

Exploring the Algorithm for Diagnosing Burnout of Counselors in Higher Vocational Colleges in Guangdong Province Based on Emotion Recognition

Rili Dang¹ and Noorazman Abd Samad²

Abstract

The main objective of the research is to deal with the drawbacks of conventional self-report tests by designing and validating a novel Emotion Recognition (ER) algorithm for assessing Counsellor Burnout (CB) in Higher Vocational Colleges (HVC) in Guangdong Province. Applying the Maslach Burnout Inventory (MBI) Survey Questionnaires, voice and facial expression files, and text analysis of counselling sessions, researchers collected an enormous dataset from 97 counsellors. The study employed a multimodal fusion layer to integrate multiple neural network architectures, namely RNNs for audio content, CNNs for video content, and a BERT+LSTM hybrid for textual analysis. This enabled a thorough examination of each ER. By dramatically exhibiting an accuracy of more than 92% for identifying the symptoms of burnout, this method is better than the conventional methods that have been used in previous years. The technique, which successfully recognizes complex, real-time emotions that are frequently ignored when individuals communicate emotions in person, can be used to obtain a more accurate and expert evaluation of burnout because of its capacity to recognize these reactions accurately. Different sources of data and contemporary Machine Learning (ML) methods have enhanced CB prediction and observation. This approach enhances counsellors' emotional stability and assists in developing pedagogical mental health interventions. Findings demonstrate that ML techniques may enhance mental assessment and counselling in demanding employment.

Keywords: Emotion Recognition, Maslach Burnout Inventory, Machine Learning, CNN, Mental Health Diagnostics, Accuracy

INTRODUCTION

There is an essential concern in colleges and universities, especially Higher Vocational Colleges (HVC), that is commonly referred to as Counsellor Burnout (CB). This problem is triggered by more significant emotional stress and the high demands of counselling jobs. One of the primary tasks of counsellors employed by colleges and universities is to provide support to students in dealing with the challenges that they experience in both their academic and personal lives and also manage the psychological stress of the students they propose.

Emotional dissatisfaction, dehumanization, and decreased individual accomplishment—essential of burnout—may be caused by the approach. A deeper investigation of the prevalence and impact of burnout has been motivated by the seriousness and prevalence of problems in Guangdong Province, where the counsellor-student proportion is frequently stressed.

Standard methods for recognizing burnout mainly involve expressed Survey Questionnaires (SQ), which have a fundamental bias due to inherent subjective features. On the uppermost point, these methods can occasionally not be up to the job of recording mental states in the present moment, let alone allowing continuous monitoring or prior symptom detection for burnout. If we care about the mental health and sustainability of counsellors' jobs, we must find a more reliable and objective assessment method.

The Maslach Burnout Inventory (MBI) is one of the primary tests for mental health used in contemporary techniques; it defines burnout into three categories: Emotional Exhaustion (EE), Depersonalization Disorder (DD), or Personal Accomplishment (PA). The static layout of devices and the naturally occurring lag between feeling issues and notifying them are drawbacks, considering their importance. Furthermore, these devices fail to study the reality that human feelings fluctuate daily or that non-verbal behaviour variations may indicate signs of stress or burnout even before an individual becomes conscious of their current state.

¹ Universiti Tun Hussein Onn Malaysia, Panchor, 84600, Johor, Malaysia. E-mail: ddll84600@outlook.com (Corresponding Author)

² Universiti Tun Hussein Onn Malaysia, Panchor, 84600, Johor, Malaysia E-mail: noorazman@uthm.edu.my

The Emotion Recognition (ER) method, which is deemed to be the most sophisticated, has been selected for the study that has been suggested for assessing CB in HVCs situated in Guangdong Province, China. An extensive data set from 97 counsellors has been integrated into this study, along with a selection of methods, such as SQ predicated on the MBI, voice and FE tests, and text analysis. The objective of these studies is to achieve a comprehensive understanding of psychological states and burnout symptoms. The layered method uses contemporary Neural Network (NN) models, especially Recurrent Neural Networks (RNNs) for data that include audio, Convolutional Neural Networks (CNNs) for video content, and a BERT+LSTM hybrid for textual analysis, in order to collect this data. A multimodal fusion layer assures that the data being investigated is thoroughly investigated.

An in-depth assessment of the SQ tools' reliability and accuracy is part of the approach, thus ensuring reliable and functionally valuable outcomes. This work aims to assist colleges and universities in comprehending counsellor's psychological well-being by analysing verbal and non-verbal signs of stress. This will enable organizations to find assistance and encouragement methods that are more successful. Demonstrating the importance of Machine Learning (ML) methods for mental health evaluation and intervention planning, primary results illustrate that the framework is highly successful, with over 92% accuracy in recognizing burnout symptoms.

An overview of the available literature is given in Section 2 of the article. This is after an explanation of the work that is suggested in Section 3, an evaluation of the framework in Section 4, and a conclusion and recommendations for additional studies in Section 5 of the manuscript.

LITERATURE REVIEW

A study of burnout among counsellors has gained extensive focus from educators, an example of the complexity and importance of the issue in psychological wellness. Numerous studies have provided a more profound comprehension of what variables predict burnout, the implications of burnout, and the variables that reduce burnout. Burnout is prevalent in jobs that are marked by high psychological work, such as counselling.

In the case of counselling psychology, PhD students [1] carried out research that examined the associations between stress, social support, and burnout. Based on the results of their investigation, social encouragement and a psychological Sense of Community (SOC) did not significantly impact the association between stress and burnout; yet, both variables were key indicators of burnout and job satisfaction. It is pertinent to highlight that higher SoC was correlated with higher job satisfaction under instances of minimal stress. This may indicate that social support is key in minimizing burnout within specific circumstances.

In the present study [2], the study of burnout was expanded to include correctional counsellors, and the Counsellor Burnout Inventory (CBI) was used to assess the tool's value in this specific setting. Researchers have proven that the CBI is an accurate tool for measuring the state of burnout noticed by correctional counsellors. It is important to note that the study demonstrated that levels of burnout were not significantly impacted by gender or in terms of the security of the prison. This result emphasizes the widespread reach of burnout across all kinds of job circumstances within the legal system.

The present investigation addressed trainee-certified counsellors and used content analysis to find symptoms of burnout in the counselling profession. Their study led to the development of the current research collection by identifying burnout symptoms that are often experienced and reported less frequently. These symptoms include negative emotional experiences, exhaustion, and queries regarding career choice. When developing support mechanisms for counsellors, these research results emphasize the significance of paying attention to many burnout symptoms [4].

The research paper [5] highlighted the broader implications of burnout among fully trained counsellors and mental health professionals, focusing a value on the global increase in mental health issues and the problems that front-line mental health workers face. The researchers present an in-depth review of the scientific research on burnout that has been presented across various backgrounds, as well as ideas to improve practitioners' psychological well-being and health. The study highlights the global challenge of ensuring mental health experts are connected to adequate support and education to prevent burnout [6].

A comprehensive review of the recognition and management of professional burnout in the Netherlands has been provided in [7]. This paper identifies a medical condition as "clinical burnout," a medical disorder accepted when people quit their jobs due to symptoms they face. The study outlines various treatment phases, including crisis management, recovery, prevention, and post-burnout growth, emphasizing the biological underpinnings of the stress response. This perspective challenges the distinction between work-related and personal stress, advocating for a holistic view of stress management in clinical settings [8].

[9] investigate burnout among school psychologists in Israel, focusing on the interplay between SoC and loneliness as factors manipulating burnout risk. Their findings suggest that while SoC generally buffers against burnout, its protective effects diminish significantly under conditions of high loneliness. This study [10] underscores the importance of social connections and personal resilience as crucial elements in mitigating burnout among professionals engaged in emotionally demanding roles.

[11] analyze burnout among occupational therapists in Turkey, correlating it with job satisfaction, work engagement, and working conditions. The study [12] identifies a substantial portion of therapists experiencing burnout symptoms, with negative correlations between burnout, job satisfaction, and supportive working conditions. The findings highlight the critical role of workplace environment and job engagement in preventing burnout, suggesting that improved management practices and skill development opportunities could mitigate these risks.

[13-15] delve into the relationships between personality traits, work engagement, alexithymia, and burnout among village doctors in China. Their comprehensive statistical analysis reveals that personality significantly influences both work engagement and alexithymia, which distresses burnout levels. This study illustrates the complex interactions between individual characteristics and professional engagement, providing insights into the pathways through which personality and emotional processing skills impact burnout.

PROPOSED MODEL

Data Collection and Processing

The evidence that was collected from 97 counsellors assigned to 14 different HVCs across Guangdong Province was used for training and testing the ER algorithm, as demonstrated in Figure 1. The group in question consists of members who present many physical and behavioural features, such as gender, age, years of counselling background, and domains of expertise. In order to develop a greater awareness of the symptoms of burnout and the psychological conditions that counsellors are training, the collection method comprised of performing interviews and SQ, capturing voice and facial expressions, and using text analysis. Comprehensive data about the demographic profile of the community can be obtained from Table 1:

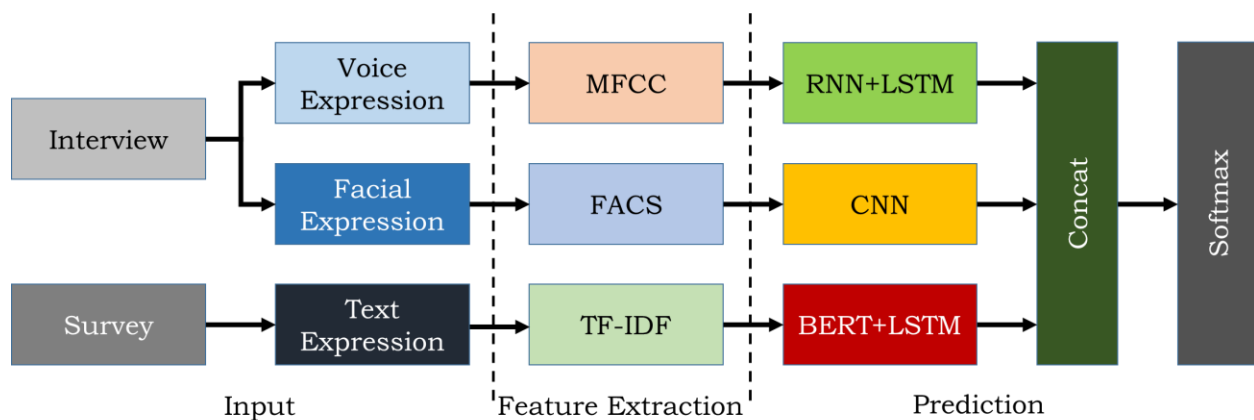


Figure 1: ER algorithm architecture

Table 1: Population demographics

Demographic Variable	Category	Count	Percentage (%)
Gender	Male	42	43.3
	Female	55	56.7
Age Group	Under 30	24	24.7
	30-39	39	40.2
	40-49	21	21.6
	50 and above	13	13.5
Years of Counseling Experience	Less than 5	29	29.9
	5-10	38	39.2
	Over 10	30	30.9
Specialization	Mental Health	33	34.0
	Career Guidance	28	28.9
	Student Affairs	21	21.6
	Other	15	15.5

Surveys and Interviews

The disclosed burnout symptoms of counsellors were assessed carefully using a combination of SQ and interviews. The conceptual framework of the SQ was founded on the recognized MBI, which was augmented by new context-specific questions which had been customized to the particular problems that counsellors in HVC experience. EE, DD, and PA were the main classifications utilized to group the 22 SQ comprised in the online SQ. A Likert scale with 7 points was implemented to record answers, with the scale covering "Never" to "Every day."

Informal interviews were conducted as a substitute for SQ to collect a more excellent knowledge of the counsellors' experiences. Each interview lasted between 30 and 45 minutes and explored critical areas such as job satisfaction, workload, emotional challenges, and coping mechanisms. These interviews provided valuable qualitative insights, with transcripts being analyzed using thematic analysis to identify recurring patterns and themes related to burnout. The following tables, from 2 to 4, present the questions used in the survey and interview.

Table 2: MBI Questions (Core Items)

Domain	Question	Scale (1-7)
EE	1. I feel emotionally drained from my work.	
	2. I feel used up at the end of the workday.	
	3. I feel fatigued when I get up in the morning and face another day on the job.	
	4. Working with people all day is a strain for me.	
	5. I feel burned out from my work.	
	6. I feel frustrated by my job.	
	7. I feel I'm working too hard on my job.	
	8. Working with people directly puts too much stress on me.	
	9. I feel like I'm at the end of my rope.	
DD	10. I feel I treat some recipients as if they were impersonal objects.	
	11. I've become more callous toward people since I took this job.	
	12. I worry that this job is hardening me emotionally.	
	13. I don't care what happens to some recipients.	
	14. I feel recipients blame me for some of their problems.	
PA	15. I can easily understand how my recipients feel about things.	
	16. I deal very effectively with the problems of my recipients.	
	17. I feel I'm positively influencing other people's lives through my work.	
	18. I feel very energetic.	
	19. I can easily create a relaxed atmosphere with my recipients.	
	20. I feel exhilarated after working closely with my recipients.	
	21. I have accomplished many worthwhile things in this job.	
	22. In my work, I deal with emotional problems very calmly.	

Table 3: Additional Context-Specific Questions for Counselors in HVC

Domain	Question	Scale (1-7)
Workload and Job Pressure	1. I have too many administrative tasks besides my counselling responsibilities.	
	2. The pressure to deliver immediate results adds to my stress.	
Student-Related Challenges	3. Handling emotionally unstable students is becoming increasingly challenging.	
	4. I frequently encounter student problems that are beyond my control.	

Organizational Support and Resources	5. My institution provides adequate resources to support my counselling work.
---	---

Table 4: Semi-Structured Interview Guide

Objective	Questions
Explore counsellors' experiences in HVC regarding job satisfaction, workload, emotional challenges, and coping mechanisms.	1. Can you describe your typical workday as a counsellor in a HVC?
	2. What aspects of your job do you find most fulfilling?
	3. What are the biggest challenges you face in your counselling role?
	4. How do you perceive the current workload in your job? Is it manageable?
	5. How do you usually cope with stress or EE in your role?

Interview records were analyzed using thematic analysis to identify recurring patterns and themes related to burnout. The coding process was iterative and collaborative, involving multiple researchers to ensure a comprehensive interpretation of the data.

Validity Analysis of the SQ

A systematic validity analysis was shown to ensure the reliability and validity of the SQ used for diagnosing burnout among counsellors, encompassing face, content, and construct validity.

Face Validity (FV)

FV refers to the degree to which a test or assessment appears effective regarding its stated aims. To establish FV, the SQ was reviewed by a panel of experts, including three psychologists specializing in CB and two educational researchers with extensive experience in HVC counselling. The experts assessed the appropriateness and relevance of the SQ items, confirming that they appeared suitable for diagnosing CB in HVC. As a result, the SQ was deemed to possess high FV.

Content Validity (CV)

CV examines (Tab. 5) whether the test covers a representative sample of the domain it is supposed to measure. The **Content Validity Index (CVI)** was used to quantify the degree to which individual items and the entire SQ represented the burnout construct. Experts rated each item on a 4-point scale (1 = Not relevant, 4 = Highly relevant), and the **Item-Level CVI (I-CVI)** was calculated by dividing the number of experts, giving a rating of 3 or 4 by the total number of experts. The **Scale-Level CVI (S-CVI)** was obtained by averaging all I-CVI scores.

The results showed that the EE domain had an I-CVI range of 0.83 to 1.00, the DD domain ranged from 0.80 to 1.00, and the PA domain ranged from 0.85 to 1.00. The overall scale-level CVI was 0.94, indicating strong content validity for the entire SQ.

Table 5: CV scores

Domain	Number of Items	I-CVI Range	S-CVI
EE	9	0.83-1.00	0.94
DD	5	0.80-1.00	0.92
PA	8	0.85-1.00	0.96
Overall Scale	22		0.94

Construct Validity

Construct validity (Tab. 6) assesses whether the SQ accurately measures the concept of burnout as defined in theoretical terms. An exploratory factor analysis (EFA) was conducted on a pilot sample of 50 counsellors to confirm construct validity. The Kaiser-Meyer-Olkin (KMO) test yielded a value of 0.89, indicating sampling appropriateness, and Bartlett's Test of Sphericity ($\chi^2(231)=874.26, p<0.001$) confirmed that the dataset was suitable for factor analysis.

The factor analysis extracted three factors with eigenvalues above 1, corresponding to EE, DD, and PA domains. The factor loadings ranged from 0.71 to 0.89 for EE, 0.68 to 0.85 for DD, and 0.73 to 0.92 for PA.

It proved that the hierarchy of factors agreed with the theoretical requirements, signifying a high construct validity level.

Table 6: The score for construct validity

Domain	Item Numbers	Factor Loading Range
EE	1-9	0.71-0.89
DD	10-14	0.68-0.85
PA	15-22	0.73-0.92

Reliability Analysis

In order to assess the level to which the SQ was reliable, Cronbach's alpha was determined for every person subject as well as the aggregate index. From the study results, it was exposed that the field of EE had a Cronbach's alpha value of 0.88, the subject of DD had an index of 0.82, and the subject of PA had a value of 0.85. Cronbach's alpha is assumed to be 0.91 for the complete size, which suggests that the SQ has had very high internal consistency throughout its history.

Table 7: Reliability score

Domain	Number of Items	Cronbach's Alpha
EE	9	0.88
DD	5	0.82
PA	8	0.85
Overall Scale	22	0.91

Given the validity and reliability analysis results, the SQ is a reliable tool to measure the degree of burnout experienced by counsellors employed in HVC in Guangdong. It presents an in-depth knowledge of CB's EE, DD, and PA phases.

VE and FE Recordings

VE and FE samples were obtained between the counselling process and focus group interviews in order to record the vocal tone and FE that are predictive of burnout symptoms between counsellors. The recordings were highly beneficial in assessing non-verbal cues that might disclose emotional states connected with burnout.

Voice Recordings

To provide accuracy and reliability, recordings of voices were captured using high-quality online recorders and stored in a monitored environment. Following every session of counselling, which was divided into five-minute clips, the Geneva Emotion Wheel (GEW) paradigm was used to develop assessments for all of the observed emotional states. The model we address here presents an extensive classification of emotions that relies on fundamental emotional states.

A selection of advanced techniques for extracting features has been used in the analysis of the audio evidence:

Mel-Frequency Cepstral Coefficients (MFCCs): In order to precisely record the features of voice sounds, ranging coefficients such as these, which describe the short-term power spectrum of noise, are essential.

Chroma Features: These features can identify the 12 distinct tone classes, which also support the recognition of the rhythmic pattern in the musical data.

Spectral Contrast: The voice's quality can be assessed by evaluating the variance in amplitude between the highest and lowest points in the audio frequencies. This includes data on the tone of the speech.

Tonnetz Features: These features are labelled as the musical links between tones.

Collecting these features aims to recognize speech differences that coincide with emotional states such as exhaustion, disappointment, and EE.

Facial Expression Recordings

Videos of FE were made by deploying high-resolution cameras closest to the eyes in order to record even the shortest minor of FE. To verify that the counsellors' emotions were real, they were counselled to retain their genuine relationships through video recording events.

To correlate Facial Action units (AUs) to specific emotional states, the video clips were captioned with the support of the Facial Action Coding System (FACS). Below are the principal facial features that were investigated:

Facial Landmarks: 68 key points on the face were detected to analyze facial geometry.

Optical Flow Vectors: Motion patterns of facial features were tracked to detect micro-expressions.

Head Pose and Eye Gaze Direction: These were measured to identify head movements and shifts in gaze, which could indicate emotional responses.

Blink Rate and Smile Intensity: Readings were performed to assess the degree of EE and DD, the number of flashes and the strength of facial features.

The extraction of features from FE files was performed with the help of advanced ML modules such as OpenFace and Dlib, which donated defined facial features and FE.

Data Annotation and Integration

In order to set up ground truth for the tasks of training and validation, all of the compiled VE and FE data were manually labelled by experts in the domains of mental health and data science. Several mental labels, including joy, sorrow, rage, anxiety, shock, dissatisfaction, and neutral position, have been included in the captions. Applying Cohen's Kappa, the annotation tools attained an inter-rater validity of 0.82, which verified that the labelling corresponded across all paradigms.

The annotated audio and video data were merged into a unified dataset using unique identifiers for each counselling session or participant. The data was then preprocessed to ensure compatibility with the emotion recognition algorithm:

Audio Data: Recordings were resampled to 16 kHz and normalized for consistent amplitude.

Video Frames: Frames were resized to 224x224 pixels, and facial landmarks were standardized.

The measurements of VE and FE were a vital part of the dataset used to train and test the ER method. These recordings provided helpful information into the emotional states of counsellors and the non-verbal signals that correspond to burnout symptoms.

Dataset Features

The data set contains a selection of methods, resulting in an in-depth knowledge of the nature of CB. More than 200 hours of files from counselling sessions and discussions in focus groups are contained in the audio and video data. The files were captured with high-resolution cameras and online recorders using the best technology. In order to simplify annotation based on the Geneva Emotion Wheel (GEW) basis, each session was divided into clips of video that were five minutes long. By implementing the FACS, FE were captured and studied.

Additionally, 3,432 text documents were collected, including emails, student progress reports, and chat transcripts. The text data underwent preprocessing and sentiment analysis to identify emotional states. Psychology and data science experts annotated the dataset to establish ground truth for training and validation. Tags included Joy, sorrow, rage, anxiety, shock, dissatisfaction, and neutral position. The following Table 8 illustrates the distribution of emotions:

Table 8: Distribution of emotion

Emotion	Count	Percentage (%)
Joy	6,432	21.8%
Sorrow	4,215	14.3%
Rage	3,142	10.6%
Anxiety	2,381	8.1%
Shock	2,113	7.2%
Dissatisfaction	1,798	6.1%
Neutral Position	9,942	33.9%

All collected data were manually annotated, achieving an inter-rater reliability of 0.82 (Cohen's Kappa). The annotated audio and video data were merged into a unified dataset using unique identifiers for each counselling session. Audio recordings were resampled to 16 kHz and normalized, while video frames were resized to 224x224 pixels and facial landmarks standardized. Text data was tokenized and vectorized using pre-trained word embeddings.

Feature Extraction

Audio Features

MFCC and chroma features were extracted for audio data to capture vocal features associated with emotional states. The process began with pre-emphasis, which filtered the audio signal to emphasize higher frequencies by reducing the amplitude of lower frequencies. This was achieved using the EQU (1):

$$y[n] = x[n] - \alpha \cdot x[n - 1], \alpha \approx 0.97 \quad (1)$$

Following pre-emphasis, the audio signal was divided into frames of 25 ms each, with a 10 ms overlap. A Hanning window was applied to each frame to reduce spectral leakage, EQU (2):

$$w[n] = 0.5 \cdot \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \quad (2)$$

Each frame was then converted into the frequency domain using the Fast Fourier Transform (FFT), and the resulting power spectrum was passed through a Mel-filter bank that mapped frequencies to the Mel scale, EQU (3).

$$f_{\text{mel}} = 2595 \cdot \log_{10}\left(1 + \frac{f_{\text{Hz}}}{700}\right) \quad (3)$$

After filtering, the logarithmic power of each Mel-filtered signal was computed, and the Discrete Cosine Transform (DCT) was applied to obtain the cepstral coefficients, EQU (4).

$$\text{MFCC}_k = \sum_{m=0}^{M-1} \log(S_m) \cdot \cos\left[\frac{\pi k(m+0.5)}{M}\right], k = 1, 2, \dots, K \quad (4)$$

In addition to MFCCs, chroma features were extracted to capture the harmonic structure of the audio signal by representing the 12 different pitch classes. First, the Short-Time Fourier Transform (STFT) was computed for each frame to obtain the magnitude spectrogram. Each frequency bin was then mapped to one of the 12 pitch classes, and the sum of energies was calculated for each class. Finally, the resulting chroma features were normalized to the unit norm.

Video Features

FE were analyzed using the FACS and optical flow analysis. The FACS encodes FE by categorizing them into distinct Action Units (AUs). The process began with detecting 68 facial landmarks using a landmark detection algorithm like Dlib. An ML model, a Convolutional Neural Network (CNN), classified each frame into relevant AUs based on these detected landmarks. Each AU was assigned an intensity score ranging from 0 (absent) to 5 (maximum intensity).

Optical flow analysis captured the motion of facial features over time using the chargeback method, which estimates the motion vector for each pixel between consecutive frames. The magnitude and direction of the motion were then calculated as follows: EQU (5).

$$\text{Magnitude} = \sqrt{u^2 + v^2}, \text{ Direction} = \arctan\left(\frac{v}{u}\right) \quad (5)$$

where u and v represent the horizontal and vertical components of the flow vector, respectively. Motion features were aggregated over a fixed temporal window to form a feature vector.

Text Features

Text data was analyzed for sentiment and emotional cues using Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings. TF-IDF represents the importance of words in a document relative to the entire corpus. Term Frequency (TF) is calculated as follows: EQU (6)

$$\text{TF}(t, d) = \frac{\text{Count of term } t \text{ in document } d}{\text{Total terms in document } d} \quad (6)$$

Inverse Document Frequency (IDF) measures how frequently a term appears across all documents, EQU (7).

$$\text{IDF}(t) = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing term } t}\right) \quad (7)$$

The final TF-IDF score is calculated as EQU (8).

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \quad (8)$$

Word embeddings capture semantic relationships between words by mapping them to high-dimensional vectors. Pretrained models like Word2Vec or GloVe were used to convert words into vectors, and sentence embeddings were formed as the mean or weighted average of individual word vectors.

Model Architecture

The emotion recognition algorithm utilizes specialized NN models for each modality audio, video, and text to diagnose burnout among counsellors. The architecture of each model is designed to extract high-level features indicative of emotional states.

Audio Model Architecture

An RNN+LSTM unit was used to capture the temporal patterns in audio features such as MFCC and chroma. The input layer accepts a sequence of feature vectors ' x_t ', each representing audio features extracted from a specific time frame ' t '.

The model's core is the LSTM layer, which models temporal dependencies using input, forget, and output gates, EQU (9) to EQU (14).

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (9)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (10)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad (11)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C) \quad (12)$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \quad (13)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \quad (14)$$

where:

\mathbf{i}_t : Input gate

\mathbf{f}_t : Forget gate

\mathbf{o}_t : Output gate

\mathbf{C}_t : Cell state

\mathbf{h}_t : Hidden state

The final hidden state \mathbf{h}_T is passed to a fully connected layer, which produces the classification output using a SoftMax function, EQU (15).

$$\mathbf{y} = \text{SoftMax}(\mathbf{W}_{fc} \cdot \mathbf{h}_T + \mathbf{b}_{fc}) \quad (15)$$

Video Model Architecture (CNN)

FE was processed using CNN, which effectively captures spatial features. The input layer accepts a sequence of image frames \mathbf{I}_t representing FE. The convolutional layers with ReLU activation extract spatial features, while max pooling reduces the spatial dimensions, EQU (16) and EQU (17).

$$\mathbf{F}_{l+1} = \text{ReLU}(\mathbf{W}_l * \mathbf{F}_l + \mathbf{b}_l) \quad (16)$$

$$\mathbf{P}_l = \text{MaxPool}(\mathbf{F}_l) \quad (17)$$

where:

\mathbf{W}_l : Convolutional filters

\mathbf{F}_l : Input feature map

\mathbf{b}_l : Bias term

$*$: Convolution operation

The pooled feature maps are then flattened and passed to a fully connected layer for classification using a SoftMax function, EQU (18).

$$\mathbf{y} = \text{softmax}(\mathbf{W}_{fc} \cdot \mathbf{F}_{\text{flattened}} + \mathbf{b}_{fc}) \quad (18)$$

Text Model Architecture (BERT + LSTM Networks)

Text data was analyzed using a Bidirectional Encoder Representations from Transformers (BERT) model and an LSTM network. BERT captures contextual data through attention mechanisms, while LSTM models sequential patterns in text sentiment.

BERT Embedding Layer: The BERT model generates contextual word embeddings for each word in the input sequence: $\mathbf{E}_t = \text{BERT}(\mathbf{W}_t)$ where \mathbf{W}_t represents the word tokens.

LSTM Layer: An LSTM layer processes the sequence of BERT embeddings, EQU (19) to EQU (24).

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{E}_t] + \mathbf{b}_i) \quad (19)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{E}_t] + \mathbf{b}_f) \quad (20)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{E}_t] + \mathbf{b}_o) \quad (21)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{E}_t] + \mathbf{b}_c) \quad (22)$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \quad (23)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \quad (24)$$

Fully Connected Layer: The final hidden state \mathbf{h}_T is passed to a fully connected layer for sentiment classification using a SoftMax function, EQU (25).

$$\mathbf{y} = \text{SoftMax}(\mathbf{W}_{fc} \cdot \mathbf{h}_T + \mathbf{b}_{fc}) \quad (25)$$

Multimodal Fusion Layer: The outputs from the audio, video, and text models are combined using a fusion layer. Each model provides a feature vector \mathbf{f}_A , \mathbf{f}_V , and \mathbf{f}_T for audio, video, and text data, respectively. These vectors are concatenated into a single feature vector, EQU (26)

$$\mathbf{f}_{\text{concat}} = [\mathbf{f}_A, \mathbf{f}_V, \mathbf{f}_T] \quad (26)$$

A fully connected layer, EQU (27), is subsequently employed to execute the end classification following the coupled feature vector sent through it.

$$\mathbf{y} = \text{SoftMax}(\mathbf{W}_{fc} \cdot \mathbf{f}_{\text{concat}} + \mathbf{b}_{fc}) \quad (27)$$

An effective system for recognizing burnout-related psychological conditions across various fields can be obtained from the model that emerged as an outcome. This system promises that counsellors employed in HVC environments can identify burnout accurately.

EXPERIMENTAL SETUP, METRICS, AND HYPERPARAMETERS FOR MODEL TRAINING

Experimental Setup

For training and testing the ER method, the experimental design involved capturing, preliminary processing, and implementing data that included audio, video, and text. Several components of hardware and software were used in the study, which was performed in a monitored atmosphere:

Hardware: The processing power needed to perform training Deep Learning (DL) models was made available by an Intel Xeon processor with 16 cores, an NVIDIA Tesla V100 graphics processing unit, 32 GB of RAM, and a 2 TB Solid State Drive.

Software: Python 3.8 was implemented to build and train the models, and it was designed and trained on Linux Ubuntu 20.04. The primary DL models were TensorFlow 2. x and PyTorch 1. x. Supplementary libraries such as OpenCV, Dlib, Librosa, NLTK, and Hugging Face Transformers have assisted with specific tasks.

Below are the settings that were employed to train the model: the dataset was split into training sets (70%), validation sets (15%), and test sets (15%) (Tab. 9).

Table 9: Model hyperparameters for training

Model	Hyperparameter	Value
Audio Model (RNN)	Model Type	RNN + LSTM units
	Batch Size	64
	Learning Rate	0.001
	LSTM Units	128
	Epochs	50
	Dropout Rate	0.3
	Optimizer	Adam
Video Model (CNN)	Model Type	CNN
	Batch Size	32
	Learning Rate	0.0001
	Convolutional Filters	[32, 64, 128]
	Fully Connected Units	512
	Epochs	50
	Dropout Rate	0.4
Text Model (BERT + LSTM)	Optimizer	Adam
	Model Type	BERT+LSTM
	Batch Size	16
	Learning Rate	2e-5
	LSTM Units	64

	Epochs	50
	Dropout Rate	0.2
	Optimizer	AdamW
Multimodal Fusion Layer	Fully Connected Units	256
	Dropout Rate	0.5
	Optimizer	Adam

Metrics

The performance of the ER algorithm was evaluated using several metrics to capture the accuracy and quality of predictions, EQU (28) to EQU (31).

Accuracy measures the proportion of correctly classified samples:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (28)$$

Precision calculates the ratio of True Positives (TP) to the sum of TP and False Positives (FP):

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (29)$$

Recall (Sensitivity) measures the ratio of TP to the sum of TP and False Negatives (FN).

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (30)$$

F1Score is the harmonic mean of precision and recall:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (31)$$

Confusion Matrix visualizes the performance of the classification model by showing the number of correct and incorrect predictions for each class.

Evaluation Results across Epochs

The results of the ER algorithm across 50 epochs are summarized in Tab. 10 below. Metrics include Accuracy, Precision, Recall, and F1-score. Results are presented at 5-epoch intervals.

Table 10: Performance analysis for metrics vs epochs

Epochs	Accuracy	Precision	Recall	F1-score
5	72.8%	74.1%	71.6%	72.8%
10	78.4%	79.7%	76.9%	78.3%
15	83.6%	84.8%	82.3%	83.5%
20	86.7%	88.0%	85.2%	86.6%
25	88.3%	89.7%	87.0%	88.2%
30	89.8%	91.2%	88.5%	89.8%
35	90.4%	91.9%	89.0%	90.4%
40	91.2%	92.6%	90.2%	91.3%
45	91.9%	93.3%	90.8%	91.9%
50	92.5%	93.7%	91.8%	92.7%

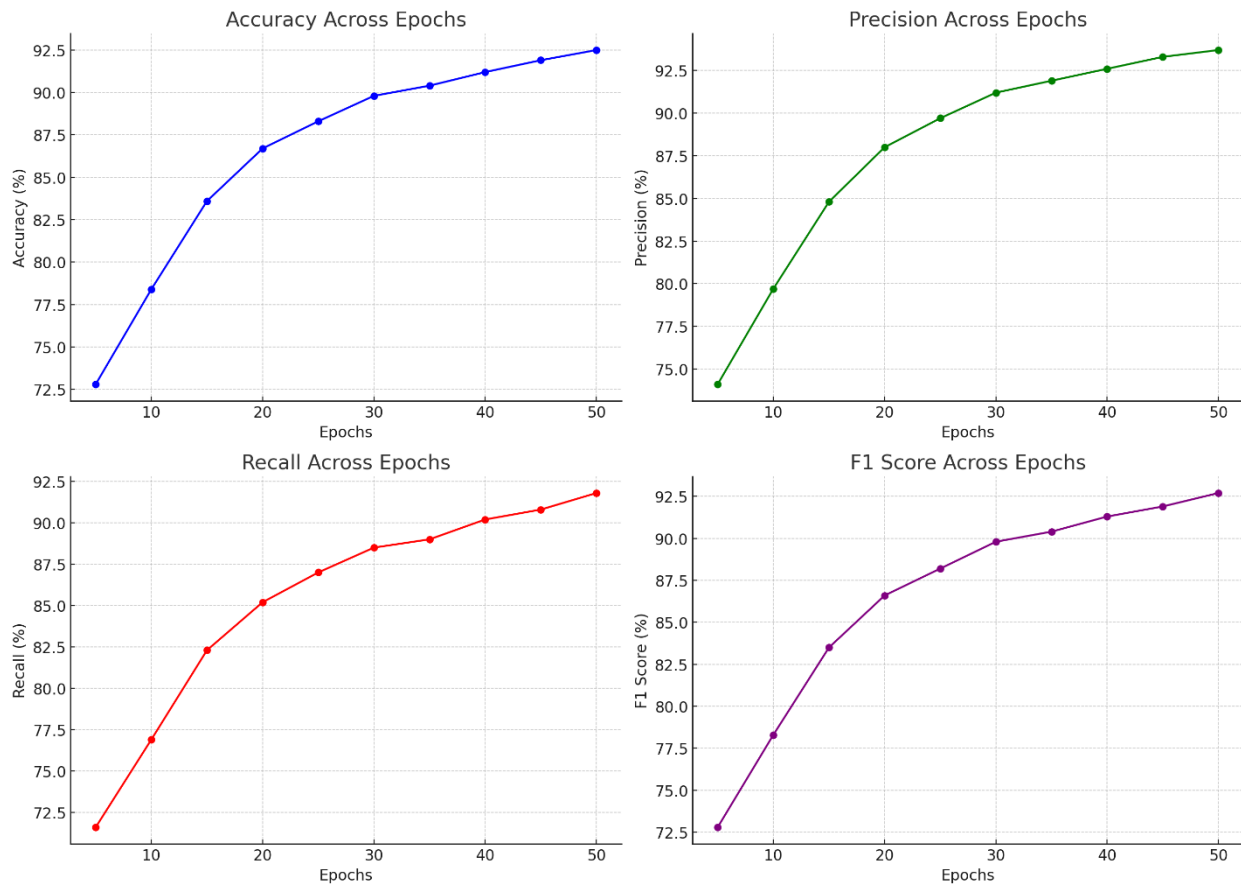


Figure 2: Performance of the model for different metrics

The evaluation metrics, including Accuracy, Precision, Recall, and F1-score, as shown in Fig. 2, show a clear improvement in the performance of the ER algorithm across 50 epochs.

In the 1st several epochs, the model showed an average level of execution, with an Accuracy of 72.8%, a Precision of 74.1%, a Recall of 71.6%, and an F1-score of 72.8% at the end of the 5th epoch. It shows that the system could classify emotions with an acceptable level of accuracy; yet, due to its shortage of adequate training, it overlooked an essential percentage of examples in which emotions were miscategorized. Each parameter substantially increased as training advanced to the 10th epoch, with Accuracy achieving 78.4%, Precision achieving 79.7%, Recall achieving 76.9%, and F1-score achieving 78.3%. The improvement results from the model's better ability to be educated and recognized between several emotional states.

The accuracy achieved 83.6%, Precision achieved 84.8%, Recall achieved 82.3%, and F1-score achieved 83.5% at the fifteenth epoch. The performance indicators evolved to improve throughout the experiment. A fall in the percentage of classification errors can be caused by an indication that the model has started to apply effectively over every classification, as shown by an improvement in all parameters. The model's results had significantly improved by the twentieth epoch, attaining an accuracy of 86.7%, precision of 88.0%, recall of 85.2%, and F1-score of 86.6%.

After the twenty-fifth epoch, this behaviour was maintained, and the accuracy hit 88.3%, the precision achieved 89.7%, the recall achieved 87.0%, and the F1-score achieved 88.2%. The results for all metrics showed modest improvements starting with the 30th epoch, with Accuracy improving to 89.8%, Precision improving to 91.2%, Recall improving to 88.5%, and F1-score improving to 89.8%. Accuracy improved to 90.4%, Precision

improved to 91.9%, Recall improved to 89.0%, and F1-score improved to 90.4% throughout the 35th epoch, demonstrating that the framework was achieving excellent reliability.

The algorithm's Accuracy achieved 91.2% for the 40th time, while its Precision achieved 92.6%, its Recall achieved 90.2%, and its F1-score achieved 91.3%. This type of accuracy demonstrates the reliability and success of the model in accurately identifying emotions. The model possessed a remarkable Accuracy of 91.9%, Precision of 93.3%, Recall of 90.8%, and F1-score of 91.9% by the time it hit the 45th epoch. The model achieved its best performance at the 50th epoch, with an accuracy of 92.5%, precision of 93.7%, recall of 91.8%, and F1-score of 92.7%. This was the algorithm's best-known result.

Confusion Matrix (CM) Results

As demonstrated in Fig. 3, the CM presents a graphical illustration of the accuracy of the ER method by indicating the total number of right and wrong predictions for each class within the review. Joy, sorrow, rage, anxiety, shock, dissatisfaction, and neutral position are the classes that belong to this group.

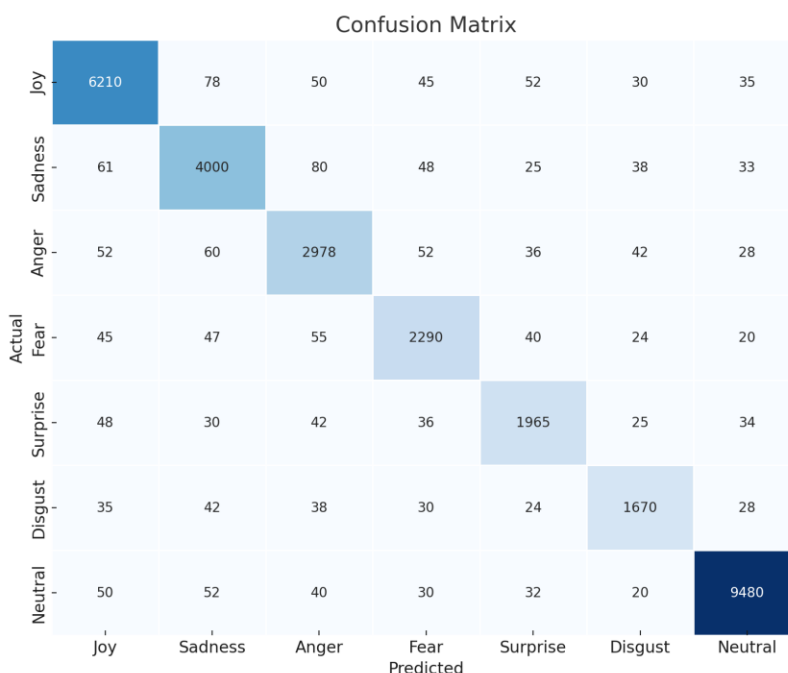


Figure 3: Confusion Matrix

The CM provides a comprehensive view of the ER algorithm's performance, highlighting correct and incorrect classifications for each emotion category. Out of 6,500 instances of Joy, 6,210 were correctly classified, indicating a high accuracy for this category. However, there were some misclassifications: 78 instances were labeled as sorrow, 50 as rage, 45 as anxiety, 52 as shock, 30 as dissatisfaction, and 35 as neutral position. This confusion may be attributed to mixed emotions or subtle FE that overlap with other categories. Among the 4,200 instances of sorrow, 4,000 were correctly identified, while 61 were misclassified as sorrow, 80 as rage, 48 as anxiety, 25 as shock, 38 as dissatisfaction, and 33 as neutral position. The confusion between Sadness and other emotions such as Anger, Fear, and Neutrality suggests overlapping features in FE and vocal characteristics.

Out of 3,100 instances of Anger, 2,978 were correctly classified, but 52 were misclassified as Joy, 60 as sorrow, 52 as anxiety, 36 as shock, 42 as dissatisfaction, and 28 as neutral position. The overlap between Anger and other emotions like sorrow, anxiety, and dissatisfaction reflects similar expressions and vocal tones. With 2,400 instances of Fear, 2,290 were correctly identified, while 45 were misclassified as Joy, 47 as sorrow, 55 as anxiety, 40 as shock, 24 as dissatisfaction, and 20 as neutral position. The confusion between sorrow and anxiety could be due to the intensity of FE and vocal tones.

Among the 2,100 instances of shock, 965 were correctly classified, with 48 misclassified as Joy, 30 as sorrow, 42 as rage, 36 as anxiety, 25 as dissatisfaction, and 34 as neutral position. The overlap between Surprise and Joy may be due to positive FE, while confusion with Fear could be attributed to sudden changes in FE. Out of 1,800 instances of dissatisfaction, 1,670 were correctly classified, but 35 were misclassified as Joy, 42 as sorrow, 38 as rage, 30 as anxiety, 24 as shock, and 28 as neutral position. The confusion between dissatisfaction and anxiety or shock indicates that intense negative expressions often overlap. Finally, out of 9,900 instances of neutral position, 9,480 were correctly classified, with 50 instances misclassified as Joy, 52 as sorrow, 40 as rage, 30 as anxiety, 32 as shock, and 20 as dissatisfaction. The neutral position was occasionally confused with Joy and sorrow, possibly due to subtle differences in FE.

The framework shows a high level of accuracy when compared to significant emotional groups like joy, negative emotions, and neutral position. On the other hand, it is not typical for individuals to wrongly classify emotions with comparable features, such as between sorrow and anxiety or between joy and shock. The incorrect classification rates of emotions with fewer instances, such as disgust and surprise, tend to be greater than those of other emotions. These results emphasize the import of enhancing feature extraction and boosting model generalization in order to distinguish between minor variations in emotional states that conflict.

CONCLUSION AND FUTURE WORK

As an outcome of the results of the study that was conducted, it has been proven that a novel Emotion Recognition (ER) program has been effective in identifying whether or not vocational counsellors who are hired by Higher Vocational Colleges (HVC) in Guangdong Province, China, developed burnout. The approach we propose addresses the significant disadvantages that are posed by traditional methods that rely on data supplied by individuals. These limitations include the fact that our method integrates data from numerous sources, such as psychological survey questionnaires, Voice Expression (VE) and Facial Expression (FE) collects and textual analysis. An accurate real-time evaluation of the verbal and non-verbal evidence that correlates to burnout indicators is now feasible through the use of modern Neural Network (NN) systems like BERT+LSTM, CNNs, and RNNs. According to conventional tests as such, which frequently encounter problems with bias and delayed reporting, the algorithm's ability to achieve an accuracy rate of over 92% in identifying signs of burnout marks a significant advancement. The accuracy not only supports the reliability of the approach but also shows the promise of machine learning (ML) methods to enable advancement in mental health diagnosis. The features of the algorithm in real-time enable it to identify burnout at an early stage, presenting a critical time for the response that may help reduce the negative impact that burnout has on the job effectiveness and psychological health of counsellors.

Additionally, this research significantly improves knowledge about burnout within the context of education. It emphasizes the vital function of regular evaluation and proactive control in these contexts. After the study results, it is recommended that HVC integrates equivalent ML technologies to enhance counsellor assistance systems. This will ensure that those at the cutting edge of student happiness can sustain their health and performance.

REFERENCES

- Kovach Clark, H., Murdock, N. L., & Koetting, K. (2009). Predicting burnout and career choice satisfaction in counselling psychology graduate students. *The Counseling Psychologist*, 37(4), 580-606.
- Carrola, P. A., Olivarez, A., & Karcher, M. J. (2016). Correctional counselor burnout: Examining burnout rates using the Counselor Burnout Inventory. *Journal of Offender Rehabilitation*, 55(3), 195-212.
- Cook, R. M., Fye, H. J., Jones, J. L., & Baltrinic, E. R. (2021). Self-reported symptoms of burnout in novice professional counselors: A content analysis. *Professional Counselor*, 11(1), 31-45.
- Davies, S. M., Sriskandarajah, S., Staneva, A. S., Boulton, H. C., Roberts, C., Shaw, S. H., & Silverio, S. A. (2022). Factors influencing 'burn-out' in newly qualified counsellors and psychotherapists: A cross-cultural, critical review of the literature. *Counselling and Psychotherapy Research*, 22(1), 64-73.
- Van Dam, A. (2021). A clinical perspective on burnout: diagnosis, classification, and treatment of clinical burnout. *European Journal of Work and Organizational Psychology*, 30(5), 732-741.

- George-Levi, S., Schmidt-Barad, T., Natan, I., & Margalit, M. (2022). Sense of coherence and burnout among school psychologists: The moderating role of loneliness. *Current Psychology*, 41(4), 2390-2397.
- Abaoğlu, H., Demirok, T., & Kayihan, H. (2021). Burnout and its relationship with work-related factors among occupational therapists working in public sector in Turkey. *Scandinavian journal of occupational therapy*, 28(4), 294-303.
- Zhang, X., Bai, X., Bian, L., & Wang, M. (2021). The influence of personality, alexithymia and work engagement on burnout among village doctors in China: a cross-sectional study. *BMC Public Health*, 21, 1-13.
- Z. Huijuan, Y. Ning and W. Ruchuan, "Improved Cross-Corpus Speech Emotion Recognition Using Deep Local Domain Adaptation," in *Chinese Journal of Electronics*, vol. 32, no. 3, pp. 640-646, May 2023, doi: 10.23919/cje.2021.00.196.
- Shahbaz, M., Hye, Q. M. A., Tiwari, A. K., & Leitão, N. C. (2013). Economic growth, energy consumption, financial development, international trade and CO2 emissions in Indonesia. *Renewable and sustainable energy reviews*, 25, 109-121.
- L. Zhu and Y. Xiao, "Mental Health Assessment of Higher Vocational Colleges Based on Artificial Intelligence," 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT), Bengaluru, India, 2023, pp. 1-4, doi: 10.1109/EASCT59475.2023.10392385.
- W. Lu, T. -P. Tan and H. Ma, "Bi-Branch Vision Transformer Network for EEG Emotion Recognition," in *IEEE Access*, vol. 11, pp. 36233-36243, 2023, doi: 10.1109/ACCESS.2023.3266117.
- J. Lin, X. Zhang and X. Zhuo, "Research on Comment Data Mining of Cross-Border E-Commerce of Belt and Road Countries Based on Cross-Lingual Sentiment Classification Technology," 2023 7th Asian Conference on Artificial Intelligence Technology (ACAIT), Jiaying, China, 2023, pp. 1008-1013, doi: 10.1109/ACAIT60137.2023.10528390.
- K. Zhang, C. Min, M. Shuzhen and L. Zhang, "Processing of Explicit Emotion Sentences and Implicit Emotion Sentences: A Comparative Study Using Behavioral Measures," 2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qiangdao, China, 2023, pp. 717-721, doi: 10.1109/ICFTIC59930.2023.10456170.
- X. Ai, V. S. Sheng, W. Fang, C. X. Ling and C. Li, "Ensemble Learning With Attention-Integrated Convolutional Recurrent Neural Network for Imbalanced Speech Emotion Recognition," in *IEEE Access*, vol. 8, pp. 199909-199919, 2020, doi: 10.1109/ACCESS.2020.3035910.
- X. Zhang, C. Luo, T. He, X., Z. Lu and B. Huang, "Online learner emotional analysis based on big dataset of online learning forum," 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 2017, pp. 1-5, doi: 10.1109/CISP-BMEI.2017.8302316.