Multimodal Analysis of Physiological Signals for Wearable-Based Emotion Recognition Using Machine Learning

Feryal A Alskafi[1], Ahsan H Khandoker[1], Uichin Lee[2], Cheul Young Park[2], Herbert F Jelinek[1]

[1]Khalifa University, Abu Dhabi, United Arab Emirates
[2]Korea Advanced Institute of Science and Technology, Daejeon, South Korea

Abstract

Recent advancements in wearable technology and machine learning have led to an increased research interest in the use of peripheral physiological signals to recognize emotion granularity. In healthcare, the ability to create an algorithm that classifies emotion content can aid in the development of treatment protocols for psychopathology and chronic disease. The non-invasive nature of peripheral physiological signals however is usually of low quality due to low sampling rates. As a result, single-mode physiological signal-based emotion recognition shows low performance. In this research, we explore the use of multi-modal wearable-based emotion recognition using the K-EmoCon dataset. Physiological signals in addition to self-reported arousal and valence records were analyzed with a battery of datamining algorithms including decision trees, support vector machines, k-nearest neighbors, and ensembles. Performance was evaluated using accuracy, true positive rate, and area under the receiver operating characteristic curve. Results support the assumption with 83% average accuracy when using an ensemble bagged tree algorithm compared to single heart rate-based emotion accuracy of 56.1%. Emotion granularity can be identified by wearables with multi-modal signal recording capabilities that improve diagnostics and possibly treatment efficacy.

1. Introduction

Understanding emotion in quantifiable terms rather than the traditional qualitative manner used by psychologists has been one of the aims of wearable, personalized technologies research as part of the IoT [1]. Though humans usually construe emotions as body language, they are essentially the result of neuronal and hormonal activity that translates into changes in diverse physiological signals such as cardiac rhythm, peripheral temperature, electrodermal activity and respiratory rhythm. Emotions and the associated physiological systems are modulated by autonomic nervous system (ANS) as well as higher cortical centres establishing specific physiological networks. Through hypothalamic-brainstem connections, the cerebral cortex and limbic system affect ANS activities linked with emotional reactions. Vasodilation of the blood vessels, fainting, cold chills, and a rapid heart rate are examples of physiological changes associated with emotional reactions [2].

Data mining has become an integral part of medicine including exploration of large clinical datasets as well as features associated with diverse physiological signals [3]. Hierarchical processing has further led to improvements in prediction accuracy for clinical decision making and emotion classification [4]. The current research applied hierarchical datamining methods to predict emotions as arousal and valence during a debate using the K-EmoCon dataset.

2. Experimental Method

2.1. Data

The K-EmoCon dataset chosen to conduct this research contains comprehensive annotations of continuous emotions during a naturalistic debate. Specifically, the context for data collection, was a semi-structured, turn-taking debate between male and female students (age range: 19 to 36) about a social topic with randomly assigned partners, allowing for gathering emotions similar to those which may naturally develop throughout the course of the day. The setting's formality and spontaneity however, forced individuals to manage their emotions in a socially acceptable manner [5]. The data used for the emotion classification was blood volume pulse (BVP), electrodermal activity (EDA), heart rate (HR), and skin temperature (SKT) signals collected using the *Empatica E4 Wristband* alongside the first-hand self-reported annotations of arousal and valence.

2.2. Emotion Annotations

All emotional states are represented by two fundamental neurophysiological attributes, one connected to valence and the other to arousal, according to Russel's circumplex

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.328

model of affect, with each emotion may be represented as a linear combination of these two dimensions [6]. Therefore, it is quite important to consider emotions in this two-dimensional space in any emotion recognition application. Arousal and valence are each considered as having either low or high states where 1 and 2 labels are considered low and 3, 4, and 5 labels are considered high. The intersection of the 2-class arousal and the 2-class valence model was considered, which creates four quadrants into which emotions are classified (Figure 1). It is important to note that the correct identification of emotions in the HALV and the LALV quadrants is essential as that is where the emotions most associated with stress and illness occur such as sadness, depression, and anxiety.

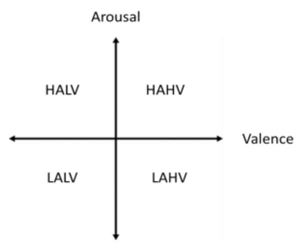

Figure 1. The intersection of low/high arousal with low/high valence creates four quadrants: High Arousal/High Valence (HAHV), High Arousal/Low Valence (HALV), Low Arousal/Low Valence (LALV), and Low Arousal/High Valence (LAHV).

2.3. Signal Pre-processing

Minimal pre-processing was applied to the physiological signals from the K-EmoCon recordings prior to input to the machine learning models. Resampling to 1 Hz was applied to the 64 Hz BVP and the 1 Hz HR signals to match the EDA and SKT signals using MATLAB retime function with a linear interpolation method and all signals were synchronized. After resampling, the four signals are stacked to form an array corresponding to a 5-second segment matching the annotation frequency. In addition, the four signals had varying means and range of values which could slow down the learning and convergence of the models as variables with larger ranges will dominate over those with small ranges leading to biased results. To overcome this issue, the training set mean, and standard deviation were used to standardize the training and testing sets (Figure 2). The minimal pre-processing approach applied here was carried out to investigate the possibility of using such models in real-time emotion recognition applications, which require low computing costs.

2.4. Classification Models and Platform

The experimental setup included the implementation of machine-learning algorithms in MATLAB. Decision trees

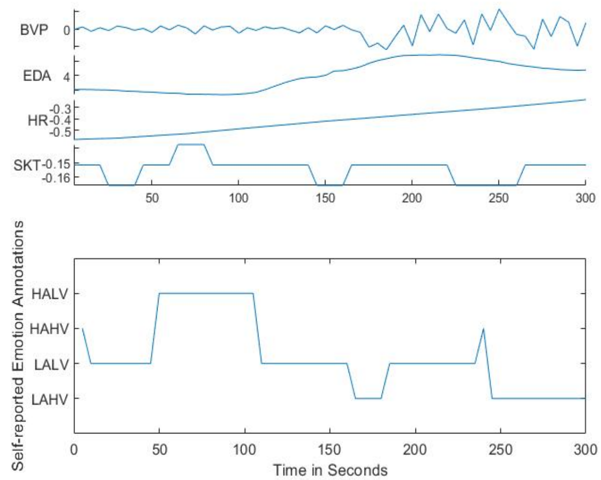

Figure 2. A 5-minute excerpt of BVP, EDA, HR, and SKT signals for one of the participants after synchronization and standardization along with corresponding emotion annotations.

(fine, medium, and coarse), support vector machines (linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian), k-nearest neighbours (fine, medium, coarse, cosine, cubic, and weighted), and ensembles (boosted trees, bagged trees, subspace KNN, and RUSBoosted Trees) classifiers were trained using the pre-set hyperparameters as a starting point coupled with a 5-fold cross-validation scheme. Signals were first used individually as predictors then combinations of two signals and then three as indicated by principal component analysis (PCA). Finally, a multi-modal approach was tested by using all four signals.

Rather than relying on the measure of conventional accuracy alone, which in class imbalance problems as found with the current dataset, varies greatly [7], other measures were also considered. True Positive Rate (TPR) was used to assess how correctly each model classified instances in each class since the performance of the model in low occurring classes is of interest. The area under the receiver operating characteristic curve (AUC) was used to compare the different machine learning models, with a value of 0.5 indicating no discrimination, 0.7–0.8 good, 0.8–0.9 excellent, and higher than 0.9 exceptional discrimination [8].

3. Results

From the accuracy results in Figure 3 and average TPR and AUC values in Table 1, using only BVP signals resulted in an average accuracy of 44.59% for all models tested. Maximum average TPR was 28.15% when Ensemble RUSBoosted Trees was used. AUC values for all the models tested averaged at 0.51 accuracy.

On the other hand, EDA signals resulted in a 48.84%

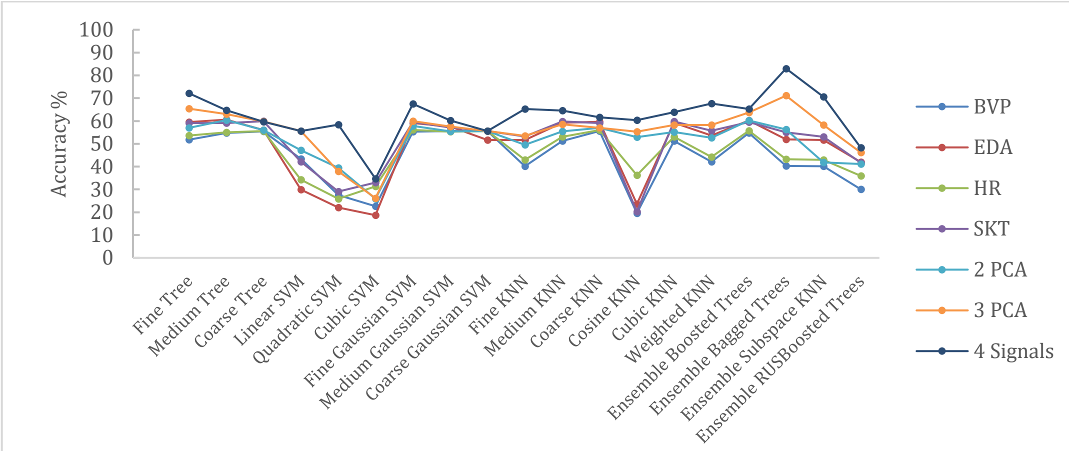

Figure 3. Average Accuracy results for individual signal analysis when using the different machine learning algorithms.

average accuracy with the highest accuracy being 60.6% when using a medium tree algorithm. Average TPR was 33.35% with maximum average TPR being 52.8% when Ensemble RUSBoosted Trees was used. AUC values ranged between 0.45 and 0.74 with an average of 0.62.

Using HR signals gave an average accuracy of 46.65%, with the highest accuracy at 56.1% when using a coarse tree algorithm. As with BVP and EDA, the highest average TPR was achieved using Ensemble RUSBoosted Trees and maximum average AUC was 0.54.

SKT gave an average accuracy of 51.14%, with the highest accuracy at 59.9% when using a coarse tree algorithm. Consistently, the highest average TPR was found using Ensemble RUSBoosted Trees, which was also reflected in the corresponding average AUC value of 0.75.

PCA was used to find the percent variability explained by the principal components of the four signals. BVP and EDA signals together explained 68.13% of all variability while BVP, EDA, and HR together explained 99.99% of all variability.

The use of the combination of BVP and EDA signals as predictors gave an average accuracy of 51.43% for all the classification algorithms tested, with the highest accuracy at 60.4% when using a medium tree algorithm. Maximum average TPR and average AUC were 52.4% and 0.74, respectively, both using Ensemble RUSBoosted Trees.

For BVP, EDA, and HR together, the average accuracy for all the classification algorithms tested was 55.84%, with the highest accuracy at 83% when using an ensemble bagged trees algorithm. Maximum average TPR was 58.05% using Ensemble RUSBoosted Trees while maximum AUC was 0.85 using ensemble bagged trees algorithm.

Using the four signals as predictors produced the best overall results. The ensemble bagged trees algorithm, in particular, was not only able to produce higher accuracy at 83%, but it also discriminated between the four emotion classes. Specifically, the true positive rate for the LALV, HALV, LAHV, and HAHV classes was 62.2%, 59%, 82.1%, and 90.5%, respectively.

4. Discussion

When using the BVP signal as the only predictor, all algorithms failed to display any discrimination between classes, indicating that on its own, the BVP signal was not enough to support emotion recognition. EDA did better in detecting the low occurring classes and in classification overall, with average AUC values of about 0.7 as opposed to 0.6 with BVP. Still, the discrimination between classes was not sufficient. The use of HR as the only predictor was similar to that of BVP in terms of not showing any discriminating abilities. The SKT signal did relatively well in detecting the LALV class, however, it came at the cost of loss of accuracy in detecting the other three classes. The combined use of BVP, EDA, HR, and SKT signals as predictors produced the best results. The ensemble bagged trees approach, not only had a high level of accuracy at 83% that was in line with accuracies of recent wearable-based emotion recognition studies [9], but also had a high level of discrimination when it came to class prediction. When fewer signals were employed, the ensemble RUSBoosted trees fared well in properly categorizing the low occurring class, but it became biased towards the other classes, which is why the overall accuracy was poor.

5. Conclusion

Studies on wearable-based emotion recognition have progressed in search of an economic system that can be utilized for non-invasive long-term monitoring. The size and power constraints of the sensors used put limits on the qualities of individual signals, resulting in poor performance when used in emotion recognition models. In this research, a multi-modal approach was proposed to overcome the aforementioned issues, by utilizing BVP, EDA, HR, and SKT signals from the K-EmoCon dataset.

Table 1. Average TPR% and AUC results across all four classes when using the different machine learning algorithms.

Classifier	BVP		EDA		HR		SKT		2 PCA		3 PCA		4 Signals	
	TPR	AUC	TPR	AUC	TPR	AUC	TPR	AUC	TPR	AUC	TPR	AUC	TPR	AUC
Fine Tree	24.18	0.51	37.18	0.72	28.90	0.55	36.68	0.72	36.70	0.70	44.63	0.76	51.80	0.81
Medium Tree	24.90	0.53	35.95	0.68	28.35	0.54	35.45	0.67	35.88	0.67	36.38	0.68	37.73	0.73
Coarse Tree	25.00	0.51	28.98	0.59	28.75	0.53	33.98	0.64	28.98	0.59	32.68	0.61	33.18	0.64
Linear SVM	25.33	0.50	24.70	0.49	24.85	0.50	24.55	0.49	24.85	0.50	25.00	0.49	25.00	0.54
Quadratic SVM	25.70	0.51	25.73	0.51	24.30	0.52	25.15	0.51	25.15	0.51	25.28	0.52	29.15	0.68
Cubic SVM	25.53	0.51	24.90	0.50	26.80	0.50	23.00	0.49	25.13	0.51	26.28	0.54	34.95	0.69
Fine Gaussian SVM	24.88	0.50	29.60	0.61	28.15	0.59	32.18	0.70	28.58	0.60	31.83	0.68	43.30	0.81
Medium Gaussian SVM	24.98	0.50	28.08	0.57	25.65	0.57	29.30	0.66	24.95	0.58	27.38	0.62	31.28	0.73
Coarse Gaussian SVM	25.00	0.52	25.00	0.59	25.00	0.53	25.00	0.57	25.00	0.56	25.00	0.60	25.00	0.66
Fine KNN	12.23	0.50	37.45	0.59	28.20	0.52	39.90	0.60	33.95	0.57	39.75	0.60	54.58	0.70
Medium KNN	24.78	0.50	47.37	0.72	27.83	0.57	40.08	0.74	30.43	0.63	34.10	0.67	44.20	0.79
Coarse KNN	25.00	0.53	32.60	0.72	27.85	0.58	32.53	0.73	27.45	0.60	27.98	0.62	33.28	0.72
Cosine KNN	25.45	0.51	22.13	0.45	23.60	0.48	22.40	0.44	27.28	0.54	29.73	0.63	39.98	0.73
Cubic KNN	24.78	0.50	38.58	0.72	27.83	0.57	40.08	0.74	30.28	0.63	33.53	0.67	43.15	0.78
Weighted KNN	25.00	0.51	38.18	0.64	28.40	0.54	40.50	0.66	33.33	0.63	38.43	0.68	52.70	0.81
Ensemble Boosted Trees	24.70	0.53	35.68	0.73	28.75	0.58	35.05	0.72	35.58	0.72	36.00	0.76	39.10	0.80
Ensemble Bagged Trees	24.80	0.50	31.35	0.65	28.30	0.54	41.33	0.69	39.05	0.69	57.63	0.85	**73.45**	**0.94**
Ensemble Subspace KNN	24.75	0.50	37.43	0.59	28.20	0.52	39.60	0.60	26.58	0.56	41.60	0.68	53.35	0.84
Ensemble RUSBoosted Trees	**28.15**	0.52	**52.80**	0.74	**36.13**	0.59	**55.23**	0.75	**52.40**	0.74	**58.05**	0.80	61.03	0.83

The results support the use of multimodal signal analysis for detection of emotion granularity defined by valence and arousal. In addition, the minimum preprocessing proposed here was sufficient for obtaining highly accurate results that pave the way for routine clinical use. The findings of this paper will be used as a basis for further testing into optimizing the hyperparameters of the models for better performance.

Acknowledgments

This work was supported by a KU-KAIST joint grant (8474000221;KKJRC-2019-Health2 (Khandoker)) from Khalifa University Abu Dhabi, UAE

References

[1] Bota PJ, Wang C, Fred AL, Da Silva HP. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. IEEE Access. 2019 Sep 26;7:140990-1020.

[2] Levenson RW. The autonomic nervous system and emotion. Emotion review. 2014 Apr;6(2):100-12.

[3] Abawajy J, Kelarev A, Chowdhury MU, Jelinek HF. Enhancing predictive accuracy of cardiac autonomic neuropathy using blood biochemistry features and iterative multitier ensembles. IEEE journal of biomedical and health informatics. 2014 Oct 20;20(1):408-15.

[4] Wang Y, Chu YM, Thaljaoui A, Khan YA, Chammam W, Abbas SZ. A multi-feature hybrid classification data mining technique for human-emotion. BioData Mining. 2021 Dec;14(1):1-20.

[5] Park CY, Cha N, Kang S, Kim A, Khandoker AH, Hadjileontiadis L, Oh A, Jeong Y, Lee U. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Scientific Data. 2020 Sep 8;7(1):1-6.

[6] Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and psychopathology. 2005 Sep;17(3):715-34.

[7] Lee CC, Mower E, Busso C, Lee S, Narayanan S. Emotion recognition using a hierarchical binary decision tree approach. Speech Communication. 2011 Nov 1;53(9-10):1162-71.

[8] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology. 2010 Sep 1;5(9):1315-6.

[9] Schmidt P, Reiss A, Dürichen R, Laerhoven KV. Wearable-based affect recognition—A review. Sensors. 2019 Jan;19(19):4079.

Address for correspondence:

My Name. Feryal Amjad Alskafi
My Full postal address. P.O. box 127788, Khalifa University, Abu Dhabi, UAE
My E-mail address. 100057997@ku.ac.ae