УДК:681.3.3.01+681.325
doi.org/10.15407/fmmit2024.39.167

Enhancing emotion classification through signal fusion and wavelet-based feature extraction

Oleksandr Sukholeister[1], Rostyslav Nakonechny[2]
[1]PhD Student, Lviv Polytechnic National University, 79013, Lviv, Stepan Bandera 12, e-mail: oleksandr.o.sukholeister@lpnu.ua
[2]Rostyslav Nakonechny, PhD, Docent, Lviv Polytechnic National University, 79013, Lviv, Stepan Bandera 12, e-mail: rostyk.ws@gmail.com

Emotion recognition from physiological signals, such as EEG, ECG, and GSR, has shown promise for various applications in affective computing. This study introduces a wavelet-based approach that leverages continuous wavelet transform (CWT) to extract both image and numerical features from multi-modal physiological data. The extracted features are utilized to train and compare classification models, including a ResNet-50-based deep learning framework and traditional machine learning models like support vector machines (SVMs). Our results demonstrate that the combination of wavelet-based feature extraction and signal fusion significantly enhances emotion classification performance. Notably, the numerical features derived from CWT achieve comparable or superior accuracy to image-based features, offering insights into the effectiveness of different feature representations. This work highlights the potential of wavelet-based methods for emotion recognition and suggests pathways for future research in optimizing multi-modal signal processing and classification.

Keywords: Emotion recognition, Continuous Wavelet Transform, EEG, ECG, GSR, Multimodal Fusion, Deep Learning, Affective Computing

Introduction. The understanding and recognition of human emotions through physiological signals have become central to advancements in affective computing and human-computer interaction. The accurate classification of emotions is critical in domains ranging from healthcare and education to entertainment and intelligent transportation systems. Physiological signals such as electroencephalography (EEG), electrocardiography (ECG), and galvanic skin response (GSR) provide reliable indicators of emotional states due to their resistance to manipulation, unlike behavioral signals such as facial expressions or speech.

Recent work in emotion recognition emphasizes the use of multimodal data fusion, which combines different physiological signals to improve the accuracy and robustness of emotion detection systems. The integration of multiple signal modalities allows for a more comprehensive understanding of human emotions by capturing complex interdependencies between physiological processes. However, effectively extracting and fusing these signals presents significant challenges, requiring sophisticated techniques for feature extraction, fusion, and classification.

1. Related Work.

The field of emotion recognition has seen the application of numerous feature extraction and classification techniques. Continuous wavelet transform (CWT) has emerged as a powerful method for analyzing non-stationary signals like EEG and ECG, providing a detailed representation in both time and frequency domains. Studies such as the one by Mohammadi et al. [4] demonstrated the efficacy of wavelet-based feature extraction, achieving high classification accuracy using discrete wavelet transform (DWT) on EEG signals for arousal and valence classification.

The use of pre-trained deep learning models, such as ResNet-50, for classifying wavelet-transformed images has been explored in several studies. Deep learning approaches leverage the rich representations of CWT images, significantly outperforming traditional classifiers in tasks requiring the understanding of complex signal patterns [5, 8]. On the other hand, support vector machines (SVMs) remain a popular choice for classifying numerical features derived from physiological signals, offering robustness and efficiency in high-dimensional spaces [5, 9].

The MAHNOB-HCI database has been extensively used for emotion recognition experiments, providing a well-structured set of multimodal physiological recordings that facilitate the development and testing of novel algorithms. The dataset's comprehensive annotations in terms of both valence and arousal make it a suitable benchmark for evaluating the effectiveness of signal fusion strategies [2, 3].

2. Challenges in Multimodal Emotion Recognition.

Despite advancements, several challenges persist in the domain of multimodal emotion recognition. The first challenge lies in the variability and complexity of physiological signals, which are affected by numerous external factors, such as individual differences and environmental conditions [7]. Signal fusion techniques, including sensor fusion, feature fusion, and decision fusion, have been developed to mitigate these challenges by leveraging the strengths of each modality. Decision-level fusion, in particular, has shown promise for combining the outputs of different classifiers to enhance overall performance, though it introduces its own set of complexities in terms of model integration and optimization [3, 5].

3. Method.

3.1. Data Acquisition

The experiments in this study leverage the MAHNOB-HCI dataset, a comprehensive multimodal database for affective computing and emotion recognition research. The dataset contains synchronized recordings of physiological and behavioral data collected from 27 participants during emotion elicitation sessions. Specifically, the MAHNOB-HCI dataset provides EEG, ECG, and GSR signals, annotated for arousal and valence levels. Using this dataset allows for a robust evaluation of the proposed methods, given its high-quality recordings and curated emotional annotations. The dataset was organized into a unified format to enable consistent preprocessing and analysis with use of EEGLAB MATLAB Add-On with BioSig extension.

3.2. Signal Selection.

The selection of EEG, ECG, and GSR signals is motivated by their established roles in reflecting emotional states. EEG captures electrical activity from the brain and has been extensively used for identifying emotional valence and arousal due to its high temporal resolution. For the purpose of some experiments in this study F3 and F4 EEG channels were pre-selected (out of 32 total EEG channels), ECG1 (out of 3) and the GSR signals. For the EEG, key regions, such as the prefrontal cortex, are especially relevant for emotion processing, with electrodes like F3 and F4 providing crucial information. F3 is generally associated with approach-related emotions and is critical in the prefrontal cortex's processing of emotional stimuli. It has shown a strong correlation with emotional valence, especially in distinguishing between positive and negative emotions. F4 is linked to withdrawal-related emotions and plays a significant role in emotional regulation and arousal detection. It is valuable for differentiating emotional states with varying levels of arousal and valence. ECG records cardiac activity, with features such as heart rate variability serving as reliable indicators of emotional arousal. GSR, or skin conductance, measures changes in sweat gland activity, which correlates with sympathetic nervous system activation and emotional arousal. Combining these modalities allows for a comprehensive representation of an individual's emotional state, leveraging the complementary nature of neural, cardiac, and autonomic signals.

3.3. Denoising.

The raw signals from the dataset were preprocessed using MATLAB signal processing toolbox. Since some records contained trailing zeroes due to padding or incomplete recordings, trailing zeroes were removed from all signals prior to filtering to ensure computational efficiency and signal integrity. The session is skipped if any of the signals become empty after trimming. Each signal type is subjected to specific denoising techniques to ensure clean and artifact-free data:

EEG: A bandpass filter is applied between 0.5 Hz and 45 Hz along the time dimension, followed by manual z-score normalization to standardize the signals.

ECG: The ECG signal is filtered using a sequence of filters:

1. A bandpass filter with a frequency range of 0.5 Hz to 40 Hz.
2. A bandstop (Notch) filter between 49 Hz and 51 Hz to remove powerline interference.
3. A high-pass filter at 0.5 Hz to eliminate low-frequency drift.
4. The signal is then normalized using z-score.

GSR: A low-pass filter with a cutoff frequency of 5 Hz is applied, followed by normalization.

3.4. Segmentation.

The denoised signals are segmented into non-overlapping windows of 5, 10, and 20 seconds with use of max 10 segments from the end of each recording. These segment lengths are selected to evaluate the impact of temporal resolution on classification performance, with shorter segments potentially capturing transient emotional changes and longer segments providing more stable representations of emotional states.

3.5. Feature Extraction.

Feature extraction from the segmented signals is performed using continuous wavelet transform (CWT).

The CWT of a signal x(t) is defined as:

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t)\,\psi^*(\frac{t-b}{a})dt$$

where ψ is the mother wavelet, a is the scale parameter, b is the translation parameter, and * denotes complex conjugation [10].

The Morlet mother wavelet was chosen for this study because of its advantageous time-frequency localization, making it suitable for analyzing the transient and non-stationary nature of physiological signals such as ECG, EEG, and GSR. For EEG, the Morlet wavelet has been shown to effectively capture brain activity linked to emotional states [11]. The Morlet wavelet's ability to provide a clear representation of signal components also makes it appropriate for ECG and GSR signals, where precise detection of spectral components is essential for evaluating heart rate variability and skin conductance patterns associated with emotional responses.

The Continuous Wavelet Transform (CWT) was used to extract features from physiological signals in two different ways: as a numerical feature vector and as an image representation. The numerical feature vector consists of dozens to hundreds of features, making it efficient for training and classification with simpler models. In contrast, the image representation creates a much larger multidimensional array with hundreds of thousands of pixels, offering a rich visual representation of the time-frequency signal representation. This representation typically requires more complex classifiers, such as deep learning models, and benefits from transfer learning, where pre-trained networks can be leveraged to improve accuracy. It comes at the cost of increased computational complexity and longer training times.

3.5.1. Numeric Feature Vector.

For each segment, CWT coefficients are computed and summarized into numerical features: Mean and standard deviation of the coefficients within specific frequency bands, energy and entropy to quantify signal characteristics. Entropy is a measure of the randomness of a signal and can act as a feature to analyze psychological time series data.

The energy of each frequency band is calculated as:

$$E = \sum_{i=1}^{N} |c_i|^2$$

where Ci represents the continuous wavelet transform (CWT) coefficients for a given frequency band and N is the total number of coefficients in that band.

The entropy for each frequency band is computed as:

$$H = -\sum_{i=1}^{N} c_i \log(c_i + \epsilon)$$

where Ci represents the continuous wavelet transform (CWT) coefficients for a given frequency band, N is the total number of coefficients in that band and ϵ is a small positive constant added to avoid taking the logarithm of zero.

The following frequency bands of interest were selected based on their relevance to emotion recognition as highlighted in scientific literature:

EEG: Emotional processing in EEG signals is commonly associated with specific frequency bands:

1. Delta (0.5–4 Hz): Linked to deep sleep and basic emotional processing.
2. Theta (4–8 Hz): Related to memory, attention, and emotional arousal.
3. Alpha (8–13 Hz): Associated with relaxation and attentional states.
4. Beta (13–30 Hz): Involved in active thinking and emotional engagement.
5. Gamma (30–45 Hz): Linked to higher cognitive processing and the integration of sensory and emotional information.

ECG: Heart Rate Variability (HRV) analysis divides the ECG spectrum into frequency bands:

1. Low Frequency (LF: 0.04–0.15 Hz): Reflects both sympathetic and parasympathetic activity.
2. High Frequency (HF: 0.15–0.4 Hz): Corresponds to parasympathetic activity, often linked to emotional arousal.
3. Very Low Frequency (VLF: 0.003–0.04 Hz): Not used in this study because reliable analysis of VLF requires at least a 2-minute segment length.

GSR: Frequencies below 5 Hz are relevant for measuring skin conductance responses, indicative of sympathetic nervous system activity.

These frequency bands have been widely utilized in emotion recognition research due to their established relationships with emotional states [1,4,5].

3.5.2. Image Representation.

The CWT coefficients were converted into RGB images with dimensions suitable for input into the ResNet-50 network (224 pixels in width and height, with 3 color channels). These images are used as input for a pre-trained ResNet-50 network, fine-tuned for emotion classification. The CWT images offer a visual representation of the signal's time-frequency content, facilitating the use of powerful convolutional neural network architectures (Fig. 1, Fig. 2).

The Image Representation of the Continuous Wavelet Transform (CWT) was created by transforming signal segments into a time-frequency representation using the CWT. Specifically:

1. Each signal was decomposed into CWT coefficients using the Morlet mother wavelet, which captures both frequency and temporal information.
2. The absolute values of the CWT coefficients were computed to form a scalogram, a visual representation of the signal's energy distribution across different scales (or frequencies) and time.
3. The scalogram was then rescaled to an intensity range suitable for image representation and mapped to a color scale (using the 'jet' colormap), resulting in a visually interpretable image.
4. To ensure uniformity, the scalogram images were resized to a predefined target size, making them suitable for input into deep learning models.

This image-based approach captures complex signal characteristics and provides a rich feature representation for subsequent classification tasks. Fig. 1, Fig. 2 and Fig. 3 represent example signals of each modality separately (ECG, EEG and GSR).

Fig. 4 illustrates the example CWT representation of four physiological signals—two EEG channels (F3 and F4), one ECG channel, and one GSR signal—arranged in a 2x2 grid format. Each subplot presents the time-frequency decomposition of the corresponding signal, highlighting the distinct patterns and features captured by the CWT. This visual layout provides an intuitive comparison across different physiological modalities, allowing for a comprehensive analysis of how each signal contributes to emotion recognition. The structured grid approach ensures that spatial relationships between the modalities are preserved, making it suitable for further processing by convolutional neural networks.

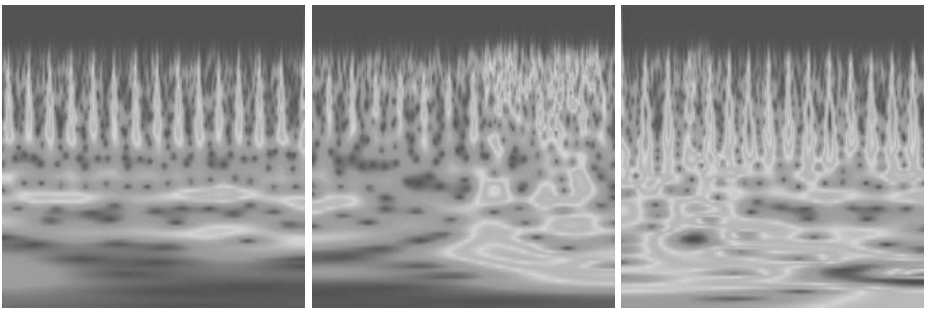

Fig. 1. Sample Image CWT Representation of ECG Signal (3 samples with 10s length)

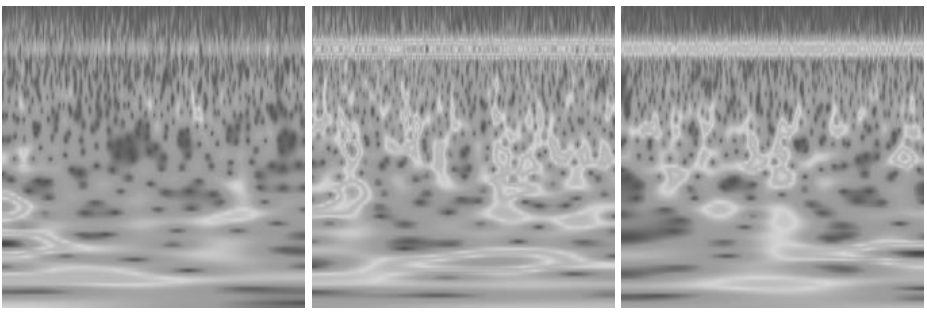

Fig. 2. Sample Image CWT Representation of EEG Signal (3 samples with 10s length)

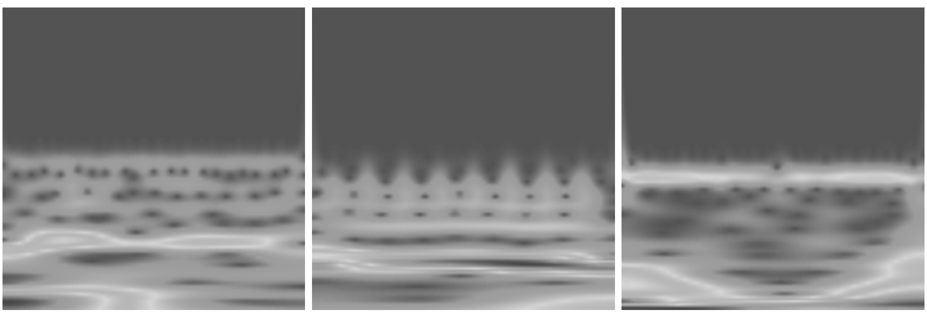

Fig. 3. Sample Image CWT Representation of GSR Signal (3 samples with 10s length)

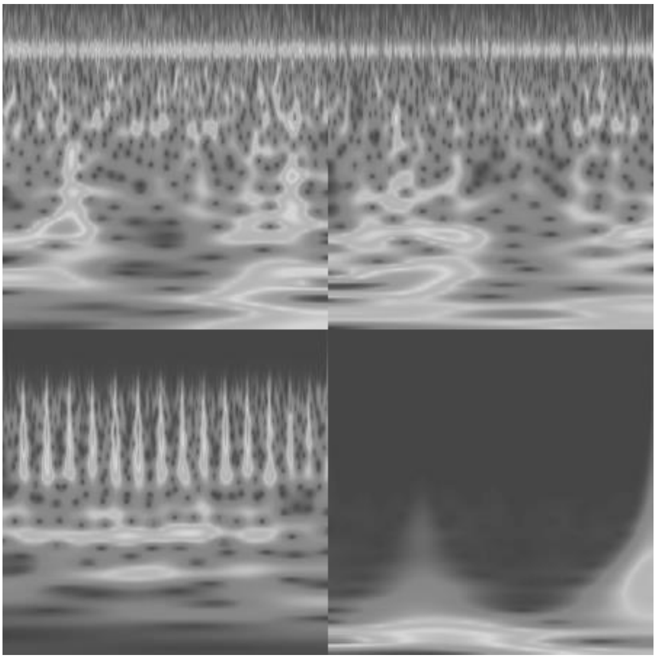

Fig. 4. Sample Image CWT Representation of 4 Signals (EEG F3, EEG F4, ECG, GSR)

This approach to image representation transforms physiological signals into detailed, visually interpretable CWT images. As a result, we get clear time-frequency patterns from each signal, organized in a format that is ideal for deep learning models, like ResNet-50. These images can then be used for emotion classification, leveraging the model's ability to learn complex features from the data.

3.6. Dataset Preparation.

Labels for arousal and valence were binarized using the median value as the threshold, converting the problem into a binary classification task. To address class imbalance, the dataset was balanced using downsampling, where the majority class was reduced to match the size of the minority class. This approach was chosen for its simplicity and to ensure that the training data did not introduce artificial patterns that could bias the model. The implications of class imbalance on the model's performance was analyzed, and downsampling was performed to reduce the overall amount of training data, potentially affecting model generalization. Future work could explore more sophisticated balancing methods, such as ensemble techniques or data augmentation, to mitigate these limitations while preserving the characteristics of physiological signals.

The balanced data is split into training, validation, and test sets using a 70/20/10 ratio. This split ensures that the model's performance is evaluated on unseen data while providing sufficient data for training and hyperparameter tuning.

3.7. Classification.

Following classification approaches are employed:

Support Vector Machine (SVM): Numeric feature vectors extracted from CWT are classified using an SVM. The SVM is chosen for its effectiveness in high-dimensional spaces and its robustness to overfitting.

ResNet-50: The modified ResNet50 network architecture starts with the original pre-trained ResNet50 convolutional layers, which are used for feature extraction. The final classification layers (fully connected, softmax, and output layers) of the original network have been removed and replaced with custom layers tailored for binary classification. These new layers consist of a fully connected layer with two output neurons (corresponding to the two classes), a softmax activation layer, and a classification output layer. This structure leverages the pre-trained weights for feature extraction while enabling the network to be fine-tuned for binary classification tasks.

3.8. Evaluation.

The performance of the models is evaluated using accuracy, precision, recall, and F1-score, computed on the test set. These metrics provide a comprehensive assessment of the models' classification capabilities, with F1-score offering a balanced measure of precision and recall.

4. Results.

This section presents the results of our experiments evaluating the effectiveness of continuous wavelet transform (CWT) features in arousal and valence classification using both numeric and image representations. The results cover various experimental conditions, including varying segment lengths, individual versus merged signal modalities, and comparisons between numeric and image-based feature representations.

Table 1.

Arousal classification accuracy (%) from numerical representation of CWT features.

Arousal Accuracy vs segment length	EEG F3	EEG F4	ECG1	GSR1	EEG F3, F4, ECG1, GSR1	32xEEG, 3xECG, 1xGSR
5s	55.47	**59.38**	47.27	54.30	61.52	69.53
10s	54.79	57.34	**57.34**	52.05	60.27	72.41
20s	**57.48**	55.72	54.30	**56.89**	**62.46**	**73.02**

Table 2

Valence classification accuracy (%) from numerical representation of CWT features. The results show better performance with longer segments.

Valence Accuracy vs segment length	EEG F3	EEG F4	ECG1	GSR1	EEG F3, F4, ECG1, GSR1	32xEEG, 3xECG, 1xGSR
5s	**57.03**	52.73	48.83	50.59	58.40	69.73
10s	55.38	**56.36**	**59.53**	53.82	60.27	**72.80**
20s	50.73	53.37	54.30	**56.89**	**60.41**	72.73

4.1. Impact of Segment Length on Classification Performance.

The experiments tested arousal and valence classification accuracy using three segment lengths: 5 seconds, 10 seconds, and 20 seconds. The segment length affects the granularity of the data, which in turn influences the classifier's ability to capture emotional variations. The results indicate that longer segment lengths generally lead to higher accuracy, likely due to the availability of more data within each segment to capture the emotional state. However, the improvement tapers off beyond 20 seconds, suggesting a trade-off between temporal resolution and classification performance.

Arousal Classification (Table 1): The performance generally increased with longer segments. For instance, when using the merged set of all signals (32 EEG channels, 3 ECG channels, and 1 GSR channel), accuracy improved from 69.53% for 5-second segments to 73.02% for 20-second segments. However, the gain was not consistent across all individual modalities. For example, using the GSR signal alone showed only minor variations across different segment lengths.

Valence Classification (Table 2): Similarly, valence classification accuracy showed improvements with longer segments. The merged signals resulted in the highest accuracy of 72.80% for 10-second segments and slightly dropped to 72.73% for 20-second segments. Notably, EEG F3 performs better at shorter segments (57.03% at 5s), indicating that valence recognition may benefit from higher temporal granularity.

These results suggest that longer segments generally facilitate more accurate classifications, particularly when using a comprehensive set of modalities. However, certain signals may exhibit diminishing returns or even performance drops as segment length increases.

4.2. Individual vs. Merged Modalities.

The classification performance was compared using single signals (e.g., EEG F3, EEG F4, ECG1, and GSR1) against combinations of these signals.

Arousal Classification (Table 1): The highest performance was consistently observed with the merged modalities, achieving an accuracy of 73.02% with 20-second segments. Individual signals, such as ECG1 and GSR1, yielded lower accuracies, often below 60%. Notably, EEG signals (e.g., F4) performed better than ECG and GSR individually but were still outperformed by the merged signal set.

Valence Classification (Table 2): The merged modalities also outperformed individual signals for valence recognition, with the highest accuracy of 72.80% at 10-second segments. Individually, EEG F3 and ECG1 showed moderate performance, while GSR1 consistently ranked lower.

These results highlight the benefits of multimodal fusion, which effectively captures the complementary information provided by EEG, ECG, and GSR signals. The improved performance with merged modalities underscores the advantage of utilizing signal interdependencies for emotion recognition.

4.3. Comparison of Numeric vs. Image-Based Features.

The study also compared the performance of numeric feature vectors extracted from CWT with image-based features processed using a ResNet-50 network. Experiments for this comparison were conducted on 10-second segments using select signal combinations (Table 3).

Table 3.

Impact of feature representation onto classification performance

Arousal Classification	EEG F4	EEG F4	ECG1	ECG1	EEG F3, F4, ECG1, GSR1	EEG F3, F4, ECG1, GSR1
Features representation	Vector	Image	Vector	Image	Vector	Image
Num. features	20	224x224x3	12	224x224x3	56	224x224x3
Classifier	SVM	ResNet50	SVM	ResNet50	SVM	ResNet50
Segment	10s	10s	10s	10s	10s	10s
Accuracy	57.34	56.79	57.34	61.49	60.27	60.18

Arousal Classification (10s Segments): Using the EEG F4 signal, the numerical features classified with SVM achieved an accuracy of 57.34%, while the image-based ResNet-50 classifier yielded 56.79%. For the merged set (EEG F3, F4, ECG1, GSR1), numerical features with SVM reached an accuracy of 60.27%, compared to 60.18% for image-based ResNet-50.

The numeric representation of CWT features generally matched or slightly outperformed image-based classifications for EEG and combined signals. Image based representation outperforms numeric representation only for ECG signal, which may be caused by low number of numerical features for ECG or better suitability of image CWT for ECG signal. This outcome may reflect the effectiveness of traditional feature extraction techniques in capturing relevant information for SVM classifiers. The comparable performance between SVM and ResNet-50 also suggests that image-based approaches require further optimization to produce consistent improvements.

Conclusion. This study explores the effectiveness of emotion classification using EEG, ECG, and GSR signals from the MAHNOB-HCI dataset, focusing on continuous wavelet transform (CWT) feature extraction and multimodal fusion. Our findings demonstrate that longer segment lengths, especially 20-second windows, generally result in higher classification accuracy for both arousal and valence. The fusion of multiple modalities outperforms individual signal analysis.

Our results suggest that numerical CWT features perform as well as or better than image-based features processed by a ResNet-50 network, indicating the robustness of traditional feature extraction methods for emotion recognition.

Future research could explore advanced deep learning architectures designed specifically for time-series analysis. For instance, temporal convolutional networks (TCNs) or recurrent neural networks (RNNs), such as long short-term memory (LSTM) or gated recurrent units (GRUs), may better capture temporal dependencies in physiological signals. These architectures could potentially enhance classification performance by more effectively modeling the dynamic nature of emotional states.

Moreover, incorporating model regularization and hyperparameter tuning could further optimize deep learning models. Building on the success of multimodal fusion in this study, experimenting with more sophisticated fusion strategies, such as attention mechanisms or transformers, could provide significant improvements. Attention mechanisms can selectively focus on the most relevant parts of each signal, enhancing the model's ability to capture complex interdependencies among EEG, ECG, and GSR signals. Additionally, exploring hierarchical fusion techniques that combine feature-level and decision-level fusion may provide a more comprehensive representation of emotional states. The integration of wavelet coherence features can be included in future work. Unlike traditional CWT features that analyze each signal separately, wavelet coherence can quantify the phase and frequency relationships between two signals over time. Analyzing wavelet coherence could provide deeper insights into the synchronization and interaction of different physiological signals, which may be critical for emotion recognition. For example, analyzing the coherence between EEG and ECG signals could reveal important neural-cardiac patterns associated with emotional states.

These results highlight the potential of multimodal approaches combined with both classical and deep learning techniques to improve emotion classification systems, contributing to advancements in affective computing and human-computer interaction.

References

1. Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A Multimodal Database for Affect Recognition and Implicit Tagging. IEEE transactions on affective computing, 3(1), 42-55.
2. Sukholeister, O., Nakonechnyi, A. (2022). Recognition of Mental Disorders from Physiological Signals Analysis. Measuring Equipment and Metrology, 83(4), 11-12
3. Sukholeister, O., Nakonechnyi, A. (2023). Strategies for Physiological Signal Fusion for Emotion Recognition. Measuring and Computing Devices in Technological Processes, Issue 4, 34-35.
4. Mohammadi, Z., Frounchi, J., & Amiri, M. (2016). Wavelet-based Emotion Recognition System Using EEG Signal. Neural Computing & Applications, 28, 1985-1990.
5. Tawsif, K., Aziz, N. A., Raja, J. E., & Hossen, J. (2022). A Systematic Review on Emotion Recognition System Using Physiological Signals: Data Acquisition and Methodology. Emerging Science Journal, 6(5), 1167-1168.
6. Wang, Y., Song, W., Tao, W., et al. (2022). A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. Affective Computing Journal, 21, 1-4.
7. Ahmad, Z., & Khan, N. (2022). A Survey on Physiological Signal-Based Emotion Recognition. Bioengineering, 9, 688.
8. Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2021). Human Emotion Recognition: Review of Sensors and Methods. Sensors, 20, 592.
9. Egger, M., Ley, M., & Hanke, S. (2021). Emotion Recognition from Physiological Signal Analysis: A Review. Electronic Notes in Theoretical Computer Science, 343, 35-55.
10. Addison, P. S. (2017) The Illustrated Wavelet Transform Handbook. 2nd edn. CRC Press.
11. Bajaj, V., Pachori, R.B. (2015). Detection of Human Emotions Using Features Based on the Multiwavelet Transform of EEG Signals. In: Hassanien, A., Azar, A. (eds) Brain-Computer Interfaces. Intelligent Systems Reference Library, vol 74. Springer, Cham.

Покращення класифікації емоцій за допомогою злиття сигналів і виділення ознак на основі вейвлетів

Олександр Сухолейстер, Ростислав Наконечний

Розпізнавання емоцій за фізіологічними сигналами, такими як електроенцефалограма (ЕЕГ), електрокардіограма (ЕКГ) та шкірно-гальванічна реакція (ШГР), показало багатообіцяюче застосування в афективних обчисленнях. Це дослідження представляє підхід на основі вейвлетів, який використовує безперервне вейвлет-перетворення (БВП) для вилучення як зображення, так і числових характеристик із мультимодальних фізіологічних даних. Вилучені ознаки використовуються для навчання та порівняння моделей класифікації, включаючи модель глибокого навчання на основі ResNet-50 і традиційні моделі машинного навчання, такі як опорні векторні машини (ОВМ). Наші результати демонструють, що поєднання вилучення ознак на основі вейвлетів і злиття сигналів значно покращує ефективність класифікації емоцій. Числові характеристики, отримані за допомогою CWT, досягають такої ж або кращої точності, ніж функції на основі зображень. Ця робота підкреслює потенціал методів на основі вейвлетів для розпізнавання емоцій і пропонує шляхи майбутніх досліджень оптимізації мультимодальної обробки та класифікації сигналів.

Отримано 16.11.24.