2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)

A novel multi-kernel 1D convolutional neural network for stress recognition from ECG

Giorgos Giannakakis
Institute of Computer Science
Foundation for Research and Technology - Hellas
Heraklion, Greece
ggian@ics.forth.gr

Eleftherios Trivizakis
Institute of Computer Science
Foundation for Research and Technology - Hellas
Heraklion, Greece
trivizakis@ics.forth.gr

Manolis Tsiknakis, Kostas Marias
Institute of Computer Science
Foundation for Research and Technology - Hellas
Department of Electric and Computer Engineering, Hellenic Mediterranean University
Heraklion, Greece
tsiknaki@ics.forth.gr, kmarias@ics.forth.gr

***Abstract*—Stress is an emotional state which although experienced in a subjective way, it shares specific common characteristics. Objective stress recognition has proven to be a complicated issue, due to the number of parameters involved. Thus, the investigation of reliable indices associated with the stress response is of utmost importance. Heart activity may provide useful information towards this goal. Traditional machine learning techniques have been used in the area of emotion recognition but they sometimes present specific limitations. The emergence of Deep Learning (DL) techniques permits the reveal underlying patterns in electrocardiography (ECG) which, otherwise, would not be easily observed. The proposed DL architecture utilizes a variety of kernels per module to compute complex feature maps and enables a multi-level modelling of the unique heart rate variability signature for stress state identification. The proposed methodology using 6-fold cross-validation outperforms single kernel networks achieving classification accuracy up to 99.1%, better overall performance (avg. F1-score 88.1%, avg. accuracy 89.8%) and more consistent behaviour across study's experimental phases.**

Keywords—stress, HRV, heart rate, ECG, Deep Learning classification, multi-kernel, convolutional neural network, biosignal analysis

I. INTRODUCTION

In recent years, stress has emerged as one of the most significant problems in modern societies. Thus, automatic stress recognition is receiving increasing research interest. Although there is a common understanding of the term stress, however stress identifying characteristics and modelling remains a difficult task. The problem becomes more complex considering the different subjective stress experience when faced with specific stressors and the fact that there exist different stress types and different stress manifestations which greatly vary. Besides, people usually make efforts to hide their stress as it is usually an unpleasant and often embarrassing situation.

The reliable and objective estimation of one's stress affect, experience and levels remains a challenge. Thus, several scientific approaches and methods have been developed for automatic stress recognition based on behavioural patterns [1], facial cues [2], upper body movements [3, 4], biosignals (electroencephalography (EEG) [5], ECG [6], electrodermal activity (EDA) [7]) or multimodal approaches [8, 9]. In some cases, stress experience may be moderated by one's defence mechanisms and stress manifestations may be hidden. Therefore, the research community pursuing reliable stress indices turn to physiological measures (biosignals) as semi- or involuntary measures which may provide a more reliable stress estimation [8]. This approach has been supported by the recent technological evolution which enables the usage of available top-quality and off-the-shelf wearable devices and sensors towards stress recognition [10].

Traditional machine learning (ML) techniques have been used in supporting stress recognition but they sometimes present limitations in terms of identifying complex patterns. In contrast, Deep Learning (DL) techniques provide an exhaustive analysis of the ECG timeseries following hierarchical approaches and revealing underlying patterns that cannot be easily observed.

To the best of our knowledge, there are few studies in the literature employing deep learning techniques in stress recognition through ECG. Masood et al. [11], proposed a single kernel 1D and a recurrent CNN in order to analyse ECG, EEG features for stress discrimination achieving up to 90% accuracy with holdout stratification. Additionally, in [12], ultra-short-term ECG analysis has been used along with DL techniques achieving accuracy up to 87.4%.

The purpose of this paper is to develop a robust stress model being able to efficiently discriminate stress states using ECG signals. Towards this objective, we investigate the most appropriate transformations for data-driven models in the context of stress classification, introduce a transparent and robust methodology by integrating the best practices for model convergence, and finally propose a computationally efficient DL architecture for automated heart activity analysis.

II. EXPERIMENTAL PROCEDURE AND DATASET

A. Experimental procedure

In order to investigate the effects of stress conditions, a stress experiment was designed and developed. It includes neutral (reference) periods and stressors that simulate a wide range of everyday life conditions that were induced to the participants. The different stressors categorized in 4 experimental phases (social exposure, recall of a stressful event, cognitive load, watching of stressful videos) in an

Prof. K. Marias and M. Tsiknakis are equally contributing last authors.
This study was supported by the Greek State Scholarships Foundation (IKY) and co-financed by the action entitled "Reinforcement of Postdoctoral Researchers" (NSRF 2014-2020).
978-1-7281-3891-6/19/$31.00 ©2019 IEEE

TABLE I EXPERIMENTAL TASKS, PHASES AND CONDITIONS EMPLOYED IN THE STUDY.

Experimental phase	Affective State	Duration (min)
Social Exposure		
1.1 Neutral (reference)	N	1
1.2 Self-describing	S	1
Stressful event recall		
2.1 Neutral (reference)	N	1
2.2 Recall anxious event	S	1
2.3 Recall stressful event	S	1
Cognitive load		
3.1 Neutral/Stressful images	S	2
3.2 Stroop Colour Word Task	S	2
Stressful videos		
4.1 Neutral (reference)	N	1
4.2 Relaxing video	R	2
4.3 Adventure video	S	2
4.4 Psychological pressure video	S	2

Note: Intended affective state (N:neutral,S:stress,R:relaxed).

attempt to cover different underlying stress types. The experimental procedure (for more information about the procedure please read [2]) along with the tasks description and their corresponding affective state is presented in Table I.

For the ECG recording, the patient's skin was prepared using prepping gel and conductive paste. Two Ag/AgCl electrodes were placed in symmetric position of the chest corresponding to the leads V1 and V2 which they are considered as the most appropriate in order to acquire bipolar ECG recording [13].

B. *Dataset of the study*

The population of this study was 24 participants (7 women, 17 men) with age 47.3±9.3 years. The study was approved by the North-West Tuscany ESTAV (Regional Health Service, Agency for the Technical-Administrative Services of Wide Area) Ethical Committee. The dataset was collected during the data acquisition campaign (SRD'15) of a research project that focused on the development of a computational platform for the unobtrusive monitoring of cardiometabolic risk [14].

C. *ECG Preprocessing*

The bipolar V1-V2 ECG signal recorded during the experiment as described in section IIA was detrended and bandpass filtered in order to remove power line noise. Spikes and artifacts (due to the subject's activity like head/body movements, chewing, etc.) were suppressed.

The R peaks of the QRS complex were detected and ectopic heartbeats were excluded, by adopting the HRV signal approach [15]. A heartbeat was determined as ectopic if there is a percentage change of 40% over the averaged previous 5 heartbeats. Then, the RR Intervals (RRI) were calculated providing multiparametric information which was used as input to our system.

D. *Baseline remove and normalization*

For each participant of the experiment, it was used either the non-normalized RRI timeseries or the normalized data considering an appropriate reference period (a neutral task at the beginning of each phase). This period corresponds to each subject's baseline which is used to normalize data for all subsequent feature analyses. This generates a common reference to each feature across subjects providing data normalization [2].

III. DEEP LEARNING METHODOLOGY

The dataset demonstrates significantly high variability among experiment phases which represent different stress types, therefore the different phases were modelled separately in order to provide a data-driven model for each stress type. Deep Learning (DL) analysis provide a unified and automatic feature extraction, selection and classification pipeline by leveraging labelled time-series via end-to-end supervised learning. This methodology exhaustively computes discriminative feature maps unveiling the complex biosignals patterns related to the emotional state of each subject [2].

A. *Proposed architecture - DWNet1D*

The proposed 1-dimensional Deep Wide Convolutional Neural Network (DWNet1D) introduces a novel approach to heart activity analysis with multiple kernels of variable length per module as depicted in Figure 1. In particular, three modules were included each followed by a pooling layer, batch normalization and ReLU activation. This architecture enables a complex and diverse inner representation of more than 1.1 million learnt parameters (number of modules, kernels, neurons) that reveal underlying patterns of the cardiac signal. The source code and the final hyper-parameters of DWNet1D are available online[1].

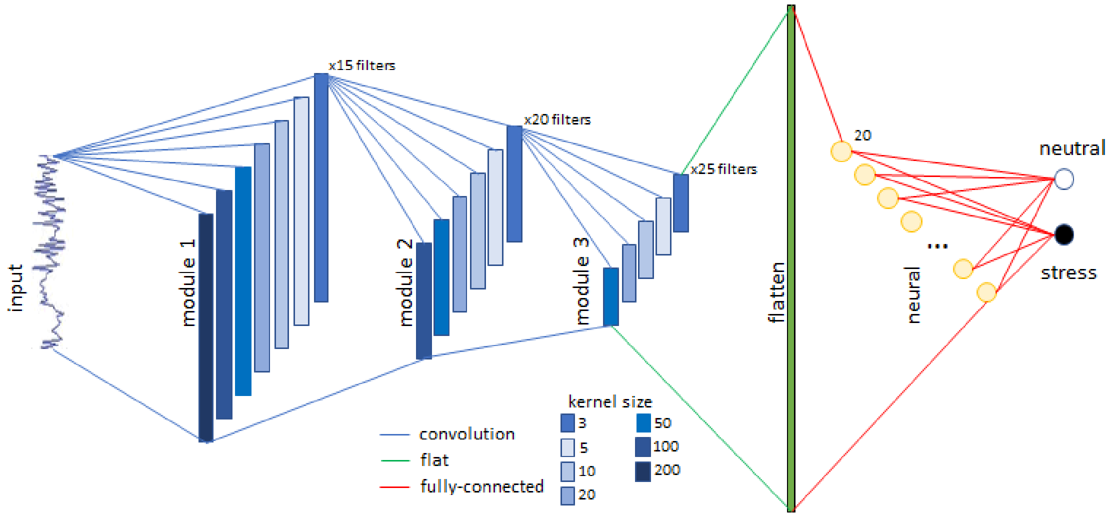

Figure 1: A detailed view of the proposed Multi-Kernel 1-D Convolutional Neural Network -DWNet1D- for ECG signal analysis.

[1] https://github.com/trivizakis/DWNet1D/

TABLE II CLASSIFICATION ACCURACY OF EACH PHASE AND TOTAL AVERAGE AMONG MULTI & SINGLE KERNEL NETWORKS

	Normalized				Non normalized				
Kernel	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Total Avg.**
DWNet1D	90.3	83.5	92.6	85.9	80.2	87.8	99.1	99	**89.8**
Single-3	91.9	87.8	92.7	85.3	74.3	85.2	99	95.3	88.9
Single-5	91.5	80.3	92.0	87.4	69	**96.3**	81.5	98.7	87.1
Single-10	**93.5**	**90.5**	**94.5**	87.3	68.4	76.7	99.5	93.8	88.0
Single-20	89.6	85.2	93.1	84.6	73.7	81.5	99.1	97.9	88.1
Single-50	87.2	78.7	92.8	84	79.9	78.5	**99.4**	93.5	86.8
Single-100	92.9	88.9	90.3	86.4	78.4	84.1	98.1	96.9	89.5
Single-200	90.6	89	88.7	**88.5**	**80.7**	83.3	88.4	97.8	88.4

B. *Data stratification*

The dataset consists of a total of 24 subjects and 11 trials which is limited for DL training and optimization process. To overcome this limitation, a 6-fold cross-validation on a subject basis fold was used for performance evaluation of the models iteratively along with shuffle hold out for model convergence. This resulted in 16 training, 4 validation and 4 testing subjects for each fold. The produced deep models adapt to infer the 2 emotional states (stress, neutral) of the participant under investigation on the convergence set (training and validation set). The convergence set was split into training and validation set by a random hold-out stratification preserving the classes balance. The final evaluation and performance metrics for the models (6-fold cross-validation averages) was performed solely on the unseen testing set.

C. *Data Augmentation*

The analysis was performed using an overlapping sliding window of a window length 60 sec and a step 3 sec, with zero padding. Augmented data with over 50% padding were rejected to avoid having spurious results of using a small ECG sample. This procedure was performed after the subject-level stratification to prevent the introduction of overfitting from the new but redundant information and avoid memorization of the dataset by the network. The stratification on a subject basis and not a sample basis provides a fair evaluation and consequently increased generalization ability for the produced.

D. *Hyperparameter Optimization*

The training set was used for model fitting and the validation set for identifying the best network parameters. This includes learnable (number of modules, kernels, number of neurons) and other fundamental parameters (learning rate, optimizer, activation, classifier). Early-stopping was performed 20 epochs after reaching the maximum validation accuracy to avoid overtraining, ensure the best model was obtained and refrain from unnecessary training iterations. Additionally, juxtaposing the training-validation accuracy curves provided insights about the fitting status for the models and assisted in the selection of the optimal hyperparameters. The learning curve should converge to a narrow generalization gap (the distance between training and validation accuracy) in order the model to achieve a robust learning performance.

E. *Performance Evaluation Metrics*

The studied classification problem was evaluated mainly by the Receiver Operating Characteristic (ROC), Area Under Curve (AUC) due to its binary nature and secondary by using other metrics including accuracy $\frac{TP+TN}{TP+TN+FP+FN}$, sensitivity $\frac{TP}{TP+FP}$, specificity $\frac{TN}{TN+FP}$, and F1-score $\frac{2TP}{2TP+FP+FN}$ where *TP*, *TN*, *FP*, and *FN* stand for true positive, true negative, false positive and false negative.

IV. RESULTS

A total of 336 models for the single kernel and 24 for multi-kernel analysis with similar parameters were tested in terms of their precision/recall. The dataset was comprised of 24 subjects with 11 tasks each. The length of each task RRI timeseries was 600 samples (or 2x600 samples for the 2-min tasks).

The proposed methodology achieves a classification accuracy of up to 99.1%. To better point out the performance of the proposed multi-kernel architecture additional single kernel networks were trained individually achieving good results in terms of the accuracy metric on the minority of phases and signal transformations (non-normalized and normalized). The multi-kernel architecture demonstrates the most balanced and consistent results (less variability) across experimental phases due to the combination of the diverse features maps calculated by multiple and varying length kernels. It achieves the best total average accuracy of 89.8% which outperforms single kernel networks. The equilibrium of accuracy differences of DWNet1D versus Single-10 (best single kernel network) is much higher in favour of the proposed architecture demonstrating a robust and stable performance across experimental phases. Additionally, the multi-kernel architecture is also better on average accuracy than the single-kernel models. The classification accuracies results categorized in experimental phases and data normalization mode are presented in Table II. Additional context for the performance of the phase-specific multi-kernel networks can be deduced observing the corresponding average 6-fold ROC AUC (Figure 3).

To verify the reproducibility of the methodology we tested multiple models with the best parameters obtained

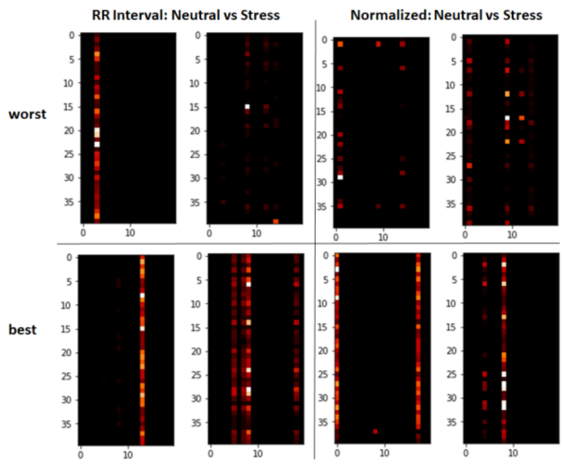

Figure 2: Visualization of activations from the neural part of the worst (top) and best (bottom) models

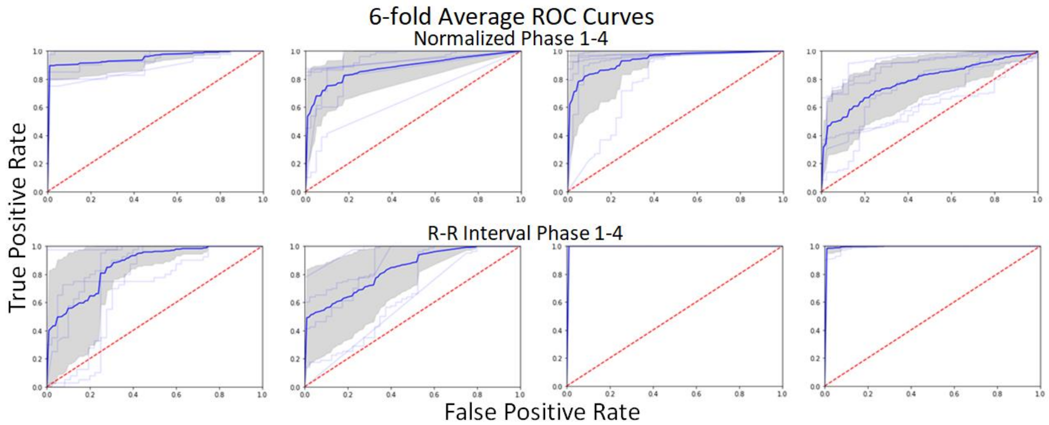

Figure 3: Average 6-fold ROC Curves for Normalized (top) and RR Interval (bottom) transformation for the examined stress types (left to right)

using the hyper-parameter optimization. There is a variance in the measured metrics ±2% which is expected due to the randomness of both the initialization of the networks and the subject stratification process. Additionally, modelling solutions with no padding were also tested resulting in similar results and behaviour for each experiment and signal transformation.

The visualized activations of a one-fold from the fully-connected layer of the DWNet1D is presented in Figure 2. Each row represents a testing set pattern (activations) of the fully-connected layer providing insights about the computed patterns of each emotional state (stress, neutral). Models with poor performance cannot easily discriminate the two states as shown in the top right of Figure 2 in contrast to the activation patterns from the best models (bottom panel of Figure 2) where it is evident that there is a stable CNN model response as testing set samples present a consistent pattern (similar row pattern in bottom panels) across the fold testing set.

V. Discussion

In this study, we propose a Deep Learning multi-kernel architecture for recognizing stress states through heart activity. As stated in the Section IV, the proposed DL methodology in terms of 6-fold cross-validation achieves a classification accuracy of up to 99.1% outperforming in terms of accuracy the related DL studies (achieving 90% in [11] and 73-87.4% in [12]) or the related ML approach using the same dataset (84% in [16]). To the best of our knowledge, the proposed methodology is the first multi-kernel 1-dimension CNN that is used in the DL literature as till now related studies use single kernel CNN. The proposed methodology reveals the role of different signal transformations for various stress types introducing a reliable experimental protocol for data-driven and learning-based models.

The limited population size used in the experiment was addressed using sliding window analysis and evaluating multiple models through 6-fold cross-validation. Another significant limitation for a data-driven approach as in this study is the inter-subject variability among the different stress types (experiment phases). To address this, an individual model was established for each of the 4 different stress types. Further studies and increased sample size would increase the model's reliability which will result in optimizing the architecture's parameters leading to even better performance and more generalized models on experimental tasks.

References

[1] C. Mohiyeddini and S. Semple, "Displacement behaviour regulates the experience of stress in men," *Stress*, vol. 16, pp. 163-171, 2013.

[2] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89-101, 2017.

[3] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, pp. 1334-1345, 2007.

[4] L. Salahuddin, J. Cho, M. G. Jeong, and D. Kim, "Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings," in *2007 29th annual international conference of the ieee engineering in medicine and biology society*, 2007.

[5] G. Giannakakis, D. Grigoriadis, and M. Tsiknakis, "Detection of stress/anxiety state from EEG features during video watching," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 6034-6037.

[6] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 2957-2960.

[7] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on Information Technology in Biomedicine* vol. 14, pp. 410-417, 2010.

[8] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis., "Review on psychological stress detection using biosignals.," *IEEE Transactions on Affective Computing*, 2019.

[9] J. Aigrain, M. Spodenkiewicz, S. Dubuisson, M. Detyniecki, D. Cohen, and M. Chetouani, "Multimodal stress detection from multiple assessments," *IEEE Transactions on Affective Computing*, 2016.

[10] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 671-676.

[11] K. Masood and M. AlGhamdi, "Modeling Mental Stress using a Deep Learning Framework," *IEEE Access*, 2019.

[12] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B.-T. Zhang, "Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," *TELEMEDICINE and e-HEALTH*, vol. 24, pp. 753-772, 2018.

[13] L. Bernardi, J. Wdowczyk-Szulc, C. Valenti, S. Castoldi, C. Passino, G. Spadacini, and P. Sleight, "Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability," *Journal of the American College of Cardiology*, vol. 35, pp. 1462-1469, 2000.

[14] Y. Andreu, F. Chiarugi, S. Colantonio, G. Giannakakis, D. Giorgi, P. Henriquez, E. Kazantzaki, *et al.*, "Wize Mirror - a smart, multisensory cardio-metabolic risk monitoring system," *Computer Vision and Image Understanding*, vol. 148, pp. 3-22, Jul 2016.

[15] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard, "Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera," in *CHI Conference on Human Factors in Computing Systems*, 2016.

[16] G. Giannakakis, K. Marias, and M. Tsiknakis, "A stress recognition system using HRV parameters and machine learning techniques," in *8th International Conference on Affective Computing & Intelligent Interaction (ACII 2019)*, 2019.