2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10386020

Transformer-Based Domain Adaptation for Multi-Modal Emotion Recognition in Response to Game Animation Videos

Jing-Yi Liu, Jia-Wen Liu, Wei-Long Zheng, Bao-Liang Lu*
Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

***Abstract*—Emotion recognition researches necessitate the strategic selection of stimuli to evoke targeted emotions for robust physiological analyses. This study pioneers the use of Genshin Impact game animation videos to induce positive and neutral emotional states. Notably, it introduces a Transformer-based feature extractor, enhancing Domain-Adversarial Neural Networks (DANN) to advance domain adaptation capabilities. Leveraging the inherent advantages of the Transformer architecture, including parallel processing and handling intricate time-series Electroencephalogram (EEG) and eye movement data, this innovation is reinforced by a discriminator with gradient reversal layers, harmonizing source and target domain distributions. Empirical results demonstrate the effectiveness of the Transformer-based DANN model in cross-subject multimodal emotion recognition, achieving a remarkable prediction accuracy of 83.38% across 59 subjects.**

***Index Terms*—emotion recognition, transformer, domain adaptation, multi-modal deep learning**

I. Introduction

In the realm of emotion recognition research, inducing precise emotional states in subjects is pivotal for analyzing physiological changes. To ensure meticulous control over experimental conditions, researchers frequently employ emotionally evocative stimulus materials, designed with precision to elicit targeted emotions. Traditional EEG datasets for emotion analysis predominantly resort to conventional sources such as movies or music as stimuli, as exemplified by well-known datasets like the SEED dataset [1] and the DEAP dataset [2], which often overlook the utilization of game-based stimuli. A singular publicly accessible dataset centered around game-oriented stimuli, denoted as the GAMEEMO dataset [3], features a collection of four computer games designed to elicit a range of emotions, including boredom, calmness, horror, and enjoyment.

Divergent from the passive observation of stimuli, computer games offer subjects the dual roles of observers and active participants, thereby facilitating heightened immersion and deeper emotional engagement. Notably, the framework of Game-Based Assessment (GBA) entails the evaluation of individuals' psychological traits based on game-related data, fostering simulated interactive scenarios while mitigating test anxiety, consequently eliciting more authentic behavioral responses [4]. This characteristic makes games a promising avenue for inducing and assessing emotions. However, recording EEG signals during gameplay is challenging due to subject movements and unknown factors. Gameplay-related videos maintain interaction realism and signal stability, corroborating the efficacy of animation in eliciting emotions. In view of these considerations, we use Genshin Impact game animation videos to induce positive and neutral emotions, selected based on online questionnaire ratings.

Among the diverse modalities of emotion recognition, EEG has demonstrated remarkable reliability owing to its comprehensive capture of emotional information and its resistance to deliberate manipulation [5]. While the precision and objectivity inherent in EEG render it a compelling choice, its inherent heterogeneity across individuals poses challenges when training a universally stable classifier [6]. In situations such as clinical diagnosis of psychological disorders, gathering individual-specific data and conducting model training become impractical due to time constraints. To address this predicament, transfer learning, particularly through the lens of Domain Adaptation (DA) [7], has emerged as an effective approach, utilizing labeled data from the source domain to optimize the classification in the target domain by learning a mapping between the two domains.

In this paper, we present an innovative Transformer-based feature extractor that extends the foundational principles of DANN [8]. The architecture encompasses a domain classifier intricately connected to a feature extractor via a gradient reversal layer, effectively ensuring domain invariance. By incorporating the inherent attention mechanism within Transformer encoders, we further augment the extraction of pertinent information from diverse input modalities. Compared to other deep networks utilized in DA methods, our model demonstrates enhanced parallelization and computation efficiency, proving adept at processing concatenated EEG and eye movement data.

II. Experiment

A. Stimuli

Our emotion elicitation comprises 32 video clips inducing positive and neutral emotions (16 per category), each lasting 1-2 minutes. These stimuli are drawn from official promotion and fan-made videos of the game Genshin Impact, signaling

* Corresponding author

979-8-3503-3748-8/23/$31.00 ©2023 IEEE

a transition from cinematic to interactive paradigms. Fig. 1 illustrates screenshots from two chosen videos, each aimed at evoking positive or neutral emotions. Positive videos feature dynamic game scenes and character actions set to lively music, while neutral videos depict the serene open-world landscape with mellow music.

(a) Positive video (b) Neutral video

Fig. 1. Examples of the selected game animation clips in the experiment.

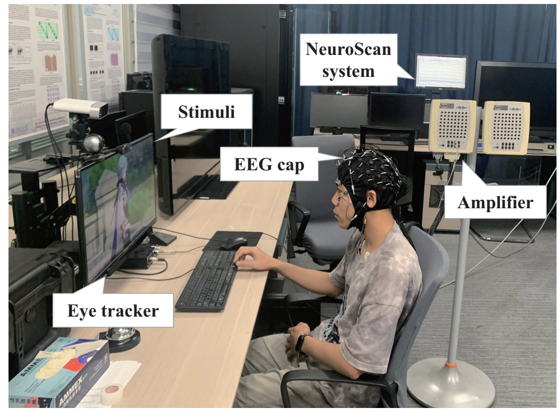


Fig. 2. An experimental scene in the emotion experiments.

Sixteen positive videos were chosen from a collection of 45 videos through a questionnaire distributed to 43 college students (mean age 21.23). The selected videos had ratings of 3.16 or higher on a 0-5 scale, signifying effective emotional elicitation. In the formal experiment, subjects assessed the emotional arousal of positive videos on a 0-1 scale, while valence scores for neutral videos ranged from 0 to 1 (0 for negativity, 1 for positivity, and 0.5 for emotional neutrality). The mean scores were 0.5966 for positive videos signifying consistent positive emotional inducement, and 0.5443 for neutral videos confirming the selected videos' efficacy in evoking intended emotional states.

B. Subjects

We recruited 59 native Chinese students (30 females, 29 males, mean age 21.44) from Shanghai Jiao Tong University with normal hearing, normal or corrected-to-normal vision, and no history of psychiatric illness. The selection of participants prioritized those with low N (emotional stability) and high E (extroversion) scores on the Eysenck Personality Questionnaire (EPQ) [9]. This strategic choice was made to facilitate the induction of intended emotions during the experiment [10], thereby enhancing the classification model's accuracy. Ultimately, subjects were balanced in terms of gender and gaming experience (15 female players, 14 female non-players, 16 male players, and 14 male non-players).

C. Protocol

The experiment maintained controlled conditions, providing a quiet environment with suitable temperature and lighting while minimizing external sound. EEG signals were recorded using the NeuroScan system at 1000 Hz, while the Tobii desktop eye-tracker captured eye movements. Fig. 2 shows the experimental scene.

In each session, 32 video clips were presented, comprising 16 positive and 16 neutral clips, with a total duration of approximately 46 minutes. After viewing, participants provided video ratings, followed by a 10-second intermission for emotional recalibration. The experimental protocol is visually depicted in Fig. 3. To prevent emotional fatigue due to sustained exposure to a particular type of material and to forestall abrupt emotional transitions, positive and neutral videos alternated every 4 segments, balancing emotional engagement while avoiding undue fluctuations. The video sequence was randomized while maintaining consistent labeling to ensure methodological rigor.

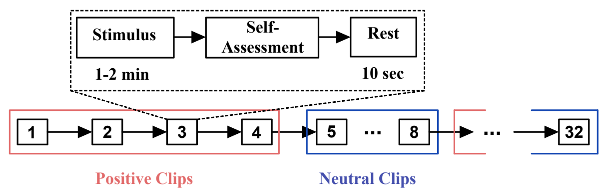


Fig. 3. The protocol used in our emotion experiment.

D. Data Processing

We employed a proven method, consistent with previous work [1], [11], to extract the 310-dimensional differential entropy (DE) features from EEG signals. Linear Dynamic System (LDS) smoothing was applied to remove non-emotion-related features [12]. For eye movement signals, principal component analysis was conducted to eliminate ambient light effects on pupil diameter changes, retaining emotion-related data. This yielded a 33-dimensional set of DE features, detailed in Table I, which was then smoothed using the LDS algorithm.

III. Method

In order to estimate our dataset's performance in emotion recognition, we implement in-subject classification and cross-subject classification with multiple models. For cross-subject classification, we use domain adaptation (DA) [7] methods to overcome subject transfer problems.

A. Model Overview

We integrate the Transformer architecture with Domain-Adversarial Neural Networks (DANN) [8]. Our model, illustrated in Fig. 4, comprises a feature extractor, classifier, and

TABLE I
SUMMARY OF EXTRACTED EYE MOVEMENT FEATURES

Eye movement parameters	Extracted features
Pupil diameter (X, Y)	Mean, STD, DE in four bands (0–0.2 Hz,0.2–0.4 Hz, 0.4–0.6Hz,0.6–1Hz)
Disperson (X, Y)	Mean, STD
Fixation duration (ms)	Mean, STD
Blink duration (ms)	Mean, STD
Saccade	Mean and STD of saccade duration (ms) and saccade amplitude (°)
Event statistics	Blink frequency, fixation frequency, fixation duration maximum, fixation dispersion total, fixation dispersion maximum, saccade frequency, saccade duration average, saccade amplitude average, saccade latency average.

discriminator. The feature extractor extracts domain-specific features, while the classifier performs emotion classification. During training, the discriminator distinguishes between source and target features, driving domain adversarial representation.

B. Transformer-Based Feature Extractor

At the core of our methodology, the feature extractor capitalizes on the Transformer architecture [13]. The self-attention mechanism intrinsic to Transformers enhances computational efficiency through parallel processing of inputs. To incorporate the temporal dimension of EEG and eye movement feature sequences, we employ position embeddings. These embeddings augment the information encapsulated within the time series. The concatenated features then traverse an encoder composed of N identical layers, each comprising a multi-head self-attention module and a forward network.

The multi-head attention mechanism facilitates the computation of scaled dot-product attention for each head. Concretely, it takes queries (Q), keys (K) of dimension d_k, and values (V) of dimension d_v as inputs. Outputs consist of weighted summations of values, the weights of which are computed through query-key interactions. This result is divided by the scale factor $\sqrt{d_k}$ and subjected to a $softmax$ function for weight computation. The attention function can be written as:

$$\text{Attention}(Q, K, V) = softmax\left(QK^T/\sqrt{d_k}\right)V. \quad (1)$$

Let O denote the output of (1). For a self-attention sublayer with h heads, queries, keys, and values are linearly projected h times with learnable projections. The multi-head attention is calculated as:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(O_{h_1}, \ldots, O_{h_h})W^O, \quad (2)$$

where $O_{h_i} = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ and the projections are parameter matrices W_i^Q, W_i^K, W_i^V and W^O.

The parallel operation of multi-head attention results in outputs that are subsequently combined and conveyed to a fully connected feed-forward sublayer. This sublayer encompasses two linear transformations interspersed with a Gaussian Error Linear Unit (GELU) activation which avoids the vanishing gradients problem [14].

C. Discriminator

Mirroring the conventional DANN model's principles, our adversarial learning network incorporates gradient reversal layers that invert gradients during backpropagation, engendering identity transformation during forward propagation. This design fosters the convergence of the distributions of x_s and x_t, diminishing their distinguishability.

In the adversarial-training phase, for every batch x_s originating in the source data, the discriminator selects an analogous batch of unlabeled target data (x_t) from the target domain. Following processing by the feature extractor (G_f), both sets of data are channeled into the domain classifier (G_d), responsible for ascertaining whether the data originates from the source or target domain. Considering the source data x_s, the adversarial loss is defined as:

$$\begin{aligned}\mathcal{L}_{adv}(G_d(G_f(x_s)), d_s) = d_s \log \frac{1}{G_d(G_f(x_s))} \\ +(1 - d_s) \log \frac{1}{1 - G_d(G_f(x_s))},\end{aligned} \quad (3)$$

where d_s symbolizes the domain label for x_s and takes the value 1 due to its source domain origin. Analogously, a parallel equation holds for the target data x_t, albeit with $d_t = 0$. The transfer loss is calculated as the mean of the adversarial losses for both source and target data:

$$\mathcal{L}_{transfer} = \frac{1}{2}(\mathcal{L}_{adv}(x_s) + \mathcal{L}_{adv}(x_t)). \quad (4)$$

D. Classifier

In the training phase, the classifier (G_y) operates on source data x_s, paired with their true labels y_s, and processed by the feature extractor (G_f). The classifier, structured as a linear network, adopts the cross-entropy loss as the classification loss metric:

$$\mathcal{L}_{clf}(G_y(G_f(x_s)), y_s) = -y_s \log G_y(G_f(x_s))_{y_s}. \quad (5)$$

The comprehensive loss function encompasses both classification and transfer losses, and is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{clf} + \lambda \mathcal{L}_{transfer}, \quad (6)$$

where λ denotes the weight attributed to the transfer loss, set to 1 for this study.

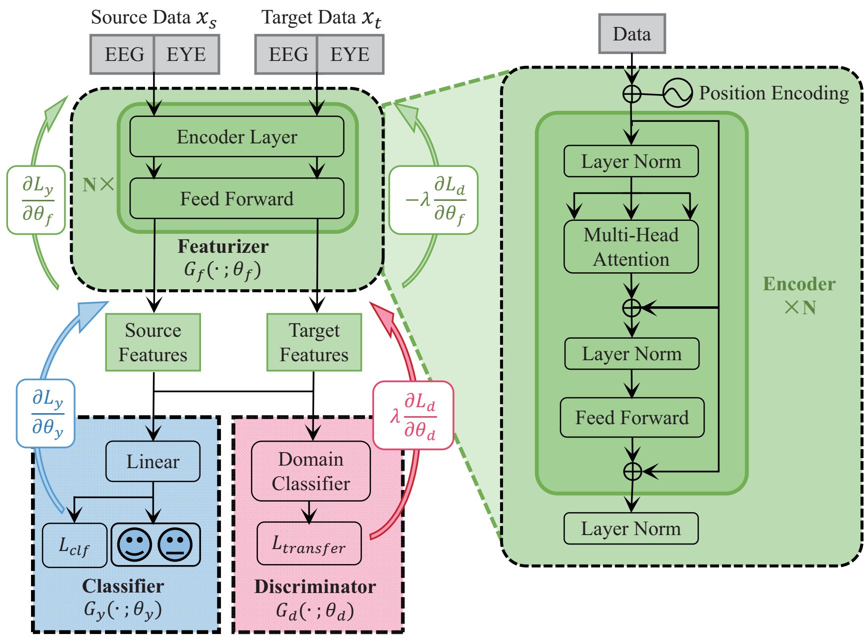


Fig. 4. Transformer-based DANN model framework.

IV. Experiment Results

This section presents the experimental outcomes of a binary recognition task involving positive and neutral emotions on our dataset. Four-fold cross-validation is conducted in Sections IV-A and IV-B. Moreover, We evaluate the cross-subject classification accuracy of Transformer-based DANN and three traditional DA methods in Section IV-C. Finally, Section IV-D operates a comparative examination between the Female/Male group and the Player/Non-player group.

A. Results of EEG-Based Emotion Recognition

We utilized a linear kernel Support Vector Machine (SVM) [15] for binary classification, determining the optimal regularization parameter C through a grid search with candidate values [0.1, 1, 10], ultimately selecting C=1 based on the results. Mean accuracies and standard deviations of SVM with DE features from various EEG frequency bands are summarized in Table II. The highest accuracy and relatively low standard deviation were achieved using DE features from all five frequency bands, with the beta and gamma bands outperforming lower frequency bands. This aligns with previous research [1], confirming the significance of EEG features from beta and gamma bands in emotion recognition.

TABLE II
Mean Accuracies and Standard Deviations of SVM with the DE Features from Different EEG Frequency Bands

Frequency band	Delta	Theta	Alpha	Beta	Gamma	Total
Acc (%)	66.08	63.83	65.79	68.23	70.68	**74.55**
Std (%)	**13.60**	15.55	14.47	14.35	15.96	13.82

Additionally, we present the distribution of normalized DE features across different brain regions in five frequency bands for positive and neutral emotions in Fig. 5. The topography reveals enhanced activation in temporal regions for positive emotion in beta and gamma bands, while the alpha band triggers more parietal activation for neutral emotions. This observation is consistent with prior findings [5], underscoring distinct neural patterns associated with different emotions.

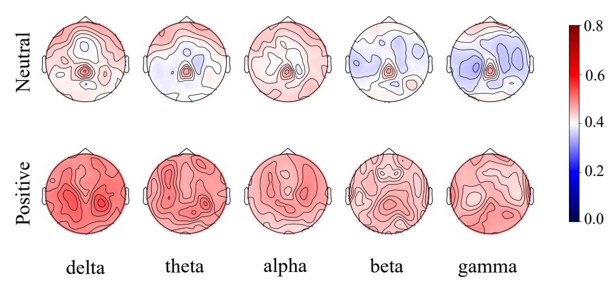


Fig. 5. Average topographical maps of brain regions in five frequency bands of positive and neutral emotions.

B. Results of Multi-Modal Affective Models

For multi-modal signals, we compare the classification accuracies of four established models:

1) MLP: Comprising two ReLU layers and a linear classifier, this model directly concatenates EEG and eye movement features. Each subject's model is trained for 70 epochs.
2) The experimental setup mirrors that of Section IV-A. The model is trained with single-modal eye movement

features and, in addition to the single-modal EEG result, concatenated EEG and eye movement features.

3) Transformer: The encoder layers are the same as the feature extractor in Section III-B. A fully connected feed-forward network serves as a classifier. The model randomly selects hyperparameters 50 times: the number of heads in a self-attention sublayer from [1, 2, 4, 8], and the number of encoder layers from [1, 2, 4]. The output dimension of the feed-forward layer is set to 512.
4) Deep Generalized Canonical Correlation Analysis with an Attention Mechanism (DGCCA-AM): As proposed by Lan et al. [16], this model extends Canonical Correlation Analysis (CCA) [17] with an attention-based multi-modal fusion, learning adaptive fusion weights for different modalities.

TABLE III
MEAN ACCURACY RATES AND STANDARD DEVIATIONS OF FOUR MULTI-MODAL AFFECTIVE MODELS ON SEED-GAME DATASET

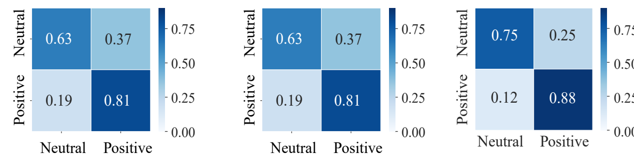
Model	Acc (%)	Std (%)
SVM-EEG	74.55	13.82
SVM-eye	80.53	11.04
SVM-multi	82.75	10.68
MLP-multi	74.04	10.61
Transformer-multi [13]	94.63	**4.84**
DGCCA-AM-multi [16]	**94.95**	5.23

Table III demonstrates DGCCA-AM as the top performer, surpassing MLP and SVM by 20.91% and 12.2% in accuracy, while reducing the standard deviation by approximately half, emphasizing the efficiency of attention-based modality fusion in emotion recognition. Notably, Transformer attains a remarkable 94.63% accuracy with the lowest standard deviation, likely attributable to the Transformer encoder layers' structure. The multi-head attention effectively models relationships between different-positioned features in the input sequence, enabling superior combination of multi-modal features compared to SVM and MLP.

Regarding modality, SVM exhibits enhanced performance with multi-modal features relative to single-modal ones. Previous research [18] corroborates that EEG and eye movements offer complementary information, resulting in improved overall performance. Fig. 6 presents the confusion matrices for different modalities, highlighting their superior recognition of positive emotions over neutral ones. In both positive and neutral emotion recognition, multi-modality outperforms single modalities, underscoring the performance-enhancing potential of combining two modalities.

C. Results of Cross-Subject Classification

We employ DA methods for cross-subject emotion recognition, which involves selecting one subject's data from the cohort of 59 subjects as the target domain, while utilizing the remaining subjects' data as the source domain. In our pre-experiment, we used DGCCA-AM as the feature extractor for DANN [8]. Despite its high accuracy in Section IV-B, this



(a) EEG (b) Eye (c) Multi modality

Fig. 6. Confusion matrices of single-modal and multi-modal SVM. (a), (b) and (c) represent respectively the results using EEG features, eye movement features, and concatenated multi-modal features.

choice resulted in subpar performance and protracted training times, preventing further discussion.

We compare our Transformer-based DANN model with other DA methods, including the original DANN, DSAN [19], DAAN [20], and Baseline. All DA models consist of a feature extractor (MLP except for our method), a classifier, and a discriminator. The Baseline method employs either MLP or Transformer as the feature extractor but lacks a discriminator. Regarding the two other DA methods, DSAN simplifies the network by eliminating adversarial components and aligns relevant subdomain distributions across different domains using local maximum mean discrepancy (LMMD). DAAN, on the other hand, incorporates a local subdomain discriminator for assessing global and local domain distributions and assigns weights to the global and local discriminator losses using a dynamic adversarial factor.

TABLE IV
MEAN ACCURACY RATES AND STANDARD DEVIATIONS OF FIVE CROSS-SUBJECT MULTI-MODAL METHODS

Model	Acc (%)	Std (%)
MLP-Based baseline	81.22	8.94
Transformer-Based baseline	81.43	8.63
DSAN [19]	81.92	8.21
DAAN [20]	82.72	8.77
MLP-Based DANN [8]	82.88	8.05
Transformer-Based DANN	**83.38**	**6.86**

Table IV demonstrates the enhanced performance of all DA methods compared to the Baseline, affirming the efficacy of the adversarial-learning module. Our Transformer-based DANN achieves the highest accuracy at 83.38% with a standard deviation of 6.86%. Remarkably, the Baseline with Transformer outperforms the Baseline with MLP, highlighting the stable advantages of Transformer networks.

D. Group Comparison and Discussion

We compare the multi-modal patterns of different groups of people based on two classification criteria: gender and Genshin Impact gaming experience. Table V summarizes in-subject classification comparison for gender and Table VI for gaming experience. In general, female subjects and Genshin Impact players consistently exhibit higher average accuracy across all classification models. Among them, there is significant difference between the player/non-player groups' data distributions (p=0.04). Particularly, DGCCA-AM and Transformer achieve accuracies above 0.93 for the non-player group and the male group as well, proving that game animation clips as stimuli have a steady effect in emotion recognition across diverse audiences.

TABLE V
EMOTION RECOGNITION ACCURACIES USING MULTI-MODAL METHODS IN DIFFERENT GENDER GROUPS

	Model	MLP	SVM	Transformer	DGCCA-AM
Female	Acc (%)	74.27	86.03	95.17	**95.46**
	Std (%)	11.95	9.93	5.01	4.92
Male	Acc (%)	73.80	78.98	94.11	94.42
	Std (%)	9.02	10.22	**4.62**	5.48

*p-value=0.42

TABLE VI
EMOTION RECOGNITION ACCURACIES USING MULTI-MODAL METHODS IN DIFFERENT GAMING EXPERIENCE GROUPS

	Model	MLP	SVM	Transformer	DGCCA-AM
Player	Acc (%)	76.68	84.63	95.46	**95.97**
	Std (%)	9.89	10.57	**4.00**	4.05
Non-Player	Acc (%)	71.12	80.28	93.75	93.81
	Std (%)	10.61	10.32	5.50	6.08

*p-value=0.04

V. CONCLUSION

In this paper, we have introduced a pioneering approach, employing game animation materials to establish a novel emotion induction paradigm that effectively elicits both positive and neutral emotional states. Empirical validation has confirmed the robustness and generalizability of this paradigm. Furthermore, our proposed Transformer-based DANN model has demonstrated exceptional performance, achieving a state-of-the-art cross-subject classification accuracy of 83.38% among a cohort of 59 subjects.

In our experimental design, we confined our subject pool to emotionally stable extroverted university students to optimize classification performance. Future research will extend the framework to include diverse individuals of varying personality types and age groups, extending the application scenarios.

VI. ACKNOWLEDGEMENTS

This work was supported in part by grants from National Natural Science Foundation of China (No. 61976135), STI 2030-Major Projects+2022ZD0208500, Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX), Shanghai Pujiang Program (Grant No. 22PJ1408600), Medical-Engineering Interdisciplinary Research Foundation of Shanghai Jiao Tong University "Jiao Tong Star" Program (YG2023ZD25), and GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine.

REFERENCES

[1] 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.

[2] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[3] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on EEG signals and various computer games–gameemo," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, 2020.

[4] M. E. DeRosier, J. M. Thomas *et al.*, "Hall of heroes: A digital game for social skills training with young adolescents," *International Journal of Computer Games Technology*, vol. 2019, 2019.

[5] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2017.

[6] W.-L. Zheng and B.-L. Lu, "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, 2016, pp. 2732–2738.

[7] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, pp. 151–175, 2010.

[8] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[9] S. B. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Personality and Individual Differences*, vol. 6, no. 1, pp. 21–29, 1985.

[10] X. Wu, W.-L. Zheng, Z. Li, and B.-L. Lu, "Investigating EEG-based functional connectivity patterns for multimodal emotion recognition," *Journal of Neural Engineering*, vol. 19, no. 1, p. 016012, 2022.

[11] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 81–84, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:6431331

[12] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 6587–6590, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:23552183

[13] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *NIPS*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:13756489

[14] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv: Learning*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:125617073

[15] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. [Online]. Available: https://api.semanticscholar.org/CorpusID:52874011

[16] Y.-T. Lan, W. Liu, and B.-L. Lu, "Multimodal emotion recognition using deep generalized canonical correlation analysis with an attention mechanism," *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221652214

[17] G. Andrew, R. Arora, J. A. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *International Conference on Machine Learning*, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:2495132

[18] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *International Joint Conference on Artificial Intelligence*, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:4473734

[19] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, "Deep subdomain adaptation network for image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 1713–1722, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:218505698

[20] C. Yu, J. Wang, Y. Chen, and M. Huang, "Transfer learning with dynamic adversarial adaptation network," *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 778–786, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202660604