Multimodal Physiological-based Emotion Recognition

Astha Sharma and Shaun Canavan

***Abstract*—In this paper, we propose a multimodal approach to emotion recognition using physiological signals by showing how these signals can be combined and used to accurately identify a wide range of emotions such as happiness, sadness, and pain. The proposed approach combines multiple signal types such as blood pressure, respiration, and pulse rate into one feature vector representation of emotion. Using this feature vector, we train a deep convolutional neural network to recognize emotions from 2 state-of-the-art datasets, namely DEAP and BP4D+. On BP4D+, we achieve an average, subject independent, emotion recognition accuracy of 94% for 10 emotions. We also detail subject-specific experiments, as well as gender specific models of emotion. On DEAP, we achieve 86.09%, 90.61%, 90.48%, and 90.95% for valence, arousal, liking, and dominance respectively, for single-trial classification. We also detail state-of-the-art results on BP4D+ and DEAP.**

***Index Terms*—Emotion recognition, deep learning, physiological data, multimodal**

I. Introduction

Emotional state is mental composition which shows our reaction to an experience. It is an integral part of human communication and behavior [1]. Recognizing emotions using machines has important roles in human-computer interaction, including applications in video games [2], assessment of multimedia technology, recommendations for multimedia content [3], pain recognition [4], and classification of Austism Spectrum Disorder [5]. While recognizing others emotions, we generally look at their facial expressions, speech or body language, however, these features can be misguiding. Facial expression, voice, and body languages can be faked. Faces can be occluded, and facial expression can be contradictory which can make feature identification difficult. For example, it has been observed that people smile during negative emotional experiences [6]. Considering this, physiological signals such as heart rate, blood pressure, respiratory signals, and Electroencephalogram (EEG) signals can be important traits for identifying emotions accurately.

There has been an increase in the works that use physiological data for emotion recognition in recent years. Mert and Akan [7], investigated empirical mode decomposition for classification of low/high arousal and valence. They also looked at the multivariate extension, which they show is useful for analyzing non-stationary EEG signals. Martinez et al. [8] combined skin conductance and blood pressure along with deep neural networks to recognize emotions. The efficiency of their model was compared with standard feature extraction and feature selection methods. They showed that their deep learning approach performed better than standard feature selection algorithms and the method shows more generalization. Sano et al. [9] used physiological data to measure stress. For 5 days, skin conductance for 18 participants was collected with wrist sensors, as well as their mobile phone usage including call, SMS, and location were monitored. A survey was done to know stress, mood, sleep, tiredness, general health, alcohol or caffeinated beverage intake and electronics usage. Correlation analysis was applied to find important features that were used to classify whether the participant was stressed or not.

A computer-aided diagnosis system was developed to automate the classification of EEG signals in three categories - normal, preictal, and seizure [10]. This method achieved 88.67%, 90.00% and 95.00% accuracy respectively. Vijayan et al. [11] used EEG signals to classify different emotions such as happiness, fear, and sadness. The experiments were performed on the DEAP dataset [12], and Shannon Entropy was used for feature extraction and a multi-class Support Vector Machine was used for training. The accuracy obtained for classification was 94.097%. Picard et al. [13] showed variations in physiological signals on a daily basis. They proposed seeding a Fisher Projection with the results of Sequential Floating, achieving an accuracy of 81% on 8 emotions. Wagner et al. [14] combined multiple physiological signals to find the affective state. A musical induction method was used to ignite real emotion in subjects for data collection. Electromyogram, electrocardiogram, skin conductivity and respiration signals were used to classify four musical emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal). A feature-based approach was proposed, obtaining a classification accuracy of 95% for subject-dependent and 70% for subject-independent experiments.

Motivated by these works we propose a method for emotion recognition using the combination of physiological signals that include heart rate, blood pressure, respiration, EDA, and EEG. We use this data to train a deep convolutional neural network to recognize a range of emotions. The contributions of this work are three-fold, and can be summarized as follows:

1) We propose a multimodal approach to recognizing emotion using physiological data and deep convolutional neural networks.
2) We detail gender-specific models of emotions showing the difficulty when training and testing across male vs. female data.
3) We compare the proposed method to state-of-the-art methods on BP4D+ and DEAP datasets, achieving state-of-the-art performance.

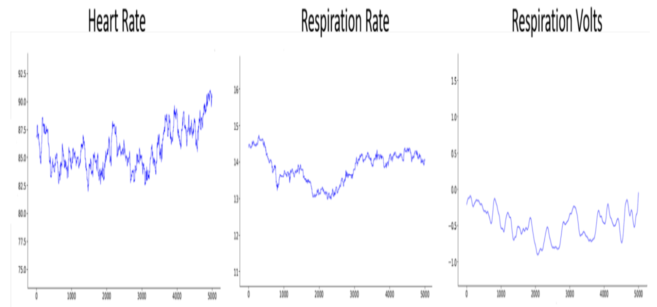

Fig. 1. Heart rate, respiration rate, and respiration volts from BP4D+, from subject experiencing a 'Happy' emotion.

Fig. 2. Fp1 EEG channel from DEAP.

II. Datasets

BP4D+. BP4D+ [15] is a multimodal spontaneous emotion corpus which includes 8 physiological signals that include blood pressure, EDA (skin conductance), heart rate and respiration. The BP4D+ also includes 2D, 3D and thermal images and videos, facial landmarks and action units. For our experiments only the physiological data was used. The huge versatility of gathered data makes BP4D+ one of the largest databases of this kind. This data was gathered from 140 subjects (58 males and 82 females) from 18 to 66 years of age. It includes the following 10 emotions: happy, sad, surprise, startle, skeptical, embarrassed, fear, pain, anger, and disgust. Each emotion was elicited through tasks such as holding hand in ice water (pain), and experiencing a smelly odor (disgust). See Fig. 1 for examples from BP4D+.

DEAP. DEAP [12] is a multimodal dataset based on the Valence-Arousal emotion model. It contains electroencephalogram (EEG) signals, as well as has sequences of peripheral signals that include EOG (eye movements), EMG (muscle movement), GSR, respiration, blood pressure and temperature. The data was collected from 32 participants (19-37 years age, 50% male and 50% female) watching 40 one-minute long music videos to elicit emotions. 32 EEG channels based on the 10-20 system [16] for recording EEG data and 8 channels for peripheral physiological data were used. These signals were recorded with a sampling rate of 512 Hz which was downsampled to 128 Hz after preprocessing. The data was labeled with arousal, valence, dominance, and liking values ranging from 1 to 9 showing intensity of each emotional state. See Fig. 2 for an example of a Fp1 EEG channel from DEAP.

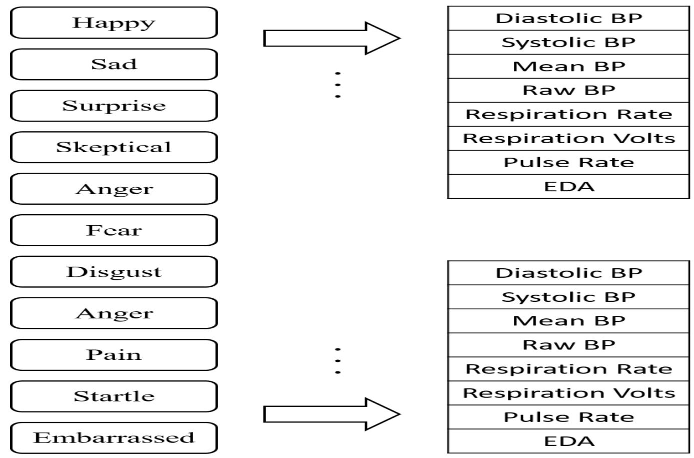

Fig. 3. Feature vector representation for BP4D+. *NOTE: 1 frame from each signal is in each feature vector (i.e. 8 frames for BP4D+).*

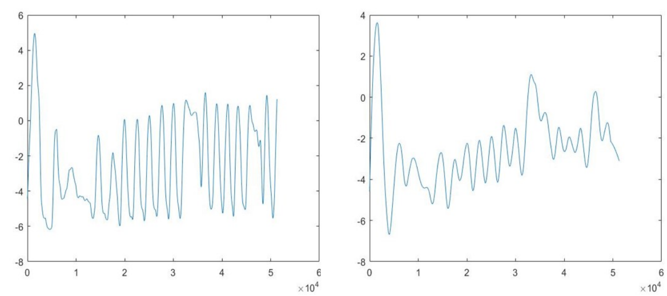

Fig. 4. Example of Savitzky-Golay smoothing on BP4D+. Left: original respiration volts signal from subject with 'Happy' emotion. Right: smoothed signal from applied filter.

III. Proposed Method

We propose to use the combination of multiple physiological signals to train a convolutional neural network for emotion recognition. In BP4D+ there are large variations in the data, therefore we first perform preprocessing on the data, specifically smoothing and scaling. We smooth the data to increase the signal to noise ratio without deforming the signal, which makes it easier to see trends in the data. In our experiments we use the Savitzky-Golay filter [17] as it has been shown to be more efficient in handling delay alignment and the transient effect at the start and end of the sequence, compared to methods such as moving average, and median filters [18]. An example, of an original signal with the smoothed signal can be seen in Fig. 4. Once the signals are smoothed, we then scale the data into the range of [0, 1] which helps with large variations in the different signals. For our experiments with the DEAP dataset, we did not perform any additional preprocessing, as this data is already preprocessed for use [12]. Given the smoothed and scaled physiological signals, for each emotion we then create feature vectors that contain 1 frame of each of the available signals which are then used to train a deep neural network. As shown in Fig. 3, for BP4D+, the feature vector includes (in order) 1 frame of data from diastolic

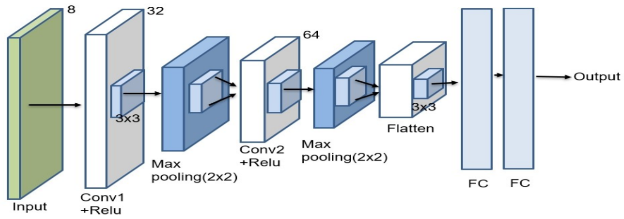

Fig. 5. Convolutional neural network architecture.

BP, systolic BP, Mean BP, Raw BP, Respiration Volts, Pulse Rate, and EDA. While this is a simple approach, we will show (Section V), that it is an effective representation of emotion outperforming current state of the art on BP4D+ and DEAP.

IV. Experimental Design

For BP4D+, all 140 subjects were used, and each task was collected over different time periods (i.e. the total number of frames is different), due to this we had approximately 450,000 feature vectors, for our experiments. For DEAP, all 32 subjects were used, with each feature vector (in the dataset) having 8064 frames. This resulted in over 10 million feature vectors for our experiments, validating the efficacy of the proposed approach for emotion recognition.

A. Deep Neural Network Architecture

In recent years, deep neural networks have proven highly efficient and have outperformed humans when classifying modalities such as audio, images, and text [19]. Deep networks have also successfully been used for classification of medical images [20], as well as prediction of future sales prices [21]. Motivated by the success of deep neural networks for a range of tasks and modalities, we train a 9-layer convolutional neural network (CNN), with the combined physiological signals (Fig. 3), to recognize emotion. The developed CNN uses two sets of convolutions, activation (ReLU) and max pooling layers. Dropout is used for regularization to help the model generalize better by reducing overfitting [22]. The RMSprop optimizer is used with a learning rate of 0.001. The network was trained using 150 epochs and a batch size of 32. See Fig. 5 for more details on the developed CNN architecture.

B. BP4D+ Experimental Design

When evaluating the efficacy of the proposed approach, on BP4D+, we are interested in answering two broad questions: (1) How does gender influence emotion recognition with the proposed method? and (2) Are there large differences in accuracy for subject-dependent vs. subject-independent with the proposed method? Considering these questions, we conducted the following experiments.

1) **(Experiment 1)** We evaluated the proposed method in a subject-dependent manner. In doing this we created 140 deep models of emotion, one for each subject in the dataset. For each model, 80% of the subject data was used to train, and 20% was used to test. Due to the subject-dependent nature of this experiment, each model was trained and tested only on the same subject. This experiment was conducted to be consistent with the experimental design when using the DEAP dataset (single-trial classification as detailed in section IV-C).
2) **(Experiment 2)** We evaluated the proposed method in a subject-independent manner. One model of emotion was created that used 80% of the data for training and 20% for testing, where the same subject did not appear in both training and testing.
3) **(Experiment 3)** We created gender-specific models of emotion. In this design, two deep models were created. For each model, 80% of the females and 80% of the males were used to train the respective gender-specific model. In this design, we test the gender-specific model across both genders (e.g. female model tested on both male and female subjects). In this experimental design, the same subject did not appear in both training and testing data.

C. DEAP Experimental Design

To evaluate the efficacy of the proposed approach on DEAP, we conducted single-trial classification experiments [12]. All experiments are done individually on each subject and evaluation is done by calculating mean and standard deviation, across all subjects, for every set of experiments. We split the data (40 channels as detailed in Section II) into three sets: (1) EEG (32 channels); (2) peripheral (8 channels); and (3) EEG and peripheral (40 channels). In total, 12 single-trial classification experiments were conducted for each subject. One experiment for each emotion label (valence, arousal, liking, dominance) resulting in 384 single-trial classification experiments over all of the DEAP dataset.

V. Results

A. BP4D+

Experiment 1 was conducted in a person-specific manner resulting in a total of 140 deep models of emotion (one for each subject). To evaluate the efficacy of these models, we detail the mean accuracy (across each subject) along with the standard deviation (Table I). This is done for three groups of subjects (1) all subjects; (2) female subjects; and (3) male subjects.

As can be seen in Table I, both male and female data had a relatively high mean accuracy, with a low standard deviation. Many of the subjects had accuracy at or near 100%, although there were a few outliers (Fig. 6). For example, subject 140 had a lower accuracy at approximately 88%. It is important to note that subject 140 was a male participant which can explain, in part, the relatively lower accuracy of male subjects compared to female subjects (Table I). Overall, these results detail the expressive power of the proposed approach for subject-specific emotion recognition. This shows that the majority of the subjects (both male and female), were recognized with high accuracy when using subject-specific models of emotion. Although the results for subject-specific

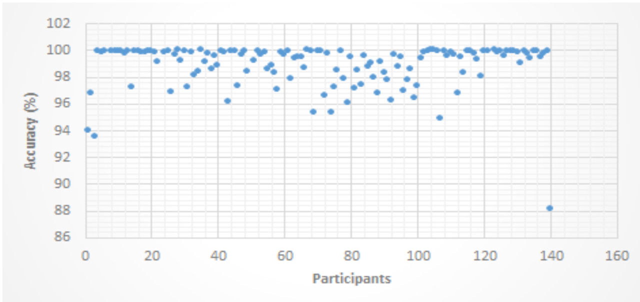

Fig. 6. BP4D+ subject-specific accuracy distribution.

emotion recognition are encouraging, in a real-wold setting it cannot be guaranteed that the test subject will appear in the training data. Considering this, we conducted Experiment 2 where the subject does not appear in the training and testing data. Experiment 2 was conducted over all emotions and all subjects (i.e. one deep model of emotion was used). Across all subjects, we achieved an accuracy of 94%. One of the questions we wanted to answer with our experiments on BP4D+, was whether the proposed method can work well on the same subject, as well as generalize to unseen subjects. These results are encouraging, as they show a relatively small difference in the emotion recognition accuracy of 4.89%, when comparing the average accuracy for subject-dependent experiments compared to the overall accuracy of the subject-independent experiment (98.89% vs. 94%).These results detail the expressive power of the proposed method for recognizing a range of emotions.

The final experiment conducted on BP4D+, was using gender specific models of emotion. In this experimental design, we evaluated both male and female models on both genders (e.g. female model on both male and female testing data). It is interesting to note that the female model outperformed the male model for both same and cross-gender testing. The gender-specific results (Tables I and II) can be explained, in part, as it has been found that some emotions are more easily recognized in female subjects compared to male [23].

TABLE I
SUBJECT-SPECIFIC BP4D+ RESULTS.

Data	Recognition Accuracy
All participants	98.89% ±1.647
Female	99.09% ±1.212
Male	98.88% ±1.916

TABLE II
EMOTION RECOGNITION ACCURACIES ON GENDER-SPECIFIC MODELS OF EMOTION.

Training Data (Model)	Testing Data	Accuracy
Female	Female	96.77 %
Male	Male	93.60 %
Female	Male	15.35 %
Male	Female	15.08 %

TABLE III
SINGLE-TRIAL CLASSIFICATION RESULTS FROM DEAP.

Emotion Category	EEG	Peripheral	Both
Valence	Mean: 60.21% ±6.306	Mean: 86.31% ±6.186	Mean: 86.09% ±5.367
Arousal	Mean: 65.03% ±9.486	Mean: 88.83% ±4.455	Mean: 90.61% ±3.579
Liking	Mean: 67.59% ±11.295	Mean: 88.38% ±6.416	Mean: 90.48% ±4.954
Dominance	Mean: 66.22% ±12.528	Mean: 89.12% ±5.484	Mean: 90.95 % ± : 4.667

As can be seen in Table II, both models performed well when tested on the same gender with female achieving 96.77% recognition accuracy on female data, and male achieving 93.6% recognition accuracy on male data. However, both male and female deep models performed poorly on the opposite gender, achieving approximate accuracies of 15%. The low recognition accuracy from cross-gender testing can be explained, in part, by the idea that neurons flow in different part of the brain in males and females during emotion elicitation. For women, these neurons connect the parts of brain that regulate internal areas of body that impacts blood pressure, respiration and hormones, while in men, these neurons connect to the areas of brain that controls vision and movement [24], [25]. It has also been noted that there are obvious differences in the emotional responses of opposite genders when analyzing facial features and physiological data [26]. An interesting application that could benefit from these results is gender classification, however, this is outside of the scope of this paper and left for future work.

B. DEAP

When analyzing DEAP data with the proposed method, we have found that the combination of both Peripheral and EEG data (40 channels) gave the highest recognition accuracy in 3 out of 4 of the emotion categories (arousal, liking, and dominance). Peripheral data alone outperformed both for valence with a mean accuracy of 86.31% compared to 86.09% with both. EEG data alone (32 channels) performed the worst in all categories (Table III). These results are supported by previous studies that have shown a multimodal approach to classification can lead to higher accuracy compared to a single modality, when using physiological data to classify infant pain [27].

As can be seen in Table III, the standard deviations for each subject are much higher compared to those found in Experiment 1 from the detailed BP4D+ experiments (Table I). This is especially true when analyzing the arousal, liking, and dominance emotion categories for EEG data alone. Arousal had the lowest standard deviation for peripheral and EEG + peripheral experiments, although it is higher compared to valence for EEG data alone. This could be partially explained by EEG data not being able to easily generalize across subjects [28]. Although the standard deviation of peripheral data is lower compared to EEG data, when both EEG and peripheral signals were combined, they gave the lowest standard deviation for

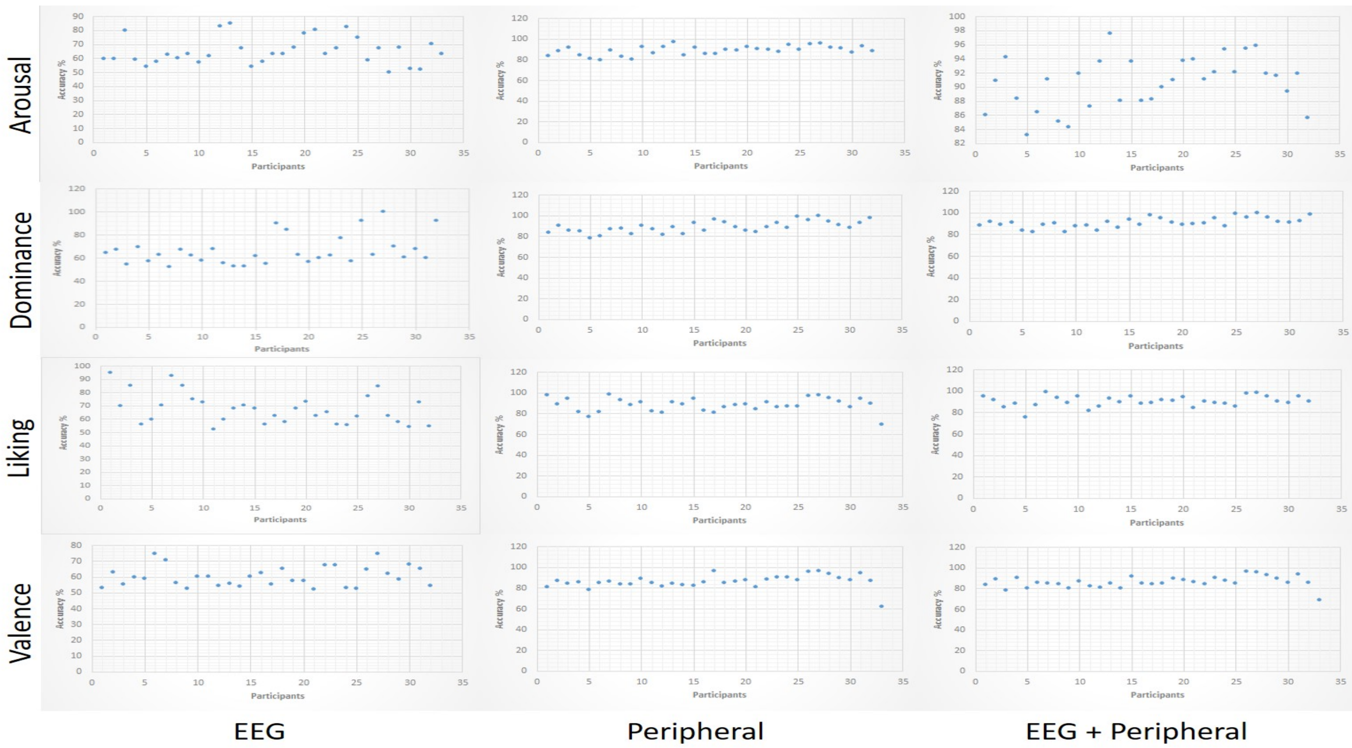

Fig. 7. Subject accuracies for dominance emotion category. From left to right: EEG data, peripheral data, EEG and peripheral data.

each emotion category. Again, this can partially be explained due to the multimodal nature of EEG + peripheral data. For dominance and liking, the standard deviation is highest for EEG data along with liking $>11\%$ and Dominance $>12\%$. Similar to arousal, this can partially be attributed to EEG data's inability to easily generalize across subjects. As can be seen in Fig. 7, with dominance, there are approximately 6 subjects that are causing the higher standard deviation (e.g. outliers). Most of the subjects having an accuracy within the range of 50%-70%, however, the 6 outliers are within the range of 80% to 100%. A question that arises from this work is: Is it possible to generalize across subjects with EEG data? With valence having the lowest standard deviation, among subjects, that can be a good starting point for future experiments to answer this question.

C. State Of The Art Comparisons

BP4D+. To the best of our knowledge Zhang et al. [15] are the first and only to detail results on physiological data from BP4D+. They randomly selected 45 subjects and conducted two experiments. First, using hand-crafted features and a RBF kernel SVM, they achieved an accuracy of 59.5% on happy, sad, startled, fear, and disgust. Secondly, they performed binary classification for low and high arousal. Using this approach, they achieved an accuracy of 60.5% over all 10 emotions in the dataset. In the proposed approach, we achieved an average accuracy of 94% on subject-independent training and testing over all subjects, and all emotions, in BP4D+.

DEAP. This dataset has successfully been used for emotion recognition since it's release by Koelstra et al. [12]. Considering this, we compare against multiple state-of-the-art approaches for recognizing the 4 emotion categories available in the DEAP dataset.

TABLE IV
COMPARISON WITH CURRENT STATE OF ART FOR DEAP.

	Valence	Arousal	Dominance	Liking
Proposed method	**86.31%**	**90.61%**	**90.95%**	**90.48%**
Liu et al. [29]	85.2%	80.5%	84.9%	82.4%
Rozgic et al. [30]	76.9%	69.1%	73.9%	75.3%
Mert and Akan. [7]	72.87%	75.0%	N/A	N/A
Daimi and Saha [31]	65.3%	66.90%	N/A	N/A
Li et al. [32]	58.4%	64.3%	65.8%	66.9%
Jirayucharoensak et al. [33]	53.42%	52.03%	N/A	N/A
Koelstra et al [12]	65.2%	63.1%	N/A	64.2%

As can be seen in Table IV, the proposed method outperforms the current state of the art on arousal, valence, liking, and dominance. The increase in accuracy over the current state of the art can be explained, in part, by the proposed method using the combination of the different signals to train a deep network (CNN) to recognize the emotions, where the compared works use classical machine learning approaches. Koelstra et al [12] used a Naïve Bayes classifier, Liu et al [29] used a linear Support Vector Machine (SVM), and Rozgic et al [30] used a combination of K-PCA and 1-NN. While Li et al [34] extracted features from a deep belief network, they used an SVM to classify the emotions.

VI. Discussion and Conclusion

We have presented an approach to emotion recognition, using physiological signals, that combines individual frames from different signal types (e.g. blood pressure and respiration rate). We tested the efficacy of the proposed approach on 2 publicly available datasets, namely BP4D+ and DEAP. The proposed method outperforms the current state of the art on BP4D+ and DEAP. It has applications in multimedia, medicine, defense and military related fields. These applications include analysis of stress and pain, lie detection, increasing soldier survivability in combat, and classification of autism in children.

Recently, gender classification has been shown to be promising with facial features and deep CNNs [35]. Considering this, we investigated generalization across genders, on BP4D+, through gender-specific models of emotion. As shown in section V-A, the gender-specific models of emotion performed poorly when cross-gender data was used to test (e.g. male deep model tested on female data). Interesting, a previous study notes that physiological signals are similar during similar emotions in male and female [36], however, our results contradict this and are supported by the work from Whittle et al. [25], as we previously mentioned in Section V-A.

An interesting application of these results, from the gender-specific models, is using the proposed method to classify a subject's gender. Due to challenges with using facial data for gender classification such as pose and lighting variation [35], physiological data could be a useful alternative as it does not suffer from those same challenges. Considering this, we will conduct experiments on gender classification using the deep gender-specific models we have developed here. We will also compare the results from other deep neural network architectures such as recurrent neural networks.

References

[1] I. Mauss and M. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[2] D. Giakoumis et al., "Automatic recognition of boredom in video games using novel biosignal moment-based features," *IEEE Trans. on Affective Computing*, vol. 2, no. 3, pp. 119–133, 2011.

[3] M. Prabhakaran B. Mariappan, M. Suk, "Facefetch: A user emotion driven multimedia content recommendation system based on facial expression recognition," *INtl. Symposium on Multimedia*, 2012.

[4] G. Zamzmi et al., "An approach for automated multimodal analysis of infants' pain," in *International Conference on Pattern Recognition*, 2016, pp. 4148–4153.

[5] H. Drimalla, N. Landwehr, I. Baskow, B. Behnia, S. Roepke, I. Dziobek, and T. Scheffer, "Detecting autism by analyzing a simulated social interaction," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2018, pp. 193–208.

[6] P. Ekman, "The argument and evidence about universals in facial expressions," *Handbook of social psychophysiology*, pp. 143–164, 1989.

[7] A. Mert and A. Akan, "Emotion recognition from eeg signals by using multivariate empirical mode decomposition," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 81–89, 2018.

[8] H. Martinez, Y. Bengio, and G. Yannakakis, "Learning deep physiological models of affect," *Computational Intelligence Magazine*, vol. 8, no. 2, pp. 20–33, 2013.

[9] A. Sano and R. Picard, "Stress recognition using wearable sensors and mobile phones," *Affective Computing and Intelligent Interaction*, pp. 671–676, 2013.

[10] U. Acharya, S. Oh, Y. Hagiwara, J. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in biology and medicine*, 2017.

[11] A. Vijayan, D. Sen, and A. Sudheer, "Eeg-based emotion recognition using statistical measures and auto-regressive modeling," *Computational Intelligence & Communication Technology*, pp. 587–591, 2015.

[12] S. Koelstra et al., "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[13] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on PAMI*, vol. 23, no. 10, pp. 1175–1191, 2001.

[14] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *ICME*, pp. 940–943, 2005.

[15] Z. Zhang et al., "Multimodal spontaneous emotion corpus for human behavior analysis," *CVPR*, pp. 3438–3446, 2016.

[16] G. Klem et al., "The ten-twenty electrode system of the international federation," *Elec. Clin. Neurophys.*, vol. 52, no. 3, pp. 3–6, 1999.

[17] W. Press and S. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.

[18] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An improved signal segmentation using moving average and savitzky-golay filter," *Journal of Signal and Information Processing*, vol. 3, no. 01, pp. 39, 2012.

[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[20] A. Ortiz, J. Munilla, J. Gorriz, and J. Ramirez, "Ensembles of deep learning architectures for the early diagnosis of the alzheimers disease," *International Journal of Neural Systems*, vol. 26, no. 07, 2016.

[21] M. Rafiei and H. Adeli, "A novel machine learning model for estimation of sale prices of real estate units," *Journal of Construction Engineering and Management*, vol. 142, no. 2, 2015.

[22] G. Dahl, T. Sainath, and G. Hinton, "Improving deep neural networks for lvcsr using rectified linear units and dropout," *Acoustics, Speech and Signal Processing*, pp. 8609–8613, 2013.

[23] J. Bailenson et al., "Real-time classification of evoked emotion using facial feature tracking and physiological responses," *International Journal of Human-Computer Studies*, vol. 66, no. 5, pp. 19–31, 2008.

[24] M. Bradley, "Emotions differences between men and women," in *HealthGuidance for better health*, 2014.

[25] S. Whittle, M. Yücel, M. Yap, and N. Allen, "Sex differences in the neural correlates of emotion: evidence from neuroimaging," *Biological psychology*, vol. 87, no. 3, pp. 319–333, 2011.

[26] C. Li, C. Xu, and Z. Feng, "Analysis of physiological for emotion recognition with the irs model," *Neurocomputing*, vol. 178, pp. 103–111, 2016.

[27] G. Zamzmi, C. Pai, R. Goldgof, D. Kasturi, T. Ashmeade, and Y. Sun, "An approach for automated multimodal analysis of infants pain," *International Conference on Pattern Recognition*, 2016.

[28] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring eeg features in cross-subject emotion recognition," *Frontiers in Neuroscience*, 2018.

[29] W. Liu, W. Zheng, and B. Lu, "Multimodal emotion recognition using multimodal deep learning," *arXiv preprint arXiv:1602.08225*, 2016.

[30] V. Rozgić, S. Vitaladevuni, and R. Prasad, "Robust eeg emotion classification using segment level decision fusion," *ICASSP*, pp. 1286–1290, 2013.

[31] G. Daimi, S.and Saha, "Classification of emotions induced by music videos and correlation with participants rating," *Expert Systems with Applications*, vol. 41, no. 13, pp. 6057–6065, 2014.

[32] X. Li, P. Zhang, D. Song, and Y. Hou, "Recognizing emotions based on multimodal neurophysiological signals," *Advances in Computational Psychophysiology*, pp. 28–30, 2015.

[33] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[34] X. Li et al., "Eeg based emotion identification using unsupervised deep feature learning," 2015.

[35] K. Zhang et al., "Gender and smile classification using deep convolutional neural networks," *CVPR Workshops*, 2016.

[36] D. Cummins, "Are males and females equally emotional?," in *Psychology Today*, 2014.