2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Arousal-Valence Classification from Peripheral Physiological Signals Using Long Short-Term Memory Networks

M. Sami Zitouni, Cheul Young Park, Uichin Lee, Leontios Hadjileontiadis, *Senior Member, IEEE*, and Ahsan Khandoker, *Senior Member, IEEE*

***Abstract*— The automated recognition of human emotions plays an important role in developing machines with emotional intelligence. However, most of the affective computing models are based on images, audio, videos and brain signals. There is a lack of prior studies that focus on utilizing only peripheral physiological signals for emotion recognition, which can ideally be implemented in daily life settings using wearables, e.g., smartwatches. Here, an emotion classification method using peripheral physiological signals, obtained by wearable devices that enable continuous monitoring of emotional states, is presented. A Long Short-Term Memory neural network-based classification model is proposed to accurately predict emotions in real-time into binary levels and quadrants of the arousal-valence space. The peripheral sensored data used here were collected from 20 participants, who engaged in a naturalistic debate. Different annotation schemes were adopted and their impact on the classification performance was explored. Evaluation results demonstrate the capability of our method with a measured accuracy of >93% and >89% for binary levels and quad classes, respectively. This paves the way for enhancing the role of wearable devices in emotional state recognition in everyday life.**

I. Introduction

Emotion classification and recognition using wearable sensors have been an emerging research topic in recent years, as emotions are fundamental in humans' daily life, and accurately detecting emotional states can revolutionize healthcare and human-machine interaction (HMI) fields. Recognizing emotions can be approached using signals of external manifestations, such as speech and facial expressions. However, these signals may not reflect the actual emotions as people can conceal or suppress them. Alternatively, emotions can be recognized through physiological signals, such as electrocardiogram (ECG) and electroencephalograph (EEG) that are unconsciously generated by the autonomic nervous system for regulating bodily functions [1].

Advancements in mobile computing and wearable technologies have enabled the continuous monitoring of the physiological signals, allowing the transformation of the traditional healthcare system by shifting from treatment to prevention [2]. Sensors of wearable devices can provide data of physiological signals, such as heart rate (HR), blood volume pressure (BVP), electrodermal activity (EDA), and temperature (T), which can reflect emotional changes. E-health applications and self-monitoring devices with reliable personalization allow the enhancement of preventive healthcare, while cloud computing can solve the issue of scalability and availability of data [3].

Developing a reliable automated system that understands emotions is challenging, since emotions are elicited in diverse contexts, and the characteristics of physiological signals are complex [4]. Additionally, laboratory settings are mostly used. Thus, a robust emotion recognition system implemented in naturalistic settings and scenarios with wearable noninvasive sensors, enables an array of novel applications in daily real-life scenarios. For example, data provided from such a monitoring system can help understanding the etiology of mental health problems, such as stress, and enable studies to improve the diagnosis and treatments of mood disorders, such as depression and post-traumatic stress disorder (PTSD) [5]. Further, HMIs may support more nuanced communications with the users, by leveraging computers' ability to differentiate human emotional states and to react accordingly, thus enhancing user experiences.

Classification can be done based on discrete emotions, such as the discrete basic emotions of the Ekman model (fear, anger, happiness, sadness, surprise, contempt, disgust) [6]. It can also be done based on a dimensional model, such as arousal, valence, dominance, and liking, whose advantage is that no prior hypothesis of emotion categorization is required [5]. Additionally, recognition of emotions is most commonly investigated using brain signals [7], mainly EEG, either using a single-modal scheme [1], [8], or a multi-modal method with other physiological signals [5], [9]. Nevertheless, for daily life self-monitoring and HMI applications, peripheral signals provided by noninvasive wearable devices are more suitable to use. Therefore, works were conducted to classify emotions using a single peripheral signals [10], [11], as well as multi-modal peripheral data [12], [13].

In this work, a framework is introduced for emotion classification in the arousal and valence space, using multi-modal peripheral physiological signals collected in a naturalistic scenario. This study is based on physiological data, including HR, EDA, and T, obtained using wearable devices, during an impromptu debate between pairs, where participants' emotions were rated frequently. A neural network classification model based on a single Long Short-Term

M. Sami Zitouni, Ahsan Khandoker, and Leontios Hadjileontiadis are with the Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, 127788, United Arab Emirates. (mohammad.zitouni,ahsan.khandoker, leontios.hadjileontiadis)@ku.ac.ae

Cheul Young Park and Uichin Lee are with the Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 34141, South Korea. cheuly@kse.kaist.ac.kr; uclee@kaist.edu

Leontios Hadjileontiadis is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece. leontios@auth.gr

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

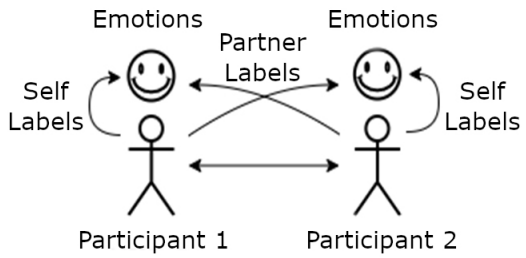

Fig. 1. A diagram of the self and partner annotation scheme during a conversation between two participants

Memory (LSTM) layer is proposed for accurate and real-time prediction of the emotions into two levels of arousal and valence, in addition to their quadrant combinations. The use of annotations provided by different rating perspectives (e.g., self vs. partner vs. others) is investigated, and their effects on the classification performance as whole and per participant, as well as the distribution of the results are demonstrated.

II. Methodology

A. Dataset

The focus of this framework is to use the physiological signals that are collected from human peripheral parts using wearable devices. Thus, data collected from 20 participants $\{P1, P5, P8, P9, P10, P11, P13, P14, P15, P16, P19, P22, P23, P24, P25, P26, P27, P28, P31, P32\}$ in the K-EmoCon [14] dataset are used in this work. K-EmoCon is a publicly available dataset with multi-modal affective information, including physiological signals, as well as audio and video recordings collected from participants engaging in naturalistic conversations, in the form of 10-minute debates between pairs on the Jeju Yemeni refugee issues. Since the focus in this framework is on signals that are collected from human peripheral parts using wearable devices, the physiological signals used are HR, EDA, and T signals, which were collected using Empatica E4 Wristband (for HR, EDA, and T) and Polar H7 Sensor (for HR signal measured from ECG signal). As a result, the chosen 20 participants, whose data are used in this work, have these signals available in the dataset with minimal distortions. The HR signals have a sampling rate of 1 Hz, while the EDA and T signals have a sampling rate of 4 Hz. EDA signal values can range between $0.01\mu S$ and $100\mu S$, whereas T signal values can range from $-40°C$ to $115°C$.Having similar sampling rates makes them suitable to be integrated in a multi-modal classification model. Furthermore, low sampling rates (1 to 4 Hz) help to implement a fast model for naturalistic scenarios where continuous monitoring is required.

B. Emotions Categorization

Emotions of the participants were annotated during the debate period every 5 seconds from different perspectives. Here, self (the participants rating themselves) and partner (the debate partners rating each other) annotations are adopted, as well as combined annotation combining both

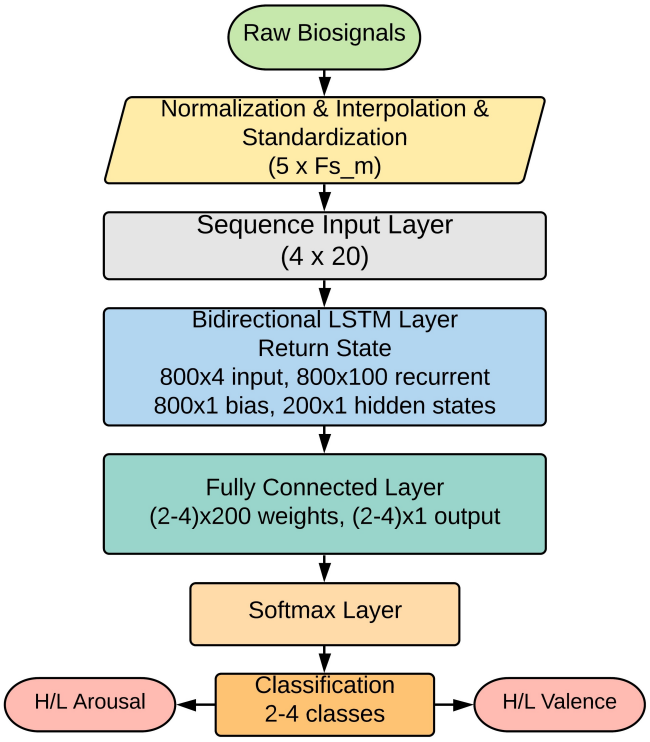

Fig. 2. Illustration of the proposed emotion classification model

ratings. Fig. 1 illustrates the annotation scheme. The emotions were annotated based on arousal and valence affective dimensional emotional model as in Russell's circumplex model of affect [15], and they were measured with a Likert scale from 1 to 5. Accordingly, the emotions are classified by the model based on the level of arousal and valence into high (H) and low (L). Moreover, the emotions are categorized into one of quad classes combining the arousal and valence levels, which are and high arousal high valence (HAHV), high arousal low valence (HALV), low arousal high valence (LAHV), low arousal low valence (LALV). Therefore, the self and partner annotations in Likert scale ratings are converted into H and L according to a mid-value of 2.5 (L: 1-2, H: 3-5). For the combined annotations, the self and partner ratings are accumulated and re-scaled into 1 to 9, then converted into H and L based on a mid-value of 4.5 (L: 1-4, H: 5-9).

C. Classification Model

To classify the emotions based on the peripheral bio-signals according to the classes aforementioned, an LSTM-based classification model is used. It is essential to have a fast and robust model suitable for real-life frequent continuous monitoring (5 seconds intervals in the K-Emocon dataset) of emotions using low frequency signals. Fig. 2 displays a block diagram that illustrates the proposed emotion classification model. First, the raw signals of each participant are normalized separately. The normalization is performed based on the collected signals during a relaxation period prior to the debate, where here the last 1.5 minutes were used. This was

TABLE I

EMOTION CLASSIFICATION ACCURACY (%) RESULTS OF THE PROPOSED MODEL AGAINST BASELINE TECHNIQUES USING DIFFERENT ANNOTATIONS

	Arousal			**Valence**			**Quad Classes**		
	GNB	*XGBoost*	*Proposed*	*GNB*	*XGBoost*	*Proposed*	*GNB*	*XGBoost*	*Proposed*
Self	62.55	79.16	**90.22**	74.86	84.82	**92.39**	45.93	68.49	**86.18**
Partner	67.76	81.26	**93.17**	74.12	84.17	**90.68**	49.17	72.88	**89.13**
Combined	65.87	77.56	**88.98**	67.88	79.12	**88.51**	44.09	66.28	**83.23**
Average	65.39	79.33	**90.79**	72.29	82.70	**90.53**	46.40	69.22	**86.18**

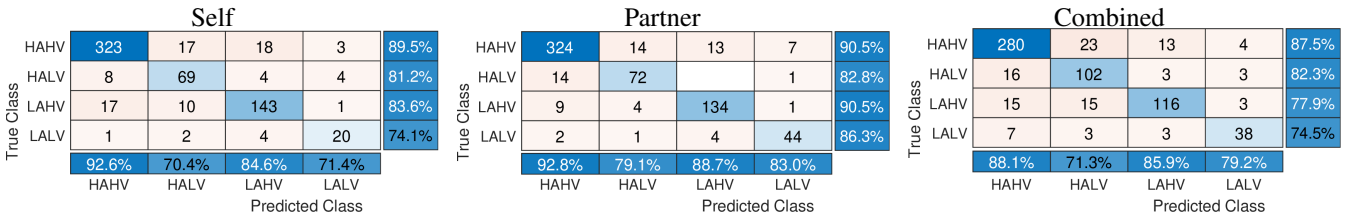

Fig. 3. Confusion matrices of emotion classification results for quad classes with different annotations

employed to remove personal bias based on the signals of the participants' natural state, which may vary due to several factors, such as age, gender, and physiological nature.

The lower frequency bio-signals (1-Hz HR signals) are then interpolated based on the highest sampling frequency used, based on the nearest-neighbor method. Then the signals are divided into segments of size $w \times Fs_m$, where Fs_m is the highest sampling frequency of the used signals (4 Hz here), and w is the annotation/classification interval (5 seconds here). This setting can be altered based on the dataset and the annotation periods of the raters.

For this model, an LSTM network is used for training and classifications. As it was proven in the literature, methods using LSTM networks were able to achieve good and robust performances when used in classification tasks of sequences from physiological signals [4], [9]. The proposed classification model includes a neural network with one bidirectional LSTM layer. Since the network is relatively shallow, training and testing can be performed considerably fast, while being able to achieve good performance. The neural network part of the model consists of a sequence input layer with a size equal to the number of signal types used (4 in this case), as well as a bidirectional LSTM layer with 100 hidden units where the return state is used, a fully connected layer, and a softmax layer for classification. The output corresponds to the level of arousal or valence. The output has two classes when trained and used for binary arousal and valence level, or four classes in case of their combination.

D. Implementation

The proposed classification model was implemented in Matlab 2020a. The options used for training are as follows. A minimum sequence length of 20, which is equal to the input sequences length ($w \times Fs_m$), to minimize the amount of padding in the mini batches. The model was trained for each experiment with a number of epochs of 500, with no shuffling of data. On a machine with 16 MB of RAM and Nvidia 980m GPU, the classification speed was 67.5 μs/prediction, which is much faster than needed for an interval of $w = 5s$, making it very suitable for real-time implementation.

III. RESULTS & DISCUSSION

The presented results of the proposed model were obtained with 4-fold cross-validation scheme, while training and testing using different annotation perspectives. Table I tabulates the emotion classification accuracy of the proposed method, in comparison to baseline techniques. The baselines are Gaussian Naive Bayes (GNB) [7], a probabilistic classifier, and XGBoost [16], an efficient high-performance tree boosting system, both trained with 30 features from the used peripheral signals following Soleymani et al.'s TEAP toolbox [17]. The K-EmoCon dataset implementation can be found in this site [18].

The results show the superiority of the proposed model with average classification accuracy of 90.79% for arousal, 90.53% for valence, and 86.18% for quad classes. In arousal classification tests, all the methods performed the highest when using the partner annotations (93.17% for the proposed), which is the same case as in quad classes classification (89.24% for the proposed). On the other hand, in valence classification experiments, the best performances were obtained using the self annotations (92.39% for the proposed). The accuracies obtained when using the combined annotations were the lowest, which may be due to the change in the data balance.

Confusion matrices of the quad classes classification results are displayed in Fig. 3. This shows the performance of the proposed model per each class. Additionally, the difference between the annotations in the distribution of the data across classes can be observed. In all cases, the data are mainly biased towards HAHV, which is the normal human state, where no negative emotion is present, and least biased towards LALV, where the emotion state is negative overall. It can be observed that the combination of the annotations leads to a more balanced distribution across the classes.

Fig. 4 shows heat-maps of inter-rater reliability between the predictions from the classifier, and the annotations used for training and testing, based on the Krippendorff's alpha. Following the experiments performed in Table I, the values are calculated for each of the 20 participants. The alpha value

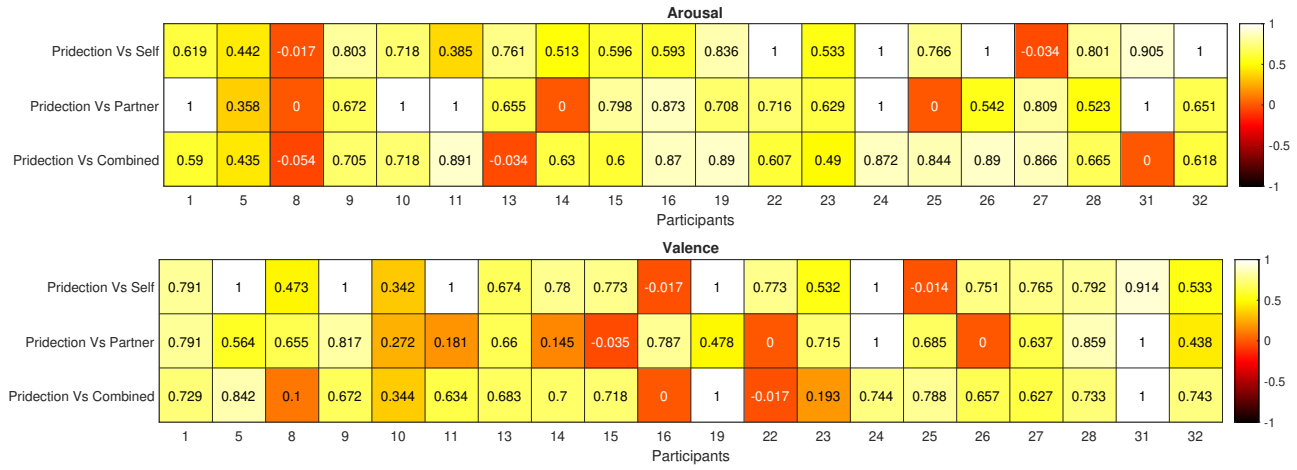

Fig. 4. Heat-maps of inter-rater reliability of model predictions against the used annotations for each participant's measured with Krippendorff's alpha

for some participants within the same test is low in comparison to the others ($P8|P27$ in arousal and $P16|P25$ in valence with self, $P8|P14|P25$ in arousal and $P15|P22|P26$ in valence with partner), which indicates that the emotion labeling or/and the participant emotion state is inconsistent with the other participants. Additionally, the alpha value varies dramatically across annotation types ($P14|P25|P27$ in arousal, $P11|P15|P16|P26|$ in valence), indicating that one labeling perspective can be more accurately representative of the emotional state that the other. Another interesting observation is that the reliability for all participants in the cases of self and combined annotations, have more consistent alpha values compared to the partner case. In other words, the standard deviation across the participants is lower (0.3011, 0.3331 and 0.3013 in arousal and valence for self, partner and combined respectively).

IV. Conclusions

This work presented a framework for emotion classification in the arousal and valance space, from peripheral physiological signals, using an LSTM neural network-based model. The emotions were categorized into two levels of arousal and balance and their quadrant classes. The peripheral signals used were collected during naturalistic conversations with various annotation schemes. The experimental results have shown increased performance (for arousal 93.17% and quad classes 89.13%-partner annotations; for valence classification 92.39%-self annotations). Classification using combined annotations resulted in more balanced results. The future work includes implementation of continuous emotional state analysis in daily-life settings, with the development of a cloud powered mobile application for personalized physical and mental health monitoring.

References

[1] Y. Luo, Q. Fu, J. Xie, Y. Qin, G. Wu, J. Liu, F. Jiang, Y. Cao, and X. Ding, "Eeg-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46 007–46 016, 2020.

[2] S. Jhajharia, S. Pal, and S. Verma, "Wearable computing and its application," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5700–5704, 2014.

[3] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion recognition from multimodal physiological signals for emotion aware healthcare systems," *Journal of Medical and Biological Engineering*, vol. 40, pp. 149—157, 2020.

[4] B. H. Kim and S. Jo, "Deep physiological affect network for the recognition of human emotions," *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 230–243, 2018.

[5] A. Anderson, T. Hsiao, and V. Metsis, "Classification of emotional arousal during multimedia exposure," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, 2017, pp. 181–184.

[6] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *International Journal of Synthetic Emotions (IJSE)*, vol. 1, no. 1, pp. 68–99, 2010.

[7] 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.

[8] J. Liu, H. Meng, A. Nandi, and M. Li, "Emotion detection from eeg recordings," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 2016, pp. 1722–1727.

[9] J. Liao, Q. Zhong, Y. Zhu, and D. Cai, "Multimodal physiological signal emotion recognition based on convolutional recurrent neural network," *MS&E*, vol. 782, no. 3, p. 032005, 2020.

[10] J. Shukla, M. Barreda-Angeles, J. Oliver, G. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Transactions on Affective Computing*, 2019.

[11] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on affective computing*, vol. 3, no. 1, pp. 102–115, 2011.

[12] M. B. H. Wiem and Z. Lachiri, "Emotion classification in arousal valence model using mahnob-hci database," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 3, 2017.

[13] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 237–249, 2011.

[14] C. Y. Park, N. Cha, S. Kang, A. Kim, A. H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, "K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," *Scientific Data*, vol. 7, no. 293, 2020.

[15] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[16] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[17] M. Soleymani, F. Villaro-Dixon, T. Pun, and G. Chanel, "Toolbox for emotional feature extraction from physiological signals (teap)," *Frontiers in ICT*, vol. 4:1, 2017.

[18] C. Y. Park. (2020) Pyteap, a python implementation of toolbox for emotion analysis using physiological signals (teap). [Online]. Available: https://pypi.org/project/PyTEAP/