EEG Headband-Based Emotion Valence Prediction Approach: CNN Model and Evaluation

Vladimir Romaniuk, Alexey Kashevnik
SPC RAS
St.Petersburg, Russian Federation
romanyuk.v@iias.spb.su, alexey.kashevnik@iias.spb.su

***Abstract*—Emotion recognition based on electroencephalography (EEG) has gained significant attention due to its potential applications in human-computer interaction, affective computing, and mental health assessment. This study presents a convolutional neural network-based approach to emotion valence prediction model development using 4-channel headband EEG data as well as its evaluation based on computer vision emotion valence recognition. We train a model on the publicly available FACED dataset and tested it on a newly collected dataset recorded using a wearable BrainBit headband. The model's performance is evaluated using both standard train-validation-test splitting and a leave-one-subject-out cross-validation strategy. Additionally, the model is evaluated on computer vision-based emotion recognition system to assess the reliability and consistency of EEG-based emotion prediction. Experimental results demonstrate that the CNN model achieves competitive accuracy in predicting emotion valence from EEG signals, despite the challenges posed by limited channel availability and individual variability. The findings show the usability of compact EEG devices for real-time emotion recognition and their potential integration into adaptive user interfaces and mental health applications.**

I. Introduction

Emotion recognition is a rapidly evolving field with significant implications for human-computer interaction, affective computing, and mental health assessment. Accurately detecting and interpreting emotional states can enhance user experiences in various applications, including personalized content delivery, adaptive user interfaces, and intelligent virtual agents. Traditionally, emotion recognition has relied on external cues such as facial expressions, speech, and body language. These methods, however, are subject to limitations, including their dependence on visual or auditory stimuli, which can be affected by lighting conditions, background noise, or cultural differences in emotional expression.

EEG is a non-invasive method that directly measures electrical activity in the brain, providing a real-time and objective window into neural processes. EEG signals reflect the brain's cognitive and emotional responses to various stimuli, making it a powerful tool for understanding underlying emotional states. Compared to other physiological signals, such as heart rate variability or skin conductance, EEG has the advantage of being highly time-sensitive and capable of capturing rapid emotional shifts in response to external stimuli.

Recent advances in wearable EEG devices and signal processing techniques have paved the way for smaller, more comfortable headsets. However, this adapting for real life comes with reduced spatial resolution. Overcoming this trade-off is a key challenge for widespread adoption in real-life emotion recognition tasks.

The paper presents a continuation of our research in the topic of 4-chanels EEG data analysis for human psychophysiological state monitoring [1], [2], [3]. In the paper we present an approach to emotion valence prediction as well as its evaluation. We develop a machine learning model that utilizes EEG data to classify emotional valence — whether the emotion is positive, negative, or neutral. Valence is a critical dimension of emotion that reflects the subjective pleasantness or unpleasantness of an emotional experience. Understanding the valence of emotions provides valuable insights into the emotional tone of an individual's state and has applications in personalized emotional support systems, mental health diagnosis, and adaptive technologies that respond to user emotions. For evaluatuion of emotion valence predition based on EEG data we propose to compare the performance of developed model with traditional computer vision-based emotion recognition techniques, which analyze facial expressions and visual cues. By comparing the two approaches, we can assess the strengths and limitations of EEG-based emotion recognition and highlight its potential as a standalone solution for emotion detection.

One of the key advantages of using EEG for emotion recognition is its non-invasiveness and ease of use. EEG-based emotion recognition systems can be integrated into wearable devices with minimal user interference, making them suitable for continuous monitoring in real-world settings. Additionally, EEG allows for the detection of emotions that may not be externally visible, such as internal emotional states, thus providing a deeper level of insight into an individual's feelings. The compact 4-channel EEG setup used in this study is an important step toward developing more accessible and cost-effective emotion detection systems. Despite the limited number of channels, the model is designed to extract meaningful patterns from the EEG signals that can accurately predict the emotional valence of an individual.

However, the use of EEG in emotion recognition presents several challenges. EEG signals are inherently noisy and prone to artifacts caused by muscle movements, blinking, and environmental factors. Moreover, individual variability in brain patterns makes it difficult to develop a model that generalizes across different people. In this study, we employ a convolutional neural network (CNN), a deep learning approach known for its ability to extract hierarchical features from data, to address these challenges. CNNs are particularly effective for processing time-series data like EEG, as they can capture both spatial and temporal dependencies within the signals. By applying this model to a dataset of 4-channel EEG recordings, we aim to achieve high classification accuracy for emotion valence recognition while minimizing the impact of noise and individual differences.

The rest of the paper is divided as follows: Section II explores related works on the similar emotion valence prediction approaches. Section III describes the proposed approach, including data collection and reprocessing, model architecture, training procedures, evaluation metrics, and encountered challenges. Section IV concludes the study and discusses future directions for research.

II. Related Work

Emotion recognition involves several theoretical and technological approaches. One theoretical view treats emotions as discrete categories. For example, Plutchik proposes a “wheel of emotions” with eight basic emotional states that differ in intensity and polarity [4]. Another approach places emotions on continuous dimensions, usually valence (pleasant–unpleasant) and arousal (high–low). Lang’s framework fits this idea, showing that each emotional experience can be placed on these two main axes [5].

Emotion recognition has been widely explored in both computer vision and physiological signal processing, with EEG-based emotion classification gaining increasing attention due to its potential for detecting internal emotional states. Various datasets have been collected to facilitate research in this field, each employing different methodologies for emotion elicitation, recording modalities, and labeling strategies.

Various methods have been used to induce emotional states in participants while recording their brain activity. The choice of stimuli depends on the desired emotional responses. A common approach to eliciting emotions is through audiovisual stimuli, where participants are shown preselected video clips or images designed to trigger specific emotional states. Carefully curated film segments or standardized image sets can evoke positive, negative, or neutral emotions, while music can further amplify emotional responses. Another method involves self-referential or autobiographical recall, where individuals are asked to remember past experiences associated with strong emotions. This technique taps into personal memories and can evoke emotions that are more representative of real-world affective states.

Emotion labeling in EEG-based datasets typically relies on self-reported assessments using scales such as the Self-Assessment Manikin, which measures valence, arousal, and dominance. Participants provide their subjective ratings immediately after stimulus exposure, ensuring that the reported emotions reflect their immediate experience. In contrast, video-based emotion datasets often employ facial expression recognition frameworks combined with expert annotations or crowd-sourced labeling.

In earlier research on EEG-based emotion recognition, many studies used traditional machine learning methods. These methods relied on hand-selected features, such as power spectral density (PSD) in the alpha, beta, gamma, delta, and theta bands, and various statistical measures derived from time-domain signals. Algorithms like support vector machines, k-nearest neighbors, discriminant analyses, or ensemble methods were then used for classification. Rahman et al. combined principal component analysis and a statistical test to refine features before training a support vector machine. Their system effectively separated neutral, positive, and negative emotional states [6]. While these methods can work well with clear data and carefully chosen features, they may not always capture complex patterns in EEG signals and often need a lot of domain knowledge to find the most useful features.

Deep learning methods have become popular in recent years because they can learn features directly from raw or slightly preprocessed EEG data. Convolutional neural networks are especially common for detecting local patterns across space (channels) and time. CNNs can learn robust representations of the data without extensive human-lead feature engineering. Yang et al. showed that a multi-column CNN could improve performance in valence and arousal classification on public EEG datasets [7].

Recurrent neural networks (RNNs) are another form of deep learning suited to sequential data like EEG. RNNs keep track of a hidden state that changes over time, which is useful for capturing the changing nature of emotional responses. Advanced types of RNNs, such as long short-term memory (LSTM) networks and gated recurrent units (GRUs), can reduce problems with vanishing or exploding gradients. Algarni et al. achieved high accuracy (96–99%) with a bidirectional LSTM model for classifying valence, arousal, and liking, which shows the importance of modeling temporal changes in EEG signals [8].

Recent work by Du et al. [9] further expands the application of deep learning in emotion recognition by focusing on music-induced emotions. In their study, a hybrid 1D-CNN-BiLSTM model was proposed to classify valence and arousal from EEG signals evoked by Chinese ancient-style music. Their approach, validated on both public datasets (DEAP and DREAMER) and a self-acquired EEG dataset from Chinese college students, demonstrated high classification accuracies – especially in negative valence detection.

Another trend in emotion recognition research focuses on making EEG recording more user-friendly by using wearable or low-channel devices. Wu et al. reached 75–76% accuracy for valence detection with only two frontal channels [10]. Moontaha et al. used wearable devices with four or eight channels and reported F1 scores of 82–87% for binary valence and arousal classification [11]. These results suggest that smaller EEG headsets can still support practical emotion recognition, especially if signal quality is managed well. However, they offer less spatial detail, so more research is needed to see how they perform in varied real-world conditions.

III. An Approach

The section includes data collection and reprocessing, model architecture, training procedures, evaluation metrics, and encountered challenges.

A. General Description

This study focuses on emotion recognition by utilizing EEG data and video-based emotion recognition models. A public dataset (FACED) containing 30-channel 250 Hz EEG data from 122 participants is used to train a convolutional neural network for emotion valence prediction. Since the recorded dataset consists of only 4-channel EEG data, a preprocessing step is applied to extract relevant features and align the data structure with the trained model.

The trained CNN model is then used to predict emotion valence for the recorded dataset based on EEG signals. In parallel, a video-based emotion recognition model from another study is used to analyze emotions from the recorded video data. The predictions from both approaches, EEG-based and video-based, are synchronized and compared to assess their consistency and reliability in emotion classification. Performance is evaluated to determine the effectiveness of EEG-based emotion recognition in comparison to the video-based model. General overview of methodolgy is shown in Figure 1.

B. Our Dataset Collection

For this study, two primary devices were employed. The first device was the BrainBit Headband, shown in Figure 2, a wearable EEG device equipped with four dry electrodes that recorded raw EEG signals at a sampling rate of 250 Hz, providing data in volts. The second device was a standard web camera, which recorded video of the participant's face in 720p resolution. Data collection setup is shown in Figure 3.

The BrainBit Headband follows the international 10-20 electrode placement system and is equipped with four active electrodes positioned at O1, O2, T3, and T4. The reference electrode for the BrainBit device is positioned on the forehead.

The study involved 7 participants who were selected without specific criteria regarding age or gender.

Data collection was performed during computer-based sessions in which participants were seated in front of a laptop. Each session lasted approximately three hours and involved a variety of tasks designed to elicit a range of eye movements. These tasks included reading passages displayed on the screen, completing standardized Landolt C tests, and playing simple computer games.

C. Public Dataset

The FACED dataset [12] comprises recordings from 123 participants, each subjected to a series of emotion-eliciting stimuli. EEG signals were captured using a 30-channel system at a 250 Hz sampling rate, and is stored in volts.

Participants were exposed to 28 distinct video clips, each selected to evoke one of nine specific emotional states: amusement, inspiration, joy, tenderness, anger, fear, disgust, sadness, and neutrality. After viewing the stimuli, participants provided self-assessments of their emotional experiences. These subjective ratings were then used to label the EEG data in terms of valence (positive, negative, neutral) and discrete emotional categories. Record overview is shown in Figure 4.

Since the EEG recordings in this dataset were captured using a high-density 30-channel system, a subset of four channels was selected to match the electrode placement of the BrainBit headband used in our study. The selected channels were O1 and O2, located over the occipital lobe, and C3 and C4, positioned over the central region of the scalp. These channels were chosen based on their proximity to the BrainBit electrodes, ensuring that the signal characteristics remained as similar as possible between the two datasets. The occipital electrodes (O1 and O2) primarily capture visual cortex activity, which is relevant given that the emotion-eliciting stimuli were presented in video format. The central electrodes (C3 and C4) are associated with motor and sensorimotor processing, which can be linked to emotional responses that involve physiological and muscular changes.

By extracting data from only these four channels, the FACED dataset was effectively adapted to the hardware constraints of the BrainBit headband, allowing the trained model to generalize better to real-world applications where compact EEG devices with a limited number of electrodes are commonly used. This channel selection also reduced computational complexity while preserving critical information relevant to emotion recognition.

D. Data Preprocessing

To ensure the quality of the EEG data, preprocessing steps were implemented to remove noise. A bandpass filter was applied, allowing frequencies between 0.5 and 40 Hz to pass through, filtering out muscle and higher-frequency noise. This filtering ensured that the EEG data retained the most relevant signal components for subsequent analysis.

The continuous EEG signals were segmented into epochs with no overlapping. For the FACED dataset, each epoch ranged inside the onsets of each video clip with 1 second window. Similarly, for our recorded dataset, we segmented signals into 1-second windows.

Consistency of data between different dataset was achieved by applying same normalization and filtering steps. Also we ensured that both datasets have equal sampling rate.

E. Model Architecture and Training

The classification tasks were performed using TSCeption [13], a deep learning model specifically designed for EEG-based time-series classification. TSCeption utilizes both temporal and spatial convolutional layers to extract multi-scale features from EEG signals, effectively capturing both short-term and long-term dependencies in the data. The model architecture consists of multiple convolutional layers that process EEG signals at different temporal resolutions, followed by feature fusion layers that integrate spatial information across channels. A final fully connected layer maps the extracted

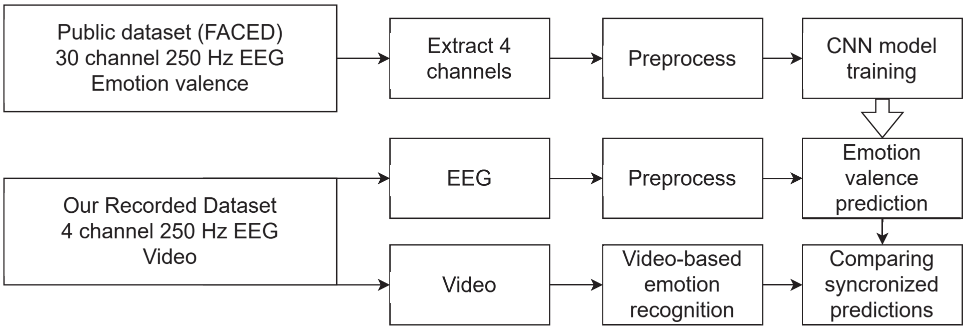

Fig. 1. Proposed Emotion Valence Classification Methodology

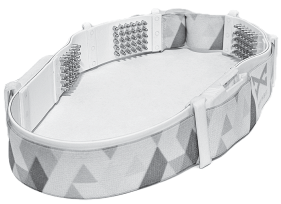

Fig. 2. The BrainBit Headband, Wearable EEG Device with 4 Dry Electrodes

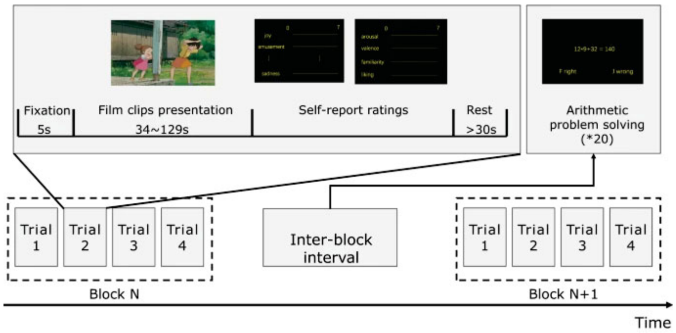

Fig. 4. FACED dataset record overview

features to the output classes. Model architecture is shown in Fig. 5.

We used the Adam optimizer with a maximum of 100 training epochs. PyTorch was used for implementing TS-Ception. The number of temporal and spatial convolutional kernels was 15 each, with hidden layer channels set to 512. A dropout rate of 0.5 was used to mitigate overfitting, while the learning rate was set to 0.001 with a weight decay of $1e^{-4}$ to introduce regularization. The loss function for training was cross-entropy, given the multi-class nature of the emotion valence classification task.

Because the dataset valence labels were unbalanced, a weighted data loader was used to adjust the sampling frequency of each class. This approach helps the model see more examples from underrepresented classes.

Two training strategies were employed to evaluate the model's generalizability. In the first approach, the dataset was randomly split into training, validation, and test sets with an 80/10/10 ratio, ensuring that data from all participants contributed to model training. Early stopping was applied based on validation performance to prevent overfitting and optimize model convergence. In the second approach, a leave-one-subject-out cross-validation (LOSO-CV) scheme was used, where the model was trained on data from 122 participants and tested on the remaining participant. This process was repeated for each participant, allowing the evaluation of subject-independent performance and robustness across individuals.

F. Evaluation Metrics

Assessment of the performance of the TSCeption model was done by accuracy and F1 score as the primary evaluation metrics. These metrics were calculated for the emotion valence classification task, comparing the model's predictions against ground truth labels.

Performance evaluation was conducted on FACED dataset under two training scenarios. In the first case, where all participants' data were mixed and split into training, validation, and test sets, accuracy and F1 scores were averaged across multiple runs to ensure stability in results. The metrics are shown in Table I. In the second case, leave-one-subject-out cross-validation (LOSO-CV) was performed, where the model was trained on 122 participants and tested on one, with results aggregated across all iterations to evaluate subject-independent performance. The metrics are shown in Fig. 6.

G. Video-Based Emotion Recognition Comparison

In this study, a video-based emotion recognition system EMO-AffectNetModel [14] was used as a benchmark for eval-

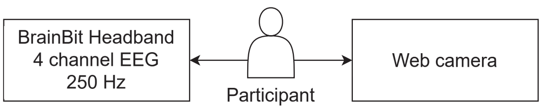

Fig. 3. Data collection setup

TABLE I. PERFORMANCE METRICS FOR TRAIN/VALIDATION/TEST SPLIT

Accuracy	F1
75.31	72.68

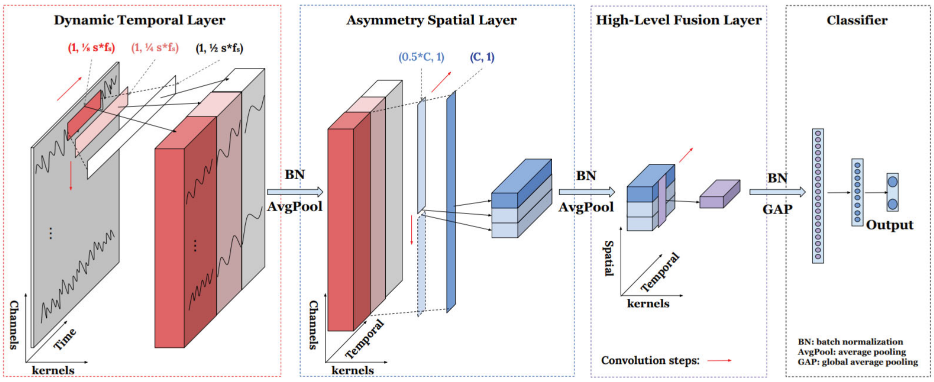

Fig. 5. TSCeption model architecture

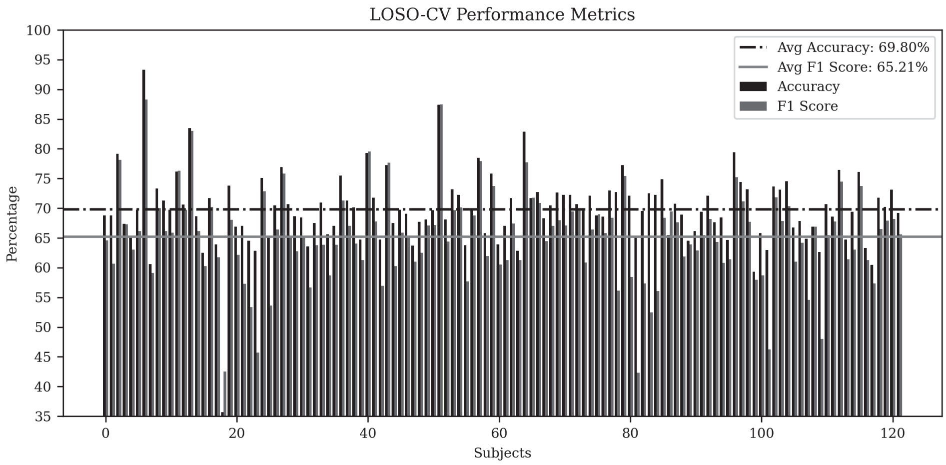

Fig. 6. Performance Metrics for LOSO-CV

uating the performance of the EEG-based model. This model analyzes facial expressions to classify emotional states. The system processes each video frame independently, extracting facial features and mapping them to an emotional category using a deep learning model trained on a large-scale facial expression dataset.

The video data of our dataset was recorded simultaneously with EEG signals, capturing the participants' facial expressions in real-time. Since both video and EEG recordings were obtained using the same device, no additional time synchronization adjustments were necessary. However, as the recordings were manually started and did not begin at exactly the same moment, a common timeline was established based on the overlapping duration of both data sources. Only time segments where both EEG and video data were available were used for further analysis to ensure consistency in emotion classification.

The recorded data was divided into non-overlapping one-second windows to allow a direct comparison between EEG-based and video-based predictions. For EEG-based recognition, a single emotion valence prediction was generated per one-second window. In contrast, video-based recognition provided frame-level predictions, meaning multiple classifications were obtained within each second. To create a comparable output, the most frequently occurring valence prediction within a one-second interval was used as the final classification for that window. This aggregation method ensured that both models produced a single, comparable prediction per time window.

To assess the consistency and reliability of EEG-based emotion recognition, its predictions were compared with those obtained from the video-based model. The analysis was conducted on a per-window basis, where each one-second segment of data contained both EEG and aggregated video-based emotion predictions. This direct comparison allowed for an evaluation of agreement between the two methods.

Performance evaluation was based on accuracy and F1-score, computed for each participant. Results are shown in Table II.

IV. Challenges and Solutions

One of the main challenges encountered in this study was the need to address the variability in EEG patterns across different individuals. Each participant's brain activity is unique, which leads to differences in the EEG signals even when exposed to the same emotional stimuli. Moreover, our dataset and public dataset have in-dataset and between-dataset differences in participants' age, gender, culture. Our dataset consist of 7 European males aged 20-41 years. Public dataset consist of 123 Chinese males and females (75 females) aged 17-38 years. This variability posed a challenge for training a generalized model that could effectively classify emotion valence across all participants. To mitigate this, a leave-one-subject-out cross-validation strategy was employed, allowing to train the model on data from multiple participants and tested on individual subjects to assess its ability to generalize across different neural patterns. Transfer learning from big public dataset ensures model generalization and stable model performance in case of expanding our dataset.

Another challenge in comparing EEG-based and computer vision-based emotion recognition methods was the imbalance in the dataset. Most participants predominantly showed neutral emotions, which could lead to inflated accuracy due to class imbalance rather than the model's actual ability to distinguish between different emotional states.

Additionally, computer vision-based emotion recognition does not achieve 100% accuracy and can not be used as 100% ground truth. Factors such as lighting conditions, facial occlusions, and individual differences in expressiveness can impact its performance. This further complicates direct comparisons between EEG-based and video-based recognition models, as inaccuracies in the video-based model could influence the evaluation of EEG-based classification.

Furthermore, our EEG-based approach have signal quality limitation. The preprocessing stage includes signal filtering using bandpass filter, but in situations were muscle activity or other artifacts highly presented in recording, this filtering may not be enough and model will fail due to highly noisy data.

V. Acknoledgements

This work was supported by the Russian State Research FFZF-2025-0003.

VI. Conclusion

The study demonstrates the potential of using CNN-based models for emotion valence recognition from 4-channel EEG data. Despite challenges related to individual variability and noise in the EEG signals, the proposed model showed promising performance in classifying emotions based on the valence dimension. Employment of a compact EEG setup offers a non-invasive and efficient approach to emotion detection, which can be beneficial in real-time applications such as adaptive user interfaces and mental health monitoring systems.

Future work will focus on several directions. First, the model's robustness and generalizability can be enhanced by incorporating advanced data augmentation and domain adaptation techniques to better address class imbalance and inter-subject variability. Additionally, integrating complementary physiological signals—such, as heart rate variability and galvanic skin response, or integration computer-vision approach in the same time with EEG-based approach, —could provide a more comprehensive understanding of emotional states. Finally, extensive testing in diverse and ambulatory settings will be essential to validate the system's performance under varied conditions.

TABLE II. Performance Metrics for EEG-based and video-based emotion recognition comparison

Subject	Accuracy	F1
1	92.46	95.45
2	91.98	95.75
3	63.57	67.95
4	88.50	92.30
5	86.75	90.50
6	89.30	93.40
7	87.80	91.60

References

[1] V. Romaniuk and A. Kashevnik, "Eye movement assessment methodology based on wearable eeg headband data analysis," in *2024 36th Conference of Open Innovations Association (FRUCT)*, 2024, pp. 675–680.

[2] V. Romaniuk and A. Kashevnik, "Intelligent eye gaze localization method based on eeg analysis using wearable headband," *Informatics and Automation*, vol. 23, pp. 521–541, 2024.

[3] A. Kashevnik, E. Glekler, A. Stankevich, M. Stradina, and E. Kucheruk, "Intelligent service for hybrid analysis of continuous mental processes based on eeg and video data," 11 2023, pp. 66–72.

[4] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, pp. 344–350, 2001. [Online]. Available: https://doi.org/10.1511/2001.28.344

[5] P. J. Lang, "The emotion probe: Studies of motivation and attention," *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.

[6] M. A. Rahman, M. F. Hossain, M. Hossain, and R. Ahmmed, "Employing pca and t-statistical approach for feature extraction and classification of emotion from multichannel eeg signal," *Egyptian Informatics Journal*, vol. 21, 2020.

[7] H. Yang, J. Han, and K. Min, "A multi-column cnn model for emotion recognition from eeg signals," *Sensors (Switzerland)*, vol. 19, 2019.

[8] M. Algarni, F. Saeed, T. Al-Hadhrami, F. Ghabban, and M. Al-Sarem, "Deep learning-based approach for emotion recognition using electroencephalography (eeg) signals using bi-directional long short-term memory (bi-lstm)," *Sensors*, vol. 22, 2022.

[9] R. Du, S. Zhu, H. Ni, T. Mao, J. Li, and R. Wei, "Valence-arousal classification of emotion evoked by chinese ancient-style music using 1d-cnn-bilstm model on eeg signals for college students," *Multimedia Tools and Applications*, vol. 82, 2023.

[10] S. Wu, X. Xu, L. Shu, and B. Hu, "Estimation of valence of emotion using two frontal eeg channels," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 1127–1130.

[11] S. Moontaha, F. E. F. Schumann, and B. Arnrich, "Online learning for wearable eeg-based emotion classification," *Sensors*, vol. 23, 2023.

[12] J. Chen, X. Wang, C. Huang *et al.*, "A large finer-grained affective computing eeg dataset," *Scientific Data*, vol. 10, p. 740, 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-02650-w

[13] Y. Ding, N. Robinson, S. Zhang, Q. Zeng, and C. Guan, "Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, 2023.

[14] E. Ryumina, D. Dresvyanskiy, and A. Karpov, "In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study," *Neurocomputing*, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222012656