

# Towards Generic Emotion Representation with Type-2 Fuzzy VAD Space: Deep Fuzzy Multimodal Fusion Framework for EEG Analysis of Emotions

Mohammad Asif , *Graduate Student Member, IEEE*, Noman Ali , Sudhakar Mishra , Anushka Dandawate ,  
Uma Shanker Tiwary , *Senior Member, IEEE*.

**Abstract**—Recently, analysis of emotional states using neurological data has drawn enough attention. However, the complex nature of emotions and the subjective biases in self-reported values of Valence, Arousal and Dominance (VAD) make the emotion model too specific to a particular experiment. This study aims to develop a generic model representing emotions using a fuzzy VAD space and improve emotion recognition by multimodal fusion with EEG data. We partitioned the crisp VAD space into a fuzzy VAD space using low, medium and high type-2 fuzzy dimensions to represent emotions. A framework that integrates fuzzy VAD space and EEG modalities has been developed to recognize emotions. The EEG features were extracted using spatial and temporal feature vectors from time-frequency spectrograms, while the subject-reported values of VAD were also considered. The study was conducted on the DENS dataset, which includes a wide range of twenty-four emotions, along with EEG data and subjective ratings. The study was validated using various deep fuzzy framework models based on type-2 fuzzy representation, cuboid probabilistic lattice representation and unsupervised fuzzy emotion clusters. These models resulted in emotion recognition accuracy of 96.09%, 95.75% and 95.31%, respectively, for the classes of 24 emotions. The study also included an ablation study, one with crisp VAD space and the other without VAD space. The result with crisp VAD space performed better, while the deep fuzzy framework outperformed both models. The model was extended to predict cross-subject cases of emotions, and the results with 78.37% accuracy are promising, proving the generality of our model. The generic nature of the developed model, along with its successful cross-subject predictions, gives direction for real-world applications in areas such as affective computing, human-computer interaction, and mental health care.

**Index Terms**—Affective Computing, Emotion Recognition, VAD, DENS Dataset, Emotion Analysis, Emotion Representation, CNN, LSTM, Multimodal Fusion, Deep Fuzzy Framework, Emotion Modelling, EEG, Neurology.

## I. INTRODUCTION

**I**DENTIFYING and understanding emotions has been a major challenge not only in the field of artificial intelligence but also in general science and psychology. Understanding emotions comes with a wide range of benefits across various aspects of life, including personal well-being, social

interactions, and mental health. It is a complex phenomenon. According to the multidimensional theory of emotions, these psychological phenomena are complex constructs that consist of different dimensions [1], which together shape the human emotional experience. These multidimensional spaces are Valence, Arousal and Dominance (VAD). In this perspective, the complexity of emotions arises from the interplay of multiple dimensions. Basic emotions can be considered as primary building blocks within a multidimensional space [2]. More subtle and complex emotions emerge through the blending, intensifying, or suppressing of these basic emotional dimensions [3]–[5]. A schematic depiction of the emotion dimensions is provided in Fig. 1. We can depict the representation of emotion- $i$  in VAD space by

$$E_i \sim f(x_V, x_A, x_D) \quad (1)$$

where  $E \in \{\text{Set of Emotions}\}$ ,  $f$  represents a function, and  $x_V, x_A, x_D$  are the crisp values of valence, arousal and dominance provided by the participants.

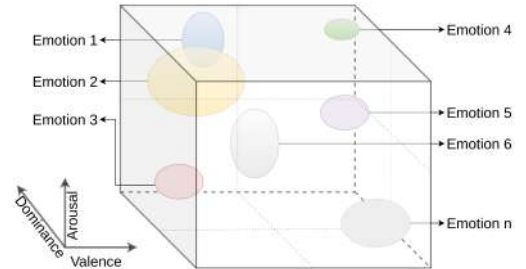


Fig. 1: Emotion Dimensions: A schematic diagram depicting emotions in VAD space.

Emotions are not simple entities but complicated combinations of subjective feelings, physiological responses, cognitive evaluations, and behavioural tendencies [6]. Moreover, the temporal dynamics of emotions should not be examined as isolated events but as evolving trajectories within this multidimensional framework. The subjective experience of an emotion is closely linked to the cognitive appraisal of a situation, the physiological arousal it induces, the subsequent behavioural expressions, and the element of perceived control or influence within the emotional experience [6]. This work also aims to explore the importance of a multidimensional theory and why valence, arousal, and dominance are important in understanding a wide spectrum of emotions. A

Mohammad Asif, Uma Shanker Tiwary, and Anushka Dandawate are with the Indian Institute of Information Technology Allahabad, Prayagraj, 211012, India (email: pse2017001@iiita.ac.in; ust@iiita.ac.in; iit2020077@iiita.ac.in).

Noman Ali is with the Indian Institute of Technology Madras, Chennai 600036, India, and also with Panjab University, Chandigarh 160014 (email: 21f1004271@ds.study.iit.ac.in, alinoman.itz@gmail.com).

Sudhakar Mishra is with the Indian Institute of Technology Kanpur, Kanpur 208016, India (email: sudhakarm@iitk.ac.in).

special emphasis will be placed on Dominance and empirically showing why it is so important while considering a complex phenomenon like emotions and how it is related to the user's subjective experience for controlling the emotions.

Human emotions frequently involve a degree of ambiguity, where individuals may experience mixed feelings or transitions between emotional states [7]. Also, emotional expression and perception vary across cultures and individuals [8]. Individuals vary in how they express and experience emotions. Some may express joy more exuberantly, while others may do so more subtly. Emotional intensity often varies gradually rather than abruptly [9]. Also, due to a limited emotional vocabulary, it is challenging to convey the nuances of their feelings accurately. The lack of precise words for certain emotional states contributes to ambiguity in self-reports [10]. The meanings of emotional terms, especially for fine-grained distinctions between similar emotions, can vary among individuals [11]. This semantic variability introduces ambiguity in the interpretation of self-reports. This ambiguity also occurs from the individuals' appraisal and interpretation of emotional stimuli differently as they may use different terms to describe similar emotional responses based on their cognitive interpretations [10]. Due to all these and because human emotions are inherently complex and often exist along a spectrum [12], crisp boundaries, being discrete and predefined, struggle to capture the full spectrum of emotional experiences. Emotions can manifest with varying intensities, shades, and combinations. Individuals rarely experience emotions as isolated events but rather undergo fluid transitions from one emotional state to another [13]. Fuzzy logic excels in handling such cases [14], and it can handle complex emotion combinations by allowing overlapping memberships. This flexibility enables the representation of mixed emotions or intricate blends, acknowledging that individuals often experience a combination of feelings simultaneously.

#### **EEG Based Emotion Recognition:**

Emotion recognition through EEG is a growing field in biomedical and neuroimaging areas that involves studying the complex neural patterns associated with human emotions by examining brain activity during emotional experiences. EEG, a non-invasive method, offers a real-time window with high temporal precision, providing a valuable approach in diverse applications. This can range from improving human-computer interaction [15] to contributing to mental health research [16] and finding neural correlates of the brain for emotions and processing.

EEG data analysis includes frequency bands, spectral power analysis, ERPs, connectivity measures, and time-frequency analysis. Emotion recognition is modelled using several methods, such as machine learning, ensemble learning, fuzzy logic, deep learning, and so on. Researchers use various models such as CNN, CNN-LSTM, CNN-GRU, attention-based hierarchical, and multi-head self-attention for emotion recognition [17]–[23].

However, there are challenges in this field, specifically defining and categorising emotions has always been a challenging task, including the complexity of emotions and the need to interpret intricate neural signals accurately [21]. EEG-

based emotion recognition provides an objective measure of emotional states, minimizing reliance on subjective self-reports. This objectivity is essential for avoiding biases and gaining a more accurate understanding of emotional responses. Additionally, there are other challenges that need to be overcome, including individual variability and integration with multimodal data like ECG, EMG, EOG and self-reported states [23]. Physiology-based emotion analysis is rare to find with fuzzy systems [24]. We will address and incorporate this issue in this work.

#### **Research Gap- Mapping The Relationship Between VAD Space and Emotion:**

In emotion studies, prevailing research often focuses on discrete emotion recognition [25] or continuous emotional dimensions recognition [26] and usually ignores the complex relationship between emotions and emotional dimensions [6]. This work proposes a deep fuzzy multimodal fusion framework that maps this relationship of emotions to multidimensional VAD space. The idea is to highlight the significance of multidimensions and emphasise why it is imperative not to overlook them when delving into the identification of complex emotions beyond discrete categories.

Another significant challenge in this field is the availability of datasets based on EEG. Generating the required amount of datasets based on EEG or any other biomedical imaging is not an easy task. The limited availability of datasets poses a significant challenge in training deep learning models effectively. Utilising all the information inherited in the datasets can help to cope with this problem, but generally, the ratings provided by the participants are omitted. However, integrating multimodalities like emotional experience in terms of three independent dimensions - Valence, Arousal, and Dominance (VAD) with neurological signals can overcome this issue. Fuzzy logic can be seamlessly integrated with deep learning models [27], combining the strengths of both paradigms. This integration allows for the development of deep fuzzy frameworks that leverage the power of neural networks while benefiting from the flexibility and interpretability of fuzzy logic [28].

#### **Main Highlights of the Work:**

- An insight into a novel representation of emotions using Interval Type-2 fuzzy VAD space.
- **Deep Fuzzy Multimodal Fusion Framework:** Proposing a multimodal fusion framework for recognition of a wide spectrum of emotions using fuzzy VAD space and EEG data.
- Significantly improving cross-subject emotion recognition with the fusion of multimodalities and fuzzy framework.
- Representation of temporally localised twenty-four emotions (in DENS Dataset), developed and validated on Indian population in our lab [21], [29].

The paper is structured into several sections, starting with an introduction to the research topic and research gaps. The following section covers the VAD space and provides an overview of the emotions that are included in this work using the DENS dataset. The third section talks about the fuzzy representation of emotion. The fourth section is about the

methodology, which discusses the complete methodology from the preprocessing of the EEG signal to the integration of fuzzy framework in the deep architecture. The next section presents the results of all the experiments done. Finally, the study concludes with a discussion in the last section.

## II. BEYOND TRADITIONAL LABELS: MAPPING EMOTIONS INTO THE VAD SPACE USING DENS

In this section, we give a description of the emotional dimensions and empirically analyse why considering emotions in an emotional dimension (VAD space) is crucial. We also present an overview of the emotions collected from the DENS dataset [21], which covers twenty-four different emotions. These emotions have concise information of the time and have EEG recordings with subjective ratings.

### A. Exploring the Dimensions of Emotional Experiences

Emotions are often represented in a three-dimensional space defined by valence, arousal, and dominance, known as the VAD model [30], [31]. The representation in this space is a potential way to capture the multifaceted nature of human emotions [32], [33]. Understanding the interaction of these three dimensions allows for a subtle and comprehensive representation of emotions and provides a geometric space where each point corresponds to a unique emotional state.

Emotions are studied in neuroscience and psychology by exploring their representation in the mind and subjective experience [34], [35]. Affective experiences are central to our understanding of emotions, and these experiences are not merely influenced by language or social desirability but are intrinsic to how emotions are mentally represented [35]. While core affect, or the basic experience of pleasure or displeasure, provides a foundational aspect of emotional experience, it is acknowledged that a complete understanding of emotions goes beyond this basic level. As mentioned, an experience of emotion is not just about feeling pleasure or displeasure; it is an intentional state that involves being directed towards or about something [36].

The level of physiological arousal (e.g., heart rate, breathing, muscle tension) is a crucial component of emotional experiences. Different emotions are associated with varying levels of arousal. For example, excitement or fear may be linked to high arousal, while calmness or sadness may be associated with lower arousal [37].

Having these two dimensions is not sufficient if we ignore the degree of control or influence one perceives they have over their emotional experience and expression. This is the Dominance dimension we refer to, and it is one of the key dimensions often used to characterise emotions, along with valence (pleasantness/unpleasantness) and arousal (intensity or activation level) [38]. On one end of the spectrum, there are emotions that are more dominant or controllable. These are the emotions where individuals feel like they have agency over their emotional experiences. They can regulate their emotions consciously and choose how to respond to a situation. On the other end of the spectrum, there are emotions that are less controllable or more dominant over an individual's subjective

experience. In these cases, it might feel like the emotion takes control rather than being consciously regulated [39].

There are individual differences in how people perceive and experience control over their emotions. Overall, recognising the dominance or control dimension adds depth to our understanding of emotions and contributes to the development of strategies for adaptive functioning.

### B. Overview of Twenty-Four Emotions and DENS Dataset

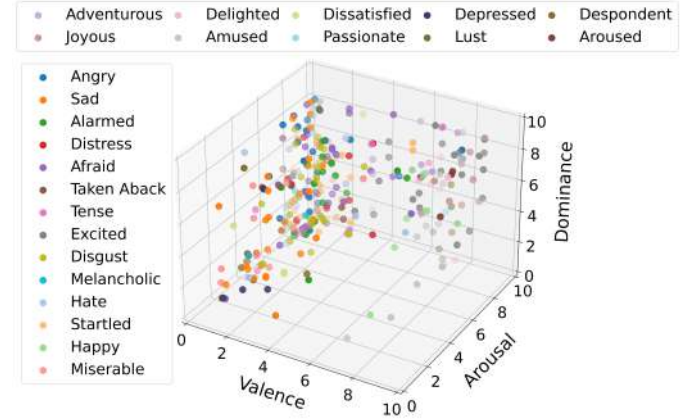


Fig. 2: Distribution of All Emotion samples in VAD Space Extracted from DENS: These emotions are utilized and further referenced in this work. Please refer to the supplemental file for more details about the DENS emotions.

The research employed naturalistic stimuli [40] to elicit emotional responses from participants. Each participant was randomly exposed to nine emotional stimuli and two non-emotional stimuli selected from a pool of 16 emotional stimuli, each lasting 60 seconds. The dataset encompasses EEG recordings from forty participants (mean age:  $23.3 \pm 1.25$ ,  $F=3$ ) while viewing emotionally evocative films. Participants were instructed to perform a mouse click upon experiencing any emotional response, referred to as an Emotional Event. After each video stimulus, participants assessed self-assessment scales measuring valence, arousal, and dominance, with valence, arousal, and dominance scales ranging from 1 to 9. After that, participants selected emotions from a predefined list or specified an emotional category aligned with their experience.

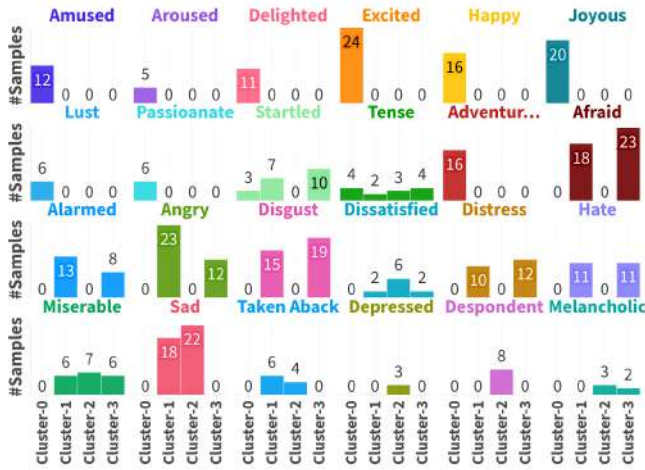
A total of 465 emotional experiences were collected, with each participant clicking at least once, averaging 1.29 clicks per participant. Though EEG recordings covered the entire stimulus duration (60 seconds), the work focused on a specific 7-second segment for each emotional event, yielding optimal results compared to alternative time frames (8s, 9s, and 10s). The recordings were sampled at 250 Hz.

The total number of labelled emotions is 24 in these 465 emotional events. The dataset also includes 10 other emotions associated with fewer emotional events. The distribution of all the twenty-four emotions in VAD space is shown in Fig.2. Additional information regarding the distributions of valence, arousal, and dominance of the 465 emotional events, along with the mean valence, arousal, and dominance values of each

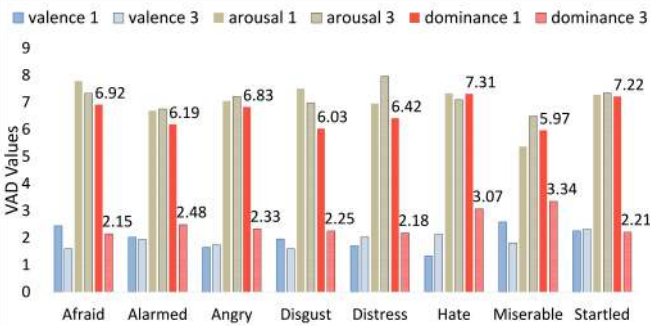
emotion and their respective standard deviations, along with centroids' locations, can be found in the Supplemental file (S-1 to S-4).

### C. Crucial Empirical Insights: Significance of Emotional Dimensions, with Emphasis on Dominance in VAD Space

To gain a better empirical analysis of the relationship between multidimensional space and emotions, we used Fuzzy C-Means (FCM) to divide the emotions into different clusters. We determined the optimal number of clusters using the fuzzy silhouette index [41] and found the optimum number of clusters is 4. FCM is applied to the VAD values for all the emotional events. We assigned all the emotional labels to these clusters based on the highest membership value associated with their respective clusters. Table I represents the centroids of the clusters. Emotion distribution to these clusters and crucial insights are shown in Fig. 3 and in the supplemental file (S-5 and S-6). Similar patterns are found in FCM with 6 clusters also (see supplement, S-7).



(a) Distribution of All the 24 Emotions in their respective Clusters and Samples.



(b) Cluster-1 vs Cluster-3: Emotions with Same Labels, but in Different Clusters- A Comparison of Emotions in two Clusters.

Fig. 3: Fuzzy C-Means with 4 Clusters: Analyzing the Same Emotion in Multiple Clusters. Many emotions are assigned to more than two clusters, with similar mean values of Valence and Arousal, but significantly different mean values of Dominance.

Clusters	Centroids			Number of Samples
	Valence	Arousal	Dominance	
Cluster-0	7.73	7.70	6.82	133
Cluster-1	2.09	7.17	6.75	140
Cluster-2	1.89	2.63	2.14	73
Cluster-3	1.86	7.05	2.57	119

TABLE I: Centroid of Clusters Received from FCM

After determining the optimal number of clusters using the fuzzy silhouette index, fuzzy C-Means is applied to the VAD values of the emotional space, resulting in their division into four clusters. The assignment of emotional labels to these clusters was based on the highest membership value in each cluster.

The analysis revealed that emotions like- Happy, Excited and Joyous predominantly belonged to cluster-0, which notably possessed a cluster centroid for valence 7.73, distinguishing it from other clusters whose centroid valence values fell within the range of 1 to 3. A clear distinction was observed between the Sad and Angry emotions. The sad emotions, which are less related to emotions causing agitation, were categorised into cluster 2. The discernible contrast in emotional agitation between the emotions within cluster 2 and those in clusters 1 and 3 is further evident in the disparity of centroid arousal values. Specifically, cluster 2 demonstrates a value of 2.63, while clusters 1 and 3 exhibit values of 7.17 and 7.05, respectively. Clusters 1 and 3 exhibited similar angry-afraid emotions, with varying combinations depending on the dominance of specific emotions. Emotion "Afraid" is found in two different clusters with almost equal numbers of emotional events. Upon analysis, it can be observed that the mean Valence and mean Arousal values are quite similar in these two clusters for this particular emotion. However, the mean Dominance value differs significantly between the two clusters. This is a similar case with many other emotions, as shown in Fig. 3b. These emotions have the same emotional labels, similar valence, and similar arousal but significantly different dominance values.

In our previous study, we found that including the Dominance dimension improves accuracy when considering only the Valence and Arousal dimensions [42]. When considering the VAD dimension, an enhanced accuracy was received compared to when using only V-A dimensions even when the number of classes increases. These differences are the results from the subjective experiences of the participants where the participant perceives emotion as other participants perceive, but the controlling mechanism of the participants differs a lot, which can only be examined with the Dominance value. It is difficult to understand why an emotion should be grouped differently into distinct categories if we don't take into account its level of dominance. Therefore, it is important to represent an emotion in terms of its Valence, Arousal, and Dominance (VAD) instead of considering it as a separate and isolated entity. In summary, mapping emotions onto the VAD space is crucial for a more accurate understanding of their nature and effects.



### III. BEYOND BINARY: PRACTICAL STRATEGIES FOR INTERPRETING HUMAN EMOTION COMPLEXITY WITH FUZZY REPRESENTATION

Emotions are subjective and can be ambiguous. They are represented in three dimensions, but due to individual differences in perception, the crisp values of subjective feelings and reporting may not accurately reflect the true nature of the emotion. Applying fuzzification to these dimensions reflects the complexity and subjectivity of emotional experiences. The concept of fuzzification deals with uncertainty and ambiguity by allowing values to be flexible and fall within a range instead of precise points. Fuzzification can indeed be applied to the dimensions of valence, arousal, and dominance in representing the complexity and variability of emotional experiences [43], [44].

Subjects feel and try to respond to their subjective feelings in terms of crisp values. Still, it will be more realistic if they represent it in continuous values such as low, medium or high and in membership functions in defining the imprecise boundaries of emotional states such as emotion belonging to a particular group or nearness to a particular group or sharing multiple groups. This approach acknowledges the fluid and subjective nature of emotional experiences and accommodates the inherent uncertainty in emotional expression [13]. This concept enables the continuous grading of emotional ratings, which traditional categorical models may struggle to capture [45].

Representing emotions through fuzzy VAD introduces the concept of soft boundaries between emotional states. This acknowledges that transitions between emotions are gradual instead of abrupt, which aligns with the dynamic and continuous nature of emotional experiences [6], [35], [45]. This soft transition allows for a more natural and realistic representation of the fluidity between different emotional states. Furthermore, it helps in reducing misclassifications in emotion recognition systems. By capturing the uncertainty and imprecision inherent in emotional expression, fuzzy concepts mitigate the risk of rigid categorisations that may lead to misinterpretations of complex emotional states [46]. This is particularly valuable in real-world scenarios where emotions are multifaceted and context-dependent [35].

Instead of strictly categorising emotions as purely positive or negative, fuzzification allows for shades of emotional experiences. Emotions may fall on a continuum of valence, acknowledging that some emotions may have mixed or nuanced valence. This idea aligns with the concept that emotions exist on a spectrum rather than being neatly categorised. Using fuzzy concepts in emotion dimensions can be a significant improvement in emotion representation methodologies, as it contains the margin for personal biases. Arousal can also be fuzzified to capture the variability in intensity experienced across different emotions. For example, an emotion could be moderately arousing rather than strictly categorised as high or low arousal [36]. This can manage uncertainty and imprecision, enabling us to capture the subjective and context-dependent nature of human emotions in a more subtle way. Applying fuzzy concepts in emotional dimensions of-

fers improved representational flexibility, allowing a deeper understanding of how emotions are experienced and expressed in different contexts. Fuzzifying the dominance dimension recognises that individuals may perceive varying degrees of control over their emotions [47]. Some emotions may be perceived as partially controllable, and fuzzification allows this gradation. Additionally, it provides a flexible framework that can adapt to individual differences in emotional expression and interpretation.

A more realistic and comprehensive emotional representation of Eq.1 depicting real-world scenarios can be represented by

$$E_i \sim f(\mu_V(x_V), \mu_A(x_A), \mu_D(x_D)) \quad (2)$$

where  $E \in \{\text{Set of Emotions}\}$ ,  $f$  represents a function, and  $\mu_V(x_V), \mu_A(x_A), \mu_D(x_D)$  are the membership degrees of valence, arousal and dominance dimensions for their respective crisp values of  $x_V, x_A, x_D$ .

#### Type-2 Fuzzy Representation of Emotions:

We defined the representation of emotion with interval type-2 fuzzy sets. The two membership functions- Lower and Upper (LMF and UMF), are based on the general meaning of the terms and the population-based meaning of the terms. UMF is defined as a generic term for all the dimensions, while LMF is set based on the statistical values of the population.

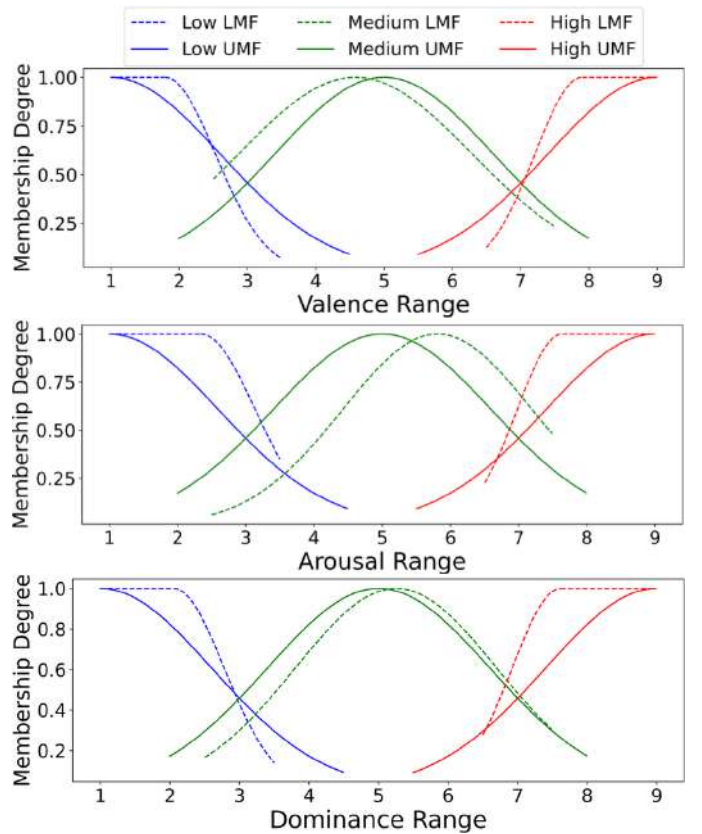


Fig. 4: A Graphical View of the Membership Functions for VAD Dimensions

In this representation, crisp values of the VAD are converted into the representations of low, medium and high. Membership

Dimension	M <sub>Low</sub> <sup>UMF</sup>	$\sigma_{Low}^{UMF}$	M <sub>Med</sub> <sup>UMF</sup>	$\sigma_{Med}^{UMF}$	M <sub>High</sub> <sup>UMF</sup>	$\sigma_{High}^{UMF}$	M <sub>Low</sub> <sup>LMF</sup>	$\sigma_{Low}^{LMF}$	M <sub>Med</sub> <sup>LMF</sup>	$\sigma_{Med}^{LMF}$	M <sub>High</sub> <sup>LMF</sup>	$\sigma_{High}^{LMF}$
Valence	1	1.2	5	1.2	9	1.2	1.96	0.92	4.30	1.97	7.63	0.92
Arousal	1	1.2	5	1.2	9	1.2	2.8	1.02	6.04	1.54	7.26	0.90
Dominance	1	1.2	5	1.2	9	1.2	2.40	1.01	5.02	1.81	7.27	0.90

TABLE II: Means and Standard Deviations Used in this Work for Various Populations of Emotion Samples. LMF means and standard deviations are based on the real population within the range defined, while UMF means and standard deviations are given based on the generic understanding of the perception.

functions are drawn in Fig. 4, and membership degrees are defined as follows:

Upper Membership Functions (UMF):

$$\mu_{Low}^{UMF}(X^{dim}) = \begin{cases} e^{-\frac{(X^{dim} - M_{Low}^{UMF})^2}{(\sigma_{Low}^{UMF})^2}} & \text{for } X^{dim} \leq Low_u^{UMF} \\ 0 & X^{dim} > Low_u^{UMF} \end{cases} \quad (3)$$

$$\mu_{Med}^{UMF}(X^{dim}) = \begin{cases} 0 & \text{for } X^{dim} < Med_l^{UMF} \\ e^{-\frac{(X^{dim} - M_{Med}^{UMF})^2}{(\sigma_{Med}^{UMF})^2}} & Med_l^{UMF} \leq X^{dim} \leq Med_u^{UMF} \\ 0 & X^{dim} > Med_u^{UMF} \end{cases} \quad (4)$$

$$\mu_{High}^{UMF}(X^{dim}) = \begin{cases} 0 & \text{for } X^{dim} < High_l^{UMF} \\ e^{-\frac{(X^{dim} - M_{High}^{UMF})^2}{(\sigma_{High}^{UMF})^2}} & \text{for } X^{dim} \geq High_l^{UMF} \end{cases} \quad (5)$$

Lower Membership Functions (LMF):

$$\mu_{Low}^{LMF}(X^{dim}) = \begin{cases} 1 & \text{for } X^{dim} \leq M_{Low}^{LMF} \\ e^{-\frac{(X^{dim} - M_{Low}^{LMF})^2}{(\sigma_{Low}^{LMF})^2}} & M_l^{LMF} < X^{dim} \leq Low_u^{LMF} \\ 0 & X^{dim} > Low_u^{LMF} \end{cases} \quad (6)$$

$$\mu_{Med}^{LMF}(X^{dim}) = \begin{cases} 0 & \text{for } X^{dim} < Med_l^{LMF} \\ e^{-\frac{(X^{dim} - M_{Med}^{LMF})^2}{(\sigma_{Med}^{LMF})^2}} & Med_l^{LMF} \leq X^{dim} \leq Med_u^{LMF} \\ 0 & X^{dim} > Med_u^{LMF} \end{cases} \quad (7)$$

$$\mu_{High}^{LMF}(X^{dim}) = \begin{cases} 0 & \text{for } X^{dim} < High_l^{LMF} \\ e^{-\frac{(X^{dim} - M_{High}^{LMF})^2}{(\sigma_{High}^{LMF})^2}} & High_l^{LMF} \leq X^{dim} < M_{High}^{LMF} \\ 1 & X^{dim} \geq M_{High}^{LMF} \end{cases} \quad (8)$$

The membership function, denoted by  $\mu$ , represents the degree of membership.  $dim$  represents the emotion dimension i.e.,  $dim \in \{V, A, D\}$ . The symbol  $X^{dim}$  represents the input value for that dimension, and  $X^{dim} \in [1, 9]$ .  $M$  and  $\sigma$  represent the mean and standard deviation of the various ranges for Gaussian distributions.

$$\begin{aligned} Range(Low^{UMF}) &\in [1, Low_u^{UMF}] \\ Range(Med^{UMF}) &\in [Med_l^{UMF}, Med_u^{UMF}] \\ Range(High^{UMF}) &\in [High_l^{UMF}, 9] \\ Range(Low^{LMF}) &\in [1, Low_u^{LMF}] \\ Range(Med^{LMF}) &\in [Med_l^{LMF}, Med_u^{LMF}] \\ Range(High^{LMF}) &\in [High_l^{LMF}, 9] \end{aligned}$$

The membership functions are shown in Fig.4. For practical purposes, we set the ranges as:  $Low_u^{UMF} = 4.5, Med_l^{UMF} = 2, Med_u^{UMF} = 8, High_l^{UMF} = 5.5, Low_u^{LMF} = 3.5, Med_l^{LMF} = 2.5, Med_u^{LMF} = 7.5, High_l^{LMF} = 6.5$ . The corresponding mean, standard deviation and ranges are shown in TableII. These membership functions are modified where they intersect to define the Footprint of Uncertainty (FoU), and the means are adjusted accordingly.

$$FoU = F(\mu^{UMF}, \mu^{LMF}) \quad (9)$$

where  $F$  represents a function to determine the FoU and coincide the means to a single point.  $\mu^{UMF}$  and  $\mu^{LMF}$  represent the UMF and LMF memberships.

## IV. METHODOLOGY

### A. Data Preparation

1) *Emotional Events Extraction and Preprocessing*: In the initial phase of EEG signal processing, the raw data, initially referenced to the Cz electrode, was imported. Subsequently, average re-referencing was applied. The EEG signal underwent filtering using a fifth-order Butterworth bandpass filter, with a low cutoff frequency of 1 Hz and a high cutoff frequency of 40 Hz. Segmentation of the filtered signal was performed to isolate event segments corresponding to the baseline state and specific emotional events (captured using mouse clicks). The baseline state was defined as the period from 10 to 70 seconds, while the click events were extracted from -6 to +1 seconds relative to the click response for the emotional event by the participants. The duration of the extracted click events spanned from 6 seconds before to 1 second after the event.

To mitigate the impact of eyeblink activity, Independent Component Analysis (ICA) was employed. This step was crucial to prevent automatic channel rejection based on high amplitudes associated with eye artefacts. After the initial automatic channel rejection, a second ICA iteration was implemented to eliminate any residual artefacts, such as muscle and heart activity, line noise, and channel noise. The ICLabel tool categorized independent components into different labels [48], and components with a brain activity probability greater than 0.3 were retained. The probabilities for all labels equal to 1 were aggregated.

Upon completion of the pre-processing steps, discrete occurrences corresponding to the baseline state and emotional responses were stored, facilitating subsequent data analysis.

2) *Feature Extraction and Input Formatting*: The Short-Time Fourier Transform (STFT) enables the analysis of a signal's frequency content as it evolves over time, which is critical for understanding dynamic processes such as changes in brain activity captured by EEG signals. Since EEG signals

are non-stationary, STFT provides a local analysis that adapts to these changes, making it well-suited for analyzing signals with varying characteristics. It offers the ability to capture temporal and frequency information simultaneously.

The data received from the  $i^{\text{th}}$  EEG channel is  $x_i$ , The STFT of  $x_i$  is represented as

$$X_i[m, n] = \sum_{k=0}^{N-1} x_i[n+k] \cdot w[k-m] \cdot e^{-j2\pi f k/N}$$

where STFT coefficient  $X_i$  represented at a specific time index  $n$  and frequency index  $m$  for the signal  $x_i[n]$ . A sliding window approach is used, where the input signal  $x_i$  is shifted by  $k$  samples. The term  $w[k-m]$  denotes the value of the window function at the relative position  $k-m$ . The application of the window function is integral in minimizing spectral leakage and refining the precision of the STFT. The complex exponential modulation term  $e^{-j2\pi f k/N}$  introduces a frequency-dependent weighting factor. Here,  $f$  denotes the frequency index,  $k$  represents the sample index within the window, and  $N$  stands for the total number of samples. This modulation is pivotal in elucidating the frequency content within the defined window. We used 50% overlapping to compute the STFT.

The spectrograms computed by,

$$S_i[m, n] = |X_i[m, n]|^2 \quad (10)$$

After this,  $l$  spectrograms are stacked further to create an input to be given in the architecture as Input1,

$$S_{l,m,n} := S_i[l, m, n] \quad (11)$$

### B. Architecture with Deep Fuzzy Framework

In Fig.5, we have presented the overall framework of our architecture, consisting of three modules. The first module captures spatial features, while the second module captures temporal features. The third module utilizes a fuzzy framework, demonstrated in Eq. 2, to provide the membership degree of emotion to the VAD space. We built three distinct models based on this framework that contribute to three different representations of emotions, which are discussed in detail in this section.

1) *Modules in the Architecture:* There are three modules in the architecture: First for spatial features, second for temporal features, and third for fuzzy representational block. Our emotion recognition architecture leverages both spatial and temporal information of EEG signals. To extract significant patterns and features from spectrograms, we used Convolutional Neural Networks (CNNs). Moreover, Long Short-Term Memory (LSTM) networks are highly effective in analyzing time-series data and modelling temporal dependencies. As a result, they are an ideal solution for this emotion recognition from EEG. Lastly, we have a parallel module for providing a VAD representational block.

a) *Spatial Module:* In this process, we use kernels in the convolution process to learn local patterns from the input data. These patterns are then combined and refined through multiple layers of CNNs to yield more complex and meaningful

features. We utilized two layers of CNN in our models. All the layers have dropout masks with a 20% dropout rate. The output of the spatial module is calculated by the following set of equations:

$$\begin{aligned} Y_{i,j,k}^{[1]} &= \text{MaxPooling}(\text{ReLU}(\text{Conv}(S, W^{[1]}, b^{[1]})), p_m, p_n) \\ Y_{i,j,k}^{[2]} &= \text{MaxPooling}(\text{ReLU}(\text{Conv}(Y^{[1]}, W^{[2]}, b^{[2]})), p_m, p_n) \\ Y_{\text{flattened}}^{[2]} &= \text{Flatten}(Y^{[2]}) \end{aligned} \quad (12)$$

where  $Y_{i,j,k}^{[1]}$  and  $Y_{i,j,k}^{[2]}$  represent the outputs of the first and second convolutional layers, respectively at position  $(i,j,k)$ .  $S$  is the Input tensor from Eq.11.  $W^{[1]}, W^{[2]}$  represent weights of the convolutional layers.  $b^{[1]}, b^{[2]}$  are the biases of the convolutional layers.  $Y_{\text{flattened}}^{[2]}$  represents the flattened output tensor from the second convolutional layer.  $(p_m, p_n)$  is the window size of the pooling layer. All the layers have dropout masks with a 20% dropout rate.

The convolution operation, denoted as  $\text{Conv}(X, W^{[1]}, b^{[1]})$ , is applied to the input tensor  $X$  using weights  $W^{[1]}$  and biases  $b^{[1]}$ . Following the convolution, the Rectified Linear Unit (ReLU) activation function,  $\text{ReLU}(\cdot)$ , is applied element-wise to the result. Mathematically, ReLU is defined as  $\text{ReLU}(x) = \max(0, x)$ . Finally, a max-pooling operation, denoted as  $\text{MaxPooling}(\cdot, p_m, p_n)$ , is performed on the output with a window size of  $p_m \times p_n$ . We used a  $(2 \times 2)$  pooling channel with a  $(3 \times 3)$  kernel size of the convolution. We used 32 and 64 kernels for first and second CNN layers, respectively.

Conv operation we used as:

$$Y_{i,j,k} = \sum_{a=1}^{k_m} \sum_{b=1}^{k_n} \sum_{c=1}^l W_{a,b,c,k} \cdot X_{i+a-1,j+b-1,c} + b_k \quad (13)$$

The symbol  $Y_{i,j,k}$  denotes the element at position  $(i,j,k)$  in the output tensor. Similarly,  $W_{a,b,c,k}$  represents the weight located at position  $(a,b,c,k)$  within the filter  $W$ . Moving to the input tensor,  $X_{i+a-1,j+b-1,c}$  corresponds to the element situated at position  $(i+a-1, j+b-1, c)$ . Additionally,  $b_k$  is the bias term associated with the  $k$ -th filter of size  $(k_m \times k_n)$ .  $Y_{\text{flattened}}$  will be used for temporal module.

b) *Temporal Module:* We use a hybrid architecture for emotion categorisation to take advantage of both Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). The hybrid CNN-LSTM model combines the CNN's ability to extract features from data with the LSTM's ability to capture both long-term and short-term dependencies.

The output from the Flatten layer, denoted as  $Y_{\text{flattened}}^{[2]}$ , is sequentially processed through two Long Short-Term Memory (LSTM) layers. Before the LSTM layers, a Repeat Sequence operation is applied. This operation, often used in sequence-to-sequence tasks, repeats the output sequence from the Flatten Vector and prepares it for further processing.  $R$  represents the number of Repeat Sequence. We used  $R=4$  for further processing of the Output received from the spatial module.

$$\begin{aligned} Y_{\text{rep}} &= \text{RepeatSequence}(Y_{\text{flattened}}^{[2]}, R) \\ H_r^{[1]}, C_r^{[1]} &= \text{LSTM}(Y_{\text{rep}}, H_{r-1}^{[1]}, C_{r-1}^{[1]}) \\ H_r^{[2]}, C_r^{[2]} &= \text{LSTM}(H_r^{[1]}, H_{r-1}^{[2]}, C_{r-1}^{[2]}) \end{aligned} \quad (14)$$

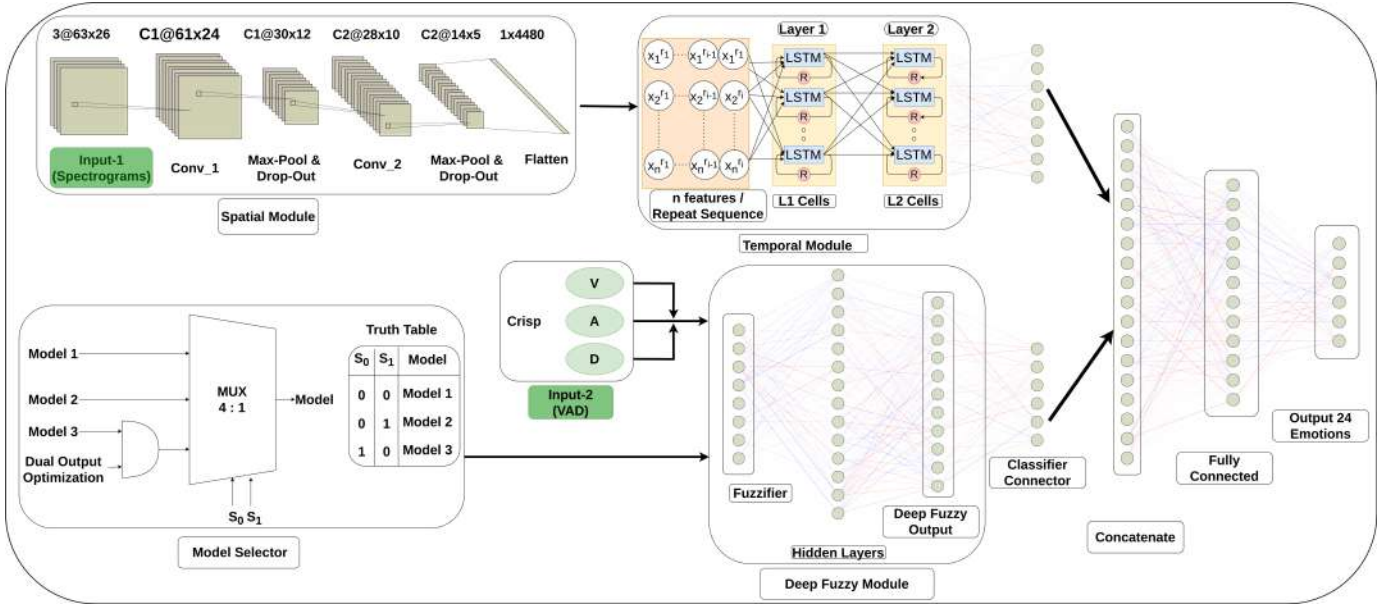


Fig. 5: Deep Fuzzy Architecture: This architecture consists of three modules. The spatial module is based on CNNs, the Temporal module is based on LSTMs, and the third module is our proposed Fuzzy Framework. This framework further represents three more models based on VAD fuzzy representations (see Fig.6 for models).

In the equations,  $Y_{rep}$  Repeat Sequence operation.  $H_r^{[1]}$ ,  $C_r^{[1]}$  represent the hidden state and cell state of the first LSTM layer at repetition  $r$ , while  $H_r^{[2]}$ ,  $C_r^{[2]}$  denote the corresponding states for the second LSTM layer. The LSTM operations involve gating mechanisms, such as input gates, forget gates, and output gates, influencing the flow of information and memory retention in the cell states. The first LSTM layer processes the repeated sequence, and its output, along with the previous hidden and cell states, becomes the input for the second LSTM layer.

In the provided equations, the final output of the second LSTM layer is denoted as  $H_r^{[2]}$ , representing the hidden state at the final time step  $r$ . This hidden state encapsulates the learned representation of the input sequence after processing through the first and second LSTM layers.  $H_r^{[2]}$  will be used as a feature representation and serves as the input for subsequent layers, which concatenates with the input from the deep fuzzy module. It captures the hierarchical dependencies and patterns in the sequential data, providing a meaningful representation for further analysis or decision-making.

2) *Integration of Deep Fuzzy Module:* In this approach, we developed three models based on fuzzy representations of emotions. The aim is to use a computational approach for decision-making in the final classification. To achieve this, we utilized membership functions, unsupervised clustering, and supervised cubical probabilities. We developed three different models based on three distinct representations of emotion: firstly, degree of membership values (Type-1 and Type-2), secondly unsupervised fuzzy clusters, and thirdly, cuboid probabilistic lattice representation of the VAD dimensions. The models are schematically represented in Fig. 6. The fuzzy module consists of a Fuzzifier sub-module, which converts the crisp VAD values into fuzzy membership degrees, a Model

Selector, according to which membership degrees are assigned (more details about the membership degrees in individual models). and Fully Connected neural network. The last layer acts as an output for this module. The membership degree of the VAD\_Input is,

$$\Gamma = \phi(VAD\_Input) \quad (15)$$

where  $\phi$  is a function Fuzzifier. This function is defined differently for each model, which is discussed later in this section.

$\Gamma$  is further processed through Dense layers with ReLU and Dropout layers inbuilt with  $(\cdot)$  product in initial layers, while the final layer is processed differently with different models.

$$\begin{aligned} O_{FC}^{[i]} &= \Gamma \cdot W_{FC}^{[i]} + b_{FC}^{[i]} \\ O_{FC}^{[f]} &= O_{FC}^{[i]} \cdot W_{FC}^{[f]} + b_{FC}^{[f]} \end{aligned} \quad (16)$$

where  $O_{FC}^{[i]}$  and  $O_{FC}^{[f]}$  are the outputs of the initial layers and final layers, respectively and  $W$  and  $b$  are the weight and bias, respectively.

a) *Model-1:* Model-1 uses the type-2 fuzzy membership function for  $\phi$  in Eq.15. These membership functions are mentioned in Section-III with the equations from Eq.3 to Eq.8.

b) *Model-2:* Model-2 is based on unsupervised clustering. We used the FCM algorithm to define the degrees of each clusters in the VAD space. Here  $\phi$  in Eq.15 is updated as,

$$u_{ij} = \left( \sum_{k=1}^c \left( \frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (17)$$

where The membership degrees are represented by a matrix  $U$  where  $u_{ij}$  denotes the degree of membership of data point  $i$  to cluster  $j$ . Here,  $i$  ranges from 1 to  $n$  (number of data points), and  $j$  ranges from 1 to  $c$  (number of clusters).



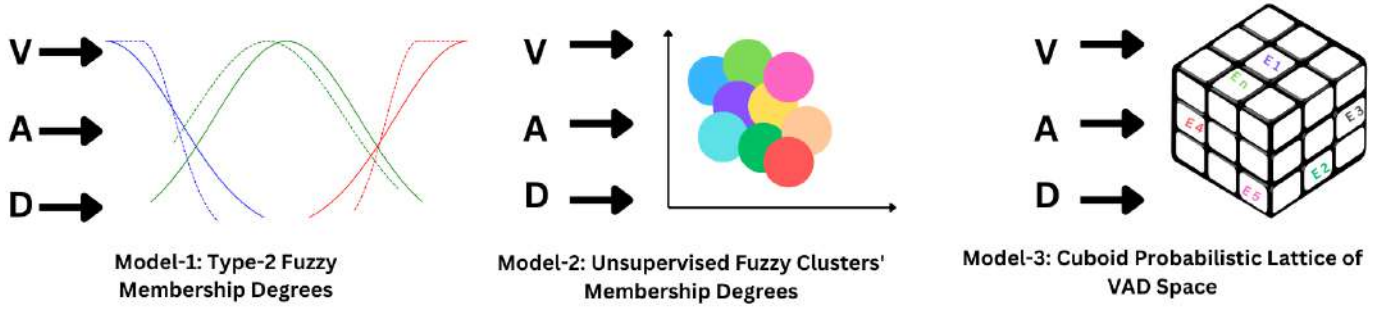


Fig. 6: A Schematic Representation of Developed Models: Models are based on various fuzzy representations. The first model is based on the type-2 membership functions outlined in the equations from Eq.3 to Eq.8. The next model is based on unsupervised learning and uses fuzzy C-Means membership degrees for various numbers of clusters. The last model is based on the cuboid probabilistic lattice representation of emotion in VAD space. Model-3 takes fuzzy input and optimises based on the probabilities of low, medium, and high VAD cuboids. It uses dual output- one for VAD and the other for the final 24 emotions classification.

c) *Model-3*: Model-3 uses the type-2 fuzzy membership function for  $\phi$  in Eq.15 with the equations from Eq.3 to Eq.8. However, the output layer of the fuzzy framework module in this model is used as a dual output and optimized simultaneously with the main output layer. The second output is trained on the classification of VAD space in 27 classes, just like a cuboid is divided into low, medium, and high values of its dimension. We used Softmax in the last layer of this fuzzy module, i.e., Softmax is applied on  $O_{FC}^{[f]}$  in Eq.16.

$$p_i = \text{Softmax}(O_{FC}^{[f]})_i = \frac{e^{o_i}}{\sum_{j=1}^k e^{o_j}} \quad (18)$$

where  $i$  represents the index of the element in the vector, and  $k$  is the number of dense units used in the final output of the layer, i.e., the number of VAD space classes used in the second output layer. And  $O_{FC}^{[f]} = [o_1, o_2, \dots, o_k]$ . We took  $k=27$ .

We call the vector  $[p_1, p_2, \dots, p_k]$  as cuboid probabilistic lattices as it contains the probability of the VAD space mapped to  $k$  cuboids. These lattices are the output of Model-3.

### C. Cross-Subject Study

Group 1		Group 2		Group 3	
Emotions	#	Emotion	#	Emotions	#
Delighted	11	Melancholic	5	Taken Aback	11
Amused	14	Depressed	6	Startled	20
Happy	5	Despondent	10	Distress	22
Adventurous	20	Dissatisfied	10	Alarmed	27
Joyous	20	Miserable	19	Afraid	46
Excited	26	Sad	45		
Total	108	Total	95	Total	126

TABLE III: Group Formation for Cross-Subject Emotion Recognition: Each group consists of several emotions, with samples mentioned against each emotion.

In order to validate our model, we extended our work to test cross-subject emotion recognition. The cross-subject task is more complex in nature [49]. Therefore, we simplified the task by grouping together a few emotions based on a tree-structured list [50], [51]. It is worth noticing here that we did not use VAD information to group these emotions to avoid creating

any biases in the model. Three distinct groups are formed. Emotions considered in these three groups are mentioned in TableIII.

We split the samples in such a way that the training and validation sets do not have data from the same participants.

## V. EXPERIMENTS AND RESULTS

We have created three models, as discussed earlier. We performed the experiments with all the models to verify the sanctity of our framework and methodology.

Model-1 is tested with type-2 fuzzy membership degrees and 24 emotion classes. Model-2, which is based on the unsupervised fuzzy clusters, we experimented with various numbers of clusters. The optimum number of clusters is 4 with the fuzzy Silhouette coefficient. We started with 4 clusters and ran the model up to 10 clusters. Results are shown in Fig.7. Model-3 is based on the cuboid probabilistic lattice and also utilizes dual output. It receives the input of type-2 fuzzy membership degrees with 24 classes. All the results of the three models are shown in TableIV with state-of-the-art works. Model-1 performed best with 96.09% accuracy, followed by Model-3 with 95.75% accuracy and the last, Model-2 with 95.31% accuracy.

We also explored cross-subject emotion classification, which is considered a relatively more challenging task. Results are shown in TableV. Results show that Group-1 vs Group-2 secures the best accuracy. Improvements when using a deep fuzzy framework are shown in Fig.8.

An ablation study with many experiments is also conducted to check the performance of our proposed deep fuzzy framework. Firstly, we evaluated the performance of our models under various conditions without the fuzzy framework and only the module working as a crisp VAD space. secondly, we removed all the VAD space and recognised the emotions with only spatial and temporal modules. Thirdly, individual UMF and individual LMF working as type-1 fuzzy are evaluated to check the performance of type-2 fuzzy sets with a comparison to type-1 sets. All the results of the ablation study are shown in TableVI.

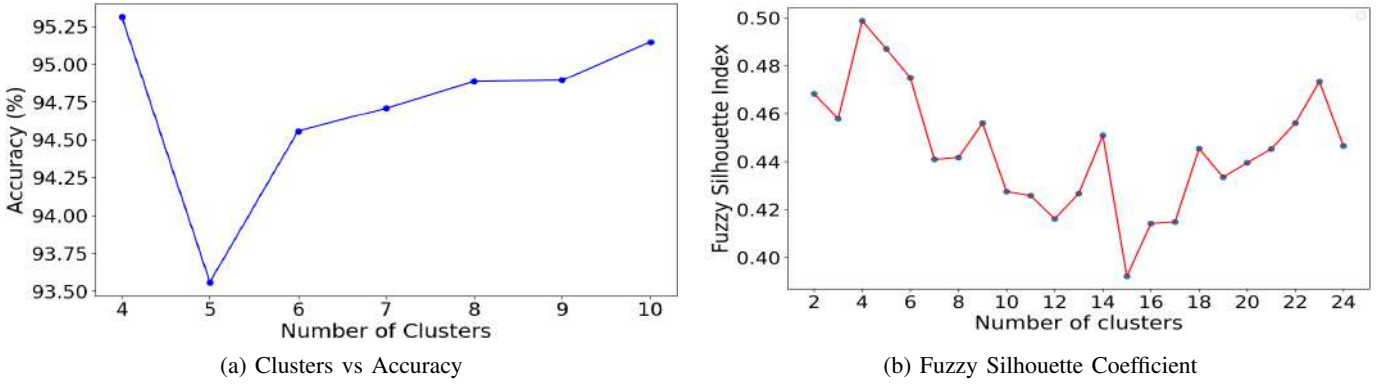


Fig. 7: Comparison of Accuracies With Model-2 Based on Fuzzy C-Means

Method	Emotions & Dataset	#Classes	Accuracy (%)
DGCNN (Directed Graph CNN) [52]	V, A, D [DREAMER]	2	V: 86.23, A: 84.54, D: 85.02
RGNN (Regularized GNN) [53]	3 Emotions (Neutral, Positive, Negative) and 4 Emotions (Neutral, Sad, Fear, Happy) [SEED]	3 & 4	SEED: 94.24 SEED-IV: 79.37
ACRNN (Attentional C-RNN) [54]	V, A [DEAP], V, A, D [DREAMER]	2	DEAP- V: 93.72, A: 93.38 DREAMER- V:97.93, A:97.98, D:98.23
CNN-LSTM Hybrid Model with STFT [21]	Low and High V & A [DENS]	4	96.82
ATDD-LSTM(Attention based LSTM) [55]	V, A [DEAP] & [CMEED]	2	DEAP- V: 90.91, A: 90.87 CMEED - V: 91.53, A: 91.55
Temporal or Spectral Features, and ZTWBES and RNN [25]	9 Emotions [DEAP]	9	Three Different Models: 87.05, 89.33, 86.53
R2G-STNN Model (region to global BiLSTM with Attention) [56]	Positive, Negative, Neutral [SEED]	3	93.38
ECLGCNN (A fusion model of GCNN + LSTM) [57]	V, A [DEAP]	2	V: 90.45, A: 90.60
<b>Our Work: Model-1 (Type-2 Fuzzy Membership)</b>	<b>24 [DENS]</b>	<b>24</b>	<b>96.09</b>
<b>Our Work: Model-2 (Unsupervised Fuzzy Clusters)</b>	<b>24 [DENS]</b>	<b>24</b>	<b>95.31</b>
<b>Our Work: Model-3(Cubical Probabilistic Lattice)</b>	<b>24 [DENS]</b>	<b>24</b>	<b>95.75</b>

TABLE IV: Results of All the Models Comparing with State of the Art Methods.

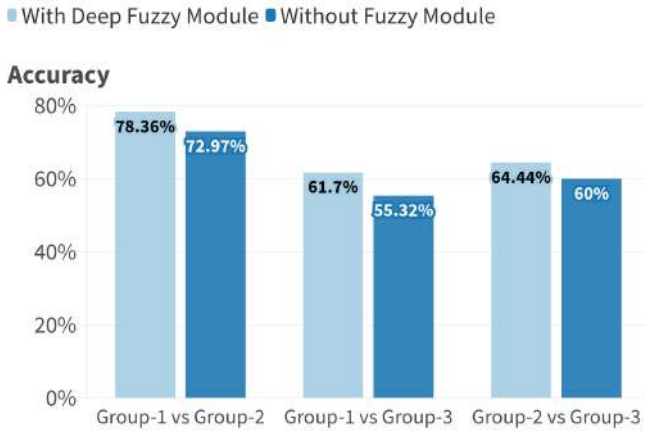


Fig. 8: Cross Subject Accuracy Comparison- with and without Deep Fuzzy Module

Groups	Model	Framework	#Classes	Accuracy
G-1 vs G-2	Model-1	<b>with fuzzy</b>	2	<b>78.35%</b>
G-1 vs G-2	Model-1	without fuzzy	2	72.97%
G-1 vs G-3	Model-1	<b>with fuzzy</b>	2	<b>61.70%</b>
G-1 vs G-3	Model-1	without fuzzy	2	55.31%
G-2 vs G-3	Model-1	<b>with fuzzy</b>	2	<b>64.44%</b>
G-2 vs G-3	Model-1	without fuzzy	2	60.00%

TABLE V: Results of Cross-Subject Emotion Recognition.

Model	Ablation Method	Classes	Accuracy
Model-1	without Fuzzy framework (with Crisp VAD values)	24	95.01%
Model-1	without Fuzzy framework (without any VAD values)	24	93.54%
Model-1	UMF working as Type-1 Fuzzy	24	95.82%
Model-1	LMF working as Type-1 Fuzzy	24	94.65%
Model-3	UMF working as Type-1 Fuzzy	24	95.16%
Model-3	LMF working as Type-1 Fuzzy	24	95.22%

TABLE VI: Results of Ablation Study

## VI. DISCUSSION AND CONCLUSION

The results mentioned in Table IV show that our models yield better accuracies even on 24 classes, while other studies mostly focus on either 2 or 3 classes only. The utilization of a fuzzy VAD space and fusion with the measured EEG data have helped in the development of a deep fuzzy framework for emotion representation and recognition. The results shown are promising, and the highest accuracy achieved is 96.09% with a type-2 fuzzy framework (Model-1) (Section-IV-B2a).

The state-of-the-art table includes the studies ranging on various datasets. The DENS dataset, unlike others, has given an option to the subjects to choose the name and time of their emotional experience in the value of V, A and D. However, consideration of raw VAD values could be an oversimplification since mental activities may not be accurately represented

using only three numbers. To address this issue, we fuzzified the VAD values and incorporated EEG (Electroencephalography) values in our deep fuzzy model. The process of fuzzifying VAD values improved the overall accuracy of the model in comparison to the base model.

By introducing the fuzziness of type-2, the model becomes more adaptive to individual variations in emotional experience, enhancing its versatility across different experiments and diverse individual responses. With the type-2 fuzzy framework, we have the leverage to take advantage of both the general representation of the emotion and the population-specific representation of the emotion if we use them as LMF and UMF. On the other hand, using type-1 representation, we obtained an accuracy of 95.82% and 94.66%, respectively, when we used these UMF and LMF separately working as type-1. This further supports our theory that this model works well as a generic model for emotion representation and recognition.

Cuboid probabilistic lattice-based representation (Model-3)(Section-IV-B2c) achieves an accuracy of 95.75% when input is given by a type-2 membership degree. However, the accuracy decreases to 95.16% and 95.22%, respectively, when independently using UMF and LMF as type-1 membership degrees. Nevertheless, this cuboid-based probability lattice is able to enhance the performance almost to the level of when UMF is used as type-1 in the previous model. This is due to the fact that this model is dual-output based, and cuboid probabilistic lattices are optimized for the next input to the EEG features, thereby improving performance.

Finally, when it comes to unsupervised learning using a fuzzy C-means cluster-based model (Model-2)(Section-IV-B2b), the model's accuracy tends to fluctuate depending on the number of clusters used. To determine the optimal number of clusters, we can use the fuzzy Silhouette coefficient. In this case, the coefficient suggests that four clusters is the optimal number. Beyond five clusters, the accuracy increases a bit but again drops.

We conducted an ablation study to evaluate the performance of our model in two scenarios: with no VAD space integrated and with crisp VAD space integrated into the architecture. The results showed that when no VAD space was used, the model's performance was as low as 93.54%. However, with the crisp VAD space, the model's performance improved to 95.01%. Hence, supporting the need for a type-2 fuzzy framework. Additionally, incorporating both physiological data (EEG features) and subjective ratings enhances the model's ability to capture the multidimensional aspects of emotions, leading to more reliable recognition outcomes.

In cross-subject analysis, our model achieves an accuracy of 78.37%, which is also on par with the other recent works [49], [58], [59]. Notably, our ablation study showed that changing no other parameters resulted in a significant improvement in accuracy from 72.97% to 78.37% (Fig.8) if we add a fuzzy framework to the existing architecture. Generally, cross-subject studies do not yield significant accuracies [58], [60], [61]. Also, a comparison of the other two groups shows a similar kind of improvement after adding the fuzzy framework to the existing architecture; however, they elicit a bit low accuracy but are still on par with the recent trends. This signifies

the generalizability of the proposed architecture, showcasing its potential application across diverse circumstances. It is a critical step toward real-world applicability, and the study's success in this aspect further validates the robustness of the developed model.

The study deals with 24 emotions at a time, which is not only challenging but also the first of its kind. Typically, emotion recognition tasks involve identifying emotional dimensions such as low or high valence and arousal. Additionally, they often focus on a few basic emotions, such as happiness, sadness, anger, and joy. The number of emotional classes is also kept low. In this perspective, the study's outcomes have significant implications for emotion recognition research and applications. The generic nature of the developed model, along with its successful cross-subject predictions, gives direction for real-world applications in the field of affective computing, human-computer interaction, mental health monitoring, etc. Future research could explore the adaptation of the model to different cultural contexts and the integration of additional modalities for an even more comprehensive understanding of human emotions, including complex emotions.

#### ACKNOWLEDGMENTS

Special thanks to Ms. Diya Srivastava, for her efforts in this work. Your assistance has been truly appreciated.

#### REFERENCES

- [1] A. Mehrabian, "Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies," 1980.
- [2] P. Saariluoma and J. P. Jokinen, "Emotional dimensions of user experience: A user psychological analysis," *International Journal of Human-Computer Interaction*, vol. 30, no. 4, pp. 303–320, 2014.
- [3] P. Ekman, "Are there basic emotions?" 1992.
- [4] P. E. Griffiths, "Basic emotions, complex emotions, machiavellian emotions," 2002.
- [5] A. Ortony, "Are all 'basic emotions' emotions? a problem for the (basic) emotions construct," *Perspectives on psychological science*, vol. 17, no. 1, pp. 41–61, 2022.
- [6] L. F. Barrett, *How emotions are made: The secret life of the brain*. Pan Macmillan, 2017.
- [7] A. Branicka, E. Trzebińska, A. Dowgiert, and A. Wytykowska, "Mixed emotions and coping: The benefits of secondary emotions," *PloS one*, vol. 9, no. 8, p. e103940, 2014.
- [8] J. J. Gross and O. P. John, "Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being," *Journal of personality and social psychology*, vol. 85, no. 2, p. 348, 2003.
- [9] P. van Rijn and P. Larrouy-Maestri, "Modelling individual and cross-cultural variation in the mapping of emotions to speech prosody," *Nature Human Behaviour*, vol. 7, no. 3, pp. 386–396, 2023.
- [10] L. F. Barrett, "Feelings or words? understanding the content in self-report ratings of experienced emotion," *Journal of personality and social psychology*, vol. 87, no. 2, p. 266, 2004.
- [11] K. E. Smidt and M. K. Suvak, "A brief, but nuanced, review of emotional granularity and emotion differentiation research," *Current Opinion in Psychology*, vol. 3, pp. 48–51, 2015.
- [12] G. A. Van Kleef and S. Côté, "The social effects of emotions," *Annual review of psychology*, vol. 73, pp. 629–658, 2022.
- [13] D. Li, M. Wu, X. Zhang, M. Wang, and J. Shi, "The roles of fluid intelligence and emotional intelligence in affective decision-making during the transition to early adolescence," *Frontiers in Psychology*, vol. 11, p. 574903, 2020.
- [14] L. Zadeh, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103–111, 1996.
- [15] J. Kumar et al., "Affective modelling of users in hci using eeg," *Procedia Computer Science*, vol. 84, pp. 107–114, 2016.

- [16] A. R. Daros, S. A. Haefner, S. Asadi, S. Kazi, T. Rodak, and L. C. Quilty, "A meta-analysis of emotional regulation outcomes in psychological interventions for youth with depression and anxiety," *Nature human behaviour*, vol. 5, no. 10, pp. 1443–1457, 2021.
- [17] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on autonomous mental development*, vol. 7, no. 3, pp. 162–175, 2015.
- [18] N. Thammasan, K.-i. Fukui, and M. Numao, "Application of deep belief networks in eeg-based dynamic music-emotion recognition," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 881–888.
- [19] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby, "Eeg-based emotion recognition using 3d convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018.
- [20] S. Sheykhiavand, Z. Mousavi, T. Y. Rezaii, and A. Farzamnia, "Recognizing emotions evoked by music using cnn-lstm networks on eeg signals," *IEEE access*, vol. 8, pp. 139 332–139 345, 2020.
- [21] M. Asif, S. Mishra, M. T. Vinodhrai, and U. S. Tiwary, "Emotion recognition using temporally localized emotional events in eeg with naturalistic context: Dens dataset," *IEEE Access*, vol. 11, pp. 39 913–39 925, 2023.
- [22] C. Li, B. Chen, Z. Zhao, N. Cummins, and B. W. Schuller, "Hierarchical attention-based temporal convolutional networks for eeg-based emotion recognition," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1240–1244.
- [23] Z. Hu, L. Chen, Y. Luo, and J. Zhou, "Eeg-based emotion recognition using convolutional recurrent neural network with multi-head self-attention," *Applied Sciences*, vol. 12, no. 21, p. 11255, 2022.
- [24] L. Chen, W. Su, M. Wu, W. Pedrycz, and K. Hirota, "A fuzzy deep neural network with sparse autoencoder for emotional intention understanding in human-robot interaction," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1252–1264, 2020.
- [25] S. Gannouni, A. Aledaily, K. Belwafi, and H. Aboalsamh, "Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification," *Scientific Reports*, vol. 11, no. 1, p. 7071, 2021.
- [26] N. Salankar, P. Mishra, and L. Garg, "Emotion recognition from eeg signals using empirical mode decomposition and second-order difference plot," *Biomedical Signal Processing and Control*, vol. 65, p. 102389, 2021.
- [27] M. Sorkhi, M. R. Jahed-Motlagh, B. Minaei-Bidgoli, and M. R. Daliri, "Hybrid fuzzy deep neural network toward temporal-spatial-frequency features learning of motor imagery signals," *Scientific Reports*, vol. 12, no. 1, p. 22334, 2022.
- [28] N. Talpur, S. J. Abdulkadir, H. Alhussian, . M. H. Hasan, N. Aziz, and A. Bamhdi, "A comprehensive review of deep neuro-fuzzy system architectures and their optimization methods," *Neural Computing and Applications*, pp. 1–39, 2022.
- [29] S. Mishra, M. Asif, and U. Tiway, "Dataset on emotions using naturalistic stimuli (dens)," 2021.
- [30] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*. the MIT Press, 1974.
- [31] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 english lemmas," *Behavior research methods*, vol. 45, pp. 1191–1207, 2013.
- [32] S. Mishra, N. Srinivasan, and U. S. Tiwary, "Cardiac-brain dynamics depend on context familiarity and their interaction predicts experience of emotional arousal," *Brain Sciences*, vol. 12, no. 6, p. 702, 2022.
- [33] —, "Dynamic functional connectivity of emotion processing in beta band with naturalistic emotion stimuli," *Brain sciences*, vol. 12, no. 8, p. 1106, 2022.
- [34] L. Feldman Barrett and J. A. Russell, "Independence and bipolarity in the structure of current affect," *Journal of personality and social psychology*, vol. 74, no. 4, p. 967, 1998.
- [35] L. F. Barrett, "Are emotions natural kinds?" *Perspectives on psychological science*, vol. 1, no. 1, pp. 28–58, 2006.
- [36] L. Barrett, K. Quigley, E. Bliss-Moreau, and K. Aronson, "Arousal focus and interoceptive sensitivity," *Journal of Personality and Social Psychology*, vol. 87, no. 5, pp. 684–697, 2004.
- [37] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.
- [38] P. Kuppens, F. Tuerlinckx, J. A. Russell, and L. F. Barrett, "The relation between valence and arousal in subjective experience," *Psychological bulletin*, vol. 139, no. 4, p. 917, 2013.
- [39] M. Jerram, A. Lee, A. Negreira, and D. Gansler, "The neural correlates of the dominance dimension of emotion," *Psychiatry Research: Neuroimaging*, vol. 221, no. 2, pp. 135–141, 2014.
- [40] S. Mishra, N. Srinivasan, M. Asif, and U. S. Tiwary, "Affective film dataset from india (afdi): creation and validation with an indian sample," *Journal of Cultural Cognitive Science*, pp. 1–13, 2023.
- [41] R. J. Campello and E. R. Hruschka, "A fuzzy extension of the silhouette width criterion for cluster analysis," *Fuzzy Sets and Systems*, vol. 157, no. 21, pp. 2858–2875, 2006.
- [42] M. Asif, M. T. Vinodhrai, S. Mishra, A. Gupta, and U. S. Tiwary, "Emotion recognition in vad space during emotional events using cnn-gru hybrid model on eeg signals," in *International Conference on Intelligent Human Computer Interaction*. Springer, 2022, pp. 75–84.
- [43] M. S. El-Nasr, J. Yen, and T. R. Iorger, "Flame—fuzzy logic adaptive model of emotions," *Autonomous Agents and Multi-agent systems*, vol. 3, pp. 219–257, 2000.
- [44] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 2022.
- [45] A. M. Ruef and R. W. Levenson, "Continuous measurement of emotion," *Handbook of emotion elicitation and assessment*, pp. 286–297, 2007.
- [46] G. Fernández-Blanco Martín, F. Matía, L. García Gómez-Escalonilla, D. Galan, M. G. Sánchez-Escribano, P. de la Puente, and M. Rodríguez-Cantelar, "An emotional model based on fuzzy logic and social psychology for a personal assistant robot," *Applied Sciences*, vol. 13, no. 5, p. 3284, 2023.
- [47] M. Siemer, I. Mauss, and J. J. Gross, "Same situation—different emotions: how appraisals shape our emotions," *Emotion*, vol. 7, no. 3, p. 592, 2007.
- [48] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "Iclabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, 2019.
- [49] M. Asif, D. Srivastava, A. Gupta, and U. S. Tiwary, "Inter subject emotion recognition using spatio-temporal features from eeg signal," in *2023 27th International Computer Science and Engineering Conference (ICSE)*, 2023, pp. 1–4.
- [50] P. Shaver, J. Schwartz, D. Kirson, and C. O'connor, "Emotion knowledge: further exploration of a prototype approach," *Journal of personality and social psychology*, vol. 52, no. 6, p. 1061, 1987.
- [51] W. G. Parrott, "Emotions in social psychology: Key readings," 2000.
- [52] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.
- [53] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2022.
- [54] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "Eeg-based emotion recognition via channel-wise attention and self attention," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 382–393, 2023.
- [55] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, and H. Wang, "An efficient lstm network for emotion recognition from multichannel eeg signals," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1528–1540, 2022.
- [56] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 568–578, 2022.
- [57] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm," *Applied Soft Computing*, vol. 100, p. 106954, 2021.
- [58] Z. Bai, F. Hou, K. Sun, Q. Wu, M. Zhu, Z. Mao, Y. Song, and Q. Gao, "Sect: A method of shifted eeg channel transformer for emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 10, pp. 4758–4767, 2023.
- [59] G. Xu, W. Guo, and Y. Wang, "Subject-independent eeg emotion recognition with hybrid spatio-temporal gru-conv architecture," *Medical & Biological Engineering & Computing*, vol. 61, no. 1, pp. 61–73, 2023.
- [60] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain-computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.
- [61] Z. Yin, L. Liu, J. Chen, B. Zhao, and Y. Wang, "Locally robust eeg feature selection for individual-independent emotion recognition," *Expert Systems with Applications*, vol. 162, p. 113768, 2020.