Calibrating the Classifier: Siamese Neural Network Architecture for End-to-End Arousal Recognition from ECG

Andrea Patanè and Marta Kwiatkowska

Department of Computer Science, University of Oxford
andrea.patane@cs.ox.ac.uk
marta.kwiatkowska@cs.ox.ac.uk

Abstract. Affective analysis of physiological signals enables emotion recognition in mobile wearable devices. In this paper, we present a deep learning framework for arousal recognition from ECG (electrocardiogram) signals. Specifically, we design an end-to-end convolutional and recurrent neural network architecture to (i) extract features from ECG; (ii) analyse time-domain variation patterns; and (iii) non-linearly relate those to the user's arousal level. The key novelty is our use of a shared-parameter siamese architecture to implement user-specific feature calibration. At each forward and backward pass, we concatenate to the input a user-dependent template that is processed by an identical copy of the network. The siamese architecture makes feature calibration an integral part of the training process, allowing modelling of general dependencies between the user's ECG at rest and those during emotion elicitation. On leave-one-user-out cross validation, the proposed architecture obtains +21.5% score increase compared to state-of-the-art techniques. Comparison with alternative network architectures demonstrates the effectiveness of the siamese network in achieving user-specific feature calibration.

Keywords: Emotion Recognition, Electrocardiogram, Siamese Neural Network, Convolutional and Recurrent Neural Network.

1 Introduction

Driven by applications in mobile mental health and human-computer interaction [1], affective analysis of physiological signals has recently grown in popularity. Since the pioneering use of electrodermal activity for arousal detection, the research has evolved to cater for a range of physiological signals, such as electrocardiogram (ECG), electroencephalogram, electromyogram, breath rhythm and skin temperature [1]. However, while much effort has focused on multi-modal sensor fusion, model performance on single signal sources is still sub-optimal. At the same time, achieving performance improvement for single sensors can push accuracy boundaries for the overall model architecture even further, potentially leading to increased wearability of emotion recognition systems.

The ECG signal, in particular, has become a focus of investigations because of its unobtrusiveness, low cost and widespread availability of ECG sensors, as well as sensitivity to both arousal and valence component of emotions [2]. Existing state-of-the-art machine learning pipelines for emotion recognition from ECG signals usually proceed by extracting the HR (Heart Rate) signal and applying sophisticated HRV (Heart Rate Variability) analysis techniques in a multi-step process. This is mainly composed of: (i) HRV feature extraction; (ii) automatic feature selection; (iii) user-specific feature calibration; (iv) hyper-parameter optimisation; and (v) model fitting. While steps (iv) and (v) are those actually involved in model estimation, the overall performance of the resulting model mainly depends upon the effectiveness of steps (i) to (iii), as testified by the extensive literature on feature extraction, selection and calibration for HRV analysis [2–6]. While the feature extraction and selection steps are focused on extracting the most informative features from the HR signal, user-specific feature calibration crucially strives to enforce *relative* variation of feature values in the model, rather than *absolute* variation, as the former are related to changes in the user's affective state. Furthermore, the features based on HRV are the *only* type of features extracted from the ECG signal, and thus affective information carried by most of the ECG signal is completely neglected [8–10].

In this work we pose the arousal recognition problem as a supervised classification problem and investigate the use of deep learning for arousal recognition from ECG. For this purpose, we design a deep Convolutional and Recurrent Neural Network (CRNN) architecture that (through end-to-end training) automatically extracts general non-linear and time-domain features from the time-series ECG signal and non-linearly relates those to specific arousal classes based on common variation patterns found. Inspired by state-of-the-art HRV-based machine learning pipelines, we propose the use of *shared-parameter siamese* neural network architectures [16], called the Siamese CRNN (S-CRNN), as a systematic way to extend and generalise feature calibration techniques into the deep learning framework. By making feature calibration an integral part of the end-to-end learning process, we allow the neural network to model general nonlinear dependencies between the user's ECG signal at rest and that during emotion elicitation experiments. Namely, at each forward and backward pass through the network one branch of the S-CRNN processes a new data sample, while the other S-CRNN branch analyses a *template* sample specific to the user's neutral affective state. We use truncated back-propagation through time and stochastic gradient descent to train the network in the classification problem associated to the user's arousal level.

We compare the S-CRNN architecture against state-of-the-art HRV analysis pipelines on the classification task associated to a dataset for arousal recognition during a real-world driving task [15]. The results obtained empirically demonstrate the advantages of the end-to-end approach for arousal recognition from the ECG signal. Namely, on leave-one-user-out cross validation settings the S-CRNN architecture obtains average AUCs percentage increase of +21.5% on the best results obtained by HRV analysis (that is, from 0.659 to 0.801). We further

analyse the proposed S-CRNN against alternative architectures and approaches for feature calibration and find that the approach based on shared parameter siamese neural networks leads to a +7.5% performance increase compared to the corresponding CRNN, at the cost of negligible increase in network parameters.

Contributions. The paper makes the following contributions.

- We propose an end-to-end classification framework for arousal recognition from ECG. We design a CRNN that automatically extracts features from ECG and analyses time patterns, among them, relating them to arousal classes.
- We investigate the use of siamese neural networks as a systematic way to implement feature calibration techniques into the deep learning framework.
- We empirically compare the S-CRNN architecture against state-of-the-art HRV analysis methods, observing a +21.5% performance improvement.
- We compare S-CRNN, models based on HRV analyses and alternative network architectures in terms of generalisation performance to new users when very few users are included in the training set. We assess the advantages of the siamese architecture in achieving personalised feature calibration.

Organisation. The remainder of the paper is organised as it follows. In Section 2 we analyse related work in emotion recognition from the ECG signal and the use of deep learning in affective computing. In Section 3 we present the S-CRNN architecture designed for arousal recognition from ECG. Empirical results evaluating the effectiveness of the S-CRNN architecture are discussed in Section 4. Finally, Section 5 completes the paper with a discussion on the method presented, and outlines future work directions.

2 Related Work

In this section, we give a brief overview of machine learning methods developed for HRV analysis and applications of deep learning for affective computing.

2.1 Heart Rate Variability Analysis for Arousal Recognition

Table 1 lists a collection of 31 features generally extracted from HR signal and used for HRV analysis for arousal recognition [2–6]. Machine learning methods based upon HRV analysis are multi-step, including feature selection and user-dependent feature calibration as crucial steps of the model learning.

In fact, Ollander *et al.* [5] investigate extensive feature selection for emotion recognition from biosignals. They extract a number of HRV features, which are then calibrated using mean and standard deviation computed from a set of user-specific neutral affective state measurements. Few of the selected HRV features actually survive the feature selection step. Zhao *et al.* [2] extract several

Domain	Name
Time	Mean, Median, SDNN, pNN50 RMSSD, SDNNi, meanRate, sdRate.
Geometrical	TINN, RRTI, HRVTi.
Frequency	Welch PSD: LF/HF, (LF+MF)/HF , peakLF, peakHF. Burg PSD: LF/HF, (LF+MF)/HF, peakLF, peakHF. L-S PSD: LF/HF, (LF+MF)/HF, peakLF, peakHF.
Poincaré	SD_1, SD_2, SD_2/SD_1.
Nonlinear	$SampEn_1$, $SampEn_2$, DFA_{all}, DFA_1, DFA_2.

Table 1: Type of HRV feature analysis employed for emotion recognition. Full details, including feature extraction algorithms, can be found in [2–5].

features from participants' HR signals and perform feature calibration on them. An SVM model is then trained on the data by using l_1 regularisation for automatic feature selection. Reportedly, only 10 out of 26 extracted features were actually used by the SVM model. Melillo *et al.* [4] extracted 13 HRV features, and applied exhaustive feature selection procedure and linear discriminant analysis. Surprisingly, the resulting classifier relied only upon three of the extracted features. In order to partially overcome the feature selection problem, Gjoreski *et al.* [7] train a multi-layer perceptron to predict arousal level from a PSD of the HR signal. They report improvements over models trained on top of HRV features, albeit the neural network proposed is constrained to use only frequency domain features, and no feature calibration procedure is implemented.

Finally, though most of the above works extract HR from ECG, HRV analysis is the only systematic method used to compute features. Thus, potentially relevant information from most of the ECG signal is ignored [8–10].

2.2 Deep Learning for Affective Computing

Many works have investigated the use of deep learning for face expression classification from images, as well as sentiment analysis of text, with deep learning approaches systematically outperforming other techniques [12].

Martinez *et al.* [11] were among the first to apply end-to-end deep learning for physiological signals' affective processing. They developed a CNN for preference learning from galvanic skin response and blood volume pulse data, and empirically demonstrated the advantages of deep features over manually designed ones. Tripathi *et al.* [17] applied CNNs for arousal recognition from EEG. Empirical results show up to $\approx 14\%$ improvement against methods based on manual feature extraction. Cho *et al.* [23] present a CNN architecture for stress recognition from breathing patterns. Emphasising data augmentation as a crucial step for model training, they obtain substantial improvements over competitive methods.

Our work is a continuation of the latter works that bring deep learning to the field of emotion recognition from physiological signals. The key novelty is the use of siamese networks as a systematic way to implement feature calibration, which is usually overlooked in deep learning frameworks for emotion recognition.

3 Methods

This section discusses our design of the neural network architecture for arousal recognition from ECG signal. First, we describe data pre-processing and augmentation used. We then present the CRNN architecture we designed for feature extraction, and describe the shared-parameter siamese version of the latter.

3.1 Preprocessing

As pre-processing steps we apply a baseline remover filter and standardisation to each ECG signal. We thus segment the signals in fixed-size time windows with 50% overlap. Based on empirical results from ultra-short term HRV analysis [18], we use time windows of 15 seconds, as these provide just enough information to extract significant features from the ECG signal. Though windows of greater size would increase model sensitivity to small feature variations, they would conflict with the practical limitations of the back-propagation through time training algorithm (i.e., increased training time and vanishing gradient problem).

3.2 Data Augmentation

Datasets for emotion recognition from physiological signals are typically of small size, and thus deep models applied to them tend to overfit [7]. Furthermore, real-world datasets related to health applications are notoriously unbalanced, with the class associated to the absence of the disorder usually greatly over-represented in the training data. This makes stochastic gradient descent somewhat challenging, as it will likely get stuck in a local optimum corresponding to a trivial majority classifier. We thus heavily rely upon data augmentation techniques in order to train our CRNN model.

First, we re-balance class labels of the training set by making multiple copies of random representatives from the minorities class until the dataset is perfectly balanced (that is, until each class is equally represented in the dataset). Then, from each signal slice, we generate n training samples. Namely, we randomly sub-sample n times the signal to a fixed size time window of m time points. In doing so, we keep the time-stamps associated to the sub-sampled signal. Hence, loss of information due to sub-sampling is mitigated, as the neural network is potentially able to partially interpolate the missing pieces of the signal. Unless otherwise specified, in the experiments of Section 4 we use $n = 20$ and $m = 1024$.

3.3 Convolutional Recurrent Neural Networks

The proposed model architecture is sketched in Figure 1 and summarised in Table 2. Inspired by state-of-the-art HRV features, the CRNN employs a 3-layer bidirectional RNN to summarise temporal patterns on top of a one-dimensional 6-layer CNN. The CRNN is designed to first extract non-linear features from the ECG signal, and then to analyse temporal information of feature variations.

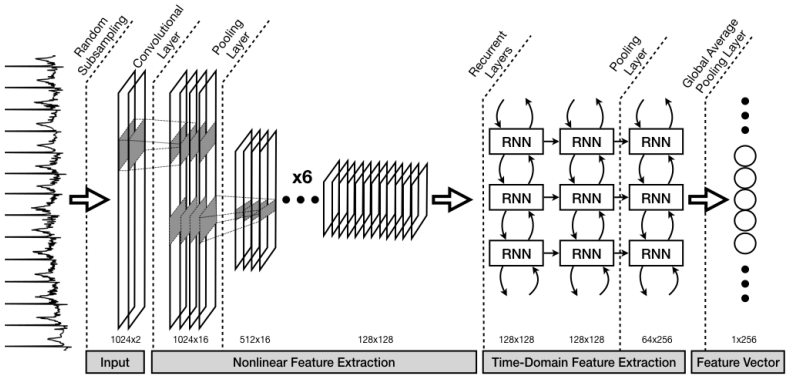


Fig. 1: Proposed CRNN architecture, consisting of 6 convolutional blocks and three stacked bidirectional recurrent neural networks.

Each convolutional block consists of a convolutional layer and a non-linear activation function layer. After every other block, we use a one-dimensional max-pooling layer to extract salient points from feature maps and compress temporal information. Crucially, we employ Parametric ReLU [19] activation functions in between convolutional layers to avoid dead ReLU problems. Parametric ReLU allows automatic learning of the activation slope for negative input, effectively avoiding the issue of fast death of units slowing down the learning procedure. Notice that, because of data augmentation applied to the training set, data distributions for the training set and the test set are systematically different, and hence we cannot rely on batch-normalisation layers (usually used to circumvent dead ReLU problems).

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
1-D Conv. Filters	16	32	32	64	128	128
1-D Conv. Kernel	11	9	9	7	7	7
1-D Max-pooling	✓	✗	✓	✗	✓	✗
Bi-RNN Units	128	128	256	✗	✗	✗
1-D Max-pooling	✗	✓	✗	✗	✗	✗

Table 2: Details of the architectures and hyper-parameters of the CRNNs designed for arousal recognition from ECG. Ticks (respectively crosses) indicates that the layer is included (not included) in the layer block.

We use vanilla RNN units, as we experimentally observe that gated recurrent layers quickly lead to overfitting problems. We speculate that this is due to the small size of the dataset used here compared to datasets usually employed to train deep LSTM and GRU recurrent networks [20, 13]. We use a one-dimensional global average pooling layer to summarise temporal patterns extracted by the recurrent layers. Finally, we interleave dropout layers in between each pair of layers, and only for the non-recurrent connections.

The final output of the CRNN is a vector of nonlinear and time domain features extracted from each time window of the ECG signal. Next, we will discuss how this is used to predict an arousal class from each signal window.

3.4 Siamese Neural Networks

We implement the CRNN inside a shared-parameter siamese architecture [16]. The outline of the siamese network is sketched in Figure 2. At each forward and backward pass through the network, a user-specific template is fed into the network along with the signal window currently analysed. The latter, and the user-specific template, are independently processed by the CRNN, which extracts two separate feature vectors from them. The resulting feature vectors are concatenated into a unique feature map and altogether processed by a fully connected layer. By relying on the fully connected layer, the siamese architecture has the capabilities to use features extracted from the user's template to systematically calibrate those extracted from the current signal sample. Finally, a soft-max layer estimates the probability of the user being in an arousal state.

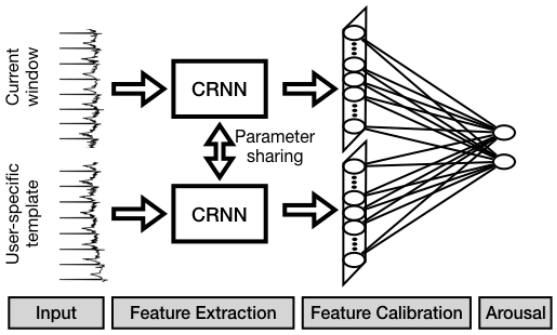


Fig. 2: Shared-parameter siamese architecture for arousal recognition. The current ECG window and the user-specific template are passed through the CRNN. The two feature maps are then concatenated and used to estimate arousal level.

Analogously to methods based on HRV analysis, for the user-specific template we employ a sample recorded from the user before the beginning of the experiment, which is assumed to be representative of the user's neutral affective state. Notice that, in order to mitigate overfitting, we apply data augmentation techniques outlined in Subsection 3.2 also to the users' templates.

4 Results

In this section we describe experiments related to the following key points:

- Comparison of HRV and S-CRNN on arousal recognition.
- Evaluation of the siamese architecture capabilities to implement feature calibration, comparing the S-CRNN with alternative network architectures.

- Analysis of the number of users included in the training set (population size) to asses the effect on the feature calibration layer.
- Sensitivity analysis of hyper-parameters included in our methodology, focusing on data-augmentation and number of convolutional/recurrent layers.

4.1 Dataset

We perform comparisons on the classification task associated to a dataset for arousal recognition made publicly available by Schneegass et al. [15]. Briefly, a set of physiological signals were recorded from 10 users during a real-world driving task. Data samples were then subjectively labelled by each user for arousal/driving workload. Among the signals included in the dataset, we focus on ECG and use the arousal labels to define a binary classification problem (low vs. high arousal).

4.2 HRV-based analyses

We train models based on HRV on the 31 features listed in Table 1 and provide results for a selection of classification methods used in the literature [2–5], that is, k-Nearest Neighbours (K-NN), Linear Discriminant Analysis (LDA), Support Vector Machine with l_1 regularisation (SVM-l1) and Random Forest (RF). We apply state-of-the-art feature selection algorithms and hyper-parameter optimisation to all the techniques based on HRV analysis on a nested cross validation setting. Namely, we use fitting and hyper-parameter optimisation routines implemented in the Matlab machine learning toolbox, and apply forward search, backward search and randomised search for feature selection. For space limitation, for each model we include results only for the best performing combination of parameters/features.

4.3 Experimental Setup

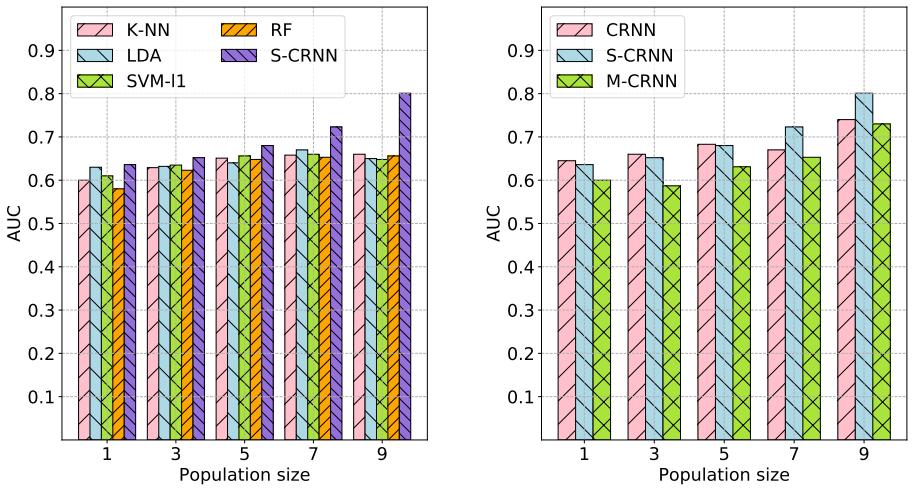
Because of strong class imbalance (only $\approx$ 6% of samples are representative of the arousal class) we compare the results based on AUC score. Results are presented for leave-one-user-out cross validation. We use Keras [21] with TensorFlow [22] backend for implementation and training of neural networks. We train the networks using Adam optimiser [24] up to a maximum of 100 epochs, and use early stopping on a validation set. We do not investigate exhaustive hyper-parameter optimisation for the S-CRNN, as it is nested in a cross validation and would thus lead to prohibitive computational times. Instead, we perform a local hyper-parameter analysis on the most sensitive hyper-parameters (Section 4.6).

We train HRV analysis models on a 2 GHz Intel Core i5 processor with a RAM of 8 GB @1867 MHz. Computational time for a full round of hyper-parameter optimisation and cross validations for each HRV model varied between about 1 and 12 hours. We train deep learning models on NVIDIA Tesla K80 GPU. Computational time for a full round of cross validation took about 60 hours.

4.4 Comparison of HRV and S-CRNN

In Figure 3a we compare average AUCs obtained by different classification models learnt on top of HRV features with the results obtained by the S-CRNN, for an increasing number of users included in the training set. Results for population sizes between 1 and 7 are averaged over 10 randomly chosen combinations of users included in the training set (consistently among models).

As expected, we observe an overall trend for all the methods to perform better as the number of users included in the training set increases. However, the performance boost obtained for all the models when increasing the number of training users from 1 to 5 seems to saturate for HRV-based methods, which fail to take advantage of such increases. On the other hand, the S-CRNN obtains additional AUC boosts when more users are included in the training set. For the largest size of the training set allowed by the dataset used here (i.e. 9 users) the S-CRNN obtains average AUCs percentage increase of +21.5% compared to the best results obtained by HRV analysis (i.e. from 0.659 to 0.801 AUC). Finally, notice that all the methods based on HRV analysis perform similarly to each other. This suggests that the low AUC reached is not related to the actual classification model used, but to the weak correlation between the HRV features extracted and the user arousal level.



(a) HRV based method and S-CRNN.

(b) Variants of S-CRNN.

Fig. 3: AUCs for increasing the number of users included in the training set.

4.5 Variations on the Architecture

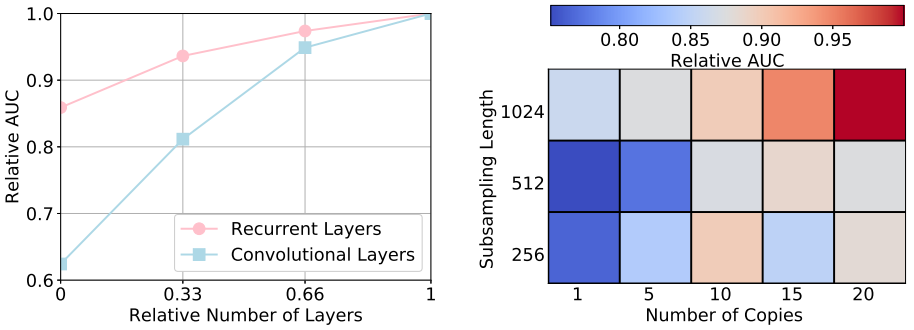
In Figure 3b we compare the S-CRNN with variants of its architecture, namely, with the CRNN model that does not benefit from the feature calibration layer,

and with a M-CRNN (Merged-CRNN). Similarly to the S-CRNN, the latter is based on two separate CRNN branches, but they do not share parameters.

Again, there is an overall trend of AUC increase as the number of training users increases. Contrary to what happens for HRV based methods, here all the models systematically get performance boost every time new users are included in the training set. This is likely to be related to the greater capacity of neural networks to use information from more data compared to manual feature extraction pipelines. Interestingly, the CRNN slightly outperforms the S-CRNN for population sizes of 1, 3 and 5. We speculate that this is because, with small population sizes, the feature calibration layer overfits to the specific training users characteristics. However, as the number of users increases, the S-CRNN is able to take full advantage of the information carried by new users' data. In fact with population size of 9, by proper calibration of the features extracted by the CRNN, the S-CRNN obtains a +7.5% percentage increase on the corresponding CRNN. Finally, notice that, even though the M-CRNN model is more general than the S-CRNN, it completely fails to improve even on the score obtained by the CRNN. This could be due to the almost double number of parameters of the M-CRNN, which quickly results in the model overfitting to the training data.

4.6 Hyper-parameters' Analysis

In Figure 4a we plot AUCs obtained for different numbers of recurrent and convolutional layers included in the S-CRNN. We analyse the effect of changing the number of layers of one type (either convolutional or recurrent), while keeping the other type of layers fixed to its nominal value. Notice that the x and y axis are normalised with respect to the S-CRNN architecture. The strongest effect is given by the convolutional layers, with the fully recurrent network obtaining only about 60% of the S-CRNN AUC. After an initial rapid increase, the AUC score saturates around the nominal S-CRNN architecture.



(a) Convolutional and recurrent layers. (b) Data augmentation hyper-parameters.

Fig. 4: Hyper-parameter analysis for S-CRNN.

Figure 4b shows the analysis results for the two hyper-parameters involved in the data augmentation phase. As expected, there is an overall trend of AUC increase as the number of copies made from each training sample is increased. However, the benefit from having more copies saturates around 15. Analogously, the more samples given as input to the S-CRNN, the higher is the AUC obtained.

5 Conclusions

We proposed a siamese CRNN architecture for arousal detection from ECG. The CRNN is explicitly designed to extract non-linear features from the ECG signal and analyse relevant time patterns using a 3-layer RNN stacked on top of a 6-layer CNN. Relying on a shared-parameter siamese architecture, we implemented feature calibration in the deep learning framework itself, which allows the neural network to model non-linear relationship between users' ECG at rest and that during emotion elicitation. We demonstrated the advantages of our approach compared to state-of-the-art HRV based methods, obtaining up to +21.5% percentage improvement on the AUC score. Further, we showed that the siamese architecture obtains +7.5% score increase compared to the CRNN.

As future work we plan to extend the S-CRNN to long-term analysis settings, and hence perform extensive comparison with techniques for medium and long-term HRV analysis. We emphasise that, though the siamese architecture was introduced for ECG, it can be generalised to most of the physiological signals used for affective state recognition. As feature calibration has proven to be a crucial step for manual feature extraction pipelines, future work will investigate whether affective computing based on deep learning architectures can benefit from the siamese network feature calibration paradigm proposed here.

6 Acknowledgements

This project was funded by the EU's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722022.

References

1. Picard, R.W.: Affective computing. Massachusetts Institute of Technology (1995)
2. Zhao, M., Adib, F., Katabi, D.: Emotion recognition using wireless signals. 22nd International Conference on Mobile Computing and Networking. 95–108 (2016)
3. Nardelli, M., Valenza, G., Greco, A., Lanata, A., Scilingo, E.P.: Recognizing emotions induced by affective sounds through heart rate variability. IEEE Transactions on Affective Computing. 6 (4), 385–394 (2015)
4. Melillo, P., Bracale, M., Pecchia, L.: Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. Biomedical engineering online. 10 (1), 96 (2011)
5. Ollander, S., Godin, C., Charbonnier, S., Campagne, A.: Feature and Sensor Selection for Detection of Driver Stress. PhyCS. 115–122 (2016)

6. Jovic, A., Bogunovic, N.: Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. Artificial intelligence in medicine. 51 (3), 175–186 (2011)
7. Gjoreski, M., Gjoreski, H., Luštrek, M., Gams, M.: Deep affect recognition from RR intervals. 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Symposium on Wearable Computers. 754–762 (2017)
8. Andrassy, G., Szabo, A., Ferencz, G., Trummer, Z., Simon, E., Tahy, A.: Mental Stress May Induce QT-Interval Prolongation and T-Wave Notching. Annals of Noninvasive Electrocardiology. 12 (3), 251–259 (2007)
9. Paoletti, N., Patanè, A., Kwiatkowska, M.: Closed-loop quantitative verification of rate-adaptive pacemakers. ACM Transactions on Cyber-Physical Systems. (2018)
10. Heslegrave, R.J., Furedy, J.J.: Sensitivities of HR and T-wave amplitude for detecting cognitive and anticipatory stress. Physiology & Behavior. 22 (1), 17–23 (1979)
11. Martinez, H.P., Bengio, Y., Yannakakis, G.N.: Learning deep physiological models of affect. IEEE Computational Intelligence Magazine. 8 (2), 20–33 (2013)
12. Kahou, S.E., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., Ferrari, R.C.: Emonets: Multimodal deep learning approaches for emotion recognition in video. Journal on Multimodal User Interfaces. 10 (2), 99–111 (2014)
13. Rosa, S., Patané, A., Lu, X., Trigoni, N.: CommonSense: Collaborative learning of scene semantics by robots and humans. In Proceedings of the 1st International Workshop on Internet of People, Assistive Robots and ThingS. 1–6 (2018)
14. Brennan, M., Palaniswami, M., Kamen, P.: Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? IEEE transactions on biomedical engineering. 48 (11), 1342–1347 (2001)
15. Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., Schmidt, A.: A data set of real world driving to assess driver workload. 5th international conference on automotive user interfaces and interactive vehicular applications. 150–157 (2013)
16. Bromley, J., Guyon, I., LeCun, Y., Sackinger, E., Shah, R.: Signature verification using a "siamese" time delay neural network. Advances in Neural Information Processing Systems. 737–744 (1994)
17. Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., Bhattacharya, S.: Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. AAAI. 4746–4752 (2017)
18. Salahuddin, L., Cho, J., Jeong, M.G., Kim, D.: Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. 29th International Conference of the Engineering in Medicine and Biology Society. 4656–4659 (2007)
19. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision. 1026–1034 (2015)
20. Panzner, M. and Cimiano, P.: Comparing Hidden Markov Models and Long Short Term Memory Neural Networks for Learning Action Representations. International Workshop on Machine Learning, Optimization and Big Data. 94–105 (2016)
21. Chollet, F. et al.: Keras. GitHub https://github.com/keras-team/keras (2015)
22. Abadi, M., Barham, P., Brevdo, E., Chen, Z. et al: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ (2015)
23. Cho, Y., Bianchi-Berthouze, N. and Julier, S.J.: DeepBreath: Deep Learning of Breathing Patterns for Automatic Stress Recognition using Low-Cost Thermal Imaging in Unconstrained Settings. arXiv:1708.06026 (2017)
24. Kingma, D.P., Jimmy, B: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)