J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 3089--3107

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Research on the design and educational application of a real-time monitoring system for mental health status based on large-scale data stream computation

Ziyan Yao[1,✉], Jin Gan[1]

[1] *Building and Road Engineering School, Guangxi Eco-engineering Vocational & Technical College, Liuzhou, Guangxi, 545004, China*

ABSTRACT
With the accelerated pace of society and the increasing pressure of competition, the issue of mental health has received increasing attention. Especially in the field of education, students' mental health status directly affects their student outcomes and overall development. The aim of this study is to design a mental health status monitoring system based on large-scale data streaming computation, to realize dynamic real-time monitoring of individual mental health through multi-source data acquisition and efficient algorithm processing, and to explore its application in educational scenarios. Sliding window algorithm and Hidden Markov Model are used to analyze and process the collected multi-source data such as physiological signals, and the experimental results show that the system is able to significantly test the difference between people with high and low scores on psychological test scales in the monitoring of mental health status, and it can provide educators with valuable decision-making support and help students' mental health education and intervention.

Keywords: large-scale data computation, sliding window algorithm, hidden Markov model, mental health status monitoring

1. Introduction

In today's fast-paced and high-stress social environment, mental health problems are becoming more and more prominent, which have an undeniable impact on the quality of life of individuals, their work efficiency, and even the harmony and stability of society [1-2]. In terms of age groups, children, adolescents, adults, as well as the elderly may face different types and degrees of psychological problems, and children may suffer from attention-deficit hyperactivity disorder and separation anxiety [3-4]. Adolescents are prone to depression and anxiety in terms of academic stress, social

✉ Corresponding author.
E-mail address: ganjin23@163.com (Z. Yao).
Received 25 October 2024; Revised 10 December 2024; Accepted 05 March 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-173

relationships, gender, and stress in the school environment [5-8]. Adults may experience burnout and psychological distress due to marital tension under the heavy pressure of work, marriage, and family responsibilities [9-11]. Elderly people, on the other hand, may experience psychological feelings such as loneliness and loss due to declining physical functioning and shrinking social circles [12-13]. From the perspective of social groups, students, working professionals, healthcare workers, and military personnel are also in need of focused attention. Student groups face the pressure of academic competition, working people may be affected by factors such as work intensity and career development bottlenecks, and healthcare workers are prone to psychological fatigue and stress reactions when they are in a high-intensity, high-risk work environment for a long period of time [14]. Military personnel may also face psychological challenges in the battlefield, during missions and training [15-16]. In addition, some special populations, in the case of illness, manifestation of their own defects, and postpartum, are prone to negative emotions such as anxiety, low self-esteem, and stress [17-19]. Traditional psychotherapy usually requires a face-to-face counselor to communicate with the patient, and the mental health status is often more difficult to self-perception and self-assessment, especially in the case of the relative lack of resources in the psychological counseling centers set up in the family or the community, or their own family economic conditions are limited, the family concept of the old-fashioned and other factors, it is difficult to understand the state of their own mental health in a timely manner [20-21]. Since changes in mental health status are affected by multiple factors, and the data it generates are diverse, such as physiological data such as an individual's heart rate, blood pressure, sleep quality, past medical history, network data such as an individual's social networking platforms and online platforms, and event data such as an individual's occurrence of a major event, emotional experiences, social interactions, and academic or work efficiency, the difficulty of monitoring escalates [22-23]. And in the past, mental health monitoring mainly relied on traditional questionnaires and clinical diagnosis, and these methods have problems such as high time cost, limited sample size, and subjective results. Therefore, it is of great practical significance to study a real-time monitoring system for mental health status.

In this paper, we design the general framework of the real-time monitoring system of mental health status from two aspects: electronic psychological profile and dynamic monitoring of individual mental health. The Apache Flink streaming computing framework is used in the data layer to analyze and process the computation of large-scale data streams. The data stream is segmented and calculated using the sliding window algorithm, and the trend analysis of the data stream after the sliding window segmentation is performed based on the least squares method. According to the relationship between emotion and mental health, the Hidden Markov Model was used to construct an emotion prediction model, and the individual mental health status was assessed by the calculation of emotion entropy. Subjects were selected in a university, psychological questionnaire scales were used to count the psychological state of the subjects, and psychological signals such as EEG and EMG were collected from the subjects. Finally, the real-time monitoring system of mental health status constructed in this paper is applied to monitor and calculate the mental health status of the subjects, and verify the effectiveness of the system of this paper in the real-time monitoring of the dynamics of mental health and the application of education.

2. Overview

Mental health monitoring lies in identifying individual mental health status, tracking, and evaluating the whole process. Literature [24] used daily voice call data to construct a mental health status monitoring system, which focuses on analyzing and evaluating voice data in order to monitor stress-induced mental health problems. Literature [25] innovated a dynamic monitoring system for the mental health of higher vocational students with the support of big data, and also updated the traditional mental health profile in conjunction with the mental health profile. Literature [26] designed a multimodal affective computing hierarchical attention network and mentioned that the design can be used to design a mental health monitoring system with high privacy and low cost with the support of wearable technology and flexible electronics. Literature [27] created an edge negative information measurement system with intelligent algorithms on the edge cloud to cognize the mental health status, thus realizing the mental health status of individuals in negative information environments for timely intervention. Literature [28] conceptualized a machine learning-enabled active social sensor service, which is a framework for proactively monitoring the mental health of twitter users. Literature [29] implemented a mental health monitoring and surveillance system for IoT data through machine learning and data mining techniques, which used depression, a common psychological problem, as a research sample, and the Random Forest algorithm in machine learning to detect the highest accuracy. Literature [30] constructed a mental health monitoring and assessment model with support vector machine and decision tree on the basis of organizing and analyzing college students' social network data by using big data analytics, while optimizing the data entered into the model with genetic algorithm. And literature [31] improves the scanning process of CM Apriori algorithm and PM_Apriori algorithm by genetic algorithm, analyzes the association rules of students' psychological symptoms with the technical support of data mining, and then analyzes the students' mental health status by using the improved Apriori algorithm, so as to realize the monitoring. Literature [32] proposed an intelligent mental health monitoring framework which analyzes the stress index in questions and answers with random forest algorithm, recognizes facial expressions with support vector machine, and also predicts the mental health status with the combination of interval type II fuzzy logic. Literature [33] proposed a mental health monitoring system by generating adversarial network algorithms and integrating IoT technology, which focuses on students' academic performance, feelings, and emotional socialization, and monitors these data in real time to assist teachers in mental health services. Literature [34] developed a lightweight convolutional neural network supported dew computing inspired mental health monitoring system framework which performs psychological monitoring from facial expressions of postpartum and sick individuals. Literature [35] proposed a real-time mental health monitoring system that monitors physiological parameters through an Arduino UNO and multi-sensors and incorporates machine learning to accomplish real-time monitoring, which has an accuracy of 96.6%. Overall, these studies have considered basic physiological data in addition to online social data, event performance situation data, facial expressions at different levels, mostly delayed monitoring.

In addition, literature [36] created a real-time mental health monitoring framework, SmartMHealth, which can be placed in a smartphone and has convenience along with real-time. Literature [37] developed a mental health monitoring system with a snowmelt-driven bi-directional fine-tuned recurrent neural network algorithm to monitor an individual's mental health status, and biosensors to collect an individual's physiological data in real-time, with the algorithm's monitoring accuracy being above 90%. For real-time monitoring, a powerful computational method is needed to face physiological data, network data, and event data, and large-scale data streaming is the technology for real-time processing and analysis, which has the advantages of real-time, streaming processing, and low latency, and it is a good choice for real-time monitoring [38].

3. System design

3.1. System architecture

3.1.1. General system architecture. The main goal of this system is to realize the all-round management and attention to mental health through the integration of two aspects, namely the establishment of an electronic psychological profile for one person's lifetime and the dynamic monitoring of individual mental health [39]. The overall architecture of the system is shown in Figure 1.

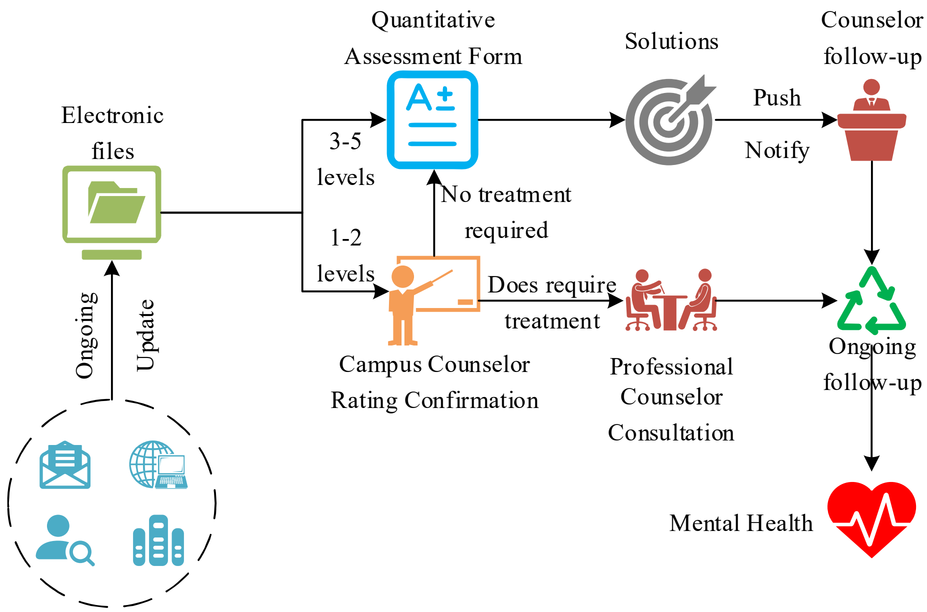


Fig. 1. Overall design

1) Electronic psychological profile. The establishment of an electronic psychological profile makes it easier to record and summarize an individual's psychological condition and experiences. When a person's psychological state fluctuates, the profile system is able to quickly capture the current situation and feed back his or her significant past experiences and psychotherapeutic content to the current psychologist. In this way, the psychologist can gain a more accurate and comprehensive understanding of the subject's psychological condition and be able to provide more effective and enriched intervention and treatment programs. In addition, the e-psychological profile can optimize subsequent psychological assessment scales for personalized and customized assessment. Through the content in the personal ePsychological profile, the assessment can be tailored to the characteristics and needs of the examinee, providing a psychological assessment that is more in line with the individual.

2) Dynamic monitoring of individual mental health. Traditional mental health monitoring often fails to be dynamic, i.e., an individual is considered to be in a normal state of mind only when he or she passes the initial psychological assessment and the values are within a reasonable range. This means that subsequent mental health monitoring is often missing. To solve this problem, this paper investigates a dynamic real-time mental health monitoring system based on large-scale data stream computing. The system highlights the characteristics of "dynamic", and through the combination of feedback information from multiple dimensions such as "home-school" and "teacher-student", the psychological state of the individual can be monitored in a full cycle, so as to achieve timely intervention and treatment.

3.1.2. Data layer design. This system uses Apache Flink streaming computing framework in the data layer [40], which is suitable for large-scale data processing and analysis, and has many advantages in real-time processing, high throughput, fault tolerance, flexible windowing, and rich API support, making Flink an ideal choice for processing complex data.

Apache Flink adopts a stream-based computing model with excellent scalability, allowing users to easily expand their computing power when processing unbounded data streams. Flink records the state of tasks by periodically generating task checkpoints, so that the system can use the most recent checkpoints to recover the state of the task in the event of a failure, thus avoiding data loss and task recalculation. Users can configure the fault tolerance level of Flink according to their actual needs, balancing fault tolerance overhead and system performance. Flink realizes efficient large-scale data processing through data parallel processing, dividing stream processing tasks into multiple subtasks each executed on an independent parallel thread, and this task parallelism allows the system to process data on multiple computing nodes at the same time, improving the overall computing power.

Apache Flink streaming computing framework, as a new distribution algorithm, divides the large-scale data processing task into a series of small, successive data streaming operations, each of which forms a computing phase that can be executed in parallel on different nodes of the cluster. Data is passed between different computing phases in the form of streams, which avoids data sharing and synchronization of displays and improves the efficiency of the whole system. Meanwhile, Flink uses an asynchronous, non-blocking message-passing model to achieve coordination between nodes through lightweight asynchronous communication. Data is loaded in the form of streams and acquired from various data sources in real time, and preprocessing includes data cleaning, conversion, and other operations. It can better ensure data quality and format consistency. The task scheduler dynamically schedules tasks according to the data flow graph and cluster state, and assigns tasks to idle compute nodes through the task manager to achieve a balanced load. On each compute node, Flink executes different data flow operations in parallel, making full use of the cluster's compute resources to achieve efficient large-scale data processing, and after the computation is completed. After the computation is completed, Flink outputs the results in a streaming manner, supporting a variety of output targets, such as robustness, database or other stream processing applications.

3.1.3. Application layer design. The ultimate goal of this system is to realize the intelligence of psychology in order to reduce labor cost and improve the efficiency of counselors and related personnel. In order to realize this goal, the system application layer mainly adopts the front-end Vue + back-end Java technology, and combines with intelligent hardware to design and realize the collection of client and PC WEB pages. The following are the main contents involved in the system:

1) Data collection and management. The system applies big data technology to collect and organize individual psychological profiles. These profiles will be managed through normalization and encryption to form an overall psychology big data analysis platform.

2) Differentiated Analysis and Personalized Push. Differentiated analysis is achieved for each individual through algorithms, and personalized psychological education content is pushed to each individual based on the analysis results. In this way, we can provide tailor-made services according to the characteristics of different individuals and better solve their psychological problems.

3) Psychological condition prediction and early warning. The algorithm is utilized to dynamically predict the psychological condition of different individuals. Based on the prediction results, the system will discover the early warning and send real-time warning notifications to relevant personnel for timely intervention and support.

3.2. Data Acquisition and Processing

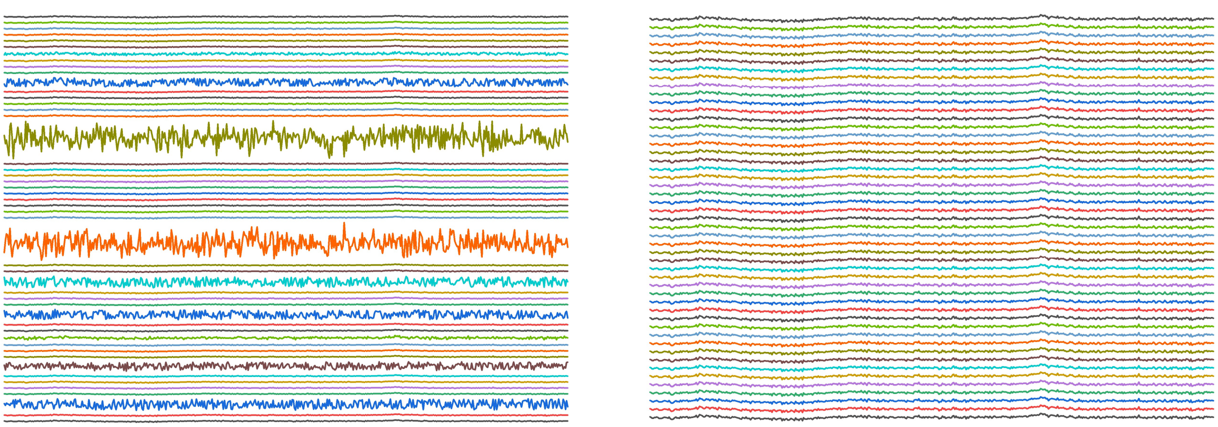
3.2.1. Data acquisition. Since the current publicly available dataset could not meet the experimental requirements of this study, the study recruited 75 volunteers to participate in this data collection and constructed the UJNSET-EMO2024 dataset. There were 60 local college students (mean age 21.48, SD=±1.87) and 15 publicly recruited off-campus volunteers (mean age 21.94, SD=±2.58). All subjects in this study were right-handed and had not taken coffee or psychotropic drugs within two weeks. 50 of them were healthy subjects (no history of mental illness, psychoactive substance abuse, or depression in themselves or in their families, and passed the SCL-90 and the physician's psychiatric diagnosis), and 25 were subjects with abnormal mental health status (11 fulfilled the Mini-

International Neuropsychiatry Criteria for the diagnosis of depression and Patient Health Questionnaire PHQ-9 score greater than or equal to 5, and 14 met Diagnostic and Statistical Manual of Mental Disorders DSM-5 criteria for assessment of anxiety disorders and Hamilton Anxiety Scale HAMA score greater than 14).

The UJNSET-EMO2022 dataset contains a total of three physiological signals, EEG, ECG, and GSR, which were collected by all subjects under the emotion-evoking paradigm. The mental health status was labeled as healthy as well as unhealthy, and the emotions evoked in the subjects were labeled as positive, medium, and negative emotion types. The emotion-evoking material used in the data collection process was evaluated according to validity, arousal, and dominance, with validity being used to differentiate between positive, neutral, and negative emotions, and arousal and dominance being used to judge the intensity of the emotion evoked, but it is worth noting that the evoked emotions were labeled as positive, neutral, and negative only, and were no longer categorized according to validity, arousal, and dominance.

A total of 75 subjects viewed five sets of emotion-evoking material, each with 25 emotion-evoking images, for a total of 9,630 multimodal physiological signals labeled with emotions and 10,458 multimodal mental health data labeled with mental health states in the full dataset.

3.2.2. Data processing. In order to minimize the interference of the surrounding environment on the physiological signals of different modalities, in this paper, the raw EEG, ECG, and GSR signals were targeted for data preprocessing respectively. For the preprocessing of EEG signals, the following operations need to be carried out, firstly, the EEG signals of 62 whole-brain channels were performed with M1M2 bilateral mastoid (average) as the reference electrode, and then a 0.1-Hz high-pass filter, a 100-Hz low-pass filter, and a 50-Hz trap filter were utilized to filter out the noise interference as much as possible. It was then necessary to separate the EEG signals of 27 pathways out of the 62 pathways of EEG signals acquired. These channels can not only maximize the retention of EEG spatial features, but also do not lose more emotional features. The EEG signal before and after processing is shown in Fig. 2, (a) and (b) represent the EEG signal data before and after processing, respectively.



(a) Before processing (b) After processing

Fig. 2. Before and after pretreatment of EEG signal

For ECG signals, the industrial frequency interference in the original signal is removed by using a 50Hz band-trap filter, then the signal is wavelet transformed, and since the wavelet decomposition coefficients of the noise are smaller in amplitude compared to the amplitude of the valuable ECG signals, they are processed by threshold quantization and the estimated wavelet coefficients are obtained, and then the signals are reconstructed by the estimated wavelet coefficients in order to eliminate the noise and to improve the quality of the signals. Finally, the baseline drift is eliminated by median filtering. For GSR signals, 0 Hz high pass filtering as well as 0.2 Hz low pass filtering is utilized to remove noise interference. The above two signals need to be cropped according to the emotion interval specified in the emotion-inducing paradigm after the noise reduction process, and then the data preprocessing is basically completed after the baseline calibration.The before and after comparison of the ECG and GSR signals are shown in Figures 3 and 4.

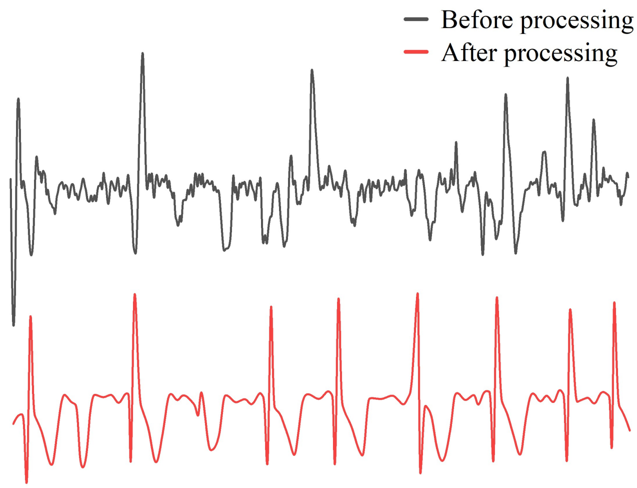


Fig. 3. Before and after pretreatment of ECG signal

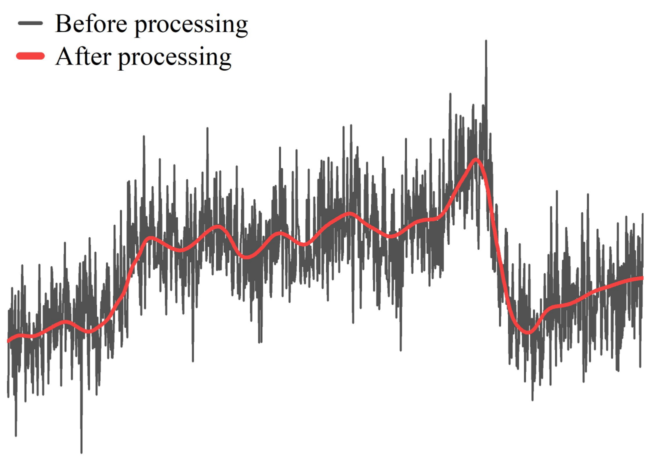
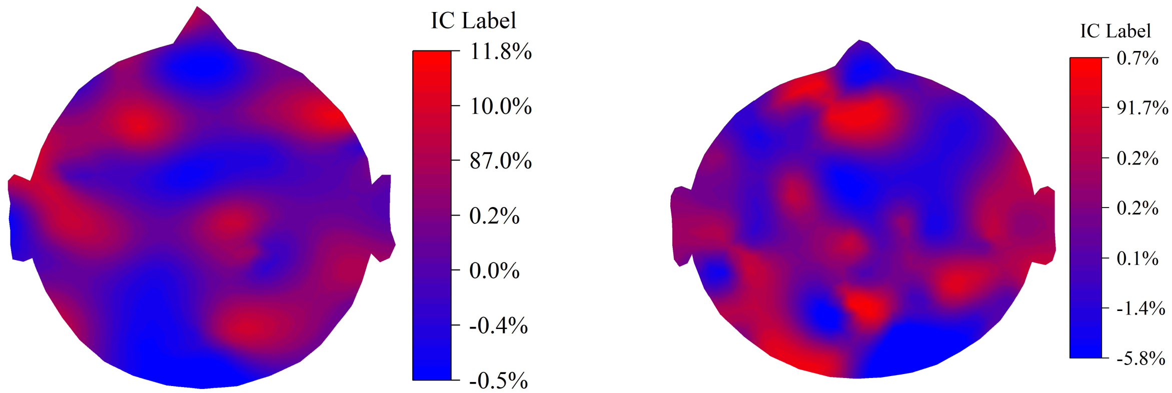


Fig. 4. Before and after pretreatment of GSR signal

After the above data processing, as the EEG signals are more likely to be interfered by other bioelectric signals, in this case this paper further purifies the EEG signals by ICA independent component analysis and manual labeling deletion. Figure 5 shows the doped EEG signals and EMG signals of a subject analyzed by ICA.



(a) EOG (b) EMG

Fig. 5. ICA's separation of EOG and EMG

3.3. Data stream processing based on sliding window algorithm

3.3.1. Data flow description based on sliding window technique. The following is a formal description of the data flow:

$$Y=\left\{y\left(t_1\right),\cdots,y\left(t_i\right),\cdots,y\left(t_p\right),\cdots\right\} \tag{1}$$

where t_p is the point in time corresponding to the data. The data stream segmentation splits Y into a series of consecutive non-empty data $\left\{Y_1,\ldots,Y_j,\ldots,Y_x,\ldots\right\}$ through the sliding window technique [41], where the jth data segment $Y_j=\left\{y\left(t_{j,1},\cdots,y\left(t_{j,t}\right),\cdots,y\left(t_{j,n_t}\right)\right)\right\}$, corresponding to the arrival time of the data $t_{j,i}\in\left\{t_1,\ldots,t_i,\ldots,t_p,\ldots\right\}$, $j=1,2,\ldots,s$, $l=1,2,\ldots,n_j$, n_j is the length of the data segment Y_j, and $t_{1,1}=t_1$, Y, is the data segment that contains the current data $y\left(t_P\right)$ and is called the current data segment.

Let the data of Y be described by a linear regression model as:

$$y(t)=f\left(t;c_j\right),t\in\left\{t_{j,1},\ldots,y_{j,n_i}\right\} \tag{2}$$

where $f\left(t;c_j\right)=\left(\alpha_j+\sigma_j(t)\right)t+\beta_f$ is the regression model for data segment Y_j, $c_j=\left[\alpha_j,\beta_j\right]^T$ is the model parameters; α, and β_i are the regression coefficients, parameter α_j is the trend eigenvalue for Y_j, and σ, is the error perturbation. The first data element $y_{\left(t_{j+1,1}\right)}$ of data segment Y_{j+1} is the split point of Y_j.

The size of the sliding window is set to be the length of Y_j, i.e., n_j. At the moment the most recent n_j data to be processed from the time of computation, i.e., the data going n_j forward in time, is to be included in the processing. t_{n_j}, α is the 0 vector. Since data streams are constantly emerging, intuitively, this pattern is like using an invariant window, data passes through the window over time, and the data that appears within the window is the data being processed.

3.3.2. Segmented trend analysis based on the least squares method. Assuming that the length of the data is N and the length of the data segment after the sliding window segmentation is s, let the functional relationship between the monitoring time t and the monitoring value be given by the theoretical equation:

$$y = f\left(t; c_1, c_2, \ldots, c_s\right) \tag{3}$$

is given, where $c_1, c_2, \ldots, c_s$ is s a parameter to be determined experimentally. For each set of observations $\left(t_j, y_j\right)$, $j = 1, 2, \ldots, N$. Corresponds to a point on the xy -plane. If there is no measurement error, all these data points fall exactly on the theoretical curve. The system of equations is obtained by simply substituting the s sets of measurements in the sliding window into equation (3):

$$y_j = f\left(t; c_1, c_2, \ldots, c_s\right) \tag{4}$$

$j = 1, 2, \ldots, s$ in Eq. The values of the s parameters are obtained by solving the joint solution of the s equations.

For simplicity, C is used below to represent $\left\{c_1, c_2, \ldots, c_s\right\}$. Considering that the measurements are independent of each other, the parameter C can be estimated from the observation $\left\{y_1, y_2, \ldots, y_N\right\}$ so that:

$$\sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - f\left(t_j; C\right)\right]^2 = \min \tag{5}$$

Taking the minimum value. Where σ, is the disturbance error of the distribution, equation (5) is called the least squares criterion. Because of the weighting factor $\omega_j = 1/\sigma_j^2$, Eq. (5) shows that estimating the parameters by the least squares method requires that the weighted sum of squares of the deviations of each measurement y_j be minimized.

According to Eq. (5), there should be:

$$\left.\frac{\partial}{\partial c_k} \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - f\left(t_j; C\right)\right]^2 \right|_{c=\hat{c}} = 0 (k = 1, 2, \ldots, s) \tag{6}$$

This results in a system of equations:

$$\sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - f\left(t_j; C\right)\right] \left.\frac{\partial f(t; C)}{\partial C_k}\right|_{c=\overline{c}} = 0 \quad (k = 1, 2, \ldots, s) \tag{7}$$

Solving the system of equations (6) gives the estimates of s parameter $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_s$, which leads to equation $f\left(t; \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_s\right)$ of the fitted curve.

The results of the segmented fit to the data are evaluated. If y_j follows a normal distribution, the fitted x^2 can be introduced:

$$x^2 = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - f\left(t_j; C\right)\right]^2 \tag{8}$$

Substituting the parameter estimate $\hat{c} = \left(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_s\right)$ into equation (8) and comparing it with equation (5) yields the smallest x^2 value:

$$x_{\min}^2 = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - f\left(t_j; \hat{c}\right)\right]^2 \tag{9}$$

It can be shown that $x_{\min}^2$ obeys a x^2 distribution with $v = N - s$ degrees of freedom, and therefore the results of the fitted data can be x^2 tested.

The expected value of the random variable $x_{\min}^2$ is known from the distribution of x^2 to be $N-s$. If $x_{\min}^2$ is close to $N-s$, say $x_{\min}^2 \leq N-s$, as calculated from equation (9), the fit is considered acceptable. If $\sqrt{x_{\min}^2} - \sqrt{N-s} > \varepsilon$, the fitted data results are considered to be in significant contradiction with the observed values. At this point it is necessary to determine whether to activate the early warning system.

3.4. Hidden Markov Model-based Assessment of Mental Health Status

Emotions are a barometer of an individual's mental health and a predictor of psychopathology. Positive emotions (e.g., optimism, enthusiasm, etc.) are conducive to individuals being in a state of mental health, while negative emotions (e.g., pessimism, depression, anxiety, etc.) are prone to triggering psychological disorders in individuals. Therefore, paying attention to the occurrence and development of individual emotions is also of great significance to mental health education, and closely linking emotion measurement with clinical diagnosis and individually appropriate treatment techniques has become a future trend in the field of psychological counseling, i.e., by predicting and interpreting emotions, the level of mental health of individuals can be better maintained.

3.4.1. The three spatial states of emotion. In this paper, emotions are divided into three state spaces: happy, calm, and unhappy, which correspond to the state number M in the hidden Markov model, and the self-rated or other-rated sentiment index is regarded as a discrete variable, corresponding to the observed values in the hidden Markov model, which requires the indexing of the emotional intensity of each state space. In the state space of "happiness", the emotion index is defined as 1~7, 1 represents mild happiness, and 7 represents mania (extreme expression of happiness); In the state space of "calm", 0 is used to represent the emotional index of calm. In the state space of "unhappyness", the emotion index is defined as -1~-7, with -1 representing mild unhappiness and -7 representing depression (extreme manifestation of unhappiness).

The classical discrete HMM was chosen for modeling mood prediction, where self- or other-rated reported mood indices appear with a certain probability under each implicit mood. Since the Hidden Markov Model consists of two parts, one of them is the Markov chain described by π and A. Obviously, different π and A will determine different shapes of the Markov chain. In an individual's emotion prediction model, the shape of an individual's inherent emotional homeostasis is different because of genetic factors and different educational environments from childhood. Starting from any moment, an arbitrary state may be reached at the next moment. Its Markov chain is the type of state traversal.

3.4.2. Construction of mood prediction models. The Baum-Welch algorithm [42] for HMM was utilized to implement the parameter estimation of the emotion transfer probability matrix and the probability vectors of different emotion outputs for a certain mood state:

$$E = -C\sum_{i=1}^{n} p_i \lg p_i \tag{10}$$

where: E is the emotional entropy, i.e., the amount of emotion. p_i is the probability of occurrence of the i rd emotional state. C is a constant related to the logarithmic base and the choice of units.

If the probability of occurrence of each emotional state in the emotional space is the same, then $p_i = \frac{1}{n^m}$, $i \in \left(1, 2, \cdots, n^m\right)$. Thus, the emotional entropy reaches a maximum value of:

$$E_{\max} = -C \lg p_i = Cm \lg n \tag{11}$$

$E_{\max}$ represents the maximum value of emotional complexity for the lifeforms in the study area. That is, $E_{\max}$ is the upper bound, and the closer to this value the more it implies an emotionally rich character. This emotional entropy is characterizing the overall performance of the constructed emotion. And the emotional stability at a given moment can be expressed in terms of emotional entropy, which is defined as:

$$e_i = -C\sum_{j=1}^{n} p_{ij} \lg p_{ij} \tag{12}$$

where: e_i is the emotional entropy in state i. p_{ij} is the probability of going to the jth emotional state in the i th. C is the constant related to the logarithmic base and unit selection.

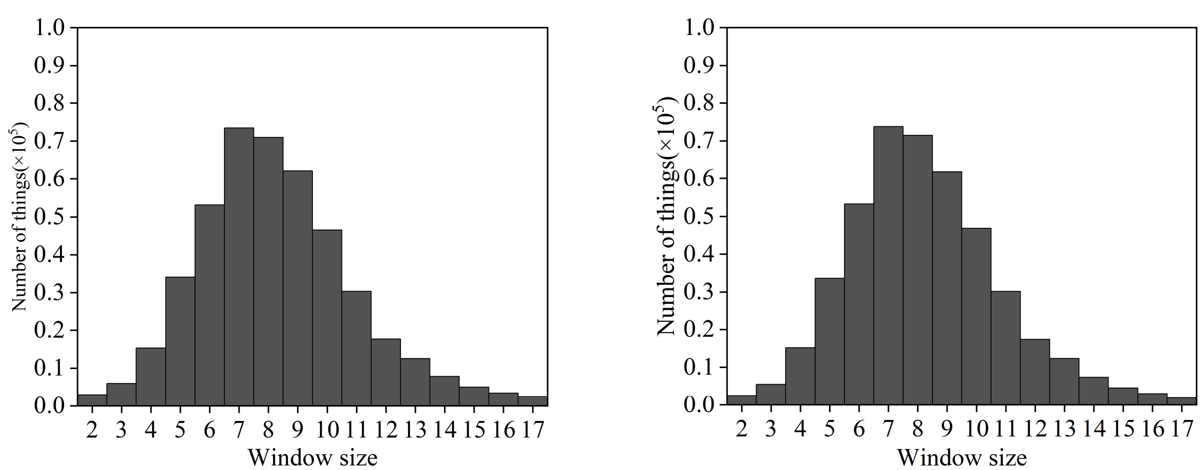
The HMM-based emotion prediction model treats the human emotional process as a two-layer stochastic process, and by adjusting the initial parameters of the model, it is able to predict the emotional output profile of individuals with different inherent emotional characteristics. As an emotion engine, the development of the emotional process is predicted in a probabilistic form.

4. Experimental results and analysis

The algorithm in this paper is written in C++ and developed using VS2023, the development environment utilized is Window10 Professional 64-bit operating system, the processor is AMD E2-3000M APU with Radeo(tm) HD Graphics 1.80GHz, and 6G RAM. Data were collected by recruiting volunteers to form the dataset UJNSET-EMO2022 with 10,458 mental health data.

4.1. *Sliding window performance analysis*

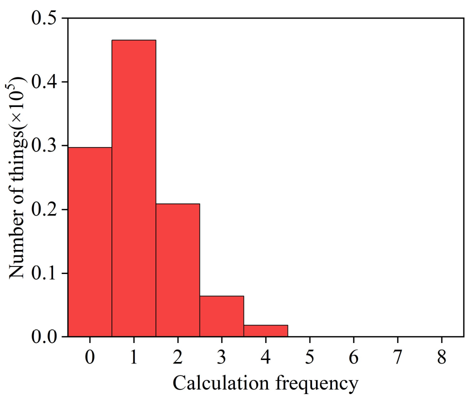
It can be proved by this experiment that the space complexity of the sliding window algorithm proposed in this paper is $O(|W|)$. The space complexity required for the operation of maintaining the window data is equal to the cost of calculating the distance between two things in the window plus the cost of the pruning operation, and the cost of calculating the distance is dependent on the window size, Fig. 6 shows the results of the distribution of the values of the sliding window size observed at different points in time as the data flow is drawn into the window, (a) represents a uniform distribution , and (b) represents binomial normal distribution. From the experiment it can be analyzed that the amount of bar data in the graph is generated by the difference between two randomly selected records. In the worst case i.e. no two identical records exist within that window, the window length is the same as the size of the crossed-in records. The experiment distributes the window size distribution by binomial normal distribution and uniform distribution. These two distributions were chosen to describe the probability densities of the results since each data in the data stream is relatively independent and there are n possibilities.



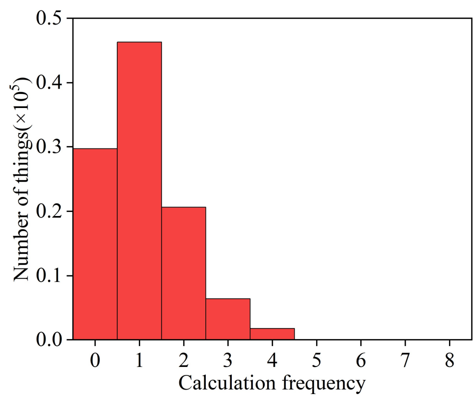
(a) Uniform distribution (b) Binomial distribution

Fig. 6. Sliding window size distribution

Fig. 7 shows the distribution of the number of operations observed at different times in the data stream through the pruning rule, (a) and (b) show the distribution of the number of computations without invoking and invoking the pruning algorithm, respectively. The experiments illustrate that when the pruning algorithm is not invoked for the removal of data in the window, the sum of the data in the window increases monotonically with time, and from the figure, it can be analyzed that the value of the window size does not increase significantly when the pruning algorithm is invoked, which is due to the removal of the transactional dataset in time for the analysis of the data, and for longer periods of time when the data in the window is less. This experiment shows that the arithmetic space complexity of the window is within the controllable range.



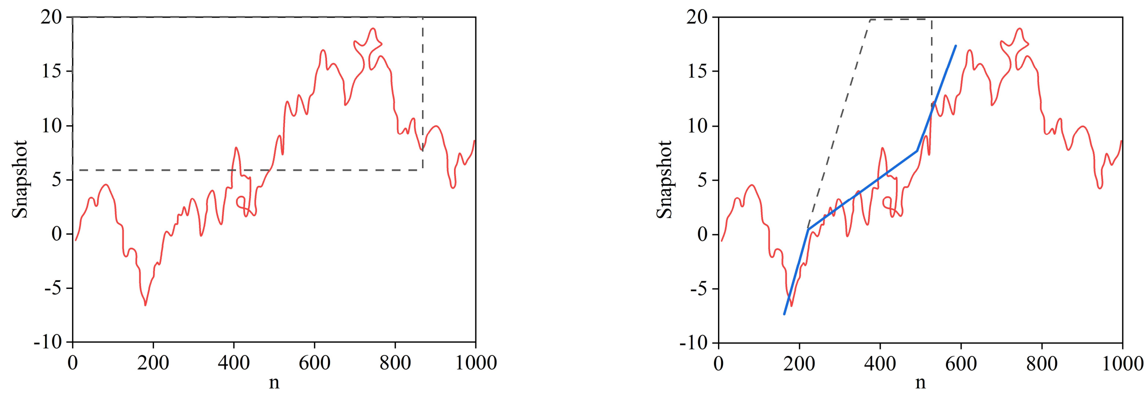
(a) No call algorithm



(b) Call algorithm

Fig. 7. Calculation frequency distribution under pruning rule

When the window's transactions are nonfrequent to prune it, set the time for the data flow to fill the current window at the standard rate of τ, in the time range of the amount of data in the window is small enough to remove the data from the window and perform a sliding operation on the window, the data points that need to be pruned are shown in Fig. 8, and (a), (b) respectively, are the data that need to be pruned under the pruning rules 1 and 2. From the experiment, it can be concluded that the dashed area is the infrequent itemset of the current window, which will be removed from the window, and the slash in the right center indicates the current skew rate, and when the slopes of the old data and the new data satisfy $P_1 > P_2$, all the transactions between the two sets are deleted. Obviously, the data itemsets between P_1 and P_2 need to be removed. Experimentally, it is proved that in the pruning rule, when there are two transactions S_m and S_n within the sliding window, transaction S_n needs to be removed when $n > m > c$ and the relative frequency of occurrence of the itemset is $\mu_{nm} \geq \mu_m$, which can respond to the data flow in a timely manner and reduce the memory consumption space.



(a) Pruning rule 1 (b) Pruning rule 2

Fig. 8. Data for pruning

4.2. Psychological state assessment analysis

The simulation of the emotion prediction model in this paper is based on the basic emotion theory, and the three basic emotions set are: happy, calm and unhappy. In the process of practical application, the emotion prediction model can be set in more detail according to different needs. According to the previous discussion of emotional entropy, the value of the emotional entropy of the sequential prediction model should theoretically be $-\lg(1/70)$ (about 1.845). In the process of simulation, the setting of the probability p_i of the appearance of a certain emotion plays an important role in the evolution of emotion prediction. Generally, the probability p_i takes a value between 0 and 1. Here, the probability $p_i = 0.33$ is set and the simulation results obtained are shown in Fig. 9. From the figure, it can be seen that after about 50 iterations, the value of emotional entropy tends to stabilize and approaches the theoretical value. It indicates that the probability values of each life form of the emotion prediction model have reached equality, and the redistribution and arrangement of emotional energy is completed. In the absence of new external stimuli, the subsequent transfer and exchange of emotional energy are carried out with the optimal strategy.

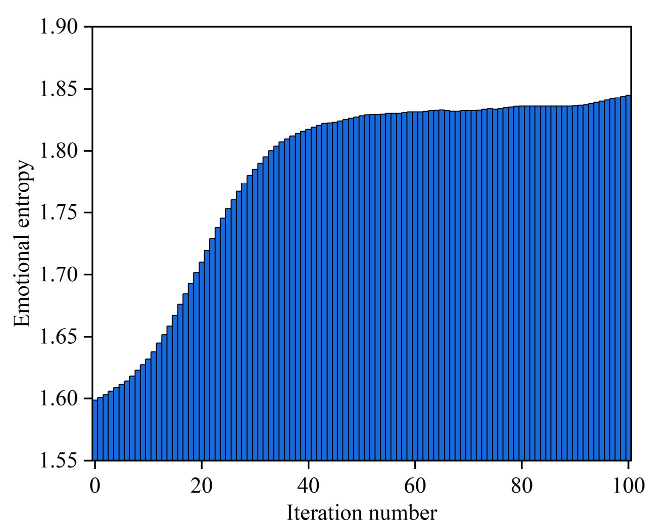


Fig. 9. The wave process of emotional entropy (From joy to peace) ($p_i = 0.33$)

Continuing to increase the value of the probability of occurrence of the emotional state so that $p_i = 0.5$, the simulation results are shown in Fig. 10. It is only after nearly 100 iterations that the emotional entropy of the emotion prediction model approaches the maximum value, and the number of iterations required is twice as many as at $p_i = 0.33$. The value of the probability of the appearance of the emotional state is defined as the sensitivity factor, and it can be seen that the influence of the sensitivity factor on the emotional entropy is very large, which is also in line with the emotional volatility of people in reality, and those who are prone to emotion are obviously very sensitive, and after modeling the person accordingly, the corresponding value can be set in the emotion prediction model to simulate the fluctuation of the person's emotion, so as to assess his mental health status.

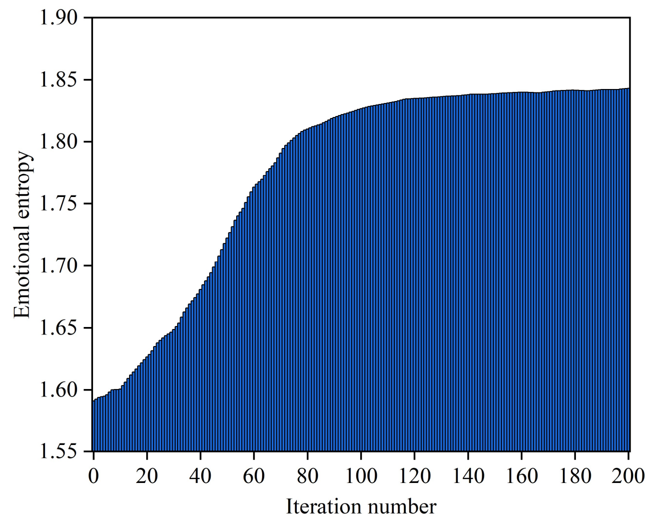
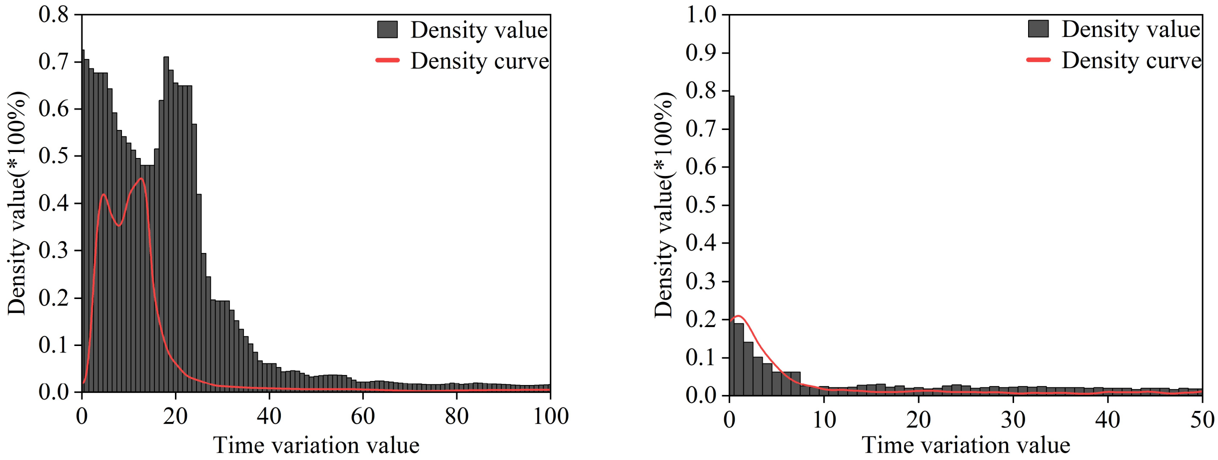


Fig. 10. The wave process of emotional entropy (From joy to peace) ($p_i = 0.5$)

4.3. Analysis of educational applications

In order to verify the monitoring effect of the mental health status monitoring system in this paper and the effectiveness of its educational application, the study designed an effect test analysis experiment. A total of 50 youths aged 19 to 22 from a university were selected, and the subjects had no history of major mental or other serious life-threatening diseases, and no history of drug or alcohol abuse. The experimenters used the system to monitor mental health status in an undisturbed, natural state, and the data were collected using a semi-automatic data collection method. Prior to the test, the psychological state of the subjects was counted using seven common questionnaire scales in psychology, and data collection of raw physiological signals from the subjects was completed. The study considered subjects with mean scores on the scales higher than or equal to 50 as psychologically unhealthy.

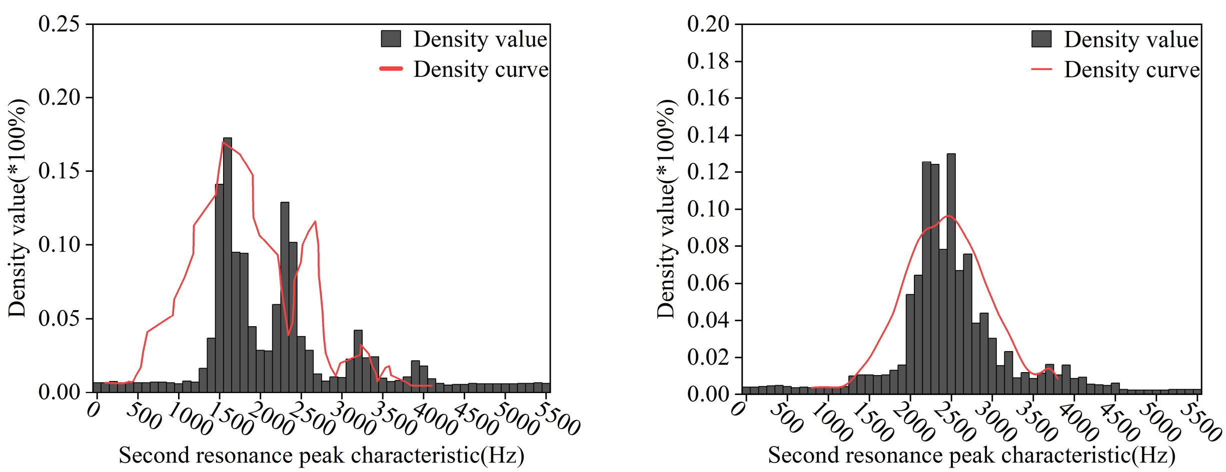
The results of the variance and density curves of the time domain in the subjects' one-day monitoring are shown in Figure 11, with (a) and (b) indicating high and low scorers, respectively. From the figure, it can be seen that the monitoring time variation values of the scale low-scoring testers are mainly concentrated in the interval between 0 and 8%, and the density curves of the monitoring time domain variance of the scale high-scoring testers have a Gaussian distribution in the interval between 0 and 60, and the maximum monitoring time variation value reaches 100, but the density value after the monitoring time exceeds 37 is less than 1%. It can be seen that the results of the calculation of the time-domain amplitude variance of the monitoring of mental health status can distinguish between subjects with high or low scores on the scale. Overall, the distribution of time-domain variance is more dispersed for those with higher scores on the psychological test scale, while the distribution of variance values for low-scoring test subjects is relatively concentrated and has smaller density values.



(a) High score (b) Low score

Fig. 11. The variance and density curve of the time domain

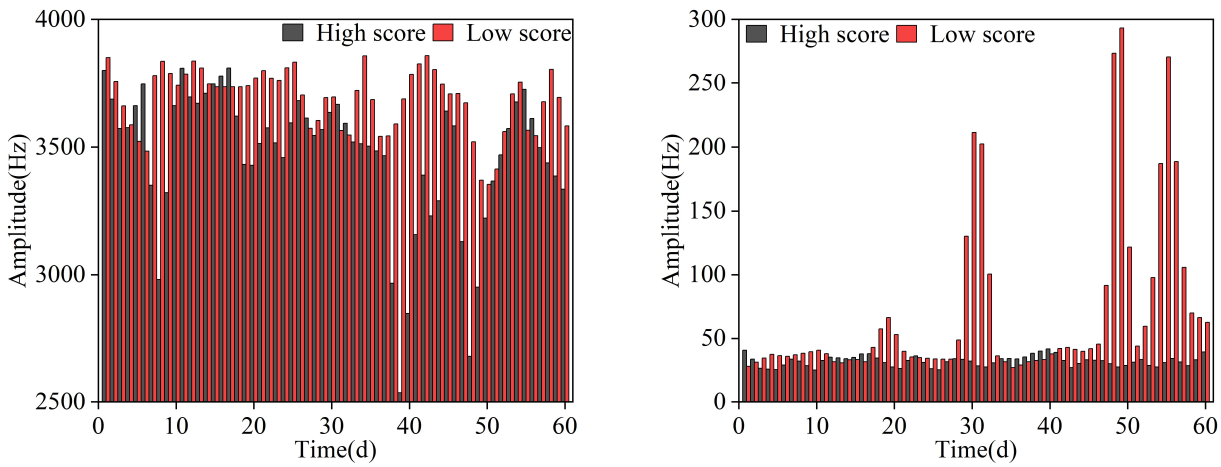
The results of the kernel density curves of the second resonance peak eigenvalue of the physiological signals in the one-day monitoring of the subjects are shown in Fig. 12, with (a) and (b) denoting the high-scorers and low-scorers, respectively. As can be seen from the figure, there is a significant difference in the distribution of the second resonance peak eigenvalue of physiological signals between the high and low scorers of the scale. The distribution of the density curves of the scale high-scoring testers has a wide range and is sparsely distributed in the range of 100 Hz to 4100 Hz. The distribution range of the density curve of the low-scoring testers is more concentrated, with Gaussian distribution in the range of 800Hz~3800Hz. It can be seen that the second resonance peak eigenvalue calculated by the heart health state monitoring system in this paper is closely related to the psychological scale score, and according to the test results of the system, it can be more obviously differentiated between the mental health state of the subjects.



(a) High score (b) Low score

Fig. 12. Second resonance peak characteristic value nuclear density curve

The subjects were monitored for 60 days for a long period of time for mental health status, and the trend of the amplitude change of the second resonance peak of the physiological signals and the time-domain maximum of the physiological signals during the monitoring time was mainly examined, and the experimental results are shown in Fig. 13, with (a) and (b) denoting the physiological signals' second resonance peak and the time-domain mean amplitude maximum, respectively. From the figure, it can be seen that the maximum value of the second resonance peak of the physiological signal of the low-scoring testers of the scale is in the range of 3350 Hz to 3858 Hz. The fluctuation of the second resonance peak of the high-scoring testers is very obvious, and the minimum value is only about 2535Hz, which has a more obvious outlier distribution. The maximum value of the time domain average amplitude shows the opposite trend of the change rule, and the fluctuation of the maximum value of the time domain average amplitude of the high-score testers is smaller, and the data are more stable. Overall comparison, the two groups of data in the numerical size distribution of low-scoring testers are greater than the high-scoring testers, in line with the actual situation of mental health status. It can be seen that the real-time monitoring system of mental health status designed in this paper has high reliability in analyzing the results of students' mental health, solves the integration and analysis problem of large-scale data flow computation, and achieves good results in the monitoring of students' mental health status in educational applications.



(a) Second resonance peak (b) Maximum average amplitude in time domain

Fig. 13. Mental health status monitoring trend changes

5. Conclusion

In this study, a real-time monitoring system for mental health status is designed and implemented on the basis of large-scale data stream computing and processing, which realizes efficient and accurate monitoring of individual mental health status through multi-source data acquisition and the application of sliding window algorithm, Hidden Markov Model and other technologies. And its application in the field of education is discussed. The experimental results show that the system demonstrates high performance and application value in mental health status monitoring, providing technical support for educators to accurately capture students' mental health status. Future research can further optimize the algorithm to improve the system's ability to identify complex mental health states. At the same time, the security and privacy protection of the system are strengthened to ensure the safety of user data during collection, transmission and storage.

Funding

This work was supported by Theoretical and Practical Research Project on Ideological and Political Education for College Students in Guangxi in 2024: "Practical Research on Integrating Excellent Traditional Culture of Guangxi into Psychological Education in Higher Vocational Education in the New Era" (Project Number: 2024SZ181).

References

[1] Bubonya, M., Cobb-Clark, D. A., & Wooden, M. (2017). Mental health and productivity at work: Does what you do matter?. Labour economics, 46, 150-165.

[2] Mason, D., Mackintosh, J., McConachie, H., Rodgers, J., Finch, T., & Parr, J. R. (2019). Quality of life for older autistic people: The impact of mental health difficulties. Research in Autism Spectrum Disorders, 63, 13-22.

[3] Battaglia, M., Garon-Carrier, G., Côté, S. M., Dionne, G., Touchette, E., Vitaro, F., ... & Boivin, M. (2017). Early childhood trajectories of separation anxiety: Bearing on mental health, academic achievement, and physical health from mid-childhood to preadolescence. Depression and anxiety, 34(10), 918-927.

[4] Song, J., Fogarty, K., Suk, R., & Gillen, M. (2021). Behavioral and mental health problems in adolescents with ADHD: Exploring the role of family resilience. Journal of Affective Disorders, 294, 450-458.

[5] Steare, T., Muñoz, C. G., Sullivan, A., & Lewis, G. (2023). The association between academic pressure and adolescent mental health problems: A systematic review. Journal of affective disorders.

[6] Cooper, K., Hards, E., Moltrecht, B., Reynolds, S., Shum, A., McElroy, E., & Loades, M. (2021). Loneliness, social relationships, and mental health in adolescents during the COVID-19 pandemic. Journal of Affective Disorders, 289, 98-104.

[7] Van Droogenbroeck, F., Spruyt, B., & Keppens, G. (2018). Gender differences in mental health problems among adolescents and the role of social support: results from the Belgian health interview surveys 2008 and 2013. BMC psychiatry, 18, 1-9.

[8] Aldridge, J. M., & McChesney, K. (2018). The relationships between school climate and adolescent mental health and wellbeing: A systematic literature review. International Journal of Educational Research, 88, 121-145.

[9] Cortès-Franch, I., Escribà-Agüir, V., Benach, J., & Artazcoz, L. (2018). Employment stability and mental health in Spain: towards understanding the influence of gender and partner/marital status. BMC public health, 18, 1-11.

[10] Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., ... & Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. Occupational and environmental medicine, 74(4), 301-310.

[11] Gammage, R. J., & Nolte, L. (2020). Family understanding and communication about an adult relative's mental health problem: A systematic narrative review. Journal of psychiatric and mental health nursing, 27(6), 763-788.

[12] Gerino, E., Rollè, L., Sechi, C., & Brustia, P. (2017). Loneliness, resilience, mental health, and quality of life in old age: A structural equation model. Frontiers in psychology, 8, 2003.

[13] Luo, M. S., Chui, E. W. T., & Li, L. W. (2020). The longitudinal associations between physical health and mental health among older adults. Aging & mental health, 24(12), 1990-1998.

[14] Suryavanshi, N., Kadam, A., Dhumal, G., Nimkar, S., Mave, V., Gupta, A., ... & Gupte, N. (2020). Mental health and quality of life among healthcare professionals during the COVID-19 pandemic in India. Brain and behavior, 10(11), e01837.

[15] Britt, T. W., McGhee, J. S., & Quattlebaum, M. D. (2018). Common mental disorders among US army aviation personnel: Prevalence and return to duty. Journal of Clinical Psychology, 74(12), 2173-2186.

[16] Kokun, O. (2025). The stability of mental health during war: Survey data from Ukraine. Journal of Loss and Trauma, 30(2), 242-263.

[17] Alene, K. A., Clements, A. C., McBryde, E. S., Jaramillo, E., Lönnroth, K., Shaweno, D., ... & Viney, K. (2018). Mental health disorders, social stressors, and health-related quality of life in patients with multidrug-resistant tuberculosis: a systematic review and meta-analysis. Journal of Infection, 77(5), 357-367.

[18] Duplaga, M., & Szulc, K. (2019). The association of internet use with wellbeing, mental health and health behaviours of persons with disabilities. International Journal of Environmental Research and Public Health, 16(18), 3252.

[19] Patabendige, M., Athulathmudali, S. R., & Chandrasinghe, S. K. (2020). Mental health problems during pregnancy and the postpartum period: A multicenter knowledge assessment survey among healthcare providers. Journal of Pregnancy, 2020(1), 4926702.

[20] Cybulski, M., Cybulski, L., Krajewska-Kulak, E., & Cwalina, U. (2017). Self-assessment of the mental health status in older adults in Poland: a cross-sectional study. BMC psychiatry, 17, 1-10.

[21] McClung, N., Glidewell, J., & Farr, S. L. (2018). Financial burdens and mental health needs in families of children with congenital heart disease. Congenital heart disease, 13(4), 554-562.

[22] Peitz, D., Kersjes, C., Thom, J., Hoelling, H., & Mauz, E. (2021). Indicators for public mental health: a scoping review. Frontiers in Public Health, 9, 714497.

[23] BinDhim, N. F., Althumiri, N. A., Basyouni, M. H., Alageel, A. A., Alghnam, S., Al-Qunaibet, A. M., ... & Ad-Dab'bagh, Y. (2020). A mental health surveillance system for the general population during the COVID-19 pandemic: protocol for a multiwave cross-sectional survey study. JMIR research protocols, 9(11), e23748.

[24] Higuchi, M., Nakamura, M., Shinohara, S., Omiya, Y., Takano, T., Mitsuyoshi, S., & Tokuno, S. (2020). Effectiveness of a voice-based mental health evaluation system for mobile devices: prospective study. JMIR formative research, 4(7), e16455.

[25] Yang, H., & Liu, Q. (2021). Innovative research of dynamic monitoring system of mental health vocational students based on big data. Personal and Ubiquitous Computing, 1-14.

[26] Dong, Z., Ji, X., Lai, C. S., Qi, D., Zhou, G., & Lai, L. L. (2022). Memristor-based hierarchical attention network for multimodal affective computing in mental health monitoring. IEEE Consumer Electronics Magazine, 12(4), 94-106.

[27] Chen, M., Shen, K., Wang, R., Miao, Y., Jiang, Y., Hwang, K., ... & Liu, Z. (2022). Negative information measurement at AI edge: A new perspective for mental health monitoring. ACM Transactions on Internet Technology (TOIT), 22(3), 1-16.

[28] Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. International Journal of Information Management Data Insights, 2(2), 100113.

[29] Mohsen, B., Al-Khaleel, O., & Alfadhly, A. (2024). Mental Health Monitoring and Detection Based on Machine Learning and IoT Data. International Journal of Information Technology & Decision Making, 23(04), 1449-1483.

[30] Shi, Q., Cai, N., & Jiao, W. (2022). Monitoring and Evaluating College Students' Mental Health Based on Big Data Analysis. American Journal of Health Behavior, 46(2), 164-176.

[31] Huang, P., & Deng, W. (2022). Application of improved apriori algorithm in mental health monitoring. Psychiatria Danubina, 34(suppl 5), 310-310.

[32] Chakraborty, A., Banerjee, J. S., Bhadra, R., Dutta, A., Ganguly, S., Das, D., ... & Saha, G. (2024). A framework of intelligent mental health monitoring in smart cities and societies. IETE Journal of Research, 70(2), 1328-1341.

[33] Liu, M. (2024). An IoT-Enabled Mental Health Monitoring System for English Language Students Using Generative Adversarial Network Algorithm. Mobile Networks and Applications, 1-22.

[34] Podder, T., Bhattacharya, D., & Majumdar, A. (2022). Dew computing-inspired mental health monitoring system framework powered by a lightweight CNN. In Disruptive Technologies for Big Data and Cloud Applications: Proceedings of ICBDCC 2021 (pp. 309-319). Singapore: Springer Nature Singapore.

[35] VENKATARAMANAIAH, B., REDDY, K. G., RAMAKRISHNAN, K., SATHISH, M., & RAJESH, S. (2024, August). Real Time Mental Health Monitoring System using Machine Learning. In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI) (pp. 806-811). IEEE.

[36] Wason, R., Rana, P. S., & Jain, V. (2021). Real-Time Mental Health Monitoring: The Indian Scenario. In Research Anthology on Mental Health Stigma, Education, and Treatment (pp. 305-314). IGI Global.

[37] Jiao, C. (2024). A mental health monitoring system based on intelligent algorithms and biosensors: Algorithm behaviour analysis and intervention strategies. Molecular & Cellular Biomechanics, 21(1), 357-357.

[38] Saini, H., Rathee, G., & Saini, D. K. (Eds.). (2020). Large-scale Data Streaming, Processing, and Blockchain Security. IGI Global.

[39] Panpan Li & Feng Liang. (2023). Construction of mental health monitoring system based on model transfer learning algorithm. International Journal of Wireless and Mobile Computing(1),58-65.

[40] Hongjian Li,Junlin Li,Xiaolin Duan & Jianglin Xia. (2025). Energy-aware scheduling and two-tier coordinated load balancing for streaming applications in apache flink. Future Generation Computer Systems107681-107681.

[41] Fatemeh Nori,Mahmood Deypir & Mohamad Hadi Sadreddini. (2013). A sliding window based algorithm for frequent closed itemset mining over data streams. The Journal of Systems & Software(3),615-623.

[42] Shiping Du,Houhui Huang & Yuming Wei. (2013). The Baum-Welch Algorithm of Mixture Coupled Hidden Markov Model. Journal of Convergence Information Technology(1),1-8.