frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 05 December 2023
DOI 10.3389/fcomp.2023.1264713

OPEN ACCESS

EDITED BY
Xipei Ren,
Beijing Institute of Technology, China

REVIEWED BY
Pallavi Pandey,
KR Mangalam University, India
Cun Li,
Jiangnan University, China

*CORRESPONDENCE
Mengru Xue
✉ mengruxue@zju.edu.cn

RECEIVED 21 July 2023
ACCEPTED 17 November 2023
PUBLISHED 05 December 2023

CITATION
Gohumpu J, Xue M and Bao Y (2023) Emotion recognition with multi-modal peripheral physiological signals.
Front. Comput. Sci. 5:1264713.
doi: 10.3389/fcomp.2023.1264713

Emotion recognition with multi-modal peripheral physiological signals

Jennifer Gohumpu[1,2], Mengru Xue[1,2]* and Yanchi Bao[1,2]

[1]Ningbo Innovation Center, Zhejiang University, Ningbo, China, [2]Zhejiang University, Hangzhou, China

Introduction: Healthcare wearables allow researchers to develop various system approaches that recognize and understand the human emotional experience. Previous research has indicated that machine learning classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Tree (DT), can improve the accuracy of physiological signal analysis and emotion recognition. However, various emotions can have distinct effects on physiological signal alterations. Therefore, solely relying on a single type of physiological signal analysis is insufficient for accurately recognizing and understanding human emotional experiences.

Methods: Research on multi-modal emotion recognition systems (ERS) has commonly gathered physiological signals using expensive devices, which required participants to remain in fixed positions in the lab setting. This limitation restricts the potential for generalizing the ERS technology for peripheral use in daily life. Therefore, considering the convenience of data collection from everyday devices, we propose a multi-modal physiological signals-based ERS based on peripheral signals, utilizing the DEAP database. The physiological signals selected for analysis include photoplethysmography (PPG), galvanic skin response (GSR), and skin temperature (SKT). Signal features were extracted using the "Toolbox for Emotional Feature Extraction from Physiological Signals" (TEAP) library and further analyzed with three classifiers: SVM, KNN, and DT.

Results: The results showed improved accuracy in the proposed system compared to a single-modal ERS application, which also outperformed current DEAP multi-modal ERS applications.

Discussion: This study sheds light on the potential of combining multi-modal peripheral physiological signals in ERS for ubiquitous applications in daily life, conveniently captured using smart devices.

KEYWORDS

multi-modalities, physiological signals, emotion recognition, machine learning, ubiquitous and mobile computing system

1 Introduction

Over the past decade, extensive research efforts have been dedicated to the development and enhancement of human ERS. Emotions play a central role in the human experience, exerting a profound influence on both physiological and psychological states. This influence holds promise for diverse applications, including Internet of Things (IoT) devices (Abdallah et al., 2020; Fodor et al., 2023), safe driving practices (Ma et al., 2021), software engineering (Fritz et al., 2014), and beyond. Research in this domain has sought to capture and interpret emotional states through a variety of signals.

One pivotal aspect of this research involves categorizing human emotion detection methods into two main groups: physical signals and physiological signals. The study of

physical signals encompasses various aspects such as facial expressions (Kuruvayil and Palaniswamy, 2022), speech (Akçay and Oğuz, 2020), text (Guo, 2022), and gestures (Zhang et al., 2020). These signals have received extensive research attention over the years due to their ease of collection and measurement. In contrast, physiological signals serve as indicators of individuals' internal states and offer a significant advantage in emotion detection due to their resistance to manipulation. This highlights the challenge of accurately determining true emotions solely based on physical signals, as individuals can intentionally conceal their emotional feelings (Ismail et al., 2022).

Empirical studies (Koelstra et al., 2012; Soleymani et al., 2012; Abadi et al., 2015; Katsigiannis and Ramzan, 2018; Schmidt et al., 2018) have extensively employed various physiological signals, including electroencephalogram (EEG), electromyography (EMG), electrocardiogram (ECG), PPG, GSR, and SKT, in the development of ERS. Among these physiological signals, the analysis of brain activity holds particular significance. Changes in brain waves provide valuable insights into real-time reactions during neurological activity, reflecting genuine human responses (Regan et al., 2010; Lan et al., 2016). While EEG stands out as the most accurate and reliable physiological signal for recognizing emotions (Qiu et al., 2018), it does come with certain drawbacks, such as the requirement of high-end devices that must be worn on the head. These drawbacks may inconvenience subjects and potentially hinder their participation in experimental activities.

While EEG has been a valuable tool in emotion recognition research, it does come with certain drawbacks that can hinder subjects' participation in experimental activities. In addition to EEG, another group of physiological signals closely correlated with human emotions are cardiac-related signals, specifically the ECG and PPG. ECG measures the electrical activity within the heart, while PPG captures changes in blood volume during heart activity.

In addition to signal selection, the methodology employed for handling single and multi-modal physiological signals in ERS is a critical factor in understanding and recognizing human emotions. For instance, a study conducted by Zhang et al. (2021) proposes an ERS based on multi-modal physiological signals using the DEAP dataset and DECAF dataset. They employ deep learning techniques to combine different modalities, such as EEG, EMG, GSR, and RES (Respiration) from the DEAP dataset, and MEG (Magnetoencephalography), EMG (Electromyography), EOG (Electrooculography), and ECG signals from the DECAF dataset. Similarly, Yan et al. (2022) conducted a study using multichannel physiological signals from the WESAD dataset, which included ECG, GSR, EMG, and BVP signals. Both of their results indicate that multi-modal ERS exhibits a high potential for achieving superior performance in contrast to single-modal ERS.

Furthermore, as technology advances, the use of advanced wearable devices equipped with unobtrusive sensors, such as smartwatches, offers researchers greater flexibility and convenience in their approach. Notably, previous work by Wang et al. (2020) proposed an adaptive Emotion Recognition System (ERS) using a sensor-enriched wearable smartwatch to explore both physiological and behavioral data across various daily activity scenes. Similarly, Quiroz et al. (2018) introduced an ERS system utilizing a smartwatch and a heart rate monitor strap to analyze human emotional states and behavioral responses. In both studies, the utilization of smartwatches demonstrated high accuracy and convincing performance in developing ERS.

Therefore, through this research, we aim to contribute to the development of an ERS model based on signals commonly found in most smartwatches, specifically the PPG, SKT, and GSR signals. We analyze the publicly available DEAP dataset, which includes these three signals, and extract relevant features. Subsequently, we apply classification to the ERS model using simple machine learning algorithms such as SVM, KNN, and DT. This choice of algorithms is made to minimize the system's footprint and processing power, especially for devices with limited resources. We evaluate the system's performance based on average accuracy and the F1 score.

The subsequent section presents a comprehensive review of previous research on ERS. Section 3 introduces the DEAP dataset, and the methodology for feature extraction and classification is explained in detail. The analyzed results are presented in Section 4. In Section 5, the discussion of findings is briefly discussed, while Section 6 serves as the conclusion of the paper.

2 Related work

In recent years, there has been a surge of research focused on understanding and recognizing user emotions by leveraging the advancements in augmented reality, virtual reality, and human-computer interaction technologies. Numerous studies have been conducted in various areas, including the development of emotion models, data collection methods, and peripheral signal-based ERS. These efforts collectively contribute to a deeper understanding of user emotions and pave the way for more effective and comprehensive emotion recognition approaches.

2.1 Emotion model

To accurately recognize emotions, it is crucial to have a well-defined and quantifiable concept of emotion. Over the past decades, psychologists from various disciplines have made attempts to define emotion. However, there is still no universally acknowledged theory of emotion. In most emotion recognition research, two common approaches have been used to define the emotion model: the discrete emotion model and the multi-dimensional emotion space model (Picard, 2000).

In the discrete emotion model, human emotional experiences are described using words rather than quantitative analysis. This approach presents limitations in analyzing complex emotions, as individuals from different backgrounds may have different emotional sensitivities (Shu et al., 2018). Therefore, it is essential for an emotion model to have a quantitative standard, especially when applied in conjunction with machine learning analysis.

To address this limitation, researchers have endeavored to develop a multi-dimensional emotion space model, which enables the measurement of emotions along different dimensions and facilitates easy comparison of varying intensities of emotional experiences (Bota et al., 2019). One notable example is the two-dimensional emotional model proposed by Lang (1995), which classifies emotions based on valence and arousal. The valence

dimension axis categorizes human emotional experiences from negative (unpleasant) to positive (pleasant), while the arousal dimension axis ranges from low (passive) to high (active). Building upon this framework, Mehrabian (1997) extended the model to a three-dimensional emotion model by introducing an additional dimension axis known as dominance. This additional axis aids in identifying emotions such as fear and anger more effectively.

2.2 Emotion recognition data collection methods

Data collection methods for emotion recognition have employed various technologies. Miranda et al. (2014) conducted anxiety detection research by combining spontaneous eye-blink rate and heart rate signals using wireless wearable products, namely the Google Glass and Zephyr HxM Bluetooth band. Similarly, Koelstra et al. (2012) utilized a high-end technology device, the Biosemi Active Two System, along with a recording PC to collect a multi-modal dataset comprising EEG signals, peripheral physiological signals, and facial video signals.

In addition to these approaches, researchers have extensively utilized open-source datasets such as the DEAP dataset (Koelstra et al., 2012), MAHNOB-HCI dataset (Soleymani et al., 2012), WESAD dataset (Schmidt et al., 2018), DECAF dataset (Abadi et al., 2015), DREAMER dataset (Katsigiannis and Ramzan, 2018), and SEED dataset (Zheng and Lu, 2015) for various types of research in the field of emotion recognition. These databases provide valuable resources for studying and analyzing emotional data.

2.3 Single-modality and multi-modalities ERS

Peripheral signals play a significant role in emotion recognition research. These signals are derived from physiological processes and provide valuable information about an individual's emotional state. Researchers have explored both single-modal and multi-modal peripheral signal-based emotion recognition systems.

In single-modal peripheral signal-based emotion recognition, researchers analyze a specific type of peripheral signal in isolation to recognize human emotions. For instance, Susanto et al. (2020) introduced an ERS based on deep hybrid neural networks that utilized GSR signals. Their study demonstrated the effectiveness of GSR in recognizing human emotions. Similarly, Zhu et al. (2019) conducted a review focused on the application of heart rate variability (HRV) in human emotion recognition. They highlighted the potential of HRV-related approaches and their prospects for broader applications in emotion recognition.

Multi-modal peripheral signal-based emotion recognition involves integrating various types of peripheral signals to gain a more comprehensive understanding of emotions. This approach combines signals such as EEG, GSR, PPG, EMG, ECG, and respiratory signals. The fusion of information from different modalities aims to improve the accuracy and robustness of emotion recognition systems. For example, Stajic et al. (2021) conducted emotion recognition research utilizing the DEAP database. They employed multiple physiological signals and analyzed them using three different machine learning algorithms: SVM, boosting algorithms, and artificial neural networks. By integrating these signals and employing advanced algorithms, their study aimed to enhance the recognition of emotions. Additionally, Lima et al. (2020) presented their research on mental stress prediction. They combined PPG and GSR signals, which were collected using their own prototype equipped with PPG and GSR sensors. Their findings indicated that simultaneously considering PPG and GSR baseline features achieved an accuracy of 77% in predicting mental stress.

These studies exemplify the application of multi-modal peripheral signals in emotion recognition and mental stress prediction. By combining different types of signals and utilizing advanced analytical techniques, researchers strive to improve the understanding and detection of emotions.

3 Method

The ERS models in this study follow the flowchart presented in Figure 1. The input data used for the models is obtained from the publicly available DEAP dataset. The data undergoes a comprehensive process, starting with feature extraction and culminating in the classification stage.

During the feature extraction stage, specific features are extracted from each physiological signal. These features (Soleymani et al., 2017) capture relevant information related to the emotional state. All the extracted features are then combined and divided into two sets: training data and testing data. The training data is utilized to train the classifiers, enabling them to learn and make predictions based on the provided features.

Three classifiers, namely SVM, KNN, and DT, are employed in this study. Each classifier uses the training data to build a model that can classify emotional states based on the extracted features. To enhance the robustness of the classification performance, we incorporated GridSearch Cross Validation and KFold Cross Validation techniques in our study. These approaches help optimize the model's hyperparameters and ensure that the performance evaluation is reliable and generalized across different subsets of the dataset.

To evaluate the performance of the models, the testing data is used. The models are applied to the testing data, and their accuracy and effectiveness in recognizing emotions are assessed. This evaluation provides insights into the performance and suitability of each classifier in the context of the ERS models developed in this study.

3.1 Public dataset—DEAP dataset

For this research, the PPG, GSR, and SKT signals from the publicly available DEAP dataset (Koelstra et al., 2012) were utilized. The DEAP dataset, developed by Koelstra et al., encompasses a wide range of physiological signals, including EEG, EOG, EMG, GSR, Respiration, PPG, and SKT. It consists of recordings from 32 subjects who were exposed to a set of 40 audio videos. The subjects provided ratings for each video in terms of arousal, valence, dominance, and familiarity levels. However, for the purpose of this

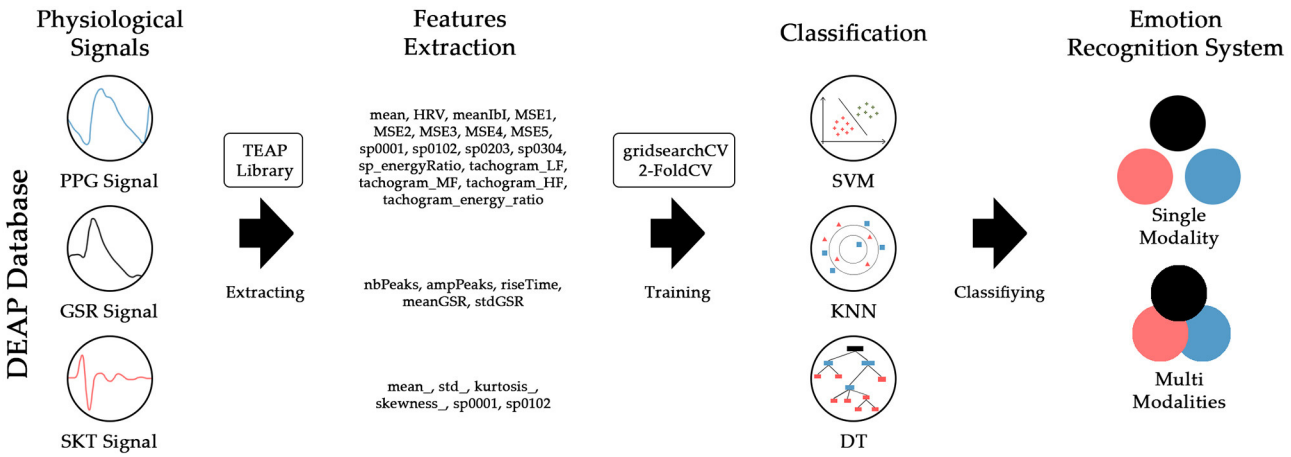

FIGURE 1
Workflow of emotion recognition model.

specific study, the focus was solely on the GSR, PPG, and SKT signals, narrowing down the scope of analysis to these specific modalities.

3.2 Feature extraction

The GSR, PPG, and SKT signals underwent pre-processing and feature extraction using PyTEAP, a Python implementation of the TEAP library. TEAP, initially introduced by Soleymani et al. (2017), is a comprehensive toolbox for analyzing various physiological signals, including EEG, GSR, PPG, and EMG. Over time, TEAP has expanded its capabilities to support additional signals such as BVP, ECG, HST, and RES. Additionally, TEAP offers pre-processing functionalities, including the application of low-pass filters to raw data.

Using PyTEAP, a total of seventeen features were extracted from the PPG signal, including mean, inter-beat interval (IBI), HRV, multiscale entropy, power spectral density, tachogram power, and the energy ratio between power spectral density and tachogram power. For the GSR signal, five features were extracted, which include the number of GSR peaks per second, average peak amplitude, average peak rise time, average GSR value, and GSR variance. Additionally, six features were extracted from the SKT signal, consisting of mean temperature, standard deviation of temperature, kurtosis of temperature, skewness of temperature, spectral power in the 0–0.1 Hz range, and spectral power in the 0.1–0.2 Hz range.

3.3 Classification

Optimizing these parameters is crucial to prevent the underfitting or overfitting of the DT model. By carefully selecting the appropriate splitting criterion, maximum depth, and minimum samples per leaf, the DT algorithm can achieve better performance and effectively capture patterns and decision rules from the data.

The selection of these three machine learning algorithms for this study was based on their specific advantages. Firstly, all three algorithms, SVM, KNN, and DT, possess versatility in handling both classification and regression problems, making them suitable for a wide range of tasks. Secondly, these algorithms demonstrate scalability, allowing them to efficiently handle smaller datasets such as the DEAP dataset used in this study. This scalability ensures that computational resources are used effectively and results can be obtained within a reasonable timeframe. Thirdly, these algorithms have relatively few hyperparameters, simplifying the parameter-tuning process. This advantage allows for greater control over the model's performance and generalization. With fewer hyperparameters to optimize, it becomes easier to find the best parameter settings that fit the model.

Additionally, in this study, the train and test data were split into an 80:20 ratio, ensuring a proper evaluation of the model's performance. Furthermore, the application of gridSearchCV in the classifier was used to systematically search for the best hyperparameter combination, optimizing the model's performance based on the given data.

By considering these advantages and employing appropriate methodologies for data splitting and hyperparameter tuning, the study aims to effectively apply these machine learning algorithms to achieve accurate and robust emotion recognition results.

3.4 Parameter tuning

The multi-modal ERS test and training sets were analyzed using the DEAP dataset and evaluated with grid-search cross-validation (GridSearchCV). The data were divided into training and testing sets, with a training size of 0.2, and the optimal hyperparameters were determined through the grid-search process.

To ensure the robustness of the classification performance, a KFold Cross-Validation technique with a specified number of folds, in this case, 2, was applied. This approach allows for the evaluation of the model's performance on different subsets of the data.

3.5 Performance metrics

Performance metrics are quantitative measurements that offer valuable insights into the effectiveness and accuracy of machine learning models. In this study, two performance metrics, namely accuracy and F1 score, were utilized to evaluate the models.

Accuracy can be expressed below:

$$Acc = \frac{(Truepositive + TrueNegative)}{(Truepositive + Truenegative + Falsepositive + Falsenegative)} \quad (1)$$

It reflects the model's capability to correctly classify emotions, providing an overall measure of its correctness in predicting the emotional state. It represents the ratio of correctly classified instances to the total number of instances.

On the other hand, the F1 score is a metric that considers both precision and recall.

$$F1\ score = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (2)$$

It is the harmonic mean of precision and recall, providing a balanced assessment of the model's accuracy. The F1 score takes into account both the model's ability to minimize false positives (precision) and its ability to capture true positives (recall).

In this study, the average accuracy and F1 score were calculated as metrics for assessing the performance of the multi-modal ERS. By averaging these metrics, a more comprehensive and robust evaluation of the model's accuracy and overall performance can be obtained.

$$Acc_{k-fold} = \frac{\sum_{i=1}^{k} Acc_i}{k} \quad (3)$$

$$F1\ score_{k-fold} = \frac{\sum_{i=1}^{k} F1\ score_i}{k} \quad (4)$$

4 Result

By conducting experiments with both single-modality and multi-modalities signal combinations, this study seeks to compare the performance and effectiveness of different approaches in recognizing and understanding emotions based on the PPG, GSR, and SKT signals from the DEAP database. Three classification algorithms are applied for the PPG, GSR, and SKT signals and analyzed using two different types of signal combinations: single-modality and multi-modalities.

In the single-modality approach, each physiological signal (PPG, GSR, SKT) is analyzed independently, focusing on the unique information contained in each signal. This allows for an understanding of the individual contributions of these signals in emotion recognition. In contrast, the multi-modalities approach integrates multiple physiological signals (PPG, GSR, SKT) to create a more comprehensive representation of the emotional state. By combining information from multiple modalities, the aim is to capture a richer and more accurate understanding of human emotions.

The classifier performance results for both single-modality and multi-modalities are shown in Table 1 and the next section presents a detailed comparison of the classification performance achieved by individual modality signals. Subsequently, a comprehensive analysis is conducted to compare the performance of single-modality signals with that of multi-modalities. Finally, a comparison is made with prior works focusing on multi-modal ERS to highlight the advancements and contributions of our study in this context.

4.1 Comparisons among single modality and dual modality

In the single-modality approach, SVM excelled in recognizing arousal, valence, and liking, while DT performed better for dominance. Among the three modalities, HRV signal classification with SVM achieved the highest accuracy: 64.5% for arousal and 65.4% for valence. For dominance, SKT signal with DT classifier reached the highest accuracy at 69.0% and HRV signals with DT classifier achieved an accuracy of 72.0% for liking.

These findings affirm the effectiveness of the SVM algorithm in single-modality emotion recognition and stress the importance of selecting the appropriate physiological signal for each emotional dimension. Specifically, the HRV signal is effective in recognizing arousal, valence, and liking emotions, while the SKT signal shows potential for dominance emotion.

In the dual-modality approach, overall accuracy improved compared to single-modality approaches, except for liking. The combination of HRV and SKT signals exhibited the highest performance with the SVM algorithm, surpassing other combinations like HRV+GSR and SKT+GSR. Notably, the HRV and SKT signal combination, with SVM, achieved accuracies of 64.7% for arousal, 65.6% for valence, 69.1% for dominance, and 71.4% for liking.

4.2 Comparisons with multi-modalities

Among the classifiers used in the multi-modalities approach, SVM exhibited the best accuracy performance compared to other classifiers, followed by DT. The multi-modalities model achieved accuracies of 66.0% for All-arousal, 66.0% for All-valence, 69.3% for All-dominance, and 71.1% for All-liking.

It is important to highlight that the multi-modalities approach demonstrated an overall performance improvement compared to the single-modality or dual-modality approaches, except for liking. By combining multiple physiological signals, the model benefitted from the complementary information provided by different modalities, resulting in enhanced accuracy in emotion recognition. Specifically, accuracy increased from 64.5% to 66.0% for arousal, from 65.4% to 66.0% for valence, and from 69.0% to 79.5% for dominance. However, the accuracy for liking slightly decreased from 72.0% to 71.1%.

Indeed, these findings emphasize the effectiveness of combining multi-modal physiological data with the SVM classifier, which

TABLE 1 The classifier performance results for both single-modality and multi-modalities.

Emotion		Arousal		Valence		Dominance		Liking	
Modalities	Classifier	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
HRV	SVM	**0.645**	**0.784**	**0.654**	**0.791**	0.681	0.810	**0.720**	0.837
	KNN	0.616	0.743	0.610	0.739	0.645	0.774	0.681	0.804
	DT	0.637	0.778	0.651	0.788	0.688	0.815	0.712	0.832
GSR	SVM	0.644	0.783	0.651	0.788	0.671	0.803	0.703	0.826
	KNN	0.602	0.739	0.587	0.715	0.623	0.751	0.679	0.805
	DT	0.631	0.773	0.640	0.780	0.680	0.809	0.701	**0.824**
SKT	SVM	0.634	0.783	0.646	0.785	0.684	0.812	0.709	0.830
	KNN	0.606	0.719	0.634	0.756	0.643	0.772	0.675	0.799
	DT	0.647	0.786	0.653	0.790	**0.690**	**0.816**	0.702	0.825
HRV + GSR	SVM	0.646	0.785	0.652	0.789	0.684	0.812	0.710	0.830
	KNN	0.607	0.730	0.610	0.735	0.652	0.780	0.682	0.807
	DT	0.648	0.787	0.651	0.788	0.686	0.814	0.711	0.831
HRV + SKT	SVM	**0.647**	**0.786**	**0.656**	**0.792**	**0.691**	**0.817**	**0.714**	**0.833**
	KNN	0.606	0.735	0612	0.743	0.661	0.787	0.684	0.807
	DT	0.647	0.786	0.653	0.790	0.686	0.814	0.712	0.832
SKT + GSR	SVM	0.638	0.779	0.646	0.785	0.679	0.809	0.713	0.833
	KNN	0.609	0.736	0.615	0.742	0.647	0.774	0.676	0.804
	DT	0.640	0.780	0.650	0.788	0.689	0.816	0.705	0.827
ALL	SVM	**0.660**	**0.795**	**0.660**	**0.795**	**0.693**	**0.813**	0.708	0.829
	KNN	0.619	0.743	0.615	0.738	0.655	0.783	0.685	0.809
	DT	0.643	0.783	0.653	0.790	0.691	0.817	**0.711**	**0.831**

Bold values indicate highest accuracy in single modality, double modalities and multi-modalities.

results in leveraging the combined information to achieve improved performance in emotion recognition.

4.3 Comparisons with prior work

Prior research investigating multi-modal ERS and utilizing the DEAP dataset consistently reveals superior performance compared to single-modality ERS. Koelstra et al. (2012) demonstrated that combining multiple modalities, which include EEG, peripheral physiological signals, and multimedia content analysis, resulted in higher performance when compared to using a single modality alone. In their study, a comprehensive comparison of F1 scores among single, double, and triple modalities combinations clearly indicated that the triple-modalities combination exhibited the most optimal performance across all emotions. These findings provide strong evidence for the effectiveness and advantage of leveraging multiple modalities in ERS for accurate and comprehensive emotion recognition.

Meanwhile, Zhang et al. (2021) conducted research by exploring different combinations of physiological signals, encompassing single, double, triple, and quadruple modalities combinations. Notably, their findings indicated that in the triple modalities, the combination of EEG, EMG, and GSR signals (EEG+EMG+GSR) demonstrated the highest performance. For the Arousal dimension, this combination achieved an accuracy of 59.0% and an F1 score of 66.2%, an accuracy of 59.5% and an F1 score of 63.1% in the Valence dimension (Table 2).

In comparison, our study utilizes a cost-effective multi-modal approach that outperforms Zhang et al.'s (2021) results. Importantly, our chosen signal combination is more accessible and practical than EEG and EMG, which require expensive medical-grade devices. This advantage enhances the feasibility of implementing ERS in real-world scenarios.

5 Discussion

In previous studies, extensive research has been conducted in the field of multi-modal ERS. For instance, Qiu et al. (2018) proposed an ERS model that leveraged deep learning techniques, incorporating EEG signals and eye movement data from the DEAP database. Similarly, Abadi et al. (2015) introduced the DECAF dataset, which included a diverse range of modalities such as MEG signals, physiological signals, face videos, and multimedia signals. Both of these studies employed advanced techniques to develop complex multi-modal ERS, resulting in improved accuracy for emotion classification. However, their studies required

TABLE 2 Comparison with prior work.

Method	Dataset	Modalities	Classifier	Arousal		Valence	
				Accuracy	F1 score	Accuracy	F1 score
Zhang et al. (2021)	DEAP	EEG + EMG + GSR	RDFKM	0.590	0.662	0.595	0.631
Our study	DEAP	HRV + GSR + SKT	SVM	0.660	0.795	0.660	0.795

the use of highly graded medical devices to measure these signals, and conducting experiments with such devices demanded controlled environments. These requirements imposed constraints on the experimental design and might have limited the practical applicability of their findings to more general situations. Therefore, in this study, we sought to explore the potential of multi-modal ERS by leveraging PPG, GSR, and SKT signals commonly obtained from wearable devices.

By utilizing these accessible and practical physiological signals, we aimed to develop an ERS that can be applied more widely in real-world scenarios. The TEAP library was employed to extract seventeen features from the PPG signal, five features from the GSR signal, and six features from the SKT signal. We utilized three machine learning algorithms, namely SVM, KNN, and DT, to analyze the DEAP dataset and train the models to recognize arousal, valence, dominance, and liking emotional states.

Upon comparing the performance of single-modality and multi-modalities ERS in our study, we observed improved accuracy in the multi-modalities approach. Specifically, for arousal, the accuracy increased from 63.4 to 63.9%, for valence, it improved from 62.8 to 63.1%, and for dominance, it rose from 66.6 to 66.9%. However, the accuracy for liking remained consistent at 69.5%. These results highlight the potential benefits of utilizing multi-modal physiological signals in Emotion Recognition Systems, as it leads to enhanced accuracy in recognizing various emotional states. While, among the machine learning algorithms, the SVM classifier exhibited the highest accuracy performance in the multi-modalities ERS. It proved to be the most suitable algorithm for leveraging the combined information from multiple modalities. This result aligns with the findings reported by Verma and Tiwary (2014) in their study, where they also observed that the SVM classifier achieved the highest accuracy in their proposed multi-modalities Emotion Recognition System utilizing the DEAP database. The consistent performance of the SVM classifier across different multi-modal approaches highlights its effectiveness in handling complex emotional data from various physiological signals.

Nevertheless, it is important to note that the accuracy performance for all emotional states, especially for dimensional emotions like dominance and liking, still fell short of being satisfactory. In the study conducted by Bălan et al. (2019), they aimed to remap the VAD (Valence-Arousal-Dominance) emotion dimensional space to the six basic emotions, namely anger, joy, surprise, disgust, fear, and sadness. They developed an ERS using the DEAP database and explored various probabilities and approaches for analyzing and featuring these emotions. Their study provided valuable insights into the complexities of emotion recognition and offered diverse methods for understanding and categorizing emotions based on the VAD dimensions.

Our study highlights the significant potential and reliability of the multi-modalities ERS model, with the SVM classifier demonstrating the highest accuracy performance among the tested algorithms. These findings are consistent with prior research (Koelstra et al., 2012; Verma and Tiwary, 2014; Liu et al., 2019; Zhang et al., 2021) in the field of multi-modalities ERS, further validating the effectiveness of combining multiple physiological signals for emotion recognition. Additionally, insights from studies conducted by Shu et al. (2018) and Bălan et al. (2019) shed light on the important features and relationships between the six basic emotions and physiological signals, contributing valuable knowledge for developing more efficient and high-performance ERS models.

5.1 Limitations and future works

Our study has identified several potential limitations. The small dataset size constrained our options for training and validating the classifier algorithm, particularly when employing deep learning techniques that often require larger volumes of data for optimal performance. The integrity of data signals collected from smartwatches or smart devices (Quiroz et al., 2018; Ismail et al., 2022) posed another limitation due to noise resulting from body movement. Future research should prioritize the development of smart devices with improved noise resilience and explore data processing methods that can effectively handle noise while preserving the signal's integrity.

To address these limitations, several avenues for future research can be considered. Firstly, exploring deep learning techniques holds promise for enhancing the accuracy of the ERS model. Secondly, future studies should focus on building customized datasets using smart devices to mitigate the limitations of the DEAP dataset. Lastly, developing pre-processing techniques and noise elimination algorithms specifically tailored for commercially-used smart devices would improve the overall quality of collected data.

By addressing these limitations, significant progress can be made in the field of emotion recognition, leading to more robust and accurate results in real-world applications. The exploration of larger datasets, advancements in noise-resistant smart devices, and the utilization of effective data processing techniques will collectively contribute to the advancement and reliability of research in this area.

6 Conclusion

This paper presents a multi-modal emotion recognition system utilizing peripheral PPG, EDA, and SKT signals from the DEAP dataset. The signals underwent feature extraction using the TEAP library and were analyzed using three machine-learning algorithms: SVM, KNN, and DT. The results of our study demonstrate improvements in accuracy performance in the multi-modal ERS compared to the single-modal approach. This highlights the viability of constructing an ERS model using this combination of multiple modalities. Additionally, the SVM classifier exhibited superior performance in accurately classifying emotions. We also discussed several areas that warrant further attention and improvement. We contribute to the field of emotion recognition and develop more robust and accurate models for real-world applications.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/download.html.

Author contributions

JG: Writing—original draft, Data curation, Formal analysis, Investigation, Methodology, Validation. MX: Conceptualization, Funding acquisition, Investigation, Project administration, Writing—review & editing. YB: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We acknowledge funding support from Ningbo Innovation Center, Zhejiang University, Ningbo.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., and Sebe, N. (2015). DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Trans. Affect. Comput.* 6, 209–222. doi: 10.1109/TAFFC.2015.2392932

Abdallah, A. S., Elliott, L. J., and Donley, D. (2020). "Toward smart internet of things (IoT) devices: exploring the regions of interest for recognition of facial expressions using eye-gaze tracking," in *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (London, ON), 1–4.

Akçay, M. B. and Oğuz, K. (2020). Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Commun.* 116, 56–76. doi: 10.1016/j.specom.2019.12.001

Bălan, O., Moise, G., Petrescu, L., Moldoveanu, A., Leordeanu, M., and Moldoveanu, F. (2019). Emotion classification based on biophysical signals and machine learning techniques. *Symmetry* 12, 21. doi: 10.3390/sym12010021

Bota, P. J., Wang, C., Fred, A. L. N., and Silva, H. P. D. (2019). A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access* 7, 140990–141020. doi: 10.1109/ACCESS.2019.2944001

Fodor, K., Balogh, Z., and Molnar, G. (2023). "Real-time emotion recognition in smart homes," in *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (Timisoara), 71–76.

Fritz, T., Begel, A., Muller, S. C., Yigit-Elliott, S., and Zuger, M. (2014). "Using psycho-physiological measures to assess task difficulty in software development," in *ICSE 2014: Proceedings of the 36th International Conference on Software Engineering* (New York, NY: IEEE Computer Society), 402–413.

Guo, J. (2022). Deep learning approach to text analysis for human emotion detection from big data. *J. Intell. Syst.* 31, 113–126. doi: 10.1515/jisys-2022-0001

Ismail, S. N. M. S., Nor, N. A., and Ibrahim, S. Z. (2022). A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG). *J. King Saud Univ. Comput. Inform. Sci.* 34, 3539–3558. doi: 10.1016/j.jksuci.2022.04.012

Katsigiannis, S., and Ramzan, N. (2018). Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* 22, 98–107. doi: 10.1109/JBHI.2017.2688239

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Kuruvayil, S., and Palaniswamy, S. (2022). Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. *J. King Saud Univ. Comput. Inform. Sci.* 34, 7271–7282. doi: 10.1016/j.jksuci.2021.06.012

Lan, Z., Sourina, O., Wang, L., and Liu, Y. (2016). Real-time EEG-based emotion monitoring using stable features. *Visual Comput.* 32, 347–358. doi: 10.1007/s00371-015-1183-y

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *Am. Psychol.* 50, 372–385.

Lima, R., Osario, D., and Gamboa, H. (2020). "Heart rate variability and electrodermal activity biosignal processing: predicting the autonomous nervous system response in mental stress," in *CCIS* (Cham: Springer), 328–351.

Liu, W., Qiu, J.-L., Zheng, W.-L., and Lu, B.-L. (2019). Multimodal emotion recognition using deep canonical correlation analysis. *arXiv preprint arXiv:1908.05349*. doi: 10.1007/978-3-030-04221-9_20

Ma, Z., Ma, F., Sun, B., and Li, S. (2021). "Hybrid mutimodal fusion for dimensional emotion recognition," in *MM '21: ACM Multimedia Conference* (New York, NY: Association for Computing Machinery, Inc.), 29–36.

Mehrabian, A. (1997). Comparison of the pad and panas as models for describing emotions and for differentiating anxiety from depression. *J. Psychopathol. Behav. Assess.* 19, 331–357.

Miranda, D., Calderon, M., and Favela, J. (2014). "Anxiety detection using wearable monitoring," in *MexIHC '14: Proceedings of the 5th Mexican Conference on Human-Computer Interaction* (New York, NY: ACM), 34–41.

Picard, R. W. (2000). *Emotions Are Physical and Cognitive*. The MIT Press. doi: 10.7551/mitpress/1140.003.0004

Qiu, J. L., Liu, W., and Lu, B. L. (2018). "Multi-view emotion recognition using deep canonical correlation analysis," in *LNCS* (Cham: Springer Verlag), 221–231.

Quiroz, J. C., Geangu, E., and Yong, M. H. (2018). Emotion recognition using smart watch sensor data: mixed-design study. *JMIR Mental Health* 5, e10153. doi: 10.2196/10153

Regan, S. O., Faul, S., and Marnane, W. (2010). Automatic detection of EEG artefacts arising from head movements. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2010, 6353–6356. doi: 10.1109/IEMBS.2010.5627282

Schmidt, P., Reiss, A., Duerichen, R., and Laerhoven, K. V. (2018). "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *ICMI '18: Proceedings of the 20th ACM International Conference on Multimodal Interaction* (New York, NY: Association for Computing Machinery, Inc.), 400–408.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors* 18, 2074. doi: 10.3390/s18072074

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Soleymani, M., Villaro-Dixon, F., Pun, T., and Chanel, G. (2017). Toolbox for emotional feature extraction from physiological signals (teap). *Front. ICT* 4, 1. doi: 10.3389/fict.2017.00001

Stajic, T., Jovanovic, J., Jovanovic, N., and Jankovic, M. M. (2021). *Emotion Recognition Based on Deap Database Physiological Signals*. Belgrade: Institute of Electrical and Electronics Engineers Inc.

Susanto, I. Y., Pan, T. Y., Chen, C. W., Hu, M. C., and Cheng, W. H. (2020). "Emotion recognition from galvanic skin response signal based on deep hybrid neural networks," in *ICMR '20: Proceedings of the 2020 International Conference on Multimedia Retrieval* (New York, NY: Association for Computing Machinery, Inc.), 341–345.

Verma, G. K., and Tiwary, U. S. (2014). Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. *Neuroimage* 102, 162–172. doi: 10.1016/j.neuroimage.2013.11.007

Wang, Z., Yu, Z., Zhao, B., Guo, B., Chen, C., and Yu, Z. (2020). Emotionsense: an adaptive emotion recognition system based on wearable smart devices. *ACM Trans. Comput. Healthcare* 1, 1–17. doi: 10.1145/3384394

Yan, M. S., Deng, Z., He, B. W., Zou, C. S., Wu, J., and Zhu, Z. J. (2022). Emotion classification with multichannel physiological signals using hybrid feature and adaptive decision fusion. *Biomed. Signal Process. Control* 71, 103235. doi: 10.1016/j.bspc.2021.103235

Zhang, J., Yin, Z., Chen, P., and Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review. *Inform. Fusion* 59, 103–126. doi: 10.1016/j.inffus.2020.01.011

Zhang, X., Liu, J., Shen, J., Li, S., Hou, K., Hu, B., et al. (2021). Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybernet.* 51, 4386–4399. doi: 10.1109/TCYB.2020.2987575

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zhu, J., Ji, L., and Liu, C. (2019). Heart rate variability monitoring for emotion and disorders of emotion. *Physiol. Meas.* 40, 064004. doi: 10.1088/1361-6579/ab1887