, i = number of input variables, N = number of neurons, w = weights, x = input variables, b = bias parameter

2.2.3. The optimized model was verified in the training dataset and testing dataset respectively, with following parameters as the assessment tool: area under receiver operating characteristic curve (AUC)[13] and confusion matrix with accuracy, sensitivity, specificity, positive predicative value, and negative predictive value.

3. Result

3.1 Characteristics of clinical information

All data records were expressed as count (%). A total of 503 patients with 13 risk factors, according to the presence or absence of the complications of VTA, were divided into the VTA group and the non-VTA group. The detailed information is shown in Table 2.

Table 2
Baseline characteristics of the risk factors and subclassification

Risk Factors	N (%)		
	VTA 74(14.71%)	non-VTA 429(85.29%)	Total 503(100%)
1.Age(years)			
< 60y	9(12.16%)	62(14.45%)	71(14.12%)
60y-75y	32(43.24%)	212(49.42%)	244(48.51%)
> 75y	33(44.60%)	155(36.13%)	188(37.37%)
2.Gender			
Female	25(33.78%)	151(35.20%)	176(34.99%)
Male	49(66.22%)	278(64.80%)	327(65.01%)
3.Type of AMI			
ST-segment elevator MI	36(48.65%)	202(47.09%)	238(47.32%)
Non-ST-segment elevator MI	38(51.35%)	227(52.91%)	265(52.68%)
4.Lesion vessels			
Single	22(29.73%)	228(53.15%)	250(49.70%)
Double	24(32.43%)	125(29.14%)	149(29.62%)
Triple	28(37.84%)	76(17.71%)	104(20.68%)
5.PCI timing after hospitalized			
Hospitalized in 24 h	29(39.19%)	197(45.92%)	226(44.93%)
Hospitalized more than 24 h	15(20.27%)	177(41.26%)	192(38.17%)
Hospitalized without PCI	30(40.54%)	55(12.82%)	85(16.90%)
6.Hypokalemia			
No	67(90.54%)	403(93.94%)	470(93.44%)
Yes	7(9.46%)	26(6.06%)	33(6.56%)
7.Diabetes			
No	51(68.92%)	336(78.32%)	387(76.94%)
Yes	23(31.08%)	93(21.68%)	116(23.06%)

Risk Factors	N (%)		
	VTA 74(14.71%)	non-VTA 429(85.29%)	Total 503(100%)
8.Hypertension			
No	46(62.16%)	252(58.74%)	298(59.24%)
Yes	28(37.84%)	177(41.26%)	205(40.76%)
9.Infection			
No	44(59.46%)	370(86.25%)	414(82.31%)
Yes	30(40.54%)	59(13.75%)	89(17.69%)
10.NT-proBNP			
Normal	12(16.22%)	125(29.14%)	137(27.24%)
Not more than 5 times threshold	33(44.59%)	231(53.85%)	264(52.48%)
More than 5 times threshold	29(39.19%)	73(17.01%)	102(20.28%)
11.EF			
> 50%	30(40.54%)	270(62.94%)	300(59.64%)
40%-50%	27(36.49%)	143(33.33%)	170(33.80%)
< 40%	17(22.97%)	16(3.73%)	33(6.56%)
12.hsTnT			
Not more than 5 times threshold	10(13.51%)	99(23.08%)	109(21.67%)
More than 5 times threshold	14(18.92%)	154(35.90%)	168(33.40%)
More than 10 times threshold	50(67.57%)	176(41.02%)	226(44.93%)
13.eGFR			
> 60 ml/min	38(51.35%)	374(87.18%)	412(81.91%)
30–60 ml/min	13(17.57%)	42(9.79%)	55(10.93%)
< 30 ml/min	23(31.08%)	13(3.03%)	36(7.16%)

3.2 Principal component analysis

The risk factors were simplified into three new principal components by PCA dimensionality reduction that include all information of the original data reflect by cumulative variance and visualization as scree plot (Figure-1a). The score values of risk factors that reflected by the new variables are shown as the clustering heat map (Figure-1b).

3.3 Model construction and parameter adjustment

We test the parameters by comparing the prediction accuracy in different neuron models and draw a conclusion that the model contains two neurons is suitable with the best accuracy reaching 87.42% (Figure-2a). Finally, We conform that the optimal artificial neural network model which includes a input layer with three variables, a single hidden layer composed of two neurons and a output layer to display the result (Figure-2b).

3.4. Evaluation of model performance

The Receiver operating characteristic (ROC) curves showed that the result of model in training dataset had a promising discrimination with AUC of 0.812 (Figure-3A). Confusion matrix was show as Table-3 with Accuracy 94.60%, sensitivity 63.04%, specificity 99.35%, positive predicative value 93.55%, negative predictive value 94.70%. In the independent testing cohort, the model displayed a decreased but medium discrimination with an AUC of 0.688 (Figure-3B), Confusion matrix was show as Table-4 with Accuracy 87.42%, sensitivity 39.26%, specificity 98.37%, positive predicative value 84.62%, negative predictive value 87.68%.

Table-3: Confusion matrix of training dataset

Training dataset predict			
Status	non-VTA	VTA	Row total
non-VTA	True Negative	False positive	306
	304	2	
VTA	False Negative	True positive	46
	17	29	
Colum total	321	31	352

Table-4: Confusion matrix of testing dataset

Training dataset predict			
Status	non-VTA	VTA	Row total
non-VTA	True Negative	False positive	123
	121	2	
VTA	False Negative	True positive	28
	17	11	
Colum total	138	13	151

4. Discusstion

VTA is a common but fatal complication of AMI. Although there are some electrocardiographic and imaging abnormalities, such as EF < 35%, long Q-T interval and R on T wave, may have a close relationship with VTA[14, 15]. Unfortunately, there is no generally accepted prediction model to identify high-risk patients accompanied with ventricle tachyarrhythmia after acute myocardial infarction. Our team attempts to construct a prediction model and make a contribution to this field.

The artificial neural network is a representative method of artificial intelligence and an ideal disease prediction model for being able to analyze data in nonlinear relationship and to predict a complex relationship between variables. The earliest application of ANN in cardiology dates back to at least 1995, with the development of computer technology and the widely application of deep learning (A new research direction in artificial intelligence).Nowadays, ANN has attracted more and more attention again[16, 17].

The suitable feature variables form the basement to the model. PCA is a multivariate statistical method by transforming multiple variables into a few less new comprehensive variables with orthogonal transformation to achieve the purpose of reducing dimension and to simplify data structure. The ANN model started from three principal components that simplify from 13 risk factors by PCA. The cumulative variance of the three principal components add up to 100% that reflects the information of the original data perfectly. We also had tried to screen feature variables by least absolute shrinkage and selection operator (LASSO) regression[18]. The result selected by Lasso are infection, eGFR, lesion vessel, hsTnT, EF, PCI timing that similar to PC1, but the rest information was missing. We prefer to screen the feature variables by PCA for the fewer variables but more comprehensive information than lasso regression.

The results of the prediction indexes in training set and testing set are considerable except for the sensitivity. According to the definition as sensitivity shown by mathematical:

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False Negative}}$$

The reason for unsatisfied sensitivity mainly because of poor true positive predictive value that caused by the imbalanced data proportional, the VTA group with 74 patients counts for 14.71% while the non-VTA group with 429 patients counts for 85.29%. In the process of prediction, the model tends to judge the patients as non-VTA group because of the higher probability accuracy. Although we have preferential model parameters and take some measures to deal with unbalanced data, such as Synthetic Minority Oversampling Technique (SMOTE)[19], but the effect was rarely. Although the high conservatism of the model to true positive results is in low sensitivity, it does not mean that the model is a failure. Once the patients were judged as true positive (VTA) by the model, the reliability (positive prediction rate) was credible, which reached 93.55% in the training set and 84.62% in the test set. This suggests that doctors need to pay more attention to these really high-risk patients.

5. Limitation

There are several limitations in the study. First of all, risk factors mainly focus on traditional clinical diagnosis and treatment for we lack of information about electrophysiological. Secondly, training and testing samples original from the same cohort. The prediction performance of the model has not been verified in other populations. Finally, the application of ANN model is not as convenient as nomogram model, because it depends on the specific computer program.

6. Conclusion

Despite of the limitations, our research introduced the method of machine learning that applied to classification into cardiovascular medicine and have developed an ANN model to predict the risk of complicating VTA after AMI. It is necessary to apply the interdisciplinary guiding ideology to medical research under the background of the rapid development in artificial intelligence area.

Abbreviations

Artificial neural network = ANN, Ventricular tachyarrhythmia = VTA, Acute myocardial infarction = AMI, ST elevate myocardial infarction = STEMI, non-ST elevate myocardial infarction = NSTEMI, Percutaneous coronary intervention = PCI, Ejection fraction = EF, N-Terminal Pro-B-Type Natriuretic Peptide = NT-proBNP, High-sensitive Troponin T = hsTnT, estimated glomerular filtration rate = eGFR, Principal component analysis = PCA, Receiver operating characteristic = ROC, Area under the receiver operating characteristic curve = AUC

Declarations

Acknowledgements

None.

Ethics approval and consent to participate

Ethical review was approved by the ethics committee of the affiliated hospital of Guangdong medical university. The research was a retrospective study and researchers would try their best to protect the information from disclosure.

Funding

None.

Competing interests

The authors declare that they have no conflicts of interest.

Authors' contributions

Xifeng Zheng and Yan He were involved in conception and design of the study. Xifeng Zheng and Weidong Nong were involved in analysis of the data. Dehui Feng and Junxian Wang were involved in collection of the data. Yan He provided scientific supervision. All authors reviewed and approved the final manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not Applicable.

References

1. Myerburg RJ and Junttila MJ. Sudden cardiac death caused by coronary heart disease. *Circulation*. 2012;125:1043-52.
2. Chugh SS. Sudden cardiac death in 2017: Spotlight on prediction and prevention. *Int J Cardiol*. 2017;237:2-5.
3. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J, Fornage M, Gillespie C, Isasi CR, Jimenez MC, Jordan LC, Judd SE, Lackland D, Lichtman JH, Lisabeth L, Liu S,

Longenecker CT, Mackey RH, Matsushita K, Mozaffarian D, Mussolino ME, Nasir K, Neumar RW, Palaniappan L, Pandey DK, Thiagarajan RR, Reeves MJ, Ritchey M, Rodriguez CJ, Roth GA, Rosamond WD, Sasson C, Towfighi A, Tsao CW, Turner MB, Virani SS, Voeks JH, Willey JZ, Wilkins JT, Wu JH, Alger HM, Wong SS, Muntner P, American Heart Association Statistics C and Stroke Statistics S. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation*. 2017;135:e146-e603.

4. Shameer K, Johnson KW, Glicksberg BS, Dudley JT and Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018;104:1156-1164.
5. Tang ZH, Liu J, Zeng F, Li Z, Yu X and Zhou L. Comparison of prediction model for cardiovascular autonomic dysfunction using artificial neural network and logistic regression analysis. *PLoS One*. 2013;8:e70571.
6. Li H, Luo M, Zheng J, Luo J, Zeng R, Feng N, Du Q and Fang J. An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. *Medicine (Baltimore)*. 2017;96:e6090.
7. Li A, Quan SF, Silva GE, Perfect MM and Roveda JM. A Novel Artificial Neural Network Based Sleep-Disordered Breathing Screening Tool. *J Clin Sleep Med*. 2018;14:1063-1069.
8. Zhao Y, Zhao L, Mao T and Zhong L. Assessment of risk based on variant pathways and establishment of an artificial neural network model of thyroid cancer. *BMC Med Genet*. 2019;20:92.
9. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, White HD and Executive Group on behalf of the Joint European Society of Cardiology /American College of Cardiology /American Heart Association /World Heart Federation Task Force for the Universal Definition of Myocardial I. Fourth Universal Definition of Myocardial Infarction (2018). *J Am Coll Cardiol*. 2018;72:2231-2264.
10. Jolliffe IT and Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374:20150202.
11. Grossi E. How artificial intelligence tools can be used to assess individual patient risk in cardiovascular disease: problems with the current methods. *BMC Cardiovasc Disord*. 2006;6:20.
12. Penny W and Frost D. Neural networks in clinical medicine. *Med Decis Making*. 1996;16:386-98.
13. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013.
14. Glinge C, Sattler S, Jabbari R and Tfelt-Hansen J. Epidemiology and genetics of ventricular fibrillation during acute myocardial infarction. *J Geriatr Cardiol*. 2016;13:789-797.
15. Bhar-Amato J, Davies W and Agarwal S. Ventricular Arrhythmia after Acute Myocardial Infarction: 'The Perfect Storm'. *Arrhythm Electrophysiol Rev*. 2017;6:134-139.
16. Renganathan V. Overview of artificial neural network models in the biomedical domain. *Bratisl Lek Listy*. 2019;120:536-540.
17. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E and Dudley JT. Artificial Intelligence in Cardiology. *J Am Coll Cardiol*. 2018;71:2668-2679.

18. Pavlou M, Ambler G, Seaman S, De Iorio M and Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016;35:1159-77.
19. Bhattacharya M, Lu DY, Kudchadkar SM, Greenland GV, Lingamaneni P, Corona-Villalobos CP, Guan Y, Marine JE, Olgin JE, Zimmerman S, Abraham TP, Shatkay H and Abraham MR. Identifying Ventricular Arrhythmias and Their Predictors by Applying Machine Learning Methods to Electronic Health Records in Patients With Hypertrophic Cardiomyopathy (HCM-VAr-Risk Model). *Am J Cardiol*. 2019;123:1681-1689.

Figures

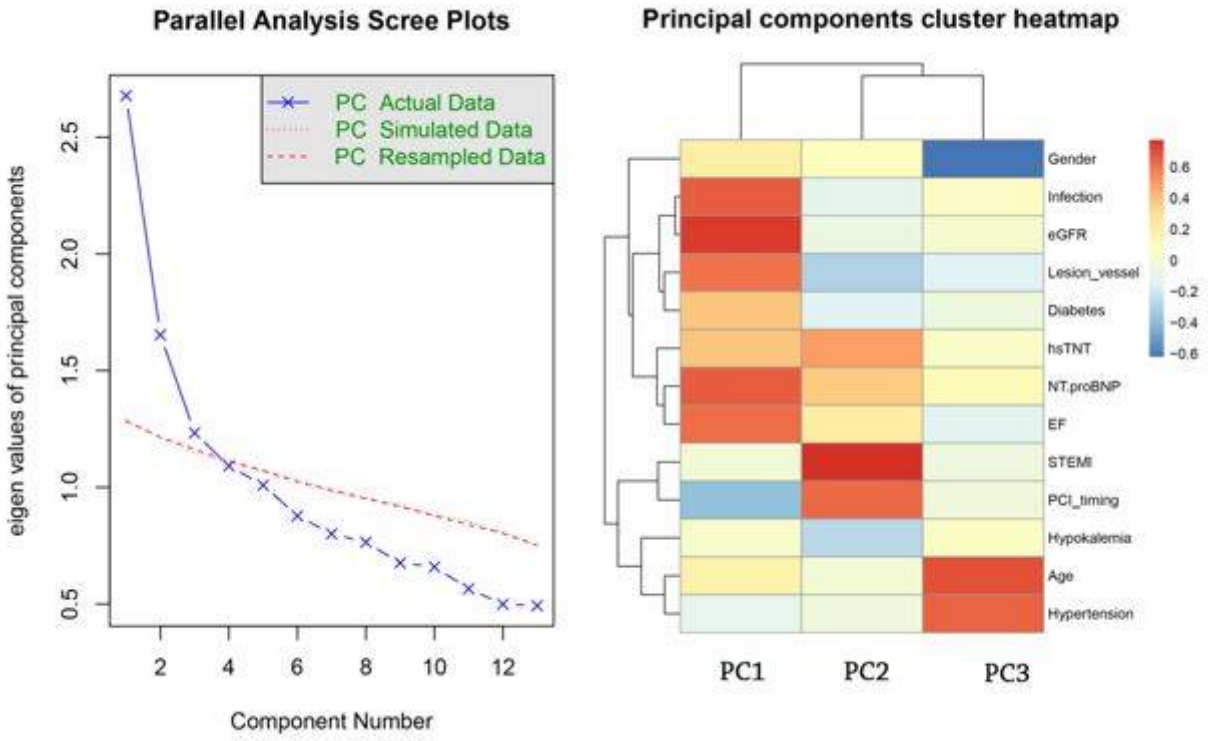

Figure 1

Scree plots and cluster heat map of PCA Note: Scree plots indicated three principal components as new comprehensive variables. PC1 reflects the information of six variables that including infection, eGFR, lesion vessel, diabetes, hsTnT, NT-proBNP and EF. PC2 reflects the information about STEMI, PCI-timing and hypokalemia while PC3 reflects the information about age, hypertension and gender.

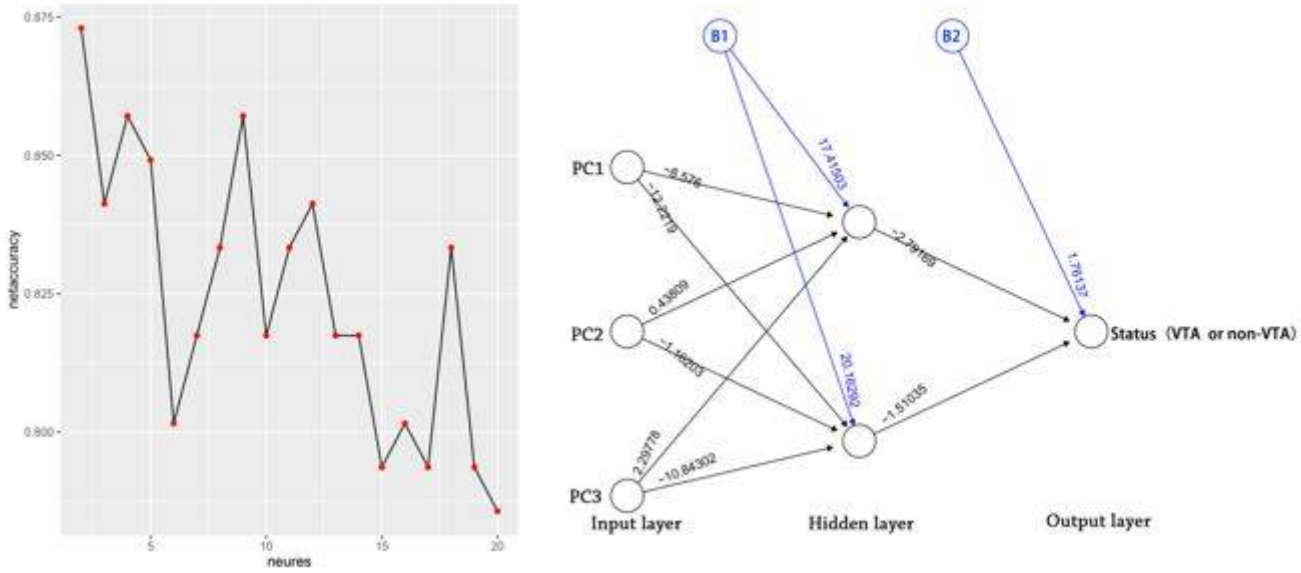

Figure 2

The optimized neurons in hidden layer and ultimate ANN model NOTE: The optimal ANN model consists of an input layer, a hidden layer, and an output layer. All the neuron connections have multiplying weights and bias parameter (B1 and B2) associated with them.

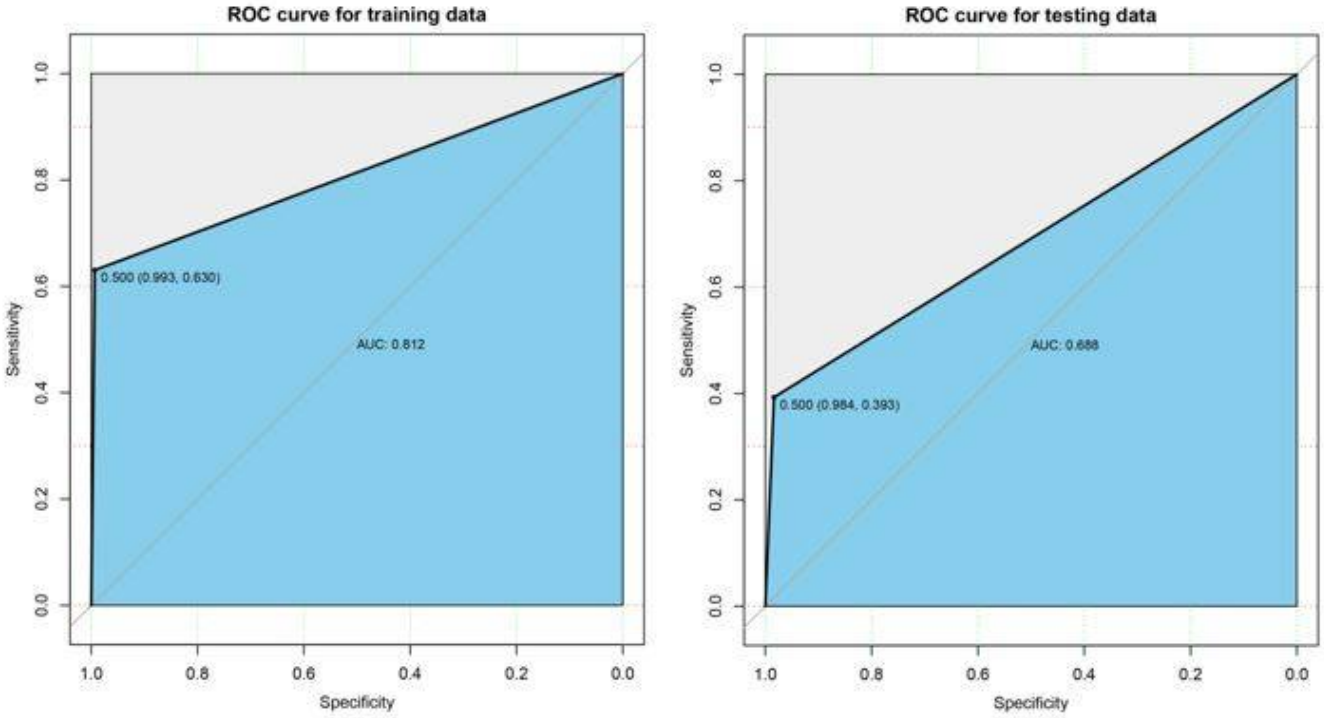

Figure 3

ROC curve for training data and testing data