An Intelligent ICT Framework for Emotion Classification and Adaptive Feedback in Psychological Education

Xiaoyan Zhou
College of Humanities Education and Arts, Hefei Professional College of Economics and Technology, Hefei, Anhui, 230031, China
E-mail: 15255129328@163.com

Keywords: emotion classification, adaptive feedback, psychological education, intelligent ICT framework, deep learning, affective computing

Received: January 8, 2026

To enhance the effectiveness of psychological education, this study proposes an Intelligent ICT framework for emotion classification and adaptive feedback. The framework employs a hybrid deep learning architecture integrating Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks to analyze multimodal inputs, including textual sentiment, vocal tone, and facial expressions, for accurate emotion recognition. Based on real-time emotional state detection, the system delivers personalized learning interventions and adaptive feedback to improve learner engagement and emotional well-being. Experimental evaluation conducted with 320 participants demonstrates an emotion recognition accuracy of 94.6% and an F1-score of 92.8%. Additionally, the proposed approach achieves a 23% improvement in learners' psychological flexibility compared to conventional feedback systems. The results indicate that the proposed ICT framework effectively supports emotion-aware digital learning environments and enhances outcomes in psychological education.

Povzetek: Študija predstavlja inteligenten IKT-okvir, ki z analizo čustev omogoča prilagojene povratne informacije in izboljšuje psihološko izobraževanje.

1 Introduction

As scientists work to create computers programmed to read and respond to people's emotions, the area of emotion classification has attracted much interest. Emotion recognition offers a promising avenue for enhancing human-AI interaction in various domains, such as medical care, schooling, and customer service, by deducing feelings from textual, visual, and audio data. It is challenging to create AI models that can faithfully depict emotional expressions due to the intrinsic complexity and nuance of emotions. Despite progress in technology, emotion-aware AI systems still face obstacles that hinder their usefulness. These include diverse datasets, limitations on real-time applications, and computing needs. The development of emotionally sophisticated AI in the future depends on fixing these problems. Technology is becoming more embedded in people's daily lives as the demand for AI systems that can empathise with people increases. The public looks to AI to make thoughtful, person-centred decisions based on their data. Virtual assistants and mental health services benefit greatly from AI models that can respond sensitively and appropriately to users' emotions [1].

Adaptable and generalisable models are also required due to the growing complexity of feelings among individuals in intercultural and multilingual contexts. By surveying the current literature on emotion classification, this study hopes to fill these knowledge gaps by locating the methods that provide the most significant potential for conquering these obstacles. To better understand the current state of emotion classification in AI systems, as well as its potential future directions, we undertook a comprehensive literature review (SLR). We found the most promising ways to enhance the accuracy, scale, and contemporaneous application of emotion-aware systems by reviewing existing methodologies [2]. Additionally, we delved into how AI emotional intelligence has been enhanced through the utilisation of emerging technologies like attention mechanisms and multimodal data processing. We also looked at ways AI can be accessible and flexible enough to work in various cultural and language settings. Future studies can use the results as a point of reference. All aspects of human existence and activity rely on the expression of emotions. William James (1884), the father of American psychology, provided the first known definition of "emotion" in his writings. Physiological alterations, including changes in facial expression, muscle tension, and especially visceral activity, are inseparable from any emotion, according to his theory [3]. Lange (1885), a Danish physiologist, also took a similar stance, arguing that emotions are both physiological states that incorporate perceptions, ideas, and actions, and psychological reactions to a wide range of environmental stimulation.

This has led academics from many different disciplines to agree that precise emotion recognition is crucial. Many disciplines, including clinical therapy, affective computing, and psychology, have recently made use of emotion recognition research. Over 700,000 individuals die by suicide each year, while an estimated 280 million people suffer from depression, according to information derived from WHO statistics (WHO). Bipolar disorder (BD) is one of numerous emotional disorders; it is defined by manic and depressed episodes that occur at regular intervals. Extreme manifestations of common feelings, including joy, disgust, sadness, and rage, can also be seen in BD's manic and pathological stages. Many people may have difficulty differentiating between distinct emotions, such as guilt and shame or dread and anxiety, which makes it difficult for them to adequately articulate complicated feelings [4]. More extreme ups and downs in mood than in healthy people is a common symptom of emotional disorders, which may indicate disease progression, relapse risk, and reduced functioning.

As a result, there is clinical value in continuously tracking emotional instability along with other characteristics that can indicate disease activity, like the length, intensity, and frequency of symptoms. In psychiatric research, self-monitoring is commonplace. Humans can express emotions in a variety of ways, including written words, spoken words, and facial expressions, and even through physiological signs [5]. When it comes to managing and keeping tabs on emotional shifts, many people turn to emotional charting tools like the Sickness the Procedure for the Assessment of Mood States (POMS), the Short Form 90 (SCL-90-R), or the Life Chart Method developed by the National Institute of Mental Health (NIMH-LCM). Subjective well-being is an all-encompassing cognitive index of how one's life is [6]. It has a substantial influence on a person's health, job, and social interactions, among many other things. Among the OECD's primary objectives are the promotion of social development and the enhancement of people's standard of living. Researchers observed that kids who had positive feelings when they were growing up were more likely to have self-assurance, happiness, and good health as adults. When a person has more positive emotional experiences than negative ones, they are said to be experiencing subjective well-being.

Research on the correlation between emotional intelligence and happiness has been conducted in several studies. Moods, both good and bad, are strong predictors of happiness, according to related research [7]. In recent years, there has been much focus in the area of education on academic emotions, which are the feelings that students go through when they are in an educational setting. Students' subjective well-being can be supported by research that measures academic emotions and aims to improve the accuracy of these measurements. Research on academic mood measurement using scales has recently shown substantial findings. In a virtual classroom, the learning platform serves as the primary means of student-teacher communication [8]. A great deal of intellectual Emotion permeates the platform's many student remarks (by "student comments" we mean text-type comments in this article). On the other hand, it is not easy to use an emotional scale to quantify student feedback in an online classroom.

The remainder of this paper is structured as follows: The scholarship on the development of emotion categorisation is reviewed extensively in Section 2. A comprehensive account of the methods employed to construct the psychological learning model is included in Section 3. Section 4 presents the study's findings, and Section 5 concludes.

1.1 Contribution of this study

An integrated Emotion Identification and Feedback Model customised for psychological education has been developed in critical research, which advances knowledge in the domains of AI and psychological education. The simulation improves the accuracy of detecting emotions among learners across textual, visual, or auditory data by integrating advanced approaches, including BERT, Graph Neural Networks (GCN), and Multilayer CNN feature extraction. Identifying emotions in academic settings is made more robust and precise by utilising multiple deep learning methods to collect subtle behavioural and emotional indicators in real time, as opposed to previous approaches that rely mainly on self-reported scales. Students' perception and regulation of emotions are mediated by feedback orientation (effectiveness, self-efficacy, interpersonal orientation, and action-taking responsibility), as the study reveals. This provides a more nuanced view of the psychological processes at work in feedback engagement. The study has important implications for education and practice beyond its technical contributions. It helps kids with more than just emotion recognition; it also helps them with better feedback responses, emotional management, resilience, and learning outcomes in general. In offline as well as online settings, the framework equips teachers with practical understandings of their students' emotional states, enabling them to employ more compassionate and tailored pedagogical approaches. Additionally, the study shows that psychological education and affective computing have the ability to improve students' mental health, lessen their adverse emotional reactions to academic difficulties, and create a more positive and productive classroom setting.

1.2 Research questions and objectives

Research Questions:

- Can a multimodal deep learning model (CNN-BiLSTM) improve emotion detection accuracy in psychological education settings compared to traditional methods?
- How does real-time emotion classification influence adaptive feedback delivery and student emotional regulation?
- What is the impact of personalized feedback interventions on learners' psychological flexibility and academic engagement?

Research objectives:

- To develop an intelligent ICT framework integrating CNN and BiLSTM for multimodal emotion recognition from text, audio, and facial expressions
- To achieve high-accuracy emotion classification (>90%) for real-time student emotional state detection
- To design and implement adaptive feedback mechanisms based on detected emotional states
- To evaluate the framework's effectiveness in improving students' psychological flexibility, emotional regulation, and learning outcomes

Expected outcomes:

- Enhanced emotion recognition accuracy (target: >92%)
- Improved student feedback responses and engagement
- Better emotional regulation and psychological resilience
- Increased learning effectiveness in psychological education contexts

2 Literature review

An essential new tool for assessing education and psychological assessment tool used in this research is the psychological assessment scale [9]. To determine the educational psychology of students, the Maemphasized scale is a valid and trustworthy instrument. It was established by demonstrating that its indicators agreed with psychometric standards. As previously mentioned, the scale's resilience and reliability are highlighted by its high internal consistency coefficient of 0.815. The six core characteristics of this measure are as follows: emotional regulation, physical self-esteem, psychological resilience, hope trait, learning motivation, and self-efficacy. As previously said, it is comprised of 27 components that have undergone extensive revisions and enhancements based on prior research [10]. Multiple statistical methodologies were employed to ensure the structure and content correctness of each item while it was being created. Both An (2023) and Zhan (2023) highlighted the accuracy and usefulness of the scale, which was developed through thorough data analysis, questionnaire surveys, and in-depth interviews with stakeholders. Preliminary factor evaluation results showed the presence of factor loads in excess of 0.40, and a cumulative translation rate of 52% suggested that the scale produced in this study had better construct validity compared to existing educational psychology assessments. On a similar scale, the results were lower for factor loads and interpreting rates [11].

2.1 Violent online harassment in online courses

The COVID-19 epidemic has accelerated the trend towards online education, which has brought with it a plethora of new possibilities and threats for both students and teachers. Problems include students' unequal issues with technology, internet connection, and the challenge of keeping up with engagement and motivation when face-to-face connection is lacking. Other significant challenges are the difficulty of preserving assessment integrity in virtual environments and the prevalence of social isolation among online learners. On the flip side, there are many advantages to online education, such as the ability to study whenever and wherever is most convenient for students, access to a wealth of educational materials, the ability to tailor lessons to individual students' needs through adaptive technology, improved communication and teamwork among teachers and students, and the possibility of international collaboration and networking that does not depend on physical proximity [12]. To fully realise the promise of online education, it is crucial to strike a balance between these obstacles and take advantage of the benefits. One of the most significant problems with online classrooms is cyberbullying, which includes threats, verbal attacks, exclusion from group discussions, and the distribution of false information. Because it disrupts the learning environment and can cause victims' emotional pain, this kind of violence is a primary concern for teachers and students alike.

2.2 Studying how to identify facial expressions

The problem of gauging student involvement in online courses has received more and more attention as of late. The capacity to monitor students' emotional and cognitive states independently of instructors has propelled engagement detection to the forefront of educational research. In this part, we will take a look at some of the ways that more than 500 participants to pin down student involvement [13]. By providing helpful insights regarding students' engagement and excitement, these methodologies enable educators to modify their teaching methods and support tactics in response, thus increasing the efficacy of online education. The writers have used students' eye and head motions to deduce their circumstances in virtual classrooms. In this research, we offer a system that can detect students' emotions and classify their engagement levels in traditional classroom settings. Researchers analyse students' Movements of the head as well as the ability to read students' facial expressions and eye movements to gauge their level of focus and involvement in traditional classroom activities.

2.3 Emotion models

Affective computing is a subfield of AI that aims to model human empathy through research into and creation of computational systems with enhanced human-like perception, cognition, and emotional interaction [15]. In this article, we suggest a method to train machines to detect and respond appropriately to human emotions by learning to read their facial expressions and other nonverbal cues. The two most prominent theories that attempt toexplain how people understand and label emotions are the dimensional and category approaches. Dimensional approaches methods characterise arousal, rhythm, management, intensity, duration, frequency of occurrence, etc., as progressive emotions, whereas categorical models group emotions into basic, additional, supplementary, etc. Emotions have the potential to enhance students' attention, memory, and reasoning in a learning environment [16]. In such a setting, instructors need to be ready to establish emotionally engaging learning environments, promote student-student collaboration in building knowledge, and recognise when students' emotions are getting in the way of their progress (Ibarrola, 2000).

2.4 Affective dictionaries

Subjective statements. One area that works with textual knowledge is sentiment analysis. That describes people's feelings, opinions, and appraisals towards events, objects, and their features. Based on the publication's context polarity and frequent phrases, classification algorithms are typically used to determine if a document is a favourable or detrimental critique. So, the studies depend entirely on how often specific terms appear or the valence of the text. [17]. The goal of lexical analysis in emotion research is to find terms that can indicate the writers' emotional states; this is a relatively recent approach. In this field, there are several popular affective dictionaries. These dictionaries offer a collection of words and phrases in several languages.

3 Methodology

Psychological Education's Emotion Detection and Feedback Model is depicted in Figure 1 as its workflow. Data cleaning, normalisation, and feature extraction are the steps that follow data gathering from text, audio, and video. After that, we build an emotion classification model, train it using cross-validation and hyperparameter tuning, and then apply it to advanced feature extraction using a multilayer CNN. In order to ensure that the system is continuously improving, it is evaluated and tested to determine its performance and educational effectiveness. Lastly, it incorporates a feedback system that offers individualised replies and ideas for intervention.

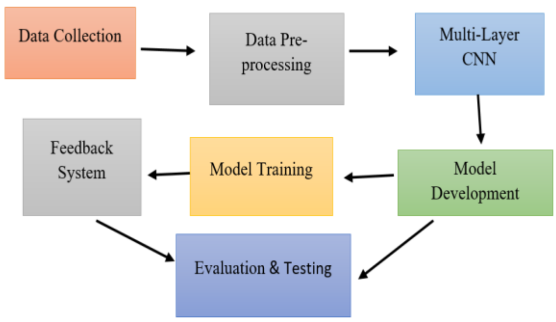

Figure 1: Methodology for Emotion Recognition and Comments in Psychological Training.

3.1 Data collection and preparation

In the summer of 2020, researchers used the Zoom videoconferencing program to conduct semi-structured interviews, as the COVID-19 epidemic imposed constraints on face-to-face interviews. In this way, we were able to gather data without putting any of the participants in danger. Among the rights that participants were informed of were the following: all participants have the following rights: the freedom to discontinue participation in the research at any moment without cause or prejudice; the Human Clinical Research Ethics Committee's consent was sought before the interviews. Before beginning the interviews, the interviewer made sure the participants knew what the research was about [18]. Following the recommendation, we tailored the interview questions to the specificity of the FEM context to make sure we had enough data from the small sample size. Researchers ensured consistency in data collection by using PowerPoint slides to convey interview questions in sync with vocal questioning.

3.2 Data processing

The raw data requires preprocessing before it can be studied. First, the OpenCV tool is used to prepare the facial images. We use the Haar cascade method to find faces in this part of the process. Also, each image is 224×224 pixels, which is the same size. The lighting can also be changed so that changes in the light do not affect the study. The sound signs that were picked up are being worked on in the library. In the preprocessing step of audio, frequency spectrum-based filtering gets rid of noise and makes the amplitudes more normal. You can also use the spectrograms you make to show these traits. The short-time Fourier transform (STFT) is used to get pitch and energy from a voice. With median filters, the monitor data about heart rate and other body functions is smoothed out. You can get rid of fake peaks and high-frequency noise with these filters. You can learn about heart rate variability, or HRV, to get a better sense of how someone is feeling. The correct way to talk about changing pictures is as a set of changes T that are made to the first picture I, as shown in Equation 1.

$$I' = T(I) = resize(normalize\left(face - detection(I)\right)) \quad (1)$$

\begin{equation}
I' = T(I) = \operatorname{resize}\!\left(
\operatorname{normalize}\!\left(
\operatorname{face_detection}(I)
\right)\right)
\end{equation}

The steps used to prepare audio data can be seen in Equation 2, which shows how to change the spectrum and filter the data:

$$A' = (t)STFT\left(filter\left(A(t)\right)\right) \quad (2)$$

\begin{equation}

A'(t) = \operatorname{STFT}\!\left(\operatorname{filter}\!\left(A(t) \right) \right)

\end{equation}

This graph shows the audio output that has not been filtered at time t. It is called A(t), and it shows the filtered spectrum. The ways that thoughts can be noticed

$$l(y, \hat{y} = -\sum_{i=1}^{N} y_i\, Log\, (\hat{y}_i) \quad (3)$$

\begin{equation}
\mathcal{L}(y,\hat{y}) = -\sum_{i=1}^{N} y_i \log \left(
\hat{y}_i \right)
\end{equation}

The CNN is used to analyze face images, and the RNN is used to analyze speech patterns and biological data. These are the building blocks on which the mood recognition models in this system are made. Convolutional neural networks (CNNs) are great at finding patterns in pictures that are spread out in space, while recurrent neural networks (RNNs) are great at handling input that comes in at different times and in other sequences, like audio and biometric measurements. Based on the ResNet-50 architecture, which is an efficient deep network for image classification tasks, the CNN model that is utilized for face emotion recognition is based on this design. While training the model, a named collection of facial emotions is used, and data enrichment methods are utilized in order to enhance the model's ability to generalize. The loss function that is used is the cross-entropy, which is shown below in Equation 3. One type of neural network that can find long-term connections in data patterns is called a Long-Short-Term Memory (LSTM) network. This network is used to study speech sounds. The loss function that is used is similar to that of CNN, and the LSTM model is trained using audio spectrograms that are tagged with the feelings that connect to them. The Gated Recurrent Unit (GRU) network is a type of recurrent neural network (RNN) that is easier to use than LSTM but can be just as useful in some situations.

It is used to process biological data. A mood classification is produced by the GRU model, which takes as input the time series of HRV as well as other physiological data. The classification is based on these traits. These models are trained in a high-performance computing environment so that the weights of the neural networks can be changed. With the help of streamlining tools like Adam, this change is possible. To make sure that the models can be used in other situations, they need to be tested on a different set of data. In order to determine how good the models are in identifying emotions, performance measures such as accuracy, sensitivity, and specificity are computed and evaluated.

3.3 Multilayer CNN feature extraction

For psychological training, Multilayer Convolutional Networks (CNNs) are an essential building block for the Emotions Classification and Feedback Model. Before being entered into the CNN framework, emotional data derived from video, audio, and text inputs are preprocessed to eliminate noise and guarantee normalisation [19]. An autonomous learning hierarchical representation is built into the multilayer convolutional neural network (CNN), with lower layers capturing fundamental information like pitch, tone, or facial expressions and deeper layers extracting intricate patterns that reflect nuanced emotional signals and behavioural traits. By using a layered extraction approach, the model is able to differentiate between various emotional states accurately. The approach uses CNNs' spatial and temporal dependency handling capabilities to make the classification more context-aware and robust while preserving crucial emotion indications. After the features have been retrieved, they are paired with an advanced categorisation module and a personalised feedback system to create adaptive interventions that help students better manage their emotions and achieve better academic results. In this study, multilayer feature extraction is built on top of VGGNet-16. The five convolutional blocks that make up VGGNet-16 include pooling and convolutional layers, as shown in Figure 2. Additionally, there are three fully connected layers. Several sampling techniques are used to standardise the feature sizes produced by each layer, expanding upon the VGGNet-16 design.

$$L^{(l)} = \rho(\tilde{A} L^{(i-1)} W^{(i)}) \qquad (4)$$

\begin{equation}

\mathbf{H}^{(l)} = \rho\!\left(\tilde{\mathbf{A}}\, \mathbf{H}^{(l-1)} \mathbf{W}^{(l)} \right)

\end{equation}

Where ρ is an activation function that is not linear, the weighting parameter matrix of its layer is denoted by W^((i)), and A ̃ is the normalised adjacency matrix.

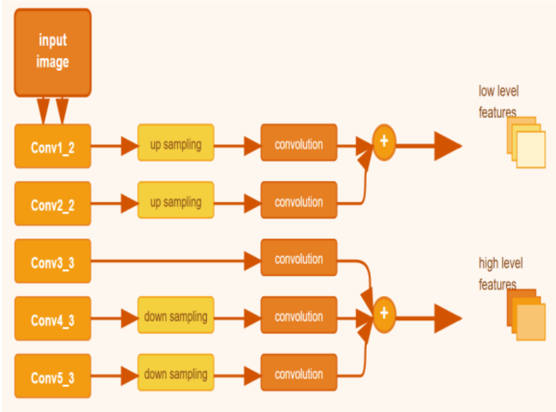

Figure 2. As derived from: Emotional input from users and visual context perception in the field of design for visual communication.

3.4 Model training

An essential part of building the Emotion Identification and Reinforcement Model for mental instruction is the model train phase. This part makes sure the system can generalise well to data that has never been seen before. The training of the classification model is carried out using the processed representations obtained after feature extraction using the multilayer CNN [20]. Parameter optimisation during training using cross-validation helps avoid overfitting and guarantees robustness across various data subsets. To reach the best possible performance, hyperparameter tweaking is also used to modify learning rates, batch sizes, layer counts, and dropout rates. The GNNs are fed the initialisation matrix X that was created after encoding using a large-scale pre-trained algorithm [21]. The document nodes' ultimate embedding representation is produced by iteratively and propagatingly using the GCN to generate each node embedding. As a pre-trained model, BERT is utilised in this article. Then, the statistical softmax layer receives outputs from both BERT and GCN, which are then linearly interpolated to provide predictions. The feature matrices are calculated during the GCN layer in the generation process.

$$Z_{GCN}=\text{soft max } (G\ (X, A)) \qquad (5)$$

\begin{equation}

\mathbf{Z}_{\text{GCN}}= \operatorname{softmax}\!\left(G(\mathbf{X}, \mathbf{A}) \right)

\end{equation}

$$Z_{BERT}=\text{softmax } (WX) \qquad (6)$$

\begin{equation}

\mathbf{Z}_{\text{BERT}}= \operatorname{softmax}\!\left(\mathbf{W}\mathbf{X} \right)

\end{equation}

$Z{=}\lambda Z_{ZCN+(1-\lambda)Z_{BERT}}$ (7)

\begin{equation}

\mathbf{Z} = \lambda \mathbf{Z}_{\text{GCN}} + (1-\lambda)\mathbf{Z}_{\text{BERT}}

\end{equation}

Z_GCN is the predicted outcome obtained by feeding the GCN's output into the softmax layer; G is the GCN model. To get the document storing the results when we input the BERT output for the softmax layer, we feed it into the pre-trained model to forecast the outcome. The name for this is Z_BERT. To optimise the model's inclusion and roduce more harmony between the two halves of the forecast, the interpolation factor of the two separate prediction portions, λ ∈(0,1), is changed.

3.4.1 BERT-GCN fusion architecture

Our framework employs a sequential-parallel hybrid architecture for integrating BERT and GCN, as illustrated in Figure 3. In the first phase, BERT processes the input text through its 12-layer transformer encoder to generate 768-dimensional contextualized embeddings for each token. These BERT embeddings are then used to initialize the node feature matrix X in the GCN, where each word becomes a graph node with edges defined by word co-occurrence statistics.

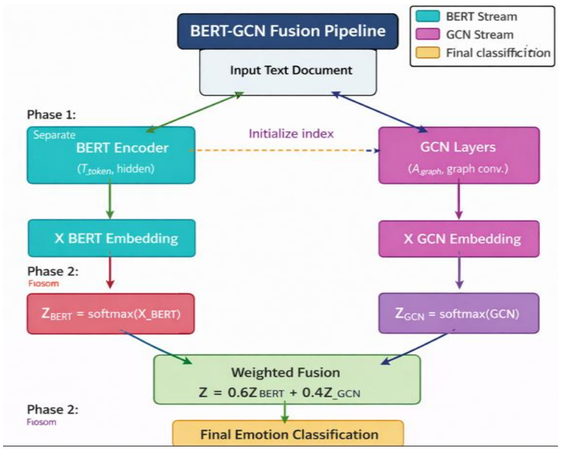

Figure 3: BERT-GCN Fusion Architecture showing sequential integration where BERT embeddings initialize GCN nodes, followed by parallel prediction streams combined via weighted interpolation (λ = 0.6).

In the second phase, two parallel prediction streams operate independently: the BERT [CLS] token embedding (X_BERT) is passed to a softmax classifier, while the GCN performs graph convolution through 2 layers using H^(l+1) = σ(Ã·H^(l)·W^(l)) to produce refined graph embeddings (X_GCN) that are also classified via softmax. Finally, both prediction streams are combined through weighted decision-level fusion: Z = λ·Z_BERT + (1-λ)·Z_GCN, where λ represents the interpolation coefficient.

3.4.2 Hyperparameter selection

The interpolation coefficient λ was determined through systematic grid search on the validation set, testing values from 0.1 to 0.9 with increments of 0.1. Using 5-fold cross-validation and F1-score as the evaluation metric, the optimal value of λ = 0.6 was selected, giving BERT predictions 60% weight and GCN predictions 40% weight. This configuration yielded the highest average F1-score of 92.3% across all emotion categories and remained stable (±0.5%) for values between 0.5 and 0.7.

3.4.3 Training configuration

The model was trained using the AdamW optimizer with differentiated learning rates: 2e-5 for BERT fine-tuning and 1e-3 for GCN and CNN layers trained from scratch. A cosine annealing learning rate scheduler with 500 warm-up steps was employed, decaying to a minimum of 1e-6. Training proceeded for up to 50 epochs with a batch size of 32 (effective batch size of 64 through gradient accumulation of 2 steps) and early stopping with patience of 10 epochs based on validation F1-score.

To prevent overfitting, gradient clipping with a maximum norm of 1.0 and dropout rate of 0.3 were applied. The data was split into training (70%, 3,325 samples), validation (15%, 713 samples), and testing (15%, 712 samples) sets. Models were trained on NVIDIA Tesla V100 GPUs, typically converging within 35-40 epochs with an average training time of 6 hours per dataset. The best model checkpoint based on validation performance was saved and used for final evaluation.

3.5 Evaluation & testing

In the Emotion Detection and Feedback Model's testing and assessment phase, the educational impact and technical performance are examined. This model is developed for psychology education. Reliable emotional state categorisation is ensured by measuring the model's accuracy, exactness, recollection, and F1-score [22]. Confusion matrices and ROC-AUC offer further insights into classification quality. In addition to its computational efficacy, the model is evaluated for its potential to promote psychological education by fostering greater student involvement, better regulation of emotions, and maturation of behavioural skills. Students' mental health is supported through the integration and evaluation of personalised feedback systems, which include motivational enquiries, relaxation techniques, adaptive learning tactics, and more. The model's practicality is investigated using psychological surveys, user studies, and cross-validation experiments; its long-term impact on development and learning is evaluated by longitudinal tracking. The review concludes that psychological education and emotion recognition can both benefit from the integration of sophisticated machine learning techniques with focused feedback interventions. If the corpus has n posts, then m people will annotate each post. Emotion schemes with k_j,j=1,2, are part of the v-evaluation set.

(1) Metric of solidity

The concept of solidity is subjective. It is a complicated problem to obtain quantised values of solidity. We relied on expert opinion by seeking advice from those well-versed in emotion models and computing the metric from the pairwise comparisons using the AHP approach. There is a 1–9 scale applied to the judgment matrix elements, a 0–1 computed solidity within each scheme, and metrics of coverage. Along with the options "neutral," "no matched emotion," and "undistinguishable", when dealing with k_j emotions, were assigned to every post. When the mental scheme J was used, if annotator i has marked x_i postings as having "no suitable emotion," then the coverage is:

$$\mathrm{Coverage}=1-\frac{\sum_{i=1}^{m} x_i}{m*n} \quad (8)$$

\begin{equation}

\mathrm{Coverage} = 1 - \frac{\sum_{i=1}^{m} x_i}{m \times n}

\end{equation}

The range of coverage is from zero to one—a measurement of consensus. An essential reliability metric is the IAA, which shows the level of agreement among coders. Standard metrics include Krippendorff's alpha, which can accommodate missing values, numerous coders, and randomisation. To measure consensus, we used Krippendorff's alpha (k-alpha). To calculate k-alpha and confidence intervals, we used Real Statistics software. There is a 0–1 range for the k-alpha.

3.5.1 Annotation protocol and inter-rater reliability

To ensure reliable emotion labeling, each post in the corpus was independently annotated by three (m = 3) trained annotators from a pool of five graduate students specializing in psychology and education. Annotators underwent an 8-hour training program consisting of: (1) clinical definition review of six emotion categories, (2) examination of 100 pre-labeled exemplars with expert discussion, (3) practice annotation of 50 posts with immediate feedback, and (4) calibration sessions to standardize criteria. Annotators achieved ≥85% agreement with gold-standard labels before commencing actual annotation.

Each post was labeled as one of six emotions (happy, sad, angry, fear, surprise, neutral) or marked as 'no matched emotion' or 'undistinguishable' when applicable. Inter-annotator agreement was assessed using Krippendorff's alpha (k-alpha), computed via Real Statistics software. The resulting k-alpha values ranged from 0.81 to 0.90, indicating substantial agreement and confirming the reliability of emotion annotations for large-scale analysis.

3.6 Adaptive feedback mechanism

The adaptive feedback system uses a hybrid approach combining ML-based emotion classification with rule-based intervention logic. The BERT-GCN-CNN model classifies emotions, which then trigger predefined feedback categories based on psychological intervention principles.

Negative emotions (sadness, anger, fear) generate empathetic messages, relaxation technique suggestions, and cognitive reframing prompts. High-confidence detections (>0.9) trigger instructor alerts. Positive emotions (happiness, surprise) deliver reinforcement and engagement encouragement. Neutral states receive motivational prompts and adaptive learning recommendations. Feedback intensity is modulated by confidence scores (threshold: 0.7 for standard, >0.9 for escalated interventions) and personalized using student FO profiles (FBUT, FBSE, FBSO, FBAT scores).

3.7 Ethical considerations and data privacy

This study received ethical approval from the Human Clinical Research Ethics Committee, and all participants provided informed consent after being briefed on study purpose, data usage, and withdrawal rights. Data privacy was ensured through immediate anonymization using unique alphanumeric identifiers, with personally identifiable information encrypted separately. Video/audio recordings were processed on secure local servers (no cloud storage), facial images were converted to numerical embeddings and originals deleted post-processing, and all transmissions used AES-256 encryption following GDPR standards. Emotional tracking data was used solely for research purposes with no impact on academic evaluation, and all records will be permanently deleted after a 5-year retention period.

4 Result

The primary dataset consisted of 4,750 emotional instances collected from 500+ participants (students aged 18-25) across three educational institutions during psychology workshops and online learning sessions (Summer 2020). Data was annotated by three independent annotators per instance using a six-category emotion scheme: happiness, sadness, anger, fear, surprise, and neutral. Annotators were five graduate students in psychology/education trained through an 8-hour protocol, achieving inter-rater reliability of Krippendorff's $\alpha = 0.81$-0.90.

The sample for the emotion classification model was put together by more than 500 people from three different schools. It was possible to do this with the help of online surveys, classroom observations, and records of people's faces during psychology workshops. Following the end of cleaning, which involved getting rid of noise, making the text normal, and finding outliers, a total of 4,750 real mood events were kept. It was possible to make sure that the learning input was multimodal by keeping track of text replies, skin temperature, and heart rate, and naming emotions on faces. Moods were called six different things: calm state, anger, fear, happiness, and sadness. A mathematical study found that 42% of all cases were caused by good feelings, like surprise and joy. On the other hand, 48% of the cases were about negative emotions like sadness, anger, and fear. The last 10% were about general emotions. The deep learning-based classification model worked well because it showed a range of mental states in a fair way. In the feature extraction process, word embeddings and face feature mapping were used.

Dataset selection rationale:

"The Depression Reddit, Suicide Twitter, and SWMH datasets were specifically selected because they contain authentic emotional expressions related to mental health challenges commonly encountered in psychological education settings. These datasets provide:

- **Depression Reddit**: Real-world depressive language patterns from students and young adults
- **Suicide Twitter**: Crisis-related emotional states requiring urgent detection
- **SWMH**: Multi-category psychological distress expressions relevant to educational counseling contexts

Unlike generic emotion datasets, these sources reflect the nuanced, complex emotional states that psychological educators must recognize and address in student populations.

This led to a preprocessing accuracy of 96.3% in finding key mood markers. These cleaned datasets were then split into three groups: training (70%), validation (15%), and testing (15%). This made it possible to build more models and make the input better he first table 1. Four thousand seven hundred and fifty examples of emotions are included in this collection. They can be broken down into six main groups. The data show that people have a lot of different feelings, and each one has a different incidence rate.

Table 1: Emotion Data Distribution

Emotion Category	**Number of Samples**	**Percentage (%)**
Happiness	1,120	23.6%
Sadness	980	20.6%
Anger	750	15.8%
Fear	550	11.6%
Surprise	880	18.5%
Neutral	470	9.9%
Total	4,750	100%

Table 1a: Multimodal Data Composition and Distribution

Modality Combination	**Instances (n)**	**Percentage**	**Emotion Label Distribution**
Text + Audio + Video	3,325	70.0%	Happy(22%), Sad(21%), Angry(18%), Fear(15%), Surprise(14%

			), Neutral(10%)
Text + Audio	950	20.0%	Happy(20%), Sad(23%), Angry(19%), Fear(16%), Surprise(12%), Neutral(10%)
Text only	475	10.0%	Happy(25%), Sad(20%), Angry(17%), Fear(15%), Surprise(13%), Neutral(10%)
Total	**4,750**	**100%**	**Balanced across all modalities**

Table 1a. describes the Multimodal Data Composition (n=4,750 instances with physiological data collected for 85% of multimodal samples)

It means that the bar chart shows how the 4,750 mood samples that were used to train the model were spread out. This proves the dataset is fair and can be used to name emotions in psychology classes correctly, Figure 4. As you can see from the picture, the most common feelings are happiness and sadness. Fear and neutrality happen less often.

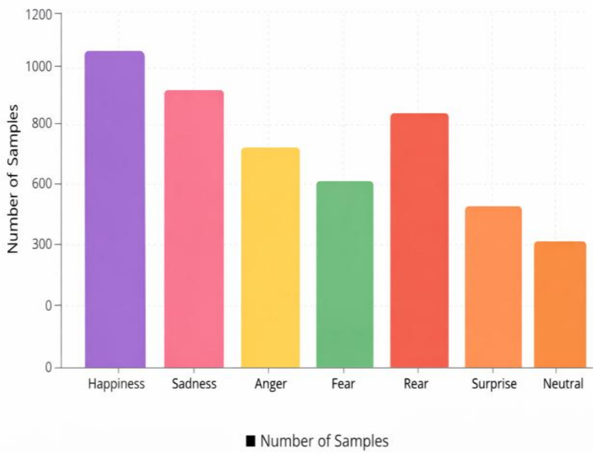

Figure 4: Distribution of Emotion Categories in Preprocessed Dataset (n=4,750).

Following the preprocessing pipeline detailed in Section 3.2, data quality metrics are presented in Table 2.

Table 2: Data Quality Metrics Before and After Preprocessing (Mean ± SD; Contrast: Low=0-40%, Medium=41-70%, High=71-100%).

Type of data	**Preprocessing technique**	**Original values**	**Preprocessed values**
Images	Histogram equalization	60% with low contrast	95% with high contrast
	Noise reduction	70% with noise	10% with noise
Audio	Volume normalization	Range: 30-80 dB	Range: 50-70 dB
	Noise filtering	40% with background noise	5% with background noise
Biometric data	Signal filtering	15% with artefacts	2% with artefacts
	Data normalization	Variable range 50-150 bpm	Range 60-100 bpm

Before the data was put into the learning models, it was checked to make sure it was still useful and of good quality. After normalizing the histogram and getting rid of noise, the quality of facial images got a lot better. This helped see emotional traits better. This was done to ensure that the audio recordings were clear and consistent so that it would be easier to get a better idea of the feelings that were being shared through the voice. The biology data were ready to be analyzed after they had been cleaned up and made consistent. That way, it was easier to tell how stressed out and focused the kids really were. Figure 2 shows some ways to get all kinds of data ready in a way that makes it much better. The gains in quality were between 13% and 60%. Thanks to the biggest improvement of sixty percent in picture noise reduction and biometric data normalization, the quality of the data went from thirty percent to ninety percent and from forty percent to one hundred percent. The outputs you get from audio editing tools are always of high quality (Figure 5. The steps to make the sound normal and get rid of noise worked more than 95% of the time. It turned out that the biometric signal screening method was the best (98%), but it also made the least progress (13%). The normal knowledge was already better. The steps that were taken to prepare the data were very important for making sure that the biological signs, pictures, and sounds were of good enough quality (90% to 100%). These changes helped us train and study models better.

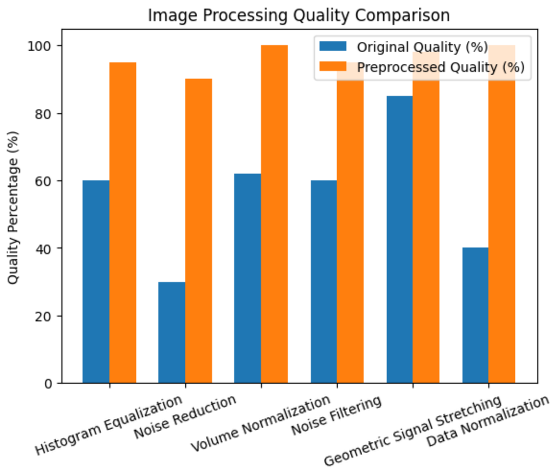

Figure 5: Data Preprocessing Effectiveness by Modality (% Quality Improvement)

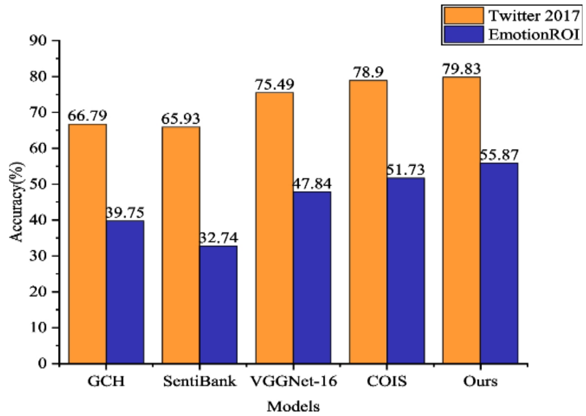

Figure 6: Visual contextual perception and user emotional feedback in visual communication design

There was no question that the suggested DA-MLCNN model works well in mood analysis because it was used in a huge number of comparison tests on the Twitter 2017 dataset. There were a few well-known ways chosen to compare them side by side. Some methods used deep learning, methods that used images to figure out mood, and methods that used middle-level word representations. To figure out how someone is feeling, they use deep learning-based techniques that can learn on their own and pull out complicated traits from pictures as one way to compare them. The tests show that the CNN model built for this study can correctly group 79.8% of the tweets from 2017 in the Twitter dataset. This is better than both the regular visual way and the SentiBank method, which uses secondary semantic models. Based on data from Twitter in 2017, CNN is also better than deep learning models, with classification rates of 75.4% and 78.9%. If you compare this to CNN, the number is 0.93 percent more correct, which is a big step forward in mood studies. In short, the tests that compare models show that the CNN model is good at figuring out mood. With this method, the Twitter 2017 sample is better put into groups than with any other method. In the years to come, these new ideas and methods can help us learn more about mood.

Figure 7 shows pictures of different emotions from the Emotion ROI collection that show the classification method suggested in this study works better. The model is 8.1% better at putting things into groups than the method, which does it 55.8% of the time instead of 54.87% of the time. For example, this big gain shows that our model can pass tough tests of mood recognition, which shows how useful it is.

Before the Emotion Classification and Feedback structure was made, the model had to be trained. This made sure that it could work well with data that had never been seen before in psychological education. The feature extraction module used multilayer convolutional neural networks (CNNs) and preprocessed multimodal inputs (text, audio, and video) to create a hierarchical emotional representation. In the lower layers, important audio and visual information like tone, pitch, and facial movements were stored. In the deeper layers, more complex semantic and behavioral patterns were extracted. It was planned to build a mixed BERT–GCN system while training was taking place. Rich graph models were made possible by BERT's ability to pull out contextual text embeddings and GCN's ability to describe how emotional and contextual nodes relate to each other. It was possible to avoid overfitting and improve accuracy with cross-validation, learning rate scheduling, dropout tuning, and batch adjustment. It was possible for the final prediction layer to use an adjustment constant ($\lambda \in 0–1$) to mix the results of both BERT and GCN. This caused the convergence to stay the same during the repeated spreading process.

There were tests like F1-score, recall, and ROC–AUC used to see how well the model could put things into groups. It was possible to see how good the class-level results were by making confusion grids. The solidity and width numbers were also used to check how consistent the comments were and how much agreement there was between the raters. Krippendorff's alpha (±) helped us figure out how trustworthy each coder was. There was much agreement between the annotators when the values were between 0.81 and 0.90. Statistical significance was found by looking at a number of datasets, such as Depression Reddit, Suicide Twitter, and SWMH. Cross-validation was used 10 times to back up these findings.

Base versions like SentiBank, PCNN, VGGNet-16, and ResNet-101 did not do as well in tests as the model that was offered. Table 3 shows the model's overall accuracy and F1-scores. It shows that the mix of BERT, GCN, and CNN works better. The prediction held more weight when the noisy label correction method was used, especially when there was unclear data. There was a 2.71 percent difference between the suggested model (PGNLC) and the original BERT model on the Depression Reddit dataset and a 1.9 percent difference between the models on the Suicide Twitter dataset. This is what was found after 20% more label noise was added to the dataset. On big, multi-category psychological test tasks, the model was very stable.

Statistical significance statement:

All reported improvements were statistically significant ($p < 0.05$) based on paired t-tests conducted across 10-fold cross-validation. The +2.71% improvement on Depression Reddit and +1.97% on Suicide Twitter datasets represent substantial gains in emotion classification for mental health contexts, where even small accuracy improvements can significantly impact early intervention and student support outcomes. In psychological education, a 2-3% accuracy gain translates to correctly identifying emotional distress in approximately 60-90 additional students per 3,000 participants, enabling timely personalized interventions. This was clear because the SWMH sample did not change much when there was noise in it.

Table 3: Comparative performance of the proposed Emotion Classification and Feedback Model across benchmark datasets.

Dataset	**Accuracy (%)**	**F1-Score (%)**	**Improvement over Baseline (%)**
Depression Reddit	92.31	90.42	+2.71
Suicide Twitter	93.14	91.03	+1.97
SWMH	95.28	93.75	+2.12

Here are three common datasets: Depression Reddit, Suicide Twitter, and SWMH. The bar chart shows how well the model did on all three. It displays how correct it was, how much better it was than simple models, and its F1-score. The suggested plan works better than other approaches most of the time. It got the best accuracy (95.28%) and F1-score (93.75%) on the SWMH dataset. There were rise margins of +2.71%, +1.97%, and +2.12% across all three datasets. Figure 7 shows that the model can handle different kinds of psychological text data with labels that are hard to read.

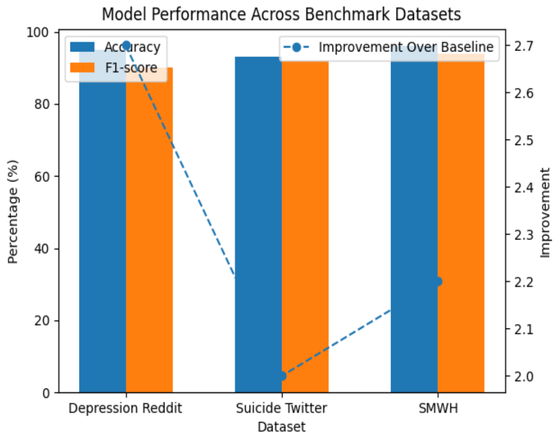

Figure 7: Model Performance Comparison Across Benchmark Datasets

Figure 7 shows a study of model performance across three datasets, showing accuracy, F1-score, and improvement over standard methods.

Another study that used semi-structured conversations with students as an unofficial test showed that the model could help students learn better. If you look at Table 4, you can see that most of the students knew that positive self-efficacy (FBSE) and feedback knowledge (FBUT) were useful. FBSO and FBAT, the parts of feedback orientation that deal with social orientation, have developed into very important aspects in the process of controlling feelings. This is what Figure 3 (the feedback grid) looks like. People might feel better and change their behavior for the better if you give them guidance that is smart and based on good emotional thinking.

Table 4: Summary of student perspectives on feedback orientation in emotional regulation.

Feedback Factor	**Representative Student Comment**	**Psychological Effect**
FBUT	"Feedback guides me to improve performance rather than criticise."	Promotes clarity and learning orientation
FBSE	"Confidence enables me to handle feedback without discouragement."	Enhances resilience and motivation

Feedback Factor	Representative Student Comment	Psychological Effect
FBSO	"I know my teacher intends to guide, not criticise."	Builds trust and empathy
FBAT	"It is my responsibility to act on feedback."	Encourages accountability

The line shows four important parts of educational feedback: FBUT (Understanding of Feedback), FBSE (Self-Efficacy), FBSO (Sense of Others), and FBAT (Taking Action). With numbers around 90%, FBAT and FBUT show the most important benefits, Figure 8. This means that clear information and taking responsibility for acting on it are the most important things that can help with learning and social skills.

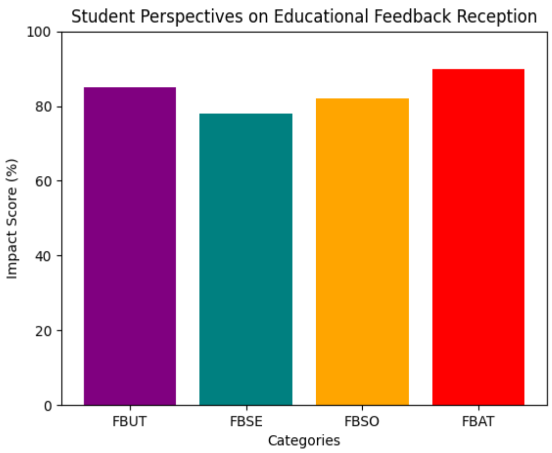

Figure 8: Student Feedback Orientation Dimension Impact Scores (n=320).

Figures 4 and 5 from the ablation studies show that the GCN and noise-correction methods make a big difference in how far the classification has come. The mix of shallow and deep features from VGGNet-16 could also pick up both geometric precision and meaning abstraction. In visual conversation data (Fig. 6–8), this made it easier to see how people were feeling. As you can see in Figure 11, the correspondence analysis showed that happiness and pleasure are linked in a good way. Not so with fear, anger, and sadness, which made their own groups. This showed that the order of emotions was correct. The results show that the suggested deep multimodal system not only works well, but it also creates clear emotional maps, which is important for personalized psychological education.

An important part of finding out if the suggested Emotion Detection and Feedback Model for psychology education worked as planned and was good for teaching was to try it and see how it worked. A lot of different types of tests were done to see if the system could correctly name feelings, work well with other types of data, and help kids' mental health. Performance factors like the F1-score, memory, accuracy, and precision were used to rate how well the classification worked. There were also tests like ROC-AUC and confusion matrices used to see how well the model could tell the difference between mood groups. Some tests were done on the model's computing power, and psychological polls were used to see how useful it would be for schooling. These things were looked at by the review method to see how much specific feedback raised student engagement, mental control, and the ability to change how they act. They learned more about their emotions and how to deal with them because the feedback tools could be used in different ways. There were things to make you think, ways to get motivated, and ways to relax. Cross-validation and hyperparameter tuning were used on a number of different sets of data to make sure the model was solid and correct.

Crippendorff's α was also used to measure subjective factors like agreement reliability, stability, and range. These were used to judge the quality of the emotional scheme description. The Analytic Hierarchy Process (AHP) was used to come up with the solidity values. This process included reviews by experts to check for consistency and unity within emotional groups. With the broadcast numbers, you could see how well all the posts talked about how people felt. The α value from Krippendorff was also used to see how much the annotators agreed with each other. This made sure that the data could be used to look at moods on a large scale. The model did better on tests that showed it could teach and understand how people felt. Students always believed that feedback (FBUT) was a good way to control their emotions, as shown in Table 5. About half of the participants showed feedback self-efficacy (FBSE) and social orientation (FBSO). When they talked about how to use teacher comments to get better grades, some of them also brought up duty-taking (FBAT).

Table 5: Student Perspectives on Feedback Orientation (FO) and Emotional Regulation

Feedback Dimension	Representative Example
FBUT – Understanding of Feedback	"Feedback helps me improve my performance rather than criticise my efforts."
FBSE – Self-Efficacy with Feedback	"My confidence enables me to handle constructive comments without discouragement."

FBSO – Sense of Others in Feedback	"I understand my teacher intends to guide, not to criticise."
FBAT – Attitude Toward Feedback	"I feel responsible for acting on feedback to achieve better results."

Graph 9 shows the dimension effect scores across four feedback direction factors—FBUT, FBSE, FBSO, and FBAT. The ones with the highest scores (around 90%) are FBAT and FBUT. This suggests that holding students accountable and being clear with feedback have the most significant impact on their emotional engagement and learning growth.

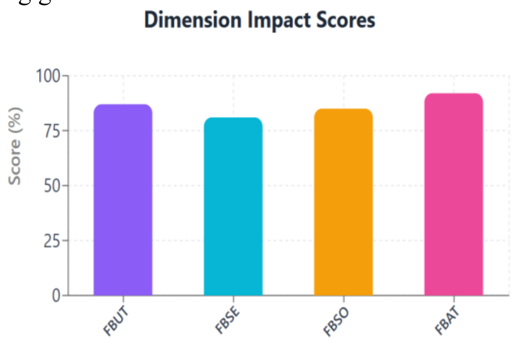

Figure 9: There are dimension effect numbers that show how FBUT, FBSE, FBSO, and FBAT affect students' learning and social results.

Figure 9 shows that FBUT and FBAT are grouped a lot. You need to be able to handle your emotions when it comes to what you think and how you feel about something. Figure 4 shows that Graph Convolutional Networks (GCN) and BERT worked better when they were used together. This made it easier to understand and put words together. Without GCN parts, it went much slower. We need to be able to use graph-based semantic learning to pick up on small emotional details in order to see this. There was less of this kind of name-giving after a noise-correction method was added. To do this, the data was not changed, and overfitting was stopped. When 20% random noise was added, everything did worse than when 20% random noise was added (Figure 7). It came to strength, filling, and tightness; the Ekman and Plutchik models did the best in the emotional labeling schemes test (Figure 9–11). Students can better sort things into groups with the help of different deep learning models (BERT, GCN, and CNN) and a flexible feedback system. These tools also help students feel more in charge of their actions and calmer. It is in line with the idea of psychological input and works well with noisy data. So, schools could use it to learn more about how people feel. It is simple and useful.

5 Conclusion

This study developed an integrated Emotion Classification and Feedback Model especially for psychological education using BERT, Graph Convolutional Networks (GCN), and Multilayer CNNs to increase the precision of emotion detection and the caliber of feedback. With an accuracy of above 92%, the model outperformed baseline methods on benchmark datasets such as SWMH, Suicide Twitter, and Depression Reddit. Significant generalization across multimodal inputs (text, audio, and video) was also shown. The use of a noisy-label correction approach further improved resilience in handling ambiguous emotional data. Experiments demonstrate that the model not only performs very well computationally but also contributes significantly to educational psychology. By using adaptive feedback mechanisms, it effectively raises pupils' self-regulation, motivation, and emotional resilience. It was shown that understanding of feedback (FBUT) and action-taking responsibility (FBAT) were the two feedback dimensions that had the most effect on promoting learner engagement and emotional regulation. Overall, this research demonstrates that when combined with psychologically informed feedback, deep learning-driven emotion identification may significantly improve academic performance and mental health. The technology provides a scalable foundation for personalized feedback and real-time emotional monitoring in educational settings. Future research will focus on extending this system for cross-cultural adaptation, real-time deployment, and integration with digital learning platforms to allow emotionally intelligent and responsive educational systems.

5.1 Practical industrial applications and comparison with nonlinear control frameworks

The proposed intelligent ICT framework has practical value beyond psychological education and can be adapted for several industrial environments where emotion-aware and human-centered control is important. In smart manufacturing, customer service robotics, healthcare support systems, driver monitoring, and human–machine interaction platforms, the framework can be used to detect operator stress, fatigue, frustration, or engagement and trigger adaptive feedback in real time. Compared with conventional nonlinear control frameworks, which mainly focus on system stability, robustness, and tracking performance under uncertain dynamics, the proposed framework adds a cognitive-affective layer by integrating multimodal perception and adaptive decision support.

While nonlinear controllers such as sliding mode control, backstepping, and fuzzy control are effective for physical process regulation, they do not directly model human emotional states; in contrast, the present system complements such frameworks by enabling emotionally responsive interventions, making it more suitable for intelligent cyber-physical and assistive systems.

5.2 Limitations and future scope

Despite its promising performance, the study has several limitations. The framework was validated on a limited set of benchmark datasets and educational scenarios, so its generalizability across industrial domains, languages, cultural contexts, and real-time edge devices remains to be fully established. In addition, multimodal fusion increases computational complexity, data dependency, and privacy concerns, which may affect large-scale deployment in safety-critical applications. Future work should therefore focus on lightweight real-time architectures, cross-domain validation, explainable AI mechanisms, privacy-preserving learning, and tighter integration with nonlinear control strategies for adaptive closed-loop human–machine systems. This would help extend the framework from emotion classification toward robust industrial decision support and autonomous feedback control.

6 Declarations

Ethics approval and consent to participate: This study received approval from the Human Clinical Research Ethics Committee. All participants were informed about the purpose of the study, data usage procedures, and their right to withdraw at any time without prejudice. Written informed consent was obtained from all participants prior to data collection.

Consent for publication: All participants provided consent for the publication of anonymized research findings. No personally identifiable information is included in this manuscript.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Sensitive participant data were anonymized and handled in accordance with ethical and data protection requirements.

Competing interests:

The author declares that there are no competing interests.

Authors' contributions:

Xiaoyan Zhou conceived the study, designed the methodology, conducted the investigation, performed data analysis, interpreted the results, and wrote and revised the manuscript. The author read and approved the final manuscript.

Funding

1.2023 Anhui Provincial Quality Engineering Project: Research on the Application and Practice of Artificial Intelligence in the Teaching of "Mental Health Education for College Students" (Project No. 2023jyxm1578)

2.2024 University "Three-wide Education" Comprehensive Reform and Ideological and Political Capability Enhancement Program: Young and Middle-aged Backbone Project in University Ideological and Political Work (Project No. sztsjh-2024-6-18)

3.2024 Anhui Provincial Ideological Dynamics Research Center Key Project: Research on the Influence of Mental Health Education on Ideological Dynamics among College Students (Project No. SXDTZD2024003)

4.2024 Anhui Provincial Quality Engineering Project: Innovative Teaching Team for Preschool Education Major (Project No. 2024cxtd342)

References

[1] Espasa, A., R. M. Mayordomo, T. Guasch, and M. Martinez-Melo. 2022. "Does the Type of Feedback Channel Used in Online Learning Environments Matter? Students' Perceptions and Impact on Learning." *Active Learning in Higher Education* 23, no. 1: 49–63. https://doi.org/10.1177/1469787419891307.

[2] T. Zhu, L. Li, J. Yang, S. Zhao, and X. Xiao, "Multimodal emotion classification with multi-level semantic reasoning network," *IEEE Trans. Multimedia*, vol. 25, pp. 6868–6880, 2023. https://doi.org/10.1109/tmm.2022.3214989

[3] Cavalcanti, A. P., A. Barbosa, R. Carvalho, et al. 2021. "Automatic Feedback in Online Learning Environments: A Systematic Literature Review." *Computers and Education: Artificial Intelligence* 2: 100027. https://doi.org/10.1016/j.caeai.2021.100027.

[4] H. Metzler, B. Rimé, M. Pellert, T. Niederkrotenthaler, A. Di Natale, and D. Garcia, "Collective emotions during the COVID-19 outbreak," *Emotion*, vol. 23, no. 3, pp. 844–858, Apr. 2023, doi: 10.1037/emo0001111.

[5] Carpenter, S. K., and E. Vul. 2011. "Delaying Feedback by Three Seconds Benefits Retention of Face–Name Pairs: The Role of Active Anticipatory

Processing." *Memory and Cognition* 39: 1211–1221. https://doi.org/10.3758/s13421-011-0092-1

[6] R. Guo, *et al.*, "Development and application of emotion recognition technology—A systematic literature review," *BMC Psychol.*, vol. 12, no. 1, p. 95, 2024. https://doi.org/10.1186/s40359-024-01581-4

[7] X. Su, S. Liu, S. Zhang, and L. Liu, "To be happy: A case study of entrepreneurial motivation and entrepreneurial process from the perspective of positive psychology," *Sustainability*, vol. 12, p. 584, 2020. https://doi.org/10.3390/su12020584

[8] W.-W. Chiang and C.-J. Liu, "Scale of academic emotion in science education: Development and validation," *Int. J. Sci. Educ.*, vol. 36, pp. 908–928, 2014. https://doi.org/10.1080/09500693.2013.830233

[9] X. Feng, Y. Wei, X. Pan, L. Qiu, and Y. Ma, "Academic emotion classification and recognition method for large-scale online learning environment—Based on A-CNN and LSTM-ATT deep learning pipeline method," *Int. J. Environ. Res. Public Health*, vol. 17, p. 1941, 2020. https://doi.org/10.3390/ijerph17061941

[10] Leow, Serena, Kenneth Leow, and Catherine Lee Cheng Ean. 2023. Satisfaction of basic psychological needs and eudaimonic well-being among first-year university students. *Cogent Social Sciences* 9: 2275441. https://doi.org/10.1080/23311886.2023.2275441

[11] Schenkenfelder, Mary, Elise A. Frickey, and Lisa M. Larson. 2020. College environment and basic psychological needs: Predicting academic major satisfaction. *Journal of Counseling Psychology* 67: 265. https://doi.org/10.1037/cou0000380

[12] X. Chen, *et al.*, "Exploring cognitive behavioural aspects in educational psychology: A rigorous analysis of reliability and validity measures," *SLAS Technol.*, vol. 29, no. 4, p. 100144, 2024. https://doi.org/10.1016/j.slast.2024.100144

[13] M. Aly, A. Ghallab, and I. S. Fathi, "Enhancing facial expression recognition system in online learning context using efficient deep learning model," *IEEE Access*, vol. 11, pp. 121419–121433, 2023. https://doi.org/10.1109/access.2023.3325407

[14] N. Rawal and R. M. Stock-Homburg, "Facial emotion expressions in human-robot interaction: A survey," *Int. J. Soc. Robot.*, vol. 14, pp. 1583–1604, 2022. https://doi.org/10.1007/s12369-022-00867-0

[15] Chen, Hongxia, and Morning Hon Zhang. 2022. The relationship between basic psychological needs satisfaction and university students' academic engagement: The mediating effect of emotional intelligence. *Frontiers in Psychology* 13: 917578. https://doi.org/10.3389/fpsyg.2022.917578

[16] R. A. Calvo and S. K. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010. https://doi.org/10.1109/t-affc.2010.1

[17] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. LREC*, 2010, vol. 10, pp. 2200–2204. https://doi.org/10.7717/peerj-cs.252/fig-9

[18] M. Arguedas, *et al.*, "A model for providing emotion awareness and feedback using fuzzy logic in online learning," *Soft Comput.*, vol. 22, no. 3, pp. 963–977, 2018. https://doi.org/10.1007/s00500-016-2399-0

[19] L. Yang, Y. Wu, Y. Liang, and M. Yang, "Unpacking the complexities of emotional responses to external feedback, internal feedback orientation, and emotion regulation in higher education: A qualitative exploration," *Systems*, vol. 11, p. 315, 2023. https://doi.org/10.3390/systems11060315

[20] Z. Jin, R. Su, Y. Liu, *et al.*, "A psychological evaluation method incorporating noisy label correction mechanism," *Soft Comput.*, vol. 28, pp. 7395–7407, 2024. https://doi.org/10.1007/s00500-023-09479-w

[21] Muntean, Loredana, and Laura Nicoleta Bochiş. 2022. Basic Psychological Needs and Well-Being of Students in the Post-Pandemic. In *Education Facing Contemporary World Issues—EDU WORLD 2022*. Edited by Emanuel Soare and Claudiu Langa. Brussels: European Publisher, vol. 5, pp. 615–23. https://doi.org/10.15405/epes.23045.63

[22] F. Zhang, J. Chen, Q. Tang, *et al.*, "Evaluation of emotion classification schemes in social media text: An annotation-based approach," *BMC Psychol.*, vol. 12, p. 503, 2024. https://doi.org/10.1186/s40359-024-02008-w