

Research Article

Application of Improved SMOTE and XGBoost Algorithm in the Analysis of Psychological Stress Test for College Students

Wei Du 

Psychological Health Education and Counseling Center, Jiaozuo Normal College, 454000 Henan, China

Correspondence should be addressed to Wei Du; jzszdw@jzsz.edu.cn

Received 7 March 2022; Revised 16 April 2022; Accepted 27 April 2022; Published 19 May 2022

Academic Editor: Wei Liu

Copyright © 2022 Wei Du. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To address the phenomenon of serious psychological stress among college students, there are problems of high cost and subjectivity in assessing psychological stress by collecting physiological data, and this paper proposes a stress assessment method (Improved SMOTE + XGBoost) based on intelligent data collection, which divides stress levels into five levels. In the process of processing a large amount of data, there will be too little data. Therefore, this paper applies the improved SMOTE method to the data preprocessing, which can reduce the difficulty of collecting psychological stress test data while ensuring the amount of data. Firstly, we extracted features from cell phone data to generate samples, processed the samples by SMOTE, and then filtered features by XGBoost algorithm to filter features; meanwhile, we trained RF, SVM, BP, and KNN with the data before and after sampling and before and after feature screening, and the results showed that Improved SMOTE + XGBoost outperformed other methods.

1. Introduction

As the pace of life accelerates, college students are increasingly likely to experience psychological problems due to the pressures of school and employment, and the America College Health Association's fall 2015 psychology report indicated that 57.7% of students felt "very anxious" at least once in the past 12 months [1]. Research has shown that perceived stress can significantly affect psychological and behavioral habits, and when people feel a great deal of stress, they often appear anxious, have insomnia, and in severe cases, can lead to mental and even physical illness. Some surveys have even pointed out that 64% of college dropouts are affected by mental illnesses. This kind of mental illness brought by stress, which is difficult to be taken seriously at the initial stage, may develop into serious problems, and then have a great impact on a person. Therefore, timely detection of psychological stress before it turns into a serious psychological problem is of great importance in terms of college students' mental health.

WHO believes that depression will be the second leading cause of serious illness worldwide after 2020 [2]. Therefore, it

is extremely important for individuals to maintain good physical and mental health if psychological stress can be assessed and given appropriate psychological guidance in a timely manner before it takes a serious toll on the body and mind. In recent years, the detection of psychological stress has received increasing attention. A lot of research studies have been carried out in the field of psychology on how to reasonably detect psychological stress in people. The most traditional method is the use of questionnaires based on psychological theories, which is still the most widely used method due to the theoretical support behind it. Secondly, human psychological stress can also be monitored by professional instruments, for example, human skin resistance can be linked to certain psychological indicators, monitoring skin resistance can be completed to detect psychological indicators, and the results are more reliable. Hosseini and Khalilzadeh [3] proposed an emotional stress recognition method, mainly by collecting electroencephalographic signals (EEG) and extracting features by wavelet coefficients to construct an Elman neural network binary classifier, and the final classification results achieved 82.7% accuracy. Jung and Yoon [4] used multimodal biosensors to extract EEG, ECG,

location, temperature, and weather features, used fuzzy logic and SVM to classify the sensory data, used DT and RF to construct a stress classifier, and used multiple physiological and environmental features to classify stress levels into four categories: normal, low, medium, and high, but no classification effect was given. Mahajan [5] extracted temporal and peak features of EEG and used MLP with 25 hidden layers to classify mental states into normal and stress. The classification accuracy was 60%. Murugappan et al. [6] extracted the features of EEG signal and used the KNN method to classify emotions into 5 categories and explored stress states based on emotions, and the classification result achieved 75.21% accuracy. Bichindaritz et al. [7] used ECG signal to classify stress levels into low, medium, and high using machine learning algorithms such as MLP and RF and classified stress levels as low, medium, and high and achieved a high accuracy rate. However, their dataset has only 67 samples.

However, physiological monitoring methods such as EEG are not suitable for daily psychological stress monitoring. Whether based on questionnaires or professional instruments, they require users to provide additional time costs to participate in the test, which is more intrusive to users and leads to less motivation to participate. Therefore, we want to find an automatic, low-cost, and low-intrusive method to achieve real-time stress monitoring. At the same time, smartphones have become a necessity in people's lives. In order to meet more needs in people's life, more and more sensing devices have been added to cell phones. In daily life, cell phones can continuously record a large amount of perceptual data related to people's daily life, including motion information, location information, cell phone usage information, and other data. For example, people tend to be less active, use mobile phones more frequently, and have lower sleep quality when they are under stress. The perceptual data provided by cell phones can reflect the behavioral habits of users, and the behavioral habits of users may have some connection with psychological stress, so we can try to use cell phone perceptual data to explore the connection between cell phone perceptual data and the psychological state of users through machine learning methods.

Psychological stress is firstly reflected in changes in mood, which is perceived from three main sources: professional equipment, social networks, and smartphone data. Changes in human psychology inevitably lead to changes in certain physiological indicators; many research studies have been devoted to the use of wearable devices to monitor daily psychological stress in humans [8]. Changes in human psychological indicators are sensed through integrated sensors. Because of the direct access to these physiological data, sensing based on wearable devices is often very convincing, but at the cost of people needing to wear these specialized devices, which poses the problem of higher costs and a greater degree of disturbance to people. From the perspective of big data, changing the test index of psychological stress from physiological to cell phone data can free the tested person from the professional equipment.

With the continuous development of Internet technology, social networks are also developing rapidly. The test for psychological stress can gradually shift from professional devices to cell phones with social network data. Lin et al. [9] used deep sparse neural networks to make judgments on users' psychological stress levels based on microblog data and used convolutional neural networks to detect users' stress. Xue et al. [10] extracted a series of features from adolescents' Twitter messages and used a classifier to understand the potential stress categories and levels of adolescents.

With the popularity of smartphones, more powerful features are provided to adapt to various usage scenarios, and more and more types of sensors are available with higher accuracy. Smartphones record users' habits of using cell phones and provide a large amount of valuable data. Mehrotra et al. [11] extracted a series of features using cell phone communication data from users' daily life and analyzed the correlation between multidimensional features and users' emotional frustration levels through a linear regression method. Xiong et al. [12] analyzed the correlation between different behavioral habits and users' social anxiety through a linear regression method by using GPS and POI data of college students. These works are a great guide to the feature extraction work in this paper; the drawback is that these works are not completed with the prediction model; this paper is completed under the inspiration of these works and constructs a model to predict the user's psychological stress. Canzain and Musolesi [13] achieved the prediction of user's frustration level by GPS data, extracted multidimensional features from user's GPS data, obtained user's frustration level from questionnaires, established the association between features and user's frustration using linear regression, and used SVM to construct a prediction model, achieving 80% accuracy. This paper uses a large amount of unlabeled data for training through cotraining, which can effectively improve the prediction accuracy of the model.

In this paper, based on the reference to the feature extraction method of psychological stress test for college students, the data were further processed with few category sampling and recursive circular screening of features. The data are preprocessed by improving the SMOTE method. In addition, the new method of machine learning is applied to the study of psychological stress, and the level of psychological stress is subdivided into more levels, which is more suitable for the actual situation of life, while the classification results achieve a high accuracy and broaden the research in the field of psychological stress.

2. Methodology

The research problem in this paper requires extracting m features from n college students' cell phone data by statistical methods to generate a data set $X_n = (x_1, x_2, \dots, x_n)^T$. The level of college students' psychological stress is recorded through a questionnaire survey, including S levels, and the stress level is used as the training category label Y_n of the classifier; using the sample data X_n and category label Y_n , a multiclassification model is built and trained by applying the

algorithm, thus making the respondents input the model by obtaining cell phone data when using cell phones and then testing the severity of the respondents' psychological stress. The population base of university students is large, and the amount of data collected during the actual psychological stress test is limited. Therefore, a range of data processing techniques were used to optimize the data structure and form the input data set for the algorithm. In this paper, the modified SMOTE algorithm was used to collect samples from a few categories and the XGBoost algorithm was selected to solve the multiclassification problem to analyze the psychological stress severity levels of college students.

2.1. Improved SMOTE. The category distribution of psychological stress-related data basically shows a normal distribution. Therefore, the category imbalance problem is inevitable. However, most classification algorithms classify the data with category imbalance, and the computational effect will be very unsatisfactory. Therefore, when collecting psychological stress data from college students, we need to balance the data set in advance and input the processed data into the model for training to achieve more satisfactory results. The methods of balancing data mainly include downsampling for multicategory data and oversampling for few category data [14].

Chawla et al. [15] proposed SMOTE (Synthetic Minority Oversampling Technique) as a sampling technique for intelligent data insertion of minority category sample data. SMOTE can increase the number of minority samples so that the classification algorithm can increase the learning of minority samples during the training of the data. Han et al. proposed an improvement method to address the drawback of inaccurate random sampling in SMOTE Borderline-Smote 1 (BS1) and Borderline-Smote 2 (BS2). BS1 is based on the original method by inserting new samples between only a few samples of the same category at the edge of the category, while BS2 generates new samples between samples of different categories that belong to the edge of the category. BS2 method causes the distance between samples of different categories to decrease, which is not conducive to machine learning for classification. However, SMOTE is a kind of overallocation.

However, SMOTE is an oversampling method, which adds noise data to the category samples and interferes with the classification of the algorithm purpose. It can effectively eliminate the noisy data caused by BS1 oversampling and prevent overfitting. Therefore, this paper combines the two methods of BS1 oversampling and TL downsampling for data sampling preprocessing. The steps of the BS1 method and TL method jointly applied are as follows.

(1) BS1 phase:

Input: dataset X_n , positive class:
 $A = \{a_1, a_2, \dots, a_{an}\}$
 Negative class $B = \{b_1, b_2, \dots, b_{bn}\}$, $an < bn$
 Output: balanced dataset

(a) Calculate the number of negative classes of a_i (m'), $m' < m$

(b) IF $0 \leq m' < m/2$, a_i is safe, continue
 (c) IF $m/2 \leq m' < m$, a_i is prone to misclassification, and put it into the danger set D_e
 (d) IF $m' = m$, a_i is noisy data, continue
 (e) For $j = 1$ to dn do: Let
 $D_e = \{a_1, a_2, \dots, a_{dn}\}$, $0 \leq dn < pn$
 (f) Randomly select similar s -nearest neighbors of a_j (s is determined by the category proportions) and calculate the deviation dif
 (g) New sample:
 $a_{new} = a_j + r \times dif$, r is random, $r \in 0, 1$

(2) TL phase:

(a) Calculate the distance from a_i to b_j , recording the minimum distance and the sample index of the minimum distance
 (b) Calculate the distance from b_i to a_j , recording the minimum distance and the sample index of the minimum distance
 (c) Compare the minimum distances of corresponding samples; if the distances are equal and the indexes are the same, they are Tomek links and delete the negative class of them

Improved SMOTE can effectively optimize the collection results of sample data and lay the foundation for subsequent algorithms for machine learning to achieve the identification and measurement of college students' psychological stress.

2.2. XGboost Algorithm. XGBoost algorithm is a gradient boosting tree algorithm that supports parallel computation and is based on the gradient boosting decision tree (GBDT) with a second-order Taylor expansion of the cost function. XGBoost shows the addition of the regularization term in the objective function, and when the base learning is CART, the regularization term is related to the number T of leaf nodes and the value of leaf nodes of the tree. XGBoost takes into account the case that the training data are sparse values, and the default direction of branches can be specified for missing values or specified values, which can greatly improve the efficiency of the algorithm [16]. Besides, it also supports column sampling, which not only reduces overfitting but also reduces computation. Its mathematical model can be expressed as follows:

$$\bar{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \bar{y}_i^{(i-1)} + f_t(x_i), f_t \in F, \quad (1)$$

$$F = f(x_i) = w_{q(x)}$$

where \bar{y}_i is the psychological pressure value of college students, x_i is the value of each evaluation index, t is the total number of submodels, $w_{q(x)}$ is the weight vector of all leaf nodes on XGBoost, f_k is the weight of each leaf node on the k th regression tree, and F is the set of all regression trees.

Define the loss function $l(y_i, \bar{y}_i) = (y_i - \bar{y}_i)^2$, which is the error dimension between the psychological stress value

and the psychological stress predicted value. The optimal solution $F^*(x) = \operatorname{argmin}_{E(x,y)} [L(y, F(x))]$ of the loss function is used to assist in selecting the appropriate number of leaf nodes, which can prevent the infinite growth of the number of leaf nodes and effectively save the model running time. The regularization term $\Omega(f_t)$, as shown in the following equation, makes the learned model simpler and prevents overfitting:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \delta \sum_{j=1}^T w_j^2, \quad (2)$$

where γ and δ are adjustment parameters to prevent the model from overfitting and T is the number of leaf nodes. And, the regularization term is positively correlated with the number of nodes. The objective function composed of the loss function and regularization term is shown in the following equation:

$$O^{(t)} = \sum_{i=1}^n (y_i, \overline{y_i^{(t+1)}} + f_t(x_i)) + \Omega(f_t) + \text{Con}. \quad (3)$$

In order to make the gradient descent of the objective function faster and more accurate, Taylor expansion is performed on it, as shown in the following equation:

$$\begin{aligned} f(x + \Delta x) &\cong f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2, \\ G_i &= \partial_{y_i^{(t-1)}} l(y_i, \overline{y_i^{(t-1)}}), \\ H_i &= \partial_{y_i^{(t-1)}}^2 l(y_i, \overline{y_i^{(t-1)}}), \\ O^{(t)*} &= \sum_{j=1}^T \left(G_j w_j + \frac{1}{2} (H_j + \delta) w_j^2 \right) + \gamma T. \end{aligned} \quad (4)$$

In order to avoid the excessive computational load caused by multiple enumerations, the “greedy algorithm” is used to find the optimal tree structure, new partitions are added to the known leaf nodes, and the gains after the partition are obtained in turn. The calculation is shown in the following equation:

$$\text{Gain} = \frac{1}{2} \left[\frac{G_K^2}{H_K + \delta} + \frac{G_L^2}{H_L + \delta} - \frac{(G_K + G_L)^2}{H_K + H_L + \delta} \right] - \gamma, \quad (5)$$

where Gain indicates that the information gain is the main reference factor for whether the tree structure branches or not, that is, when the information increment generated by the new segmentation reaches the depth limit of the tree or $\text{Gain} < 0$, the tree stops segmenting, so as to achieve the simulation effect of fast and good fitting while preventing overfitting.

Therefore, based on the data extraction of psychological stress influencing factors, this paper uses the XGBoost algorithm with high precision and strong stability as a training model to apply it to the overall evaluation of college students' psychological stress. The algorithm process is shown in Figure 1.

3. Experimental Results and Analysis

The model constructed in this paper requires the use of a portion of the sensing data from the open dataset “Studentlife” [17], which will be the Sensing EMA data in the dataset. The features were extracted from the Sensing data as the model samples, which includes eight categories of data including Activity, Music, Movie, Bluetooth, Conversation, Phone-charge, Phone-lock, and WiFi. The data of stress in EMA is used as sample labels, as shown in Table 1, and then, extract the corresponding features based on the sample labels. In this paper, the pressure is divided into five levels, denoted by the number from 0 to 4.

Human psychological stress is closely related to time, space, and life environment. Therefore, the results of the questionnaire reflect the psychological situation in which the students were at the time of filling it out. Therefore, this paper takes the time point in the data label stress as the center, extracts the Sensing data corresponding to the event window of 12 hours before and after this time point as the feature, generates the model data samples, and labels the samples as the corresponding label data stress. For example, for the data of Activity, the cell phone records the results as still. For example, for the data of Activity, the mobile phone records four states: still, walking, running, and undetected. Considering that the data related to the above four states can constitute a distribution, this paper uses the entropy table to represent the data distribution of the four states, while the data in a day are divided into daytime and nighttime to extract features, daytime refers to the time period from 8:00 am to 18:00 pm, and nighttime refers to 18:00 pm to 8:00 pm the next day. A total of 8 features are extracted and labeled accordingly to get a sample. A similar method is used to extract features for Music, Movie, Bluetooth, Conversation, Phone-charge, Phone-lock, and WiFi to generate samples. In this paper, 64 features were extracted and 2876 samples were obtained with labels as shown in Table 1.

The experimental environment of this study is the Win10 operating system, and the code is implemented on the Jupyter Notebook platform through Python3.7. Based on this environment configuration, the training and validation sets were divided in a ratio of 3:1. The model prediction results of the XGBoost algorithm are compared with the prediction results of mainstream models that have been proven to be more effective, namely, the Random Forest (RF) algorithm, Support Vector Machines (SVM) algorithm, Backpropagation (BP) neural network algorithm, and K-Nearest Neighbor (KNN) algorithm. Since the parameters configured by each algorithm are different, the grid search method is used to adjust the main parameters of each method. The configuration results of each model parameter are shown in Table 2.

The optimal parameter configuration of RF consists of the number of subtrees ($n_estimators$), the maximum growth depth of the tree (max_depth), and the minimum number of samples of leaves ($min_samples_split$). As for SVM, they are the penalty coefficient (C), the kernel function ($kernel$), and the distance error ($epsilon$). After the polynomial kernel function $poly$ is selected as the kernel function

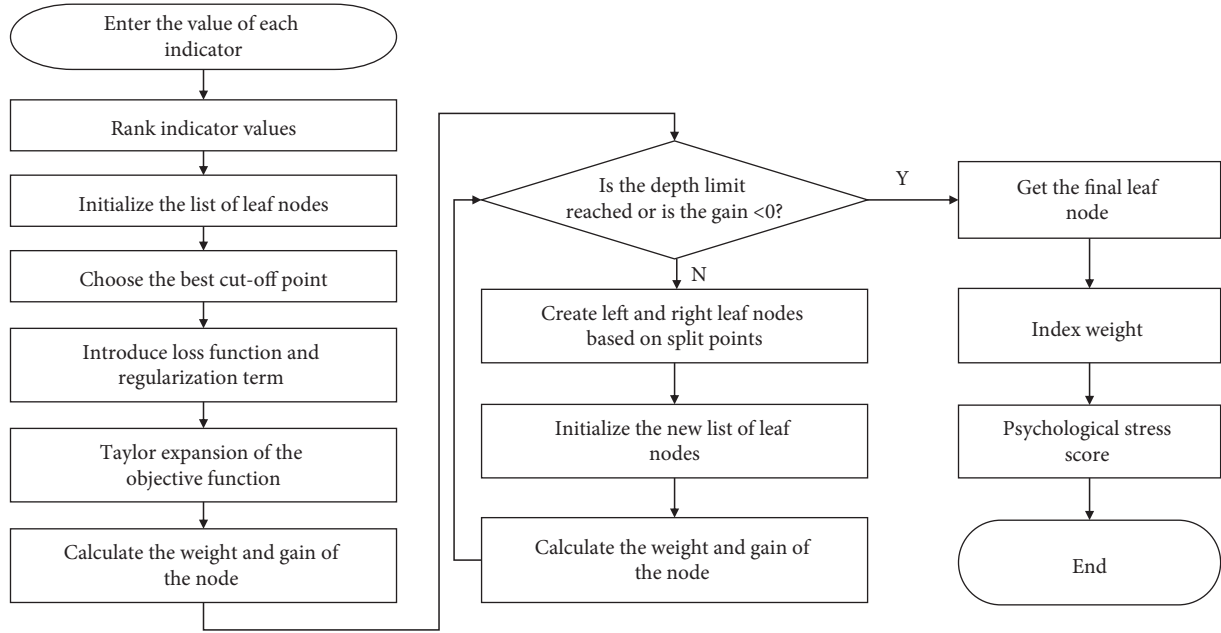


FIGURE 1: The process of the XGBoost algorithm.

TABLE 1: Stress labels.

Level	Number	Explanation
0	163	No stress
1	596	Basically no stress
2	1302	Mild stress
3	459	Moderate stress
4	356	Severe stress

TABLE 2: Parameters configuration.

Algorithm	Best parameter configuration
RF	n_estimators = 300, max_depth = 18, and min_samples_split = 4
SVM	C = 1150, kernel = 'poly', gamma = 0.4, degree = 2, and epsilon = 0.01
BP	nb_epoch = 200, batch_size = 16, optimizer = 'Adam', and activation = 'relu'
KNN	n_neighbors = 5, leaf_size = 15, and algorithm = 'auto'
XGBoost	max_depth = 6, learning_rate = 0.05, and n_estimators = 300

for the SVM algorithm, the parameters of this function implicitly (*gamma*, *degree*) determine the distribution of the data mapped to the new feature space. The number of iterations (*nb_epoch*), the number of samples selected for one training session (*batch_size*), the optimizer (*optimizer*), and the activation function (*activation*) constitute the parameter configuration of BP. The parameters of KNN include the number of neighbors (*neighbors*), the number of leaf nodes to stop building subtrees (*leaf_size*), and the specified algorithm for calculating the nearest neighbor (*algorithm*). The optimal parameter configuration of the final XGBoost algorithm contains the maximum depth of each binary tree (*max_depth*), the learning rate (*learning_rate*), and the number of iterations (*n_estimators*).

In order to compare the experimental results of each model more clearly, this study uses Accuracy (Acc),

Precision (Pr), Recall (Rc), and F_β ($\beta = 0.5$) as the verification indicators, as shown in the following equation:

$$\begin{aligned}
 \text{Acc} &= \frac{y_{\text{pred}}}{\sum \text{Data}} \\
 \text{Pr} &= \frac{y_{\text{pred}} \cap y_{\text{true}}}{y_{\text{true}}} \\
 \text{Rc} &= \frac{y_{\text{pred}} \cap y_{\text{true}}}{y_{\text{pred}}} \\
 F_\beta &= (1 + \beta^2) \frac{\text{Pr} \cdot \text{Rc}}{\beta^2 \text{Pr} + \text{Rc}}
 \end{aligned} \tag{6}$$

Figure 2 shows the comparison before and after sampling the data with the improved SMOTE method. After

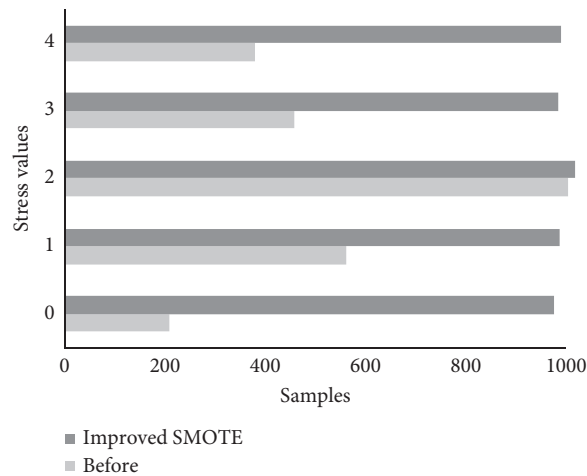


FIGURE 2: Data balance processing comparison.

TABLE 3: Comparison before and after sampling.

Classification method	Acc	Classification method	Acc
RF	0.492	RF + improved SMOTE	0.748
SVM	0.469	SVM + improved SMOTE	0.741
BP	0.461	BP + improved SMOTE	0.712
KNN	0.447	KNN + improved SMOTE	0.703
XGB	0.525	XGB + improved SMOTE	0.775

oversampling, 5370 samples are obtained, and after down-sampling, 4980 samples with labels are obtained. In order to reduce the gap caused by the extracted feature values, this paper normalizes the final data.

In order to demonstrate the enhancement effect of the sampling method of improved SMOTE on the classifier, the original dataset and the sampled dataset are compared in this paper, and the results are shown in Table 3. The results show that, before balancing, the accuracy rates of different classification methods are almost all below 50%. However, after the improved SMOTE sampling, each method displays a good classification effect. Among them, XGBoost has the best accuracy.

In this paper, the filtering method is adopted, which can reduce the problem of slow training due to too many features. If the recursive elimination method (RFE) is directly used to filter all the features, it will increase the time cost. RFE is used to remove one feature at a time to select the subset of features that gives the highest accuracy of the test set with as few features as possible. The experimental accuracy results for filtering the number of features from 50 to 20 are shown in Figure 3. In this paper, we use the test set data to show the accuracy of the model. The accuracy and the number of features in the graph show an irregular inverted U shape, and the accuracy is higher and relatively stable between the number of features 25 and 40. Even though the accuracy is 75.6% for both the number of features 25 and 29, the final number of features screened is 25, adhering to the principle of a faster rate with fewer numbers.

The XGBoost algorithm is used to filter 25 features for training, and the feature importance reflects which features are closely associated with the psychological stress state of

college students. In this paper, the top 5 features in terms of importance and their practical significance are listed in Table 4.

The data of the filtered 25 features are added into 6 classifiers of RF, SVM, BP, KNN, and XGB in order of importance to classify psychological stress into 5 states. The performance of the experimental results on Acc, Pr, Rc, and F_β is shown in Figure 4. The results show that the improved SWOTE + XGB method proposed in this paper achieves better accuracy for the classification of pressure states into five categories even when the number of features is small and has a greater advantage than other algorithms.

With the popularity of smartphones, the data on college students' cell phone usage increasingly reflects their living habits and psychological states. The pressures faced by college students are also varied. Anxiety about their studies, parents, and even marriage will be reflected in their mobile phone use status, which is the feature of the data collected in this paper. Taking "Activity" as an example, the more stressful and unsettled the students are, the more their cell phone time fluctuates. And, a correct understanding of the psychological stress level is the first step to alleviate the anxiety of college students. With the methodological evaluation in this paper, college students and their counselors will have a better understanding of the students' psychological state and will be able to provide stress relief. From a psychological point of view, having a clear goal will focus the mind on one place, thus diverting attention and weakening the adverse effects of psychological stress. In addition, having a clear goal is also an internal driving force, which makes people become positive and thus more conducive to overcoming various psychological stresses.

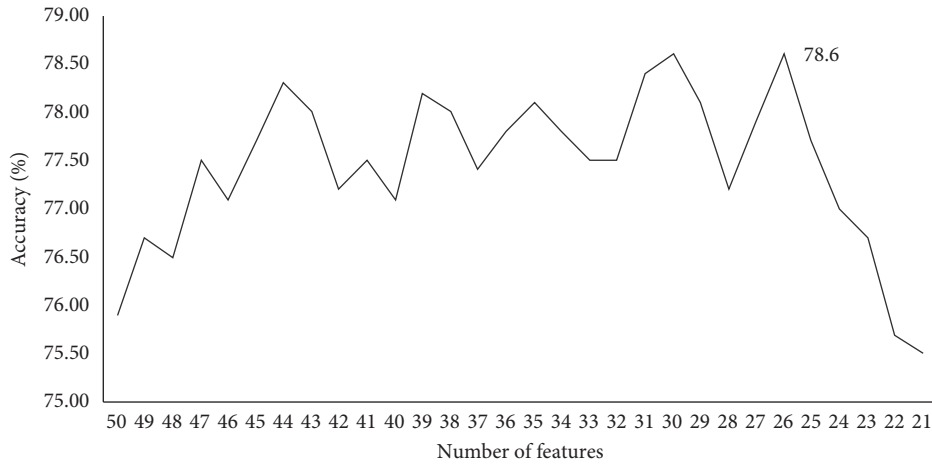


FIGURE 3: Feature screening process.

TABLE 4: Practical significance of important features.

Features	Name	Practical significance
0	Activity	The time the respondent remained still, walked, and was not tested
1	Music	The time respondents listen to music
2	Conversation	Time spent chatting on social networks
3	Phone-charge	Expenses paid by respondents for communicating over the phone
4	WiFi	How long do respondents use WiFi

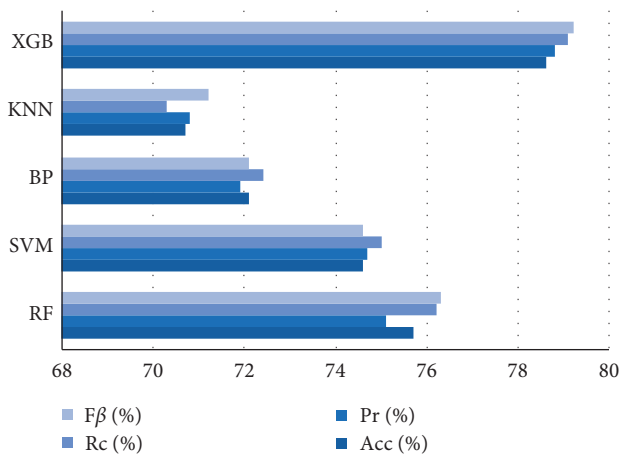


FIGURE 4: Comparison of final training results.

4. Conclusion

This paper addresses the disadvantages of using questionnaires and physiological data collection methods to assess psychological states, such as time-consuming and labor-intensive, subjective, and expensive, improves the use of smartphone mobile devices to collect data without the user’s awareness to assess the psychological stress of college students, and accurately assesses psychological stress into five levels. This paper investigates a new method IMPROVED SMOTE + XGBoost that can classify psychological states into multiple levels, which can also effectively solve the problems of unbalanced psychological stress data categories and feature redundancy. First, the features are extracted from the

perceptual data recorded in smartphones to obtain a sample set that can be used for training, oversampling and normalization are performed for the category imbalance problem of the dataset, and filtering and circular screening are performed for the feature redundancy phenomenon extracted from cell phone data. Secondly, RF, SVM, BP, KNN, and XGBoost methods are used to train the data before and after sampling and before and after feature filtering, and the results are compared to illustrate the effect of combining improved SMOTE methods with feature filtering and XGBoost. A multiclassifier is constructed based on feature screening, and the evaluation accuracy of 78.6% is achieved. Compared with the other four methods, the comprehensive approach in this paper has the best results and broadens the research on the psychological pressure of college students and the application of the XGBoost algorithm.

Although the comprehensive method in this paper has obtained good results in the multilevel classification of the psychological stress of college students, there are still some problems, for example, the waste of unlabeled data and the insufficient mining depth of psychological stress characteristics. In future work, more features related to college students’ stress need to be explored. In addition, college students should be subdivided according to age and major, and corresponding solutions should be given from a more detailed perspective.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Psychological Health Education and Counseling Center.

References

- [1] A. C. H. Association, *American College Health Association-National College Health Assessment II*, Reference Group Executive Summary, 2013.
- [2] S. Saxena, M. Funk, and D. Chisholm, "WHO's Mental Health Action Plan 2013-2020: what can psychiatrists do to facilitate its implementation?" *World Psychiatry*, vol. 13, no. 2, pp. 107-109, 2014.
- [3] S. A. Hosseini and M. A. Khalilzadeh, "Emotional Stress Recognition System Using EEG and Psychophysiological Signals: Using New Labelling Process of EEG Signals in Emotional Stress State," in *Proceedings of the 2010 International Conference on Biomedical Engineering and Computer Science (ICBECS)*, IEEE, Wuhan, China, April, 2010.
- [4] Y. Jung and Y. I. Yoon, "Multi-level assessment model for wellness service based on human mental stress level," *Multimedia Tools and Applications*, vol. 76, no. 9, pp. 11305-11317, 2017.
- [5] R. Mahajan, "Emotion recognition via EEG using neural network classifier," *Advances in Intelligent Systems and Computing*, vol. 58, no. 3, pp. 429-438, 2018.
- [6] M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *Journal of Biomedical Science and Engineering*, vol. 03, no. 04, pp. 390-396, 2010.
- [7] I. Bichindaritz, C. Breen, E. Cole, N. Keshan, and P. Parimi, "Feature selection and machine learning based multilevel stress detection from ECG signals," *Innovation in Medicine and Healthcare 2017*, Springer, vol. 71, no. 10, pp. 202-213, 2018.
- [8] O. M. Mozos, V. Sandulescu, S. Andrews, and D. N. R. J. M. Ellis, "Stress detection using wearable physiological and sociometric sensors," *International Journal of Neural Systems*, vol. 27, no. 02, pp. 3-4, Article ID 1650041, 2017.
- [9] H. Lin, J. Jia, and G. Quan, "Psychological Stress Detection from Cross-media Microblog Data Using Deep Sparse Neural Network," in *Proceedings of the IEEE International Conference on Multimedia & Expo*, IEEE, Chengdu, China, July, 2014.
- [10] Y. Xue, Q. Li, and L. Jin, "Detecting Adolescent Psychological Pressures from Micro-Blog," in *Proceedings of the International Conference on Health Information Science*, Springer, Melbourne, VIC, Australia, October, 2014.
- [11] A. Mehrotra, V. Pejovic, and J. Vermeulen, "My phone and me: understanding user's receptivity to mobile notifications," *ACM*, in *Proceedings of the CHI'16*, pp. 1021-1032, San Jose, USA, May, 2016.
- [12] H. Xiong, Y. Huang, and L. E. Barnes, "Sensus: A Cross-Platform, General-Purpose System for Mobile Crowdsensing in Human-Subject Studies," in *Proceedings of the Acm International Joint Conference on Pervasive & Ubiquitous Computing*, ACM, Heidelberg, Germany, September, 2016.
- [13] L. Canzian and M. Musolesi, "Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces analysis," in *Proceedings of the 2015 ACM International Joint Conference*, ACM, Osaka Japan, September, 2015.
- [14] S. García, S. Ramírez-Gallego, J. Luengo, and J. M. F. Benítez, "Big data preprocessing: methods and prospects," *Big Data Analytics*, vol. 1, no. 1, p. 9, 2016.
- [15] N. V. Chawla, K. W. Bowyer, and L. O. W. P. Hall, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321-357, 2002.
- [16] E. Roger, L. Torlay, J. Gardette, and C. S. L. P. M. Mosca, "A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy," *Neuropsychologia*, vol. 142, Article ID 107455, 2020.
- [17] R. Wang, F. Chen, and Z. Chen, "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the Acm International Joint Conference on Pervasive & Ubiquitous Computing*, pp. 1-14, New York; NY, USA, September, 2014.