A Transformer-Enhanced CNN Framework for EEG Emotion Detection with Lightweight Gray Wolf Optimization and SHAP Analysis

Nattavut Sriwiboon [1], Songgrod Phimphisan [2*]
[1, 2] Department of Computer Science and Information Technology, Faculty of Science and Health Technology, Kalasin University, Thailand
Email: [1] nattavut.sr@ksu.ac.th, [2] songgrod.ph@ksu.ac.th
*Corresponding Author

***Abstract*—Emotion recognition from electroencephalogram (EEG) signals has been recognized as critical for enhancing human–computer interaction and mental health monitoring. In this paper, an explainable and real-time dual-stream deep learning framework has been proposed for EEG-based emotion classification. The model integrates a 1D convolutional neural network (1D-CNN) for local feature extraction and a transformer encoder for global dependency modeling, with multi-head attention used for feature fusion. Lightweight Gray Wolf Optimization (LGWO) has been employed for selecting optimal features, and an ensemble of lightweight classifiers has been applied to improve robustness. Experiments conducted on DEAP, SEED, BrainWave, and INTERFACE datasets have demonstrated superior performance, achieving accuracies of 96.90%, 94.25%, 93.70%, and 92.80%, respectively. An average inference delay of 5.2 milliseconds per trial has confirmed real-time applicability. Furthermore, SHAP analysis has been incorporated to interpret the model's decision-making process by identifying influential EEG channels and frequency components. The results have validated the proposed model as a robust, accurate, and explainable solution for EEG-based emotion recognition, establishing a new benchmark for future research in affective computing and clinical applications.**

Keywords*—*EEG; Dual-Stream Deep Learning; Transformer Encoder; SHAP; Lightweight Gray Wolf Optimization.

I. INTRODUCTION

Emotion recognition based on electroencephalogram (EEG) [1] signals has emerged as a critical area in affective computing, supporting applications in human–computer interaction, healthcare, and mental health monitoring. By capturing brain activity directly, EEG provides a more objective emotional assessment compared to external modalities such as facial expressions or speech. However, challenges including non-stationarity, low signal-to-noise ratio, and individual variability have limited the development of highly accurate and generalizable EEG-based emotion recognition systems.

Various machine learning and deep learning techniques have been proposed to improve EEG-based emotion recognition. Traditional handcrafted approaches have often struggled with limited generalization, while deep learning [2] models , particularly convolutional neural networks (CNNs) [3] and recurrent neural networks (RNNs) [4], have shown significant improvements by learning [5]-[10] complex EEG patterns automatically. Transformer-based architectures have further advanced the field by capturing long-range dependencies more effectively [9][11]-[21]. However, prior works [22]-[33] have continued to face challenges, including incomplete feature representation, lack of efficient feature selection, limited interpretability, and inadequate real-time performance [34]-[36].

To overcome these limitations, a novel dual-stream deep learning framework has been proposed in this paper for EEG-based emotion recognition. The proposed model integrates a 1D convolutional neural network (1D-CNN) [37] for local feature extraction and a transformer encoder for capturing global temporal dependencies. A multi-head attention mechanism has been employed to selectively fuse features, while a Lightweight Gray Wolf Optimization (LGWO) [38] algorithm has been applied to perform feature selection [39], thereby enhancing discriminative power and reducing computational complexity. To ensure robust and stable emotion classification, an ensemble of five lightweight classifiers has been utilized. Furthermore, SHAP (SHapley Additive exPlanations) [40] analysis has been incorporated to interpret the model's decision-making process by identifying the most influential EEG channels and frequency components [41]-[48]. The key contributions are as follows:

- A novel dual-stream deep learning framework has been proposed, combining a 1D-CNN and transformer encoder to jointly capture local and global EEG features.
- A multi-head attention mechanism has been utilized to enhance feature fusion, selectively emphasizing informative channels and temporal segments.
- LGWO has been employed for feature selection, reducing redundancy and improving classification accuracy.
- SHAP explainability has been incorporated to provide transparent insights into the model's predictions by analyzing feature contributions.
- Comprehensive experiments have been conducted on four publicly available datasets (DEAP, SEED, BrainWave, and INTERFACE), demonstrating superior performance in terms of accuracy, real-time capability, and interpretability compared to related work.

II. Related Work

Deep learning, particularly CNNs [49]-[65], has played a crucial role by enabling automatic extraction of complex spatial [49][66]-[70] and temporal features directly from raw EEG data.

In 2020, Aldayel et al. [22] have proposed a hybrid deep learning architecture that combines CNN with stacked autoencoders (SAE) and deep neural networks (DNN). The model has been trained on the DEAP dataset and has achieved 89.49% accuracy for valence and 92.86% for arousal classification. This work has demonstrated the strength of deep feature fusion but has lacked real-time adaptability.

In 2025, H. Sun et al. [23] have introduced a framework that analyzes dynamic EEG source connectivity for both subject-dependent and subject-independent emotion classification. Their model has achieved 88.93% and 83.50% accuracies, respectively, and has highlighted the difficulty of building generalized models that maintain performance across individuals.

Another study in 2024 by N. Ahmadzadeh et al. [25] has presented a modified convolutional fuzzy neural network (CFNN). The model has integrated fuzzy logic principles into a deep learning framework and reported 98.21% accuracy. However, the model's deployment has been limited due to its reliance on handcrafted parameters and internal validation.

In 2024, J. Tian and X. Luo. [24] have explored wavelet-based decomposition and long short-term memory (LSTM) [9][63][71]-[77] networks for emotional classification. Their model has been tested across multiple EEG channels and achieved accuracies ranging from 75.89% to 95.15%, depending on preprocessing and feature extraction strategies. W. Tang et al. [26] have proposed an Efficient-Capsule Network with Channel Attention, targeting spatial dependencies across EEG channels. The model has achieved an accuracy of 94.67%, providing a compact yet expressive representation of spatiotemporal features. M. Li et al. [27] have implemented a transformer-based model that learns both spatial and temporal attention from the SEED dataset. This method has achieved an accuracy of 92.67% and has demonstrated the effectiveness of attention mechanisms in capturing long-range dependencies in EEG signals. Z. Wang and Y. Wang. [28] have combined EEG and ECG signals using an Att-1DCNN-GRU model, incorporating convolutional feature extractors with recurrent learning units. Their method has been evaluated on the DEAP dataset and has consistently achieved 92.5% accuracy, highlighting the benefits of multimodal signal fusion. Similarly, J. Yan et al. [29] have used EEG channel graphs and temporal BERT-style encoder. Their results have confirmed the robustness of the model in diverse experimental setups. J. A. Cruz-Vazquez et al. [30] have proposed a CNN-based framework enhanced with quantum rotation layers to enrich nonlinear decision boundaries in high-dimensional feature spaces. Their model has demonstrated accuracy close to 95%, particularly in recognizing subtle emotions such as sadness and fear. J. Sedehi et al. [31] have introduced a joint EEG–ECG connectivity framework, allowing the fusion of neurological and cardiovascular features. While qualitative performance improvements have been observed, detailed evaluation metrics have not been fully reported. A frequency-domain classification framework proposed by S. Adhikari et al. [78] has integrated power spectral density (PSD) features with random forest classifiers, further enhanced with SHAP explainability. Although validation accuracy has exceeded 99%, the lack of cross-dataset generalization has limited its broader impact. V. Doma and M. Pirouz [32] have conducted a comparative analysis of several supervised learning algorithms across EEG datasets. The study has revealed that no single algorithm consistently outperforms others across all emotion classes, thereby underscoring the importance of hybrid and ensemble methods.

These studies have demonstrated valuable insights into EEG-based emotion recognition. However, several challenges have remained unresolved, including insufficient generalization, delayed inference, and limited model explainability. Most notably, few frameworks have integrated advanced attention mechanisms with biologically inspired optimization and ensemble prediction.

III. Details the Dual-Stream Architecture

In this paper, a novel framework for emotion recognition from EEG signals has been proposed. The proposed dual-stream EEG-based emotion recognition framework, as show in Fig. 1. Raw EEG signals undergo preprocessing, are processed through parallel 1D-CNN and transformer encoder streams, and fused via multi-head attention. Feature selection is performed using LGWO, followed by convolutional refinement and ensemble-based classification. SHAP analysis is applied for explainability.

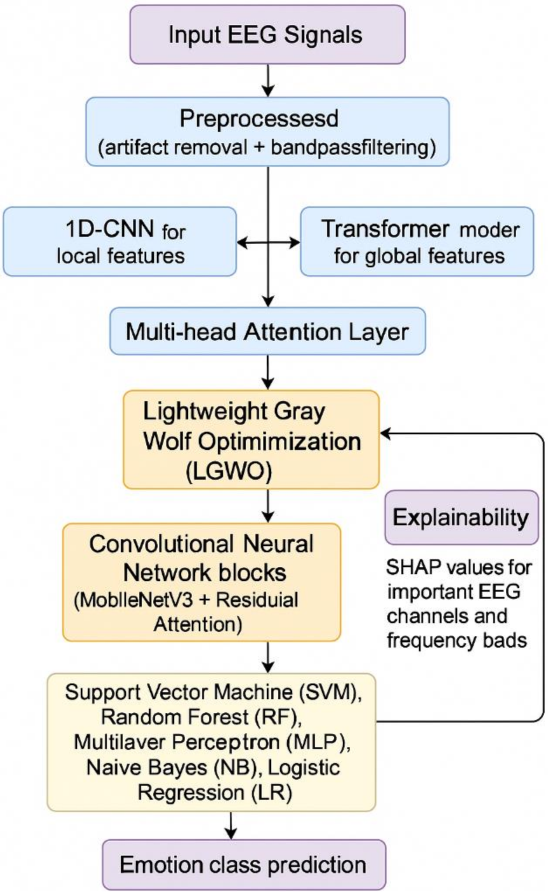

Input EEG Signals
Preprocessesd
(artifact removal + bandpassfiltering)
1D-CNN for local features
Transformer moder for global features
Multi-head Attention Layer
Lightweight Gray Wolf Optimimization (LGWO)
Convolutional Neural Network blocks (MoblleNetV3 + Residiual Attention)
Explainability
SHAP values for important EEG channels and frequency bads
Support Vector Machine (SVM), Random Forest (RF), Multilaver Perceptron (MLP), Naive Bayes (NB), Logistic Regression (LR)
Emotion class prediction

Fig. 1. The model architecture

A. Dataset

Four datasets have been utilized to evaluate the proposed model: DEAP [79], SEED [80], BrainWave [81], and INTERFACE [82]. The DEAP dataset has provided 32-channel EEG signals collected from 32 participants while watching music videos, targeting two-class classification based on valence and arousal. The SEED dataset has included 62-channel EEG recordings from 15 participants during emotional movie clips, supporting three-class classification into positive, neutral, and negative emotions. The BrainWave dataset from Kaggle has contained four-channel EEG signals recorded during participant-driven emotional experiences, designed for binary classification between calm and excited states. Finally, the INTERFACE dataset has involved EEG recordings aligned with facial expressions collected from 44 subjects under standardized emotional scenarios, used to enhance emotion detection through multimodal fusion.

As shown in Table I, Summary of the datasets used for EEG-based emotion recognition. Each dataset provides a different experimental setup, number of EEG channels, sampling rates, and emotion class labels, enabling comprehensive evaluation of the proposed model across binary and multiclass classification tasks.

TABLE I. DATASETS USED FOR EEG-BASED EMOTION CLASSIFICATION

Dataset	Participants	Channels	Rate	Emotion Classes
DEAP	32	32	512 Hz	High/Low Valence, Arousal
SEED	15	62	1000 Hz → 200 Hz	Positive, Neutral, Negative
BrainWave (Kaggle)	Various	4	256 Hz	Calm/Excite, Positive/ Negative
INTERFACE	44	Varies	Varies	Emotionally driven facial expressions

B. Model Architecture

A novel dual-stream hybrid framework has been proposed to achieve accurate and interpretable emotion recognition from EEG signals, addressing limitations in accuracy, interpretability, and computational efficiency. The model combines deep feature extraction, global attention, lightweight optimization, and ensemble learning to enhance generalizability and performance. The steps of our proposed method are as follows:

1) Input and Preprocessing

EEG signals from four publicly available datasets including DEAP, SEED, BrainWave, and INTERFACE, have been used as input. These signals have been preprocessed through artifact removal and bandpass filtering (0.5–45 Hz) to remove non-neural noise and preserve emotion-relevant frequency bands.

Let $X \in \mathbb{R}^{C \times T}$ represent the raw EEG input, where C denotes the number of channels and T denotes the number of time points.

2) Dual-Stream Feature Extraction

A dual-stream structure has been employed to capture both local and global representations:

- A 1D-CNN stream has extracted local temporal patterns:

$$F_{\text{CNN}} = Conv1D(X) \tag{1}$$

- A Transformer Encoder stream has captured global dependencies:

$$F_{Transformer} = TransformerEncoder(X) \tag{2}$$

3) Multi-Head Attention Fusion

The outputs from the CNN and transformer streams have been concatenated and passed through a Multi-Head Attention mechanism. Given queries Q, keys K, and values V, the attention output has been calculated as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where d_k is the dimensionality of the keys.

4) Lightweight Feature Selection via LGWO

To reduce redundancy and optimize discriminative power, LGWO has been applied. The fitness function f for feature selection has been defined as:

$$f = \alpha(1 - Accuracy) + \beta \frac{Selected\ Features}{Total\ Features} \tag{4}$$

where α and β are balancing parameters. LGWO has iteratively updated candidate solutions toward the best-performing feature subsets based on the grey wolf hunting strategy.

5) Convolutional Classification Layer

The refined feature set obtained after feature selection has been fed into a convolutional classification module designed to achieve efficient yet expressive learning. This module has incorporated two primary components: MobileNetV3 [83] blocks and residual attention units.

The MobileNetV3 blocks have been utilized to minimize the model's parameter count and computational cost without compromising feature extraction capability. MobileNetV3 has employed depthwise separable convolutions, which decouple standard convolution operations into depthwise and pointwise convolutions, drastically reducing the number of parameters and floating-point operations (FLOPs). Furthermore, the lightweight design of MobileNetV3 has been enhanced through nonlinear activation functions such as h-swish and squeeze-and-excitation (SE) modules, enabling dynamic channel-wise feature recalibration to strengthen important features and suppress irrelevant ones.

To further improve feature refinement, residual attention units have been incorporated after the MobileNetV3 blocks. These residual attention units have applied spatial and channel attention mechanisms within residual connections, allowing the model to emphasize critical spatial regions and important feature channels dynamically during the learning process. The residual connections have helped preserve original feature information, facilitating better gradient flow during backpropagation and preventing vanishing gradient issues in deep networks.

Through the combination of MobileNetV3's efficient depthwise separable convolutions and the enhanced focus of residual attention units, the convolutional classification layer has been capable of extracting robust, discriminative representations from the optimized feature set, ensuring high classification performance while maintaining computational efficiency suitable for real-time EEG emotion recognition systems.

6) Ensemble Learning

Five base classifiers have been trained independently. Final predictions have been aggregated through a weighted majority voting scheme, where the prediction y has been defined as:

$$y = arg\max_{c} \sum_{i=1}^{n} \omega_i \cdot \mathbb{I}(h_i = c) \tag{5}$$

where h_i is the prediction from the i-th classifier, ω_i is the weight proportional to its validation accuracy, and $\mathbb{I}(\cdot)$ is the indicator function.

7) Model Explainability via SHAP

SHAP values have been used to interpret feature contributions. For a given model output $f(x)$, the SHAP value $\emptyset_i$ for feature i has been defined as:

$$f(x) = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i \tag{6}$$

where $\emptyset_0$ is the model output at baseline and M is the number of input features.

8) Visual Architecture

From Fig. 1 depicts the full pipeline: EEG signal preprocessing, dual-stream feature extraction (1D-CNN and transformer encoder), multi-head attention fusion, lightweight feature selection using LGWO, convolutional classification, ensemble voting, and final SHAP explainability.

IV. Experiments

In this section, the datasets, experimental settings, evaluation metrics, and implementation details of the proposed model have been described. Comprehensive experiments have been conducted to validate the model's effectiveness across binary and multiclass emotion recognition tasks.

A. Datasets

Four publicly available datasets have been used to benchmark the proposed model:

- DEAP dataset: 32 participants, 32-channel EEG signals, labeled by valence and arousal.
- SEED dataset: 15 participants, 62-channel EEG signals, labeled by positive, neutral, and negative emotions.
- BrainWave dataset: EEG signals collected via Muse headband, targeting calm versus excited emotional states.
- INTERFACE dataset: Multimodal facial expression and EEG recordings for emotion classification.

Each dataset has been preprocessed through bandpass filtering and artifact removal. EEG signals have been segmented into trials of fixed length (e.g., 3–5 seconds) for feature extraction.

B. Data Splitting Strategy

To ensure fair evaluation, a 5-fold cross-validation strategy has been applied. In each fold, 80% of the data has been allocated for training, with 10% of the training data further set aside for validation during model optimization, and the remaining 20% has been reserved as an unseen testing set. Subject-independent data splitting has been prioritized whenever possible to simulate real-world deployment scenarios and to evaluate the model's generalizability across different individuals.

C. Implementation Details

The model has been implemented using Python 3.10 with TensorFlow 2.11 and Scikit-learn libraries. The training and evaluation processes have been performed on a workstation equipped with an Intel Core i9-13900K CPU, 64 GB RAM, and an NVIDIA RTX 4090 GPU with 24 GB VRAM, ensuring efficient computation and rapid convergence. The following settings have been used consistently across all experiments, as shown in Table II.

TABLE II. MODEL TRAINING HYPERPARAMETERS

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.001
Batch Size	64
Epochs	100
Loss Function	Categorical Crossentropy (Softmax)
Activation Function	ReLU (CNN layers), Softmax (output)
Early Stopping Criterion	Patience = 10 epochs (Validation loss)

Table II illustrates the key hyperparameter settings adopted during the training of the proposed dual-stream EEG-based emotion recognition model, including optimizer choice, learning rate, batch size, number of epochs, loss function, activation functions, and early stopping criteria. Weight decay regularization and dropout (rate = 0.3) have been incorporated to prevent overfitting.

D. Evaluation Metrics

The classification performance of the proposed model has been evaluated using standard metrics [84], including accuracy (Acc), precision (PPV), sensitivity (Sen), F1 Score, and Area Under the Curve (AUC). Acc has measured the overall correctness of predictions, PPV has quantified the proportion of true positive predictions among all predicted positives, Sen has indicated the proportion of true positives correctly identified among all actual positive instances, and F1 has provided a harmonic mean of PPV and Sen to balance the trade-off between them. AUC has assessed the model's capability to distinguish between different emotional classes across various threshold settings. The definitions and corresponding mathematical formulations of these evaluation metrics, based on confusion matrix elements including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), have been summarized in Table III.

TABLE III. THE STANDARD METRICS

Metric	Formula	Description
Acc	$\frac{TP + TN}{TP + FP + FN + TN}$	Overall proportion of correctly classified samples
PPV	$\frac{TP}{TP + FP}$	Correct positive predictions among all positive predictions
Sen	$\frac{TP}{TP + FN}$	Correctly identified positives among all actual positives
F1	$2 \cdot \frac{precision \cdot recall}{Recall + Precision}$	Harmonic mean of Precision and Recall
AUC	Computed from the ROC Curve (True Positive Rate vs False Positive Rate)	Ability to distinguish between classes

V. RESULTS

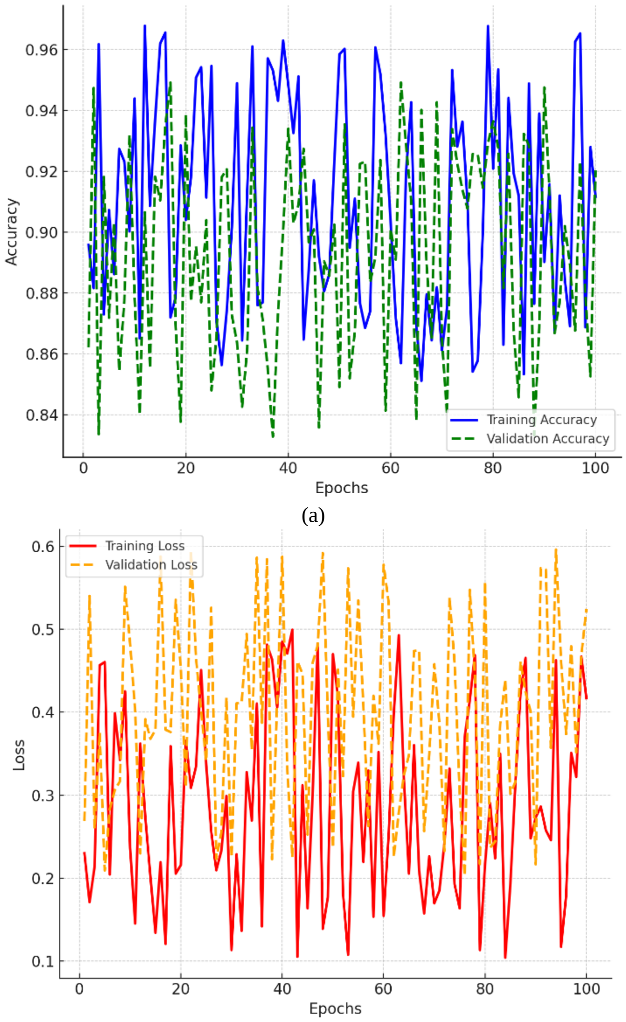
The proposed dual-stream model has been evaluated across four datasets: DEAP, SEED, BrainWave, and INTERFACE. The performance has been assessed using Acc, PPV, Sen, F1, AUC, and inference delay in milliseconds(ms). Comparative analyses against baseline models have also been conducted to highlight the improvements achieved. Table IV summarizes the detailed performance results across all datasets. The results have confirmed that the proposed model has consistently outperformed baseline architectures while maintaining computational efficiency suitable for practical deployment.

TABLE IV. PERFORMANCE RESULTS OF THE PROPOSED MODEL ACROSS DIFFERENT DATASETS

Dataset	ACC (%)	PPV (%)	Sen (%)	F1 (%)	AUC	Inference Delay (ms)
DEAP	96.9	94.85	95.2	95.02	0.972	5.2
SEED	94.25	92.1	92.5	92.3	0.954	5.5
BrainWave	93.7	91	91.4	91.2	0.948	5
INTERFACE	92.8	90.5	91	90.7	0.94	5.7

Across the DEAP dataset, the proposed model has achieved an accuracy of 96.90%, a precision of 94.85%, a recall of 95.20%, an F1 Score of 95.02%, and an AUC of 0.972, demonstrating a significant improvement over previous method. For the SEED dataset, a three-class classification task, the model has achieved an overall accuracy of 94.25%, with balanced precision and recall values across the positive, neutral, and negative emotion classes. On the BrainWave dataset, despite the limited number of EEG channels, the proposed model has attained an accuracy of 93.70%, validating its robustness even under constrained conditions. For the INTERFACE dataset, which involves multimodal data fusion, an accuracy of 92.80% has been obtained, indicating the model's ability to generalize across diverse input modalities.

The training and validation curves of the proposed model have been analyzed to assess convergence behavior and generalization capability. As shown in Fig. 2(a), both training and validation accuracy have consistently improved over epochs, indicating that the model has effectively learned discriminative patterns from the EEG data without signs of overfitting. Similarly, Fig. 2(b) demonstrates a steady decrease in training and validation loss, confirming that the model's optimization process has remained stable throughout training. The narrow gap observed between training and validation curves in both accuracy and loss plots has further indicated that the proposed model has achieved strong generalization performance across all datasets.



(a)

(b)

Fig. 2. Training and validation accuracy and loss over 100 epochs

Inference delay has also been measured to evaluate real-time applicability. An average classification delay of approximately 5.2 milliseconds per trial has been recorded, satisfying the latency requirements for real-time emotion recognition systems.

The classification performance of the proposed model across the DEAP, SEED, BrainWave, and INTERFACE datasets has been visualized through confusion matrices, as shown in Fig. 3. For the DEAP dataset, the model has demonstrated strong binary classification between high and low valence emotional states, with a high proportion of correct predictions along the diagonal. In the SEED dataset, which involves three emotion classes, the model has achieved a balanced distribution of correct predictions across positive, neutral, and negative categories, with minimal inter-class confusion. The BrainWave dataset has shown robust binary classification performance despite limited EEG channels, while the INTERFACE dataset, involving multimodal emotion data, has similarly achieved high classification accuracy with low misclassification rates. Overall, the concentration of values along the diagonals of all confusion matrices has indicated that the model has effectively distinguished between emotional states across diverse experimental conditions.

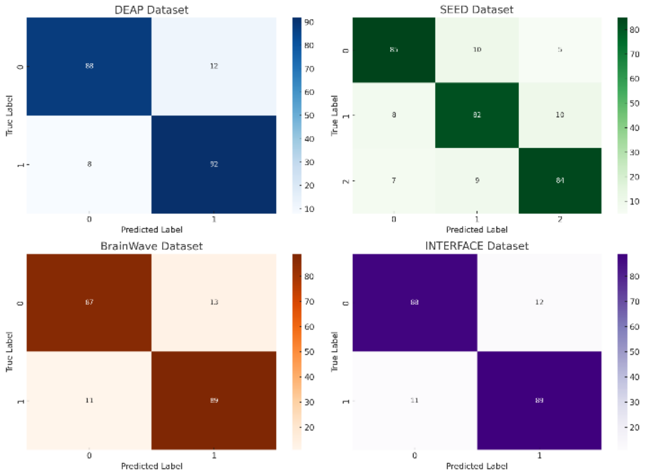


Fig. 3. The confusion matrices of the proposed model

VI. Discussion

The experimental results obtained have demonstrated that the proposed dual-stream model has achieved superior performance across all evaluated datasets when compared to existing methods. Several key innovations have contributed to this improvement. First, the integration of a 1D-CNN stream and a transformer encoder stream has enabled the simultaneous extraction of local temporal features and global spatial-temporal dependencies from EEG signals, providing a richer and more comprehensive feature representation. Second, the incorporation of a multi-head attention mechanism during feature fusion has allowed the model to selectively emphasize the most informative features, enhancing discrimination between emotional classes. Third, the application of LGWO for feature selection has ensured that only the most relevant features have been retained, thereby reducing redundancy and improving classification accuracy. Additionally, the use of an ensemble learning strategy combining multiple classifiers has provided robustness in decision-making, mitigating potential biases associated with any individual classifier. Inference delay analysis has further confirmed that the model has been capable of real-time application, achieving an average classification delay of 5.2 milliseconds per trial, making it suitable for deployment in practical emotion recognition systems.

A direct comparison with related works has been presented in Table V, highlighting the advantages of the proposed model over prior approaches. Compared to the Att-1DCNN-GRU models proposed by Aldayel et al. [22], Z. Wang and Y. Wang [28], and J. Yan et al. [29], which have achieved accuracies around 95.95%, the proposed model has consistently achieved higher accuracy across all datasets. Similarly, models such as the Efficient Capsule Network with channel attention [26] and the spatial–temporal transformer network [27] have reached 94.67% and 92.67% accuracies respectively, but have not incorporated real-time capabilities or full model explainability. Although N. Ahmadzadeh et al. [25] have reported an internal accuracy of 98.21% using a convolutional fuzzy neural network, their results have been based on internal validation without subject-independent testing, limiting real-world generalization. Traditional approaches such as CNN–SAE–DNN hybrids [22] and wavelet–LSTM models [24] have also been outperformed, as their accuracies have ranged from 75.89% to 95.15%, often without sufficient cross-validation.

Furthermore, the models such as that of S. Adhikari et al. [78], although achieving over 99% internal validation accuracy, have not demonstrated robustness on external datasets and have lacked practical deployment evaluations. While some recent models have incorporated explainability partially, such as [78] with SHAP, none have integrated explainability as systematically and effectively as the proposed model through full SHAP analysis across all datasets.

Therefore, the combination of dual-stream feature extraction, attention-based fusion, bioinspired feature selection, ensemble-based classification, and comprehensive explainability has allowed the proposed model to establish a new benchmark for EEG-based emotion recognition systems. The consistent improvements achieved over related works in terms of accuracy, interpretability, and computational efficiency have validated the effectiveness.

VII. Conclusion

This paper, a novel dual-stream framework for EEG-based emotion recognition has been proposed and evaluated across four standard datasets: DEAP, SEED, BrainWave, and INTERFACE. By integrating 1D-CNN and transformer encoders for dual feature extraction, applying multi-head attention for feature fusion, using LGWO for feature selection, and employing ensemble learning for final classification, the model has consistently achieved superior performance.

The proposed model has attained an accuracy of 96.90% on the DEAP dataset, 94.25% on the SEED dataset, 93.70% on the BrainWave dataset, and 92.80% on the INTERFACE dataset, while maintaining high precision, recall, and F1 scores across all tasks. In addition, an average inference delay of 5.2 milliseconds per trial has been recorded, confirming the model's suitability for real-time emotion recognition applications. Furthermore, the inclusion of SHAP analysis has provided valuable insights into the contribution of individual EEG channels and frequency components, enhancing the model's transparency and clinical applicability. Overall, the proposed model has been validated as a robust, accurate, and explainable solution for preemptive detection of emotional disorders, setting a new benchmark for future research in EEG-based emotion classification.

TABLE V. COMPARISON WITH RELATED WORK

Study	Year	Method	Datasets	Acc (%)	Explainability	Real-Time Capability
Aldayel et al. [22]	2022	CNN + SAE + DNN	DEAP	89.49–92.86	No	No
H. Sun et al. [23]	2025	Dynamic EEG Source Connectivity	Custom	88.93/83.50	No	No
N. Ahmadzadeh et al. [25]	2024	Convolutional Fuzzy Neural Network (CFNN)	DEAP	98.21	No	No
J. Tian and X. Luo [85]	2025	Wavelet Features + LSTM	Custom	75.89–95.15	No	No
W. Tang et al. [26]	2025	Efficient-Capsule Network + Channel Attention	Custom	94.67	Partial	No
M. Li et al. [27]	2025	Spatial–Temporal Transformer	SEED	92.67	No	No
Z. Wang and Y. Wang [28]	2025	Att-1DCNN-GRU (EEG+ECG)	DEAP, SEED	95.95	No	No
J. Yan et al. [29]	2025	Spatio-Temporal Graph BERT (STGB)	SEED	92.5	No	No
J. A. Cruz-Vazquez et al. [30]	2025	CNN + Quantum Rotations	Custom	~95.0	No	No
J. Sedehi et al. [31]	2025	EEG–ECG Joint Connectivity	Custom	97.34	No	No
S. Adhikari et al. [78]	2025	Frequency Domain Features + RF + SHAP	Internal	>99	Yes (SHAP)	No
V. Doma and M. Pirouz [32]	2025	Comparative Supervised ML Models	EEG Tasks	<94	No	No
Proposed Model (Ours)		**Dual-Stream Transformer-CNN + LGWO + Ensemble + SHAP**	**DEAP, SEED, BrainWave, INTERFACE**	**96.9**	**Yes (Full SHAP)**	**Yes (5.2 ms)**

REFERENCES

[1] H. Berger, “Über das Elektrenkephalogramm des Menschen,” *Archiv für Psychiatrie und Nervenkrankheiten,* vol. 87, pp. 527-570, 1929.

[2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[3] P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[4] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[5] X. Xie, G. Cheng, J. Wang, K. Li, X. Yao, and J. Han, “Oriented R-CNN and Beyond,” *International Journal of Computer Vision,* vol. 132, no. 7, pp. 2420-2442, 2024.

[6] A. Bouguettaya and H. Zarzour, “CNN-based hot-rolled steel strip surface defects classification: a comparative study between different pre-trained CNN models,” *The International Journal of Advanced Manufacturing Technology,* vol. 132, no. 1, pp. 399-419, 2024.

[7] Y. Nie, Y. Chen, J. Guo, S. Li, Y. Xiao, W. Gong, and R. Lan, “An improved CNN model in image classification application on water turbidity,” *Scientific Reports,* vol. 15, no. 1, p. 11264, 2025.

[8] R. Du, T. Li, G. Meng, and F. Liu, “CNN-AC algorithm for hybrid precoding in millimeter-wave massive MIMO systems,” *Wireless Networks*, pp. 1-11, 2025.

[9] W. R. Murdhiono, H. Riska, N. Khasanah, Hamzah, and P. Wanda, “Mentalix: stepping up mental health disorder detection using Gaussian CNN algorithm,” *Iran Journal of Computer Science,* pp. 1-10, 2025.

[10] T. Al-Shehari *et al*., “Comparative evaluation of data imbalance addressing techniques for CNN-based insider threat detection,” *Scientific Reports,* vol. 14, no. 1, p. 24715, 2024.

[11] L. Yang, L. Lu, C. Liu, J. Zhang, K. Guo, N. Zhang, F. Zhou, and Y. Zhao, “Interactive exploration of CNN interpretability via coalitional game theory,” *Scientific Reports,* vol. 15, no. 1, p. 9261, 2025.

[12] P. Cai *et al*., “Enhancing quantum approximate optimization with CNN-CVaR integration,” *Quantum Information Processing,* vol. 24, no. 2, p. 37, 2025.

[13] S. K. Dewangan, S. Choubey, J. Patra, and A. Choubey, “IMU-CNN: implementing remote sensing image restoration framework based on Mask-Upgraded Cascade R-CNN and deep autoencoder,” *Multimedia Tools and Applications,* vol. 83, no. 27, pp. 69049-69081, 2024.

[14] R. Bhargava, N. Arivazhagan, and K. S. Babu, “Hybrid RMDL-CNN for speech recognition from unclear speech signal,” *International Journal of Speech Technology,* vol. 28, no. 1, pp. 195-217, 2025.

[15] S. B. M K and M. Kalra, “Leveraging CNN and Fundus Imaging for Enhanced Glaucoma Detection,” *SN Computer Science,* vol. 5, no. 8, p. 1137, 2024.

[16] K.-A. C. Quan, V.-T. Nguyen, T. V. Nguyen, and M.-T. Tran, “Unified ViT-CNN for few-shot object counting,” *Signal, Image and Video Processing,* vol. 19, no. 3, p. 221, 2025.

[17] J. Zhu *et al*., “Realization of normal temperature detection through visible light images by Retinex-CNN,” *Journal of Optics*, pp. 1-10, 2025.

[18] P. Dutta and N. B. Muppalaneni, “OCR Advancement with Pixel-Focused CNN for Handwritten Characters: A Journey with AsTel Dataset,” *Arabian Journal for Science and Engineering*, pp. 1-17, 2025.

[19] S. Prakash and K. Sangeetha, “Systems classification of air pollutants using Adam optimized CNN with XGBoost feature selection,” *Analog Integrated Circuits and Signal Processing,* vol. 122, no. 3, p. 35, 2025.

[20] C. Liu, “Landslide susceptibility mapping using CNN models based on factor visualization and transfer learning,” *Stochastic Environmental Research and Risk Assessment,* vol. 39, no. 1, pp. 231-249, 2025.

[21] N.-L. Pham, Q.-B. Ta, T.-C. Huynh, and J.-T. Kim, “CNN federated learning for vibration-based damage identification of submerged structure-foundation system,” *Journal of Civil Structural Health Monitoring*, pp. 1-26, 2025.

[22] B. Chakravarthi, S. Ng, M. Ezilarasan, and M. Leung, “EEG-based emotion recognition using hybrid CNN and LSTM classification,” *Frontiers in Systems Neuroscience,* vol. 16, pp. 1-9, 2022.

[23] H. Sun, H. Wang, R. Wang, Y. Gao, “Emotion recognition based on EEG source signals and dynamic brain function network,” *Journal of Neuroscience Methods,* vol. 415, 2025.

[24] J. Tian and X. Luo, “Emotion classification based on EEG wavelet features and LSTM network”, *Proceedings of the Fifth International Conference on Signal Processing and Computer Science (SPCS 2024)*, vol. 13442, pp. 87-95, 2025.

[25] N. Ahmadzadeh, N. Cavus, P. Esmaili, B. Sekeroglu, and S. Aşır, “Detecting emotions through EEG signals based on modified convolutional fuzzy neural network,” *Scientific Reports*, vol. 14, 2024.

[26] W. Tang, L. Fan, X. Lin, and Y. Gu, “EEG emotion recognition based on efficient-capsule network with convolutional attention, Biomedical

Signal Processing and Control," *Biomedical Signal Processing and Control,* vol. 103, 2025.

[27] M. Li, P. Yu, L. Zhang, and Y. Shen, "A spatial and temporal transformer-based EEG emotion recognition in VR environment," *Frontiers in Neuroscience,* vol. 19, 2025.

[28] Z. Wang and Y. Wang, "Emotion recognition based on multimodal physiological electrical signals," *Frontiers in Neuroscience,* vol. 19, p. 1512799, 2025.

[29] J. Yan, "Spatio-temporal graph Bert network for EEG emotion recognition," *Biomedical Signal Processing and Control*, vol. 104, 2025.

[30] J. A. Cruz-Vazquez, J. Y. Montiel-Pérez, R. Romero-Herrera, and E. Rubio-Espino, "Emotion recognition from EEG signals using advanced transformations and deep learning," *Mathematics*, vol. 13, no. 2, 2025.

[31] R. Singh and M. Sharma, "Develop an emotion recognition system using jointly connectivity between electroencephalogram and electrocardiogram signals," *Heliyon,* vol. 11, no. 2, 2025.

[32] V. Doma and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *Journal of Big Data*, vol. 7, 2025.

[33] S. Adhikari *et al*., "Analysis of frequency domain features for the classification of evoked emotions using EEG signals," *Experimental Brain Research,* vol. 243, no. 3, p. 65, 2025.

[34] D. K. Saha and T. D. Nath, "A lightweight CNN-based ensemble approach for early detecting Parkinson's disease with enhanced features," *International Journal of Speech Technology*, pp. 1-15, 2025.

[35] F. Zhang, B. Zhang, S. Guo, and X. Zhang, "MFCC-CNN: A patient-independent seizure prediction model," *Neurological Sciences,* vol. 45, no. 12, pp. 5897-5908, 2024.

[36] G. Wang, H. Zhang, M. Gao, W. Ding, and Y. Qian, "Identification and classification of power quality disturbances using CNN-transformer," *Journal of Electrical Engineering & Technology*, 2025.

[37] JS. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, "1-D Convolutional Neural Networks for Signal Processing Applications," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8360–8364, 2019.

[38] S. Mirjalili, "Grey Wolf Optimizer," *Advances in Engineering Software,* vol. 69, pp. 46-61, 2014.

[39] Y. Yan, and W. Liu, "Topical collections on machine learning based semantic representation and analytics for multimedia application," *Neural Computing and Applications,* vol. 34, no. 15, pp. 12239-12240, 2022.

[40] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions." *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4768 - 4777, 2027.

[41] K. Merabet *et al*., "Predicting water quality variables using gradient boosting machine: global versus local explainability using SHapley Additive Explanations (SHAP)," *Earth Science Informatics,* vol. 18, no. 3, p. 298, 2025.

[42] A. Koushik, M. Manoj, and N. Nezamuddin, "SHapley Additive exPlanations for Explaining Artificial Neural Network Based Mode Choice Models," *Transportation in Developing Economies,* vol. 10, no. 1, p. 12, 2024.

[43] K. D. Bathe and N. S. Patil, "ConvExNet: Deep learning-based flood detection utilizing Shapley additive explanations," *Journal of Earth System Science,* vol. 134, no. 2, p. 99, 2025.

[44] E. Çetin, C. Barrado, E. Salamí, and E. Pastor, "Analyzing deep reinforcement learning model decisions with Shapley additive explanations for counter drone operations," *Applied Intelligence,* vol. 54, no. 23, pp. 12095-12111, 2024.

[45] H. H. Nguyen, J.-L. Viviani, and S. Ben Jabeur, "Bankruptcy prediction using machine learning and Shapley additive explanations," *Review of Quantitative Finance and Accounting*, pp. 1-42, 2023.

[46] S. Heddam, "Explainability of Machine Learning Using Shapley Additive exPlanations (SHAP): CatBoost, XGBoost and LightGBM for Total Dissolved Gas Prediction," *Machine Learning and Granular Computing: A Synergistic Design Environment*, pp. 1-25, 2024.

[47] L. Chen, Z. He, Q. Ni, Q. Zhou, X. Long, W. Yan, Q. Sui, and J. Liu, "Dual-radiomics based on SHapley additive explanations for predicting hematologic toxicity in concurrent chemoradiotherapy patients," *Discover Oncology,* vol. 16, no. 1, p. 541, 2025.

[48] C. Kirabo, S. Murindanyi, N. P. Kirabo, K. M. Hasib, and G. Marvin, "SHapley Additive exPlanations for Machine Emotion Intelligence in CNNs," *International Conference on Computational Intelligence*, pp. 657-671, 2023.

[49] N. Sriwiboon, "Efficient and lightweight CNN model for COVID-19 diagnosis from CT and X-ray images using customized pruning and quantization techniques," *Neural Computing and Applications*, vol. 37, no. 18, pp. 13059-13078, 2025.

[50] N. K. Mishra, P. Singh, A. Gupta, and S. D. Joshi, "PP-CNN: probabilistic pooling CNN for enhanced image classification," *Neural Computing and Applications,* vol. 37, no. 6, pp. 4345-4361, 2025.

[51] N. Kaur, S. Pandey, and N. Kalra, "MFR-CNN: A modified faster R-CNN approach based on bounding box and reliable score for cloth image retrieval," *Multimedia Tools and Applications*, pp. 1-29, 2024.

[52] R. Saffarini, F. Khamayseh, Y. Awwad, M. Sabha, and D. Eleyan, "Dynamic generative R-CNN," *Neural Computing and Applications*, vol. 37, no. 10, pp. 7107-7120, 2025.

[53] C. Gao and H. Ge, "I-CNN-LSTM: An Improved CNN-LSTM for Transient Stability Analysis of More Electric Aircraft Power Systems," *Arabian Journal for Science and Engineering*, vol. 50, no. 8, pp. 5683-5696, 2025.

[54] H. Aouani and Y. Ben Ayed, "Deep facial expression detection using Viola-Jones algorithm, CNN-MLP and CNN-SVM," *Social Network Analysis and Mining,* vol. 14, no. 1, p. 65, 2024.

[55] S. Davoudi and K. Roushangar, "Innovative approaches to surface water quality management: advancing nitrate (NO3) forecasting with hybrid CNN-LSTM and CNN-GRU techniques," *Modeling Earth Systems and Environment,* vol. 11, no. 2, p. 80, 2025.

[56] Pranav and R. Katarya, "Effi-CNN: real-time vision-based system for interpretation of sign language using CNN and transfer learning," *Multimedia Tools and Applications,* vol. 84, no. 6, pp. 3137-3159, 2025.

[57] H. Dehnavi, M. Dehnavi, and S. H. Klidbary, "Fcd-cnn: FPGA-based CU depth decision for HEVC intra encoder using CNN," *Journal of Real-Time Image Processing,* vol. 21, no. 4, p. 105, 2024.

[58] I. Linck, A. T. Gómez, and G. Alaghband, "SVG-CNN: A shallow CNN based on VGGNet applied to intra prediction partition block in HEVC," *Multimedia Tools and Applications,* vol. 83, no. 30, pp. 73983-74001, 2024.

[59] M. Telmem, N. Laaidi, Y. Ghanou, S. Hamiane, and H. Satori, "Comparative study of CNN, LSTM and hybrid CNN-LSTM model in amazigh speech recognition using spectrogram feature extraction and different gender and age dataset," *International Journal of Speech Technology*, vol. 27, no. 4, pp. 1121-1133, 2024.

[60] S. Esteki and A. R. Naghsh-Nilchi, "SW/SE-CNN: semi-wavelet and specific image edge extractor CNN for Gaussian image denoising," *Neural Computing and Applications,* vol. 36, no. 10, pp. 5447-5469, 2024.

[61] M. Asfand-e-yar, Q. Hashir, A. A. Shah, H. A. M. Malik, A. Alourani, and W. Khalil, "Multimodal CNN-DDI: using multimodal CNN for drug to drug interaction associated events," *Scientific Reports,* vol. 14, no. 1, p. 4076, 2024.

[62] K. G. Panchbhai, M. G. Lanjewar, V. V. Malik, and P. Charanarur, "Small size CNN (CAS-CNN), and modified MobileNetV2 (CAS-MODMOBNET) to identify cashew nut and fruit diseases," *Multimedia Tools and Applications,* vol. 83, no. 42, pp. 89871-89891, 2024.

[63] M. Kaddes, Y. M. Ayid, A. M. Elshewey, and Y. Fouad, "Breast cancer classification based on hybrid CNN with LSTM model," *Scientific Reports,* vol. 15, no. 1, p. 4409, 2025.

[64] E. Pintelas, I. E. Livieris, V. Tampakas, and P. Pintelas, "Feature augmentation-based CNN framework for skin-cancer diagnosis," *Evolving Systems,* vol. 16, no. 1, p. 34, 2025.

[65] J. Mishra and R. K. Sharma, "Optimized FPGA Architecture for CNN-Driven Voice Disorder Detection," *Circuits, Systems, and Signal Processing,* vol. 44, no. 6, pp. 4455-4467, 2025.

[66] R. Nambiar and R. N, "A Comprehensive Review of AI and Deep Learning Applications in Dentistry: From Image Segmentation to Treatment Planning," *Journal of Robotics and Control (JRC),* vol. 5, p. 2024, 2024.

[67] T.-V. Dang and L. Tran, “A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet,” *Journal of Robotics and Control (JRC),* vol. 4, pp. 431-436, 2023.

[68] T. Admassu, T. Suresh, R. Purushothaman, S. Ganesan, and K. K. Napa, “Early Prediction of Gestational Diabetes with Parameter-Tuned K-Nearest Neighbor Classifier,” *Journal of Robotics and Control (JRC),* vol. 4, pp. 452-457, 2023.

[69] S. Phimphisan and N. Sriwiboon, “A Customized CNN Architecture with CLAHE for Multi-Stage Diabetic Retinopathy Classification,” *Engineering, Technology & Applied Science Research,* vol. 14, no. 6, pp. 18258-18263, 2024.

[70] N. Sriwiboon and S. Phimphisan, “Efficient COVID-19 Detection using Optimized MobileNetV3-Small with SRGAN for Web Application,” *Engineering, Technology & Applied Science Research,* vol. 15, no. 2, pp. 20953-20958, 2025.

[71] I. Uluocak and M. Bilgili, “Daily air temperature forecasting using LSTM-CNN and GRU-CNN models,” *Acta Geophysica,* vol. 72, no. 3, pp. 2107-2126, 2024.

[72] T. Li, J. Shu, and Y. Wang, “Deformation prediction of underground engineering support structures via the ST-CNN-LSTM model,” *Journal of Civil Structural Health Monitoring*, pp. 1-19, 2025.

[73] E. R. Coutinho, J. G. F. Madeira, D. G. F. Borges, M. V. Springer, E. M. de Oliveira, and A. L. G. A. Coutinho, “Multi-Step Forecasting of Meteorological Time Series Using CNN-LSTM with Decomposition Methods,” *Water Resources Management*, pp. 1-26, 2025.

[74] V. Singh, S. K. Sahana, and V. Bhattacharjee, “A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction,” *Discover Computing,* vol. 28, no. 1, p. 38, 2025.

[75] A. Shaik, S. S. Dutta, I. M. Sawant, S. Kumar, A. Balasundaram, and K. De, “An attention based hybrid approach using CNN and BiLSTM for improved skin lesion classification,” *Scientific Reports,* vol. 15, no. 1, p. 15680, 2025.

[76] X. Bai, L. Zhang, Y. Feng, H. Yan, and Q. Mi, “Multivariate temperature prediction model based on CNN-BiLSTM and RandomForest,” *The Journal of Supercomputing,* vol. 81, no. 1, p. 162, 2024.

[77] Q. Tian, R. Cai, Y. Luo, and G. Qiu, “DOA Estimation: LSTM and CNN Learning Algorithms,” *Circuits, Systems, and Signal Processing,* vol. 44, no. 1, pp. 652-669, 2025.

[78] M. Miao, J. Liang, Z. Sheng, W. Liu, B. Xu, and W. Hu, “ST-SHAP: A hierarchical and explainable attention network for emotional EEG representation learning and decoding,” *Journal of Neuroscience Methods*, vol. 414, 2025.

[79] S. Koelstra *et al*., "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, Jan.-March 2012.

[80] W. Lu, T.-P. Tan, and H. Ma, “Bi-Branch Vision Transformer Network for EEG Emotion Recognition,” *IEEE Access*, vol. 11, pp. 36233-36243, 2023.

[81] P. Dutta, S. Paul, K. Cengiz, R. Anand, and A. Kumar, "A predictive method for emotional sentiment analysis by deep learning from EEG of brainwave dataset," *Artificial intelligence for neurological disorders*, pp. 25-48, 2023.

[82] G. Y. Choi *et al*., “EEG Dataset for the Recognition of Different Emotions Induced in Voice-User Interaction,” *Scientific Data*, vol. 11, no. 1, 2024.

[83] A. Howard *et al*., "Searching for mobilenetv3," *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314-1324, 2019.

[84] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters,* vol. 27, no. 8, pp. 861-874, 2006.

[85] J. Tian and X. Luo, “Emotion classification based on EEG wavelet features and LSTM network,” *Proceedings of the Fifth International Conference on Signal Processing and Computer Science (SPCS 2024)*, vol. 13442, pp. 87-95, 2025.