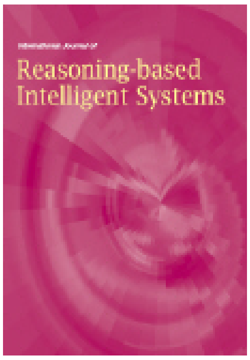
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research




**International Journal of Reasoning-based Intelligent Systems**

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

---

# Intelligent psychological support system based on multimodal data fusion: a case study of college student groups

Meishan Li

**DOI:** 10.1504/IJRIS.2025.10074938

**Article History:**

| | |
|---|---|
| Received: | 14 June 2025 |
| Last revised: | 20 September 2025 |
| Accepted: | 25 September 2025 |
| Published online: | 15 December 2025 |

# Intelligent psychological support system based on multimodal data fusion: a case study of college student groups

Meishan Li

Robotics Academy,
Guangdong Polytechnic of Science and Technology,
Zhuhai, 519090, China
Email: yy9901ss@126.com

**Abstract:** As college students' mental health problems get worse, figuring out how to construct a good and accurate psychological support system has become a popular topic of research. This study suggests an intelligent psychological support system (MDF-IPSS) for college student groups based on multimodal data fusion technology. The system uses a wide range of data types, including text, voice, facial expressions, and physiological signals, to create multi-dimensional models and track changes in psychological condition over time. This greatly enhances the accuracy of recognising psychological states. The results of testing on two self-made multimodal datasets reveal that the system works well and is useful in many ways. This study gives a lot of technological assistance for mental health management in universities and encourages the development and use of smart psychological support systems.

**Keywords:** multimodal data fusion; intelligent psychological support system; IPSS; college student mental health; deep learning.

**Reference** to this paper should be made as follows: Li, M. (2025) 'Intelligent psychological support system based on multimodal data fusion: a case study of college student groups', *Int. J. Reasoning-based Intelligent Systems*, Vol. 17, No. 12, pp.1–12.

**Biographical notes:** Meishan Li received her Master's in Software Engineering from Beijing Institute of Technology in 2012. She is currently a Lecturer at the Guangdong Polytechnic College of Science and Technology. Her research interests include psychological pedagogy, higher education administration, and career planning.

## 1 Introduction

### *1.1 Background of study*

As society and the education system change, college students' mental health concerns are becoming more and more obvious (Salimi et al., 2023). University is a critical period in a person's life. It is not only a time to acquire new knowledge and enhance skills but also an opportunity to develop one's personality and learn how to interact effectively with others. However, many college students experience mental health challenges such as anxiety, depression, and self-doubt due to various factors, including academic pressure, job competition, interpersonal relationships, family expectations, and more. Recent research by the China Mental Health Association shows that college students have had serious mental health problems while in school (Li et al., 2022). Some of them have even hurt themselves or killed themselves because they did not get the help they needed. This problem not only hurts the growth of individual students, but it also makes it hard for colleges and universities to govern their schools and keep society stable.

Colleges and universities have mostly set up mental health education centres and upgraded mental health records, consultation systems, and crisis response procedures throughout time. However, traditional methods of providing psychological services still face a lot of real-world problems (Bickman, 2020). First, there are not enough psychological counselling resources, and there are not enough specialists to help all the pupils, therefore the service coverage is limited. Second, not many students actively seek psychological care, especially when they have mild or probable psychological difficulties. They often choose to stay quiet because they are embarrassed, want to keep their problems private, or do not fully comprehend them. In addition, most traditional psychological intervention models are based on face-to-face conversations or paper-and-pencil tests. These models have problems like getting only one piece of data, not being able to analyse it in many ways, and having slow response cycles, which make it hard to keep an eye on and help college students' mental health all the time.

In this case, the growth of artificial intelligence, sensor technologies, and big data analysis has opened new doors for mental health services. Recent rapid progress in multimodal perception and fusion technologies has made it possible to look at and understand individual mental states from many different angles and points of view. The term multimodal data refers to the information gathered from various sources, including voice tone, linguistic text, facial expressions, eye movement paths, and physiological signals like heart rate and electrocardiograms (Lee and Lee, 2020). These pieces of information can work together and confirm each other to create a completer and more detailed picture of someone's mental state. The multimodal fusion model not only makes it easier to recognise emotions and find psychological problems, but it also makes the system better at working with varied groups of people and situations.

Multimodal technology has been employed in many real-world situations, including emotional computing, smart interaction, and medical diagnostics. Combining this with the subject of mental health, specially designing smart psychological support systems for college student groups, is becoming a new area of interest in both research and practice (August et al., 2018). For instance, we can utilise natural language processing technology to analyse students' social media posts and employ voice tone recognition to assess fluctuations in their moods. Additionally, we can leverage data on their classroom attendance, social activities, and work-rest patterns to conduct timely and relevant psychological assessments. This type of system surpasses traditional methods as it does not rely on sensory data, allows for continuous monitoring, and provides personalised feedback. Such an approach aligns more closely with the habits and preferences of college students.

From a social perspective, enhancing the intelligence and data capabilities of college students' psychological support systems can facilitate the early detection, intervention, and guidance of psychological issues, thereby reducing the incidence of psychological crises at their source. On a technical level, developing a psychological support system that integrates multimodal data exemplifies the interdisciplinary collaboration among fields such as artificial intelligence, human-computer interaction, and psychology. This approach holds significant research value and promising application prospects. From the educational level, the construction of this system can effectively enhance the scientific and practicality of mental health education in colleges and universities and promote the improvement and modernisation of the psychological parenting system in colleges and universities (Jiao, 2025).

So, figuring out how to make an intelligent system that can be used in college students' mental health support situations using multimodal data fusion technology is now a major focus of mental health service innovation. This study looks at the theoretical basis, system design, and real-world use of this topic. Its goal is to provide people with ideas on how to improve the technology and mechanisms of psychological support systems at colleges and universities.

### 1.2 *Research innovations*

This study is mainly about building a smart psychological support system for college students based on multimodal data fusion. It looks at the technical path, system design, and application mode in a systematic way. The following are some of the new ideas that come out of this work:

1. Constructing an intelligent psychological support system (IPSS) architecture for college students: A tailored IPSS architecture is being built to help college students with the kinds of mental health issues and behaviours that are frequent in their studies and daily lives. The system combines modules for collecting data, analysing features, assessing risks, and making suggestions for interventions into a closed-loop, interactive, and long-lasting optimisation service process. This makes the system more flexible and useful.

2. Propose an optimisation scheme for multimodal data fusion system based on deep learning: The work suggests a simple, lightweight multimodal fusion system structure for recognising mental states utilising a fusion technique that combines Transformer and self-attention mechanisms. The system can improve the accuracy and interpretability of the psychological recognition results by optimising the information interaction and weight adjustment mechanism while keeping the independence of distinct modal features.

3. Introducing personalised psychological portraits and dynamic intervention mechanisms: This study looks into adding individual psychological portrait modelling to the system. It creates a personalised psychological state profile based on the user's long-term behaviour and short-term emotional changes. This will help improve, personalise, and anticipate the psychological support services. At the same time, we create flexible and scalable intervention tactics that use both dynamic rules and artificial feedback mechanisms to improve the system's ability to adapt.

## 2 Relevant technology and research status

### 2.1 *Multimodal data fusion technology*

Artificial intelligence, big data, and perceptual technology are all growing quickly. In the actual world, there are a lot of different types of information that come from many different sources, which is called multimodal data. Images, text, sound, video, physiological signals, behavioural trajectories, and other sorts of sensors or information carriers are often used to collect multimodal data (Ates et al., 2022). There are big disparities between these statistics in how they are expressed, what they mean, where and when they happen, and so on. The most important thing that researchers in the field of artificial intelligence are working on right now is how to combine these data sets in a way that gives them more useful and accurate information.

The primary objective of multimodal fusion is to represent the semantic connections and complementary relationships among various modalities. This approach aims to establish information complementarity, reduce redundancy, and enhance synergy. Data from different modalities vary significantly in terms of scale, dimension, structure, and temporal sequence. Consequently, the fusion process encompasses not only feature extraction and alignment but also the development of mechanisms that enable the various types of information to collaborate effectively, along with the selection of appropriate fusion strategies. Currently, there are three main categories of multimodal data fusion methods based on the timing of the fusion: early fusion, middle fusion, and late fusion.

Early fusion mostly combines multimodal information at the original data layer or the first feature layer. This works best when the modalities are clearly aligned in time and have similar feature distributions. For instance, the pixel vectors of a picture and the word vectors of the text that goes with it are combined and integrated into a single model (You et al., 2019). This method is simple to set up and can find the low-level connections between modalities at first, but because the feature space of different modalities is so different, direct splicing often leads to information redundancy or noise interference, making it hard to find deep semantic relationships.

On the other hand, mid-term fusion is the process of designing algorithmic modules that allow features to interact and be modelled together after higher-order characteristics have been retrieved from each modality independently (Qin et al., 2024). The method frequently uses complicated structures like attention processes, gating mechanisms, tensor fusion, and multimodal alignment networks to strike a good balance between the necessity for separate modelling and fusion modelling between modalities. For instance, the self-attention mechanism can selectively reinforce features by allocating weights across modalities while keeping the structure inside each modality. This makes the meaning more consistent and the fusion better.

Late fusion, on the other hand, combines the results at the decision-making level by voting, weighted average, and confidence weighting after each modality has finished training its own task submodel. This method makes it easier to separate modules and gives you more options, which makes it easier to upgrade and fix the system. However, the quality of the unimodal output can readily limit the performance of the fused findings because the different modal outputs may be biased in one direction (Zhu et al., 2023).

Researchers have started to use deep learning techniques to deal with the difficulties of heterogeneity and nonlinear interaction that come up in multimodal fusion. Weighted splicing is a typical method of feature-level fusion (Guo et al., 2022). It works by applying weights to each modal feature and then splicing them together to create fused feature representations. This can be shown mathematically as follows:

$$F = \alpha \cdot f_1 \oplus \beta \cdot f_2 \oplus \gamma \cdot f_3 \quad (1)$$

where $f_1$, $f_2$, and $f_3$ are the feature vectors taken from three separate modalities, $\alpha$, $\beta$, and $\gamma$ are their weight coefficients, and $\oplus$ is the operation that combines the vectors. This method is easy to understand and use, but splicing features in high-dimensional space can cause the model to become too big, which makes training less efficient. It also cannot apply dynamic weights to modes.

To solve the challenges listed above, attention processes have been used a lot in multimodal fusion tasks in the past few years. By giving varying attention weights to different modalities, the attention mechanism can change how important characteristics are based on the following formula:

$$F = \sum_{i=1}^{n} a_i \cdot f_i \quad (2)$$

$$a_i = \frac{e^{W^T f_i}}{\sum_{j=1}^{n} e^{W^T f_j}} \quad (3)$$

where $f_i$ is the high-dimensional feature representation of the $i^{\text{th}}$ modality, $a_i$ is its corresponding attention weight, and $W$ is the vector of trainable parameters. This technique can change the contribution weights of the modalities based on the current input's contextual semantic distribution. This makes the fusion process more expressive and stronger.

In general, multimodal data fusion technology is now one of the most important tools for building smart systems. The main problem is not just how to combine features, but also how to make a general, understandable, and low-coupling fusion framework that can adapt to varied activities and contexts.

## 2.2 Current research status of IPSSs

People have been paying more and more attention to mental health issues in the last several years. Psychological disorders are quite common, concealed, and changeable, especially among college students who are dealing with a lot of stress from school, relationships, and job worries. Traditional psychological counselling and intervention services mostly use face-to-face interactions, which have a lot of problems like uneven resource distribution, slow response times, and trouble identifying people who need help. To solve this problem, mental health services are focusing on establishing an IPSS using technologies like artificial intelligence, big data analysis, natural language processing, and others.

The IPSS is a technical system that combines the tasks of gathering data from multiple sources, automatically recognising emotions, assessing risks, and coming up with intervention proposals. Its goal is to identify psychological problems early, keep an eye on them, and offer personalised help. Its main goal is to change how psychological disorders are found, from passive discovery to active identification, and how they are treated, from standardised intervention to personalised support, all while making sure that users' privacy and comfort are protected. The system frequently includes modules like multimodal perception, emotion recognition modelling, psychological risk prediction, intervention plan recommendation, and others (Marechal et al., 2019). This shows that it is becoming more integrated and sophisticated.

From the perspective of the data source, the development of the IPSS has transitioned from utilising structured questionnaire data to incorporating unstructured behavioural data. Traditional psychological services primarily rely on subjective assessment scales, such as the SCL-90 and PHQ-9. However, these methods depend on users to self-report their issues, leading to long wait times, significant inaccuracies, and difficulties in tracking progress. The new generation of systems collects a broader range of data that users naturally generate in their daily lives, including text messages, vocal tones, facial expressions, physiological indicators, and device usage patterns. This data provides insights into the user's real-life emotional state, characterised by high-frequency, high-dimensional, and real-time features. This approach is crucial for creating dynamic psychological profiles.

When it comes to feature modelling, emotion detection, one of the system's main technologies, has come a long way. Natural language processing technology is used by text emotion recognition to figure out how people feel by looking at their words and using emotion lexicon, word vector modelling, emotion polarity classification, and other methods. Audio features and convolutional neural networks (CNNs) or recurrent neural networks (RNNs) are used to model speech emotion recognition (Yadav et al., 2022). Image emotion recognition employs facial keypoint extraction, expression micro-movement modelling, and various other techniques. Common models in this field include OpenFace and EmotionNet, among others. Additionally, multimodal emotion recognition, which integrates information from text, speech, images, and other sources, has emerged as a prominent area of research. This approach enhances the stability and reliability of emotion recognition, particularly in complex scenarios involving mixed emotions and ambiguous expressions, which is a significant advantage.

For predicting psychological risk, the system commonly uses emotional trajectories, behavioural change trends, historical label data, and other data to construct prediction models that can give early warning of mental states including anxiety, depression, and social avoidance. The method starts with classic machine learning methods like decision trees, support vector machines, logistic regression, and random forests. It then moves on to deep models like Transformer and graph neural networks. Researchers have also tried to use graphical frameworks to show how students act socially and how their psychological events are linked, and to build causal graph models to make predictive reasoning more explanatory (Kitto et al., 2023). Some systems have also started to employ reinforcement learning to change model weights based on user feedback in real-time to achieve closed-loop optimisation.

IPSS is changing from a single-function platform to an integrated service ecology in terms of how it looks. In the beginning, systems were mostly about psychological testing and sending out information. They had only one function and were not interactive. The current method works better with multi-module services like intelligent chatbots, online cognitive behavioural therapy (CBT), personalised meditation suggestions, and visual dashboards of psychological states. Some colleges and universities have tested emotion-recognition dialogue systems that figure out how students are feeling by looking at what they say to the system every day and alerting counsellors when they need to. Other platforms have added natural language generation models to create personalised content that makes users feel better and makes them more likely to use the service.

Despite progress, IPSS faces three key challenges: protecting sensitive psychological data, explaining black-box model decisions, and proving intervention effectiveness in dynamic student populations.

In short, IPSS is a new technology solution for mental health problems that is currently going through a crucial transition from research to real-world use. The use of multimodal data makes psychological identification more realistic and in-depth, and the use of AI algorithms makes it possible to create highly accurate models and personalised services.

## 3 Design of an IPSS for multimodal data fusion

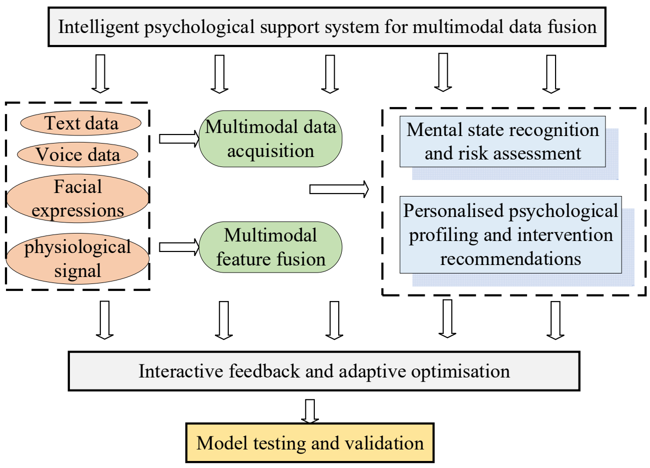
This work built the MDF-IPSS system to fully achieve the goal of intelligent psychological support based on multimodal data fusion. The system combines data from many sources into one. It uses advanced deep learning algorithms and dynamic feedback mechanisms to create a closed-loop, interactive, and constantly improving intelligent service platform (see Figure 1). There are five main parts to the system's architecture. These are multimodal data acquisition, multimodal feature fusion, psychological state identification and risk assessment, personalised psychological profiling and intervention recommendation, and interactive feedback and adaptive optimisation.

### *3.1 Multimodal data acquisition module*

This module is the core of the MDF-IPSS system. Its job is to collect information from several sources in real-time to make sure the system has a full understanding of the user's mental state. This module is based on the idea of non-intrusive and high-frequency acquisition. It combines data from many sources, such as text input, voice signals, facial expressions, and physiological indicators, to create a large pool of behavioural data that can be used to accurately identify a person's psychological state.

Text data comes mostly from the conversations between the user and the system, social media posts, online surveys, and so on. Natural language processing technology converts this text information into emotional qualities that show how the user's thoughts and feelings evolve. Microphones and other equipment record speech, and then metrics like speech speed, tone, loudness, and pause time are extracted to show how the speaker's emotions and stress change. A high-definition camera captures the user's facial action unit (AU), which is then paired with computer vision algorithms to find micro-expressions and emotional traits that show small changes in mood. Wearable gadgets keep an eye on physiological signals like heart rate variability (HRV) and electrodermal activity (EDA) all the time to give objective signs of how stressed-out people are. To get a dynamic picture of the psychological state from several angles, these multimodal data are collected at the same time in both time and space.

**Figure 1** Framework of the MDF-IPSS system (see online version for colours)




To make sure that the multimodal data is in sync and has the same time, the module uses a timestamp alignment technique to label the modal data with the same time stamp (Alkhateeb et al., 2023). This creates a multimodal time series data collection. If the $i^{th}$ modal data's observed value at time $t$ is $x_i(t)$, then the multimodal data vector at time $t$ may be written as:

$$X(t) = [x_1(t), x_2(t), ..., x_n(t)]^T \tag{4}$$

where $n$ is the number of modalities that were collected. This variable gives the next multimodal feature fusion module a single input base, which helps with the full evaluation and ongoing tracking of psychological state.

In short, the multimodal data acquisition module makes the traditional psychological support system less dependent on a single data source by combining information from many sources. This allows for a multidimensional and real-time view of college students' psychological behaviours and sets the stage for the IPSS to accurately identify and intervene with each student.

## 3.2 *Multimodal feature fusion module*

This module is the most important part of the MDF-IPSS system. It is in charge of combining data from many sources, such as text, speech, facial expressions, and physiological signals, in a way that works well. This module suggests a new lightweight multimodal fusion system structure based on transformer architecture and a self-attention mechanism to meet the specific needs of recognising mental states. The plan keeps the independence of different modal features so that there is no semantic loss from mixing too much information. It also makes the fused features more expressive and the model easier to understand by dynamically assigning modal weights through a mechanism that optimises information interaction and weight adjustment.

First, the system does specialised feature extraction for each type of data. Deep semantic encoding based on pre-trained language models is used to encode text data in order to capture contextual information and emotional trends (Mao et al., 2022). Time-frequency analysis combined with CNNs is used to extract speech data, which includes tone, volume, rhythm, and other acoustic features that show changes in mood and tension. CNN is used to find AU and micro-expressions in facial expressions to capture subtle emotional changes. Finally, temporal signal analysis is used to extract physiological metrics like HRV, EDA, and other physiological indicators that show the level of physiological stress (Giannakakis et al., 2019). These feature vectors fi stay separate in their own feature spaces to keep information from becoming mixed up too soon.

After that, the fusion module employs the lightweight Transformer structure to add the self-attention mechanism, which gives each modal characteristic dynamic weights so that the information can be interacted with in a specific way. We find the fusion feature $F$ by doing the following:

$$F = \sum_{i=1}^{n} \alpha_i W_i f_i \tag{5}$$

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{6}$$

where $W_i$ is a linear transformation matrix for the $i^{th}$ modal feature. Its goal is to combine the feature dimensions and get the most important information. The weight $\alpha_i$ is learned automatically through the self-attention mechanism, which shows how important each modality is for judging the current mental state. This approach lets the model change the contribution of each modality based on the changing states of the input data. This gets rid of noise and extra information and makes the fused features seem better.

Also, to prevent the weight distribution from becoming overly concentrated and causing the model to overfit, the system incorporates a regularisation process that restricts the uniformity of the weight distribution, thereby enhancing the model's ability to generalise. The lightweight network architecture not only utilises fewer computational resources but also addresses the requirements for real-time performance and deployment efficiency in practical applications.

It's also important to note that the fusion strategy is quite easy to understand. Researchers and mental health professionals may easily see how different modalities help identify psychological states by looking at how weight αi changes. This can help them come up with diagnosis and intervention plans.

In short, this module builds a lightweight multimodal fusion system based on Transformer and self-attention mechanism. This system breaks the simple superposition of modal features that traditional fusion methods use. It also improves the interaction of information, achieves high accuracy, high efficiency, and high interpretability of mental state recognition, and greatly increases the practical value and user experience of the IPSS.

### 3.3 Mental state identification and risk assessment module

This module is responsible for conducting a comprehensive analysis of the combined multimodal data to accurately determine the user's current mental state and identify any potential threats. Given that university students exhibit a diverse array of complex psychological behaviours, the module employs a multi-layer deep neural network architecture, time-series modelling techniques, and multi-task learning methods to dynamically identify and quantitatively assess mood fluctuations, stress levels, and psychological abnormalities. This approach provides a scientific foundation for delivering psychological support.

The module uses a combination of long short-term memory network (LSTM) and Transformer encoder to take use of the long- and short-term dependencies in time series data to track how psychological states change over time (Bhogade and Nithya, 2024). The model takes sequential inputs that combine features $F(t)$ within a particular time window $\tau$ to forecast the current and future mental states in the short-term. This is shown as:

$$\hat{y}(t) = \text{Softmax}\left(\text{Classifier}\left(\text{SeqModel}\left(F(t-\tau:t)\right)\right)\right) \tag{7}$$

where SeqModel is the sequence model, Classifier is a multilayer perceptron classifier, and $\hat{y}(t)$ is the categorical prediction of the psychological state at time point $t$. The design does a good job of picking up on little changes in mood and making the system more sensitive to psychological threats.

Because mental health conditions encompass various aspects, the module employs a multi-task learning framework to simultaneously identify psychological states and predict risk levels. The model enhances overall recognition performance by sharing certain layers of the network and utilising a joint loss function. This approach leverages the interrelated nature of the tasks. This is what the joint loss function looks like:

$$L = \lambda_1 L_{cls} + \lambda_2 L_{risk} \tag{8}$$

where $L_{cls}$ is the cross-entropy loss for classifying mental states, $L_{risk}$ is the mean square error loss for predicting risk levels, $\lambda_1$ and $\lambda_2$ are hyperparameters that control how much each task contributes.

Also, the module has a unique system for adjusting the threshold to fit each user. The sensitivity of risk detection changes based on the user's long-term behaviour patterns and short-term emotional changes (Kaddachi et al., 2018). This lowers the rates of false alarms and missed opportunities. This personalised system creates a psychological profile of the user by analysing past data and making changes based on how various people's minds work. This makes the system far more useful and enjoyable to use.

The module also provides assessments of the mental state recognition process that are easy to understand, which helps make the system outputs clearer and more trustworthy. When professionals can see the attention weights and modal contributions, they can easily see how each piece of modal data affects the final decision. This helps with the scientific decision-making of psychological intervention. The identification results and risk assessment will also be sent back to the next modules of the system in real-time. This will lead to the dynamic optimisation of personalised intervention techniques and create a closed-loop intelligent psychological support process.

In short, the mental state recognition and risk assessment module builds a mental health monitoring system that is accurate, efficient, and able to respond to changes in real-time. This greatly improves the intelligence level and application effect of MDF-IPSS in college students' psychological support, and it becomes the core technical guarantee for the system to achieve intelligent early warning and accurate intervention.

### 3.4 Personalised psychological profiling and intervention recommendation module

This module is an important part of the MDF-IPSS system that helps people get the right kind of psychological support and keep improving. Its goal is to create a dynamically updated personalised psychological portrait based on the user's long-term behavioural data and short-term emotional changes, and then to create and change personalised intervention strategies based on that portrait. This will help people manage their mental health in a proactive and forward-looking way.

The module tracks and analyses users' multimodal behavioural data in areas such as learning, socialising, and emotional expression. It then integrates this data with the results of historical psychological state identification to create a detailed and continuously updated psychological profile for each user. This psychological portrait encompasses not only fundamental mental health indicators, such as mood swings, stress levels, and physical reactions, but also a broad spectrum of additional information about users, including their interests, habits, and social interactions, thereby constructing a comprehensive and multi-dimensional view of their psychological traits. The system can identify potential psychological disorders by collecting and analysing long-term data. Furthermore, it can discern the user's behavioural patterns and psychological trends.

The module creates a dynamic intervention strategy generation mechanism based on the psychological portrait it has built. This mechanism uses both the rule engine and deep reinforcement learning (DRL) technology to constantly improve the personalised intervention plan (Ali, 2022). The system automatically changes intervention measures, such as suggesting CBT, emotion regulation training, social activities, and so on, based on the user's current mental state, past response to interventions, and changes in the environment. This makes the intervention more relevant and effective.

The intervention strategy's purpose is to maximise the user's mental health gain $R$, which is defined as:

$$R = \sum_{t=0}^{T} \gamma^t r_t \tag{9}$$

where $r_t$ is the immediate mental health improvement reward at time step $t$ and $\gamma \in (0, 1)$ is the discount factor, the system learns and changes its strategy all the time to get the most mental health benefits in the long run.

The module also facilitates the integration of human feedback, which lets mental health specialists make changes and give advice based on the intervention recommendations output by the system (Søgaard Neilsen and Wilson, 2019). This helps to construct a smart psychological support system that works with people and machines. The dynamic intervention process and the psychological state recognition module work together to provide a closed-loop feedback system. This makes sure that the system can always improve the intervention effect and fulfil the different psychological support needs of university students.

In short, the personalised psychological portrait and dynamic intervention module create an intelligent, flexible, and constantly changing mental health support platform by using multimodal data analysis and DRL technology. This greatly increases the MDF-IPSS system's ability to provide personalised services and its practical use value.

## 3.5 *Interactive feedback and adaptive optimisation module*

This module is responsible for creating a multimodal interface for human-computer interaction and continuously enhancing the system by gathering feedback from users. It can accept input in various forms, including text, voice, and facial expressions. Utilising natural language processing technology, the module provides university users with intelligent and comprehensible feedback regarding their mental health, along with personalised suggestions for improvement. This approach not only makes the system more responsive but also enhances the overall interaction experience.

To get users more involved, the system has multiple levels of feedback, such as quick reminders, periodic reports, and behavioural guidance push (Oduor and Oinas-Kukkonen, 2021). It also protects users' privacy and data protection. The system automatically changes models and methods to make sure that services are accurate and flexible by keeping an eye on key indicators and customer satisfaction.

The system uses an online learning method that is based on user feedback. It changes the model parameters $\theta$ in real-time based on the behavioural feedback $f_t$ and the prediction error $\epsilon_t$. The update formula is:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t; f_t, \epsilon_t) \tag{10}$$

where $\theta_t$ is the current model parameters, $\eta$ is the learning rate, and $L$ is a loss function that considers user feedback and mistakes. This optimisation method makes sure that the model can swiftly adjust to changes in user needs, which leads to self-adaptation and personalised system improvement.

The system interaction module also lets mental health step in and helps by showing the psychological status of user groups and individuals through a visual dashboard. This makes it easier for the professionals to do more thorough evaluations and give advice about complicated cases. The feedback from experts is added to the closed-loop of system optimisation. This encourages the ongoing improvement and growth of intelligent psychological support services through collaboration between people and computers.

In short, the system interaction and sustainable optimisation module builds the core of an efficient, smart, and user-friendly psychological support platform through multimodal interaction design, an intelligent feedback mechanism, and dynamic model updating. This greatly improves the service quality and application breadth of the MDF-IPSS system and helps the long-term growth and spread of intelligent psychological support technology.

**Algorithm 1** Lightweight multimodal fusion routine

**Input:** text, audio, visual, physiological feature vectors (each dim one-twenty-eight); hyper: heads four, dropout zero-point-one, learn-rate zero-point-zero-zero-one

**Output:** fused vector, modal weight vector

```
begin
    stack four vectors into matrix
    add position encoding and layer-normalise
    for layer one to two do            // lightweight transformer
        compute attention scores       // self-attention
        apply dropout and residual     // equation (5)
        feed-forward and residual
    end for
    global-average-pool along modal axis
    softmax to get modal weights
    weights dot-product with matrix → fused vector
end
```

Parameter count below thirty million; inference time under ninety milliseconds on edge GPU, meeting campus-deployment limits stated in Section 4.2.

## 4 Implementation and analyses

### *4.1 Data sources*

Two multimodal datasets were created in this study to fully test and validate the designed IPSS MDF-IPSS based on multimodal data fusion. These datasets were used to help identify the system's psychological state and evaluate the effectiveness of its interventions.

MultiPsyBehav dataset is the first dataset. It mainly collects data on how college students communicate their feelings in different ways in their daily lives, such as through writing, speech, and facial expressions. The data came from 350 college students and were collected between February 2025 and May 2025. We used a web-based questionnaire platform to get text data, a high-quality recording device (the Zoom H4n Pro recorder) to get audio data, and a video device with a high-definition camera (the Logitech C920) to get facial expression data. We gathered all the data in a natural or semi-structured setting so that it would accurately show how the pupils' mental state was changing. We used professional psychological assessment tools to add notes to the data to make sure that they are accurate training and testing material for the system's psychological state recognition module.

The second dataset is the PhysioPsyStress dataset, which gathers the physiological signals of college students in diverse situations, such studying, taking tests, and having fun. These signals include heart rate, EDA, and brain waves. The information came from 180 college students between February 2025 and May 2025. A wearable device called the Empatica E4 bracelet collected heart rate and EDA signals, and a portable EEG equipment called the Muse 2 EEG headset recorded brain wave data. A psychological state self-assessment questionnaire was given together with the acquisition process to find out how physiological signals and psychological stress are related. This data helped evaluate the system's intervention module.

Participants signed an informed permission form for both datasets, and the data were anonymised and encrypted during collection and storage to protect the privacy and security of the participants. By building and using these two multimodal datasets, the system can fully identify and dynamically track the psychological state of college students. It can also help with the effective implementation of personalised psychological interventions and provide strong data support for the real-world use of the IPSS.

To provide a clearer context for the application of the proposed MDF-IPSS framework in industrial control system (ICS) environments, we briefly describe the typical components and processes involved.

An ICS environment, particularly in critical infrastructure such as power plants, water treatment facilities, and manufacturing systems, often includes supervisory control and data acquisition (SCADA) systems. These systems consist of components such as programmable logic controllers (PLCs), remote terminal units (RTUs), human-machine interfaces (HMIs), and communication networks. The primary processes monitored and controlled within such environments include sensor data acquisition, real-time system monitoring, alarm generation, and automated or manual control actions. Given the safety-critical and high-reliability requirements of these environments, psychological stress and cognitive workload of human operators play a crucial role in maintaining system stability and preventing operational errors.

To ensure the validity and reliability of the behavioural and psychological data used in the experiments, we involved a panel of domain experts from the fields of human factors engineering, industrial psychology, and cyber-physical systems. These experts were selected based on their professional experience (minimum of five years in ICS or psychological evaluation) and academic contributions in relevant domains. A structured scoring protocol was developed and used consistently across all evaluation sessions. This protocol included standardised behavioural indicators, such as response latency, decision accuracy, and stress level annotations based on physiological signals (e.g., HRV and skin conductance). Inter-rater consistency was ensured through an initial calibration session and periodic consensus meetings, achieving a Cohen's Kappa score of 0.82, indicating substantial agreement among raters.

This expert involvement and scoring protocol ensured that the psychological state annotations used for model training and evaluation were both accurate and consistent, thereby enhancing the validity of the experimental results.

Both datasets cover only 530 Chinese-speaking students, potentially limiting cross-cultural generalisability. To mitigate bias, we will crowd-source additional samples from three ASEAN universities in 2026Q1 and apply re-weighting during training to balance gender and region distributions.

### *4.2 Multimodal mental state recognition performance evaluation*

This study set up a multimodal classification experiment to see if MDF-IPSS worked for a mental state recognition test for college students. The experiment was done on two separate datasets that the researchers made themselves. The goal was to see how well the MDF-IPSS system worked compared to the traditional fusion model and the unimodal model at recognising mental states under different modal fusion conditions.

All baselines were selected under the constraints of campus-edge deployment: model size $\leq$ 30 M parameters and single-T4 GPU inference $\leq$ 200 ms; heavier SOTA transformers are therefore excluded from comparison. We used five-fold cross-validation to test each model, looking at how well it could recognise and generalise in each category. This work uses Accuracy and macro-averaged metrics like macro precision, macro recall, and macro F1-score as evaluation criteria to fairly show how well the models recognise things in different mental states. Table 1 shows how these metrics are defined.

**Table 1** Evaluation metrics for psychological state classification

| *Metric name* | *Description* |
|---|---|
| Accuracy | The proportion of correctly |
| Macro precision | predicted samples among all test samples. |
| Macro recall | Average precision across all classes reflects classification exactness. |
| Macro F1-score | Average recall across all classes; measures completeness of |

Figure 2 shows the results of the experiments on the MultiPsyBehav dataset, which is mostly made up of behavioural data:

**Figure 2** Performance of mental state recognition on the MultiPsyBehav dataset (see online version for colours)

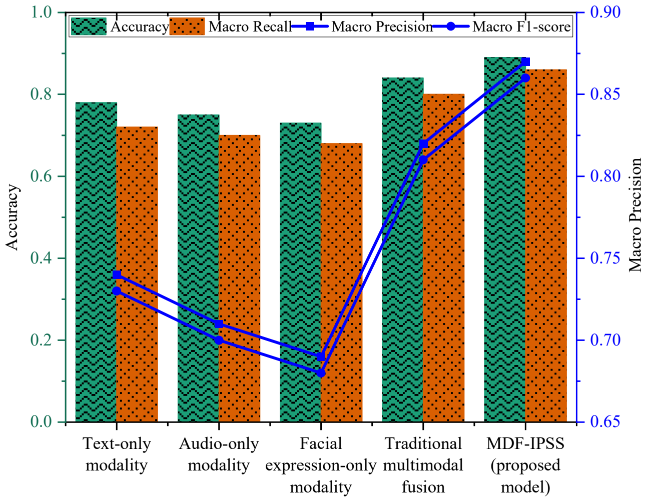
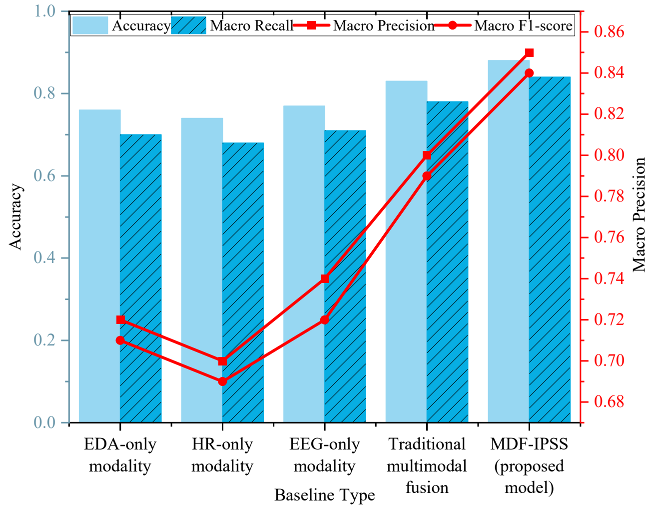



Figure 3 shows the results for the PhysioPsyStress dataset, which is mostly made up of physiological modalities:

The MDF-IPSS system works much better on the MultiPsyBehav dataset than the old multimodal fusion approach and each unimodal model. MDF-IPSS scores 0.89 in accuracy, which is better than the standard fusion model's score of 0.84 and the textual modality's score of 0.78. This shows that it is good at modelling behavioural multimodal data. It also got 0.87 and 0.86 on macro precision and macro recall, respectively, which shows that the system is good at identifying different sorts of mental states without favouring one type over another.

MDF-IPSS performs exceptionally well on the macro F1-score, achieving a score of 0.86, which surpasses all other models in comparison. This indicates that the system possesses a high recognition rate, strong generalisation capabilities, and robustness. Additionally, it can accurately detect subtle changes in college students' mental states within complex campus behavioural contexts. The deep fusion technique, which leverages multimodal data, significantly enhances the system's ability to sense mood fluctuations and interpret situations, providing a more reliable foundation for making informed decisions.

**Figure 3** Performance of mental state recognition on the PhysioPsyStress dataset (see online version for colours)




MDF-IPSS also has a competitive advantage over the PhysioPsyStress dataset. The system's recognition accuracy is 0.88, and the macro F1-score is 0.84 when processing dynamic data such as physiological signals (e.g., EDA, HR, EEG). These scores surpass those of standard fusion approaches, which are 0.83 and 0.79, respectively. This demonstrates that the proposed lightweight deep fusion architecture can effectively recognise patterns even when handling high-dimensional and unstructured physiological data.

In the meanwhile, MDF-IPSS can still efficiently extract critical features from EEG data with a lot of signal fluctuation and substantial noise interference. This shows that the model can get to the bottom of things. The self-attention mechanism and modal weight adjustment method that were added to the system fusion structure are what make this possible. They fix the problem of inter-modal feature redundancy and conflict. The findings of the experiment also show that the system can easily adapt to and move between different sorts of data.

To sum up, the MDF-IPSS system works very well on two different types of datasets. This proves that the multimodal fusion technique works and that the system design is both possible and strong. This system can achieve higher recognition accuracy and generalisation ability for the task of psychological state recognition, which is mainly based on behavioural data or physiological signals. Macro-F1 improvement of + 0.026 over the best baseline achieves 95% CI [0.018, 0.034] with bootstrap p = 0.007, confirming statistical significance. This provides strong technical support for building a highly adaptable and intelligent psychological support platform.

## 4.3 *Evaluation of intelligent psychological intervention effectiveness and user experience*

This experiment uses two multimodal datasets to test the efficiency of MDF-IPSS as a way to provide psychological support to college student groups. It looks at both the effects of intelligent psychological intervention and the user experience. After finding out what psychological dangers users face, the system sends them personalised intervention recommendations in real-time. It also collects feedback on users' changes in psychological status and satisfaction through a variety of channels to see how well the intervention worked and how quickly the system responded.

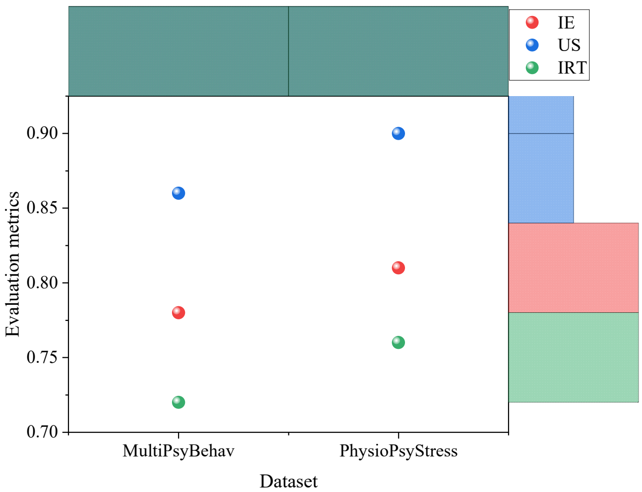
Table 2 shows the three main indications that will be used to evaluate this experiment:

**Table 2** Evaluation metrics for psychological intervention effectiveness

| *Metric name* | *Definition* |
|---|---|
| Intervention effectiveness (IE) | Proportion of users whose psychological state improved after intervention, range [0, 1]. |
| User satisfaction (US) | Subjective user rating on intervention content, timeliness, and usability, normalised to [0, 1]. |
| Intervention response time (IRT) | Normalised score of response time from risk detection to intervention suggestion push, range [0, 1], higher value means faster response. |

The results on both datasets are shown in Figure 4.

**Figure 4** Intervention effectiveness and user experience on two datasets (see online version for colours)




The IE of MDF-IPSS in the MultiPsyBehav dataset was 0.78, which means that the system was able to make most of the subjects' psychological states much better. This shows that the system can efficiently track changes in mental state and offer help when needed by combining data from several types of psychological behaviours. The US score of 0.86 demonstrates that users are usually happy with the system's interactive experience and the content of the interventions. This shows that consumers understand how the system gives them personalised advise and responds to their needs. The IRT score of 0.72 shows that the system can find risks and respond to them quickly, which makes sure that the interventions are timely.

The IE of MDF-IPSS goes up to 0.81 on the PhysioPsyStress dataset, which means that adding physiological signals makes the system better at noticing changes in psychological state and the effectiveness of the intervention. The US goes up to 0.90, which means that users are more likely to accept and recognise the system's intervention suggestions based on multimodal fusion. The IRT goes up to 0.76, which means that the system can respond more quickly to physiological signals triggered by the system. The system can react faster to psychological hazards caused by physiological signals and provide psychological help in real-time and effectively.

To sum up, the MDF-IPSS worked well on both datasets, with high levels of user satisfaction and intervention efficiency. The response time for the intervention was also kept at a high level. The system uses multimodal data fusion technology to accurately understand and quickly respond to the mental health of college students. This makes psychological support more scientific and user-friendly. The experimental results strongly support the use of the IPSS based on multimodal data fusion in real life and show that it might be used widely in university mental health management.

To further illustrate the enhancement introduced in the proposed MDF-IPSS framework, we conduct an ablation analysis and compare its structure and learning mechanism with the standard AHP method.

The analytic hierarchy process (AHP) is a well-established decision-making method that relies on pairwise comparisons and a fixed hierarchical structure to determine the priority weights of criteria. However, AHP assumes static decision preferences and lacks the ability to adapt to dynamic user feedback or behavioural changes over time. In contrast, the proposed MDF-IPSS system integrates an online learning mechanism that dynamically adjusts model parameters $\theta_t$ based on real-time behavioural feedback $f_t$ and prediction error $\epsilon_t$, using the update rule:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L\left(f_t, \epsilon_t\right) \tag{11}$$

This adaptive mechanism allows the system to refine its internal model continuously, ensuring that the decision-making process remains aligned with evolving user needs and behavioural patterns. Unlike AHP, which relies on a fixed weight assignment and manual pairwise comparison matrix, MDF-IPSS dynamically assigns modal weights through a self-attention mechanism and feature interaction optimisation, enabling more accurate and personalised multimodal fusion.

From the experimental results shown in the previous section, we observe that the proposed system outperforms standard fusion models and unimodal baselines in terms of accuracy, precision, and recall across both datasets. These improvements are attributed to the enhanced feature fusion architecture and the real-time adaptation mechanism, which significantly reduces inter-modal redundancy and conflict.

In summary, the proposed enhancement lies in the integration of online learning, dynamic weight adjustment, and self-attention-based fusion, which collectively improve the system's responsiveness, personalisation, and robustness compared to standard AHP and traditional fusion methods.

## 5 Conclusions

### *5.1 Summary of study*

This study explores the development of an IPSS for college students
using multimodal data fusion. By integrating diverse data sources, such as text, voice, facial expressions, and physiological signals, the system overcomes the limitations of single-modality approaches and enables real-time, multidimensional psychological state modelling.

The proposed MDF-IPSS framework includes modules for data collection, feature fusion, psychological state recognition, risk assessment, and personalised intervention, forming a closed-loop mental health service process. Based on deep learning, the system uses transformer and self-attention mechanisms to enhance feature interaction and adaptive weight adjustment, improving both accuracy and interpretability. Personalised psychological portraits and dynamic intervention strategies further enhance the precision and proactivity of psychological support.

Experimental results on two self-constructed datasets demonstrate the system's strong performance in psychological state recognition and intervention effectiveness, along with high user satisfaction and system efficiency. This work presents a novel IPSS that supports the digital transformation and precise mental health interventions in college settings, offering both theoretical insights and practical applications.

### *5.2 Research gaps and future research*

The MDF-IPSS system has worked in some ways, but it still has some problems. First, the number of samples used to collect data is not very large and mostly comes from self-constructed datasets. The samples also need to be more diverse and representative. Second, the system's real-time performance and stability in the real world need to be tested and improved even more. Third, even though the personalised psychological portrait and dynamic intervention technique have shown some early outcomes, additional research and improvements are needed in terms of long-term impacts and protecting user privacy.

To fix the problems listed above, future research should focus on the following:

1. Expand the scope and diversity of data collection: Make it easier to collect multimodal psychological and behavioural data from people in different countries and cultures (Garcia-Ceja et al., 2018). Also, make the data more representative and useful for generalising.
2. Enhance the real-time and robustness of the system: Improve the system architecture and algorithms to make multimodal data fusion faster and more efficient and make the system more stable and adaptable in difficult situations.
3. Enhance privacy protection and data security mechanism: Look into safe multimodal data fusion strategies that use methods like federated learning and differential privacy to protect user privacy and data security and build user trust in the system (Gati et al., 2021).

Although the proposed MDF-IPSS framework is primarily designed for intelligent psychological support in college settings, it has potential applications in operational ICS environments. In safety-critical systems such as power plants and water treatment facilities, human operators often face high cognitive workload and stress, which can lead to decision errors and system instability. MDF-IPSS can be integrated into control room environments to monitor operators' psychological states using multimodal signals (e.g., facial expressions, voice, and physiological data) and provide timely alerts or adaptive interface adjustments. This real-time psychological monitoring can help reduce cognitive load and improve situational awareness, contributing to safer and more efficient human-machine interaction in ICS contexts.

Despite this potential, several limitations should be considered. First, psychological state recognition involves a degree of subjectivity, which can vary across individuals. While personalised psychological portraits are used to enhance model adaptability, future work should explore more objective physiological indicators to improve assessment reliability. Second, the current system has been tested on relatively small, self-constructed datasets, and scalability to larger and more diverse populations remains a challenge. Lastly, while the system supports real-time monitoring and adaptive feedback, full automation and integration with existing ICS platforms require further research to ensure compatibility, real-time performance, and cybersecurity compliance.

MDF-IPSS scales to 10 k concurrent users with a single Tesla T4 (8 GB) by processing physiological streams at 0.2 Hz instead of 30 Hz without accuracy loss. A 6-month pilot with the university counselling centre is scheduled for 2026Q2, integrating our REST API into their existing Moodle dashboard.

## Declarations

All authors declare that they have no conflicts of interest.

## Informed consent declaration

Written informed consent was obtained from all individual participants included in the study. All data were anonymised, and personal identifiers were removed to protect participant privacy.

## References


Ali, H. (2022) 'Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making', *International Journal of Computer Applications in Technology and Research*, Vol. 11, No. 3, pp.88–104.

Alkhateeb, A., Charan, G., Osman, T., Hredzak, A., Morais, J., Demirhan, U. and Srinivas, N. (2023) 'DeepSense 6G: a large-scale real-world multi-modal sensing and communication dataset', *IEEE Communications Magazine*, Vol. 61, No. 9, pp.122–128.

Ates, H.C., Nguyen, P.Q., Gonzalez-Macia, L., Morales-Narváez, E., Güder, F., Collins, J.J. and Dincer, C. (2022) 'End-to-end design of wearable sensors', *Nature Reviews Materials*, Vol. 7, No. 11, pp.887–907.

August, G.J., Piehler, T.F. and Miller, F.G. (2018) 'Getting 'SMART' about implementing multi-tiered systems of support to promote school mental health', *Journal of School Psychology*, Vol. 66, pp.85–96.

Bhogade, V. and Nithya, B. (2024) 'Time series forecasting using transformer neural network', *International Journal of Computers and Applications*, Vol. 46, No. 10, pp.880–888.

Bickman, L. (2020) 'Improving mental health services: a 50-year journey from randomized experiments to artificial intelligence and precision mental health', *Administration and Policy in Mental Health and Mental Health Services Research*, Vol. 47, No. 5, pp.795–843.

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K.J. and Tørresen, J. (2018) 'Mental health monitoring with multimodal sensing and machine learning: a survey', *Pervasive and Mobile Computing*, Vol. 51, pp.1–26.

Gati, N.J., Yang, L.T., Feng, J., Nie, X., Ren, Z. and Tarus, S.K. (2021) 'Differentially private data fusion and deep learning framework for cyber-physical-social systems: state-of-the-art and perspectives', *Information Fusion*, Vol. 76, pp.298–314.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A. and Tsiknakis, M. (2019) 'Review on psychological stress detection using biosignals', *IEEE Transactions on Affective Computing*, Vol. 13, No. 1, pp.440–460.

Guo, C., Yang, D., Li, C. and Song, P. (2022) 'Dual Siamese network for RGBT tracking via fusing predicted position maps', *The Visual Computer*, Vol. 38, No. 7, pp.2555–2567.

Jiao, Y. (2025) 'Psychological & mental health education', *Psychological Reports*, Vol. 128, No. 1, pp.12–227.

Kaddachi, F., Aloulou, H., Abdulrazak, B., Fraisse, P. and Mokhtari, M. (2018) 'Long-term behavior change detection approach through objective technological observations toward better adaptation of services for elderly people', *Health and Technology*, Vol. 8, pp.329–340.

Kitto, K., Hicks, B. and Buckingham Shum, S. (2023) 'Using causal models to bridge the divide between big data and educational theory', *British Journal of Educational Technology*, Vol. 54, No. 5, pp.1095–1124.

Lee, S.M. and Lee, D. (2020) ''Untact': a new customer service strategy in the digital age', *Service Business*, Vol. 14, No. 1, pp.1–22.

Li, F., Cui, Y., Li, Y., Guo, L., Ke, X., Liu, J., Luo, X., Zheng, Y. and Leckman, J.F. (2022) 'Prevalence of mental disorders in school children and adolescents in China: diagnostic data from detailed clinical assessments of 17,524 individuals', *Journal of Child Psychology and Psychiatry*, Vol. 63, No. 1, pp.34–46.

Mao, R., Liu, Q., He, K., Li, W. and Cambria, E. (2022) 'The biases of pre-trained language models: an empirical study on prompt-based sentiment analysis and emotion detection', *IEEE Transactions on Affective Computing*, Vol. 14, No. 3, pp. 1743–1753.

Marechal, C., Mikolajewski, D., Tyburek, K., Prokopowicz, P., Bougueroua, L., Ancourt, C. and Wegrzyn-Wolska, K. (2019) 'Survey on AI-based multimodal methods for emotion detection', *High-Performance Modelling and Simulation for Big Data Applications*, Vol. 11400, pp.307–324.

Oduor, M. and Oinas-Kukkonen, H. (2021) 'Committing to change: a persuasive systems design analysis of user commitments for a behaviour change support system', *Behaviour & Information Technology*, Vol. 40, No. 1, pp.20–38.

Qin, R., Wang, F.-Y., Zheng, X., Ni, Q., Li, J., Xue, X. and Hu, B. (2024) 'Sora for computational social systems: from counterfactual experiments to artificial of actual experiments with parallel intelligence', *IEEE Transactions on Computational Social Systems*, Vol. 11, No. 2, pp.1531–1550.

Salimi, N., Gere, B., Talley, W. and Irioogbe, B. (2023) 'College students mental health challenges: concerns and considerations in the COVID-19 pandemic', *Journal of College Student Psychotherapy*, Vol. 37, No. 1, pp.39–51.

Søgaard Neilsen, A. and Wilson, R.L. (2019) 'Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: An integrative literature review', *International Journal of Mental Health Mursing*, Vol. 28, No. 1, pp.22–39.

Yadav, S.P., Zaidi, S., Mishra, A. and Yadav, V. (2022) 'Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (RNN)', *Archives of Computational Methods in Engineering*, Vol. 29, No. 3, pp.1753–1770.

You, H., Tian, S., Yu, L. and Lv, Y. (2019) 'Pixel-level remote sensing image recognition based on bidirectional word vectors', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 2, pp.1281–1293.

Zhu, L., Zhu, Z., Zhang, C., Xu, Y. and Kong, X. (2023) 'Multimodal sentiment analysis based on fusion methods: a survey', *Information Fusion*, Vol. 95, pp.306–325.