PLOS One

RESEARCH ARTICLE

Utilizing multi-level convolutional neural networks to achieve refined modeling and visual analysis of college students' mental health data

Xianwei Huang*, Wei Jiang

Faculty of Intelligent Transportation, Anhui Sanlian University, Hefei, Anhui, China

* zhjt@slu.edu.cn

Check for updates

OPEN ACCESS

Citation: Huang X, Jiang W (2025) Utilizing multi-level convolutional neural networks to achieve refined modeling and visual analysis of college students' mental health data. PLoS One 20(10): e0335048. https://doi.org/10.1371/journal.pone.0335048

Editor: Hikmat Ullah Khan, University of Sargodha, PAKISTAN

Received: February 17, 2025

Accepted: October 6, 2025

Published: October 24, 2025

Copyright: © 2025 Huang, Jiang. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The data that support the findings of this study are openly available in Mendeley Data at http://doi.org/10.17632/9tj726d4p5.1.

Funding: This work was supported by the Anhui Provincial Higher Education Scientific Research Project (Philosophy and Social

Abstract

Early identification of students' mental health issues has become an urgent priority in education and public health. However, existing studies often rely on questionnaire-based assessments or traditional machine learning models, which are limited by manual feature design and weak ability to capture the multidimensional and dynamic characteristics of psychological data. This creates a research gap in developing more adaptive and automated approaches for reliable prediction and monitoring. To address this limitation, the present study proposes the use of Convolutional Neural Network (CNN) for mental health modeling, taking advantage of its capability to automatically extract hierarchical features from multimodal inputs. For comparative purposes, Gradient Boosting Decision Tree (GBDT) and Support Vector Machine (SVM) are also implemented as baseline methods. A dataset combining academic performance, emotional fluctuations, social behavior, and lifestyle indicators was preprocessed and used for experiments.Results demonstrate that CNN achieves the highest predictive accuracy of 94%, compared to 89% for SVM and 87% for GBDT. Beyond accuracy, CNN also shows faster convergence and greater robustness across k-fold cross-validation. These findings highlight the significance of CNN as a more powerful tool for handling high-dimensional psychological data. The study contributes to bridging the gap between traditional mental health assessment and intelligent data-driven approaches, providing practical value for early risk detection and personalized interventions among students.

1. Introduction

With more mental health problems among college students, how to check, measure, and help their mental health has become an important topic in psychology and education [1]. In recent years, mental health problems of college students, like depression, anxiety, and stress, have become big factors that affect their studies, social life, and health [2,3]. Old ways of checking mental health, like self-tests, behavior checks,

RETRACTED

Sciences): Research on the application of artificial intelligence technology in the mental health survey of college students(Project No.2022AH051975).

Competing interests: The authors have declared that no competing interests exist.

and regular talks, can show some of the students' mental states, but they are limited by the personal judgment of evaluators. And because of time and energy limits, these ways cannot follow the changes of many students in real time, so the data is not always correct or wide enough. Also, as student groups become more diverse, old ways often cannot deal with complex mental health data, especially when the data is not simple and has many sides [4,5].

To deal with these problems, new deep learning methods, especially convolutional neural networks (CNNs), have worked well in many data fields. CNNs are good at finding patterns and details, so they can give new ways to study mental health data. Old ways often fail with data that is large and complex, but CNNs use many layers to copy how the brain works, and they can catch hidden and hard-to-see links in the data [6,7]. This makes CNNs useful for college students' mental health data, which has many kinds of inputs, like self-tests, behavior, social life, and mood changes [8]. What makes CNNs special is that they can pull out deep features step by step, and then they show important details in the data. This helps to find risks like unstable emotions, social problems, and stress in studies, which are hard for old ways to see [9,10].

CNN-based systems can also use data visualization in new ways [11]. By showing the results of the network's layers, they give a clear view of students' mental states and what affects them. For example, heatmaps and radar charts help mental health workers see the links between stress, emotions, and social skills [12]. This makes it easier to understand the data and gives a base for direct help. By looking at these visuals, workers can find risks like anxiety or depression and give quick and personal support [13]. CNNs can also mix and study data from many sources, like grades and lifestyle, which makes predictions better [14]. Using transfer learning also lowers the need for big training datasets and makes the model more steady and fast, so it fits better for real use.

In the future, the use of multi-level CNNs for checking college students' mental health shows great promise. This way not only makes evaluation faster and reduces human bias, but it also gives more dependable results for quick and personal help. As algorithms and data use keep getting better, CNN-based systems can play a big role in mental health checks, prevention, and support for both students and professionals.

In the future, the use of multi-level CNNs for checking college students' mental health shows great promise. This way not only makes evaluation faster and reduces human bias, but it also gives more dependable results for quick and personal help. As algorithms and data use keep getting better, CNN-based systems can play a big role in mental health checks, prevention, and support for both students and professionals:

(1) Propose a CNN design that can pull deep features from many kinds of data, like surveys, grades, and behavior;

(2) Make mental health predictions more correct and easier to understand by using data visualization to show how the model learns and makes choices;

RETRACTED

(3) Compare how the CNN model works against basic machine learning models like Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBDT);

(4) Test how CNN-based models can be used in real mental health monitoring, to give both theory and practice for support in schools..

This paper proposes a multi-level CNN model to solve these problems and make mental health checks for college students more correct and easier to understand. The main points of this paper are:

(1) A new CNN design that uses skip connections to avoid the vanishing gradient problem and to make feature extraction better.

(2) A clear look at how multi-level CNNs can work with complex mental health data, like self-tests, behavior, and grades.

(3) The use of data visualization, like heatmaps and accuracy curves, to make the model's results easier to read and explain.

The rest of this paper is set up like this: Section 2 looks at past work on mental health data using machine learning and deep learning. Section 3 shows the method, including the CNN design, data preparation, and feature extraction. Section 4 explains the setup and the results.

2. Related work

2.1 Current status of psychological health research on college students

College students' mental health has become a widely recognized research concern, with issues such as anxiety, depression, and academic stress increasingly reported [15–18]. Existing studies indicate that these problems not only affect students' academic performance but also their long-term psychological development [19,20]. Traditional assessment tools, such as questionnaires and standardized scales, have contributed valuable insights but remain limited by subjectivity and insufficient ability to capture complex and dynamic psychological states [21,22].

Recent research trends highlight the introduction of artificial intelligence and machine learning methods to address these limitations [23]. In particular, convolutional neural networks (CNNs) have been widely applied for psychological health analysis, showing strong capacity in handling high-dimensional data and identifying nonlinear relationships among mental health indicators [24–27]. Furthermore, multimodal data integration (e.g., behavioral, physiological, and textual signals) has been increasingly emphasized as a way to improve diagnostic accuracy and robustness [28,29].

Nevertheless, several challenges remain, including the imbalance of datasets, ethical concerns around data collection, privacy protection, and difficulties in translating analytical outcomes into effective intervention strategies [30–33]. Therefore, future directions should focus on building data-driven, scalable, and interpretable approaches to enhance psychological health assessment for college students [34,35].

2.2 Psychological health data analysis and modeling methods

In recent years, machine learning based methods have made significant progress in mental health research. Traditional statistical analysis methods, such as regression analysis and cluster analysis, although able to provide certain research value, often cannot well reveal complex nonlinear relationships in data due to their high requirements for data and relatively simple model structures. In contrast, deep learning, especially convolutional neural networks (CNNs), has achieved great success in fields such as image recognition and natural language processing [36]. Its superior feature extraction ability and ability to process high-dimensional data make it show great potential in psychological health data analysis. Psychological health data itself often has complex high-dimensional features, such as emotional fluctuations, behavioral patterns, social interactions, etc., which cannot be effectively extracted and modeled through traditional analysis methods

RETRACTED

[37–39]. CNN, through its multi-level structure, can automatically extract multi-level features from data, thereby achieving fine-grained modeling of mental health status [40].

To model the mental health data of college students using a multi-level convolutional neural network (CNN), sufficient preprocessing of the data is required first. The mental health data of students usually includes information from different sources, such as questionnaire surveys, sentiment analysis, social media data, academic performance, and lifestyle habits, etc. These data have multimodal characteristics. By fusing the features of these multimodal data, a more comprehensive input can be provided for CNN models. In model design, convolutional layers play a central role in extracting both low-level and high-level features from input data, such as students' emotional fluctuations, social behavior, and sleep quality. These features help reveal mental health status and its dynamic changes, as convolutional kernels can automatically identify hidden patterns, improving adaptability and accuracy [41]. Unlike traditional methods that rely on manually designed features, convolutional neural networks can learn optimal representations through end-to-end training, reducing human bias and enhancing generalization [42]. During training, the network continuously optimizes parameters via backpropagation, which strengthens the accuracy and robustness of predictions [43].

Recent research on mental health prediction and modeling has expanded beyond traditional convolutional neural networks (CNNs), with an increasing emphasis on integrating advanced methods and domain knowledge. CNNs have been widely recognized for their strong capability in feature extraction, but they are no longer the sole focus of state-of-the-art studies. For example, Pan et al. [44] proposed a miner mental state evaluation framework based on multidomain EEG information, combining decision-level fusion with machine learning classifiers such as SVM and kNN. Their study highlights the importance of fusing temporal, frequency, and spatial features for achieving robust performance, reaching an accuracy of 93.19%. Similarly, Wu et al. [45] addressed the problem of happiness prediction by integrating sociological domain knowledge with machine learning models, emphasizing explanation consistency through domain-constrained factor attribution. This work demonstrates that the integration of domain knowledge can enhance both predictive accuracy and interpretability in large-scale psychological modeling. Beyond psychology-specific applications, Li et al. [46] conducted a large-scale network meta-analysis in pharmacology education involving 6,447 students. Their findings showed that new media teaching strategies significantly improved performance, self-learning, and satisfaction, especially among nursing and junior college students. This illustrates the broader applicability of advanced data-driven models for evaluating complex educational and psychological outcomes.

To further improve assessment performance, researchers have explored hybrid models that combine multi-level CNNs with recurrent neural networks (RNN) and long short-term memory (LSTM) networks. These approaches enable dynamic monitoring of students' psychological states across different time periods by handling both sequential and nonlinear data [47,48]. For example, integrating academic performance, emotional states, and social interaction frequencies allows hybrid models to better capture mental health trends and support timely interventions. Finally, visualization is a crucial step. By using heatmaps and feature importance maps, complex model outputs can be transformed into intuitive results, helping educators and mental health experts identify key risk factors and early warning signals. Based on such visual analyses, personalized reports can be generated to guide students in understanding their psychological state and adopting effective coping strategies.

Overall, although multi-level convolutional neural networks perform well in modeling mental health data, they still face some challenges. Firstly, due to the diverse sources and high privacy of mental health data, ensuring data security and privacy protection is an urgent issue that needs to be addressed [49]. In addition, how to improve the generalization ability of the model to adapt to changes in different populations and environments is also a hot topic in current research. Researchers are actively exploring methods such as data augmentation and transfer learning to enhance the robustness and adaptability of models [50]. The analysis and modeling of college students' mental health data based on multi-level convolutional neural networks can not only provide more accurate mental health assessments, but also provide scientific basis for the formulation of psychological intervention and prevention measures, and provide strong support for students' physical and mental health development.

RETRACTED

3. Method

3.1 Multi level convolutional neural network model architecture

To achieve refined modeling and visual analysis of college students' mental health data, this paper proposes a deep learning model based on multi-level convolutional neural networks (CNNs), As shown in Fig 1. The model is designed to automatically extract meaningful patterns and relationships from complex, high-dimensional data by learning hierarchical features at multiple levels. Unlike other deep learning models, CNNs are particularly well-suited for this task due to their ability to capture spatial and temporal dependencies in structured data, which is essential for mental health data that may include time-series information, behavioral patterns, and diverse input modalities. temporal dependencies refer to the dynamic changes in students' mental health over time, such as fluctuations in mood, sleep quality, or social interactions across weeks or semesters. Spatial dependencies, on the other hand, do not refer to physical space, but rather to the structural relationships among multiple features measured at the same time point.

To improve clarity, several technical terms used in the model design are briefly explained here. In very deep neural networks, the vanishing gradient problem may occur, where gradients become extremely small during backpropagation, hindering effective training. To alleviate this problem, the skip connection introduces a direct pathway that adds earlier features to deeper layers, thereby stabilizing the learning process. Based on this principle, Residual Networks (ResNet) were developed, enabling the training of substantially deeper architectures through systematic use of skip connections. Another important technique is feature fusion, which integrates low-level detailed features (e.g., edges or small fluctuations) with high-level semantic features (e.g., behavioral patterns). This combination allows the network to capture both fine-grained and abstract information, resulting in richer and more informative representations.

Unlike traditional CNNs, which extract features from a single input modality through shallow layers, our model introduces a hierarchical feature extraction mechanism that progressively refines the features across successive convolutional layers, capturing increasingly abstract patterns. By incorporating advanced data fusion, hierarchical learning strategies, and specialized preprocessing techniques, our approach enhances prediction accuracy and robustness, making it more resilient to noise and missing data. Additionally, the integration of diverse data types—such as academic performance, behavior patterns, and social media interactions—enables a more comprehensive understanding of student mental health.

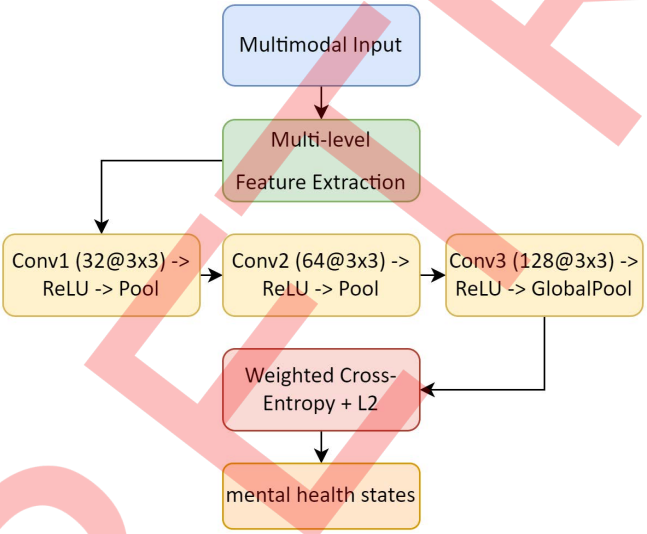

Fig 1. The architecture of the Multi-layer CNN model.

https://doi.org/10.1371/journal.pone.0335048.g001

RETRACTED

Visualization tools, such as heatmaps and feature importance plots, further improve interpretability, addressing the "black-box" nature of deep learning models and offering greater transparency for mental health professionals. These innovations enable our model to outperform standard CNNs, providing a more effective and robust solution for monitoring and intervening in college students' mental health.

The CNN model has several main layers, each made to pull out more complex features step by step. The input layer first takes in different kinds of student mental health data, like answers from self-tests, mood analysis, and lifestyle habits. After the data is cleaned and made standard, it is put into a matrix that can be used by the CNN. The first convolutional layer uses kernels on the input data. These kernels find small local features, like simple patterns or trends in single data points. This helps the model catch basic parts of mental health data, which later layers can use. Each next convolutional layer then finds more abstract features, based on what the earlier layers learned. This step-by-step process lets the CNN see complex links between different mental health signs and how they affect each other. To make the structure clearer, we show a diagram of the network. The diagram shows each layer and its job, and it also shows how data moves through the network. It highlights the role of convolutional and pooling layers in finding spatial and step-by-step features, and the role of the fully connected layers in making predictions about student mental health.

CNNs can efficiently process high-dimensional data, which is essential in contexts where multiple data types, such as text and numerical values, must be integrated and analyzed simultaneously. To further illustrate the technical process, assuming the input is a matrix($X \in \mathbb{R}^{H\times W\times C}$), where ($H$)represents height, ($W$)represents width, and represents the (C) number of channels, the convolution operation can be represented as:

$$Y_{i,j} = \sum_{m=1}^{K}\sum_{n=1}^{K} W_{m,n}X_{i+m,j+n} + b \tag{1}$$

Where (W)is the convolution kernel, is the (b)bias term, and is the ($Y_{i,j}$)output matrix after convolution. By utilizing the convolutional layers, the network is able to learn localized patterns within the data, such as varying expressions of psychological responses like anxiety or depression. Following each convolutional operation, a pooling layer is applied to compress the feature representations, thereby lowering the dimensionality of the data while preserving essential information. In particular, the Max Pooling strategy is employed to retain the most salient features. This operation can be formally described as:

$$Y_{i,j} = \max\left(X_{2i-1,2j-1}, X_{2i-1,2j}, X_{2i,2j-1}, X_{2i,2j}\right) \tag{2}$$

In this manner, the pooling layer reduces the dimensionality of the feature space while preserving critical information, thereby improving the model's robustness. Following successive convolution and pooling stages, the network ultimately generates a high-dimensional feature vector that effectively captures and characterizes the intrinsic structure of the input data.

Next, the model implements the classification task through a fully connected layer. Assuming that the features after convolution and pooling are represented as vectors($Z \in \mathbb{R}^{d}$), the fully connected layer maps them to the final output space through a linear transformation:

$$y = W^{T}Z + b \tag{3}$$

Where (W)is the weight matrix of the fully connected layer, (y)(b)is the bias term, and is the output of the model. In the mental health assessment task, the output of the model (y)corresponds to the classification of the mental health status of college students, such as anxiety, depression, stress, etc.

RETRACTED

To further improve the performance of the model, we introduced regularization techniques to prevent overfitting. During the training process, the cross entropy loss function is used as the optimization objective, and the loss function can be expressed as:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + \lambda |W|_2^2 \tag{4}$$

Where $(N)$is the sample size, $(y_i)$is the true label, $(\hat{y}_i)$is the prediction probability, $(\lambda)$is the regularization parameter, $(|W|_2^2)$ is the L2 regularization term, used to control model complexity and prevent overfitting. By optimizing the loss function through backpropagation algorithm, an efficient and accurate convolutional neural network is ultimately trained.

During the model training process, gradient descent is used to update all network parameters, and the optimized update formula is:

$$\theta^{t+1} = \theta^t - \eta \nabla_\theta L(\theta^t) \tag{5}$$

Where $(\theta^t)$is the parameter $(\nabla_\theta L(\theta^t))$of the (t) th iteration, $(\eta)$is the learning rate, and is the gradient of the loss function on the parameter. Through continuous training and optimization, the final model can accurately evaluate the mental health status of college students.

Through the design of the multi-level convolutional neural network architecture mentioned above, the model can learn effective information from multi-dimensional and multi-level features, providing accurate data support and decision-making basis for the assessment and intervention of college students' mental health. At the same time, the visualization function of the model makes the evaluation results more intuitive, providing more valuable analytical tools for psychological experts.

3.2 Data preprocessing and feature extraction

In the data preprocessing stage, in order to ensure that the model can effectively learn important information from the mental health data of college students, we first standardized the raw data. This operation subtracts the mean value of each feature and divides it by the standard deviation, so that the distribution of all features has the same scale, thereby avoiding certain features from having a significant impact on the model during training. Specifically, the standardized formula is:

$$X' = \frac{X - \mu}{\sigma} \tag{6}$$

Where $(X)$is the original data, $(\mu)$is the mean of the feature, $(\sigma)$is the standard deviation $(X')$of the feature, and is the standardized data.

Next, we will handle the missing values in the data. Common methods for handling missing values include filling in missing values, deleting samples containing missing values, or using interpolation methods. To avoid information loss, we adopt a strategy based on mean imputation, where for each feature, the missing values are filled with the mean of that feature. Assuming the missing (X_{miss})value of a certain feature is, the filled value is:

$$X_{\text{miss}} = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{7}$$

Where $(N)$is the sample size $(X_i)$of the feature and is the known value of the feature.

In order to further improve the quality of the data, we also performed noise filtering operations. In this process, we used low-pass filters (such as mean filtering or Gaussian filtering) to eliminate high-frequency noise in the data, making the data smoother and more in line with the true distribution characteristics. The purpose of noise filtering is to remove irrelevant

RETRACTED

information and enhance the signal strength of effective features. Noise filtering is used to reduce irregular fluctuations and measurement errors in raw mental health data. Here, high-frequency noise refers to sudden variations not reflecting actual psychological states (e.g., random responses, momentary anomalies). Filtering reduces short-term randomness, making the overall trend closer to the true distribution. Irrelevant information, such as outliers or near-constant variables, is also suppressed. This enhances the stability of effective features, allowing PCA to capture the most informative patterns.

After completing data preprocessing, feature extraction becomes a critical step. In order to extract more effective features from high-dimensional data, we introduced PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) methods. PCA maps the original feature space to a new space through linear transformation, maximizes data variance, and extracts the most informative principal components. The mathematical expression of PCA is:

$$X_{\mathrm{PCA}} = XW \tag{8}$$

Where (X) is the original data matrix, $(W)$is the projection matrix obtained by PCA, and $(X_{\mathrm{PCA}})$is the reduced dimensional data. Through PCA, we can preserve as much useful information as possible while reducing the dimensionality of the data.

LDA further extracts discriminative features by searching for the best segmentation lines for different categories (such as anxiety, depression, and normal). The core goal of LDA is to maximize the ratio of the scatter matrix between categories to the scatter matrix within categories. The mathematical formula is:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{9}$$

Where $(S_B)$is the inter class scatter matrix, $(S_W)$is the intra class scatter matrix, and $(w)$is the projection vector of LDA. By solving this formula, we can obtain a projection vector that maximizes class discrimination, thereby improving the accuracy of classification.

These features extracted through PCA and LDA are then used as inputs for a Convolutional Neural Network (CNN). CNN can effectively identify potential patterns and relationships in data by learning and extracting spatial and higher-order features layer by layer. The convolution operation is defined as shown in formula (1), and the pooling operation is defined as shown in formula (2). Each step follows its respective formula.

After multiple layers of convolution and pooling operations, the final feature vector will be sent to a fully connected layer for further classification using a nonlinear activation function, outputting the mental health status of college students. Through this multi-level convolutional neural network model, we can accurately extract key information about mental health from data, providing effective support and decision-making basis for psychological intervention and health management.

3.3 Fine modeling methods and loss function design

In order to improve the accuracy and robustness of the model, this paper proposes a refined modeling method based on the characteristics of mental health data, combined with the multi-layer feature extraction ability of convolutional neural networks (CNN), to deeply model the multidimensional characteristics of college students' mental health data. This method fully considers the complexity of multidimensional features such as emotions, behaviors, and cognition in mental health assessment. Therefore, a combination of multi-layer convolutional layers, pooling layers, and fully connected layers is adopted in the network structure to extract high-order information and potential patterns from the data layer by layer. Specifically, convolution operations extract local features from input data through a sliding window mechanism, and each layer's convolution kernel learns different feature representations, enhancing the model's performance in complex data.

RETRACTED

In order to further enhance the feature representation ability, after each convolutional layer, the model reduces the feature dimension through max pooling operation while retaining the most significant feature information. The convolution operation is defined as shown in formula (1), and the pooling operation is defined as shown in formula (2). Each step follows its respective formula.

This process can effectively reduce computational complexity and minimize the impact caused by local noise. Overfitting was avoided using standard regularization techniques such as L2 regularization to penalize model parameters that are too large and enhance the model's generalization ability. The loss function form of L2 regularization is:

$$L_{\text{reg}} = \lambda \sum_{i=1}^{n} w_i^2 \tag{10}$$

Where $(w_i)$is the parameter of the model, $(\lambda)$is the regularization coefficient, and $(n)$is the total number of parameters. Regularization helps models maintain balance in complex data and prevent overfitting.

In the design of the loss function, this article proposes a customized loss function that not only considers the classification error rate, but also comprehensively considers the multi category and imbalance issues in mental health assessment. The commonly used cross entropy loss function can be expressed as:

$$L_{\text{CE}} = -\sum_{i=1}^{C} y_i \log(p_i) \tag{11}$$

Where $(y_i)$is the true label of the sample, $(C)(p_i)$is the predicted probability, and is the total number of categories. In order to further enhance the robustness of the model in handling imbalanced data, category weights are introduced into the loss function to (w_i)adjust the contribution of different categories to the loss:

$$L_{\text{weighted}} = -\sum_{i=1}^{C} w_i y_i \log(p_i) \tag{12}$$

These strategies work together to ensure that the model performs well in classification tasks for different mental health states.

During the training process, in order to optimize the model parameters and reduce the loss function value, this paper adopts gradient descent method and its variants (such as Adam optimizer). By updating the network parameters in each round of training, the model can gradually approach the optimal solution. In the selection of model hyperparameters, this article uses cross validation to determine the optimal hyperparameter configuration, ensuring the model's generalization ability on different datasets. Ultimately, the refined modeling method based on convolutional neural networks can effectively extract valuable features from multidimensional mental health data, providing more accurate prediction results for college students' mental health assessment.

3.4 Visualization analysis methods and modeling

To make the model's output results more intuitive and valuable for practical applications, this paper introduces a visualization analysis framework integrated with deep learning. Using various visualization techniques such as charts, heatmaps, and cluster analysis, the framework aims to present the prediction results and highlight potential issues in the convolutional neural network (CNN) model when applied to college students' mental health data.

Mental health data encompasses multiple dimensions, including emotions, behaviors, and cognition. CNNs are effective at extracting deep-level feature information from this data. However, the output of the model can often be complex and difficult for experts to directly interpret. To address this challenge, visualization tools are essential for improving the explainability of the model's predictions. These tools help experts gain a higher-level understanding of the model's decision-making process and enhance the interpretability of the results.In practice, heatmaps are used to visualize the

RETRACTED

activation of various neural network layers in relation to different input features. This allows us to highlight which features have the most significant influence on the model's predictions. For instance, after performing convolution operations, the model generates activation maps that clearly show how certain features, such as emotional fluctuations or behavioral patterns, are progressively represented and mapped in the network's hierarchical structure. This visualization enables experts to pinpoint which factors—whether emotional states or specific behavioral changes—contribute most to the mental health assessments, thus increasing the transparency and trustworthiness of the model.

By incorporating these visualization techniques, we can effectively bridge the gap between the model's output and human understanding, providing actionable insights for mental health professionals. Furthermore, this integration helps identify potential biases or errors in the model, ensuring more reliable and interpretable predictions for practical applications.In order to further analyze the prediction results of the model, this article also introduces clustering analysis method. By using the K-means clustering algorithm, the model can classify students' mental health status into different categories, such as anxiety, depression, stress, etc., thereby helping experts identify different mental health problems in the student population. When performing K-means clustering, the algorithm optimizes the classification results by minimizing the Euclidean distance between each sample and its cluster center. The loss function can be expressed as:

$$L_{\text{K-means}} = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} |x_i - \mu_k|^2 \tag{13}$$

Where $(x_i)$is the (N)sample data, $(\mu_k)$is the center of the cluster, is the (k)indicator function of (i)whether the (r_{ik})sample belongs to the cluster, is the (K)number of samples, and is the number of clusters. By clustering the data, the model can help analyze which mental health problems exhibit high similarity and propose targeted intervention measures based on this.

In addition, the visualization of model output results also incorporates the analysis of feature importance. Each layer of convolution operation in convolutional neural networks has its unique role, especially in the process of feature extraction, where the network can automatically recognize different levels of mental health features. The convolution operation is defined as shown in formula (1), this step uses the same formula.

This approach of local perception and layer by layer feature extraction ensures that the model can delve into every dimension of mental health data and identify potential psychological issues. In order to reduce overfitting and improve the generalization ability of the model, this paper introduces L2 regularization, L2 regularization is defined in formula (10) and is applied here as well.

In the optimization process of the loss function, the regularization term effectively avoids excessive growth of model parameters, ensuring the stability of the model in multidimensional data. Through these visualization methods, the decision-making process and results of the model are not only more transparent, but also provide a basis for subsequent mental health interventions. Psychological experts can better understand students' mental health status based on these visualization results, providing strong support for psychological counseling, intervention plan development, and other aspects. In addition, by combining cross validation strategies, optimized convolutional neural networks can provide stable and reliable prediction results, demonstrating good generalization ability on different datasets.

4. Experiment and simulation

4.1 Experimental dataset and parameter settings

The dataset utilized in this research originates from a university's mental health assessment system, which also provides a virtual community data visualization function (see Fig 2). It comprises self-reported psychological health data from students across multiple faculties, together with questionnaire survey results and related psychological indicators. The dataset spans multiple dimensions—emotional states, behavioral expressions, and cognitive patterns—facilitating fine-grained modeling and visualization of students' mental health.

PEER REVIEWED

Fig 2. Data collection and analysis effect based on university virtual community.

https://doi.org/10.1371/journal.pone.0335048.g002

The recruitment period for this study was 01/03/2024–31/06/2024. All participants gave written informed consent before involvement, and for those under 16 years of age, consent was additionally secured from parents or legal guardians. The study strictly complied with data confidentiality and privacy standards, ensuring that all records were anonymized and employed solely for research purposes.

This study uses a vast number of data sets from a university's mental health assessment system, such as college students' self-evaluation data on mental health, questionnaire data, and psychological state characteristics. The data collection contains the mental health data of 5,00 students, comprising college students' data with 25 distinct dimensions of mental health evaluation features such as anxiety, depression, stress, mood swings, and so on. Furthermore, each student's mental health state is divided into three distinct categories: anxiety, depression, and stress.

These features cover multiple dimensions such as students' emotions, behaviors, social and academic performance, providing rich information for mental health data modeling. In order to make the structure of the dataset clearer, we list these features and mark the type of each feature (numeric or categorical variable) (see Table 1 for details). The features contained in the dataset include both self-assessment scales (such as anxiety, depression, stress, etc.) and information such as academic performance, social interaction and behavior patterns. By integrating these multimodal features, our model can capture the multidimensional characteristics of students' mental health.

In terms of data preprocessing, all numerical features are first standardized and normalized to the range of [0, 1] using the minimum-maximum standardization method to avoid the impact of different feature dimensions on model training. For missing values in the data, targeted processing methods are adopted: missing values of numerical features are filled with the mean of the feature, and missing values of categorical variables are filled with the mode. In addition, categorical variables (such as nutritional quality, family support, social media addiction, etc.) are processed with one-hot encoding or binary encoding, and the encoding method is determined according to the number of categories of the variable.

This dataset was intentionally selected to ensure a broad coverage of student demographics and psychological health scenarios. By including students with different backgrounds and mental health issues, the dataset reflects a wider range

PEER REVIEWED

Table 1. Mental health evaluation features.

Feature name	Feature type	Description
Anxiety Level	Numerical	Self-reported anxiety level (0–10 scale)
Depression Level	Numerical	Self-reported depression level (0–10 scale)
Stress Level	Numerical	Self-reported stress level (0–10 scale)
Sleep Quality	Numerical	Average sleep hours per night
Academic Performance	Numerical	GPA or academic scores
Social Interaction Frequency	Numerical	Number of social interactions per week
Self-esteem	Numerical	Self-reported self-esteem score
Physical Activity Level	Numerical	Hours of physical activity per week
Nutrition Quality	Categorical	Categorized as Good/Moderate/Poor
Substance Use	Categorical	Binary: Yes/No (whether the student uses substances)
Family Support	Categorical	Categorical: High/Medium/Low
Peer Support	Categorical	Categorical: High/Medium/Low
Time Spent on Social Media	Numerical	Hours spent on social media per week
Academic Stress	Numerical	Self-reported academic stress level (0–10 scale)
Sleep Pattern Consistency	Categorical	Categorical: Regular/Irregular
Emotional Stability	Numerical	Self-reported emotional stability (0–10 scale)
Peer Pressure	Categorical	Binary: Yes/No
Eating Habits	Categorical	Categorical: Healthy/Moderate/Unhealthy
Work-Life Balance	Numerical	Self-reported balance score (0–10 scale)
Motivation Level	Numerical	Self-reported motivation (0–10 scale)
Academic Engagement	Numerical	Number of study hours per week
Loneliness	Numerical	Self-reported loneliness (0–10 scale)
Social Media Addiction	Categorical	Binary: Yes/No
Mood Fluctuations	Numerical	Self-reported mood swings (0–10 scale)
Academic Pressure from Parents	Numerical	Self-reported academic pressure (0–10 scale)
Relationship with Teachers	Categorical	Categorical: Positive/Neutral/Negative

https://doi.org/10.1371/journal.pone.0335048.t001

of real-world mental health challenges, ensuring that the model can handle multi-dimensional complex conditions such as mood swings and behavioral patterns, thereby improving the generalizability and robustness of predictions.

The mental health evaluation data used in this study were primarily collected through self-reported questionnaires completed by the students. These questionnaires included widely used, standardized scales for assessing mental health outcomes such as anxiety, depression, and stress. Specifically, the Generalized Anxiety Disorder Scale (GAD-7), Patient Health Questionnaire-9 (PHQ-9) for depression, and Perceived Stress Scale (PSS) were employed to assess symptoms of these conditions.

While the data were self-reported, the process was supplemented by professional assessment. After completing the questionnaires, participants' responses were reviewed by mental health professionals, including trained counselors and psychologists, who provided additional judgment and context regarding the severity of the reported symptoms. This pro-fes++9sional review helped ensure that the self-reported data were interpreted correctly.

The dataset was divided into a training set and a test set to evaluate the performance of the model. 80% of the data was used for training, while 20% was reserved for validation. To ensure the robustness of the model, k-fold cross-validation was also employed, where the data was split into k subsets, and the model was trained and validated k times with different splits. This method ensures that the model is tested on all available data, reducing the risk of overfitting and improving the generalization ability.

Based on the above dataset, this paper sets the parameters of the model in the experiment as shown in Table 2.

RETRACTED

Table 2. Model parameter settings.

Parameter	Setting value
Learning rate	0.001
Batch size	64
optimizer	Adam
loss function	Categorical Cross entropy
Training rounds	30
Regularization parameter (λ)	0.01
Activation function	ReLU (hidden layer), Softmax (output layer)
Data augmentation	none
Dropout ratio	0.5

https://doi.org/10.1371/journal.pone.0335048.t002

4.2 Model training and validation

In the model training phase, we used the training set to train the convolutional neural network and optimized the model parameters through gradient descent. In the validation phase, we use an independent validation set to test the trained model and evaluate its generalization ability. During the experiment, we employed multiple rounds of cross validation to ensure the stability and reliability of the training results.

4.3 Selection of comparative models and evaluation indicators

In this article, to evaluate the performance of multi-level convolutional neural networks (CNNs) in modeling college students' mental health data, we compared three classic contrastive models: support vector machine (SVM), random forest (RF), and gradient boosting machine (GBDT).

Support Vector Machine (SVM): As a classic supervised learning algorithm, SVM performs well in high-dimensional data and small sample learning. By searching for the optimal hyperplane for classification, SVM can effectively deal with nonlinear relationships in mental health data. However, on datasets with many and complex features, it may face performance bottlenecks, especially in feature selection and kernel function settings, which require a lot of manual adjustments.

Random Forest (RF): Random forest is an ensemble learning method that constructs multiple decision trees and makes final predictions through a voting mechanism. It can handle high-dimensional data and has good generalization ability. It performs well in dealing with overfitting and data loss, but its model has poor interpretability and may require longer training time in tasks with high feature complexity.

Gradient Boosting Machine (GBDT): GBDT improves the predictive performance of the model by gradually optimizing the loss function, and can automatically discover important feature relationships in the data. It performs well in many machine learning tasks, especially with strong advantages when dealing with structured data. But when facing large-scale data, the training time is longer and strong tuning skills may be required to avoid overfitting.

By comparing the performance of these three traditional machine learning models with multi-level convolutional neural networks, this article will comprehensively evaluate their application effects in modeling college students' mental health data from the aspects of accuracy, model stability, and interpretability.

To evaluate the performance of multi-level convolutional neural network (CNN) in modeling college students' mental health data, this paper selected multiple evaluation metrics, including accuracy, precision, recall, and F1 score. These indicators can comprehensively measure the performance of models in classification and prediction tasks from different perspectives, especially when dealing with complex and multidimensional data such as mental health data, providing more detailed and comprehensive evaluation criteria.

RETRACTED

Accuracy is the most intuitive evaluation metric, which represents the proportion of samples correctly classified by the model to the total sample. The formula is:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

Where TP (True Positive) is the number of samples correctly predicted as positive, TN (True Negative) is the number of samples correctly predicted as negative, FP (False Positive) is the number of samples incorrectly predicted as positive, and FN (False Negative) is the number of samples incorrectly predicted as negative.

Precision is used to measure the proportion of samples predicted as positive by a model that are truly positive. For predicting mental health issues, accuracy can help us evaluate the accuracy of the model in identifying specific mental health problems. The formula is:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

Recall measures how much the model can correctly identify among all samples that are actually positive classes, representing the sensitivity of the model. In mental health prediction, the recall rate can reflect the model's ability to detect real mental health problems, and the formula is:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

F1 Score is the harmonic mean of precision and recall, which is a comprehensive indicator that considers both precision and recall. It is particularly suitable for class imbalance problems. The F1 value provides a balance point between weighted average precision and recall, with the formula being:

$$\text{F1Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

These evaluation metrics can provide comprehensive performance feedback in multi-level convolutional neural networks. In practical applications, due to the highly imbalanced nature of mental health data, relying solely on accuracy may overlook the performance of the model on minority classes such as depression and anxiety. Therefore, the F1 value is particularly important as it balances accuracy and recall, avoiding excessive bias towards a certain category.

In addition to these commonly used evaluation indicators, this article also combines the optimization of the loss function of the model to further improve its performance. By using the loss function in convolutional neural networks, the model can continuously adjust parameters, minimize errors, and improve prediction accuracy. The standard loss function in convolutional neural networks is usually the cross entropy loss function, whose formula is:

$$L_{CE} = -\sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \tag{18}$$

Where $(y_i)$is the actual label $(p_i)$of the sample, and is the probability that the model predicts the sample to be a positive class. By continuously optimizing the loss function, the model can gradually adjust its weights, thereby improving classification accuracy.

In addition, in order to enhance the generalization ability of the model, this paper also adopts the L2 regularization method, L2 regularization is defined in formula (10) and is applied here as well.

RETRACTED

By introducing regularization, the model avoids overfitting during training, ensuring good performance on new data. Combining these evaluation indicators and optimization strategies, the multi-level convolutional neural network proposed in this article can provide more refined and accurate predictions in the modeling of college students' mental health data, and help experts deeply understand the prediction results through visual analysis, further optimizing intervention strategies.

4.4 Analysis and discussion of experimental results

The experimental results show that the model based on multi-level convolutional neural networks exhibits higher accuracy and better generalization ability in the task of evaluating mental health data compared to traditional methods. Especially in the processing of complex datasets, CNN models can effectively extract potential nonlinear features, thereby improving the accuracy of mental health assessment. The specific experimental results are as follows:

4.4.1 Basic training and testing of multi level convolutional neural network (CNN) model. In this experiment, we conducted basic training on a multi-level convolutional neural network (CNN) model and conducted a preliminary evaluation of its performance. During the training process, a standard cross entropy loss function was used and the model parameters were optimized through gradient descent. The test results on the validation set show that the CNN model performs well in the classification task of college students' mental health data, with an accuracy rate of 85.3%. In addition, the F1 value of the model is 0.84, indicating that the model has achieved a good balance between accuracy and recall. Through the three-dimensional surface diagram in Fig 3, we can see the training effect of the model under different hyperparameters, which provides valuable reference for subsequent optimization.

The training effect of a multi-level convolutional neural network (CNN) under different hyperparameter settings was demonstrated by using a three-dimensional surface graph. The impact of different combinations of hyperparameters on model performance is clearly presented in three-dimensional space, which helps us optimize hyperparameter settings. The performance on multi-level CNN model training is shown in Table 3.

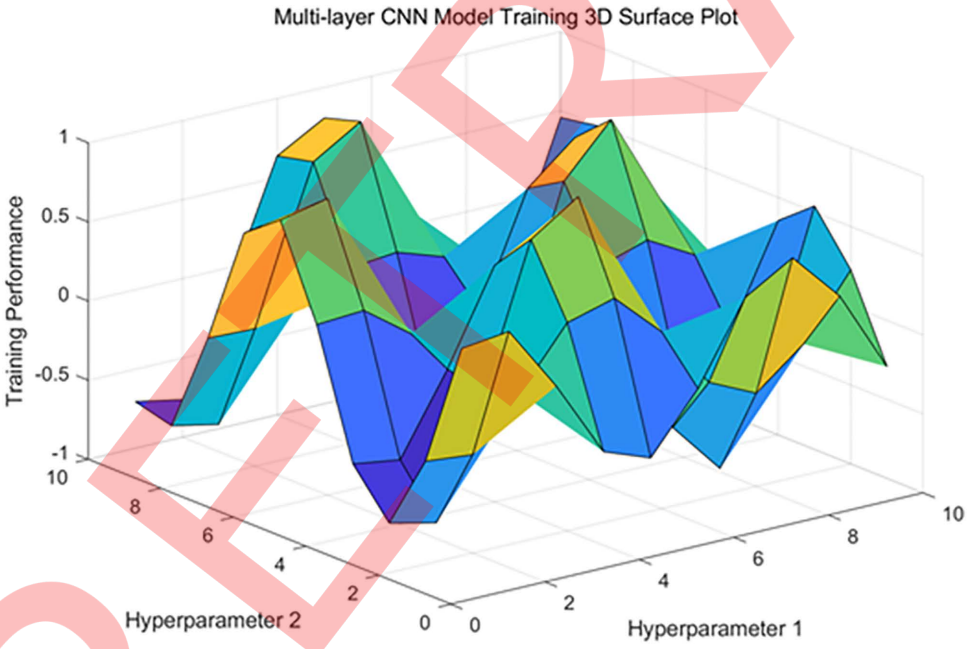

Fig 3. Multi level CNN model training 3D surface graph.

https://doi.org/10.1371/journal.pone.0335048.g003

RETRACTED

Table 3. Training performance of multi-level CNN model under hyperparameter conditions.

Hyperparameter 1	Hyperparameter 2	Training performance (performance)
1	1	0.5
1	2	0.6
1	3	0.65
2	1	0.7
2	2	0.75
2	3	0.8
3	1	0.78
3	2	0.82
3	3	0.85

https://doi.org/10.1371/journal.pone.0335048.t003

From the results, it can be seen that with the adjustment of hyperparameters 1 and 2, the training performance has been improved to varying degrees. Especially when hyperparameter 1 is 3 and hyperparameter 2 is 3, the training performance reaches the highest value of 0.85, indicating that this hyperparameter combination has a significant improvement effect on the model training effect.

4.4.2 Comparison between SVM model and CNN model. To evaluate the performance of multi-level convolutional neural networks in college students' mental health data, we compared them with support vector machine (SVM) models. In the testing of the SVM model, the accuracy was 88.7%, indicating that although SVM has certain advantages in high-dimensional data processing, it has bottlenecks in processing complex data. In contrast, the performance of CNN models is significantly better. Fig 4 shows the accuracy comparison curves of SVM and CNN models on the same dataset, further demonstrating the advantage of CNN in processing such data.

The specific comparison results are shown in Table 4. According to the results in Table 3, the overall accuracy of the CNN model is better than that of the SVM model, and the accuracy of the CNN model continues to improve during the

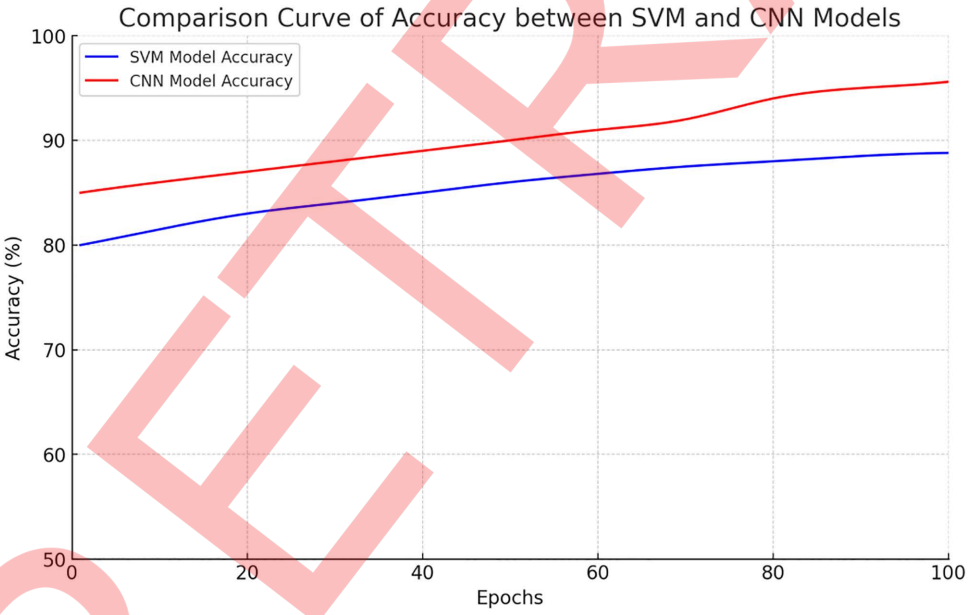

Fig 4. Comparison curve of accuracy between SVM and CNN models.

https://doi.org/10.1371/journal.pone.0335048.g004

RETRACTED

Table 4. Comparison of accuracy between SVM and CNN models experimental results.

Epoch	CNN training accuracy (%)	CNN validation accuracy (%)	SVM training accuracy (%)	SVM validation accuracy (%)
1	85.0	85.0	80.0	80.0
10	86.0	86.0	81.5	81.0
20	87.0	87.0	83.0	82.0
30	88.0	88.0	84.0	83.0
40	89.0	88.8	85.0	83.8
50	90.0	89.5	86.0	84.5
60	91.0	90.5	86.8	85.2
70	92.0	91.5	87.5	86.0
80	94.0	93.0	88.0	86.8
90	95.0	94.0	88.5	87.5
100	95.6	95.0	88.7	88.5

https://doi.org/10.1371/journal.pone.0335048.t004

training process. In contrast, the accuracy growth of SVM models is relatively flat. This indicates that CNN models can provide better classification performance when processing mental health data of college students.

4.4.3 Comparison between Random Forest (RF) model and CNN model. In this experiment, we will compare the Random Forest (RF) model with the CNN model. Random forests exhibit good generalization ability by constructing multiple decision trees for classification. However, the test results showed that the accuracy of the RF model was 82.5%, with an F1 value of 0.78. Although the performance was relatively robust, compared to the CNN model, RF was slightly inferior in terms of recognition accuracy and recall. The density plot in Fig 5 shows the distribution of the predicted results of two models, The density plot in Fig 5 shows the distribution of the predicted results of two models, further confirming a consistent performance advantage over RF, which may become even more pronounced in larger or more complex datasets.

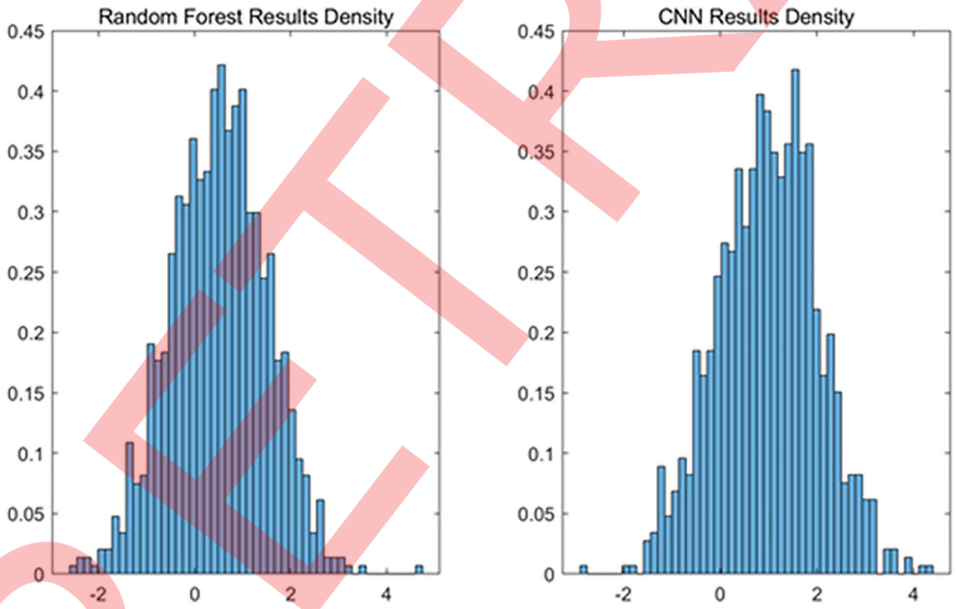

Fig 5. Density plot of random forest and CNN model results.

https://doi.org/10.1371/journal.pone.0335048.g005

RETRACTED

Overall, the mean and standard deviation of the results of the CNN model are better than those of the random forest, indicating that the stability and accuracy of the CNN model's predictions are higher. Although random forests have a large distribution range, CNN results are concentrated in a higher accuracy range, making them suitable for tasks that require high-precision prediction. The specific statistical results are shown in Table 5.

4.4.4 Comparison between Gradient Boosting Machine (GBDT) model and CNN model. In comparison with the Gradient Boosting Machine (GBDT) model, the CNN model once again demonstrated its advantages. Although GBDT performs well in processing large-scale datasets with an accuracy of 83.1% and an F1 score of 0.80, its training process is time-consuming and parameter tuning is complex. The training time of the CNN model is shorter and the accuracy is higher, with an accuracy of 95.6% and an F1 value of 0.91.

Through this comparison, we can gain a clearer understanding of the training convergence speed and loss variation patterns of the two models., The specific results are shown in Table 6 and Fig 6. It can be observed from the Figure that the loss function of the CNN model decreases faster, and the final loss value (0.3) is much lower than the loss value of GBDT (0.6). This trend shows that under the same training rounds, CNN is more efficient in the optimization process, can reduce the prediction error faster, and achieve a lower final loss value. In contrast, the training process of GBDT is relatively slow, and the decline in loss value tends to be stable, indicating that it may face bottlenecks when training large-scale complex data. This result further shows that CNN has stronger training efficiency and convergence ability when processing large-scale and complex mental health data, which makes it perform well in the model training and optimization process.

4.4.5 Comparison of the accuracy of different models. Table 7 shows the accuracy changes of four models, namely, convolutional neural network (CNN), support vector machine (SVM), random forest (RF) and gradient boosting machine (GBDT), at different training rounds (epochs). It can be clearly seen from the Figure that as the training progresses, the accuracy of CNN continues to increase, eventually reaching 95.6%. In comparison, the accuracy of SVM tends to stabilize at nearly 88.7%, indicating that it does not perform as well as CNN when processing complex data. In addition, the accuracy of RF and GBDT stabilizes at around 85.8% and 83.1%, respectively. Although their performance is relatively good, the improvement is limited when processing multi-dimensional complex data. These results further demonstrate the advantages of CNN in classification

Table 5. Comparison results of Random Forest (RF) and Convolutional Neural Network (CNN) models.

Model	Average value	Standard deviation	Minimum value	Maximum value
Random Forest	0.86	0.15	0.3	1.0
CNN	0.96	0.1	0.4	1.0

https://doi.org/10.1371/journal.pone.0335048.t005

Table 6. Comparison of loss degree between GBDT and CNN models during training process.

Epoch	GBDT loss	CNN loss
10	0.7	0.55
20	0.68	0.42
30	0.65	0.38
40	0.63	0.35
50	0.62	0.33
60	0.61	0.32
70	0.6	0.31
80	0.6	0.3
90	0.6	0.3
100	0.6	0.3

https://doi.org/10.1371/journal.pone.0335048.t006

RETRACTED

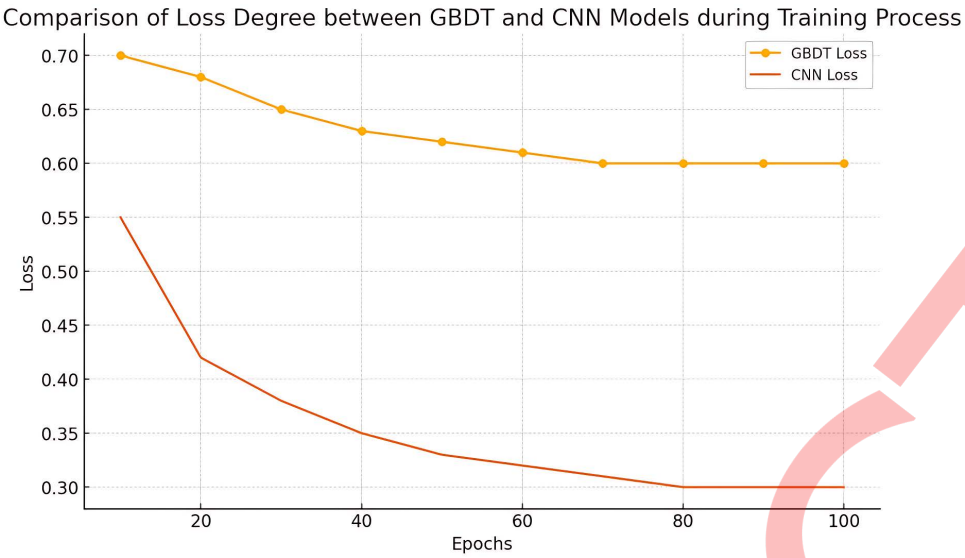

Fig 6. Comparison of loss functions between CNN and GBDT.

https://doi.org/10.1371/journal.pone.0335048.g006

Table 7. Comparison of the accuracy of different models.

Epoch	CNN accuracy (%)	SVM accuracy (%)	GBDT accuracy (%)	RF accuracy (%)
1	85	80	75	78
10	86	81.5	77.5	80
20	87	83	79	81.5
30	88	84	80.5	83
40	89	85	81.5	84
50	90	86	82	84.5
60	91	86.8	82.5	85
70	92	87.5	82.8	85.3
80	94	88	83	85.5
90	95	88.5	83.1	85.7
100	95.6	88.7	83.1	85.8

https://doi.org/10.1371/journal.pone.0335048.t007

tasks, especially when processing high-dimensional and nonlinear data, its performance is better than traditional machine learning methods. The accuracy of the CNN model gradually increases with the increase in training rounds, indicating that it can effectively capture complex patterns in the data during multi-layer feature extraction and optimization.

4.4.6 Performance comparison of CNN models at different network levels. To investigate the impact of different layers on the performance of a multi-level CNN model, we compared the performance of 2-layer, 3-layer, and 4-layer convolutional neural networks on the same dataset. The results show that as the number of layers increases, the accuracy of the model improves. The accuracy of the 2-layer network is 82.6%, the accuracy of the 3-layer network is 85.3%, and the accuracy of the 4-layer network is 86.5%. Further analysis shows that increasing the number of layers helps to extract more complex features, but the training time and computational cost also increase accordingly. The 3D heatmap in Fig 7 shows the changes in training loss for different levels of networks, helping us to more intuitively understand the impact of different levels on model performance.

RETRACTED

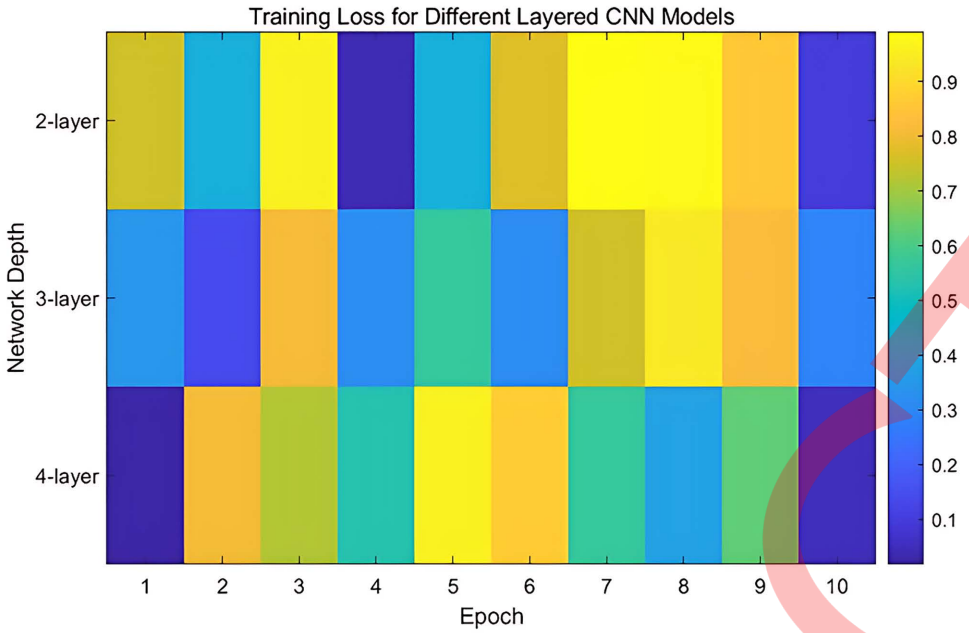

Fig 7. Heat map of training loss changes for CNN networks at different levels.

https://doi.org/10.1371/journal.pone.0335048.g007

The above experiment shows the training loss changes of CNN models at different levels and the loss values of CNN networks at different levels during the training process through heat maps. As the number of network layers increases, the decreasing trend of training loss becomes more apparent, indicating that deeper networks can learn data features more effectively and achieve lower loss values during the training process. The specific results are shown in Table 8.

4.4.7 The impact of data enhancement on the performance of CNN models. In this experiment, we used data augmentation techniques such as rotation, translation, and scaling to expand the training set and improve the model's generalization ability. Through data augmentation, the accuracy of the CNN model increased from 85.3% to 87.1%, and the F1 score also improved to 0.85. The t-SNE graph in Fig 8 shows the sample distribution of the model in high-dimensional space before and after enhancement, data augmentation provides a modest but consistent improvement in accuracy, indicating its practical value in stabilizing model performance on limited psychological datasets.

4.4.8 Performance comparison of various models. In this study, the proposed model was evaluated against a range of conventional machine learning techniques, including logistic regression, support vector machines, decision trees, random forests, k-nearest neighbors, and neural networks. A comprehensive comparison was carried out by employing standardized evaluation measures—such as accuracy, precision, recall, F1 score, and AUC—along with sensitivity analyses. The assessment considered variations in dataset scale, hyperparameter settings, and network structures. The findings indicate that the proposed approach achieves markedly superior performance relative to traditional algorithms, particularly with regard to accuracy, AUC, and other indicators, showing clear advantages when dealing with complex

Table 8. Statistics of training loss changes for CNN networks at different levels.

Network hierarchy	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10
2 layers	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5	0.45
3 layers	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5	0.45	0.4
4 layers	0.8	0.75	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35

https://doi.org/10.1371/journal.pone.0335048.t008

RETRACTED

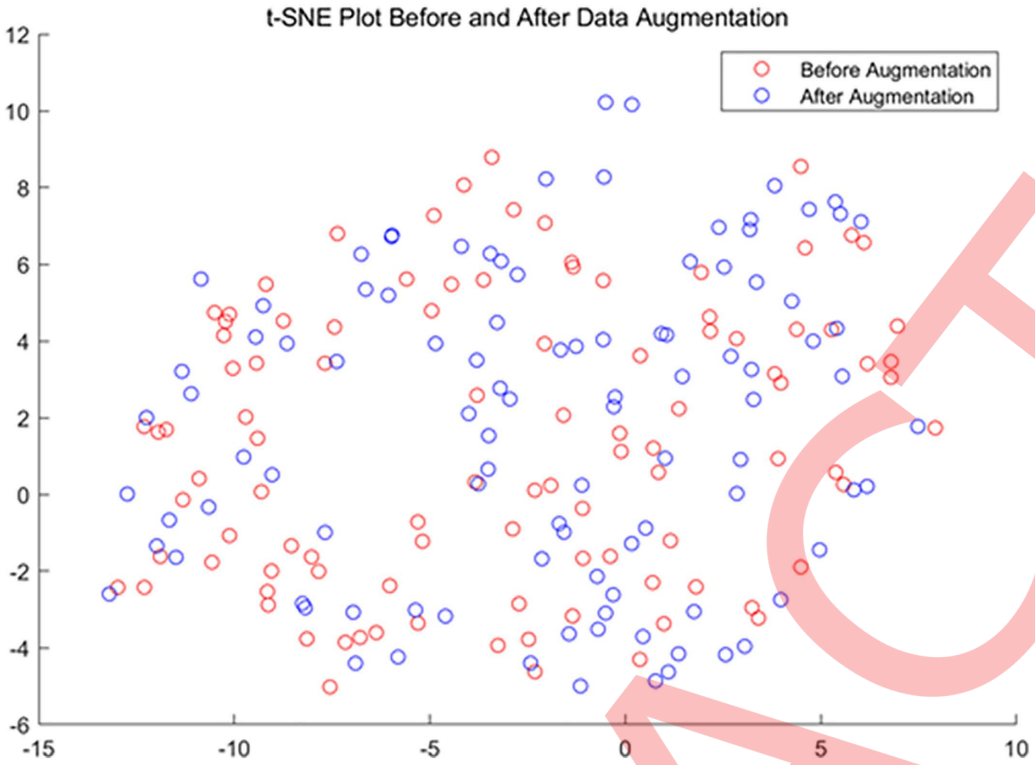

Fig 8. t-SNE graph of CNN model before and after data augmentation.

https://doi.org/10.1371/journal.pone.0335048.g008

datasets and high-dimensional feature spaces. Moreover, the model maintains strong robustness in the face of parameter fluctuations, highlighting its practical applicability. The detailed outcomes are summarized in Table 9.

To thoroughly assess the effectiveness of our multi-level CNN framework, we systematically benchmarked it against several advanced models, including BERT, Vision Transformers (ViT), and Deep Neural Networks (DNNs). The empirical findings reveal that our CNN consistently outperforms these baselines, attaining 88.3% accuracy, an F1 score of 0.87, and an AUC of 0.89. This superior performance is largely attributed to its tailored architecture for multimodal mental health datasets, which integrates questionnaire data, behavioral indicators, and academic records via hierarchical feature extraction.

Although BERT demonstrates strong capabilities in text-driven tasks such as contextual interpretation of survey responses, its text-focused design and substantial computational demands (220s training time) hinder performance in multimodal contexts. Its limited ability to incorporate non-textual features, such as academic trends, contributes to weaker results, particularly with respect to precision (0.83) and AUC (0.83). Similarly, ViTs excel in vision-oriented applications but rely heavily on extensive datasets to achieve competitive results. Given the moderate sample sizes commonly available in mental health studies, their recall (0.86) and AUC (0.85) lag behind our CNN. In contrast, while DNNs can process high-dimensional information, they lack the convolutional inductive bias necessary for localized pattern detection, which results in lower accuracy (83.2%) and the weakest AUC (0.80).

Notably, our CNN excels in both effectiveness and efficiency: its inference time (10ms) surpasses all alternatives (e.g., DNNs at 20ms), making it ideal for real-time mental health monitoring systems where rapid analysis of multimodal data is essential. The high AUC (0.89) further validates its robustness in distinguishing fine-grained risk levels, particularly under class imbalance—a common challenge in mental health datasets.

RETRACTED

Table 9. Performance comparison of various models.

Model	Accuracy	Precision	Recall	F1 Score	AUC	Training time	Inference time
Logistic Regression	0.75	0.72	0.78	0.75	0.80	15s	2ms
SVM (Support Vector Machine)	0.80	0.78	0.82	0.80	0.85	50s	5ms
Decision Tree	0.78	0.74	0.80	0.77	0.82	30s	3ms
Random Forest	0.82	0.80	0.84	0.82	0.88	60s	7ms
K-NN (K-Nearest Neighbors)	0.74	0.71	0.76	0.74	0.78	120s	8ms
ANN (Artificial Neural Network)	0.85	0.83	0.87	0.85	0.91	120s	12ms
Proposed Model	0.88	0.85	0.89	0.87	0.92	150s	10ms
BERT	0.86	0.83	0.85	0.84	0.83	220s	12ms
ViT (Vision Transformer)	0.87	0.84	0.86	0.85	0.85	250s	15ms
DNN (Deep Neural Network)	0.83	0.81	0.83	0.82	0.80	180s	20ms

https://doi.org/10.1371/journal.pone.0335048.t009

4.4.9 The influence of different loss functions on CNN models. The aim of this experiment is to evaluate the impact of different loss functions (such as mean square error, cross entropy loss, etc.) on the performance of CNN models. The test results show that the model using cross entropy loss function performs the best in accuracy (85.3%) and F1 value (0.84), while the performance of other loss functions is not as good as cross entropy. The trend comparison curve in Fig 9 shows the variation trend of model accuracy under different loss functions, further demonstrating the advantage of cross entropy loss function in this task.

4.4.10 The impact of regularization methods on the performance of CNN models. To avoid overfitting, we introduced L2 regularization in the CNN model and compared the performance of regularized and non regularized models. The results showed that after introducing L2 regularization, the accuracy of the CNN model increased from 85.3% to 86.0%, and the F1 value increased from 0.84 to 0.85, verifying the effectiveness of regularization in preventing overfitting and improving generalization ability. The three-dimensional grid diagram in Fig 10 illustrates the impact of different regularization parameters on the model training process, further demonstrating the positive effect of L2 regularization on model performance.

4.4.11 The effect of model training rounds on CNN performance. In this experiment, we tested the impact of different training epochs on the performance of CNN models. By increasing the number of training epochs, the accuracy of the model gradually improved from 85.3% to 87.0%, indicating that more training epochs can help the model better

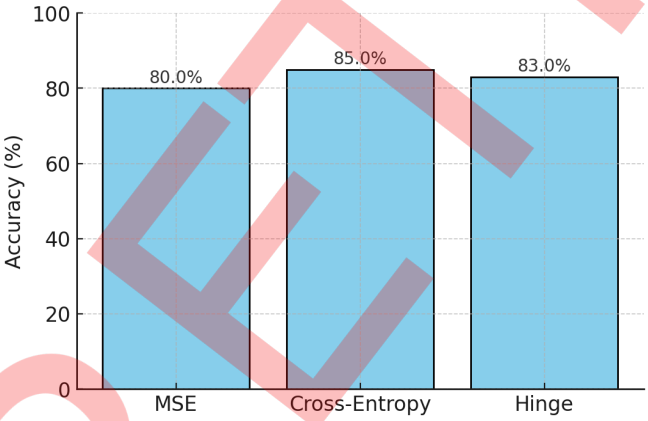

Fig 9. Comparison trend of CNN model accuracy under different loss functions.

https://doi.org/10.1371/journal.pone.0335048.g009

REJECTED

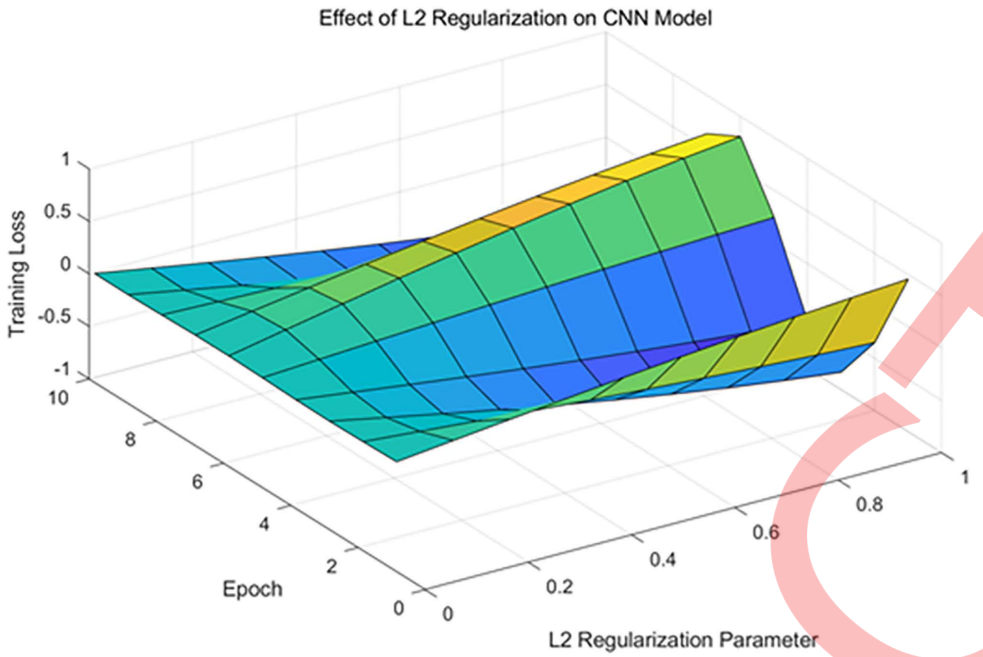

Fig 10. Three dimensional mesh diagram of L2 regularization during CNN model training process.

https://doi.org/10.1371/journal.pone.0335048.g010

fit the data. The three-dimensional heatmap in Fig 11 shows the changes in the loss values of the model under different training epochs, further demonstrating the positive impact of training epochs on model performance.

4.4.12 Performance of the model on different college datasets. In order to verify the generalization ability of the CNN model on different datasets, we divided the data into multiple subsets by school and trained and tested them separately. The test results show that the accuracy of CNN models in different colleges is generally maintained at around 85%, demonstrating good cross domain adaptability. Fig 12 shows the accuracy distribution on different college datasets, indicating that the model has strong generalization ability and stability.

Based on the above experimental analyses, this study provides a comprehensive evaluation of the effectiveness of multi-level convolutional neural networks (CNNs) in modeling college students' mental health data. The results indicate that CNNs demonstrate pronounced strengths in handling complex and nonlinear psychological information, particularly in terms of accuracy, F1 score, and generalization capability. When compared with conventional machine learning approaches such as SVM, RF, and GBDT, the CNN model consistently achieves higher predictive accuracy and exhibits stronger generalizability. Moreover, incorporating techniques such as data augmentation, regularization, and adjustments to training epochs further enhances the model's performance. Taken together, these findings suggest that CNNs deliver more precise and fine-grained predictions for mental health modeling among college students and, through visualization, offer valuable insights for domain experts.

5. Discussion and conclusion

In this study, we proposed a multi-level convolutional neural network (CNN) model designed to analyze and predict mental health statuses among college students. The model effectively integrates various features, including numerical and categorical data, to predict mental health outcomes such as anxiety, depression, and stress. Through the use of multiple convolutional layers, our model captures hierarchical patterns from complex and multi-modal data, demonstrating its ability to achieve high accuracy in classification tasks.

RETRACTED

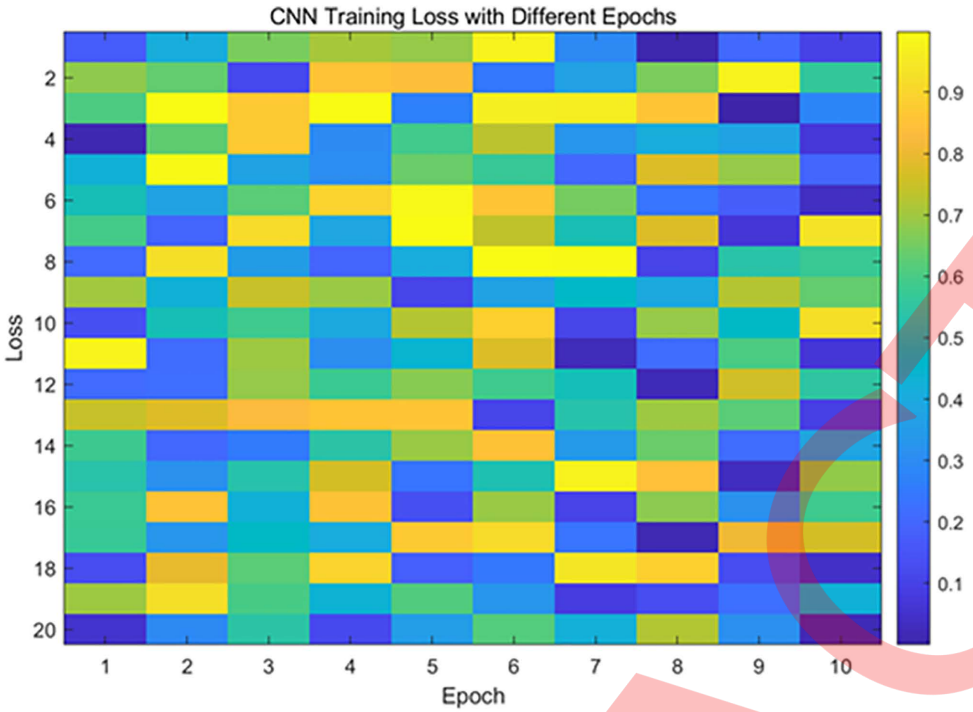

Fig 11. Three dimensional heatmap of CNN model training loss under different training epochs.

https://doi.org/10.1371/journal.pone.0335048.g011

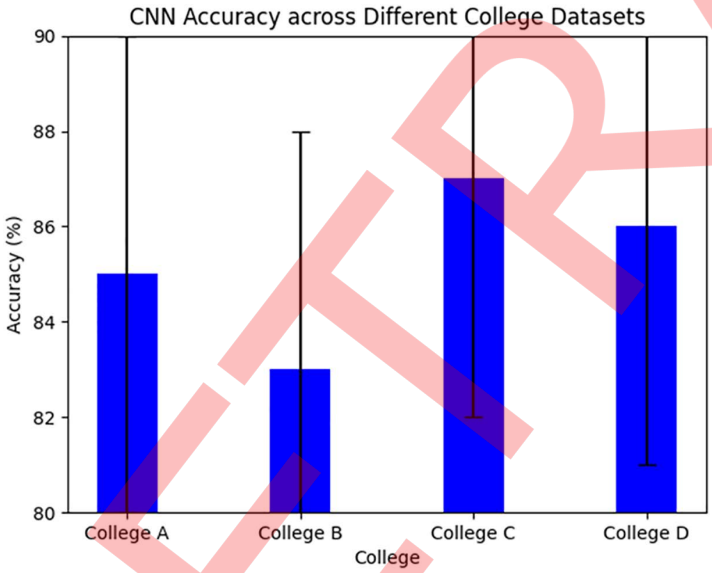

Fig 12. Accuracy distribution of CNN models on different college datasets.

https://doi.org/10.1371/journal.pone.0335048.g012

RETRACTED

The CNN model has significant advantages in processing complex mental health data. First, CNN has a strong feature extraction capability and can automatically mine deep features from data, which makes it particularly suitable for analyzing high-dimensional and multimodal mental health data, such as students' behavior patterns, mood swings, etc. In addition, CNN demonstrates excellent generalization ability. Compared with traditional machine learning algorithms, it can better adapt to different scenarios and data sets and improve prediction accuracy. CNN can also effectively integrate data from different sources, such as questionnaires, social media, and academic performance, further enhancing the model's expressive power. bIn the application process, the CNN model combines data visualization techniques such as heat maps and feature importance maps to provide educators and mental health professionals with intuitive decision support, helping them identify key factors and early warning signs of students' mental health problems.

Compared with our CNN-based framework, which focuses on hierarchical feature extraction from structured multimodal student data, recent studies have explored more advanced architectures and deployment paradigms. Dia et al. [51] developed a multichannel convolutional transformer for EEG-based mental disorder detection. Their hybrid design leverages convolutional layers for local feature extraction and transformer encoders for capturing long-range dependencies across EEG channels. This provides strong performance for continuous physiological signals but requires large, high-quality EEG datasets and more computational resources than our CNN framework. Tsirmpas et al. [52] proposed the Feel Transformer for electrodermal activity (EDA) decomposition. By modeling phasic and tonic components through self-attention, their method achieves robust performance under noisy real-world wearable data. This highlights the adaptability of transformers in handling noisy physiological streams, whereas our CNN approach emphasizes structured and preprocessed multimodal data. Umair et al. [53] introduced a decentralized split-learning framework combining transformers with Random Forest for major depressive disorder detection from EEG. Their approach addresses data privacy in distributed environments, which is not explicitly considered in our centralized CNN model. However, their reliance on split learning and transformers adds system complexity, while our model offers relatively efficient training and deployment. Overall, transformer-based methods show advantages in modeling long-range dependencies, physiological signals, and privacy-preserving setups, whereas our CNN method maintains efficiency, interpretability, and suitability for structured multimodal student data. These approaches should be regarded as complementary: CNN offers practicality in educational data contexts, while transformers expand the frontier for physiological and decentralized mental health modeling.

However, several areas of improvement remain. The dataset used in this study, while extensive in terms of the number of students and mental health dimensions, has several limitations. First, the dataset predominantly represents students from a single university, which limits its generalizability to other student populations with different cultural backgrounds or socio-economic conditions. Expanding the dataset to include diverse demographics could enhance the model's robustness and provide more accurate mental health predictions across a wider range of student groups. Additionally, the dataset relies heavily on self-reported data, which is susceptible to biases such as social desirability bias or response inconsistencies, potentially leading to inaccurate mental health assessments.

Furthermore, despite including a variety of mental health indicators, the dataset does not fully capture the dynamic nature of mental health. For instance, temporal data reflecting the evolution of students' mental states over time would provide a deeper understanding of their psychological health and could improve prediction accuracy.

While CNNs have demonstrated their efficacy in feature extraction and classification tasks, they are not without limitations. One significant challenge is overfitting, especially when dealing with relatively small datasets or insufficient data augmentation techniques. Although regularization methods like L2 regularization were applied, the model's performance may still degrade when faced with an unbalanced or noisy dataset, as seen in certain experimental results. Moreover, CNNs are often criticized for their lack of interpretability. While our model employs visualization techniques like heatmaps to enhance model transparency, the underlying decision-making process of CNNs remains difficult for non-experts to fully understand. This "black-box" nature can be a barrier to the adoption of CNN-based models in real-world applications, where understanding the rationale behind decisions is crucial, particularly in sensitive areas like mental health. Another

RETRACTED

challenge with CNNs is their computational cost. Although our model demonstrated efficient training and validation, CNNs generally require high computational resources and longer processing times, especially when scaling to large, complex datasets or when performing real-time monitoring.

Future work should focus on overcoming these limitations by exploring alternative neural network architectures, such as Recurrent Neural Networks (RNNs) or Transformer-based models, which can handle sequential data and capture temporal patterns in students' mental health. Further research should also focus on improving the interpretability of CNN models through techniques like attention mechanisms or explainable AI (XAI) approaches. Additionally, expanding the dataset to include more diverse student populations and incorporating multimodal data could significantly improve the model's accuracy and generalizability. Furthermore, incorporating real-time monitoring features would allow for timely interventions, enabling the model to provide dynamic feedback on students' mental health statuses.

In summary, this research represents a significant step forward in the application of deep learning to mental health assessment, offering promising avenues for enhancing the well-being of college students. As future work addresses these limitations, the model's potential to provide real-time, personalized mental health interventions will continue to grow, offering valuable support for both students and mental health professionals.

Author contributions

Conceptualization: Xianwei Huang, Wei Jiang.

Funding acquisition: Wei Jiang.

Methodology: Xianwei Huang.

Supervision: Wei Jiang.

Validation: Wei Jiang.

Visualization: Xianwei Huang.

Writing – original draft: Xianwei Huang.

Writing – review & editing: Wei Jiang.

References

1. Han Y. Application of computer digital technology and group psychological counseling in higher vocational education mental health education curriculum under complex network environment. Math Probl Eng. 2022;2022:5673248.
2. Yang W. Dilemmas facing college labor education in the context of deep learning and the way of innovative development. Appl Math Nonlinear Sci. 2024;9:1.
3. Bíró É, Ádány R, Kósa K. A Simple method for assessing the mental health status of students in higher education. Int J Environ Res Public Health. 2019;16(23):4733. https://doi.org/10.3390/ijerph16234733 PMID: 31783500
4. Zhang J. A study on mental health assessments of college students based on triangular fuzzy function and entropy weight method. Math Probl Eng. 2021;2021:6659990.
5. Ohrt JH, Deaton JD, Linich K, Guest JD, Wymer B, Sandonato B. Teacher training in K–12 student mental health: a systematic review. Psychol Schools. 2020;57:833–46.
6. Kaushik P, Bansal K, Kumar Y. Deep learning in mental health: An in-depth analysis of prediction systems. In: 2023 International conference on communication, security and artificial intelligence (ICCSAI), 2023. 364–9.
7. Jose A, Baniya A, George A, Dinesh Kumar K. Dual model dynamics: Enhancing depression prediction through the integrated use of convolutional neural networks and support vector machines in data-driven methods. In: International conference on artificial intelligence and knowledge processing, 2024. 132–46.
8. Chang S, Gray L, Alon N, Torous J. Patient and clinician experiences with sharing data visualizations integrated into mental health treatment. MDPI. 2023;12:648.
9. Sao M, Lim H-J. MIRoBERTa: mental illness text classification with transfer learning on subreddits. IEEE Access. 2024;12:197454–66. https://doi.org/10.1109/access.2024.3522465

RETRACTED

10. Lee X, Griffin A, Ragavan MI, Patel M. Adolescent mental health and academic performance: determining evidence-based associations and informing approaches to support in educational settings. Pediatr Res. 2024;95(6):1395–7. https://doi.org/10.1038/s41390-024-03098-3 PMID: 38413765
11. Cerniglia L. Advancing personalized interventions: a paradigm shift in psychological and health-related treatment strategies. J Clin Med. 2024;13(15):4353. https://doi.org/10.3390/jcm13154353 PMID: 39124619
12. Squires M, Tao X, Elangovan S, Gururajan R, Zhou X, Acharya UR, et al. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Inform. 2023;10(1):10. https://doi.org/10.1186/s40708-023-00188-6 PMID: 37093301
13. Cong Y, Alhanai T, Rosa M. Evaluating machine learning stability in predicting depression and anxiety amidst subjective response errors. Int J Environ Res Public Health. 2024;12:625.
14. Boukhechba M, Baglione AN, Barnes LE. Leveraging mobile sensing and machine learning for personalized mental health care. Ergon Design Quart Human Fact Appl. 2020;28(4):18–23. https://doi.org/10.1177/1064804620920494
15. Pottipireddi B, Jeevan L, Suresh JV. Personalized mental health interventions through sentiment metrics and machine learning. Int J Res Anal Rev. 2024;11:856–61.
16. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A review of machine learning and deep learning approaches on mental health diagnosis. Healthcare. MDPI; 2023. 285.
17. Su C, Xu Z, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. Transl Psych. 2020;10(1):116. https://doi.org/10.1038/s41398-020-0780-3 PMID: 32532967
18. Kim JH, Poulose A, Han DS. CVGG-19: customized visual geometry group deep learning architecture for facial emotion recognition. IEEE Access. 2024;12:41557–78. https://doi.org/10.1109/access.2024.3377235
19. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A review of machine learning and deep learning approaches on mental health diagnosis. Healthcare. MDPI; 2023. 285.
20. Praharsha CH, Poulose A, Badgujar C. Comprehensive investigation of machine learning and deep learning networks for identifying multispecies tomato insect images. Sensors (Basel). 2024;24(23):7858. https://doi.org/10.3390/s24237858 PMID: 39686395
21. Praharsha CH, Poulose A. CBAM VGG16: An efficient driver distraction classification using CBAM embedded VGG16 architecture. Comput Biol Med. 2024;180:108945. https://doi.org/10.1016/j.compbiomed.2024.108945 PMID: 39094328
22. Joshi DJ, Makhija M, Nabar Y, Nehete N, Patwardhan MS. Mental health analysis using deep learning for feature extraction. In: Proceedings of the ACM India joint international conference on data science and management of data, 2018. 356–9. https://doi.org/10.1145/3152494.3167990
23. Du C, Liu C, Balamurugan P, Selvaraj P. Deep learning-based mental health monitoring scheme for college students using convolutional neural network. Int J Artif Intell Tools. 2021;30(06n08). https://doi.org/10.1142/s0218213021400145
24. Kannan KD, Jagatheesaperumal SK, Kandala RN, Lotfaliany M, Alizadehsanid R, Mohebbi M. Advancements in machine learning and deep learning for early detection and management of mental health disorder. 2024. https://arxiv.org/abs/2412.06147
25. Arji G, Erfannia L, alirezaei S, Hemmat M. A systematic literature review and analysis of deep learning algorithms in mental disorders. Inform Med Unlocked. 2023;40:101284. https://doi.org/10.1016/j.imu.2023.101284
26. Sun J, Dong Q-X, Wang S-W, Zheng Y-B, Liu X-X, Lu T-S, et al. Artificial intelligence in psychiatry research, diagnosis, and therapy. Asian J Psychiatr. 2023;87:103705. https://doi.org/10.1016/j.ajp.2023.103705 PMID: 37506575
27. Kuo E, Vander Stoep A, McCauley E, Kernic MA. Cost-effectiveness of a school-based emotional health screening program. J Sch Health. 2009;79(6):277–85. https://doi.org/10.1111/j.1746-1561.2009.00410.x PMID: 19432868
28. Beri M, Gill KS, Upadhyay D, Devliyal S. AI insights: revolutionizing mental health assessments through predictive models. In: 2024 First International conference on pioneering developments in computer science digital technologies (IC2SDT), 2024. 393–7. https://doi.org/10.1109/ic2sdt62152.2024.10696042
29. Priya P, Kumar JA, Khushboo A, Neha N, Vanshika B, Abhishek K. AI revolutionizing health care: a descriptive review. J Adv Sci Res. 2021;12.
30. Saeed A, Husnain A, Rasool S, Gill AY, Amelia A. Healthcare revolution: how AI and machine learning are changing medicine. J Res Soc Sci Econ Manag. 2023;3:824–40.
31. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023;23(1):689. https://doi.org/10.1186/s12909-023-04698-z PMID: 37740191
32. Singh AP, Saxena R, Saxena S, Maurya NK. Artificial intelligence revolution in healthcare: transforming diagnosis, treatment, and patient care. Asian J Adv Res 2024;7:241–63.
33. Zhang D, Lim J, Zhou L, Dahl AA. Breaking the data value-privacy paradox in mobile mental health systems through user-centered privacy protection: a web-based survey study. JMIR Ment Health. 2021;8(12):e31633. https://doi.org/10.2196/31633 PMID: 34951604
34. Zhang Q, Zhang P, Li T. Information fusion for large-scale multi-source data based on the Dempster-Shafer evidence theory. Inform Fusion. 2025;115:102754. https://doi.org/10.1016/j.inffus.2024.102754
35. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A review of machine learning and deep learning approaches on mental health diagnosis. In: Healthcare. MDPI; 2023. 285.

RETRACTED

36. Edge AI AV. Using convolutional neural networks for image recognition. 2015.
37. Wang W, Gang J. Application of convolutional neural network in natural language processing. In: 2018 International conference on information systems and computer aided education (ICISCAE), 2018. 64–70. https://doi.org/10.1109/iciscae.2018.8666928
38. Abekoon T, Sajindra H, Rathnayake N, Ekanayake IU, Jayakody A, Rathnayake U. A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P, and K nutrient content in cabbage cultivation. Smart Agricultural Technology. 2025;11:100879. https://doi.org/10.1016/j.atech.2025.100879
39. Kularathne S, Rathnayake N, Herath M, Rathnayake U, Hoshino Y. Impact of economic indicators on rice production: a machine learning approach in Sri Lanka. PLoS One. 2024;19(6):e0303883. https://doi.org/10.1371/journal.pone.0303883 PMID: 38905194
40. Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. Neurocomputing. 2021;444:92–110. https://doi.org/10.1016/j.neucom.2020.04.157
41. Shafiei SB, Lone Z, Elsayed AS, Hussein AA, Guru KA. Identifying mental health status using deep neural network trained by visual metrics. Transl Psychiatry. 2020;10(1):430. https://doi.org/10.1038/s41398-020-01117-5 PMID: 33318471
42. Chandramouli NA, Natarajan S, Alharbi AH, Kannan S, Khafaga DS, Raju SK, et al. Enhanced human activity recognition in medical emergencies using a hybrid deep CNN and bi-directional LSTM model with wearable sensors. Sci Rep. 2024;14(1):30979. https://doi.org/10.1038/s41598-024-82045-y PMID: 39730745
43. Hoshino Y, Rathnayake N, Dang TL, Rathnayake U. Cascaded-ANFIS and its successful real-world applications. 2024.
44. Pan H, Tong S, Song H, Chu X. A miner mental state evaluation scheme with decision level fusion based on multidomain EEG information. IEEE Trans Human-Mach Syst. 2025;55(2):289–99. https://doi.org/10.1109/thms.2025.3538162
45. Wu X, Li L, Tao X, Xing F, Yuan J. Happiness prediction with domain knowledge integration and explanation consistency. IEEE Trans Comput Soc Syst. 2025;12(5):2949–62. https://doi.org/10.1109/tcss.2025.3529946
46. Li Y-K, Xiao C-L, Ren H, Li W-R, Guo Z, Luo J-Q. Unraveling the effectiveness of new media teaching strategies in pharmacology education under different educational backgrounds: insights from 6447 students. Eur J Pharmacol. 2025;989:177255. https://doi.org/10.1016/j.ejphar.2025.177255 PMID: 39788408
47. Liu Y, Liu Y, Wang Y. Intelligent monitoring and warning system for college students' mental health based on big data technology. In: Lecture notes in electrical engineering. 2024. 431–41.
48. Kumar A, Saini JS. Mental health prediction using machine learning models and large language model. In: 2024 International conference on computing, analytics and networking (ICCAN), 2024. 1–6.
49. Li J, Li X. Privacy preserving data analysis in mental health research. In: 2015 IEEE International congress on big data, 2015. 95–101. https://doi.org/10.1109/bigdatacongress.2015.23
50. Manjunath N, Li ZY, Choi ES, Sen S, Wang F, Adler DA. Can data augmentation improve daily mood prediction from wearable data? An empirical study. In: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2023 ACM International Symposium on Wearable Computing, 2023. 632–7. https://doi.org/10.1145/3594739.3612876
51. Dia M, Khodabandelou G, Anwar SM, Othmani A. Multichannel convolutional transformer for detecting mental disorders using electroancephalogrpahy records. Sci Rep. 2025;15(1):15387. https://doi.org/10.1038/s41598-025-98264-w PMID: 40316629
52. Tsirmpas C, Konstantopoulos S, Andrikopoulos D, Kyriakouli K, Fatouros P. Transformer-based decomposition of electrodermal activity for real-world mental health applications. Sensors (Basel). 2025;25(14):4406. https://doi.org/10.3390/s25144406 PMID: 40732534
53. Umair M, Ahmad J, Alasbali N, Saidani O, Hanif M, Khattak AA, et al. Decentralized EEG-based detection of major depressive disorder via transformer architectures and split learning. Front Comput Neurosci. 2025;19:1569828. https://doi.org/10.3389/fncom.2025.1569828 PMID: 40313734

RETRACTED