Journal of Infrastructure, Policy and Development 2025, 9(1), 9288.
https://doi.org/10.24294/jipd9288

Article

Influence of environmental renewal on students' mental health development under the background of long-term online teaching

Ligui Zhu*, **Boli Li**

International College, Krirk University, Bangkok 10220, Thailand
*** Corresponding author:** Ligui Zhu, zzaixx100@sina.com

CITATION

Zhu L, Li B. (2025). Influence of environmental renewal on students' mental health development under the background of long-term online teaching. Journal of Infrastructure, Policy and Development. 9(1): 9288. https://doi.org/10.24294/jipd9288

ARTICLE INFO

Received: 15 March 2024
Accepted: 25 September 2024
Available online: 25 January 2025

Abstract: In the wake of the COVID-19 pandemic, the prevalence of online education in primary education has exhibited an upward trajectory. Relative to traditional learning environments, online instruction has evolved into a pivotal pedagogical modality for contemporary students. Thus, to comprehensively comprehend the repercussions of environmental changes on students' psychological well-being in the backdrop of prolonged online education, this study employs an innovative methodology. Founded upon three elemental feature sequences—images, acoustics, and text extracted from online learning data—the model ingeniously amalgamates these facets. The fusion methodology aims to synergistically harness information from diverse perceptual channels to capture the students' psychological states more comprehensively and accurately. To discern emotional features, the model leverages support vector machines (SVM), exhibiting commendable proficiency in handling emotional information. Moreover, to enhance the efficacy of psychological well-being prediction, this study incorporates an attention mechanism into the traditional Convolutional Neural Network (CNN) architecture. By innovatively introducing this attention mechanism in CNN, the study observes a significant improvement in accuracy in identifying six psychological features, demonstrating the effectiveness of attention mechanisms in deep learning models. Finally, beyond model performance validation, this study delves into a profound analysis of the impact of environmental changes on students' psychological well-being. This analysis furnishes valuable insights for formulating pertinent instructional strategies in the protracted context of online education, aiding educational institutions in better addressing the challenges posed to students' psychological well-being in novel learning environments.

Keywords: online teaching; mental health; CNN; SVM; multimodal features; psychological states environmental renewal

1. Introduction

In recent years, Internet technology has been widely popularized, and the influence of information network on people has deepened. In the integration of information technology and education, the role of network in teaching cannot be ignored (Basic, 2017). Against the backdrop of the novel coronavirus pandemic, the popularity of online teaching in basic education has risen. Through online teaching, students can not only obtain diverse learning resources conveniently but also communicate with other learners. Compared with the traditional learning environment, online teaching is favored by a growing number of students, and it has become an important learning method for modern students (Orchard and Fullwood, 2010). The age of students is an important factor of their physiological and psychological development. Students' mental health level shows a downward trend with the increase of age, so we should pay enough attention to students' mental health. Students are active participants and promoters of online life, and they can make good use of the convenience and advantages of the network. However, excessive Internet use can lead to a series of psychological and behavioral problems, which can easily cause social problems (Weinstein and Lqoyeu, 2010). The excessive use of the network will make people mentally dependent, which may not only cause plant nerve disorders but also lead to a decline in immunity. At the same time, the excessive use of the Internet will make individuals seem introverted, feel a sense of inferiority, and refuse to communicate with others, which will likely cause anxiety and depression and induce other emotions. Therefore, the lack of research on the impact of environmental renewal on students' mental health development under the background of long-term online teaching is a prominent problem that needs to be solved urgently. The current educational practice also brings new problems and challenges to online teaching.

At present, in-depth research is conducted on teenagers' mental health under the network environment. The network learning environment hinders the socialization of individuals, which limits the communication between learners, and the communication between teachers and learners is insufficient (Cho et al., 2012). As such, the impact of online teaching environment renewal on students' mental health needs to be further studied. Artificial intelligence is a branch of computer science, which can simulate people's thinking process (Romeo, 2016). Computer hardware have been constantly updated, and the Internet has also brought massive data sets, thereby increasing the accuracy of artificial intelligence models (Valeria, 2017). The biggest advantage of artificial intelligence is that it can save on substantial labor costs, because it requires less physical labor and more intellectual labor. Artificial intelligence can also be used for various types of tasks, including fact-based decision making, rather than emotion-based decision making, which is beneficial for business decisions. When artificial intelligence is applied to the prediction of students' mental health, the features can be changed from high dimension to low dimension by multilayer neural network, and the optimal solution of the model can be found by layer-by-layer training method (Huang et al., 2020). This process involves extracting the multilayer features of a neural network by selecting the appropriate network structure, mapping the multilayer features on the subspace, then performing adaptive weighted fusion through the defined central variables in each subspace. Applying artificial intelligence technology to predict students' mental health under the background of long-term online teaching can improve the accuracy of the prediction, which can be important reference for formulating online teaching strategies.

Therefore, based on artificial intelligence technology, this study puts forward a model that can accurately predict students' mental health. The existing research on mental health prediction is used for reference, and a prediction model of students' mental health based on artificial intelligence technology and multimodal fusion method is constructed. The model introduces the multimodal fusion method of attention mechanism to construct the mental health model and identifies the emotional characteristics of psychological counseling on the basis of SVM. Finally, the model is tested and verified.

2. Literature and definitions

2.1. Relationship between environment renewal and students' mental health

Through the collation and analysis of relevant literature, it is discerned that researchers predominantly concentrate on two aspects: firstly, the correlation between the utilization of online teaching and students' psychological well-being, and secondly, the psychological health education strategies for students within the online teaching environment (Gunes and Piccardi, 2007). A comprehensive examination of the relationship between online teaching and students' psychological well-being becomes particularly imperative, as the viewpoints of numerous researchers harmoniously underscore the intricacies inherent in this realm of inquiry. Paudel (2021) accentuates the notable advantages of online teaching in terms of flexibility and educational accessibility, yet issues a cautionary note regarding potential challenges such as social isolation and their impact on students' psychological well-being. Furthermore, Wilbraham et al. (2024) directs attention to the promotion of students' self-directed learning abilities through online teaching, deeming it pivotal for positive psychological development. Emphasizing the potential role of personalized learning tools in alleviating academic pressures, the study highlights their significance. Hilliard et al.'s (2020) research focuses on the nexus between technological challenges stemming from online learning and student anxiety, underscoring the necessity for educators to prioritize mental health support services to ensure students receive essential assistance during online learning endeavors. Lastly, Turk et al. (2022) underscores the potential influences of the online learning environment on students' social skills and emotional expression, advocating for the incorporation of collaborative learning and team projects to foster communication and cooperation among students. The convergence of these diverse perspectives accentuates the multifaceted linkages between online teaching and students' psychological well-being. While online teaching contributes to enhancing students' confidence and sparking creativity, excessive internet use also yields adverse effects. In the protracted milieu of online teaching, students are prone to experiencing an array of psychological issues.

A scrutiny of the pertinent literature on the relationship between the online teaching environment and psychological well-being reveals a dearth of research on the updates to the online teaching environment and its impact on students' mental health (Wu et al., 2020). Furthermore, there exists a lack of a specific definition for the variable of the online teaching environment, coupled with an absence of quantitative analysis. Some studies overlook the affirmative aspects of the online teaching environment, such as the beneficial utilization of information resources for alleviating learning anxiety. This oversight impedes a comprehensive understanding of the online teaching environment and hinders the establishment of corresponding intervention mechanisms. Hence, researching the influence of environment updates on the psychological development of students in the context of prolonged online teaching is deemed imperative.

2.2. Students' psychological analysis based on artificial intelligence

In recent years, artificial intelligence technology was widely developed, which also attracted increasing research attention. The states can be expressed in various ways, and multimodal emotion recognition has always been an important branch in the field of mental state prediction. The multimodal emotion recognition has also attracted growing attention. Multimodal fusion is a key research point in multimodal research, which integrates information extracted from different modalities into a stable multimodal representation. A clear link exists between multimodal fusion and representation, and if a process is focused on the use of a certain architecture to integrate different single-modal representations, then it is classified under the fusion class. At present, multimodal fusion methods are applied in different scenes such as visual question and answer, behavior recognition, image subtitle generation, mental state recognition and emotion analysis, and they rather vary in different application scenes (Lin et al., 2018). In the scenario of mental state prediction, to improve the accuracy and timeliness of mental state prediction, scholars put forward a multimodal emotion recognition method that combines acoustic and video images. Analyzing people's psychological state only by integrating all the ways people show their psychological state is the most reasonable approach (Zhang et al., 2019).

In the realm of online education, multimodal learning data encompasses objective and observable physiological data, including electroencephalography and electrocardiography, as well as subjective and indirectly observable psychological data through self-administered questionnaires, and behavioral data, such as body movements and learning logs. The collection of diverse multimodal learning data is intricate, given variables such as varying levels of student knowledge, classroom discipline, technical malfunctions, and unforeseen disruptions. Educational data generated during the learning process is dynamic and continuously evolving, with measurements of learners' psychological and physiological data changing in real-time in conjunction with their cognitive abilities, learning environments, and attitudes.

With the evolution of deep learning technologies, sentiment analysis has transcended traditional text-based research boundaries. Presently, various multimodal sentiment analysis models, based on neural networks, have been proposed and achieved significant success. Yu et al. (2016) employed convolutional neural networks for both images and text, extracting features separately and then concatenating these modal features to train a logistic regression model for sentiment classification. In sentiment classification, text and image information complement and interact with each other. Ju et al. (2021), in a recent end-to-end Multimodal Aspect-Based Sentiment Analysis task, observed that text generally plays a more crucial role than images, while images may provide vital cues for text (Yang et al., 2022). Consequently, Xu et al. (2018) introduced a co-attention mechanism to model the interaction between text and images.

Based on artificial intelligence technology, we can predict students' psychological state combined with various modal information, and we can get more accurate results. Therefore, with this technology as basis, we study the impact of environmental renewal on students' healthy development under the background of long-term online teaching.

3. Method

3.1. Model framework

When examining the impact of environmental renewal on students' healthy development under the background of long-term online teaching, we must first build a model that can accurately predict students' psychological state. This study puts forward a prediction model of students' mental health on the basis of artificial intelligence technology that has high accuracy for recognizing students' mental health state. By using the method of cross-modal attention mechanism, we fuse the multimodal data to recognize students' psychological state. The overall model process is divided into five steps, as shown in **Figure 1**.

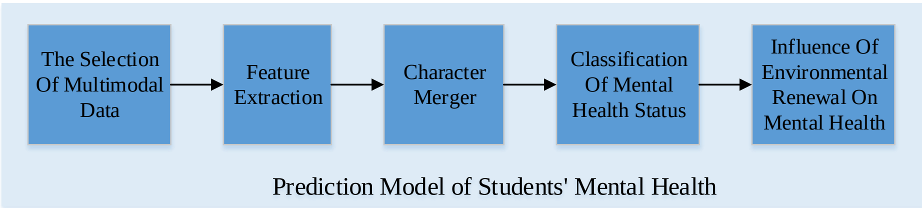

Figure 1. Overall process of prediction model.

In the prediction model of students' mental health based on artificial intelligence technology, three modal data are used as the input of the model. The input includes three low-level feature sequences from image, acoustics, and text. The features of each mode and their methods vary; different neural networks are used to extract the features.

3.2. Data preprocessing

This paper collected the behavior data of 300 students from a university during their online learning of linear algebra courses in MOOCs from January to June 2022, totaling more than 1 million online learning behavior records, and the students' personal information was encrypted. Moocs is a distance education learning platform based on microservice architecture, which can provide long-term learning environment for learners. Moocs provides activity module, data module and homework module for students and teachers. Students can learn courseware and video resources on the platform, teachers can set class check-in tasks and upload courseware resources on the platform, and the platform also provides teaching administrators with students' learning logs, which record the detailed operations of students after logging in to the platform.

Given the influence of environmental factors on video capture, video data may contain noise. To extract more precise facial and gait key point information and ensure model accuracy, it is necessary to preprocess the original video signal, including denoising and resampling.

The first step involves denoising. Since the captured video data predominantly consists of low-frequency data, this study employs low-pass filtering to denoise the original low-frequency signal. Specifically, a sliding window is used for Gaussian filtering to eliminate noise from the video data. The study utilizes the GaussianBlur function from the OpenCV library to perform denoising on the video data, setting the Gaussian kernel size to 3 × 3.

After the original text and video data are preprocessed, text set F and video image set G are obtained, as shown in Equations (1) and (2) respectively.

$$F = \{f_1, f_2, \dots, f_t\} \tag{1}$$

$$G = \{g_1, g_2, \dots, g_t\} \tag{2}$$

The *t* value depends on the size of the number of video frames, where f_t and g_t are single-layer text and video images, respectively, and their data storage format in the computer is matrix.

To standardize the format of audio data, we converted all audio files to WAV format. Next, to ensure consistency in the voice samples, we uniformly trimmed the audio files. Considering that subjects require an adaptation period when entering the context of voice capture, we selected the middle portion of the voice data as experimental data, extracting 90-second audio segments from the original audio files as subsequent inputs.

Furthermore, denoising the audio signals is essential. In this study, we employed Wiener filtering for audio denoising. Wiener filtering removes random noise, or environmental noise, from the audio signal through the use of the impulse response function. By synthesizing voice signals with added white noise, the study constructs the RXX matrix and RXS vector from the synthetic audio signal. Multiplying the inverse of the RXX matrix by the RXS vector yields the filter matrix, which is then applied to the original voice signal to obtain the denoised audio signal. We used the wiener function from the scipy.signal package to implement the aforementioned Wiener filtering process for audio denoising.

After the raw speech data is preprocessed, we divide the speech data into multiple speech sequences at 1-second intervals. The resulting audio set V is shown in Equation (3).

$$V = \{v_1, v_2, \dots, v_t\} \tag{3}$$

where the *t* value depends on the size of the speech length, the data storage format of v_t in the computer is a matrix.

3.3. Feature extraction

BERT, by utilizing the multi-head self-attention mechanism, is capable of learning both local and global features in text while taking into account the contextual information surrounding the text. This study employs a pre-trained BERT model (BERT-Chinese-case) to extract textual features and obtain corresponding text vectors. The text definition within students' online learning dynamics is defined as F, which may encompass multiple sentences S_i.

$$F = [S_1, S_2, S_3, \cdots, S_M] \tag{4}$$

For each sentence S_i consists of a sequence of words, see Equation (5).

$$S = [w_1, w_2, w_3, \cdots, w_N] \tag{5}$$

BERT acts as the text encoding layer BERT Encoder to encode the text vector and obtain the text vector H, see Equation (6).

$$H = BERT \text{ Encoder } (F) = [h_1, h_2, h_3, \cdots, h_k] \tag{6}$$

This study uses the VGG16 model trained on the ImageNet dataset as a pre-trained model to acquire image features. For a given set of matching images in the dynamic data released by university students G , which consists of 1 to J. It consists of 1 to 2 images.

$$G = [g_1, g_2, g_3, \cdots, g_j] \tag{7}$$

After the VGG16 network, the feature vector of the image can be obtained at the last pooling layer POOL H^g, for each image in the set G For each image in the set g_j POOL can get the feature vector of the response for each image in the set H_j^g. The last pooling layer of the VGG16 network

$$H_j^g = \text{POOL}\left(g_j\right) = \left[h_1^g, h_2^g, h_3^g, \cdots, h_k^g\right] \tag{8}$$

3.4. Feature fusion

After obtaining the text, speech and facial features, the attention mechanism is used to fuse these features. First, we cascade fusion of three modal features, then introduce the attention mechanism to the fused features and calculate the attention coefficient of each modality, then multiply the modal features with their corresponding attention coefficients and carry out the second cascade fusion, then we get the final multimodal fusion features.

First we combine textual features F_{fea}, speech features V_{fea} and image features G_{fea} After fusion, we get the feature matrix Fus satisfies Equation (9), and the matrix size is 103×4. The size of the matrix is

$$\text{Fus} = \left[F_{fea}, V_{fea}, G_{fea}\right] \tag{9}$$

Next, the multimodal features are input as the Key in the attention mechanism, and the Query is the output vector of the mental recognition based on the fusion feature matrix Fus. The calculation process of the attention coefficients is as follows: firstly, the corresponding weight coefficients are calculated according to Query and Key, and then the corresponding Value and the calculated weight coefficients are weighted and summed up, the specific calculation steps are as follows.

To calculate the similarity between Query and Key, the calculation of similarity consists of two aspects, namely, the dot product between two vectors and the cosine similarity, which are shown in Equations (10) and (11), respectively.

$$\text{Similarity1}\left(\text{Query}, \text{Key}_i\right) = \text{Query} \cdot \text{Key}_i \tag{10}$$

$$\text{Similarity2}(\text{Query, Key }_i) = \frac{\text{Query} \cdot \text{Key}_i}{\| \text{ Query } \| \cdot \| \text{ Key}_i \, \|} \tag{11}$$

The similarity parameter is normalized to obtain the weight coefficients.

$$\alpha_i = \text{softmax}\,(\text{Sim}_i) = \frac{e^{\text{Sim}_i}}{\sum_{j=1}^{L_x} e^{\text{Sim}_i}} \tag{12}$$

Weighted summation of the weighting factors is calculated by Equation (13).

$$\text{Attention (Query, Source)} = \sum_{i=1}^{L_x} \alpha_i \cdot \text{Value}_i \tag{13}$$

3.5. Mental health classification

Finally, the classifier classifies the mental health status according to the fusion features, and the mental health recognition results are the output. The ultimate goal of using the SVM method to train the model is to optimize a plane so that all the data or support vectors in the two classification sets can be furthest away from the plane. The ultimate goal of using the SVM method to train the model is to optimize a plane so that all the data or support vectors in the two classification sets can be furthest away from the plane.

We use the Gaussian function as the kernel function of the SVM classifier, and the Gaussian function is defined as shown in Equation (14).

$$K(x, z) = \exp\left(-\frac{\parallel x - z \parallel^2}{2\sigma^2}\right) \tag{14}$$

where z denotes the center value of the Gaussian kernel function, and σ is the width parameter of the function. In this paper, we choose to call the svm method in the sklearn package to train the model, and the parameters are set as follows: the classification kernel function is set as a Gaussian function, the penalization coefficient of the error term is set as 1, and the degree of the polynomial function inside the SVM is set as 3. The dataset is divided into training set, validation set $6:2:2$ The data set is divided into training set, validation set and test set, and the model is trained by ten-fold cross-validation method. Through the above training process, six SVM classifiers were obtained for recognizing the six psychological indicator states.

After the fusion features are obtained by the multimodal fusion method with attention mechanism, we use SVM to train the mental health recognition model, and the SVM classifier outputs the mental health status of the tested samples. The output mental health categories include depression, learning anxiety, fearful emotion, hostile emotion, obsessive-compulsive disorder, and normal. **Table 1** lists each mental state and its representative meaning.

In this study, six classifiers are obtained by training. These six judgment functions are used to judge different mental health conditions. Each mental health index has two states, namely, positive and negative. The positive indicates that the tested sample suffers from the mental illness corresponding to the index, whereas negative indicates that the tested sample does not.

Table 1. Mental states and their representative meaning.

Mental state	**Representative meaning**
Depression	It is a mental illness with abnormal low spirit as the main clinical manifestation.
Learning anxiety	Learning anxiety is often manifested as restlessness, inferiority, self-blame, headache, dizziness, panic, and impatience.
Fear	The fear caused by online learning is often manifested as weariness.
Hostility	It refers to the conflict of interests.
Obsessive-compulsive disorder	It is a group of neuropsychiatric diseases with forced thinking and forced behavior as the main clinical manifestations.
Normal	It indicates that the students' mental state is normal during online teaching.

4. Results and analysis

4.1. Model training

In the course of experimentation, the selection of parameters plays a pivotal role in the training performance of the model. To ascertain the optimal parameters for the model, this paper conducted comparative experiments based on the number of iterations within the model. While maintaining other parameters constant, the training of the model was adjusted by varying the number of iterations.

As the number of training iterations changed, the loss function of the training data exhibited a trend, as illustrated in **Table 2**.

Table 2. The loss function value.

Number of iterations	5	10	15	20	25	30	35	40
Loss function value	0.3255	0.2587	0.2349	0.2158	0.1958	0.1768	0.1772	0.1759

The number of training iterations has a discernible impact on the final accuracy of the model. In a proper training process, each completed training round should result in a reduction in the loss function and an increase in accuracy. After a certain number of iterations, the loss function should reach its minimum, and accuracy should attain its maximum.

According to **Table 3**, as the number of iterations in the online learning behavior dataset increased, the loss value continuously decreased from 0.3255 until reaching approximately 0.1768 at 30 iterations, where it then plateaued. The loss function saw significant reductions, and the model's accuracy markedly improved with an increase in iterations. However, beyond a certain point, further increases in iterations yielded diminishing returns in terms of accuracy.

Notably, after a relatively few iterations, the model demonstrated commendable accuracy, indicating that it possesses strong performance capabilities. The comparison of different models in **Table 3** demonstrates the performance variations in predicting student mental health using various architectures. The table presents the accuracy rates of different models, offering a comparative perspective on their efficacy.

Table 3. Model comparsion.

Models	Accuracy/%
BiLSTM (Gui et al., 2019)	70.48
CNN (Cai et al., 2019)	62.40
Concat (Zeng et al., 2020)	69.85
HFM (Truong and Lauw, 2019)	66.73
Ours	73.49

The BiLSTM model achieves an accuracy of 70.48%, indicating a reasonable performance in capturing sequential data patterns. However, the CNN model lags behind with an accuracy of 62.40%, suggesting that its approach may not be as effective for this specific task, possibly due to its focus on spatial features rather than temporal patterns. The Concat model shows an accuracy of 69.85%, which is slightly lower than BiLSTM. This suggests that the model's concatenation approach may not leverage the strengths of individual modalities as effectively as expected. The HFM model attains an accuracy of 66.73%, placing it between CNN and Concat. This may imply that the hierarchical fusion approach, while capturing different levels of abstraction, might not optimize the integration of diverse data types for this task. The proposed model ("Ours") outperforms the other models with an accuracy of 73.49%. This indicates its superiority in predicting student mental health, potentially due to its advanced integration of multimodal data sources and optimized model architecture.

4.2. Correlation analysis

To illustrate the validity of the proposed prediction model, an experimental study is carried out on the model. This study comprehensively considers the educational informatization level of alternative schools, age group, and school region to select the research object. According to the structural equation model scores provided by the references, the educational informatization level of alternative schools is divided into three levels, namely, high, medium, and low. According to the differences of students' physical development and cognitive level, the candidates are divided into four age groups. Significant differences are also noted in students' physical development level and cognitive level in the same age group and different regions. Therefore, when choosing research objects, the school areas are divided into three types according to the local economic development level, namely, high, medium and low. To further study the influence of online teaching environment on students' mental health, the online learning environment is divided in this study, as shown in **Table 4**.

At home and abroad, the network learning environment has been examined in depth, and network learning environment standards have been clearly defined. According to actual needs, this study adjusts some topics and options in the references, and the scale is composed mainly of the physical environment and human environment. According to the definition of network learning environment, divided into several factors, namely mobile phone and network equipment, ways and methods of online learning, information resources, online learning time, students online learning assessment, online teaching interaction and online learning self-control ability, which will provide the basis for subsequent analysis.

Table 4. Influence factors of network learning environment.

Subscale	Factors
Physical environment	Mobile phone and network equipment Ways and methods of online learning Information resources
Human environment	Online learning time Students' online learning assessment Online teaching interaction Online learning self-control ability

Table 5 presents the correlation analysis of students' online learning environment and their mental health. The physical environment and humanistic environment are significantly positively correlated with the mental health of students, among which the physical environment is the highest and the humanistic environment is the lowest. In terms of learning anxiety, the physical environment is significantly positively correlated with it, and the humanistic environment has the lowest correlation. In terms of hostility, the humanistic environment is significantly positively correlated with it, whereas the physical environment has the lowest correlation. In terms of fear, only the physical environment is significantly positively correlated with it, whereas depression is the most correlated with online learning environment. The Internet time is significantly correlated with other mental health dimensions except obsessive-compulsive disorder. No significant correlation is observed between online learning time and mental health, so reducing online time other than learning is the key to improving mental health.

Table 5. Correlation analysis between online learning environment and mental health.

Correlation	Physical environment	Human environment
Depression	Significant positive correlation	Significant positive correlation
Learning anxiety	Significant positive correlation	Lower positive correlation
Fear	Significant positive correlation	Negative correlation
Hostility	Significant positive correlation	Negative correlation
Obsessive-compulsive disorder	Significant positive correlation	Negative correlation
Normal	Positive correlation	Negative correlation

The results of the student mental health prediction model based on multimodal analysis **(Table 6)** reveal notable characteristics, but also highlight certain limitations. Firstly, the model performs well in predicting fear and hostility, achieving an accuracy of 77.57% and 76.73% respectively. This may be attributed to the distinctiveness of these emotions in students, which allows for easier identification through multimodal data.

However, the model's accuracy in predicting learning anxiety is relatively low, only 68.49%. This could be due to the complex nature of learning anxiety, which poses challenges for the model's prediction. This result suggests that the model may need further optimization when addressing more complex and subtle mental health issues.

Table 6. Model prediction accuracy.

Mental health indicators	Accuracy/%
Depression	74.64
Learning anxiety	68.49
Fear	77.57
Hostility	76.73
Obsessive-compulsive disorder	74.37
Normal	74.56
Overall	73.49

4.3. Discussion

Overall, the model's accuracy in overall prediction is 73.49%, indicating that multimodal analysis holds potential in predicting student mental health. Nevertheless, this level of accuracy leaves room for improvement. The significant differences in prediction accuracy across various mental health indicators could be attributed to factors such as data quality, sample bias, and model design.

Demographic differences are also between students' mental health and online learning environment. Specifically, girls have more mental health problems than boys. The configuration of boys' network equipment is better than that of girls', but there are also more network problems than that of girls. The students in senior grades have better network hardware facilities and richer network learning resources than those in lower grades, and the key schools are better than non-key schools in the construction and utilization of network information and network learning resources.

The correlation between the learning environment and students' mental health is significant, and the physical environment has the highest correlation. Significant positive correlation are also observed with mobile phone network equipment, ways and means of surfing the Internet, students' self-control ability, and mental health. On the whole, mobile phone and network equipment, Internet access methods, and students' self-control ability can positively predict the mental health of students, whereas information resources can negatively predict it. Although students' online learning conditions are related to all dimensions of their mental health, most of their mental health problems stem from the incorrect use of the network. This finding also shows that configuring Internet equipment is appropriate for parents and schools, which should also teach students scientific ways to use the Internet.

The use of information resources can predict the improvement of mental health, so students should be guided to actively use online learning resources and strengthen the construction of online learning resources. Firstly, the online learning resource utilization needs to be improved. As students' guide in their learning and in life, teachers can not only instruct students in using the network to assist in their learning in the classroom but also pay attention to giving full play to the role of the network outside the classroom. Second, the construction of network learning resources should adapt to the development trend of mobile Internet. The traditional educators should keep up with the development of the times, which should also actively integrate traditional educational institutions and learning resources. Finally, parents and teachers should control the time for students to use the Internet for entertainment, they should also encourage students to use the Internet to obtain information and learn knowledge.

Students’ good self-control ability can significantly predict the improvement of their mental health, which indicates that strengthening the cultivation of students’ self-control ability is an important factor to improve their mental state.

The improper ways and means of surfing the Internet can lead to high learning anxiety of students, which then lead to self-blame, terror, and loneliness. If students’ online learning styles and methods are incorrect, the less time they spend on learning, and the more likely they are to have learning anxiety. Bad online learning methods can result in horrible thoughts or behaviors of students. Therefore, by strengthening online learning guidance to develop good online learning methods, students’ learning anxiety and other mental health problems can be alleviated. Based on the SVM algorithm, this study can achieve the accurate identification of six psychological characteristics but cannot describe individuals’ mental health level effectively, which must be divided comprehensively. Future research could focus on enhancing the model’s integration and processing of multimodal data to improve its accuracy in predicting complex mental health issues such as learning anxiety. Additionally, ensuring data quality and diversity is crucial to the model’s robustness and generalizability. With these improvements, the model will be better positioned to provide accurate and reliable predictions for students’ mental health.

5. Conclusion

This study employs an artificial intelligence-based prediction model to investigate the impact of environmental renewal on students’ mental health during an extended period of online teaching. The model incorporates a multimodal fusion approach with an attention mechanism, and emotional characteristics of psychological counseling are identified using the SVM. The results indicate that the model effectively identifies six psychological features, and the integration of the attention mechanism in feature fusion enhances overall performance. The study reveals that inappropriate internet usage can elevate students’ learning anxiety, resulting in self-blame, terror, and loneliness. However, the research suggests that providing robust online learning guidance can assist students in developing positive online learning styles, subsequently reducing mental health problems. This exploration of the influence of environmental renewal on students' mental health provides valuable insights for formulating teaching strategies amid the context of prolonged online education. The implications of this research extend to educational practitioners and policymakers. The identification of psychological features through advanced AI models can aid in the early detection and intervention of mental health issues among students. The emphasis on the role of online learning guidance underscores the importance of proactive measures to foster healthy learning environments, especially in situations where online education becomes a long-term or prevalent mode of instruction.

Furthermore, the study contributes to the growing body of literature addressing the psychological aspects of online learning, providing a nuanced understanding of how environmental factors, specifically internet usage, can impact students’ mental well-being. As educational institutions continue to navigate the challenges of online teaching, these findings offer actionable insights for adapting teaching strategies to promote positive mental health outcomes among students.

Conflict of interest: The authors declare no conflict of interest.

References

Basic, S., Markovic, I., Sporis, D., et al. (2017). Psychogenic non epileptic seizure status – diagnostic and treatment challenge. Psychiatria Danubina, 29(1), 87–89. https://doi.org/10.24869/psyd.2017.87

Cai, Y., Cai, H., & Wan, X. (2019). Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.18653/v1/p19-1239

Cho, S.-M., Sung, M.-J., Shin, K.-M., et al. (2012). Does Psychopathology in Childhood Predict Internet Addiction in Male Adolescents? Child Psychiatry & Human Development, 44(4), 549–555. https://doi.org/10.1007/s10578-012-0348-4

Gui, T., Zhu, L., Zhang, Q., et al. (2019). Cooperative Multimodal Approach to Depression Detection in Twitter. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 110–117. https://doi.org/10.1609/aaai.v33i01.3301110

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications, 30(4), 1334–1345. https://doi.org/10.1016/j.jnca.2006.09.007

Hilliard, J., Kear, K., Donelan, H., et al. (2020). Students' experiences of anxiety in an assessed, online, collaborative project. Computers & Education, 143, 103675. https://doi.org/10.1016/j.compedu.2019.103675

Huang, H., Hu, Z., Wang, W., et al. (2020). Multimodal Emotion Recognition Based on Ensemble Convolutional Neural Network. IEEE Access, 8, 3265–3271. https://doi.org/10.1109/access.2019.2962085

Ju, X., Zhang, D., Xiao, R., et al. (2021). Joint Multi-modal Aspect-Sentiment Analysis with Auxiliary Cross-modal Relation Detection. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.18653/v1/2021.emnlp-main.360

Orchard, L. J., & Fullwood, C. (2009). Current Perspectives on Personality and Internet Use. Social Science Computer Review, 28(2), 155–169. https://doi.org/10.1177/0894439309335115

Paudel, P. (2020). Online Education: Benefits, Challenges and Strategies During and After COVID-19 in Higher Education. International Journal on Studies in Education, 3(2), 70–85. https://doi.org/10.46328/ijonse.32

Romeo, V. (2016). Can Compulsive Internet Use Affect Adolescent Mental Health. Psychology Today, 1(06).

Shu, L., Xie, J., Yang, M., et al. (2018). A Review of Emotion Recognition Using Physiological Signals. Sensors, 18(7), 2074. https://doi.org/10.3390/s18072074

Truong, Q.-T., & Lauw, H. W. (2019). VistaNet: Visual Aspect Attention Network for Multimodal Sentiment Analysis. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 305–312. https://doi.org/10.1609/aaai.v33i01.3301305

Turk, M., Heddy, B. C., & Danielson, R. W. (2022). Teaching and social presences supporting basic needs satisfaction in online learning environments: How can presences and basic needs happily meet online? Computers & Education, 180, 104432. https://doi.org/10.1016/j.compedu.2022.104432

Valeria, D. V. (2017). The Draft of the ICD-11 Chapter on Mental Disorders: A Report for WPA Constituencies. Psychiatria Danubina, 29(01), 96-100.

Weinstein, A., & Lejoyeux, M. (2010). Internet Addiction or Excessive Internet Use. The American Journal of Drug and Alcohol Abuse, 36(5), 277–283. https://doi.org/10.3109/00952990.2010.491880

Wilbraham, S. J., Jones, E., Brewster, L., et al. (2024). Inclusion or Isolation? Differential Student Experiences of Independent Learning and Wellbeing in Higher Education. Education Sciences, 14(3), 285. https://doi.org/10.3390/educsci14030285

Wu, Z., Pan, S., Chen, F., et al. (2021). A Comprehensive Survey on Graph Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4–24. https://doi.org/10.1109/tnnls.2020.2978386

Xu, N., Mao, W., & Chen, G. (2018). A Co-Memory Network for Multimodal Sentiment Analysis. In: Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. https://doi.org/10.1145/3209978.3210093

Yang, L., Na, J.-C., & Yu, J. (2022). Cross-Modal Multitask Transformer for End-to-End Multimodal Aspect-Based Sentiment Analysis. Information Processing & Management, 59(5), 103038. https://doi.org/10.1016/j.ipm.2022.103038
Yu, Y., Lin, H., Meng, J., et al. (2016). Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks. Algorithms, 9(2), 41. https://doi.org/10.3390/a9020041
Zeng, Q., Li, X., & Lin, H. (2020). Concat Convolutional Neural Network for pulsar candidate selection. Monthly Notices of the Royal Astronomical Society, 494(3), 3110–3119. https://doi.org/10.1093/mnras/staa916
Zhang, S., Tong, H., Xu, J., et al. (2019). Graph convolutional networks: a comprehensive review. Computational Social Networks, 6(1). https://doi.org/10.1186/s40649-019-0069-y