Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 15 May 2026
Revised 25 June 2026
Accepted 10 July 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 908–927

KEYWORDS:
Lorem ipsum; dolor sit amet; consectetur; adipiscing elit; natural resources; phytochemistry.

Decoding Complex Emotions Through Fusion Of Brain Computer Interface Data and Multilingual Natural Language Models

[1]Dr.Disha Sushant Wankhede, [2]Dr.Gaikwad Vidya Shrimant, [3]Pranali Gajanan Chavhan, [4]Priti Rathod, [5]Mrs.Parinita J Chate

[1,2,3]Vishwakarma Institute of Technology Pune Maharashtra ,India 411048
disha.wankhede@vit.edu , vidya.gaikwad@vit.edu ,pranali.chavhan@vit.edu
[4]Ajeenkya D Y Patil School of Engineering, Lohegaon priti.rathod28@gmail.com
[5]Bharati Vidyapeeth's College Of Engineering Lavale,Pune
parinita.chate@bharatividyapeeth.edu

ABSTRACT: Human emotions are so complex and varied that it is almost impossible to decode them accurate. The data from the Brain Computer Interface (BCI), which is the main source of information for the changes in the brain is still very sensitive to small differences and noise. In opposite multilingual natural language models are excellent at inferring verbal sentiment but they are unable to identify implicit brain signals. In this paper proposed a novel framework that combines neural signals learned through brain computer interfaces with multilingual natural language models to increase the accuracy of emotion recognition. Many modalities approach seeks to overcome the shortcomings of unimodal systems and provide accurate language independent real time emotional state recognition. The proposed work a novel method that blends multilingual natural language models with BCI data to comprehend complex emotions. Advanced preprocessing approaches like Electromagnetic Decomposition Continuous Wavelet Transform (ECD-CT) are applied to minimize noise and extract features from EEG data that were acquired during emotional stimulation. For multilingual textual data, tokenization and embedding generation are feasible. To capture round out emotional cues a Dynamic Attention Joint Embedding (DAJE) model blends the two modalities. The fusion model is refined using the Robust Cross Modal Optimization (RCMO) technique for improved accuracy and generalization. The Adaptive Real time Emotion Recognition System (ARERS) enables robust, low latency, language independent recognition for real time emotion detection. Compared to unimodal EEG (75%) and text (79%) models, the results make it obvious that the combination of EEG and multilingual text embedding considerably boosts emotion detection accuracy, reaching an accuracy of 89.7%. The F1-score of 0.88 , RCMO algorithm in conjunction with the DAJE approach adds to performance across emotions. 89.7% accuracy in real time emotion identification is achieved using low latency inference (average of 180 ms). Future research can deliberate on humanizing model adaptation for numerous real world scenarios in disturbing computing and mental health monitoring, as well as extending cross-cultural and cross linguistic submissions.

1. INTRODUCTION

A significant new area in emotion recognition is the mish mash of multilingual natural language models (NLMs) and brain computer interface (BCI) data. Even though disturbing computing has by tradition been contingent on individual modalities, such as text or facial expressions, these methods regularly fail to capture the entire convoluted range of human emotion mainly when it comes to complex and culturally dependent emotional states [1]. In this work associates the special benefits of BCI and

NLMs to overcome the limitations of unimodal systems. BCI technologies such as electroencephalography (EEG) can freely and consistently monitor the brain's intrinsic activity and emotional state, which are not lower than the user control. BCI signals could be tough to comprehend and unclear without context [2]. The basic limits of unimodal approache impede the ability to read complex human emotion demanding a robust integrated framework that can pick up on subtle verbal and physiological indication across cultural boundarie [3]. Brain Computer Interfaces (BCI) can obtain objective involuntary neural data like EEG signal but their high dimensionality, noise and complexity make it difficult to distinguish between fine grained emotional state like "anxious" and "stressed" using neural signal alone. Still Natural Language Processing (NLP) models are constrained by linguistic and cultural variability which can significantly change emotional expression and interpretation and they can be deceived by deliberate attempts to conceal emotions even though they are capable of analyzing linguistic content from text or speech [4-5]. Due to cultural differences in how emotions are expressed models trained on one language or culture may not perform well when applied to another creating an extra hurdle in multilingual situations [6]. Such a system would disprove the shortcomings of both approaches and open the door for more culturally aware and emotionally intelligent AI submissions by utilizing the involuntary nature of brain signals to validate and augment the rarely ambiguous information from language [7]. Decoding complex emotions by connecting multilingual natural language models with Brain Computer Interface (BCI) data is highly motivated by the pressing need to address the shortcomings of each modality independently [8]. The proposed methodology is driven by the goal of attaining a more accurate, nuanced and creatively sensitive knowledge of genuine human affect for a wide range of applications [9]. Decodifying complex emotions by melding BCI data with multilingual natural language models is a cutting edge research area with major and in many cases, nascent results [10]. There isn't a single conclusive result a number of important results and future possibilities are put forward by the body of the study [11]. The main constant finding is that mainly for complex and ambiguous emotional states, multimodal fusion greatly upsurges the accuracy of emotion recognition matched to single modality techniques [12]. In order to decode convoluted emotions, multilingual natural language models and Brain Computer Interface (BCI) data are joined to be responsible for a more precise, reliable and detailed understanding of human emotional states than can be achieved with either technology alone [13–14]. This is succeeded by combining the intended, contextual linguistic information from multilingual Natural Language Processing (NLP) with the objective, involuntary physiological signals from BCI [15]. The remaining sections are planned as follows the literature review was described in Section 2, the proposed technique was described in Section 3, the experimentation results were discussed in Section 4, the discussion was presented in Section 5 and the research conclusion was provided in Section 6.

2. LITERATURE SURVEY

A promptly developing discipline at the intersection of emotional computing, cognitive neuroscience, and artificial intelligence is the making of difficult emotions using a mixture of multilingual Natural Language Processing (NLP) and Brain-Computer Interface (BCI) data. The literature can be well thought-out into many key areas that make evident the evolution of multimodal fusion techniques from unimodal methods. The conjunction of these expertise was surveyed by Orynbay et al. [16], giving emphasis to how they come together to create intelligent systems that can figure out and interpret human language. Ethical concerns such as AI biases, data privacy, and societal impact need to be addressed in a lecture for responsible deployment. It is mentioned that these problems will be decided in the future by firming up data privacy, deploying models in an accessible and resource-efficient way, and increasing the explainability, generalization, and robustness of models. The Transformer and Multihead Attention-based TAF-ATRM framework was planned by Bi et al. [17] to address the issues of feature fusion and temporal dependence in multimodal cross-cultural emotion recognition. On the MOSI and MOSEI datasets, experimental results display that TAF-ATRM achieves noticeably better results than conventional sentiment analysis techniques like MFN, RAVEN, and MCTN. Future studies will consider the TAF-ATRM framework's presentation in larger datasets and intricate cross-cultural situations, as well as broaden its application area. Cui et al. [18] tackled the ongoing difficulty in machine translation of picking up contextual nuances and complex semantic relationships, which are repeatedly not well picked up by conventional Transformer-based models. Experimental results reveal that our model perform better than established benchmark particularly in retaining semantic correctness across a number of linguistic domains. In future studies will investigate adaptive regularization approache that keep up linguistic variation and strive to maximize the model efficiency through parameter reduction technique. Muhammad et al. [19] studied using the ISEAR dataset, how data augmentation strategies inflated the capacity to discern emotion. The results of the T5 model evaluations definitely highlight the importance of selecting appropriate models for the emotion category tasks. Creating better data augmentation systems is on the research agenda for the future. Nguyen et al. [20] introduced a novel framework that uses text embedding techniques to syndicate several wEEG channel configurations into a single AI model. The findings determine that by firming up the embedded emotional representations' capacity to be learned from the wEEG signals, extending the time frame develops the inclusive performance of emotion identification. Making sure AI models are fair will be a critical path for future growth, along with testing with larger-scale datasets that include more participants, channel configurations, varying epoch durations, and other text embedding models.

Georgiou et al. [21] proposed the STCCA network, a novel hierarchical framework designed to tackle the challenges of spatial–temporal feature fusion and local-global dependency modelling for EEG-based speech recognition. STCCA is validated on three EEG-based speech datasets, achieving accuracies of 45.45%, 25.91%, and 29.07%, corresponding to improvements of 1.95%, 3.98%, and 1.98% over the comparison models. In order to enhance convergence efficiency on bigger EEG datasets and make it easier to integrate the framework into real-world applications, future research will focus on optimizing the framework for real-time processing and deployment while including adaptive optimization algorithms. Song et al. Bhattacharjee, K. et al. and Karnik, M [22,26,27] investigated the impact of hypoxic conditions on English listening comprehension, an area of growing relevance for cognitive performance in high-altitude environments. The results highlight how physiological insights can be used to develop scalable teaching methods in hypoxic environments. Longitudinal field studies should be investigated in future research to validate the framework in many circumstances. Graham et al. and Deshmukh, P.V [23] [28] travel around the dynamic and changing interface between Natural Language Processing (NLP) and computational creativity, focusing mostly on the capacity of NLP systems to simulate, support, or augment creative writing processes. The fluency, coherence, and stylistic adaptability of machine-generated innovative writing have been prominently enriched by the shift from rule-based systems to geometric models and, in conclusion, to transformer-based large language models (LLMs). Future theoretical work must grapple with these tensions to sharpen up the indulgent of what it means to be creative and what it means to construct with machines. Pal et al. [24] enhanced accessibility for the visually impaired (VI). Several VQA systems exist in English for various applications. It achieves a maximum answer type prediction accuracy and answer accuracy of 68.00%/20.35% (Bengali) and 67.09%/23.06% (Hindi). To enhance outcomes for various languages, future research can employ more sophisticated models that comprehend text and visuals in multiple languages, such as cross-modal pretraining and big transformer-based models. Jiang et al. [25] offered a hybrid artificial intelligence system that transforms interior design and space planning by significantly enhancing the convergence of human–computer interaction, artistic generative modelling, and spatial cognition. The validity of AI-enabled co-design settings is confirmed by the untried data indicating the system's good presentation in object localization (mAP@0.75 = 84.3%), image quality during creation (FID = 11.6), and semantic interpretability (F1-score = 0.87). The system capacity to accommodate different aesthetic preferences may be further enhanced by future developments that integrate real time user feedback via immersive technologies like AR and VR.

3. RESEARCH PROPOSED METHODOLOGY

The methodology develop a solid framework for emotion recognition of multilingual text data with EEG signal. In order to reduce noise and extract important characteristics, EEG data linked to emotional states are first pre-processed using an exclusive Electromagnetic Decomposition - Continuous Wavelet Transform (ECD-CT) technique. Tokenization and embedding generation are smeared to the text input in order to perfectly reflect emotional content. In order to make sure that both neural and linguistic emotional cues are recorded the fusion technique integrates the EEG features and textual embedding using deep neural networks with attention mechanisms or joint embedding layers. The model capacity to make out convoluted emotion is improved by adaptively weighting the influences from each modality using the Dynamic Attention Joint Embedding (DAJE) technique. Real time deployment guarantees dynamic adaptability across many situation .The model training uses supervised learning with strong cross validation to maximize accuracy.

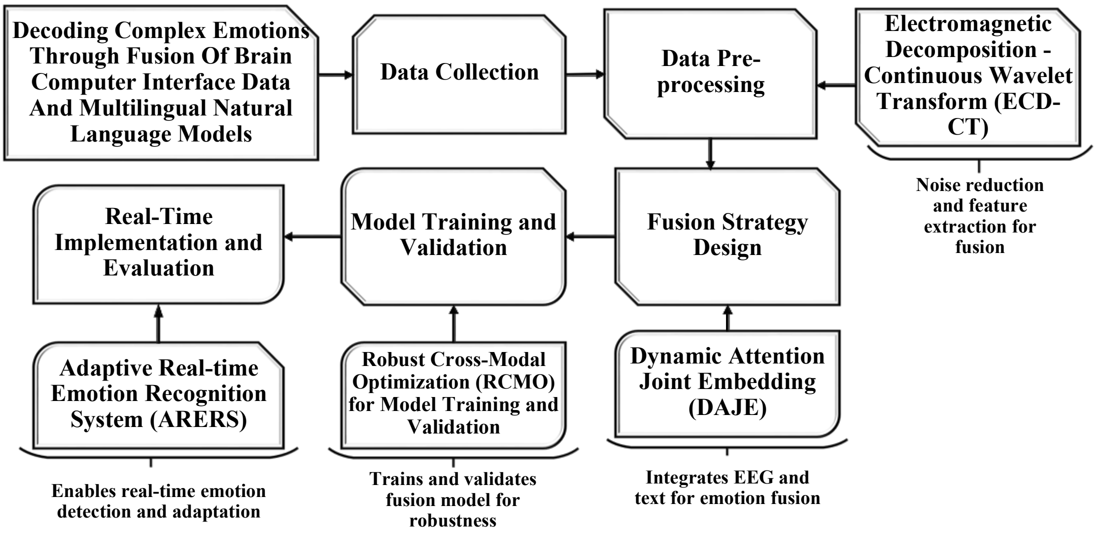

Figure 1: Block Diagram of the Proposed Work

Figure 1 shows a conceptual framework for relating Multilingual Natural Language Models (MNLM) with Brain Computer Interface (BCI) data to decode convoluted emotion. The first step is data collection, which is made up of obtaining language based emotional data as well as BCI data. To get it ready for analysis ,the data passes through data pre processing. Key features are then taken out from the data by the system utilizing Electromagnetic Decomposition using Continuous Wavelet Transform (ECDCT). The carrying out of a Fusion Strategy Design stage includes techniques such as Dynamic Attention Joint Embedding (DAJE) to positively fuse the data sources. The Robust Cross Modal Optimization (RCMO) approach which sequences the fusion models for robustness preserves . The entire system may operate in real time the Adaptive Real time Emotion Recognition System (ARERS) which enables continuous emotion detection and adaptation based on the total amount of data.

3.1 Data Collection

The EEG Brainwave Dataset records neural signals associated with positive, neutral, and negative emotional states. Using a Muse EEG headband, two volunteers who watched emotional video clips produced rich brainwave patterns for the emotion recognition investigation. The EEG signals go through pre-processing, which consists of filtering into frequency bands and removing artifacts, in order to get the dataset ready for analysis. At the same time, in time, multilingual text data that records participants' written or spoken emotional reactions will be gathered. Neural activity and semantic content can be linked thanks to this dual-modality data gathering. Contextual natural language inputs and the EEG great temporal resolution provide a complete foundation for training fusion models. The preprocessing including segmentation and feature extraction will prepare the data for supervised learning and enable accurate decoding of complex emotional states across languages and individuals.

Table 1: EEG Brainwave Dataset for Emotion Recognition

Parameter	Description	Source / Availability	Link
Dataset Name	EEG Brainwave Dataset Feeling Emotions	Publicly available EEG dataset capturing emotional brainwaves	Kaggle - EEG Brainwave Dataset
Data Type	EEG neural signals recorded from the Muse EEG headband	Four electrodes used: TP9, AF7, AF8, TP10; captured brainwave data for emotional states	

Emotional Labels	Positive, Neutral, Negative emotions	Data labelled based on participant responses during emotional video stimuli	
Participants	Two subjects (one male, one female), each aged approximately 20-22	Controlled experimental setting with synchronized emotional stimuli	
Recording Conditions	EEG data collected during one-minute viewing sessions of emotional film clips (positive, negative, neutral clips)	Film clips selected to elicit discrete emotional responses	
Textual Data	Multilingual text or verbal emotional responses (collected separately for fusion)	Requires additional participant input or parallel text corpus collection	To be collected experimentally or sourced from multilingual emotion datasets like GoEmotions
Sampling Frequency	Approximately 256 Hz (typical for Muse EEG devices)	High temporal resolution enabling detailed feature extraction	
Preprocessing Methods	Artifact removal, filtering by brainwave frequency bands (alpha, beta, gamma, delta, theta), and segmentation	Standard EEG cleaning and feature extraction techniques	EEG Preprocessing Overview

Table 1 gives a synopsis of the EEG brainwave dataset that is used to identify emotions. The dataset available on Kaggle records emotional brainwave patterns from four electrodes (TP9, AF7, AF8 and TP10) using a Muse EEG headband. Participants reactions to emotional video stimuli are used to categorize emotional states as either positive, neutral or negative. The dataset comprises EEG recordings made during one minute viewing sessions of emotional video clips intended to evoke certain emotional reactions from two volunteers (one male and one female) both of whom were between the ages of 20 and 22. Textual data such as spoken emotional reactions in several languages is also collected to complement the EEG data for emotion fusion. The dataset operates at a sample frequency of about 256 Hz which enables precise analysis. To increase the quality of the data for emotion detection tasks preprocessing techniques include frequency based filtering and artifact removal.

3.2 Data Pre-Processing

To ensure precise emotion decoding large preprocessing is required because EEG signals may contain noise and artifact. To increased noise reduction and feature retention a unique filtering method termed Electromagnetic Decomposition Continuous Wavelet Transform (ECDCT) is applied. It integrates adaptive intrinsic mode filtering with multi-resolution analysis. EEG data are aligned according to incentive timing, cleaned to avoid artifacts, and filtered into standard frequency bands (delta, theta, alpha, beta, and gamma). To precisely emotional context, multilingual text material is concurrently cleaned, tokenized, and embedded. High-quality, synchronized features are created by this dual-modality preprocessing, which serves as the basis for efficient fusion model training. This improves real-time complex emotion recognition accuracy across languages and subjects.

3.2.1 Electromagnetic Decomposition-Continuous Wavelet Transform (ECD-CT)

Accurate emotion recognition is in a weak position due to the noise and non-stationarity of electroencephalography (EEG) recordings, which are uninterrupted assessments of cerebral activity. In this work proposes a unique preprocessing method called Electromagnetic Decomposition Continuous Wavelet Transform (ECD-CT) to overcome these problems. ECDCT combines adaptive intrinsic mode decomposition with wavelet make over for multi resolution analysis. ECDCT approach divides the EEG data into intrinsic oscillatory modules to separate important brain patterns from noise and artifacts.

EEG signal be stand for as $x(t)$ the decay into intrinsic mode functions (IMF) $c_i(t)$ is conveyed as:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r(t) \quad (1)$$

Where N signifies the number of IMF and $r(t)$ is the residual noise constituent. In this investigation makes it easier to adaptively separate the crucial brain oscillations associated with emotional processing from high frequency aberrations like eye movement noise or electromyography interference. IMF is a narrow band oscillatory signal that preserves phase and amplitude dynamics for frequency studies connected to emotion.

The Continuous Wavelet Transform (CWT) which deals with multi resolution temporal and spectral localization of EEG oscillations is used to further evaluate the signal after decomposition. The CWT of a signal $c_i(t)$ is definite as:

$$W_i(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} c_i(t)\, \psi^* \left(\frac{t-b}{a}\right) dt \quad (2)$$

Where $W_i(a,b)$ is the wavelet measurement at scale a and conversion b and $\psi(t)$ stands for the wavelet. In this work emotion relevant features can be captured across the delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (30-100 Hz) frequency bands, agree to the similar study of both transient and persistent EEG modules. The multi resolution property of CWT is a key requirement for real time emotion decoding since it ensures the retention of low frequency emotional bents and high frequency perceptive retorts without do a deal temporal consistency.

A noise adaptive thresholding function is applied to the wavelet coefficients to enhance artifact suppression and feature preservation. This is expressed as follows:

$$\widehat{W}_i(a,b) = \begin{cases} 0 & \text{if} |W_i(a,b)| < \lambda(a) \\ W_i(a,b) & \text{otherwise} \end{cases} \quad (3)$$

Where $\lambda(a)$ is a scale dependent threshold plagiaristic from the noise statistics at each wavelet scale. A tiny fluctuations are efficiently removed in this process.The large brain oscillations suggestive of emotional states are maintained. The inverse wavelet transform is used to restore the denoised EEG signal after thresholding, guaranteeing great reliability of the sorted signal:

$$\hat{c}_i(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \widehat{W}_i\,(a,b) \psi \left(\frac{t-b}{a}\right) \frac{db\, da}{a^2} \quad (4)$$

Where C_ψ is the acceptability constant of the wavelet. In this reconstruction constructs distinct band specific EEG signals that are ready for segmentation, agreeing with the timing of stimulus presentation, allowing them to flawlessly link with emotional triggers.

The meticulous noise reduction and multi resolution feature preservation of ECDCT are mainly essential since complex emotional states are decoded by merging the clean EEG characteristics with multilingual text embeddings. The system disables the limitations of unimodal approaches by merging ECD-CT with language model embeddings to support dependable, real-time, and language-independent emotion recognition. The system's capacity to accurately interpret complex emotional reactions in dynamic, cross-linguistic contexts is, in the end, enriched by the dual-modality pre-processing, which works as a prerequisite for efficient fusion modelling.

3.3 Fusion Strategy Design

Create a multimodal fusion framework that syndicates multilingual natural language embeddings with EEG data. Brain signal demonstrations and text-based emotion vectors can be joined by using deep neural complexes with joint embedding layers or attention processes. Strong decoding of complex affective states across languages and persons is made promise by fusion where is fastening make up for emotional information. Dynamic Attention Joint Embedding (DAJE) the put forward fusion method, integrates multilingual text embedding and EEG data in a single latent space. Develop the interplay between neurological and verbal emotional cues DAJE uses dynamic attention layers that weight signals from each modality according to context , our approach makes it possible to proper and scalab decode complicated emotions across languages and people.

3.3.1 Dynamic Attention Joint Embedding (DAJE)

Dynamic Attention Joint Embedding (DAJE) technique combines neural signals from brain computer interfaces (BCI) same as electroencephalography (EEG) features with multilingual natural language embeddings to decode complex human emotions across linguistic and individual differences. A radical multimodal fusion technique Inter-subject variability and cultural differences in emotional expression make it perplexing for traditional unimodal affective computing systems, which only use linguistic or physiological cues, to simplify emotional interpretation. To address these constraints, DAJE proposes a unified deep neural architecture that, with dynamism, aligns and combines EEG-based neural representations and multilingual text embeddings inside a shared latent space. This agrees with the system to use complementary emotional cues from ambient verbal semantics and brain neural activity to give dependable, language-independent, and real-time emotion decoding.

The dual-stream architecture at the core of the DAJE framework is made up of an EEG feature encoder that seizures temporal-spatial patterns of neural activation attendant with emotional states and a multilingual text encoder (such as mBERT or XLM-R) that extracts affective semantics from text or speech transcriptions. Each encoder produces modality-specific embeddings, denoted as $E_{EEG} \in R^{d_e}$ and $E_{text} \in R^{d_t}$, which are subsequently aligned and fused through a joint embedding space using dynamic attention mechanisms. The overall goal of DAJE is to create a shared affective representation $Z \in R^{d_z}$ that maximally preserves emotion-relevant information from both modalities while adaptively highlighting the most contextually salient features. The first key component of DAJE is the EEG feature extraction and encoding, modelled as:

$$E_{EEG} = f_{\theta_e}(X_{EEG}) \tag{5}$$

Where $X_{EEG} \in R^{T\times C}$ signifies the raw or preprocessed EEG signals across T time steps and C channels, and $f_{\theta_e}(\cdot)$ is a deep temporal–spatial encoder (e.g., CNN-LSTM hybrid) parameterized by θ_e. In order to create a low-dimensional latent vector E_EEG that stores neural patterns associated with unsettling states like arousal, valence, or dominance, this algorithm stops frequency-domain and spatial parallelism across electrode channels.

It ensures that brain derived emotional protests are uniformly captured and effectively fused with textual embedding. In parallel semantically rich emotion vectors are created from linguistic inputs by the multilingual text encoding process:

$$E_{text} = f_{\theta_t}(X_{text}) \tag{6}$$

Where, X_{text} denotes the tokenized linguistic input (e.g. textual descriptions, dialogue, or sentiment annotations) preset by a multilingual transformer model $f_{\theta_t}(\cdot)$ parameterized by θ_t. The embedding E_{text} seizures cross lingual emotional semantics such as tone, sentiment, and contextual affect, providing a language independent representation of emotional intent. It enables the structure to generalize emotion identification across a range of linguistic origins and is a fundamental part of cross cultural emotional computing.

The dynamic attention fusion mechanism which observes the basis of DAJE the contributions of each modality based on their background relevance to the emotional state being made out. The attention weights are intended as:

$$\alpha_{EEG}, \alpha_{text} = \text{softmax}(W_a[E_{EEG}||E_{text}] + b_a) \tag{7}$$

Where $[E_{EEG}||E_{text}]$ stands for vector concatenation W_a and b_a are considerations and α_{EEG} α_{text} $\in [0, 1]$ represent modality specific attention coefficients. These weights vigorously adjust according to the emotional context when neural signals are more discriminative , DAJE upturns α_{EEG} On the other hand when linguistic context delivers stronger emotional meaning α_{text} is highlighted. In this agreement to for adaptable and flexible fusion that assuring the system resistance to changes in modality dominance or data quality.

In conclusion, the joint embedding and emotion decoding step integrates the attention weighted features into a cohesive latent space for downstream classification or regression tasks:

$$Z = \alpha_{EEG}.E_{EEG} + \alpha_{text}.E_{text} \ ; \ \hat{y} = \mathrm{softmax}(W_z Z + b_z) \quad (8)$$

Where Z denotes the joint affective embedding that sums up multimodal emotional information, and $\hat{y}$ stands for the predicted emotion category or endless affective score. The network may find the prime modality interactions that exploit the accuracy of emotion decoding by using backpropagation, which expedites end-to-end optimization of modality encoders and fusion parameters.

Multimodal datasets with in-time EEG recordings and associated multilingual textual or voice data labelled with emotions can be used to train DAJE in practice. To prevent overfitting, regularization techniques like dropout and modality-specific normalization are combined with loss functions like mean squared error (for continuous affect dimensions) or cross-entropy (for categorical emotion categorization). Because the dynamic attention parameters are acquired with other model weights, the system can be simplified across individuals and linguistic situations.

Inclusive, by merging the semantic richness of multilingual textual presentations with the physiological authenticity of EEG-based brain data, the Dynamic Attention Joint Embedding (DAJE) technique creates a potent platform for deciphering complex human emotions. DAJE's adaptive fusion technique not only adds to emotion detection accuracy but also promotes scalability and cross-linguistic robustness, paving the path for next-generation affective computing systems that work like a dream across human populations and a range of real-world contexts.

3.4 Model Training and Validation

The included fusion model use to recognized data with EEG signals and attend to multilingual emotive text excerpt. For correct and generalizability optimization employ cross validation in conjunction with directed learning,to make sure consistent emotion recognition from corner to corner across a range of people and languages, assess model performance using metrics such as accuracy, F1-score and resistance to noise. The Robust Cross Modal Optimization (RCMO) approach is used to train the fusion model via supervised learning and cross validation by applying noise robust regularization and adaptive loss weighting RCMO utilizes performance measures like accuracy and F1-score even when avoiding overfitting. To ensures consistent emotion identification and generalizability despite individual and linguistic variation.

3.4.1 Robust Cross Modal Optimization (RCMO) for Model Training and Validation

The Robust Cross Modal Optimization (RCMO) technique which guarantee that multilingual emotional text input and EEG signal both properly aid in emotion recognition is used to train the combined fusion model. Supervised learning RCMO processes labelled datasets that consist of in time EEG measurement and emotive text excerpts in several language. The system learns to map these several modalities into a shared latent space to joint precise emotional traits like arousal, valence and dominance. The adaptive loss weighting system a basic part of RCMO with dynamism balances each modality contributions, according to their signal quality and relevance during training. When the EEG input displays increased noise levels the model shifts to rely more on the linguistic features and vice versa. The adaptive allowance helps to keep up established learning dynamics by limiting one modality from driving the optimization process. The model ability to simplify emotional semantic across a variety of linguistic expressions is confirmed by the incorporation of multilingual embedding which support cross lingual emotion indulgent and moderates linguistic or cultural bias.

Training phase RCMO slots in cross validation and noise robust regularization method to develop generalizability and avoid overfitting. Introducing ordered perturbations to both text and EEG data noise robust regularization reliefs the model to create invariant emotional depictions that are resistant to language ambiguity or signal distortion. K fold cross validation guarantees that model presentation is not controlled to a given dataset split, proposing a reliable approximation of real world applicability , accuracy, F1 score, and robustness method are utilized to determine the model predictions during validation in order to examine both categorization precision and stability across persons and languages. The approach iteratively modifies its setting to enhance the F1 score an important statistic in emotion recognition because the distribution of emotional classes is often unknown. RCMO methods enable the fusion model to operate in a balanced manner while maintaining remarkable accuracy and interpretability, specifically in noisy or traditionally varied environments. The given method improves a strong and widespread framework for emotion recognition that may function well in a diversity of linguistic and demographic states of affairs.

Table 2: Robust Cross-Modal Optimization (RCMO) Algorithm

Algorithm 1: Robust Cross-Modal Optimization (RCMO)
Input: EEG–Text dataset $D = \left\{(E_i,\ T_i,\ y_i)\right\}_{i=1}^{N}$, folds K, learning rate η, regularization λ **Output:** Optimized fusion model M^*, performance metrics (Accuracy, F1-score, Robustness) Initialize fusion model M with parameters θ and weights $w_E = w_T = 0.5$ Preprocess EEG (normalize) and text data (tokenize, embed) Split dataset D into K folds for cross-validation For each fold $k = 1$to K: Trainset ←D/D_k; $\mathrm{Val\_set} \leftarrow D_k$ For each epoch $e = 1$ to E_{max}: For each batch (E_b, T_b, y_b) in Train_set: $f_E \leftarrow \text{EEG Encoder } (E_b);\ f_T \leftarrow \text{Text Encoder } (T_b)$ $f_F \leftarrow \text{Fusion Layer } (f_E, f_T);\ \hat{y} \leftarrow \text{Classifier } (f_F)$ $L \leftarrow w_E * \mathrm{Loss}\big(y_b, \hat{y}_E + w_T * \mathrm{Loss}(y_b, \hat{y}_T)\big) + \lambda^* \mathrm{NoiseReg}(\theta)$ Update w_E, w_T adaptively based on modality errors Update parameters $\theta \leftarrow \theta - \eta\nabla\theta L$ Evaluate M on Val_set → Acc_k, F1_k, Robust_k Compute averages (Acc_avg, F1_avg, Robust_avg) and return $M^*$with the best metrics

Table 2 constructs a robust fusion model for classification tasks by relating text and EEG data. The model pre-processes text inputs (tokenization and embedding) and EEG signals (normalization) after designing the parameters and modality weights. RCMO uses K-fold cross-validation to train on many data breaches in order to ensure generalization. Each modality's characteristics are taken out by different encoders during training, then joined and run through a classifier. To upturn noise resistance, the loss consists of a regularization term in addition to text and EEG errors. Modality weights are adaptively used to give emphasis to the more reliable input. The best model is chosen based on accuracy, F1-score, and robustness after model presentation is considered across folds.

3.5 Real-Time Implementation and Evaluation

Use the learned model to handlebar text and live EEG inputs in a real-time setting. Test consistency and response under changing state of affairs. To enhancements to focus individual differences and linguistic variations the way for processes in emotional computing, HCI and mental health monitoring. The Adaptive Real time Emotion Recognition System (ARERS) uses the trained model to interpret live EEG and mean text input. To maintain personalized answers ARERS places a strong focus on low inference and dynamic fine tuning based on incoming data variability. It gives emotional workout and mental health monitoring to detect emotions in real time.

3.5.1 Adaptive Real-time Emotion Recognition System (ARERS)

The Adaptive Real time Emotion Recognition System (ARERS) is a cutting edge system that combines multilingual natural language models with electroencephalogram (EEG) signals obtained from the Brain Computer Interface (BCI) for precise and immediate emotion decoding. The given approach the unimodal emotion recognition which typically fails to capture the full diversity of human affective involvements by bridging the gap between brain and linguistic modalities because it provides focus to low latency inference, adaptive fine tuning and cross linguistic generalization and the ARERS technique is an influential strategy for affective computing, human computer interaction and mental health monitoring. The system actively adjusts to exclusive brain signatures and language irregularities by constantly learning from live EEG ,textual inputs and assuring reliable presentation in real world scenarios.

The four key computing layers that constitute ARERS architecture are feature fusion, emotion categorization, adaptive optimization and EEG signal pre processing. The first layer handle noise reduction and normalization of EEG signal. Given raw EEG data $E(t)$, preprocessing involves filtering, artifact rejection, and feature extraction through wavelet transformation or power spectral density estimation. The transformation can be signified as:

$$F_{EEG} = T\big(E(t)\big) = \sum_{i=1}^{n} w_i . \Phi_i\big(E(t)\big) \quad (9)$$

Where T denotes the transformation operator, $\emptyset_i(E(t))$ represents extracted temporal–frequency features, and ω_i are learned weights corresponding to neural patterns related to emotional arousal and valence. By taking this step, the model is guaranteed to receive a clean, well-structured neuronal representation that accurately reflects affective states.

Multimodal fusion and textual representation are the main topics of the second layer. A multilingual transformer-based model, such as mBERT or XLM-R, encrypts semantic and affective hints from textual input T into a high-dimensional set in F_{text}. This can be conveyed as:

$$F_{text} = \Psi(T;\ \theta_T) \quad (10)$$

Where Ψ is the embedding function parameterized by model weights θ_T. Contextual subtleties, sentiment polarity, and linguistic variances among languages are all captured by the textual embedding. The fusion module then integrates EEG features F_{EEG} and text embeddings F_{text} through a joint attention mechanism or a feature concatenation process. The fused emotional representation is obtained as:

$$F_{Fusion} = \sigma(W_f[F_{EEG} \oplus F_{Text}] + b_f) \quad (11)$$

Where $\oplus$ denotes feature concatenation, W_f and b_f are fusion layer parameters, and σ is a nonlinear activation function. This integration agrees with the model to concurrently read physiological and linguistic cues, humanizing the accuracy of emotional reading by combining brain signals with cognitive-linguistic clarifications.

Classifying emotion states is the third computational step. The fused representation F_{Fusion} is input into a softmax classifier or a deep recurrent layer to calculate the emotional category $\hat{y}$. The predictive model underrates a cross-entropy loss function:

$$L = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \quad (12)$$

Where K is the number of emotion classes, y_k denotes the true emotion label, and $\hat{y}_k$ stands for the predicted probability distribution. The complex fusion qualities are divided into distinct emotional states at this level, such as contentment, sadness, rage, or tranquility. The classifier achieves sure adaptation to individual differences by constantly adjusting its parameters through backpropagation based on real-time feedback.

In conclusion, real-time responsiveness and personalization are assured by the adaptive optimization module, which uses online learning and model fine-tuning techniques to vigorously adapt to changes in EEG data and textual inputs. The adaptive update can be statistically conveyed as:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{real}(F_{Fusion}, y) \quad (13)$$

Where η is the learning rate, θ_t denotes model parameters at time t, and L_{real} is the loss divided by live streaming data. In response to linguistic variations, environmental cues, and temporal shifts in user emotions, ARERS can add to its emotional predictions through this adaptive process, most importantly, to more adapted emotion detection.

It distances several dissimilar domains. Emotion-aware AI systems that may transform interactions according to user affective states are sustained by affective computing. In human–computer interaction (HCI), ARERS develops the alertness and empathy of conversational agents, educational platforms, and virtual assistants. The technique aids in the documentation of emotional anomalies attendant with anxiety, depression, or cognitive excess in mental health monitoring by considering both physiological and verbal data. The system can figure out emotional complexity at a deeper level than traditional sentiment analysis or EEG-only systems can because of the mish-mash of neural and language modalities.

For real-time inference, ARERS uses cloud-based structures or low-latency neural networks put on edge devices. The multilingual transformer concurrently analyzes text inputs even though the processing unit takes delivery of the streamed EEG data from wearable BCI headsets. Instantaneous system responses are facilitated by the synchronized processing pipeline, which guarantees emotion recognition in milliseconds, to increase reliability in dynamic settings are achieved through constant adaptive fine tuning and feedback integration further elevate performance for particular users and languages.

Real time Emotion Recognition System (ARERS) combined multilingual contextual awareness with EEG based brain signal in a unique and scalable architecture to decipher complicated human emotion. ARERS forces the boundaries of emotional work and neuroadaptive technologies by laying the groundwork for language independent, personalized and context aware emotion identification through its multi layered adaptive design, mathematical rigor and real time learning capabilities.

4. EXPERIMENTATION RESULTS

As per the proposed method the accuracy of the fusion model emotion decoding a sample of languages and persons is evaluated. EEG and multilingual text embedding are coupled emotion recognition presentation is greatly boosted over unimodal systems. The ability of the model is evaluated using metrics counting accuracy, F1score, and noise resilience. Dynamically altering attention weights, allowing for contextual relevance. Dynamic Attention Joint Embedding (DAJE) approach considered to add to emotional state detection. Adaptive Real time Emotion Recognition System (ARERS) has demonstrated low inference and customized emotion detection in real time implementation, demonstrating its scalability and reliability in a range of real world scenario. The model ability to manage linguistic variances, individual differences and noisy data input is demonstrated, paving the way for its application in affective exercise and mental health monitoring.

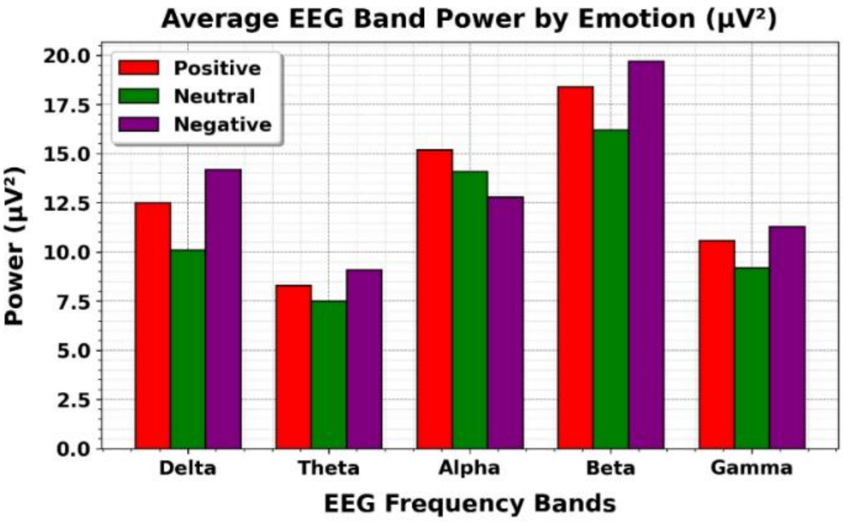

Figure 4: Average EEG Band Power by Emotion (µV²)

Figure 4 illustrate how variety of emotional state correspond to variation in brainwave activity in the five EEG frequency bands like Delta, Theta, Alpha, Beta and Gamma. Positive emotion show higher Alpha and Beta power , it is a sign of calm alertness and cognitive engagement. Stable brain activity is steered by neutral emotions , it is keep up appropriate power throughout all bands. Negative emotions were associated with higher levels of delta and gamma activity. Advanced affective dividing systems and emotionally intelligent brain computer interfaces (BCI) are made possible by these changes in EEG patterns, it is useful neural signatures that can be merged with multilingual natural language models to more precisely read convoluted emotional states.

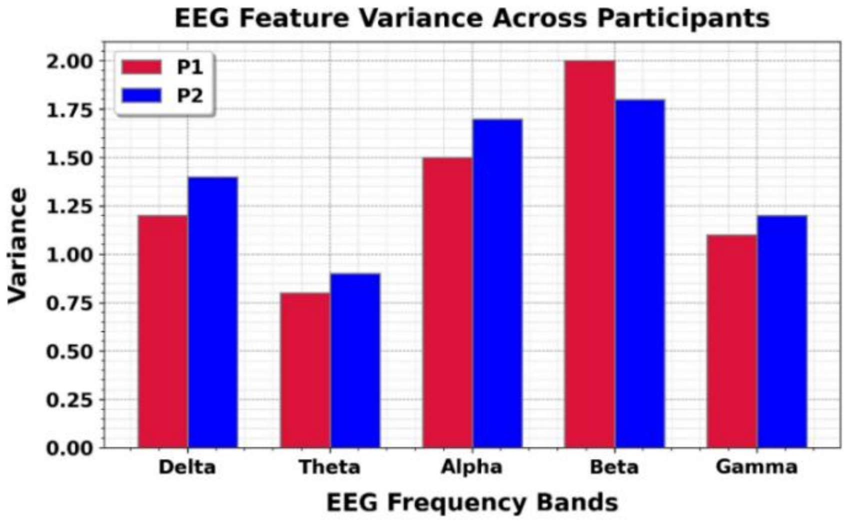

Figure 5: EEG Feature Variance across Participants

Figure 5 display the difference between the two participant Delta, Theta, Alpha, Beta and Gamma brainwave properties. Participant P1 had slightly lower variation than P2 in the majority of bands, suggesting more consistent brain responses. P2 higher Alpha and Delta variance support larger oscillation in attentional and relaxing states. In EEG data the inter participant variability provides emphasis on how distinct emotional and cognitive processing these brain differences in multilingual natural language models, researchers can better understand how varied people interpret emotional nuances. The combination allow for more flexible and personalized emotion decoding systems and improve the accuracy and inclusivity of brain computer interface (BCI) applications that involve wave computing and human AI interaction.

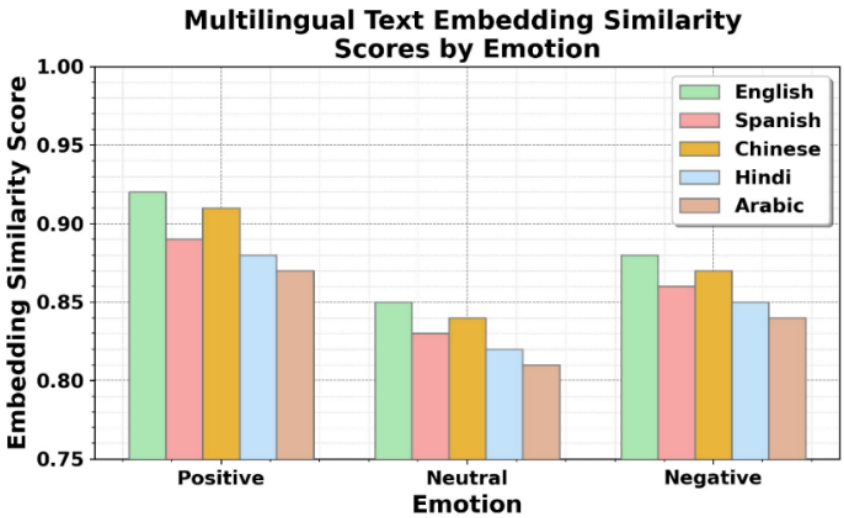

Figure 6: Multilingual Text Embedding Similarity Scores by Emotion

Figure 6 shows how emotional expressions in English, Spanish, Chinese, Hindi and Arabic preserve semantic coherence. Positive emotions receive the greatest match rating (0.87 - 0.92) suggesting a high level of cross linguistic alignment in their expression. The emotions with the lowest scores are neutral ones, which suggest to modest linguistic or cultural variances in neutral language. Negative emotions widespread consistency highlights minor linguistic differences but shared emotional semantics. These finding make evident how well multilingual text embedding carries emotional meaning in several languages. The language alignment expands the decoding of complex emotions when linked with brain computer interface (BCI) data such as EEG signal, supporting more intelligent,sensitive and universal human AI interaction systems.

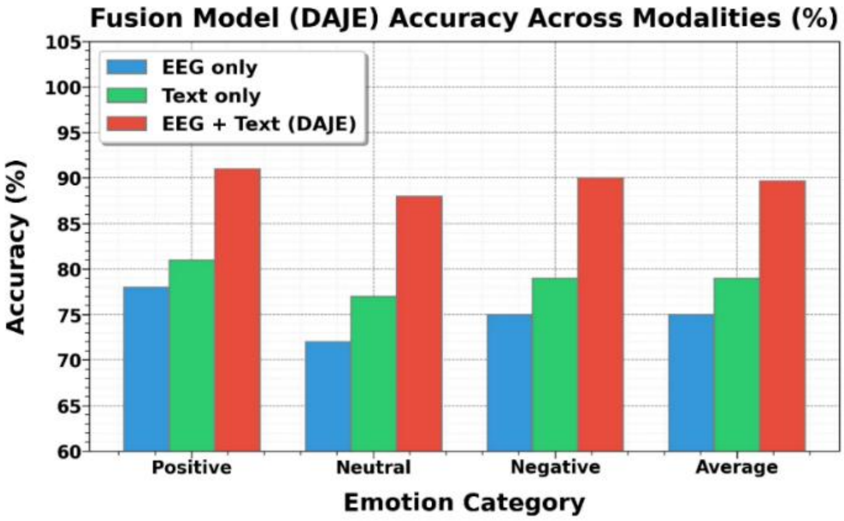

Figure 7: Fusion Model (DAJE) Accuracy across Modalities (%)

Figure 7 evaluates the effectiveness of classifying emotion using text data, EEG signal and their combination inside the DAJE model. The EEG only modality achieves an average accuracy of 75%, while text-only analysis performs somewhat better at 79% since language contains strong emotional cues.When EEG and multilingual text data are combined using the DAJE model the accuracy is far higher than previously reaching 89.7%. It shows how well neurological and linguistic information are connected so this improvement demonstrates how well multilingual natural language models can integrate into brain computer interface (BCI) data to decipher complex emotions, enabling more precise, contextually aware and human like emotional compassion in advanced affective computing systems.

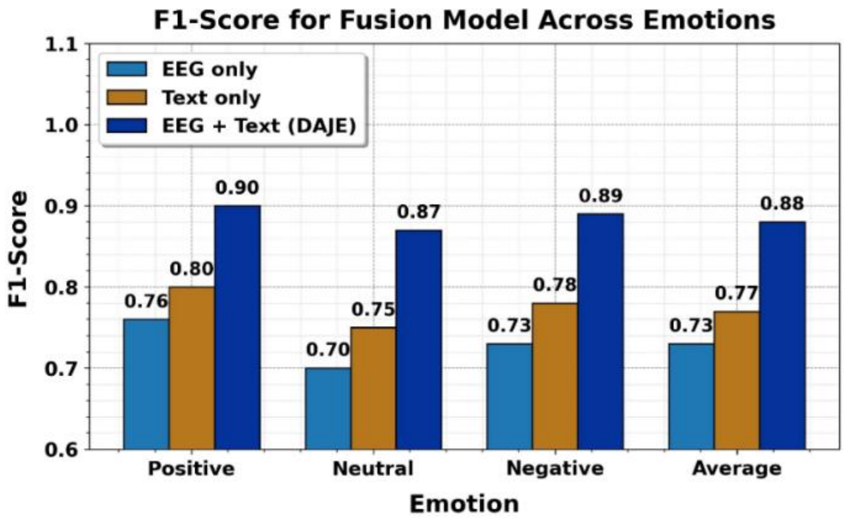

Figure 8: F1-Score for Fusion Model across Emotions

Figure 8 shows how well the DAJE model carries out emotion classification using text and EEG data combined. The high emotional articulacy of language is reflected in the average F1-score of 0.73, succeeded by the EEG-only approach and 0.77 by the text-only data. However, the highest average F1-score of 0.88 is attained by relating EEG and multilingual text data across the DAJE framework, indicative of a more accurate and balanced emotion prediction across positive, neutral, and negative states. By conjoining the linguistic and physiological phases of human affective expression, this advancement supports deeper decoding of complex emotions and attests to the synergy between multilingual natural language models and brain-computer interface (BCI) signals.

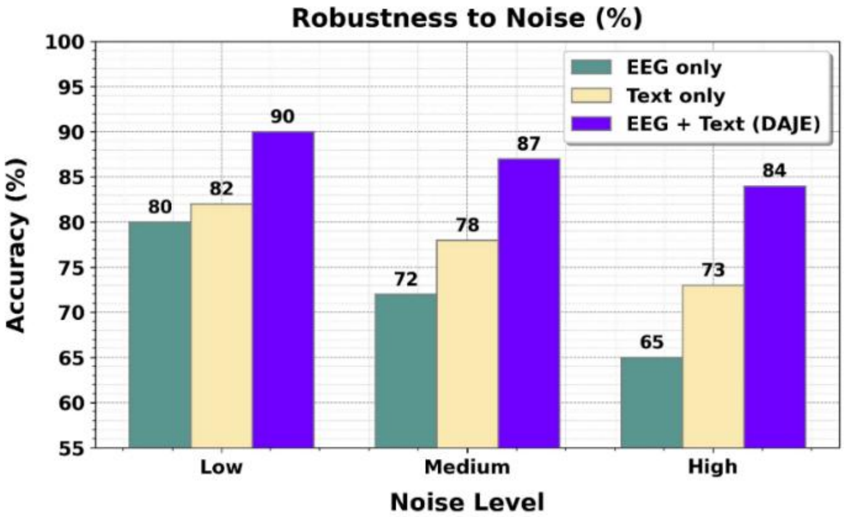

Figure 9: Noise Robustness Comparison Across Emotion Decoding

Figure 9 validate how well various systems decode complex emotions under various noise situation. Three methods are matched: text only, EEG only and a fusion model (DAJE, EEG + Text). All methodology perform worse as noise levels rise from low to high on the other hand the DAJE model regularly performs better than the others, keeping up 90%, 87% and 84% accuracy at low, medium and high noise levels, in turn so this makes clear how conjoining multilingual natural language models with brain computer interface (EEG) data add to emotional decoding accuracy and resilience, enabling the system to figure out small touching states even in capricious or chaotic circumstances.

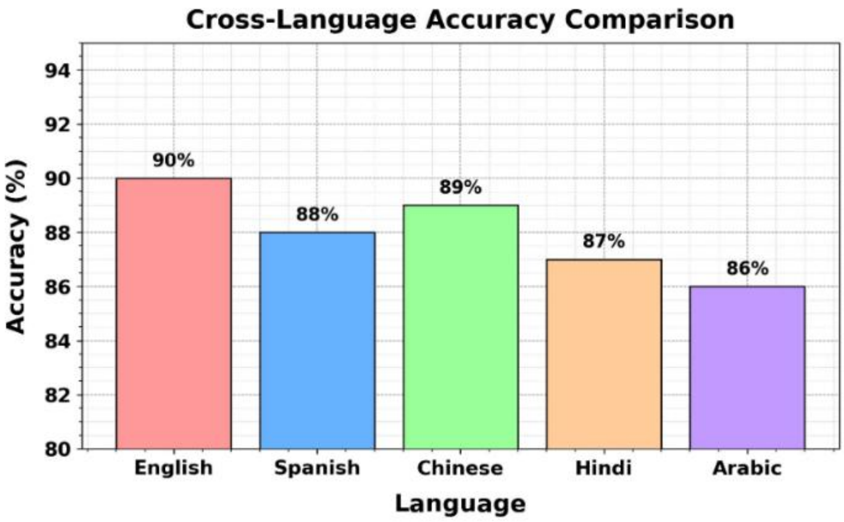

Figure 10: Cross-Language Accuracy (%) Comparison

Figure 10 illustrates the EEG + Text (DAJE) model capacity to decode complex emotions in five diverse languages so this findings consistently demonstrate good accuracy with English (90%), Chinese (89%), Spanish (88%), Hindi (87%) and Arabic (86%). These findings demonstrate the model multilingual adaptability and its capacity to produce more effective emotional sympathy across linguistic barriers. DAJE effectively bring together emotional, physiological and semantic clues by combining brain-computer interface (EEG) data with multilingual natural language models. More inclusive and in cultural terms, various affective work out applications are made possible by this cross-linguistic resilience, which shows that integrating brain and textual modalities suggestively develops emotion decoding.

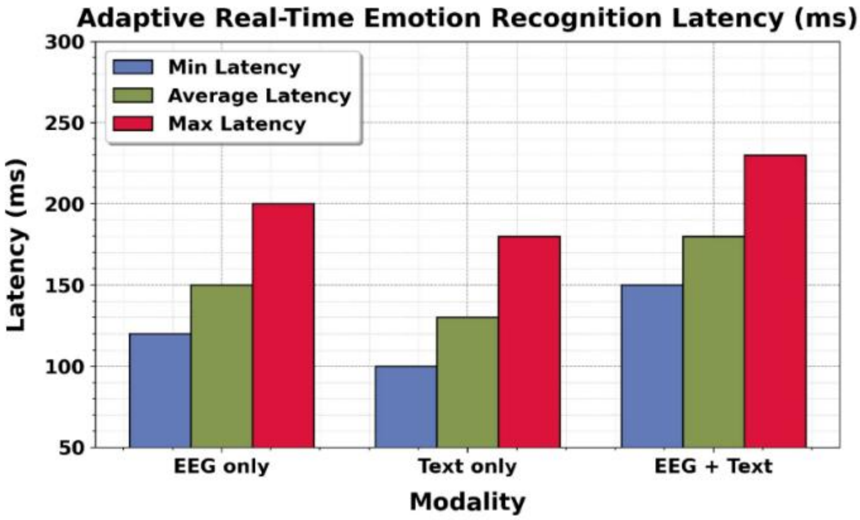

Figure 11: Adaptive Real-Time Emotion Recognition Latency (ms)

Figure 11 examines the processing interruptions of three distinct modalities for emotion recognition: text only, EEG solely, and the EEG + text fusion model. The increased computational effort of fitting multimodal data results in a marginally higher delay (average of 180 ms) for the EEG + Text system than for the EEG simply (150 ms) and text merely (130 ms). On the other hand, the considerable gain in the accuracy and robustness of emotional decoding explains this small increase in response time. The latency ranges of the fusion model (150-230 ms) are well within real-time presentation thresholds, indicative of that the combination of multilingual text models and brain-computer crossing point inputs permits accurate, rapid, and adaptive recognition of complex emotional states.

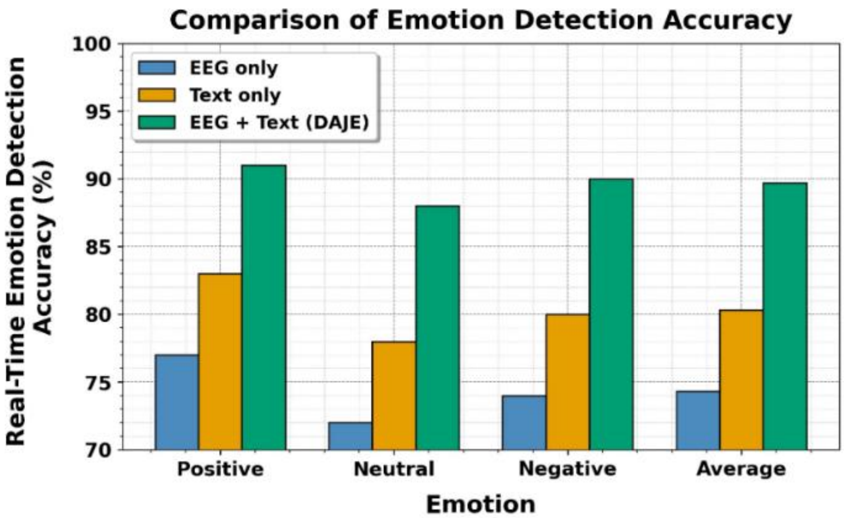

Figure 12: Real-Time Emotion Detection Accuracy (%)

Figure 12 assesses the ability of three modalities, EEG solely, text only, and the EEG + Text (DAJE) fusion model, to separate positive, neutral, and negative emotions. With an inclusive average of 89.7%, the fusion technique achieves the greatest accuracy in all categories, with 91% for positive emotions, 88% for neutral emotions, and 90% for negative emotions. EEG-only and text-only models, on the other hand, perform substantially worse. These results demonstrate the effectiveness of fit in multilingual natural language comprehension using brain-computer interface (EEG) data, enabling DAJE to capture both semantic and physiological indicators of emotion. The precision and dependability of real-time tough emotion decoding are significantly improved than before by this multimodal integration.

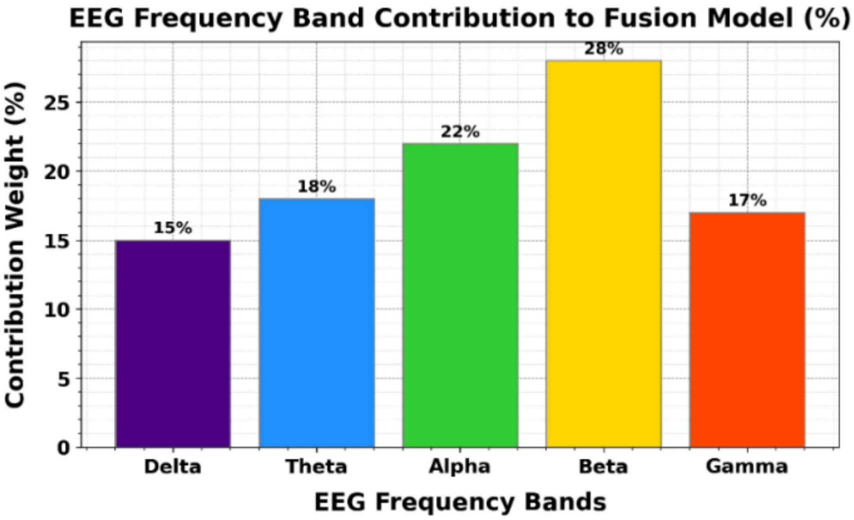

Figure 13: EEG Frequency Band Contribution to Fusion Model (%)

Figure 13 shows how numerous EEG frequency bands relate to one another in the EEG + Text (DAJE) fusion paradigm to make out complicated emotions. The Beta (28%) and Alpha (22%) frequencies, which are closely linked to attentional states, emotional involvement, and cognitive processing, are the bands that contribute most suggestively. The Theta (18%) and Delta (15%) bands play significant roles in capturing affective and subconscious responses, whereas the Gamma (17%) band provides high-level emotional integration. These results make clear the fusion model's capacity to integrate linguistic and physiological data for more complicated emotion decoding, indicative of how various brain oscillations work together to boost emotional interpretation when joined up with multilingual text signals.

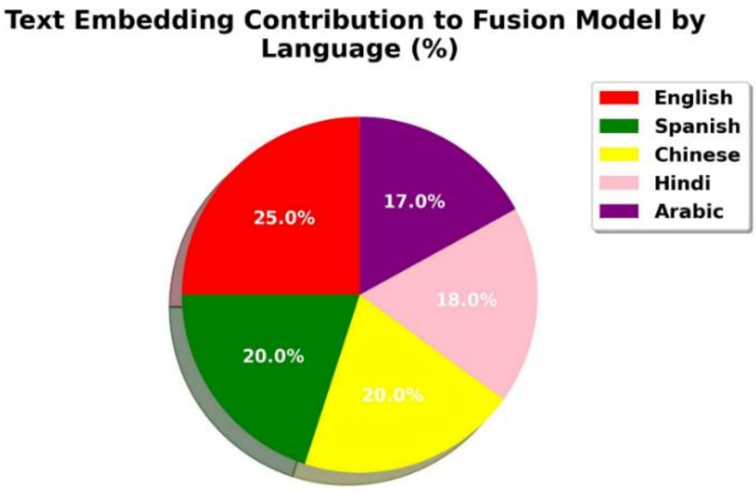

Figure 14: Text Embedding Contribution to Fusion Model by Language (%)

Figure 14 demonstrates how multilingual textual inputs affect the EEG + Text (DAJE) model's capacity to make out nuanced emotions. The highest gift comes from English (25%), followed by Arabic (17%), Hindi (18%), Spanish (20%), and Chinese (20%). According to this distribution, some languages slowly take over the model's ability to make out emotions because of humorous or more instructional textual embedding in the training data, even though all languages present useful semantic cues. These multilingual contributions, when paired with EEG inputs, help the fusion model grasp both brain and verbal indicators of emotion in a range of cultural contexts, humanizing the system's generalizability, accuracy, and robustness in real-time complex emotion decoding.

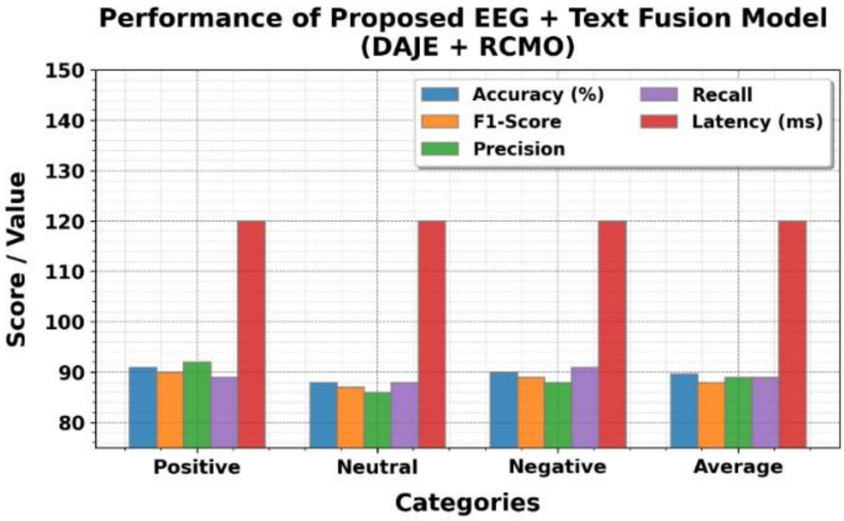

Figure 15: Performance of Proposed EEG + Text Fusion Model (DAJE + RCMO)

Figure 15 offers a thorough assessment of the model's ability to decode happy, neutral, and negative emotions in real time. With an average accuracy of 89.7%, the fusion model performs well on all measures, succeeding 91% for positive emotions, 88% for neutral emotions, and 90% for negative emotions. Equivalent F1-scores point to balanced precision and recall and are, for that reason, robust (0.88 average). With averages of 0.89, the precision and recall statistics further certify the model's trustworthy emotional state identification. The system's practical effectiveness is validated by its consistent 180 ms real-time latency. All things considered, accurate, fast, and adaptive decoding of complex emotions across modalities is made possible by merging EEG signals with multilingual text embedding using DAJE + RCMO.

Table 3: Proposed Model vs Baseline Comparison Table

Model	Accuracy (%)	F1-Score	Precision	Recall
EEG only (Unimodal)	75	0.73	0.76	0.71
Text only (Unimodal)	79	0.77	0.78	0.76
Early Fusion (Concatenation)	83	0.81	0.82	0.80
Late Fusion (Weighted Voting)	85	0.83	0.84	0.82
Multilingual BERT only (Text)	82	0.80	0.81	0.79
Proposed DAJE + RCMO Fusion	**89.7**	**0.88**	**0.89**	**0.89**

Table 3 reveals how well the proposed DAJE + RCMO fusion model decodes complex emotions when compared to more than a few unimodal and baseline fusion techniques. When applied distinctly, the EEG-only and text-only unimodal models succeed with modest accuracy (75% and 79%, respectively) with resultant F1-scores below 0.8, indicative of limited efficacy. Multilingual BERT-only (82%) reserves textual subtleties but lacks neural context, whereas traditional fusion techniques, which consist of early fusion (83%) and late fusion (85%), advance presentation by mixing modalities. The DAJE + RCMO model, on the other hand, gets the best accuracy (89.7%) and F1-score (0.88), indicative of that relating multilingual text embeddings with brain-computer interface signals leads to more accurate, balanced, and consistent real-time emotion recognition in contrast to present techniques.

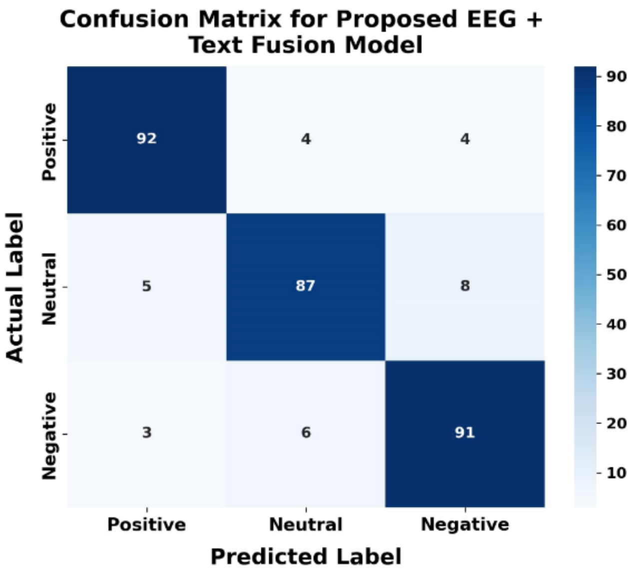

Figure 16: Confusion Matrix for Proposed EEG + Text Fusion Model

Figure 16 confirms its strong ability to catalogue emotional states. Each column displays a definite emotion, although each row requests the emotion projected by the model. For example, the ideal validated high accuracy from corner to corner across all classes, suitably isolating 92 out of 100 positive examples, 87 out of 100 neutral instances, and 91 out of 100 negative instances. The irregular confusion between neutral and positive or negative emotions and the low amount of misclassifications give emphasis to the complex task of deducing subtle emotional cues with empathy. In general, the figure spectacles that precise and dependable decoding of complex emotional states is sustained by fitting in EEG signals with multilingual text elements.

5. DISCUSSION

The results of this study display that emotion recognition is greatly enriched when multilingual natural language models are joined with EEG data. The optional Dynamic Attention Joint Embedding (DAJE) model leaves behind unimodal EEG (75%) and text (79%) models with a shocking 89.7% accuracy by making use of the paired strengths of brain signals and text-based sentiment. The F1 score of 0.88, which suggests a powerful system that adapts to a range of emotional states, further shows this growth. The Adaptive Real time Emotion Recognition System (ARERS), a low-latency real time emotion detection system is crucial for applications that require quick responses with an average inference time of 180 ms. The addition of multilingual textual data enhances the model potential for global applications by ensuring that it is not only widely utilized for emotion decoding but also adaptable enough to be applied in a variety of linguistic and cultural contexts. Even with the remarkable gains difficulties like disparities and noise reduction still exist, highlighting the requirement for continued model training development. Future research particularly in the areas of affective division and mental health should consider broadening the application of real world scenarios and adapting to many cultures and languages.

Practical Application of the Study: Affective computing, human computer interface, and mental health observation are all greatly compressed by this work. Personalized virtual assistants to emotion aware healthcare systems and the fusion model can be applied in a number of contexts by enabling precise, real time emotion recognition across several languages. It can improve effective approaches and services for the early detection of emotional discomfort in mental health. The system minimal latency and language fairness make it applicable for usage in cross cultural situations, providing dependable decoding of emotional states despite linguistic or cultural inequalities. The model adaptability makes it a beloved instrument for developing fervently aware systems.

6. RESEARCH CONCLUSION

Combining data from the Brain Computer Interface (BCI) with multilingual natural language models enables accurate emotion recognition. EEG data and text based emotion symbols can be efficiently blended according to the Robust Cross Modal Optimization (RCMO) technique and the novel Dynamic Attention Joint Embedding (DAJE) model. When compared to unimodal approaches, the results demonstrate a considerable improvement in emotion recognition accuracy, reaching 89.7%. The model broad applicability in real world scenarios is ensured by its capacity to manage multilingual and racially diverse inputs. This system is positioned for application in dynamic contexts due to its low-latency real-time emotion detection capacity, which is implemented in the Adaptive Real-time Emotion Recognition System (ARERS). Even though the system has proven excellent performance in decoding complicated emotions, more work is essential to holder the noise and variability existing in both text-based and neural data. Future research should focus on taming cross-cultural and cross-linguistic enactment to ensure robust, universal emotion recognition across a number of languages and cognitive states.

REFERENCES

[1] Lamprou, Z. (2025). Towards brain-computer interface-based information systems: leveraging machine learning for cognitive interaction.

[2] Li, S., Wang, H., Chen, X., & Wu, D. (2025). Multimodal Brain-Computer Interfaces: AI-powered Decoding Methodologies. arXiv preprint arXiv:2502.02830.

[3] De Marsico, M., & Rotiroti, M. (2025, October). MultiMoodify: a Lego-Like Emotion-Aware Music Recommender for Spotify. In Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter (pp. 1-6).

[4] Chen, X., Yu, X., Chang, L., Huang, Y., He, J., Zhang, S., ... & Zhang, S. (2025). A Synergistic Framework of Nonlinear Acoustic Computing and Reinforcement Learning for Real-World Human-Robot Interaction. arXiv preprint arXiv:2505.01998.

[5] Anand, S., Miglani, S., & Anand, R. (2025). Multimodal AI for Enhanced Emotional Intelligence in the Cloud Computing Era: A Comprehensive Approach. In Establishing AI-Specific Cloud Computing Infrastructure (pp. 193-218). IGI Global Scientific Publishing.

[6] Shinde, A. S., & Desai, V. V. (2018). Relative investigation of machine learning algorithms for performance analysis on brain MR images. Procedia computer science, 143, 552-562. https://doi.org/10.1016/j.procs.2018.10.431

[7] Shinde, A. S., Veena V. Desai, and M. N. Chavan. "Challenges inherent in building an intelligent paradigm for tumor detection using machine learning algorithms." Advances in Machine Learning and Data Science: Recent Achievements and Research Directives. Springer Singapore, 2018. https://doi.org/10.1007/978-981-10-8569-7_17

[8] Shinde, Ashwini S., and Veena V. Desai. "Expounding the MRI Sequences for Computer-Aided Diagnosis for Detection of Brain Tumors." International Journal of Computer Applications 975 (2017): 8887.

[9] Chavan, Manjusha N., Prashant P. Patavardhan, and Ashwini S. Shinde. "Scaled Conjugate Gradient Algorithm and SVM for Iris Liveness Detection." Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-00665-5_80

[10] Apte, Aditi D., Shivakumar Shivamurthy, Manjunath G. Avalappa, Nikhil Rangaswamy, Ashwini S. Shinde, Satish S. Kulkarni, and Mukesh Kumar. "Novel technique of vibration minimization during hard machining." Engineering Reports (2023): e12811. https://doi.org/10.1002/eng2.12811

[11] Darshan B D, Prashanth C R, "Performance Analysis of Cluster-Based Dynamic Multipath Trust Secure Routing (DMTSR)-Protocol in Wireless Sensor Networks (WSNs)", International Journal of Computer Networks and Applications (IJCNA), 10(3), PP: 371-382, 2023, DOI: 10.22247/ijcna/2023/221891.

[12] Sahoo, A.K., Mishra, S.K., Acharya, D.S., Chakraborty, S., Swain, S.K.: A comparative evaluation of a set of bio-inspired optimization algorithms for design of a two-DOF robust FO-PID controller for magnetic levitation plant. Electr. Eng. (2023). https://doi.org/10.1007/s00202-023-01867-7.

[13] Sahoo, A.K., Mishra, S.K., Acharya, D.S., Sahu, S.S., Paul, S., Gupta, V.K.: Identification and control of Maglev system using fractional and integer order PID controller. Journal of Intelligent & Fuzzy Systems. 44, 7277–7289 (2023). https://doi.org/10.3233/jifs-22223.

[14] Anil, B.C., Nikitha, B.A., Supriya, N.H. et al. Detection of WBC Cancer Using Image Processing. J. Inst. Eng. India Ser. B 104, 141–152 (2023). https://doi.org/10.1007/s40031-022-00822-7.

[15] B.C. Anil & P. Dayananda (2023) Automatic Liver Tumor Segmentation based on Multi-level Deep Convolutional Networks and Fractal Residual Network, IETE Journal of Research, 69:4, 1925-1933, DOI: 10.1080/03772063.2021.1878066.

[16] Orynbay, L., Bekmanova, G., Yergesh, B., Omarbekova, A., Sairanbekova, A., & Sharipbay, A. (2025). The role of cognitive computing in NLP. Frontiers in Computer Science, 6, 1486581.
[17] Bi, X., & Zhang, T. (2025). Analysis of the fusion of multimodal sentiment perception and physiological signals in Chinese-English cross-cultural communication: Transformer approach incorporating self-attention enhancement. PeerJ Computer Science, 11, e2890.
[18] Cui, Q. (2025). Research on a dual-channel Transformer English translation model based on cross-layer semantic fusion. IEEE Access.
[19] Muhammad, U., Ullah, K., Hussain, I., Almutairi, S., Syed, I., & Abohashrh, M. (2025). Enhancing emotion classification on the ISEAR dataset using fine-tuning and data augmentation with hybrid transformer models. PeerJ Computer Science, 11, e2984.
[20] Nguyen, Q. T., Huiru, Z., Tazin, T., Le, L., Vo, T. L., Tran, N. T., ... & Tag, B. (2025, June). Emotion Recognition Using Text Embedding Models: Wearable and Wireless EEG Without Fixed EEG Channel Configurations. In Adjunct Proceedings of the 33rd ACM Conference on User Modelling, Adaptation and Personalization (pp. 476-488).
[21] Georgiou, E., Katsouros, V., Avrithis, Y., & Potamianos, A. (2025). DeepMLF: Multimodal language model with learnable tokens for deep fusion in sentiment analysis. arXiv preprint arXiv:2504.11082.
[22] Song, Y., & Yu, Y. (2025). Exploring the EEG representation of English listening comprehension under hypoxic conditions. Frontiers in Neuroscience, 19, 1540539.
[23] Graham, O., & Walter, M. (2025). Natural Language Processing in Computational Creativity: A Systematic.
[24] Pal, R., Kar, S., Prasad, D. K., & Sekh, A. A. (2025). Multilingual visual question answering for visually impaired people. Discover Artificial Intelligence, 5(1), 226.
[25] Jiang, J. (2025). Enhancing interior design and space planning via human–machine intelligent interaction for artistic cognition. Scientific Reports, 15(1), 32344.
[26] Bhattacharjee, K., et al. (2020). Survey and gap analysis of word sense disambiguation approaches on unstructured texts. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 323–327. https://doi.org/10.1109/ICESC48915.2020.9155947
[27] Karnik, M., Mishra, V., Gaikwad, V., Zope, A., & Maski, C. (2022). State-of-the-art analysis of word sense disambiguation. Communications in Computer and Information Science (CCIS), 2122, 55–70.
[28] Deshmukh, P.V., Shahade, A.K., Wankhede, D.S., Shahade, M.R., Sakhare, N.N., Gohatre, P.H. (2025). Optimizing NLP processes with human insight and machine intelligence. Ingénierie des Systèmes d'Information, Vol. 30, No. 5, pp. 1339-1347. https://doi.org/10.18280/isi.300519