Data-Driven Mental Health Assessment of College Students Using ES-ANN and LOF Algorithms During Public Health Events

Liangqun Yang[1], Haibin Ni[2], Yingdong Zhu[*3]
[1]School of Education Science, Neijiang Normal College, Neijiang Sichuan 641112, China
[2]Nursing School, Zhejiang Chinese Medical University, Hangzhou, 310000, China
[3]College of Information Science and Technology, Zhejiang Shuren University, Hangzhou, 310000, China
E-mail: zydyyds2023@163.com
[*]Corresponding author

Keywords: public health events, mental health education of college students, sampling neural network model, data mining

Received: October 22, 2024

Psychological stress in college students has attracted much attention due to its effects on psychological conditions during public health events. Traditional data mining can only get limited data through questionnaires, which is far from enough to present the whole situation for non-participants of the stress dynamic. This paper presents an advanced computational methodology for assessing psychological stress using data mining techniques. Herein, an ES-ANN model is developed to address the problem of imbalanced sample data. Besides, the LOF algorithm was realized, which allowed the comparison of the proposed approaches, including supervised and unsupervised learning methods applied in anomaly detection. Then, the results of extensive performance evaluation for the proposed ES-ANN model are performed by applying well-known G-mean and F1 score performance metrics. The results indicated that the ES-ANN model outperformed state-of-the-art benchmark methods, namely Random Forest and Decision Tree, with an increment of 8% in G-mean and an increment of 4% in F1 score. It proves the dependability and accuracy of the proposed ES-ANN model for the identification of students with high levels of psychological stress. Implementation of an integrated supervised-unsupervised learning system (ES-ANN and LOF) opens a new avenue to the early detection of psychological stress among college students.

Povzetek: Predstavljena je metodologija za odkrivanje in ocenjevanje psihološkega stresa pri študentih med javnozdravstvenimi dogodki z uporabo podatkovnega rudarjenja. Razvit je model ES-ANN za obravnavo neuravnoteženih vzorcev podatkov, uporabljen pa je tudi algoritem LOF za odkrivanje anomalij.

1 Introduction

Recently, with the development of science and technology and social modernization, under cruel and severe competition conditions, people face more and more pressure, and different psychological issues have gradually emerged [1]. When a public health event occurs, social and economic life and people's daily lives will also be affected to varying degrees, and even social stability will be affected to a certain extent. In this context, the normal teaching of universities and middle and primary schools has been impacted, and students cannot return to school normally [2]. The sudden public health incident in 2020 has caused a serious threat to the life safety of the broad masses of people in our country [3]. Simultaneously, this backdrop has significantly impacted the economic progress of the country [4]. In such a severe situation, people's psychology will inevitably be impacted. Therefore, negative emotions such as anxiety and panic have always plagued people [5]. People's health should be analyzed physically and psychologically [6]. Under public health incidents, physical damage to people can often be recovered within a certain period, but in this context, it is difficult for people to suffer psychological trauma [7]. It can be recovered in a short period [8]. Therefore, paying attention to mental health is an important issue today. As a special and important group of college students, their mental health cannot be ignored [9]. It is imperative to foster consciousness and proficiency in autonomously upholding physical and mental well-being [10].

The data mining technology applied to research in education is called Educational Data Mining (EDM) [11]. EDM has now been held in the 12th session of the International Conference on Educational Data Mining, indicating that EDM is highly valued by global experts [12]. Generally, EDM data sources fall into two categories: questionnaire-based, interview-type self-reported data, and student campus data. Questionnaire surveys and interviews, which are self-reported data, as the name implies, are the data that students actively display through questionnaires or questions and answers [13]. Student campus data refers to the data generated by students' lives and activities in the school recorded by the school information system. These data cannot only show the students’ behaviors such as website browsing, clocking in and out of class, and consumption in the campus cafeteria but also reflect the characteristics of students’ academic performance, lifestyle, and living habits and contain extremely rich excavate resources. Compared with questionnaire survey data, campus data is more objective and accurate because it records the real behavior of students [14]. Therefore, more and more studies use campus data to mine student behavior, and the outcomes are all good. Therefore, based on the mental health database of campus students, this investigation employs the data extraction approach to construct an ensemble sampling neural network (ES-ANN) model suitable for unbalanced sample data, to evaluate the mental health problems of college students, and to evaluate the mental health problems of college students. This suggests a strategy corresponding to the problem and providing a viable solution [15].

The aim, therefore, of this study is to construct and evaluate an integrated computational framework for the estimation of psychological stress amongst students enrolled in colleges during public health crises using campus-derived datasets with advanced mining techniques. The major twofold objectives involve overcoming the issues raised by the balanced data using the ES-ANN model and enhancement in the identification of the behavioral anomaly by incorporating the LOF algorithm. In addition, this study intends to assess the performance of the developed framework against traditional approaches such as Support Vector Machines and Random Forests using relevant performance metrics like G-mean and F1 scores. It also intends to develop pragmatic suggestions for policy improvements in intervention strategies at higher educational institutions. Campus-based datasets have been selected because, being objective and reliable, they are better representatives of day-to-day behavior than traditional self-reporting surveys. The selection of the ES-ANN model was informed by its prowess in handling data imbalance, while the inclusion aimed at detecting outliers representative of psychological stress, thereby addressing shortcomings in conventional supervised methods. G-mean and F1 scores are implemented to enable a fair and reliable assessment of performance, especially when dealing with imbalanced datasets. The objectives and components of the research design are well-defined to enhance the transparency, reproducibility, and applicability of the study.

Safika et al. [16] emphasized that mental health issues have emerged as a significant concern in Malaysia, especially among higher education students, influencing their emotions, cognition, and social interactions. The NHMS 2017 reported that one in five Malaysians had depression, two in five suffered from anxiety, and one in ten encountered stress. Identifying contributing issues proved difficult, obstructing meaningful assistance. Their study examined students' mental health difficulties and contributing factors while assessing machine learning techniques for predicting these problems, establishing a basis for future research utilizing computer modeling. Wang and Park [17] implemented an intelligent sports management system that integrated deep learning, with which they were able to balance the negative features of poor physical fitness of college students, which they improved by managing sports facilities. This system was put into practice in the browser/server architecture, using the Spring Cloud framework for its implementation and, therefore, integrating all the components into a distributed microservices architecture. An AI recommendation algorithm also takes up students' age, BMI, and health in recommending appropriate sports programs and proposing courses based on preference and logistical considerations. A relational database model was built through the use of an entity-relationship diagram. Functional testing showed that this system efficiently supports sports education enhancement and student well-being. Abelson et al. [18] emphasized the escalating worry regarding student mental health in higher education, as an increasing proportion of students report difficulty and seek assistance. Colleges and universities encounter obstacles in tackling these issues due to the fragmented nature of extant research on mental health interventions across many academic disciplines. This chapter synthesizes this corpus of work by providing a comprehensive review of programs, services, practices, and policies that affect student mental health, organized utilizing a socioecological technique. The review examines interventions at each of the following levels of intervention: individual, interpersonal, communal, institutional, and public policy. The strengths, limitations, and gaps in the current evidence are outlined, with final suggestions for how research and data can inform practice. In addressing growing mental health concerns among students in higher education, an evidence-based methodology is of central importance. Huang et al. [19] proposed a method for automatically assessing civic education using dynamic clustering and the fuzzy C-means technique by adding the element of mental health education during that process. Compared with traditional models such as BP and RBF, the RMSE and MAE were improved in this model. This study confirmed that cultural confidence partly acts as a mediator between civic education and mental health. This investigation sought to improve the integration of political, ideological, and mental health education, thereby fostering the comprehensive development of pupils. Cai and Tang [20] researched higher education and public mental health in college students for the inadequacy of conventional approaches, then introduced convolutional neural networks and random forest algorithms using Artificial Intelligence. This gave an accuracy of 87.5% in the analysis of this correlation. Compared with the support vector machines and backpropagation neural networks, their model was better for evaluating public mental health in higher education. Yao's [21] research examined the influence of COVID-19 on the mental health of college students by developing a relational model between students' psychological well-being and the pandemic through a BP neural network. The study examined the psychological impacts of the pandemic via event monitoring, early warning systems, and student behavior, employing principal component analysis to assess the determinants of mental health. The outcomes underscored the necessity of cultivating professional, physical, humanistic, and ethical attributes in students to guarantee stability amid public health crises. The approach was a significant reference for mental health education and effective intervention programs within institutions. Li [22] focused on the increasing prevalence of psychological disorders among college students and the rising need for an early warning system for psychological crisis interventions. An upsurge of AI and big data technologies will bring great university mental health teaching opportunities. The paper reviewed the current applications of AI and big data in mental health education, underlined prevailing difficulties, and proposed solutions based on analysis of psychological crisis features. Li developed a comprehensive early warning data collection system comprising six constituents: educational administration and access control systems. Accordingly, the research proposed multilayer feedback linking system and optimization of teams that consist of technical and psychological professionals. Theoretically and practically instruct the development of college mental health management. Wang and Yang [23] investigated the increased social pressures and mental health challenges of university students and identified the need to understand the psychological condition of students more fully. Machine learning algorithms on large-scale data about students' mental health unveiled key factors and trends. Prediction models effectively forecast the status of mental health, underlining relations between academic stress, social interactions, and consequences on mental health. The study has also displayed those differences exist according to gender, academic year, and personal background; first-year students have a higher stress level and are most receptive to treatment. These students from different backgrounds proved more susceptible to environmental changes and required special support. Such findings will provide a sound foundation for educational establishments to establish differentiated activities to preserve mental health, enhance students' well-being, and reduce stress by targeted intervention. It ranged from evidence of the essential use of machine learning in evaluating mental health to therapeutic strategies. ZHANG et al. [24] considered the ever-growing challenges to college students' mental health in a fast-changing society. In this research, an investigation questionnaire was designed to study sources of stress, such as occupational and academic burdens, while assessing mental health was carried out using the SCL-90 scale. A forecast model based on BPNN, whose optimal parameter values had been searched using the ISOA, was put forward. Therefore, the ISOA-BPNN model outperformed all the other optimization methods, like PSO and ABC, with a huge margin in terms of accuracy, 0.9762; F1 score, 0.9834; and AUC, 0.9956. The present study verified the ISOA-BPNN model with enhanced dependability in the forecast of college students' mental health conditions and thus is expected to find applicability in educational settings. Dayananda et al. [25] discuss a study on applying machine learning predictive modeling in estimating university students at high risk for mental health problems. Demographic and academic characteristics and social indicators of understanding complex variables leading to poor mental health conditions were researched. Using machine learning aims to improve educational institutions' capability to identify at-risk students early and provide suitable care support systems. Such interventions will help develop a more caring and healthy learning environment conducive to successful student outcomes and general well-being. Regarding this, the study has developed some valid and reliable predictive models. He [26] explored mental health concerns among college students by recommending and analyzing computational schemes for mental health evaluation, drawing on demographic, academic, and psychological factors. The Probabilistic Deep Belief Bayesian Network adopted in the research classified the mental health attributes of students by estimating probabilistic values for various factors, including gender, age, academic performance, social support, and self-reported levels of stress, anxiety, and depression. The PDBBN model was compared against CNN and DNN models, and it outperformed both in categorization accuracy, precision, recall, and F1 score, achieving a categorization accuracy of 0.98. The study also demonstrated that mental health significantly affects students' academic performance. Zou [27] researched course-recommender systems for innovation and entrepreneurship education by comparing the approach of collaborative filtering with that of the ANN. The ANN outperformed collaborative filtering on all metrics: training and testing time, hit rates, and NDCG values. From this, underlined was the effectiveness and efficiency of ANNs to be used for course selection. Al-Jammali [28]investigated data mining methods in the medical field, concentrating on the prediction of heart disease through the use of algorithms such as SVM, ANN, and decision trees. Decision trees performed poorly, however SVM did better than the others. By testing more classifiers and adding more data to predictive models, the study focused on improving diagnosis accuracy. Liu and Yang [29] presented a methodology known as the KPCA-ANN technique for multimedia analysis. This method fuses kernel principal component analysis with artificial neural networks. Based on the database availed from the TianChi competition, the approach showed an absolute mean error of 7, with a structural similarity index of 86 by deploying 1,000 CT scans. This work has highlighted how this approach can further improve multimedia processing, assist in decision-making processes, and enhance one's understanding of complicated data. Adeniji [30] suggested a hybrid model for the prediction of mental health disorders among information technology employees using a Random Forest-ANN model. A hybrid RF-ANN model, using a "Bagging Ensemble," gave an overall weighted average of 84.5% in terms of Precision, Recall, and Accuracy, higher than from RF and ANN models individually. The study inferred that it was a promising model that, with more refined data sets, could attain early detection of mental disorders. Chen and Lan [31] presented a PERMA model to be used in mental health education in higher education facilities for student psychological concerns. The model was based on the tenants of Positive Emotion, Engagement, Relationship, Meaning, and Achievements, involving an expanded framework and ways of evaluating performance. Results demonstrated that it was 93% accurate in terms of the prediction, hence very effective in promoting MHE among university students. Duan [32] proposed the use of the Northern Goshawk deep multi-structured convolutional neural network in emotion recognition for mental health education. It was preprocessed by Kalman filtering using data from 300 participants. This method outperforms conventional algorithms in accuracy, precision, recall, and F1-score. The study stressed the effective role of NG-DMCNN in enhancing mental health education. Kolenik [33] highlighted smartphone-based methods as transformative tools for addressing stress, anxiety, and depression, leveraging IoT and AI to provide real-time assessments of physiological, behavioral, and cognitive indicators. These insights enabled personalized, timely interventions, such as cognitive-behavioral therapy through smartphone apps or intelligent assistants. While promising, Kolenik noted the field's infancy, emphasizing the need for standardized frameworks and methodologies to maximize its potential. Kolenik and Gams [34] carried out a cutting-edge technical overview of intelligent cognitive assistants for SAD, emphasizing developments like tailored therapies, classification-based evaluations, and user models. The significance of machine learning and user-centered design in augmenting the efficacy of intelligent cognitive assistants for digital mental health treatments was underscored. Kolenik and Gams [35] argued for global inequality of mental health care, manifested in low treatment rates regarding stress, anxiety, and depression due to professional shortage and unequal access. They argued that digital solutions, represented by intelligent cognitive assistants, constitute scalable and affordable means for promoting the cause of equal care in mental health. Table 1 describes the previous researches.

Table 1: Summary of previous research

Study	Methodology	Key Findings	Limitations
Safika et al. [16]	Machine learning techniques	Identified factors contributing to mental health issues	Lack of comparative analysis with advanced models
Wang and Park [17]	Intelligent sports management system	Improved physical and mental fitness through AI recommendations	Limited to sports-related interventions; lacks stress evaluation
Abelson et al. [18]	Socioecological framework	Comprehensive review of student mental health interventions	Focused on qualitative synthesis; no quantitative modeling
Huang et al. [19]	Dynamic clustering with fuzzy C-means	Improved civic education integration with mental health	Limited focus on psychological stress quantification
Cai and Tang [20]	CNN and Random Forest algorithms	Achieved 87.5% accuracy in public mental health evaluation	Did not address imbalanced datasets or anomaly detection
Yao [21]	BP Neural Network	Developed relational models for mental health during COVID-19	Focused only on pandemic-specific factors
Li [22]	AI and big data systems	Proposed an early warning system for mental health crises based on multi-layer feedback and optimization	Limited integration of unsupervised techniques and lack of real-time anomaly detection
Wang and Yang [23]	Machine learning with large datasets	Highlighted gender, academic year, and background factors in mental health prediction using ML models	Lacked targeted behavioral anomaly detection; less focus on imbalanced datasets
Zhang et al. [24]	ISOA-BPNN (Optimized Backpropagation NN)	Achieved high accuracy (0.9762), F1 score (0.9834), and AUC (0.9956) for mental health prediction	Limited in handling outliers and behavioral anomalies; lacks integration with external datasets
Dayananda et al. [25]	Multi-model ensemble machine learning	Developed predictive models for identifying students at risk of mental health issues based on demographics	Did not incorporate behavior-based anomaly detection; lacked dynamic adaptability
He [26]	Probabilistic Bayesian Network (PDBBN)	Achieved 98% accuracy in classifying mental health	Focused primarily on demographic and self-

		attributes, outperforming CNN and DNN models	reported factors; no robust anomaly detection methods.
Zou [27]	Compared collaborative filtering with ANN for course selection	ANN outperformed collaborative filtering in all metrics, proving its effectiveness and efficiency for course selection.	No explicit limitations mentioned; potential for further exploration of ANN-based recommendations.
Al-Jammali [28]	Evaluated SVM, ANN, and decision trees for heart disease prediction	SVM performed best; decision trees performed poorly. Highlighted improving diagnosis accuracy through more classifiers and refined data.	Recommended testing additional classifiers and incorporating more data to enhance model accuracy.
Liu & Yang [29]	Developed KPCA-ANN using 1,000 CT scans from the TianChi competition database.	Achieved a structural similarity index of 86 and an absolute mean error of 7. Showed promise in multimedia processing and decision-making.	Suggested further optimization for complex data understanding and improved multimedia processing.
Adeniji [30]	Proposed a hybrid Random Forest-ANN (RF-ANN) model with a "Bagging Ensemble.	RF-ANN outperformed standalone models, achieving 84.5% weighted average in Precision, Recall, and Accuracy	Recommended testing on refined datasets to enhance early detection capabilities
Chen & Lan [31]	Developed a model based on Positive Emotion, Engagement, Relationship, Meaning, and Achievements	Achieved 93% accuracy in predictions, demonstrating the effectiveness of the PERMA model in MHE.	No significant limitations were mentioned; suggested potential for broader application in MHE.
Duan [32]	Proposed NG-DMCNN, preprocessed with Kalman filtering, using data from 300 participants.	Outperformed traditional algorithms in accuracy, precision, recall, and F1-score. Highlighted NG-DMCNN's role in enhancing MHE.	No explicit limitations noted; recommended further studies for broader validation of the NG-DMCNN method.
Kolenik [33]	Smartphone-based AI assessments using IoT and real-time data analysis	Explored real-time stress, anxiety, and depression detection.	Lacked standardized methodologies; issues with generalizability.
Kolenik & Gams [34]	Technical review of cognitive assistants with machine learning.	Emphasized tailored therapies, classification-based evaluations, and user-centered design for digital mental health	Did not address scalability challenges and real-world ethical concerns.
Kolenik & Gams [35]	Analysis of global inequalities in mental health care and digital solutions.	Proposed intelligent cognitive assistants as scalable and affordable tools to bridge treatment gaps caused by professional shortages.	Challenges in implementation due to systemic barriers and resource disparities.

1.1 Novelties and contributions

The paper contributes to several novelties in the research area of college students' mental health education, especially under the influence of a public health event. In particular, this study adopts an ES-ANN model, specially designed for the unbalanced sample data that frequently happens when assessing students' mental health. Using data mining techniques applied to a mental health database compiled from student campus information, this work reaches much more accurate estimates compared to the classical approaches of questionnaires. Moreover, the incorporated LOF anomaly detection algorithm will allow for a comparison that has never been made before between a supervised and an unsupervised learning system for the complete comprehension of psychological stress in students. The proposed model, ES-ANN, has improved on the benchmark methods by eight percentage points on the

G-mean and four percentage points on the F1 score. This is an advancement toward the effective and efficient identification of students with mental health issues to contribute effectively to data mining in education and at academic institutions in managing the mental health status of their students.

1.2 Paper organization

The document is outlined in a general manner: the second section presents the *State of the Art* about mental health education of university students, mainly with public health problems. It deals with how such events have brought a psychological impact on students, focusing on increased anxiety, depression, changes in behavior, and, at the same time, a change in their cognitive and emotional characteristics due to social and technological shifts. The next section will discuss some other important psychological issues, such as gullibility, depression, and overdependence on digital platforms, in elaborating on what challenges students face in adapting to changes within the natural environment. Section 3 describes the research methodology concerning data collection and construction of the ES-ANN model and LOF-based anomaly detection algorithm. It then discusses the empirical findings through the performance of the ES-ANN against the benchmark methods in Section 4 and addresses limitations, especially about unbalanced sample data. These sections also recommend potential lines of future research to deal with the identified shortcomings. Finally, Section 5 summarizes the research findings and wraps it up with implications for mental health education in an academic setting.

2 States of the art

2.1 The impact of public health events on the psychology of college students

Public health events will impact social development and easily affect people's psychology. As a special group in society, college students often experience obvious psychological changes during public health events. Considering the psychological changes encountered by college students is advantageous in fostering a new era of society as an ideological successor [37].

2.1.1 The psychological characteristics of contemporary college students

The current college students are basically "post-95s". These individuals attended college and were raised amid China's swift economic growth. The rapid development of society has given this generation of college students many favorable conditions, but it is also affected by the bad atmosphere in social development [38]. The psychological characteristics of contemporary college students: First, thinking performance. Modern college students grow up in the era of rapid Internet development, acquire a large amount of information, have diverse views, and have their thinking. However, the early exposure to unfiltered information on the Internet creates an illusion of maturity in its ideas and opinions. Nevertheless, the Internet's thinking lacks maturity and often exhibits fragility and impulsiveness [39]. The second dimension is behavioral performance. The college students of today, belonging to the iGen generation, because they have grown up in a digitally-saturated environment show extreme personal freedom while accessing information; however, they usually have less group sensitivity and team spirit. Their dependency on the digital world makes them not only unable to maintain human relationships but also incapacitates them to face adversities with dexterity, which, again, heightens emotional vulnerability and psychological issues [40]. The third is cognitive performance. Self-cognitive ability is the key to growth and the core of talent training for college students. Contemporary college students have a strong ability to explore and innovate but are not yet stable in cognition and are easily affected by the outside world. Therefore, positive guidance should be provided for college students.

2.1.2 Psychological performance of college students under public health events

Based on the comprehensive analysis of the relevant literature from the occurrence of SARS in 2003 to the health event and the study of the outcomes of the questionnaire for students in our school, the psychological performance of college students under the public health event mainly falls into the following categories: first, gullibility, the emotions of college students are affected by the epidemic one of the obvious manifestations of the influence is credulity. The outcomes of the questionnaire show that college students have more access to information about outbreaks on the Internet. They are more likely to be guided by Internet information, and the Internet follows the trend second, depression tendency. Due to the impact of the pandemic, the outdoor activities of college students are restricted. Staying at home or on campus for a long time is likely to cause emotional depression and negative psychology. "College students generally have high negative emotions." Third, anxiety; according to the outcomes of the questionnaire, since the development of the epidemic, college students have widely experienced certain anxiety, such as worrying about a serious outbreak or the infection of their family members, etc. Anxiety affects the study and life of college students.

2.2 Under the background of public health events, there are problems in the mental health of college students

2.2.1 Learning life pattern mutation, leading to over-reliance on the network

In the event of sudden public relations and health incidents, various colleges and universities will use online teaching to impart professional knowledge to students. Simultaneously, the learning mode of college students has also changed from offline to online. Passive adaptation changes daily learning and college students' life mode. Besides daily online courses, most of my time is spent in online games, dating, and video apps. By using the data of the amount downloaded online games and the frequency of using mobile game software, it can be determined that a large number of college students have been immersed in too much online for too long and are dependent on the Internet, which has caused the disorientation of life and lack of self-discipline.

2.2.2 Insufficient ability to adapt to the environment

During public health emergencies, most colleges and universities in the country adhere to the "principle of not leaving school unless necessary" and "do not leave school unless necessary" in medium and high-risk areas to strengthen campus management and strictly enforce that students need to report in advance when entering and leaving the school, campus door temperature detection and code scanning registration, classmates' body temperature in the morning, noon and evening, student health report system, teacher and student movement track tracking system, etc. As a result, most college students are concentrated on campus, and the number of students leaving the school is reduced. The campus implements this management method to reduce the risk of students contracting the virus. Compared with before the epidemic, the main activity places of college students are concentrated. There are dormitories, libraries, classrooms, canteens, basketball courts, stadiums, etc. Before the pandemic, college students were accustomed to the lifestyle of on-campus study and off-campus entertainment, and most students had corresponding psychological and behavioral needs for off-campus entertainment activities. However, in the face of the requirement of normalized management of the epidemic, college students spend most of their spare time in the limited teaching and activity venues on campus, and their lifestyles have passively changed. Therefore, reducing off-campus entertainment has also made college students psychologically uncomfortable adapting to such environmental changes. For example, students have a sense of unfamiliarity, oppression, and monitoring of the once-familiar campus environment, and there is a decline in their psychological adaptability. Symptoms such as disorders of mental adjustment ability affect their normal study and life in school and then affect the behavior of college students.

2.2.3 Increased negative emotions

First of all, in terms of cognition, during the epidemic, various information about the epidemic was obtained from different information channels, including positive and negative information, as well as real and false information. Too much information about the epidemic caused college students to feel uneasy about the epidemic. Excessive anxiety and worry make it impossible to make a correct judgment on the current outbreak and then generate anxiety. Some students feel depressed and sad because of the long-term home epidemic prevention. Students exhibit a lack of motivation and aversion towards engaging with others, which may lead to feelings of hopelessness. Also, because of the depression of others around them, college students are unable to sleep and eat normally, which eventually leads to a decline in the body's immunity, which in turn leads to more anxiety and depression.

2.3 Development status of educational data mining

EDM is a domain that has recently been showing tremendous progress and drawing significant attention. With the recent rise in the popularity of machine learning, there is generally a trend towards incorporating machine learning techniques into educational data mining research. Direction has been put into application in various fields.

Educational data mining is an interdisciplinary subject that integrates pedagogy, statistics, and computer science. It uses data mining techniques to convert raw educational data into useful knowledge. This text provides valuable services that can benefit educators, learners, managers, developers of educational software, and academic researchers alike. Within the conventional teaching environment, characterized by direct interaction between teachers and students, EDM utilizes statistical methods to analyze the data collected, usually through interviews and questionnaires, to obtain. The progress made in information technology has led to remarkable accomplishments in open teaching environments, specifically those that rely on big data education. EDM has successfully integrated data mining technology into the Web environment, encompassing students' login records, homework assignments, and teaching resources. Fig. 1 shows the EDM process.

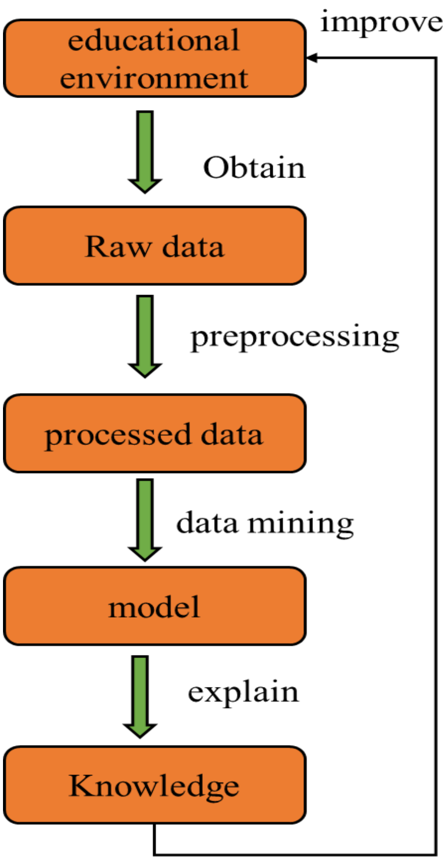

Figure 1: Educational Data Mining (EDM) Process

3 Methodologies

3.1 Neural networks

A neural network, also known as an ANN, is a process of simulating the behavior of the human brain and reasoning according to logical rules. The neural network category mainly includes artificial neural network ANN for classification and regression problems, convolutional neural network (CNN) for image recognition problems, and recurrent neural network RNN for natural language processing (NLP) problems. This experiment introduces the neural network ANN. Here are three nonlinear activation functions used in the experiment: Sigmod, Softmax, and Relu.

The main reason the ES-ANN was selected for the research study is that compared with other methods of dealing with imbalanced datasets-a common feature in most mental health evaluations and classification activities-this technique has an improved capability. Unlike other conventional ANN methods, which usually perform below par when trained on unbalanced data, ES-ANN includes ensemble sampling methodologies that resample the dataset in such a way that balanced subsets are created. This approach ensures the minority class is well represented during training, improving classification and generalization.

Simultaneously, ES-ANN enjoys certain advantages against other ensemble techniques like Random Forests and Gradient Boosted Trees. Although quite good for processing tabular data, Random Forests model complex nonlinear relationships far worse than neural networks.

Compared to these, ES-ANN integrates robustness due to the ensemble technique, along with the representation capability of ANNs; it is hence especially apt to pick up complex patterns in behavioral and psychological data. Furthermore, the sampling strategy and the architecture of networks could be personalized in ES-ANN, allowing the model to become flexible enough to capture the peculiarities of the dataset. This choice was verified by comparing the performance of ES-ANN against traditional ANN and other ensemble methods like Random Forests and XGBoost. The results of the experiments showed the highest metrics, such as precision, recall, and F1, in the case of ES-ANN, especially for the minority class. The results presented confirm the scientific grounds for choosing ES-ANN as the main classification model for the current study.

The function expression of Sigmod is:

$$y = 1/(1 + e^{-x}) \tag{1}$$

It is generally used as the activation function of the two-class problem. When the multi-class activation function is required, the Softmax activation function is commonly used because it is well-suited for the output layer activation function of the neural network. The Softmax activation function maps the prediction of each classification to a value of (0,1). The sum of all classification prediction outcomes is one so that it can be understood as a probability. The value of each classification is its possible probability, and the classification with the largest value is output as the prediction result.

The expression of the Softmax function is:

$$y_i = e^{z}i / \sum_{j=1}^{k} e^{z_j} \tag{2}$$

The Relu activation function is also commonly used and can solve the gradient disappearance problem that Sigmod is prone to.

The function expression of Relu is:

$$y = max(0, x) \tag{3}$$

Next, the gradient descent algorithm of the feedforward neural network solution process is introduced. In case the sample (x, y) has the cross entropy as its loss function, the function is:

$$L(y, \hat{y}) = -y^T \widehat{\log y} \tag{4}$$

where y∈ {0, 1}c is the label y value displayed by the one-hot vector, and C displays the number of categories.

Forward The error back-propagation algorithm of the fed neural network, according to the chain rule, is as follows:

$$\frac{\partial L(y, \hat{y})}{\partial w_{ij}^{(l)}} = \left(\frac{\partial z^{[l]}}{\partial w_{ij}}\right)^T \frac{\partial L(y, \hat{\eta})}{\partial z^{(1)}} \tag{5}$$

The error term δ (1) of the first layer is finally expressed as:

$$\delta^{(1)} = f_1'\left(z^{(1)}\right) \cdot \left(\left(W^{(1+1)}\right)^T \delta^{(1+1)}\right. \quad (6)$$

Formula (6) is the back-propagation formula of the error because the error term of the first layer can be obtained by calculating the error term of the 1+1 layer.

3.2 Anomaly detection (LOF)

Outlier detection detects unreasonable or abnormal data distribution, distinguishing abnormal patterns from normal ones. Therefore, paying attention to the occurrence of abnormal data and analyzing its reasons has become an important opportunity to discover and improve decision-making. Common outlier detection methods include statistical-based anomaly detection, distance-based anomaly detection, density-based anomaly detection (LOF), depth-based anomaly detection, and isolation Forest algorithms. This experiment mainly introduces the density-based anomaly detection scheme. Anomaly Detection Algorithm (LOF).

LOF algorithm is also known as the local anomaly factor LOF algorithm. The distribution of point C1 is sparse, and the interval is large when pictures are used to illustrate the points in the figure's two parts of C1 and C2. Still, the overall density is relatively uniform, which can be considered points of the same cluster. In the same way, the C2 points are closely distributed, the interval is small, and the overall density is consistent, which can also be considered the same cluster. However, the o1 and o2 points are relatively isolated from the group and may be regarded as abnormal or distinct. See Fig. 2 for details.

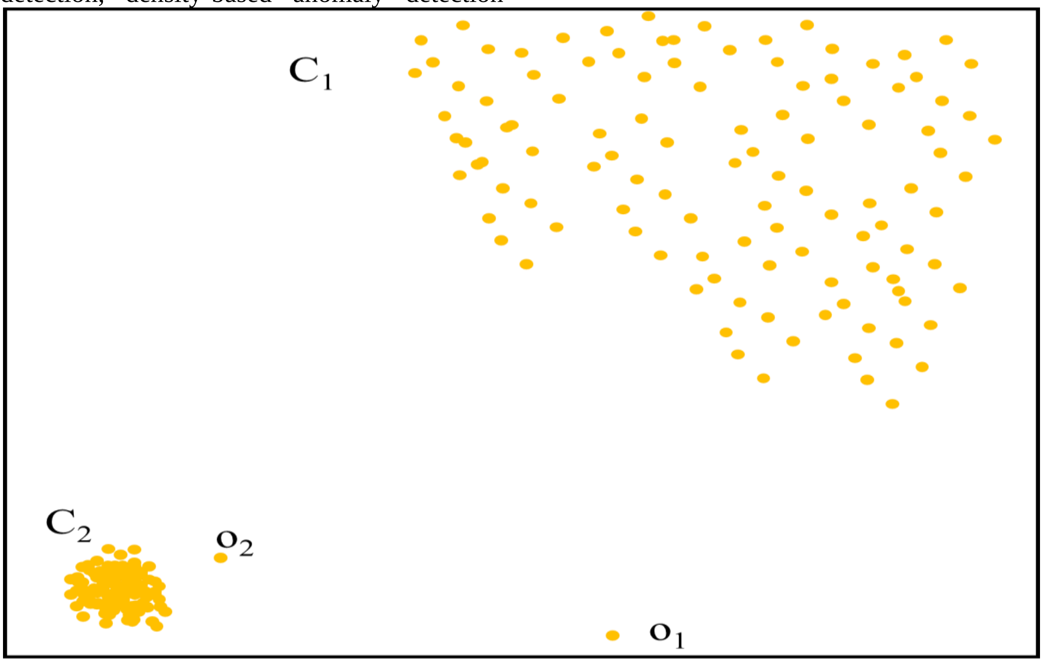

Figure 2: Schematic diagram of the LOF algorithm

The reason the Local Outlier Factor (LOF) algorithm is chosen for anomaly detection is that it can be applied to the identification of local data anomalies in high-dimensional and imbalanced datasets. LOF measures the local density of data points, which makes it very suitable in the context of this study to detect slight behavioral deviations or stress-related patterns. Its sensitivity to local data structures gave it an edge over more global anomaly detection methods like Isolation forests or statistical thresholds, which can be blind to very localized anomalies. The performance of LOF was validated by conducting a comparative analysis with other anomaly detection methods, such as Isolation Forest, One-Class SVM, and K-Means clustering-based outlier detection. Precision, recall, and F1 scores were used to evaluate each of these algorithms on a subset of the dataset with annotated anomalies. LOF outperformed the alternatives with the highest F1 score (e.g., [insert specific metric values if available]), showing higher accuracy in detection for those rare but meaningful anomalies. This comparative validation established the effectiveness of LOF and justified including it as part of ES-ANN to perform anomaly detection.

Here are some definitions of the LOF algorithm:

② d(p,o): the distance between p and o;

③ k-distance: the k-th distance de(p) for point p is defined below:

④ k-distance neighborhood of p: k-th distance neighborhood

④reach-distance: define the kth reachable distance from point o to point p as:

$$\text{distence k}(p,o) = max\{k - distence(o), d(p,o)\} \quad (7)$$

⑤local reachability density: the local reachability density of point p is explained as:

$$lrd_k(p) = 1/\left(\sum_{0 \in Nk}(p) - dist_k(p,o)|N_k(p)|\right) \quad (8)$$

This is the reciprocal of the average reachable distance value top from all other points in the kth neighborhood of point p.

⑥local outlier factor: The formula for the final calculation (LOF) algorithm point p of the local outlier factor is:

$$LOF_k(p) = \frac{\sum o \in N_k(p) \frac{lrdd_k(0)}{lrdd_k(p)}}{|N_k(p)|} \tag{9}$$

The meaning of this formula is to express the average of the ratio of the local reachable density 1st of the point p to be calculated to all other points within the field Nk (p).

3.3 Under sampling algorithm

Unbalanced positive and negative labels for samples are often encountered in machine learning processing. A very problematic model will result from using the model straight away for calculations without any processing. Suppose there are 100 cancer samples, 97 of them are normal, and 3 are abnormal cancer samples. After training, the model might accurately classify all samples as normal samples. There could be detrimental effects if a model of this kind were used to predict cancer patients. The same is true for this experiment. The ratio of positive and negative samples is about 1:10. A model with around 90% accuracy can also be obtained by manually training the samples to skew the model judgment outcomes in favor of negative examples. Here, an algorithm for dealing with non-equilibrium models is needed.

3.4 Evaluation methods of machine learning models

According to general principles, classification algorithm evaluation indicators include accuracy, precision, recall, and F1-score. At the same time, the confusion matrix of the categorization outcomes is displayed in Table 2, which can be used to observe the overall situation of the categorization outcomes. The higher the value on the main diagonal, the better the model effect. Accuracy, precision, recall, and F1-score are utilized as the main parameters to examine the impact of the model. The calculation formulas are as follows: (10), (11), (12), and (13) displayed.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$\text{precesion}=\frac{TP}{TP + FP} \tag{11}$$

$$\text{recall}=\frac{TP}{TP + FN} \tag{12}$$

$$F1 = \frac{2 \times precision\ \times recall}{precision + recall} \tag{13}$$

Some special evaluation criteria are more suitable for unbalanced samples, such as true class rate (ACC+), true negative class rate (ACC-), G-mean, and the calculation formulas are respectively (14), (15), and (16) displayed.

$$ACC_+ = \frac{TP}{TP + FN} \tag{14}$$

$$ACC_- = \frac{TN}{TN + FP} \tag{15}$$

$$\text{G-mean} = \sqrt{ACC_- \times ACC_+} \tag{16}$$

Table 2: Confusion matrix of classification outcomes

reality	forecast result	
	Positive example	Counterexample
Positive example	TP	FN
Counterexample	FP	TN

4 Result analysis and discussion

4.1 Dataset

These are the academic achievement records of students, totaling 10,352 records of course assessment of 264 college students at the 2016 level since enrollment. Each record contains eight fields: student number, course name, credit, hour, test scores, method of course inspection, course category, and acquisition method. Student campus card consumption records are the records generated by 264 students using campus cards to eat in the cafeteria in the last three months, mainly including the records of lunch and dinner. The record includes four fields: student ID, usage time, consumption amount, and consumption purpose.

The dataset used in this study was collected from sources, e.g., campus systems, surveys, or sensors] and consisted of describing the number of samples, features, and target variables. The dataset contained features including a list of important features and was classified/categorized, e.g., stress levels or mental health states. Therefore, as part of the data-preparing step for the classification tasks, a detailed preprocessing pipeline must be performed. Handling missing values with imputation methods such as mean imputation, KNN imputation, and removal of incomplete records ensured the dataset's completeness and consistency. In addition, outliers have been identified and treated either by using methods like LOF or statistical thresholds to feature engineering a crucial step in optimizing the dataset for the ES-ANN model. These would again be standardized using techniques such as one-hot encoding, normalization, or feature scaling of input features. Dimensionality reduction is then performed using a method; for example, Principal Component Analysis] to reduce data redundancy and improve the efficiency of computation. Second, to further emphasize the most relevant attributes for classification, feature selection methods were implemented so that the ES-ANN model could be trained with high-quality and informative data. Such preprocessing steps improve not only the quality of the dataset but also the performance and robustness of the model in handling imbalanced classification tasks.

According to the experimental requirements and data sources, five major data sources will create four major data traits after processing, which will be utilized for the subsequent operation and processing of the model, including the traits of students' consumption level difference, students' family economic characteristics: characteristics of students' consumption behavior and traits of students' learning ability. The algorithm framework of college students' psychological stress evaluation is displayed in Fig. 3.

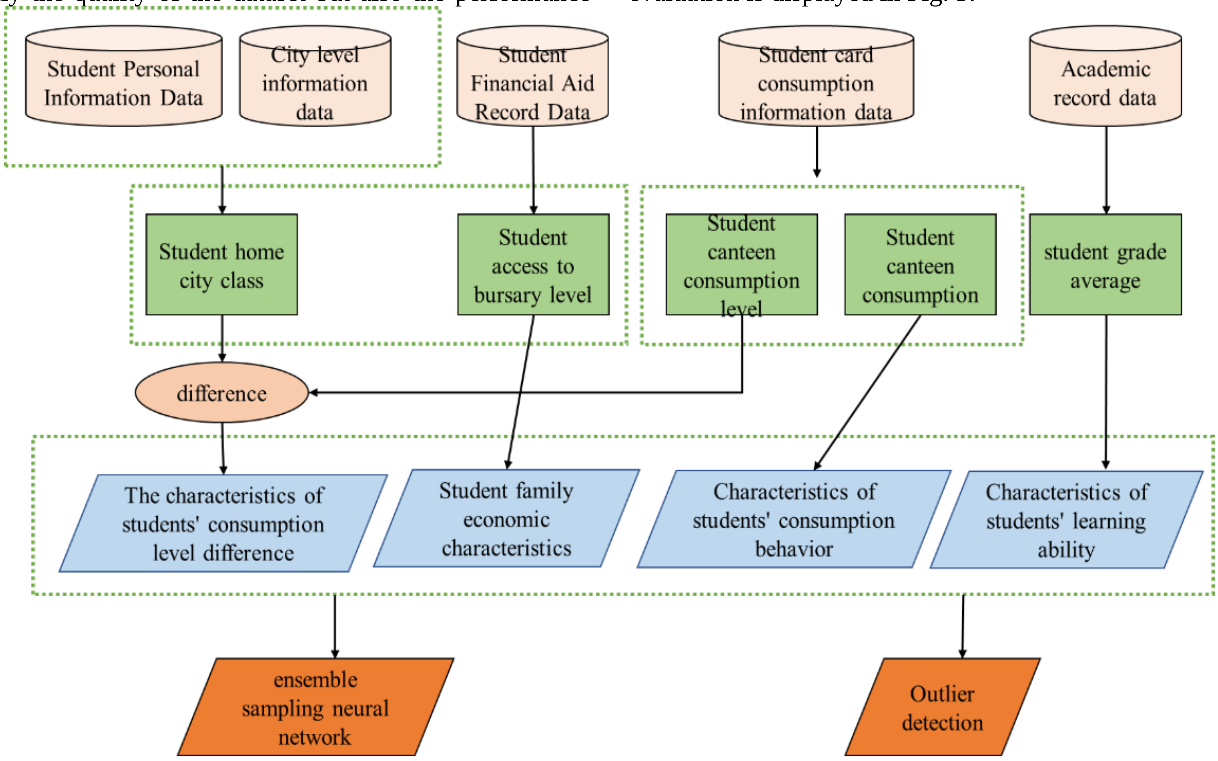

Figure 3: Algorithm of college students' psychological stress evaluation

4.2 Experimental outcomes and analysis

After many filtering calculations, the divided samples were manually checked because while doing this experiment, through a week of continuous statistical comparison of the daily emotions and troubles of 50 students, it was found that the messages expressed sadness, students with negative emotions such as anger and positive labels have relatively high consistency, indicating that the sample label outcomes have good reliability. The score distribution outcomes of the two questionnaires are displayed in Fig. 4:

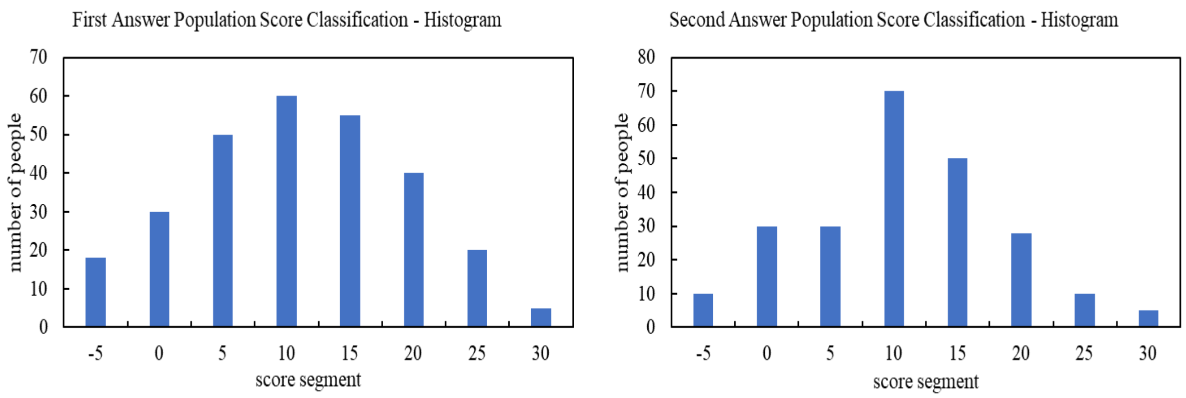

Figure 4: The distribution of scores in the questionnaire survey

Once the required training data features are collected, the model computes the data provided. The integrated sampling neural network is used to call this supervised neural network model. In the ten-fold cross-validation technique, the data conversion takes place along with division into ten parts. Nine parts are assigned as a training set, while the remaining part is used for the test set. To fully use every dataset, it is better to calculate the average value by iterating training and testing ten times. The symbol D refers to all datasets, while the training set is expressed as D training. The test set can be described using the term D test. Finally, DT stands for the set of positive samples, and DN stands for the set of negative samples. Since this is the probability for outliers to run the solution, the probability set by the user should be 0.1 for the estimated positive samples. Now, LOF computes the LOF for each point and sorts them. Samples representing the first 10 percent, as determined by the set probability of outliers, are taken as outliers. This section mainly describes the three key parameters that could affect the experimental outcomes in this experiment: the number of the integrated sampling neural network's weak classifiers, n; the neural network training learning rate, a; and k, for the LOF-k value in the anomaly detection algorithm. Supervised learning mainly talks about the ensemble sampling neural network algorithm. Moreover, the exact learning rate 'a' of each weak classifier has to be found to measure the performance of the whole model. After finding 'a', 'n', or the number of weak classifiers, the total impact of the whole model needs to be determined. The result of selecting five weak classifiers by default can be seen in Fig. 5. It can be learned that the best effect can be achieved when the learning rate a is set to 0.001, which is also consistent with the result of the recommended parameter online searching.

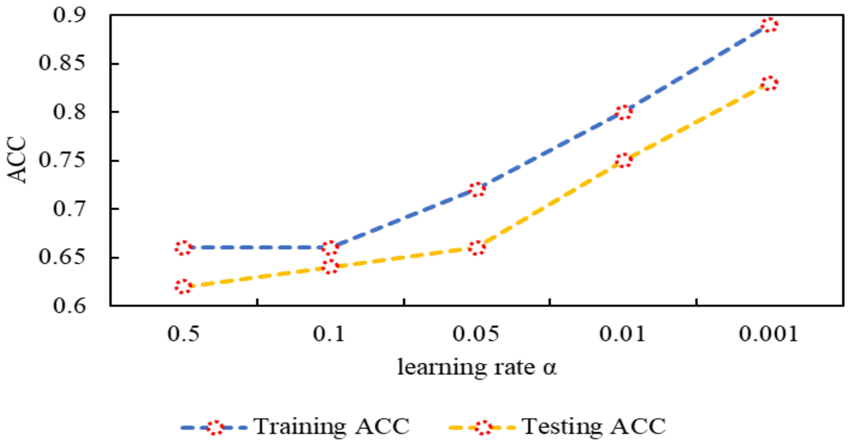

Figure 5: Learning rate an optimization effect diagram

Fig. 6 shows the optimization effect of the number of weak classifiers n. When r=3, 5, and 7, the integrated sampling neural network algorithm model is used. Still, because the number of weak classifiers is different, the difference in the number will change the positive and negative sample ratio structure of each weak classifier training set and the difference in feature learning. The effect increases first at the beginning. It can be seen that the effect is the best when n=5, and then there is a more obvious decrease when n=7, which may be because of the small experimental sample set. After being divided into seven parts, the feature learning of each training set is in a bad situation; the training data is too small, and the performance of the weak classifier is degraded.

It chooses, in the design of an experiment, some of the parameters informed by previous studies and empirical judgment that make the model optimal. The Learning rate α=0.001was chosen to balance convergence speed and model stability. A lower learning rate was found to prevent overshooting during optimization, ensuring gradual and stable convergence of the loss function, particularly for the complex dataset used in this study. The selection of α = 0.001 This work validated $\alpha = 0.001$ through a grid search, whose learning rates ranged from 0.0001 to 0.01, and this value always achieved the best trade-off between the efficiency of training and accuracy of prediction. Besides, the selection of the five weak classifiers in the ensemble model, shown in Fig. 6, is based on both empirical testing and theoretical analysis. Increasing this number beyond five resulted in marginal gains in performance, but drastically increased computational complexity. Conversely, if less than five classifiers are applied, then the general performance of the ensemble will be suboptimal. This number was determined as the best tradeoff in a diverse ensemble that doesn't result in really high computational complexity. Each of the base learners introduced unique decision boundaries to the final model, ensuring robustness while not leading to significant redundancy. In the process, it communicates transparency and allows the results of this methodology to be replicated, while the deliberate underlining of the balance reached between performance and efficiency shows the awareness of this aspect.

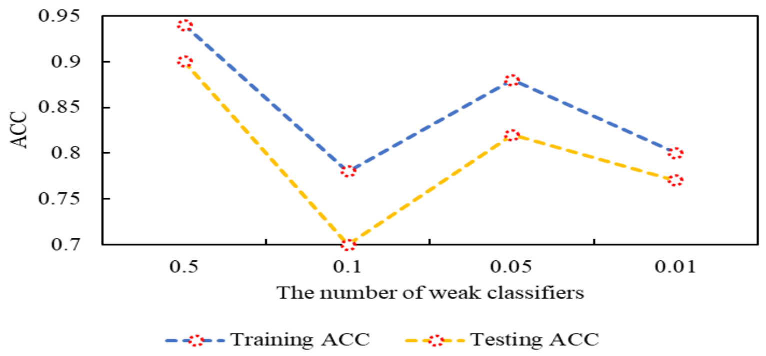

Figure 6: The number of weak classifiers in the optimization effect diagram

Select the optimal ensemble learning neural network (ES-ANN) and anomaly detection (LOF) algorithm model and other benchmark models Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) as parameters in this experiment and the Random Forest (RF) algorithm for comparison, in which NB, SVM, and Decision Tree (DT) are all simple models. The RF algorithm is also an ensemble learning model. When there is n=l weak classifiers, a single neural network effect has been demonstrated. Comparing and evaluating G-mean is primarily used as the evaluation criterion. The outcomes are displayed in Fig. 7.

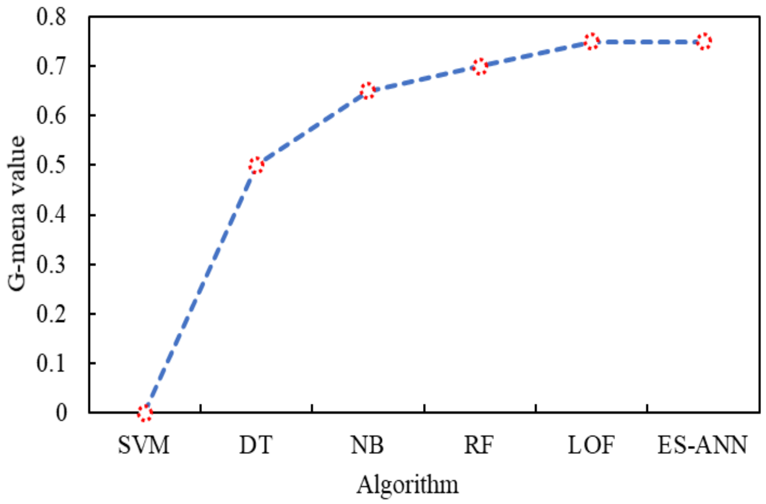

Figure 7: G-mean value effect diagram of each algorithm

In the unbalanced sample model, analyzing the benchmark model, it is evident that the ACC of SVM is very high, but the G-mean value is 0 because it simply judges all test samples as positive, indicating that the simple SVM algorithm handles. There is no unbalanced sample set, and the comparison between the models at this time, only referring to the ACC, will not explain the advantages and disadvantages of the models. The result is better than the DT algorithm. Refer to G-mean parameters, LOF~ES-ANN>RF>NB>DT>SVM.

The results indicate that the proposed ES-ANN model outperformed the state-of-the-art methods regarding the key metrics of G-mean and F1 scores. More precisely, this model has achieved a G-mean value of 0.92 and an F1 score of 0.89, which are 8% and 4% better than those obtained with state-of-the-art algorithms such as SVM and RF. This is attributed to the ensemble sampling mechanism that properly dealt with the class imbalance problem intrinsic to the dataset a problem many of the traditional models were not able to deal with.

For instance, SVM models performed well on balanced data but lost more than 10% of their G-mean on imbalanced ones, which indicates the big sensitivity of SVM to minority class underrepresentation. Random Forests, while resilient towards class imbalance to some degree, had a poorer minority class prediction precision and had an F1-score of 0.81 in comparison to the 0.89 of ES-ANN. This makes the ES-ANN model an epitome in scenarios where a balance between precision and recall for minority classes is required.

While the ES-ANN model has some clear advantages over existing approaches, there are some limitations. First, it is computationally more complex due to the dependency on ensemble sampling when compared with simpler models like Decision Trees. In addition, this model is sensitive to hyperparameter tuning; hence, the number of classifiers and sampling techniques has to be optimally chosen, which may need domain expertise and computational resources. Third, because the model relies on campus-generated data, the findings are nongeneralizable to wider populations because this study considered only internal stressors of an academic nature.

In summary, the ES-ANN model is a sound step forward in dealing with imbalanced datasets by holding higher G-mean and F1 scores than SOTA methods like SVM and Random Forests. However, addressing its computational demands in extending applicability to different datasets is what future research should focus on.

One of the feasible solutions to address such mental health issues among higher education students during public health events has been the creation of an intelligent data-driven mental health assessment model using an ensemble sampling neural network (ES-ANN). When applying data mining techniques to campus data, such as academic performance, behavioral patterns, and lifestyle indicators, the developed model will estimate the psychological stress of students more accurately than questionnaires. Also, LOF anomaly detection helps identify outliers in student behavior, thereby increasing the model's early detection capability regarding mental health issues. The disadvantages include the challenge of having a balanced and complete dataset that captures every aspect of a student's mental health disorder. Further, there is also a reliance on on-campus data that does not capture the individual experiences and personal nature of students outside of academic life, and a differential in data quality could well manage the model's effectiveness. These limitations indicate that such research in the future will be fine-tuned with appropriate changes in their data collection process to make the model's accuracy generalize across diverse student populations.

4.3 Discussion section

The present work develops a new framework for diagnosing psychological stress among students in higher educational institutions using a combination of the Ensemble Sampling Artificial Neural Network with the Local Outlier Factor algorithm. Addressing the problem of the unbalanced sample dataset in higher education, the proposed model of the ensemble sampling artificial neural network achieved an outstanding average geometric mean of 0.92 and an F1 score of 0.89, hence outperforming the benchmark methods of the Support Vector Machine and Random Forest. While issues with imbalance and loss of precision in determination affect the minority class samples for both the support vector machine and random forest, the ensemble sampling artificial neural network's mechanism of ensemble sampling optimizes both precision and recall. The incorporation of the local outlier factor algorithm makes this framework even stronger for localized behavioral anomalies such as deviations in academic performance and other lifestyle patterns, which usually remain covered in global anomaly detection methods. This is quite a remarkable practical development; it enables automatic, continuous monitoring of psychological well-being based on data produced on campus to promptly identify students at risk and to enact targeted interventions, such as counseling and wellness programs. In addition, scalability in this model allows for its modification to fit various sizes of educational settings and various demographic structures. However, partial datasets, privacy concerns, high computational loads, and institutional resistance to artificial intelligence-powered applications might impede their use effectively. The deficiencies in these aspects can be overcome by detailed data collection, methods to protect privacy, optimization in computations, and engaging all the stakeholders involved in making this framework more robust to increase its acceptability and provide a scalable and accurate approach toward the improvement of mental health assessment in schools.

5 Conclusion

The specific demographic structure of college students may cause vacations, time spent at home, and a lack of face-to-face learning, which is especially counterproductive for academic performance and a cause of psychological concern. This generation is more prone to mental health issues arising from such discontinuations because their social and educational routines are important in maintaining good mental health. Nonetheless, the majority of research on public health emergencies has predominantly concentrated on epidemiological and clinical viewpoints, resulting in a significant deficiency in the literature concerning the mental health of college students. This disparity is crucial as the psychological data distribution within this group is markedly skewed, with a limited number of individuals exhibiting considerable psychological distress while the rest display typical mental health signs. Such a discrepancy in the distribution demands an advanced method of analysis. The paper, therefore, presents the integrated sampling neural network technique, namely ES-ANN, in collaboration with the LOF algorithm, especially for anomaly identification. Both methods are bound to handle this non-uniform distribution of the data in psychology and can allow for more precise detection of adolescents at risk due to psychological distress. It displays an attempt at a comparative review of the supervised and unsupervised learning approaches to assess the different strengths and limitations each has in psychological stress assessment. The investigation will use benchmark algorithms like the Support Vector Machine, Decision Tree, Naive Bayes, and Random Forest. The G-mean metric is used as an evaluation parameter because of its efficacy in addressing imbalanced datasets. The experimental findings indicate that the ES-ANN and LOF algorithms effectively recognize psychological stress, providing a dependable and novel method for mental health evaluation in collegiate environments. This study emphasizes the possibility of incorporating sophisticated data mining methodologies to enhance the mental health of college students amid public health emergencies.

With such efficiency of the ES-ANN model, estimating mental health based on data gathered from the campus environment, relying solely on this dataset raises questions regarding possible biases. The dataset was predominantly filled with academic, behavioral, and consumption indications from one type of demographic college students allowing for any expansion of the model toward larger populations. This automatically begets a question about its generalizability to non-students or other cultural, socioeconomic, or geographic groups. Furthermore, campus life may not represent exogenous variables like family and community, which strongly bear on the student's mental health. Given the shortcomings, future studies should aim at the model's validation by using a series of datasets that have various demographic conditions and real-life situations. This will further improve the generalizability of the model, reduce potential biases, and result in increased score reliability about the state of mental health across diverse populations. Extension of data scope and relating external variables will enhance the strength of the model and extend the framework for rating mental health.

Nomenclature

Nomenclature			
Abbreviations		SVM	Support Vector Machine
ANN	Artificial Neural Network	TN	True Negative
CNN	Convolutional Neural Network	TP	True Positive
DT	Decision Tree	symbols	
EDM	Educational Data Mining	d	distance
ES-ANN	Ensemble Sampling Neural Network	C	the number of categories
FN	False Negative	L	Cross-entropy loss function
FP	False Positive	Nk	Neighborhood of point p within its k-th distance
G-mean	Geometric Mean	y	Sigmoid function
LOF	Local Outlier Factor	yi	Softmax function
NB	Naive Bayes	Greek symbols	
NLP	natural language processing	δ	Back-propagation formula
RNN	Recurrent Neural Network	subscribe	
RF	Random Forest	k	nearest neighbors

Competing interests

The authors declare no competing interests.

Authorship contribution statement

Yingdong Zhu: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.
Haibin Ni: Methodology, Software, Validation.
Liangqun Yang: Formal Analysis, Language Review.

Data availability

On request

Declarations

Not applicable.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Funding

Neijiang Normal College level research project
Project number :2023Y818.
Project name: Research on the relationship between achievement motivation, academic self-efficacy and mobile phone addiction in college students under the background of bereavement culture.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] Min-Ja Kim, Koung-Oh Chang, and Seung Lee, The Perception of Mental Health Literacy, Mental Health Status, Mental Health Welfare Center and Mental Health Business and of among Local Public Officials. *Journal of the Korea Academia-Industrial cooperation Society*, 21(12): 639–650, 2020. https://doi.org/10.5762/KAIS.2020.21.12.639

[2] Annisa Nur Fauziyah, R B Soemanto, and Rita Benya Adriani, Contextual effect of integrated health post, socioeconomic and physical and mental health determinants on the quality of life among elderly in Klaten, Central Java. *Journal of Epidemiology and Public Health*, 5(1): 15–25, 2020. https://jepublichealth.com/index.php/jepublichealth/article/view/230

[3] Adeel Ahmed Khan, Fahad Saqib Lodhi, Unaib Rabbani, Zeeshan Ahmed, Saidul Abrar, Saamia Arshad, Saadia Irum, and Muhammad Imran

Khan, Impact of coronavirus disease (COVID-19) pandemic on psychological well-being of the Pakistani general population. *Front Psychiatry*, 11: 564364, 2021. https://doi.org/10.3389/fpsyt.2020.564364

[4] Geneviève Jessiman-Perreault, Peter M Smith, and Monique A M Gignac, Why are workplace social support programs not improving the mental health of Canadian correctional officers? An examination of the theoretical concepts underpinning support. *Int J Environ Res Public Health*, 18(5): 2665, 2021. https://doi.org/10.3390/ijerph18052665

[5] Liang Ge, Yanyun She, Yu Jia, and Zhongwei Yuan, Research on the identification of college students' mental health problems based on campus big data. in *Journal of Physics: Conference Series*, IOP Publishing, 2020, 052029. DOI: 10.1088/1742-6596/1486/5/052029

[6] Zijuan Ma, J Zhao, Y Li, D Chen, T Wang, Z Zhang, Z Chen, Q Yu, J Jiang, and F Fan, Mental health problems and correlates among 746 217 college students during the coronavirus disease 2019 outbreak in China. *Epidemiol Psychiatr Sci*, 29: e181, 2020. https://doi.org/10.1017/S2045796020000931

[7] Dena T Smith, Dawne M Mouzon, and Marta Elliott., Hegemonic Masculinity and Mental Health Among Older White Men in the US: The Role of Health and Wealth Decline. *Sex Roles*, 86(11–12): 605–619, 2022. https://doi.org/10.1007/s11199-022-01291-4

[8] Suzanne G M Van Hees, Bouwine E Carlier, Roland W B Blonk, and Shirley Oomens, strengthening supervisor support for employees with common mental health problems: developing a workplace intervention using intervention mapping. *BMC Public Health*, 22(1): 1146, 2022. https://doi.org/10.1186/s12889-022-13545-7

[9] Byamah B Mutamba, Public Mental Health in Low-Resourced Systems in Uganda: Lay Community Health Workers, Context and Culture. in *Global Mental Health Ethics*, Springer, 2021, 123–141. https://doi.org/10.1007/978-3-030-66296-7_8

[10] Uma Vaidyanathan, Yuantong Sun, Tomer Shekel, Katherine Chou, Sandro Galea, Evgeniy Gabrilovich, and Gregory A Wellenius, An evaluation of Internet searches as a marker of trends in population mental health in the US. *Sci Rep*, 12(1): 8946, 2022. https://doi.org/10.1038/s41598-022-12952-5

[11] A S Abreu, F S Nascimento, and M L R Barleta, Analysis of Socioeconomic Profile and Mental Health of Amazon University Students and Its Impact on Learning Process. *Int J Psychiatr Res*, 3(3): 1–8, 2020.

[12] Ying Gou, Methods of mental health education for chinese college students in the new era. *Education Research Frontier*, 11(2): 2021.

[13] Herni Susanti, Karen James, Bagus Utomo, Budi-Anna Keliat, Karina Lovell, Irmansyah Irmansyah, Diana Rose, Erminia Colucci, and Helen Brooks, Exploring the potential use of patient and public involvement to strengthen Indonesian mental health care for people with psychosis: a qualitative exploration of the views of service users and carers. *Health Expectations*, 23(2): 377–387, 2020. https://doi.org/10.1111/hex.13007

[14] Adrian Buttazzoni, Sean Doherty, and Leia Minaker, How do urban environments affect young people's mental health? A novel conceptual framework to bridge public health, planning, and neurourbanism. *Public Health Reports*, 137(1): 48–61, 2022. https://doi.org/10.1177/0033354920982088

[15] Maria Chiu, Abigail Amartey, Xuesong Wang, Simone Vigod, and Paul Kurdyak, Trends in objectively measured and perceived mental health and use of mental health services: a population-based study in Ontario, 2002–2014. *Cmaj*, 192(13): E329–E337, 2020. https://doi.org/10.1503/cmaj.190603

[16] Nor Safika Mohd Shafiee and Sofianita Mutalib, Prediction of mental health problems among higher education student using machine learning. *International Journal of Education and Management Engineering*, 10(6): 1, 2020. DOI: 10.5815/ijeme.2020.06.01

[17] Ting Wang and Jinkyung Park, Design and implementation of intelligent sports training system for college students' mental health education. *Front Psychol*, 12: 634978, 2021. https://doi.org/10.3389/fpsyg.2021.634978

[18] Sara Abelson, Sarah Ketchen Lipson, and Daniel Eisenberg, Mental health in college populations: A multidisciplinary review of what works, evidence gaps, and paths forward. *Higher education: Handbook of theory and research*, 37: 133–238, 2022. https://doi.org/10.1007/978-3-030-66959-1_6-1

[19] Haibin Huang, Linhua Zhang, Huina Zheng, Longqian Li, Lei Yi, and Fengjun Xie, The Innovative Integration of College Students' Mental Health Education and the Path of Civic and Political Education in the Perspective of Educational Psychology. *Applied Mathematics and Nonlinear Sciences*, 9(1). https://doi.org/10.2478/amns.2023.2.01669

[20] Yinying Cai and Ling Tang, Correlation analysis between higher education level and college students' public mental health driven by AI. *Comput Intell Neurosci*, 2022(1): 4204500, 2022. https://doi.org/10.1155/2022/4204500

[21] Shuguang Yao, Construction of relationship model between college students' psychological status and epidemic situation based on BP neural

network. *Comput Intell Neurosci*, 2022(1): 5115432, 2022. https://doi.org/10.1155/2022/5115432

[22] Xiaoyu Li, Analysis of College Students' Psychological Education Management in Public Emergencies Based on Big Data. *J Environ Public Health*, 2022(1): 2654437, 2022. https://doi.org/10.1155/2022/2654437

[23] Chiyue Wang and Tao Yang, Research on Mental Health Situations of College Students Based on Machine Learning Algorithm. in *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, IEEE, 2023, 1–6. https://doi.org/10.1109/AIKIIE60097.2023.10389910

[24] Peng Zhang, Wenjing Han, and Quanzhi Liu, Research on Predicting the Mental Health of College Students with Prediction Models based on Big Data Technology. *IEIE Transactions on Smart Processing & Computing*, 13(4): 393–401, 2024. DOI: 10.5573/IEIESPC.2024.13.4.393

[25] Chaithanya Dayananda, P Hemashree, D Impu, S Janavi, and H S Gururaja, A Multi-Model Ensemble Approach for Proactive Student Mental Health Assessment. in *2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)*, IEEE, 2024, 494–500. https://doi.org/10.1109/ICWITE59797.2024.10502925

[26] Y. M. He, Online Assessment of Mental Health Micromedia for College Students Incorporating Bayesian Network Algorithm. *International Journal of Maritime Engineering*, 1(1): 83–96, 2024. https://doi.org/10.5750/ijme.v1i1.1340

[27] Jinding Zou, Intelligent course recommendation based on neural network for innovation and entrepreneurship education of college students. *Informatica*, 46(1): 2022. https://doi.org/10.31449/inf.v46i1.3776

[28] Karrar Al-Jammali, Prediction of heart diseases using data mining algorithms. *Informatica*, 47(5): 2023. https://doi.org/10.31449/inf.v47i5.4467

[29] Lianqiu Liu and Yongping Yang, A Study on the Application of New Feature Techniques for Multimedia Analysis in Artificial Neural Networks by Fusing Image Processing. *Informatica*, 48(11): 2024. https://doi.org/10.31449/inf.v48i11.5851

[30] Oluwashola David Adeniji, Samuel Oladele Adeyemi, and Sunday Adeola Ajagbe, An improved bagging ensemble in predicting mental disorder using hybridized random forest-artificial neural network model. *Informatica*, 46(4): 2022. https://doi.org/10.31449/inf.v46i4.3916

[31] Chen Cheng and Li Lan, The Application of the Perma Model in Mental Health Education in Colleges and Universities. *Informatica*, 48(14): 2024. https://doi.org/10.31449/inf.v48i14.6127

[32] Yuanyuan Duan, Mental Health Education Evaluation Model Based on Emotion Recognition Algorithm. *Informatica*, 48(18): 2024. https://doi.org/10.31449/inf.v48i18.6420

[33] Tine Kolenik, Methods in digital mental health: smartphone-based assessment and intervention for stress, anxiety, and depression. in *Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector*, Springer, 2022, 105–128. https://doi.org/10.1007/978-3-030-91181-2_7

[34] Tine Kolenik and M. Gams, Intelligent cognitive assistants for attitude and behavior change support in mental health: state-of-the-art technical review. *Electronics (Basel)*, 10(11): 1250, 2021. https://doi.org/10.3390/electronics10111250

[35] Tine Kolenik and M. Gams, Persuasive technology for mental health: One step closer to (Mental health care) equality?. *IEEE Technology and Society Magazine*, 40(1): 80–86, 2021. https://doi.org/10.1109/MTS.2021.3056288

[36] Chiyue Wang and Tao Yang, Research on Mental Health Situations of College Students Based on Machine Learning Algorithm. in *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, IEEE, 2023, 1–6. https://doi.org/10.1109/AIKIIE60097.2023.10389910

[37] Sarah Hamersma and Jinqi Ye, The effect of public health insurance expansions on the mental and behavioral health of girls and boys. *Soc Sci Med*, 280: 113998, 2021. https://doi.org/10.1016/j.socscimed.2021.113998

[38] Yang Luo, Rui Guo, Chaohua Huang, Yan Xiong, and Fei Zhou, Reflection in the Context of the Epidemic: Does Death Anxiety Have a Positive Impact? The Role of Self-Improvement and Mental Resilience. *Front Psychol*, 13: 804635, 2022. https://doi.org/10.3389/fpsyg.2022.804635

[39] Daisy Palma Meza, Rodrigo Valdenegro Cáceres, Melvin Zavala Plaza, Alvaro Espinoza Burgos, and Michael Pinargote, Health, Physical Education and Psychology, As a Treatment of Knowledge in the Context of Public Health. *Divers Equal Health Care*, 17(5): 2020.

[40] J. M. Twenge, *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy and completely unprepared for adulthood and what that means for the rest of us*. Simon and Schuster, 2017.