Multimodal Fusion of Physiological Signals and Facial Action Units for Pain Recognition

Saurabh Hinduja[1], Shaun Canavan[1] and Gurmeet Kaur[2]
[1]Department of Computer Science and Engineering, [2]MUMA College of Business, University of South Florida, Tampa, Florida

Abstract— **In this paper, we propose a method for pain recognition by fusing physiological signals (heart rate, respiration, blood pressure, and electrodermal activity) and facial action units. We provide experimental validation that the fusion of these signals results in a positive impact to the accuracy of pain recognition, compared to using only one modality (i.e. physiological or action units). These experiments are conducted on subjects from the BP4D+ multimodal emotion corpus, and include same- and cross-gender experiments. We also investigate the correlation between the two modalities to gain further insight into applications of pain recognition. Results suggest the need for larger and more varied datasets that include physiological signals and action units that have been coded for all facial frames.**

I. Introduction

Assessing pain can be difficult and subject self-report is the most widely used form of assessment. When it is reported in a clinical setting, the measures are subjective and no temporal information is available [10]. Automatic methods to recognize pain have the potential to improve quality of life and provide an objective means for assessment. Considering this, in recent years there has been encouraging works, in this direction, that include analysis of facial expressions, physiological signals, and kinematics and muscle movement.

To facilitate advances in automatically recognizing pain, Aung et al. [1] developed the EmoPain dataset which contains multi-view face videos, audio, 3D motion capture, and electromyographic signals from back muscles. The dataset contains 22 patients (7 male/15 female) with chronic back pain and 28 healthy control subjects (14 male/female). They released baseline results on both facial expressions and muscle movements. Using this dataset, Wang et al. [18] proposed the deep learning architecture, BodyAttentionNet, to recognize protected behavior [17]. This architecture learns temporal information including which body parts are better suited to this task. They showed improved recognition accuracies, as well as a model with 6 to 20 times fewer parameters compared to state of the art. Olubbade et al. [13] proposed a method for automatically recognizing pain levels using kinematics and muscle activity. They trained a random forest and support vector machine, with data from the EmoPain dataset, showing accurate results for recognizing low and high pain, as well as as healthy control levels.

Fabiano et al. [8] proposed a method for fusing physiological signals for emotion recognition. They developed a weighted fusion approach, showing that using the fused signals can accurately detect pain, in the BP4D+ dataset [19], 98.48% of the time. Olugbade et al. [14] investigated how affective factors in chronic pain interfere with daily functions. They found that movement data can recognize distinct distress and pain levels. Lucey et al. [10] proposed an active appearance model-based approach to detect pain in videos, from the the UNBC-McMaster shoulder pain dataset [11], using facial action units (FACS) [7]. Their approach is encouraging, showing that action units can be used to detect pain, motivating the use of this modality here.

Motivated by these works, we propose a method for recognizing pain with the fusion of physiological data (e.g. heart rate, respiration, blood pressure, and electrodermal activity) and facial action units [7]. Using the BP4D+ dataset [19], we fuse action units from the most expressive part of the face, along with physiological signals that have been synced to have a one-to-one correspondence with each facial image that contains action units. We conduct experiments to validate the utility of the proposed fusion method, showing the fused signals result in a positive impact to the accuracy of detecting pain. We further analyze the correlations between the modalities giving insight into future applications of pain recognition. The contributions from this work are 3-fold and can be summarized as follows:

1) A method for recognizing pain through the fusion of physiological signals and action units is proposed.
2) Insight into the correlations between physiological signals and action units is shown.
3) Cross-gender (female vs. male) experiments are conducted for pain recognition, as well as correlations between modalities.

II. Fusion and Experimental Design

To recognize pain, we propose to fuse physiological signals and action units (AUs) from the most facially expressive segments of sequences of tasks meant to elicit emotion. The most expressive segment is defined as the frames where Facial Action Units (AUs) [7] have been manually annotated by experts. To facilitate our experimental design, we use the BP4D+ multimodal emotion corpus [19].

A. Multimodal Dataset (BP4D+)

BP4D+ [19] is a multimodal dataset that include thermal and RGB images, 2D and 3D facial landmarks, manually coded action units (33 total), 4D facial models, and 8 physiological signals (*diastolic blood pressure, mean blood pressure, EDA, systolic blood pressure, raw blood pressure,*

TABLE I: Pain recognition results using physiological signals, action units, and the fusion of both. M=Male; F=Female.

Modalities	All Subjects		Male		Female		Trained (M) / Tested (F)		Trained (F) / Tested (M)	
Metric	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score
Physiological	77.70%	0.3	75.25%	0.285	76.98%	0.269	69.14%	0.35	75.43%	0.219
Action Units	89.02%	0.734	88.00%	0.668	90.73%	0.778	88.27%	0.725	87.50%	0.753
Fused	89.20%	0.75	88.58%	0.689	91.35%	0.787	88.27%	0.725	86.31%	0.75

pulse rate, respiration rate, and respiration volts). There are 140 subjects (58 male and 82 female) with an age range of $18-66$. Action units are manually coded on the most expressive frames of each sequence (task). Emotions are elicited through 10 tasks that the subjects perform. In this work, as we are focusing on pain, our experiments focus on the cold compressor task (i.e. the subjects place their hand in a bucket of ice water for an extended time). We use the physiological signals, and AUs for 139 of the subjects in our experimental design as Subject F082 has some missing features, therefore this subject is removed which follows the experimental design from Fabiano et al. [8]. Approximately 20 seconds of the most expressive frames are coded, and four tasks, that is, the target emotions of happy, embarrassment, fear and pain are coded for action units. Considering this, we use these four emotions for our experiments.

B. Syncing Physiological Signals with AUs

The frame rate of the video recordings in BP4D+ are 25 frames per second (fps) where as the sampling rate of the physiological signals is approximately 1000 frames per second. Due to this difference in sample rate between the AU frames and physiological signals, synchronization is needed to perform fusion of the modalities. To do this, we first calculate the number of frames in the sequence and then down sample the physiological signals, to that same number, using the one step bootstrapping technique [15]. Using this technique, on physiological signals, reduces the number of samples while retaining the important information in the signals [8]. This then gives us a one-to-one correspondence (i.e. synced) between each video frame and physiological signal. It is important to note that the video sequences are not the same length (i.e. different number of frames across subjects). Considering this, we do a final re-sample, of the most expressive segments, to 5000 frames. We use the physiological signals and AUs from these 5000 synced frames to facilitate our fusion approach.

C. Fusion of Physiological Signals and AUs

Given the 5000 frames of synced physiological signals and AUs, from the most expressive segments, we concatenate them to form a new feature vector. For each frame, we concatenate 33 AUs and 8 physiological signals, giving the feature vector, $f = [AU_1, \ldots, AU_{33}, Phys_1, \ldots, Phys_8]$. Where AU_i are the 33 AUs and $Phys_j$ are the 8 physiological signals as detailed in Section II-A. As we want to incorporate temporal information into our pain detection approach, we then concatenate *f* from each of the 5000 synced frames giving us our final feature vector, $F = [f_1, \ldots, f_{5000}]$. This final concatenation results in a feature vector of size $205,000$ (41×5000), which is then used to recognize pain. The proposed fusion results in 1 feature vector for each sequence giving us a total of 556 feature vectors across the 4 tasks (139 subjects $\times 4$ tasks).

D. Pain Recognition Experimental Design

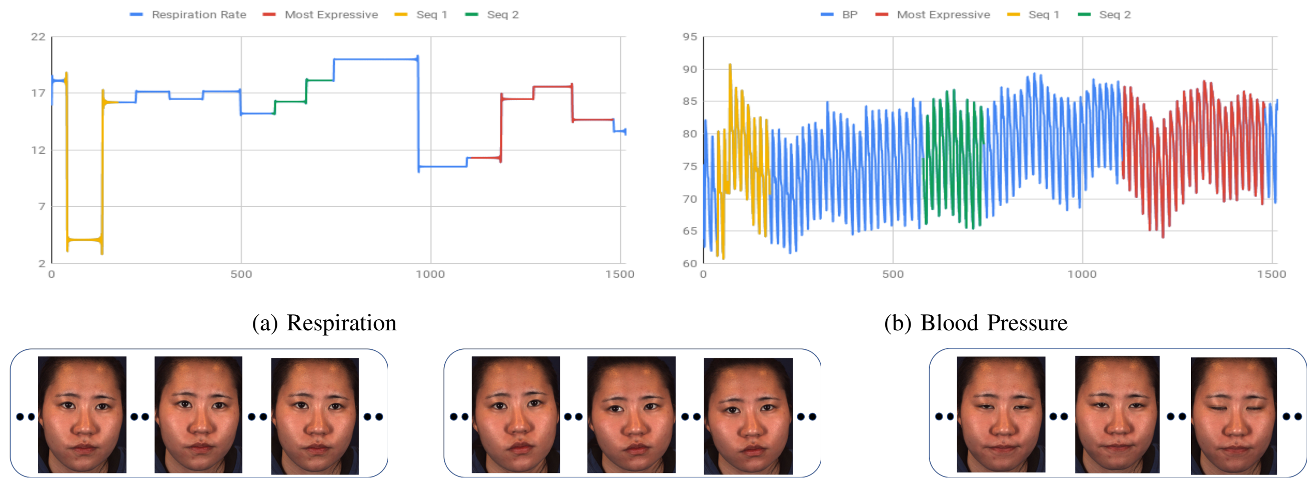
As detailed in Section II-A, BP4D+ has sequences from 4 tasks that have been manually AU coded (happy, embarrassment, fear and pain). As we are interested in recognizing pain, we treat the pain sequences as our positive classes (i.e. pain), and data from the other 3 sequences as our negative class (i.e. no pain). This experimental design is consistent with other works that have investigated pain recognition on BP4D+ [8]. Using this experimental design, we train a random forest classifier [5], using the feature vectors, F, to recognize pain vs. no pain. The random forest consisted of 275 trees, and we validated our approach using subject-independent 10 fold cross-validation on all subjects (139). Along with training and testing on all subjects, we all performed cross- and gender-specific experiments. For each of these experiments, we evaluated each independent modality, that is, physiological signals and action unit independently, as well as the proposed fusion approach (Section II-C).

III. Results

A. Pain Recognition

For our pain recognition experiments, we report overall accuracy, as well as the F1-score for each experiment. For our subject independent experiments on all tested subjects, we achieved an accuracy of 89.2% and F1-Score of 0.75 when fusing the physiological signals and action units. Although physiological signals performed reasonably well as a single modality, with 77.7% accuracy, their F1-Score is low with 0.3. Using this modality, the majority of instances were classified as no pain which is the class containing more samples. It is interesting to note though that action units alone achieved an accuracy of 89.02% and F1-Score of 0.734. These results agree with the literature that AUs can be a powerful representation for automatically recognizing pain [10].

For our gender-based experiments, similar results are obtained. When training and testing on male subjects or female subjects the accuracy of the physiological signals is approximately 76%, while the F1-Score is approximately 0.28. Again with the majority of classifications being labeled as no pain. In both cases (male or female) action units performed well with an accuracy of approximately 89% and F1-Score of 0.668 and 0.778 for male and female experiments, respectively. Along with testing and training



(a) Respiration

(b) Blood Pressure

(c) Seq 1

(d) Seq 2

(e) Most Expressive

Fig. 1: Visual comparison between reparation and blood pressure, and facial expressions (i.e. action units) occurring at same time. During Seq 1, the subject has placed their hand in the bucket of ice water. Seq 2 is approximately 20 seconds after placing hand in bucket, and the most expressive sequence occurs after 40 seconds of hand being in bucket.

on male or female, we also conducted cross-gender experiments (e.g. trained on female and tested on male). In these experiments, again the results are similar where physiological has a low F1-Score of approximately 0.28. The interesting result is that the fusion did not help in this scenario. The fusion for training on male and testing on female resulted in the same accuracy of 88.27% and F1-Score of 0.725. When training on female and testing on male, the fusion caused a slight decrease in accuracy of 1.19% compared to action units alone. This can partially be explained by the differences in correlations between male and female subjects when analyzing the most expressive parts of the sequences (See Figs. 2b and 2c). See Table I for all experimental results.

B. *Comparison to State of the Art*

To the best of our knowledge, there is no direct comparison to this work using the selected modalities from BP4D+. Fabiano et al. [8] investigate pain recognition using fused physiological signals from the BP4D+ dataset, however, their approach was not subject independent where we used a subject-independent experimental design. As this is the closest approach to ours, we detail their results compared to ours. Using their proposed approach and a feed-forward neural network, they achieved pain recognition accuracies of 98.48%, however, they also evaluated their method using support vector machine (92.64%), random forest (90.27%), and naïve Bayes. (89.77%). Using a random forest trained on physiological signals only, with a subject-independent approach, we achieved an accuracy of 77.7%. When fusing the physiological signals with AUs, the proposed approach achieves an accuracy of 89.2% which is comparable to the results from Fabiano et al. when using a random forest.

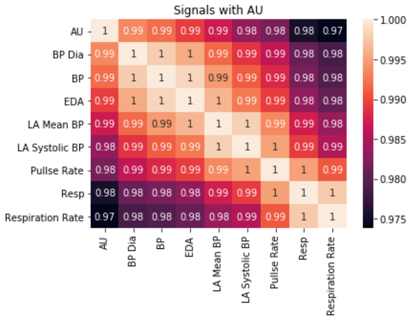
IV. Discussion

From our analysis of physiological signals and AUs, with respect to pain, we make two observations.

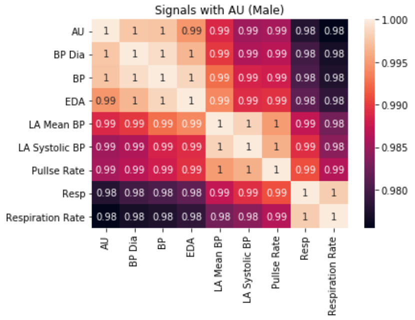
1) The peak for physiological signals and facial expressions occur at different times. (see Fig. 1)
2) Physiological signals are highly correlated during the most expressive parts of the sequence, however, not when the entire signal is analyzed (Fig. 2).

In Fig 1 we have have shown the two physiological signals, respiration (Fig. 1a) and blood pressure (Fig. 1b), and the facial expressions of three segments of the physiological signal. The instant where we see the most variance in physiological signals (Seq 1) there is no facial motor response (Fig 1c) and when the facial expression occurs (Fig 1e), the physiological signals are leveled and there is little variance in the signals. To further analyse this effect of variance of signals at different times during the task, we calculated the correlation [3] between the signals during the most expressive sequence, as well as the entire signal (Fig. 2). As seen in Fig. 2d, the physiological signals during the entire sequence are not well correlated, but when we sample the physiological signals for the most expressive part of the sequence the physiological signals become highly correlated, as seen in Fig. 2a. This increase in correlation can be seen across gender as well. One possible explanation for the low accuracy of pain recognition, using physiological signals, can be this change in variance and correlation of physiological signals during the most expressive part of the video. It has been shown that physiological signals with higher variance tend to perform better at recognizing pain [8].

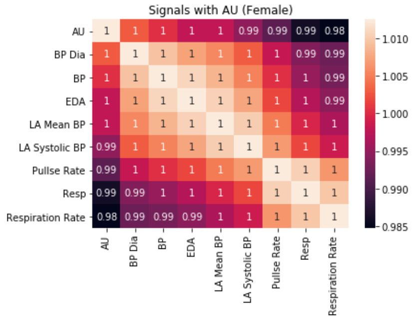
We have seen that exposure to ice cold water is seen to cause physiological changes in the autonomic nervous system – metabolic rate, circulatory system and respiratory system (Fig. 1). This analysis is consistent with results from the medical literature. Circulatory dynamics show effects of different temperatures on the cardiovascular system (Fig. 1b). Immersion in ice cold water stimulates thermoreceptors, activates different regulatory systems (sympathetic nervous



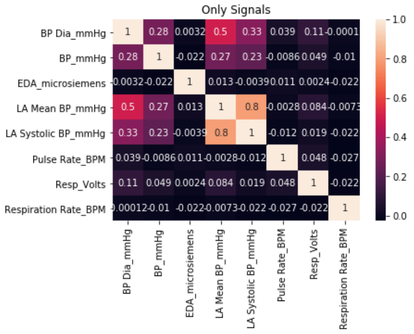
(a) Physiological with avg AU occurrence (Most Expressive Sequence)



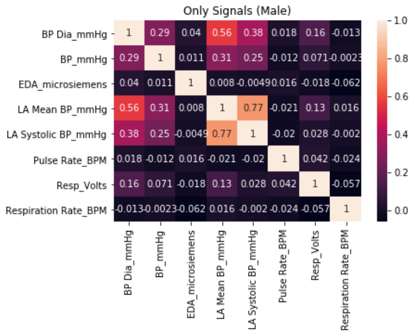
(b) Physiological with avg AU occurrence (Most Expressive Sequence) - Male



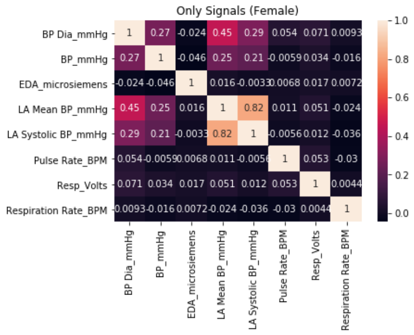
(c) Physiological with avg AU occurrence (Most Expressive Sequence) - Female



(d) Complete Physiological Signal



(e) Complete Physiological Signal - Male



(f) Complete Physiological Signal - Female

Fig. 2: Correlation matrices of physiological signals and action units. Top row, of matrices, shows the most expressive parts of the sequence. Bottom row, of matrices, shows the entire physiological signal, without AUs.

system and endocrine function), different effector mechanisms and increases heart rate, systolic blood pressure and diastolic blood pressure [16]. Increase in heart rate is seen to continue or show maintenance until the exposure to ice cold water is discontinued [12]. 1-mm immersion of hand or foot may show increase in systolic and diastolic blood pressure by 10 - 20 mm/Hg [9]. Immersion in ice cold water has significant effects on the autonomic nervous system. Breathing slows down to retain carbon dioxide and result in acidosis when exposed to low temperatures. This occurs due to mild alterations in the brain stem neuronal systems [6]. Voluntary movements which include motor actions and emotional expression are dependent on the sensory stimulus which occurs due to the peripheral connection with the brain [2]. Objects emotionally felt are the result of action -reaction of the cerebral cortex and diencephalon [2].

It is seen that when subjects are exposed to ice cold water, a spike in physiological changes occurs first, that is increase in heart rate, blood pressure and decrease in respiratory rate which is followed by motor response, that is facial expressions. This is due to the reflex arc mechanism which explains the synapse between sensory and motor neurons. When exposed to external stimuli, sensory signals are sent to the central nervous system through the nerve pathway which causes changes at the physiological levels and relays back the signals to the motor neurons and the motor actions occur [4]. Considering this, these results suggest the need for physiological signals and the coded AUs for the entire sequences, for improved pain recognition applications.

V. Conclusion

We have proposed a multimodal approach to detect pain, using physiological signals fused with action units. We have conducted experiments, on the BP4D+ dataset, using single modalities, as well as the fusion of both. Experiments also include same- and cross-gender validation. Using the most facially expressive segments from sequences shows that physiological data results in a lower F1-Score, while action units give significantly improved results, and the fusion of both further improve accuracies when evaluations include all subjects or same gender. Along with our experimental design, we have also given an in-depth analysis of the correlation between the physiological signals and the action units. This analysis shows that there is a high correlation between these modalities during the most expressive segments of the sequences, however, when the entire sequence is analyzed the correlation decreases. Due to the temporal nature of emotion (Fig. 1) and the differences in correlations between most expressive sequences and others, we suggest larger and more varied datasets that include physiological signals, as well as facial images that have action units coded for all sequences.

Acknowledgment

This material is based on work that was supported in part by an Amazon Machine Learning Research Award.

References

[1] M. S. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, et al. The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset. *IEEE transactions on affective computing*, 7(4):435–451, 2015.

[2] P. Bard. The neuro-humoral basis of emotional reactions. *Handbook of General Experimental Psychology*, 1934.

[3] J. Benesty, J. Chen, Y. Huang, and I. Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[4] E. Bredo. Evolution, psychology, and john dewey's critique of the reflex arc concept. *The Elementary School Journal*, 98(5):447–466, 1998.

[5] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] A. Datta and M. Tipton. Respiratory responses to cold water immersion: neural pathways, interactions, and clinical consequences awake and asleep. *Journal of Applied Physiology*, 100(6):2057–2064, 2006. PMID: 16714416.

[7] P. Ekman. What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (facs). *Oxford University Press*, 1997.

[8] D. Fabiano and S. Canavan. Emotion recognition using fused physiological signals. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 42–48. IEEE, 2019.

[9] W. Lovallo. The cold pressor test and autonomic function: A review and integration. *Psychophysiology*, 12(3):268–282, 1975.

[10] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(3):664–674, 2010.

[11] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Face and Gesture 2011*, pages 57–64. IEEE, 2011.

[12] L. Mourot, M. Bouhaddi, and J. Regnard. Effects of the cold pressor test on cardiac autonomic control in normal subjects. *Physiological research*, 58(1), 2009.

[13] T. A. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. C. Williams. Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 243–249. IEEE, 2015.

[14] T. A. Olugbade, A. Singh, N. Bianchi-Berthouze, N. Marquardt, M. S. Aung, and A. C. D. C. Williams. How can affect be detected and represented in technological support for physical rehabilitation? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(1):1–29, 2019.

[15] W. R. Schucany and S. Wang. One-step bootstrapping for smooth iterative procedures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):587–596, 1991.

[16] P. Šrámek, M. Šimečková, L. Janský, J. Šavlíková, and S. Vybíral. Human physiological responses to immersion into water of different temperatures. *European Journal of Applied Physiology*, 81(5):436–442, Feb 2000.

[17] M. J. Sullivan, P. Thibault, A. Savard, R. Catchlove, J. Kozey, and W. D. Stanish. The influence of communication goals and physical demands on different dimensions of pain behavior. *Pain*, 125(3):270–277, 2006.

[18] C. Wang, M. Peng, T. A. Olugbade, N. D. Lane, A. C. D. C. Williams, and N. Bianchi-Berthouze. Learning bodily and temporal attention in protective movement behavior detection. *arXiv preprint arXiv:1904.10824*, 2019.

[19] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. Cohn, Q. Ji, and L. Yin. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.