Electrodermal Activity for Emotion Recognition Using CNN and Bi-GRU Model

Lili Zhu, Petros Spachos, and Stefano Gregori
School of Engineering, University of Guelph, Guelph, ON, Canada

ICC 2023 - IEEE International Conference on Communications | 978-1-5386-7462-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICC45041.2023.10278603

***Abstract*—Several signals can be collected from wearable devices containing important physiological and psychological information. Understanding various physiological signals is significant for computers to recognize human emotional states. Electrodermal Activity (EDA), originating from the spontaneous activation of sweat glands in the skin, is closely related to mood, arousal, and attention and is the most widely used measurement in the physiological response system for emotional state detection. However, extracting valuable features from EDA signals and making accurate emotional classification predictions has always been challenging. With the continuous development of models with representation learning capabilities, the use of deep learning models to automatically learn physiological signal features and perform classification learning is promising. In order to improve the shortcomings of traditional emotion recognition methods, which require a deep understanding of physiological signals and artificial extraction of relevant features, this paper proposed a Recurrent Neural Network (RNN) -based method for automatic feature extraction from EDA's spectrograms. A Convolutional Neural Network (CNN) is used to learn the extracted features further and output the determined emotional state. The results show that the classification accuracy for arousal and valence has reached 83.4% and 81.2%, respectively, which is promising in extracting features automatically and tackling the emotional state classification problem.**

***Index Terms*—Human emotional states, Stress detection, Galvanic Skin Response (GSR), Deep learning, CNN, RNN, IoT, Multimodal methods.**

I. Introduction

Fast-paced work and study are a hallmark of modern life. In such a living environment, stress is something that everyone has to deal with regularly. A small amount of stress can help people focus and overcome obstacles. However, the effects of chronic stress on the body can have many serious consequences. It affects the mind and body in countless ways, increasing the risk of more harmful diseases like obesity or depression. One of the keys to properly handling stress is understanding how it arises and affects the body. If people can fully recognize the visible signs of stress, they can assess themselves or others more quickly and accurately to avoid physical or mental diseases.

Electrodermal activity (EDA), also known as Galvanic Skin Response (GSR), may be one of the most sensitive emotional feedbacks. EDA originates from the autonomous activation of sweat glands in the skin [1]. Sensors for EDA measure the skin's resistance and can detect slight variations in conductivity when people engage in different actions or experience emotions. For example, EDA sensors can detect changes in skin conductance that occur when people are active or stressed [2]. EDA sensors are becoming increasingly popular because they offer multiple benefits to users. These sensors use a small electrical current to detect changes in skin conductivity levels and then use this information to estimate the user's emotional response. Some typical applications for EDA sensors include gaming, health monitoring, and security purposes. That is why EDA sensors are often found in wearable technology products such as smartwatches and fitness trackers. They provide accurate monitoring of these response parameters without manually taking multiple measurements, especially when other physiological signal sensors are also integrated into the device, such as Photoplethysmography (PPG) and Electrocardiogram (ECG). These devices allow users to understand their physical or emotional reactions during gameplay or when interacting with sensitive medical data. The technology is used in many applications, from self-assessment to health monitoring and stress and anxiety management. It becomes an important tool that can be used to improve overall health.

The essence of emotion recognition by analyzing EDA signals is to select supervised machine learning or unsupervised machine learning classifiers and train the samples to learn the extracted representative features to achieve the goal of determining various emotional states [3]. Shallow classifiers are generally used in emotion recognition research, which requires manual feature selection and extraction. However, whether these features can reflect emotional information stably and accurately depends mainly on researchers' professional knowledge and long-term experience accumulation. In the research of emotion recognition based on EDA, it is arduous to use traditional classifiers for application. The main reason is that the boundaries of different emotions are relatively blurred. With the rapid development of machine learning in recent years, researchers have gradually applied new and efficient deep learning algorithms to EDA decoding and demonstrated their advantages over traditional machine learning. Deep learning is a technology that performs nonlinear transformation or representation learning on the input of a neural network with no less than two hidden layers, emphasizing end-to-end learning directly from the original data rather than learning starting from artificially extracted features. Currently, there are some multimodal EDA emotion databases for analyzing human emotional states, such as CLAS [4], WESAD [5], DEAP [6], Amigos [7], etc. These databases usually use the three dimensions of Valence, Arousal, and Dominance as the subject's emotional evaluation indicators, that is, by scoring the three dimensions to determine the subject's current emotional state.

This work aims to compare valence and arousal classification accuracy based on an automatic feature extraction method and a deep learning model with other algorithms based on the EDA signal in the Amigos dataset. The contribution of this study lies in the potential of adopting an improved Recurrent Neural Network for automatic EDA feature extraction and improving a deep learning model for analyzing EDA signals.

The rest of this paper is organized as follows: A brief discussion on the related work is in Section II. Section III introduces the materials and methodologies for this study, followed by Section IV which presents the results and discussions of the experiment. Finally, Section V concludes this work.

II. Related Works

With the help of EDA, emotional changes evoked by any content, product, or service can be tested, such as thought experiments, videos, images, sounds, smells, and other sensory stimuli. In [8], they experimented with discriminating work-related stress by monitoring the EDA when subjects were performing different arithmetic problems. When assessing consumer preferences in marketing, EDA can be measured to track products with high emotional arousal to consumers. In [9], they used EDA to examine whether arousal provoked by a cause-related product can predict the product's potential in the market depending on whether the product is hedonic or utilitarian. In [10], they conducted experiments to identify which stimuli can elicit students' emotions when programming and if EDA can reflect the students' feelings. In user experience research, monitoring EDA can provide insight into how users feel when interacting with new websites or software content. In [11], EDA was collected when users were online shopping, and the data analysis validated that EDA can predict users' satisfaction with online shopping.

EDA is adopted for emotional state recognition with different machine learning methods achieving classification accuracy between 64% to 96% depending on datasets [12]–[16]. In [17], they used EDA, ECG, and PPG data from the WESAD and CLAS datasets to examine the impact of unimodality and multimodality data on stress detection. A Stacking Ensemble Learning (SEL) model gained an accuracy of 86.4% based on only EDA, which outweighs other methods and signals.

Using deep learning models to extract EDA features and classify emotional states is also promising. In [18], a Convolutional Neural Network (CNN) was utilized for human emotion recognition with EDA in MAHNOB [19] and DEAP datasets. In [20], they used one-dimensional CNN-bidirectional LSTM (Long short-term memory) for low and high arousal classification based on EDA and achieved a 91.02% F1-score. Especially when research uses EDA and other modalities, deep learning architectures are more welcome than traditional machine learning models due to the complexity of data fusion. In [21], they proposed an algorithm called DRER (driver's real emotion recognizer) based on deep learning to detect drivers' emotional changes during driving with face image and EDA, and the accuracy is 86.8%.

The Amigos dataset used in this study has been widely used in research that has achieved remarkable results in classifying emotional states with their respective methods. In [22], they applied a CNN to extract features automatically from biosignals in Amigos, then fed the extracted features into machine learning models. The results show that the emotion recognition accuracy of EDA signals can achieve 71% and 75% for arousal and valence, respectively. In [23], they concatenated EDA's statistical features and spectrogram features by manual calculation and VGG-16 network-based method, respectively, then assessed uni-modality and multi-modality's performance for emotion recognition. EDA provided an accuracy of 82.74% for the arousal category. In [24], they extracted 621 EDA features from the time domain, frequency domain, and time-frequency domain. After the extraction, three feature selection methods, Joint Mutual Information (JMI), Conditional Mutual Information Maximization (CMIM), and Double Input Symmetrical Relevance (DISR), were applied to select the most significant features. Support Vector Machine (SVM) performed the classification and offered an average accuracy of 85.75% for arousal and 83.9% for valence. In the recent study of [25], they adopted an unsupervised representation learning method to reconstruct a set of the most significant features from the input.

III. Materials and Methods

In this study, we proposed an emotion recognition system combining RNN-based automatic representation learning methods and a CNN architecture classification model. The introduced framework is shown in Fig. 1.

A. Electrodermal Activity

The human body consists of about 30,000 sweat glands. By regulating the balance of negative and positive ions in the exudate, electrical current can flow more efficiently, resulting in detectable skin conductance changes. This change in skin conductance is called Electrodermal Activity (EDA), which means, EDA is an electrical manifestation of the sympathetic innervation of sweat glands [26]. Skin conductance is regulated solely by autonomic sympathetic activity to drive physical, cognitive, and emotional states as well as entirely subconscious cognitive levels. It is difficult for people to control skin conductivity consciously, so it is possible to test the actual mental state with EDA under unconscious behaviour without subjective cognitive state control.

EDA can be measured from skin conductance data because skin conductance is proportional to sweat secretion [27]. This makes skin conductivity values an ideal measure of sympathetic nervous system activation. The units of measurement for conductance are microSiemens (μS). EDA measurement includes tonic skin conductance level and phasic skin conductance response, as shown in Fig. 2. Skin conductivity level (SCL) refers to the average conductivity obtained from the EDA signal's tonic component and measures the EDA signal's slow and smooth change over the measurement time. The dramatic changes or spikes in EDA's phasic data associated

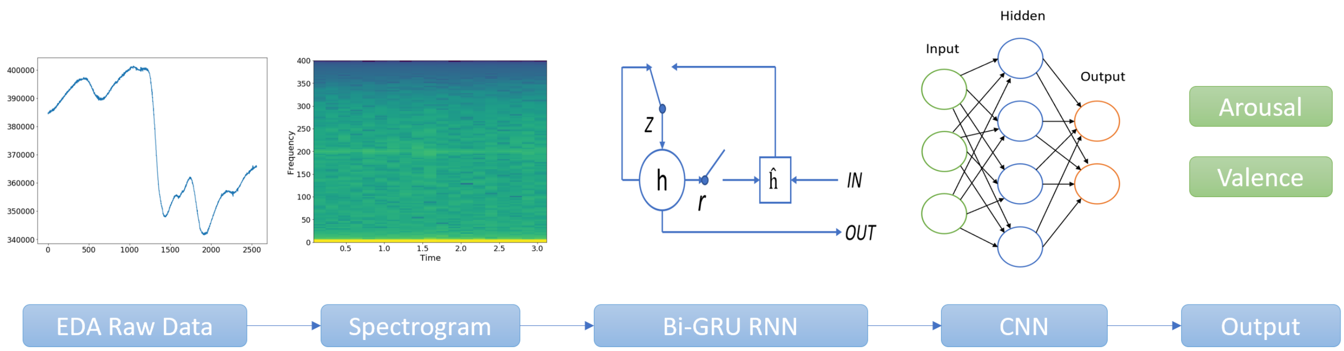

Fig. 1: The framework of the proposed emotion recognition system.

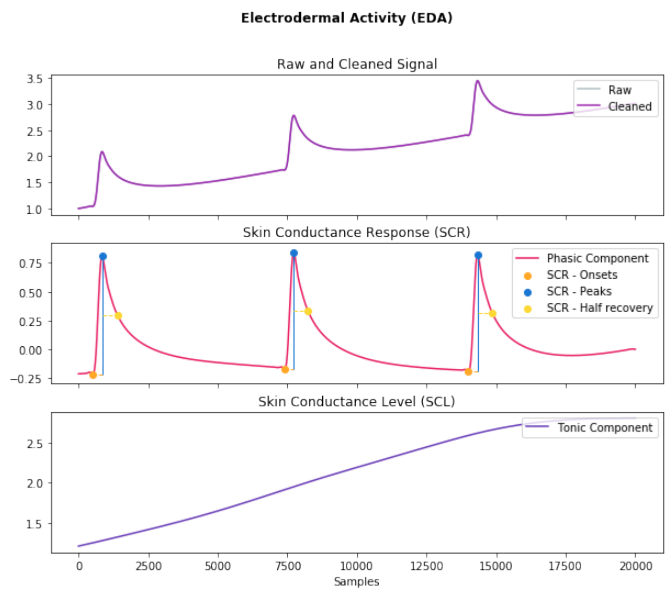

Fig. 2: An example of the tonic and phasic component of EDA and four parameters of SCR.

with the response to stimuli are known as skin conductance responses (SCR). SCR is sensitive to specific emotionally arousing stimulus events (event-related SCR, ER-SCR). The parameters that define SCR are latency, amplitude, rise time, and recovery time. These data spikes occur 1-5 seconds after stimuli. Nonspecific skin conductance responses (NSSCRs) refer to the number of SCRs over time that is not associated with any particular event. They are spontaneous fluctuations that occur at a rate of 1 to 3 times per minute.

B. Dataset

The dataset used in this study is the Amigos dataset [7]. By collecting EEG, ECG, and EDA signals, Amigos studies personality, affect, and mood. Unlike other databases, short and long videos are used in two configurations, one for a single viewer and one for groups of viewers. This dataset allows for a multimodal study of individuals' emotional responses to personality and mood and analyses how these responses are influenced by individual/group configuration and the videos' duration (short vs long). The physiological data were collected in two experimental settings. In the first experiment, 40 participants watched 16 short emotional videos while watching alone. In the second experiment, the same participants watched 4 long videos. Some played individually, and the rest in groups. The participants' EEG, ECG, and EDA were recorded in both settings using wearable sensors. Frontal, full-body and depth videos were also recorded. Participants were profiled for personality using the Big-Five personality traits model and for mood using the Positive Affect and Negative Affect Schedules. Participants' emotions during the experiments were annotated with two methods, self-assessment and external observers. Only the EDA data collected from the short video experiment was used in this study.

C. Data Preprocessing

Signals can be divided into stationary and non-stationary signals according to their time-varying characteristics. For non-stationary signals, the frequency characteristics of the signal at any time are important, and more is needed to analyze only the time domain or frequency domain. Typically, a limitation of time-domain and frequency-domain feature extraction methods is that significant features with high resolution are discarded since the information is computed from only one domain. For example, time-domain characteristics do not provide oscillation information. Meanwhile, in the case of frequency analysis, no detailed information on the variation of spectral signals is provided over time. However, these are significant concerns for the study of physiological signals. Therefore, it is essential to select and improve the time-frequency method, to combine the time domain and the frequency domain to describe and observe the time-frequency joint characteristics of the signal, and to form the time-frequency spectrum of the signal. In this study, for the EDA signals in the Amigos dataset, we adopted Short-time Fourier Transform (STFT) to transform the signals to spectrogram images after a moving window of 0.5 seconds to smooth the signal. The Hamming window function with a 1-second window length and the 1-second rate was applied to STFT.

D. Feature Extraction

After transforming the EDA signals to spectrogram images, an RNN-based method is ready for feature extraction.

RNN is a type of neural network with short-term memory capability, which is suitable for processing video, speech, text, and other time-series-related problems. In an RNN, neurons can receive information from other neurons and their own information to form a network structure with loops. If $h_t^{(l)}$ denotes to the hidden state of the l th layer at time t, $h_t^{(l)}$ is jointly determined by the hidden state of the l th layer at time $t-1$ and the hidden state of the $(l-1)$ th layer at time t:

$$h_t^{(l)} = f(U^l h_{t-1}^{(l)} + W^l h_t^{(l-1)} + b^l), \quad (1)$$

in which U^l and W^l are the weighing matrix, b^l is the bias, $h_t^{(0)} = x_t$ and x_t is the input at time t. When y_t is the output at time t, if the prediction y_t depends on the input x_{t-k} at time $t-k$, but the time interval k is relatively large, the problem of vanishing and exploding gradients is prone to occur. As a result, it is difficult for the RNN to learn such long-term input information. In this case, the long-term dependency problem arises when the current prediction needs to use more distant information. To solve these two problems, a gate mechanism is introduced to control the accumulation speed of information, including selectively adding new information and forgetting previously accumulated information. The common gated RNNs include the LSTM and Gated Recurrent Unit (GRU).

Compared with the classical RNN, they can effectively capture the semantic association between long sequences and alleviate the gradients vanishing and exploding problems. Unlike classical RNN, which has one hidden state, LSTM has two hidden states in the last recurrent layer. One is the long-term state C_t, and the other is the short-term state h_t. The long-term state is activated by the $Tanh$ function and then filtered through the output gate to obtain the short-term state. The short-term state is for connecting the fully connected layer with the input to obtain the output of the model. In LSTM, three gates f_t, i_t, and o_t have different functions. The forget gate f_t controls how much information needs to be forgotten in c_{t-1}; the input gate decides which information will be saved in the candidate state $\tilde{c}_t$ at the current moment; the output gate o_t determines the information in c_t needs to output to h_t. In contrast, the improvement of the GRU network to the LSTM network has two aspects: one is combining the forget gate and the input gate into one update gate z_t, and the other gate is called the reset gate r_t; the other one is that without introducing an additional internal state c, a linear dependency is directly introduced between the current state h_t and the historical state h_{t-1}. The reset gate r_t is used to control whether the calculation of the candidate state depends on the state $_{t-1}$ at the previous moment:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \quad (2)$$

where the candidate state is:

$$\tilde{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (3)$$

The update gate z_t is used to control the information the current state h_t needs to retain from h_{t-1} and the new information that will be received from the candidate state $\tilde{h}_{t-1}$:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (4)$$

then calculate the hidden state h_t:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t, \quad (5)$$

when $z_t = 0$ and $r_t = 1$, the hidden state h_t is the same as in Eq. 1, so that the GRU becomes the simple RNN:

$$h_t = tanh(W_h x_t + U_h h_{t-1} + b_h) \quad (6)$$

Moreover, in the rearranged equation for the hidden state h_t, there is a linear relationship between ht and ht-1, and there is a nonlinear relationship as well, which can alleviate the problem of vanishing and exploding gradients:

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ tanh(W_h x_t + U_h(r_t \circ h_{t-1}) + b_h) \quad (7)$$

Both LSTM and GRU have a further bidirectional form, which means that without changing their internal structure, the model is applied twice in different directions. The results obtained from the two directions are combined as the final output. However, compared with LSTM and GRU, since GRU has fewer parameters and a faster convergence speed, the actual training time is less than LSTM, which can significantly speed up the iterative process. The Bi-GRU network is illustrated in Fig. 3. In this study, the Bi-GRU algorithm is used to extract features from the spectrogram images of EDA.

E. Classification

The learned time-frequency features from the previous step are then fed into a CNN model for the final emotional state recognition. CNN is a neural network specialized for processing data with known, grid-like topology. It provides an end-to-end learning model. The parameters in the model can be trained by the traditional gradient descent method. The trained convolutional neural network can learn the features of the input data and complete the feature extraction and classification. CNNs are mainly composed of these layers: input layer, convolutional layer, ReLU layer, pooling layer, and fully connected layer. A complete CNN can be constructed by stacking these layers together. In practical applications, the convolution layer and the ReLU layer are often referred to as the convolution layer. After the convolution operation, the convolution layer must go through the activation function. Specifically, the convolutional and fully connected layers use activation functions and multiple parameters, such as the weights w and biases b of neurons, to transform inputs. Meanwhile, the ReLU and pooling layers perform fixed-function operations. The parameters in the convolutional and fully connected layers are trained with gradient descent so that the classification score computed by the CNN matches the label of each sample in the training set. Since the Bi-GRU networks already extract the time-frequency features, a

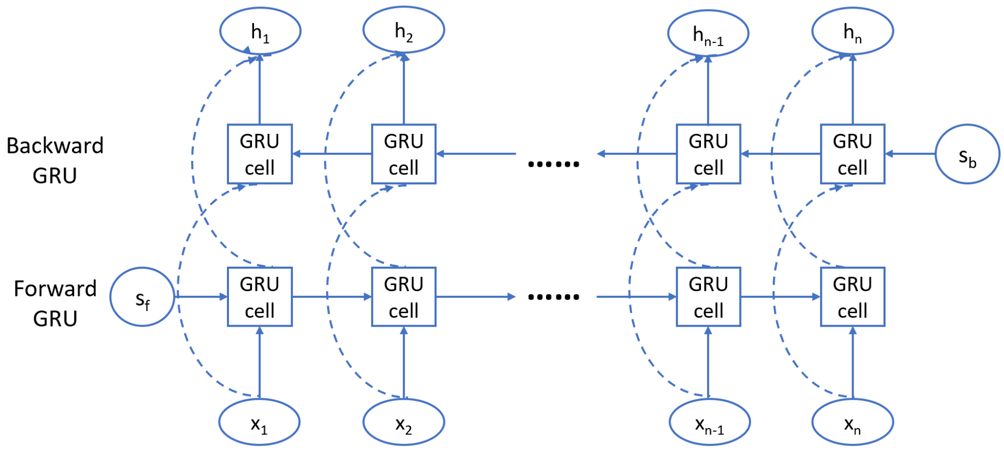

Fig. 3: The structure of Bi-GRU Networks.

Methodology	Features	Classifier	Accuracy	
			Arousal	Valence
Gjoresk et al. [28]	Hand-crafted	AdaBoost DT	0.56	–
Siddharth et al. [23]	Hand-crafted	ELM	0.81	0.81
Granado et al. [22]	Hand-crafted	DCNN	0.71	0.75
Ross et al. [25]	Convolutional Autoencoder	Random Forest	0.64±0.03	–
Proposed method	Bi-GRU RNN	CNN	0.83	0.81

Table I: Comparison of the methods used for emotion recognition based on EDA from the Amigos dataset.

lightweight CNN architecture is used for the classification in this study. For validation, the Leave-One-Subject-Out method is used in this study, meaning one subject will be left as testing set while the rest are training set.

IV. Results and Discussion

The classification results of the model are shown in Table I. Our proposed method achieved an accuracy of 0.83 and 0.81 for arousal and valence, respectively. We compared our performance with other research, which also uses the Amigos dataset and has a single modality of EDA as the physiological signal input. From the comparison, it can be noticed that our proposed method outweighs the other existing research. In [23], although having similar accuracy, the method of automatically extracting features we propose can avoid ignoring hand-crafted features that can be important due to bias. Also, this method of automatic feature extraction is more suitable for real-time systems.

In this study, feature extraction happened in the time-frequency domain, which differs from most other studies that extract features from the time or/and frequency domain. The raw EDA data is a limited discrete time series data composed of many sample points. Therefore, the raw EDA signal is data in the time domain. Time domain analysis focuses on the change of EDA amplitude over time and can quickly obtain the change in amplitude caused by an event (stimulus). The advantage of time-domain analysis is that its calculation is simple and fast. However, more than time-domain analysis is needed to fully reflect the information contained in EDA signals, such as the gradient component's detection of SCR [29]. However, the Fourier transform is limited since it only applies to stationary data, while EDA data are non-stationary data. In addition, frequency domain analysis cannot reflect frequency changes over time. Therefore, only time-domain or frequency-domain analysis cannot fully reflect the signal characteristics, resulting in time-frequency analysis being required. The results of this study demonstrate the feasibility of analyzing EDA signals using features from the time-frequency domain.

Time series is also an essential characteristic of physiological signals such as EDA. Traditional machine learning algorithms, or manually extracting features, miss important sequence information. On the contrary, RNN solves this problem. They are networks with loops that keep information passing forward. Therefore, the introduction of RNN to extract the features of the EDA signal can make the features of a sequence position mathematically related to the previous sequence. However, although RNNs can connect previous features to the current point, some relevant information cannot be successfully connected to the current point due to the considerable distance. Under this circumstance, LSTM and GRU are created based on RNN to learn about long-distance dependencies. LSTM is more often used in processing time-series signals. However, the structure of GRU models is less uncomplicated than standard LSTM models since GRU lacks forget gates and cell states. As a result, architectures with GRU have more potential to be integrated into applications for real-time physiological signal analysis. To our knowledge, this study is the first to investigate the feasibility of adopting GRU to extract features from EDA automatically. The promising results prove that the method is competitive. However, an issue of automatic feature extraction is that the signals should be clean since noises and motion artifacts could interfere with the correlation between features in time series signals. Especially in EDA signals, the sensors are usually placed on users' wrists, fingers, or palms so that motion artifacts are common to appear. Moreover, motion artifacts have similar features as SCR peaks caused by stimuli. Consequently, appropriate motion artifact detection and removal methods are essential before extracting the features.

V. Conclusion

In this study, we proposed a combination method for emotional state recognition. A bidirectional GRU-RNN network is used for feature extraction from EDA signals' spectrograms, while a CNN structure tackled the final classification task and achieved satisfactory accuracy. Compared with other research conducted on the Amigos dataset, the experiments and results show that the proposed method is promising for emotion recognition. For future studies, the performance of the proposed method will be further investigated on multi-modality signals, and the possibility of the fusion of different modalities at the feature level and decision level can be assessed. Furthermore, the potential of multi-dataset fusion can be explored.

References

[1] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the 24th australian computer-human interaction conference*, 2012, pp. 420–423.

[2] P. J. Lang, M. M. Bradley, B. N. Cuthbert *et al.*, *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention Gainesville, FL, 2005.

[3] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, p. 479, 2020.

[4] V. Markova, T. Ganchev, and K. Kalinkov, "Clas: A database for cognitive load, affect and stress recognition," in *2019 International Conference on Biomedical Innovations and Applications (BIA)*. IEEE, 2019, pp. 1–4.

[5] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

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

[7] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2018.

[8] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.

[9] J. Guerreiro, P. Rita, and D. Trigueiros, "Attention, emotions and cause-related marketing effectiveness," *European Journal of marketing*, 2015.

[10] J. Gorson, K. Cunningham, M. Worsley, and E. O'Rourke, "Using electrodermal activity measurements to understand student emotions while programming," in *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*, 2022, pp. 105–119.

[11] Y. Wu, Y. Liu, N. Su, S. Ma, and W. Ou, "Predicting online shopping search satisfaction and user behaviors with electrodermal activity," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 855–856.

[12] A. Greco, G. Valenza, J. Lázaro, J. M. Garzón-Rey, J. Aguiló, C. De-la Camara, R. Bailón, and E. P. Scilingo, "Acute stress state classification based on electrodermal activity modeling," *IEEE Transactions on Affective Computing*, 2021.

[13] O. N. Rahma, A. P. Putra, A. Rahmatillah, Y. S. K. A. Putri, N. D. Fajriaty, K. Ain, and R. Chai, "Electrodermal activity for measuring cognitive and emotional stress level," *Journal of Medical Signals and Sensors*, vol. 12, no. 2, p. 155, 2022.

[14] V. Sharma, N. R. Prakash, and P. Kalra, "Audio-video emotional response mapping based upon electrodermal activity," *Biomedical Signal Processing and Control*, vol. 47, pp. 324–333, 2019.

[15] S. Amalan, A. Shyam, A. Anusha, S. Preejith, A. Tony, J. Jayaraj, and S. Mohanasankar, "Electrodermal activity based classification of induced stress in a controlled setting," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018, pp. 1–6.

[16] A. A. S., S. P., S. V., S. K. S., S. A., T. J. Akl, P. S. P., and M. Sivaprakasam, "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 92–100, 2020.

[17] L. Zhu, P. Spachos, and S. Gregori, "Multimodal physiological signals and machine learning for stress detection by wearable devices," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2022, pp. 1–6.

[18] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors*, vol. 19, no. 7, 2019.

[19] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.

[20] R. Sánchez-Reolid, F. L. de la Rosa, M. T. López, and A. Fernández-Caballero, "One-dimensional convolutional neural networks for low/high arousal classification from electrodermal activity," *Biomedical Signal Processing and Control*, vol. 71, p. 103203, 2022.

[21] G. Oh, J. Ryu, E. Jeong, J. H. Yang, S. Hwang, S. Lee, and S. Lim, "Drer: Deep learning–based driver's real emotion recognizer," *Sensors*, vol. 21, no. 6, p. 2166, 2021.

[22] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos)," *IEEE Access*, vol. 7, pp. 57–67, 2018.

[23] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Transactions on Affective Computing*, 2019.

[24] M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "Cnn and lstm-based emotion charting using physiological signals," *Sensors*, vol. 20, no. 16, p. 4551, 2020.

[25] K. Ross, P. Hungler, and A. Etemad, "Unsupervised multi-modal representation learning for affective computing with multi-corpus wearable data," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–26, 2021.

[26] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[27] C. W. Darrow, "The rationale for treating the change in galvanic skin response as a change in conductance," *Psychophysiology*, vol. 1, no. 1, pp. 31–38, 1964.

[28] M. Gjoreski, B. Mitrevski, M. Luštrek, and M. Gams, "An inter-domain study for arousal recognition from physiological signals," *Informatica*, vol. 42, no. 1, 2018.

[29] Y. Shimomura, T. Yoda, K. Sugiura, A. Horiguchi, K. Iwanaga, and T. Katsuura, "Use of frequency domain analysis of skin conductance for evaluation of mental workload," *Journal of physiological anthropology*, vol. 27, no. 4, pp. 173–177, 2008.