Chapter 5

Discussions

This study analyzes EDA signals using ML models to enhance emotion recognition techniques. Our primary objective is to achieve effective classification of emotional states through a series of meticulous optimizations:

1. Initially, the EDA, tonic, and phasic signals were optimized by employing two decomposition methods: cvxEDA and BayesianEDA.

2. Next, phasic segments such as First-half, Second-half, and Whole phasic signals were optimized using two time-frequency representations: STFT and MFC.

3. Finally, various time-encoded methods such as GADF, GASF, MTF, and RP were employed to optimize the windowing approach, utilizing Five and Nine windows on the Second-half phasic signal.

To facilitate these optimizations, a publicly accessible CASE dataset was used, leveraging its repository of EDA signals spanning four distinct emotional states such as amusing, boring, relaxing, and scary. Further, three ML methods, SVM, RF, and XGB, were used to assess the optimization's performance. The following discussions compared the optimization strategies and outcomes with the literature, assessing their effectiveness and implications for emotion recognition.

5.1 Effect of Decomposition Methods

The EDA signals obtained from the CASE dataset underwent pre-processing and decomposition into tonic and phasic signals using two methods, cvxEDA and BayesianEDA. Then, ten temporal features were extracted from these signals, including the EDA, tonic, and phasic. These features were utilized to optimize signals and compare the efficacy of the two decomposition techniques. Further, the significance of these features was assessed using the KW test to determine their statistical significance. Subsequently, the resultant significance features were applied to three ML models, such as SVM, RF, and XGB, and evaluated their performance using various metrics.

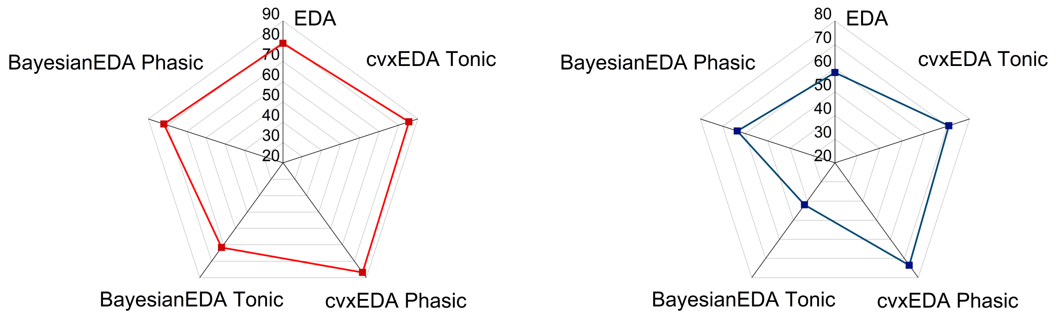
The study explored the behavioral patterns of EDA signals corresponding to four categorical emotions: amusing, boring, relaxing, and scary. Its findings revealed distinct characteristics in the EDA signals across these emotional states. Specifically, EDA signals exhibited minimal variability and a downward trend during relaxing and boring situations. In amusing scenarios, EDA signals displayed low variability with a gradual increase. During subject encounters with scary stimuli, EDA demonstrated heightened variability and an upward trend. Moreover, it was observed that skin conductance levels were notably elevated in response to fear-inducing stimuli compared to other emotions. This response is likely attributed to eccrine sweat glands' activation (Boucsein 2012). Similar observations were reported in previous studies examining stressed and distressed subjects using EDA (Bornoiu and Grigore 2014). The tonic component of EDA signals resembles the overall EDA pattern, characterized by low-frequency components. In contrast, phasic signals represent rapid variations in the EDA signal. These variations were most pronounced during scary situations, followed by amusing scenarios. Conversely, the lowest fluctuations occurred in the phasic signals corresponding to relaxing and boring experiences, nearly approaching baseline levels.

Figure 5.1 illustrates the classification accuracy and F1-score obtained from the XGB model, revealing notable disparities in classification performance across different signals and decomposition methods. Specifically, utilizing EDA signals achieved an accuracy of

78.96% and an F1-score of 58.17%. Employing the tonic signals derived from the cvxEDA decomposition method yielded an accuracy of 85.41% and an F1-score of 70.77% while utilizing the phasic signals from the same decomposition method resulted in even higher accuracy at 86.87% and an F1-score of 73.58%. Similarly, employing tonic signals derived from the BayesianEDA method led to an accuracy of 71.67% and an F1-score of 41.94%, whereas utilizing phasic signals from BayesianEDA achieved higher accuracy at 81.87% and an F1-score of 63.53%.

The findings of this study underscore the significant contribution of the phasic component to emotion classification compared to utilizing the EDA signal. This superiority can be attributed to phasic EDA signals reflecting dynamic changes in skin conductance associated with emotional arousal, offering richer temporal dynamics than tonic signals, representing baseline levels. This difference enables temporal features extracted from phasic signals to capture transient physiological responses indicative of emotional states more effectively. Additionally, by capturing rapid fluctuations in skin conductance, phasic signals provide valuable information about the timing and intensity of emotional onset, which tonic signals may not adequately represent. Consequently, temporal features extracted from phasic EDA signals, with their ability to capture immediate and sustained physiological responses, emerge more effective in discerning emotional states than those from tonic EDA signals. In literature, researchers have utilized features from EDA (Shukla et al. 2019), tonic (Greco, Valenza, Lázaro, et al. 2021), and phasic (Zangróniz et al. 2017) for emotion classification. Still, these studies have yet to compare the effectiveness of using tonic and phasic in classifying emotions. On the other hand, a study reported that emotional patterns align with phasic signals, which matches the results (Posada-Quintero, Reljin, et al. 2019).

The results of this study also suggest that the phasic components derived from the cvxEDA decomposition method exhibited superior classification results compared to those obtained from the BayesianEDA method. It reveals that the cvxEDA is often more robust to emotion patterns in EDA signals. It can effectively handle signal variations, leading to a cleaner decomposition than BayesianEDA. While a study reported similar accuracies in emotion classification using cvxEDA and BayesianEDA methods (Veeranki, Diaz, et al. 2024), another study compared cvxEDA and BayesianEDA, reporting higher performance with BayesianEDA (Sriram Kumar et al. 2023). This discrepancy may be attributed to differences in signal length and the types of features utilized in each study. In summary, the optimized signal for emotion classification is the phasic component of EDA, with the cvxEDA decomposition method emerging as the preferred choice for signal decomposition.



(a) (b)

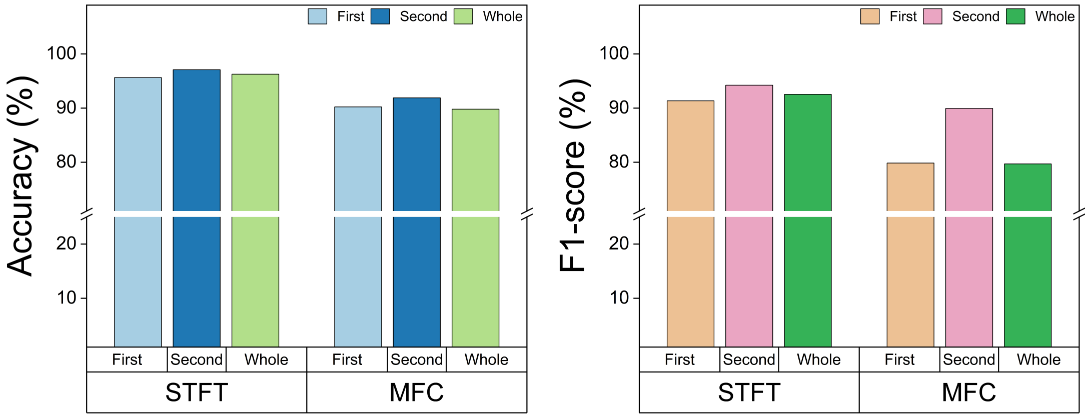
Figure 5.1: Radar diagram representation for the optimization of signals and decomposition methods using performance metrics (a) Accuracy, and (b) F1-score.

5.2 Effect of Phasic EDA Segments

The study aimed to enhance emotion recognition by optimizing the phasic EDA segments, which represent fluctuations in skin conductance associated with emotional arousal. Initially, the phasic signals were segmented into two segments: the First-half phasic (Initial point to the median signal level) and the Second-half phasic (signal starts from the median point to the end of the signal). Following segmentation, the First-half, Second-half, and Whole phasic signals were transformed into spectrograms using two distinct time-frequency domain representations: STFT and MFC. Further, the GLCM and GLRLM features were extracted from these spectrograms. These extracted features were then subjected to the KW test, allowing for the identification of statistical significance (p<0.05).

Subsequently, the significant features were applied to the three ML models, SVM, RF, and XGB, to classify the four emotional states. This comprehensive methodology investigated various aspects of phasic EDA signal processing and feature extraction to improve emotion recognition accuracy.

Our findings revealed varying performance metrics across different segments and types of spectrograms. For instance, when utilizing STFT spectrograms, the SVM classifier achieved high accuracy and F1-score across the different phasic segments: accuracy of 95.62% and F1-score of 91.33% for the First-half segment, accuracy of 97.08% and F1-score of 94.22% for the Second-half segment, and accuracy of 96.25% and F1-score of 92.53% for the Whole phasic signal. However, when employing MFC spectrograms, the SVM classifier demonstrated comparatively lower performance, with accuracy ranging from 89.79% to 91.88% and F1-score ranging from 79.69% to 89.93%.



(a) (b)

Figure 5.2: Bar diagram representation for the optimization of phasic segments using performance metrics: (a) Accuracy, and (b) F1-score.

Our findings shed light on the critical role played by the Second-half segment of the phasic signal, particularly when compared with both the First-half and the Whole phasic signal. The First-half segment of the phasic signal captures the immediate physiological responses that coincide with the onset of emotions. However, this initial phase may be susceptible to interference from external factors such as noise or artifacts, potentially leading to less reliable classification outcomes. On the other hand, the latter segment of the phasic signal encapsulates more enduring physiological responses and a subsequent phase of recovery following the initial emotional arousal. Within this signal, discernible patterns emerge, providing valuable cues essential for accurately classifying emotions. This sustained response period offers a more prosperous and stable source of emotional information, thus enhancing the accuracy of emotional classification models. Our findings align with prior research conducted by Martinez et al. 2019 and SK and JF n.d., who explored the latter half of physiological signals in the context of emotional classification. Their study highlighted the wealth of emotional information in the latter segment, aligning closely with our observations. Similarly, Polo et al. 2021 focused on the final 100 seconds of signals from varying lengths (ranging between 120 to 190 seconds) for three-class categorical emotional classification. Our research highlights the importance of segment selection for optimal emotional classification, demonstrating that the Second-half phasic segment contains significant emotional information, yielding superior results compared to utilizing the Whole or First-half segments of the EDA signal.

The study observed that STFT consistently outperformed MFC spectrograms across emotional categories. This difference in performance could be attributed to various factors, including the inherent characteristics of STFT and MFC spectrograms and their ability to capture relevant features from EDA signals. STFT, with its capability to analyze frequency content over short intervals, was better at capturing temporal variations in the EDA signals, leading to more accurate emotional state classification. The window size for STFT was chosen and optimized through a trial-and-error method to balance frequency and time resolution. Specifically, the window size for the first-half and second-half segments is 50 samples long, while the whole phasic signal uses a window size of 100 samples. On the other hand, MFC spectrograms provide a multi-scale representation of signal features but may require additional techniques to effectively capture nuanced temporal dynamics in EDA data, resulting in comparatively lower performance. N. Kumar et al. 2021 used three techniques, namely MFC coefficients, Croma, and STFT, for emotional speech analysis and found that MFC coefficients give better results than the other two techniques. Ganapathy, Veeranki, and Swaminathan 2020 used STFT for valence and arousal classification; however, they never compared the results with other time-frequency representation methods. Another study utilized STFT, CWD, and smoothed pseudo-Wigner-Ville Distribution (SPWVD) distribution for signal representation in the time-frequency domain, reporting that SPWVD achieved the highest classification results for valence and arousal (Rao Veeranki, Ganapathy, and Swaminathan 2021). However, it's worth noting that their dataset, feature set, and classification methods differed from those used in the present study. To summarize, the Second-half phasic EDA signal emerges as the optimized segment for emotion classification, with the STFT time-frequency method being the preferred choice.

5.3 Effect of Segment Windows

This study optimized the windowing approach for the Second-half phasic signals. Initially, the optimization efforts focused on extracting Second half-phasic signals derived from preceding EDA data. These signals were then partitioned into intervals representing 33.33% and 20% of the signal duration with a 50% overlap, resulting in five and nine windows, respectively. Following this windowing approach, the signals were transformed into two-dimensional images using techniques such as GADF, GASF, MTF, and RP. A diverse set of features was extracted from these transformed images, including GLCM, GLRLM, FDTA, FOS, HM, and ZM. These extracted features were then subjected to significance testing utilizing the KW test to identify features that exhibited statistically significant differences ($p<0.05$). Subsequently, the significant features were applied to three ML models, SVM, RF, and XGB, to facilitate the classification of emotional states. This comprehensive approach aimed to refine the windowing technique specifically tailored to the latter phase of EDA signals, thereby enhancing the accuracy and robustness of emotional state classification.

The experimental results with the five-window approach yielded diverse performance outcomes across different methods and classifiers. Notably, employing the GADF method with the XGB classifier yielded an accuracy of 66.64% and an F1-score of 35.04%. Similarly, utilizing the GASF method alongside the SVM classifier resulted in an accuracy of 67.04% and an F1-score of 34.2%. Subsequently, employing the MTF method in conjunction with the SVM classifier led to a notable improvement in performance, achieving an accuracy of 81.33% and an F1-score of 62.68%. Furthermore, utilizing the RP method with the XGB classifier resulted in an accuracy of 81.58% and an F1-score of 62.44%.

Transitioning to the nine-window approach, variations were observed in performance metrics across the different methods and classifiers. Specifically, implementing the GADF method with the XGB classifier led to an accuracy of 67.04% and an F1-score of 37.48%. Meanwhile, employing the GASF method with the SVM classifier yielded an accuracy of 70.53% and an F1-score of 41%. Notably, utilizing the MTF method with the XGB classifier resulted in a substantial performance improvement, achieving an accuracy of 88.24% and an F1-score of 76.37%. Lastly, employing the RP method alongside the SVM classifier led to an accuracy of 84.75% and an F1-score of 68.45%. This windowing approach utilizing 20% signals, still this study obtained significant results in teams of classification performance compared with the Second-half phasic signasl.

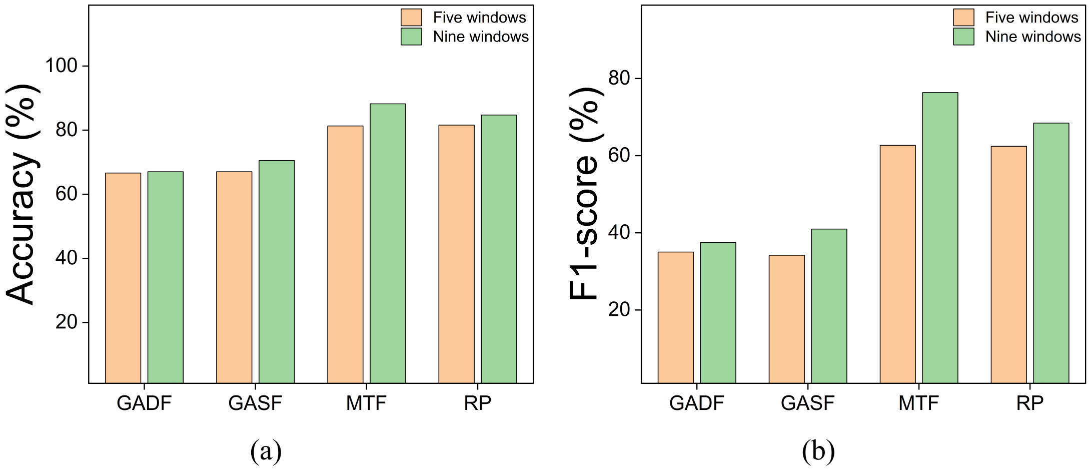


Figure 5.3: Bar diagram representation for the optimization of windowing approach using performance metrics: (a) Accuracy, and (b) F1-score.

The outcomes of this study show that MTF performed better, followed by RP, GASF, and GADF. The superior performance of MTF compared to RP, GASF, and GADF across the metrics in both the five-window and nine-window approaches can be attributed to the following factors. Firstly, MTF captures the transitional probabilities between signal values, providing a compact representation of temporal dynamics that captures both short-term and long-term dependencies in the data. This enables MTF to extract features that capture the underlying patterns associated with emotional states. In contrast, RP, GASF, and GADF methods may focus more on visualizing recurrent patterns or geometric transformations of the signal, potentially missing crucial transitional information. Secondly, MTF's representation in the frequency domain allows it to efficiently capture signal characteristics across different scales, facilitating the extraction of discriminative features that capture the nuanced variations in the EDA signals indicative of different emotional states. This adaptability and richness in feature representation contribute to the superior performance of MTF in emotional state classification tasks compared to RP, GASF, and GADF methods. In contrast, this study's results do not match the literature. A recent study that used time-encoded GASF, GADF, MTF, and RP on EEG signals for brain-computer interface applications reported that RP performs better than other methods (Paula et al. 2023). A study used GAF for emotion classification and reported that GASF outperformed earlier studies (J.-L. Qiu, X.-Y. Qiu, and K. Hu 2018). Another study utilized GAF, MTF, and RP to classify ECG heartbeat. The results were inconsistent between the methods and datasets used for the analysis. Furthermore, MTF failed to outperform GAF and RP in any of the analyses (Ahmad et al. 2021). A recent study utilized STFT and RP for valence and arousal classification using EDA, PPG, and ECG and reported that RP performed better than the STFT (Elalamy, Fanourakis, and Chanel 2021). The better performance of RP in the literature compared to other methods may depend on the modality, application, and sample size.

In this study, the nine-window approach outperformed the five-window approach, indicating the significance of the finer segmentation of EDA signals for improved classification accuracy. The windowing approach followed in this study utilizes 20% signals and still produced significant classification performance compared with the entire segment of Second-half phasic signals. By dividing the signals into smaller windows, the nine-window approach allows for a more granular analysis of the temporal dynamics inherent in the data. This finer windowing enables the classification model to capture more detailed variations in physiological responses associated with different emotional states. Consequently, the enhanced resolution provided by the nine-window approach facilitates a more accurate characterization of the underlying patterns in the EDA signals, ultimately leading to improved performance in emotional state classification tasks.

5.4 Effect of Process-Pipeline

Table 5.1 compares the results of this study with the state of the arts using the CASE dataset for emotional classification. This analysis yielded an impressive accuracy rate of 97.08% and F1-score of 94.22% using the pipeline involving Second-half phasic signal, STFT spectrograms, GLCM, and GLRLM features followed by classification using SVM, with a robust 10-fold cross-validation. The proposed pipeline in this study outperformed the accuracies reported in the literature. However, these studies might use different modalities and emotion paradigms for their analysis.

Dissanayake et al. 2022 explore multiple datasets, including CASE and K-EmoCon, employing diverse modalities such as EDA, BVP, and SKT. While their study encompasses various emotional classes, the classification approach lacks categorical emotional classification.

Polo et al. 2021 focuses on the CASE dataset and employs ECG, BVP, and EDA modalities, categorizing emotions into amusing, relaxing, and scary. However, their approach excludes the boring emotion, and they refrain from employing deep learning techniques.

Zhang et al. 2020 utilize a comprehensive set of datasets, including CASE and MERCA, alongside modalities such as EDA, ECG, HR, BVP, and SKT. Their approach involves correlation-based feature extraction followed by a broad learning system. However, their study lacks categorical emotional classification.

Further, Awais et al. 2020 explore various datasets, including CASE, focusing on modalities like EDA, BVP, SKT, and HR. They employ deep learning techniques but do not conduct categorical emotional classification. Hinduja, Kaur, and Canavan 2021 utilize CASE data and multiple modalities, employing long short-term memory-based deep learning techniques. However, their study lacks ML algorithms and cross-validation involving complex sensing modalities.

Zhang et al. 2022 explore CASE and MERCA datasets, emphasizing EDA as the modality. Although they achieve a commendable accuracy rate, their study exhibits variations in accuracies using different sensing modalities. Compared to earlier works, Sriram Kumar et al. 2023 achieved 78.96% accuracy using temporal morphological features, while P Sriram Kumar et al. 2023 improved upon this with 95.83% using time-frequency domain features. Another study by KumarP and Ronickom 2023 explored two-class categorical emotional classifications, achieving 84.17%. In contrast, SK and JF n.d. focused on spectrogram-based features, yielding a lower accuracy of 62.54% for four-class dimensional emotional classification. Our study maintains focus on four emotional states and similar feature sets yet surpasses previous accuracies, reaching 97.08% accuracy using SVM, RF, and XGB algorithms. In summary, the proposed pipeline in this thesis stands out for its meticulous optimization of EDA components, phasic EDA segments, and windowing approach, resulting in robust emotion recognition accuracy. While other studies explore diverse modalities and employ sophisticated techniques, they often need more categorical emotional classification or suffer from limitations such as excluding specific emotions. Our study's focus on a single modality enables a detailed investigation into the efficacy of EDA signals for emotion recognition, thereby contributing valuable insights to the field.

Table 5.1: Summary of various frameworks for the classification of emotions using the CASE dataset

Author	Dataset	Modality	Class	Feature extraction	Cross validation	Classification model	Performance (%)	Limitation
Zhang et al. 2020 (2020)	CASE, MERCA	RSP, EDA, ECG, EMG, SKT, BVP	A-2, V-2, A-3, V-3, VA-4	Correlation-based feature extraction	10-fold	ML, DL, CorrNet	77.01 (V-2), 80.11 (A-2), 61.83 (V-3), 62.03 (A-3), 69.36 (VA-4)	No categorical emotion classification was attempted.
Hinduja, Kaur, and Canavan 2021 (2021)	CASE	BVP, ECG, EDA, RSP, SKT, EMG	11 classes	-	10-fold	RF, FFNN	89.2	No 2, 3, 4 VA class implemented.
Polo et al. 2021	CASE	ECG, BVP, EDA	Amusing, Relaxing, Scary	Time domain, statistical	10-fold	LDA	72 (Amusing), 67 (Relaxation), 89 (Scary)	Not implemented 2, 3, 4 VA class.
Dissanayake et al. 2022 (2022)	CASE, CLAS, K-EmoCon, WESAD	EDA, BVP, SKT	A-2, V-2, A-3, V-3	-	LOSOCV	SigRep	76.30 (A-2), 74.08 (V-2), 65.07 (A-3), 64.83 (V-3)	No 4 VA class was attempted.
Zhang et al. 2022 (2022)	CASE, MERCA, CEAP-360VR	EDA, BVP HR, SKT	A-3 V-3	-	-	EDMIL	75.63 (V-3), 79.73 (A-3)	Did not work on 4 VA class.
Bhatti et al. n.d. (2022)	CASE WESAD SWELL-KW	ECG, EDA BVP, RESP SKT	Arousal Stress NAS	-	LOSO	AttX	66.72	Not implemented 2, 3, 4 VA class.
Y. Wu, Daoudi, and Amad 2023 (2023)	CASE WESAD K-EmoCon	EDA BVP Temp	V-2 A-2	-	-	Self supervised learning	77.49 (V-2), 73.67 (A-2)	Did not work on 3, 4 VA class.

Table 5.2: Summary of various frameworks for classifying emotions using the CASE dataset.

Author	Dataset	Modality	Class	Feature extraction	Cross validation	Classification model	Results (%)	Limitation
Sriram Kumar et al. 2023 (2023)	CASE	EDA	Amusing Boring Relaxing Scary	Temporal Morphological	6-fold	LR SVM RF	78.96	Binary Accuracy
P Sriram Kumar et al. 2023 (2023)	CASE	EDA	Amusing Boring Relaxing Scary	Time Frequency Time-Frequency	10-fold	LR, SVM RF, XGB MLP	95.83	Binary Accuracy
KumarP and Ronickom 2023 (2023)	CASE	EDA	Amusing Boring Relaxing Scary	Temporal features	10-fold	LR RF XGB	84.17	2, 3 VA Class not implemented.
SK and JF n.d.	CASE WESAD	EDA	HVHA HVLA LVHA LVLA	GLCM GLRLM FDTA ZM, HM FOS	10-fold	LR RF XGB	62.54 (VA-4)	2, 3 VA Class not implemented.
Our study	CASE	EDA	Amusing, Boring, Relaxing, Scary	GLCM GLRLM FDTA ZM, HM FOS	10-fold	SVM RF XGB	**97.08**	2, 3 VA Class not implemented.

Note: A-2: Arousal 2 class; V-2: Valence 2 class; A-3: Arousal 3 class; V-3: Valence 3 class; VA-4: Valence Arousal 4 class; Cornet: Correlation-based emotion recognition algorithm; LSTM: Long short-term memory; FFNN: Field forward neural network; LDA: Linear discriminant analysis; LOSOCV: Leave-one-subject-out cross-Validation; SigRep: self-supervised representation learning mechanism; EDM IL: Emotion recognition algorithm based on deep multiple instance learning; NAS: Neutral, Amusement, Stress; LOSO: Leave-one-subject-out; AttX: Attentive cross-modal connection; MLP: Multi-layer perceptron.