Innovation and Technology Studies
ISSN: 3007-6927
DOI: https://doi.org/10.61784/its3010

PTSD CHARACTERISTICS OF RESCUERS BASED ON DECISION TREE

WenBei Zheng*, Xi Yang, WenYun Xia
Guangxi Normal University, Guilin 541000, Guangxi, China.
Corresponding Author: WenBei Zheng , Email: 1181122711@qq.com

Abstract: Post-traumatic stress disorder (PTSD) is a mental disorder caused by traumatic events, characterized by recurrence of memories and nightmares, avoidance of stimulation, difficulty in regulating emotions, and persistent hypervigilance. Due to the particularity of their work, rescuers are often more likely to suffer from PTSD. Identifying and clarifying the influencing factors of PTSD, establishing an effective prediction model for rescuers, and providing effective evaluation tools to provide targeted treatment and support measures for rescuers are of great significance in helping rescuers cope with and prevent PTSD. Based on the data set provided by the PLA Medical College, this study used a decision tree classification model to construct an effective prediction model for rescuers, identify important variables that affect rescuers' PTSD, and evaluate the prediction efficiency of the model by the receiver operating characteristic curve (ROC). Results: The accuracy was 95.56%, the sensitivity was 93.75%, the specificity was 95.69%, the false positive rate was 4.31%, the F1 score was 75%, and the AUC value was 94.72%. The features classified by the model were ranked according to their importance. The top eight features of the decision tree model were: ASD alertness, ASD avoidance, ASD re-experience, ASD separation, ASD nature, smoking status, psychological resilience, and age. Conclusions: The decision tree model has high accuracy and stability and can be used to guide clinical prevention and treatment.
Keywords: PTSD; Decision tree model; Influencing factors; Feature selection

1 INTRODUCTION

Post-traumatic stress disorder(PTSD) refers to a strong psychological reaction such as helplessness, fear, anxiety or disgust caused by an abnormally threatening or catastrophic event[1]. A traumatic event is an experience that threatens the safety or death of an individual or others. Nearly 70% of the population is exposed to at least one traumatic event in their lifetime, and 31% will experience about 4 traumatic events[2]. A considerable number of people will suffer from PTSD, with a lifetime prevalence rate between 6.1% and 9.2%. The nature of the work of rescue workers often exposes them to extreme environments and traumatic events. If they are not intervened and treated in time, they are often more likely to suffer from PTSD[3,4]. Data on post-traumatic stress disorder (PTSD) among rescuers vary by region, year, and research method. Here are some examples of relevant research results: a) In a study of American firefighters, about 7% of firefighters were diagnosed. b) In Canada, a study of firefighters and emergency personnel found that about 10% of respondents were diagnosed. c) A study of Australian rescuers found that about 5% of rescuers suffered from post-traumatic stress disorder. The actual prevalence may vary due to multiple factors, such as the selection of research samples, differences in diagnostic criteria, etc. In addition, many rescuers may not actively seek help or be correctly diagnosed, so the actual prevalence may be underestimated. The nature and severity of the traumatic event, the individual's psychological characteristics and coping ability, social support and environmental factors, and biological factors are all important factors affecting PTSD[5]. To date, there have been many studies on the factors affecting PTSD, but there are very few studies on the importance of these factors.
Machine learning (ML), as one of the core technologies of artificial intelligence, is a method of summarizing features and patterns of large amounts of data[6]. Due to its powerful data processing and mining capabilities, machine learning can help identify the most relevant features and evaluate the importance of features, thereby determining which factors have a significant impact on a phenomenon or result. Therefore, this paper aims to use the decision tree classification model to construct an effective prediction model for rescuers and identify important variables that affect rescuers' post-traumatic stress disorder, provide targeted treatment and support measures for rescuers, and help rescuers cope with and prevent PTSD.

2 THEORETICAL OVERVIEW

Decision tree is an important classification and regression method in data mining technology. It is a prediction model expressed in the form of a tree structure. It recursively divides the features in the data set to build a tree-like decision model to achieve the classification or numerical prediction of samples[7]. Each leaf node in the tree represents a classification result, and different branch paths represent different classification choices[8]. The main strategic idea of building a decision tree is to divide and conquer from top to bottom, that is, in the recursive process from root to leaf, find a "partition" attribute at each intermediate node to divide the data set into smaller and smaller subsets until the conditions are met. Decision tree structure diagram can be seen in figure 1.

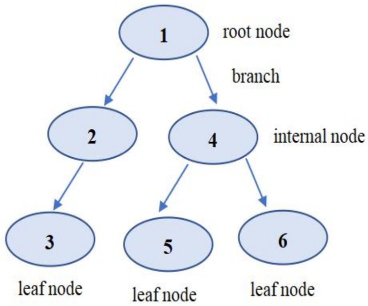

Figure 1 Decision Tree Structure Diagram

In summary, the core of the decision tree algorithm is to select the optimal attribute division. In order to achieve the optimal attribute division, the samples contained in the internal nodes of the decision tree are usually as much as possible belonging to the same category, that is, the "purity" of the node is as high as possible. Information entropy is the basic indicator for measuring the "purity" of a sample set, and the calculation formula is as follows:

$$Ent(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \tag{1}$$

It is assumed that the proportion of the k-th class samples in the current sample set D is P_k . The smaller the value of information entropy is, the higher the "purity" is.

3 DATA DESCRIPTION

The data used in this paper is provided by the PLA Medical College. This dataset contains the basic information of rescuers, with a total of 903 records, each of which represents the personal data of a rescuer. The variables in the dataset include but are not limited to demographic information, traumatic experience, psychological state, and behavioral habits.

Before analysis and modeling, the data is first processed appropriately. In order to enable the model to effectively identify and utilize all non-numerical variables, a label encoding strategy will be adopted instead of one-hot encoding, which introduces high-dimensional sparse features and may lead to dimensionality disasters. Label encoding divides the variables into levels according to the health risk or severity they reflect, allowing the model to more accurately identify the key factors affecting PTSD. For continuous variables, standardization will be used, 0 and 1 label encoding will be used for binary variables, and sequential integer encoding will be used for multi-classification variables. For example, the gender variable is coded as {1, male} and {0, female}, which can simplify data processing and is more convenient in statistical analysis. The educational level variable is coded as {3, college and above}, {2, high school}, {1, technical secondary school and below}. This coding method reflects the increasing relationship of education level and is helpful for analyzing the impact of education level on PTSD.

For the study of PTSD characteristics, the recursive feature elimination cross-validation (RFECV) method will be used for feature selection. First, the comprehensive variable ASD total score and the response variable PCL total score are deleted. This is because the ASD total score is the sum of other ASD-related variables (such as ASD dissociation, ASD re-experience, ASD avoidance, and ASD vigilance), which contains information about other ASD variables. In addition, the PCL total score has been converted into a PTSD nature variable and used as a response variable. Therefore, in order to avoid information duplication and multicollinearity, it was decided not to include these two variables in the feature selection process. Then, recursive feature elimination cross-validation (RFECV) was performed to obtain the optimal number of features and the corresponding accuracy of the model. Figure 2 shows the change in model accuracy as the number of features increases. It can be seen that as the number of features increases, the accuracy of the model shows an overall upward trend. When 25 features are used, the accuracy is the best, which is 0.96457. According to the result data, the accuracy of the model is significantly improved for the first time when 8 features are used, reaching 0.96455, and remains relatively stable at the subsequent number of features. The accuracy of the model reached the highest when using 25 features, but it was very close to the accuracy when using 8 features. In order to study the different effects of more variables on PTSD, we chose to retain all the variables in the model.

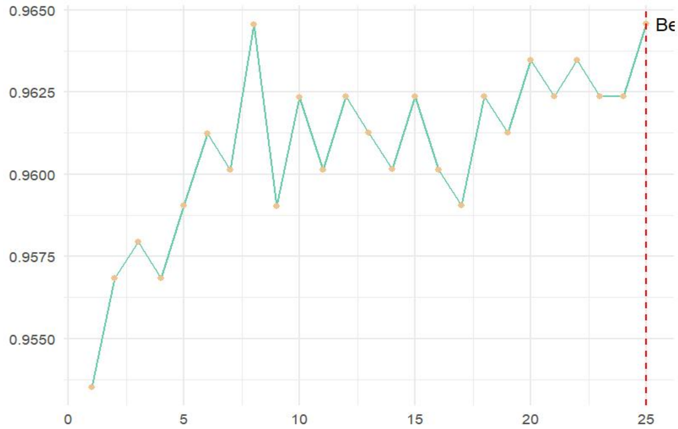

Figure 2 The Optimal Number of Features for the Model

Finally, the data set is divided reasonably, and the oversampling technique is applied to the training set. First, the createDataPartition function is used to divide the data set into a training set and a test set at a ratio of 75%. In the original data set, there are 837 individuals without PTSD and 66 individuals with PTSD. After the division, the training set contains 628 individuals without PTSD and 50 individuals with PTSD, while the test set contains 209 individuals without PTSD and 16 individuals with PTSD. Since the number of individuals with PTSD in the training set is relatively small, in order to avoid the impact of class imbalance on model training, the training set is oversampled. The ovun.sample function is used to increase the number of minority class (PTSD) samples in the training set to 272, so that the number of samples without PTSD and with PTSD in the training set is 628 and 272 respectively. Through this data balancing strategy, the model can fully learn the characteristics of minority class samples during the training process, improve the recognition ability of minority classes and the overall prediction performance.

4 MODEL ANALYSIS

After preprocessing the above data sets, eight common machine learning methods such as decision tree, logistic regression, and support vector machine were used to build models to find the best model in PTSD research. The six indicators of accuracy, sensitivity, and specificity of each model were comprehensively considered. Finally, it was found that the decision tree classification model performed best among the eight machine learning models, with an AUC value of 0.9472. Classification performance evaluation indicators of each model can be seen in table 1.

Table 1 Classification Performance Evaluation Indicators of Each Model

Model	Accuracy	Sensitivity	Specificity	False_Positive_Rate	F1_Score	AUC
Decision Tree	0.9556	0.9375	0.9569	0.0431	0.7500	0.9472
LR	0.9289	0.8125	0.9378	0.0622	0.6190	0.8751
SVM	0.9200	0.8750	0.9234	0.0766	0.6087	0.8992
RF	0.9644	0.7500	0.9809	0.0191	0.7500	0.8654
NN	0.9289	0.6250	0.9522	0.0478	0.5556	0.7886
Adaboost	0.9511	0.8750	0.9569	0.0431	0.7180	0.9160
XGBoost	0.9556	0.8125	0.9665	0.0335	0.7222	0.8895
Stacking Ensemble	0.9644	0.6250	0.9904	0.0096	0.7143	0.8077

The ROC curve is used to evaluate the classification performance of the decision tree model on the training set and the test set. As can be seen from Figure 3, the AUC values of the decision tree model on the training set and the test set reached 0.97 and 0.95 respectively, indicating that the model not only has very high classification ability, but also shows good generalization performance. This means that the performance of the model on different data sets is very consistent and can effectively distinguish between high-risk and low-risk individuals. It is particularly noteworthy that the shape of the curve is close to the ideal state, highlighting the superior performance of the model in sensitivity and specificity, ensuring that individuals with a higher risk of PTSD can be accurately captured in practical applications.

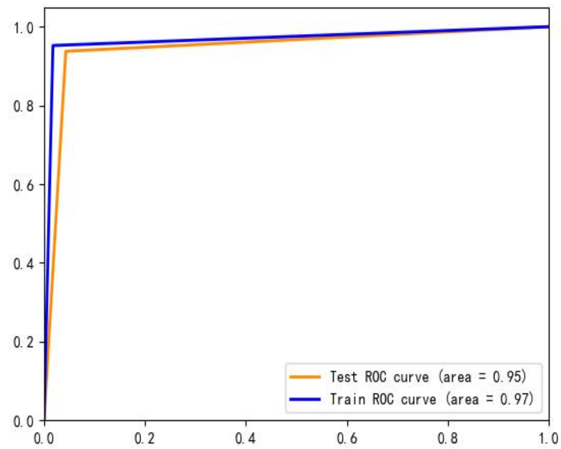

Figure 3 Decision Tree Classifier ROC Curve

The confusion matrix of the decision tree model on the training set and the test set can intuitively show the classification effect and misclassification of the model. In the training set, the decision tree model correctly classified 617 negative observations and 259 positive observations, and generated 11 false positives and 13 false negatives; in the test set, it correctly classified 200 negative observations and 15 positive observations, and generated 9 false positives and 1 false negative. Overall, the model performed well in classification accuracy and stability, which provides strong support for its application in practical operations. Confusion matrix can be seen in table 2.

Table 2 Confusion Matrix

(a) training set

		Reference	
		Positive	Negative
Prediction	Positive	617	13
	Negative	11	259

(b) test set

		Reference	
		Positive	Negative
Prediction	Positive	200	1
	Negative	9	15

The importance of each feature in the evaluation model can be seen from Figure 4. Its importance ranking from high to low is ASD alertness, ASD avoidance, ASD re-experience, ASD separation, ASD nature 1, smoking status 2, psychological resilience, age, income 3, income 4, weight, disaster scene 1, BMI, gender 1 and height. Among them, ASD alertness and ASD avoidance are the most important influencing factors, significantly affecting the risk of PTSD. ASD re-experience and ASD separation also play an important role in the model. In contrast, factors such as smoking status and psychological resilience are relatively minor, but still contribute to the risk of PTSD. Indicators such as age, weight and income are relatively less important, but still have a certain influence in specific situations.

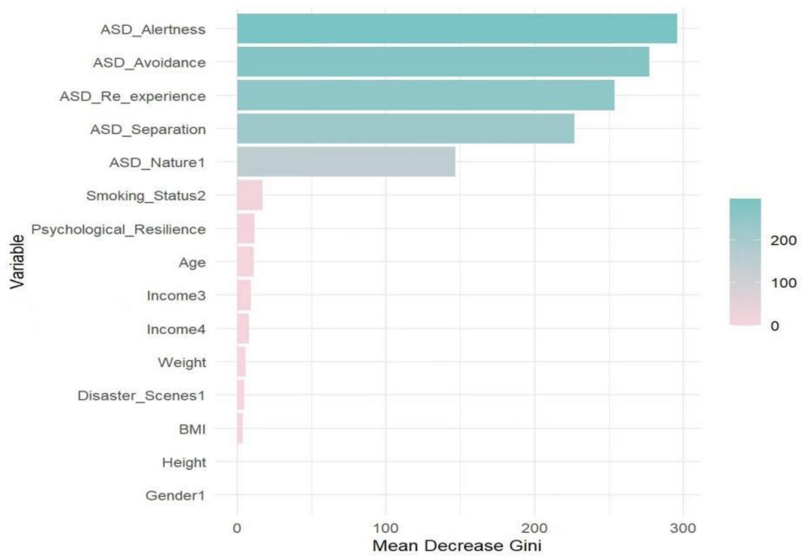

Figure 4 The Optimal Number of Features for the Model

5 CONCLUSION

Artificial intelligence (AI) has made rapid progress in the research and application of psychiatry. As an important technology in the field of artificial intelligence, machine learning has shown great recognition function in the field of medical diagnosis, providing a new research approach for its medical diagnosis. This study uses a decision tree model to construct an effective prediction model for rescuers and identify important variables that affect rescuers' post-traumatic stress disorder. The research results have high accuracy and reliability. For the features obtained by model classification, they are ranked according to their importance. The top eight feature variables are ASD alertness, ASD avoidance, ASD re-experience, ASD separation, ASD nature 1, smoking status 2, psychological resilience and age. It can be found that ASD alertness, ASD avoidance and ASD re-experience are the most critical influencing factors. It is recommended that rescuers should go to the hospital for help in time when they have symptoms of ASD alertness, ASD avoidance or ASD re-experience to prevent further development into PTSD, which will have a profound impact on the mental health and daily life of rescuers. For doctors, when conducting preliminary ASD (acute stress disorder) assessments on rescuers, psychological counseling should be implemented as soon as possible according to the ASD score to effectively prevent the occurrence and development of PTSD.

Through this study, we hope to provide valuable reference and guidance for research and practice in related fields, help rescuers cope with and prevent PTSD, and thus improve their long-term health and quality of life. The PTSD prediction model based on the decision tree will become a powerful tool for doctors in clinical practice, helping them to identify and intervene in PTSD early, and provide rescuers with more comprehensive and personalized psychological support.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] First MB. Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility. J Nerv Ment Dis, 2013, 201(9): 727-729. DOI: 10.1097/NMD.0b013e3182a2168a.

[2] Asmundson GG, Stapleton JA. Associations between dimensions of anxiety sensitivity and PTSD symptom clusters in active-duty police officers. Cognitive Behavior Therapy, 2008, 37(2): 66 – 75.

[3] Oniszczenko W. Temperamental correlates of trauma symptoms in firemen, policemen and soldiers. Int J Occup Med Environ Health, 2014, 27(4): 599-607.

[4] Witt M, Stelcer B, Czarnecka-Iwanczuk M. Stress coping styles in firemen exposed to severe stress. Psychiatria Polska, 2018, 52(3): 543-555.

[5] Hui Wuli, Shi Wanghong, Xu Dezhong, et al. Analysis of influencing factors of PTSD in rescue soldiers from shipwreck in Yantai. Chinese Journal of Mental Health, 2001, 15(3): 148-149.

[6] Liu Xiaohan, Chen Minglong, Guo Jing. Application of machine learning in the identification and prognosis prediction of post-traumatic stress disorder in children. Advances in Psychological Science, 2022, 30(4): 851-862.

[7] Tehrany MS, Pradhan B, Jebur MN. Spatial prediction of flood susceptible areas using rule-based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. Journal of Hydrology (Amsterdam), 2013, 504: 69-79.

[8] Liu Siyu, Nadalin. Application of decision tree analysis in English teaching. Medical Education Management, 2023, 9(S1): 54-60.