

Synthesizing Physiological and Motion Data for Stress and Meditation Detection

Md Taufeeq Uddin and Shaun Canavan

Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA

mdtaufeeq@mail.usf.edu, scanavan@usf.edu

Abstract—In this work, we present the synthesis of physiological and motion data to classify, detect and estimate affective state ahead of time (i.e. predict). We use raw physiological and motion signals to predict the next values of the signal following a temporal modeling scheme. The physiological signals are synthesized using a one-dimensional convolutional neural network. We then use a random forest to predict the affective state from the newly synthesized data. In our experimental design, we synthesize and predict both stress and meditation states. We show the utility of our approach to data synthesis for prediction of stress and meditation states through two methods. First, we report the concordance correlation coefficient of the synthetic signals compared to the ground truth. Secondly, we report prediction results on synthetic data that are comparable to the original ground-truth signals.

Index Terms—synthesis, stress, meditation, physiological, motion, prediction

I. INTRODUCTION

Stress is traditionally diagnosed by psychologists using offline procedures such as asking questions [10], however, more objective measures such as physiological readings can be used to diagnose stress in place of subjective questions [7]. Considering this, there has been interesting work in automatically recognizing stress in recent years. Jebelli et al. [8] investigated recognizing stress in workers using electroencephalogram (EEG) data. The EEG data was collected, from construction workers, across 14 different channels using a wearable EEG headset. After removal of noise artifacts in the data, they trained a deep convolutional neural network (CNN) to recognize stress. They achieved a maximum accuracy of 86.62%. Also, using EEG data, Liao et al. [12] trained a deep neural network to recognize stress. They collected EEG data from 7 subjects that listened to music for 10 minutes. Using an 80/20 split of the data (train on 80% and test on 20%), they trained a deep neural network achieving an accuracy of 80.13%.

Although there has been success in recognizing stress using EEG data, another interesting modality that has been used is thermal data. Cho et al. [4] detected breathing patterns, in the thermal data, by analyzing temperature changes around the subject's nostrils. They transform the 1D breathing signals into 2D respiration variability spectrograms, which are then used to train a CNN. They also propose a data augmentation technique to train on a small sample size. They are able to recognize between two and three levels of stress with 84.50% and 56.52% accuracy, respectively.

Although these results are promising, there are two challenges that we aim to address in this paper. First, these works focus only on stress, where we predict stress, amusement and meditation (i.e. relaxed) states. Investigating mediation along with stress and amusement is important as it has been found that when meditation is practiced for 15 minutes, two times a day, stress and anxiety are reduced due to being in a state of deep rest according to the subject's physiological state [5]. Secondly, all of them utilize data that has been collected, processed and then classified. We are interested in the prediction of mental states, especially in a real-time medical setting, as we aim to predict the subject's mental state ahead of time. Due to this, the signals from collected datasets will not be sufficient for our experimental needs. Considering this, we propose a method for synthesizing the next step (i.e. time) of physiological and motion signals using a one dimensional CNN (1DCNN). Given these synthesized signals, we then predict stress, amusement and meditation states using a random forest classifier. The major contributions of this work are two-fold and given below:

1. We propose a temporal modeling approach for synthesizing a given modality (e.g. physiological signals) from raw signals to model affective states ahead of time. This allows us to effectively predict the subject's mental state ahead of time, which is a specific aim of this work.
2. To the best of our knowledge, we also provide the first baseline results on meditation state prediction from physiological and motion data from the publicly available dataset, wearable stress and affect detection (WESAD) [15]. This investigation into mediation has potential applications for ultimately treating affective disorders such as stress.

II. PHYSIOLOGICAL AND MOTION DATA SYNTHESIS

In this work, we propose synthesizing futuristic signals for detecting affective states beforehand. First, we consecutively use a sequence of previous and current observation of a signal to predict the next observation of signal. As a result, we obtain synthetic data about futuristic affective states even before the occurrence of the affective states. We then use the synthesized data to train a machine learning classifier to recognize an affective state, effectively predicting the new state, ahead of time.

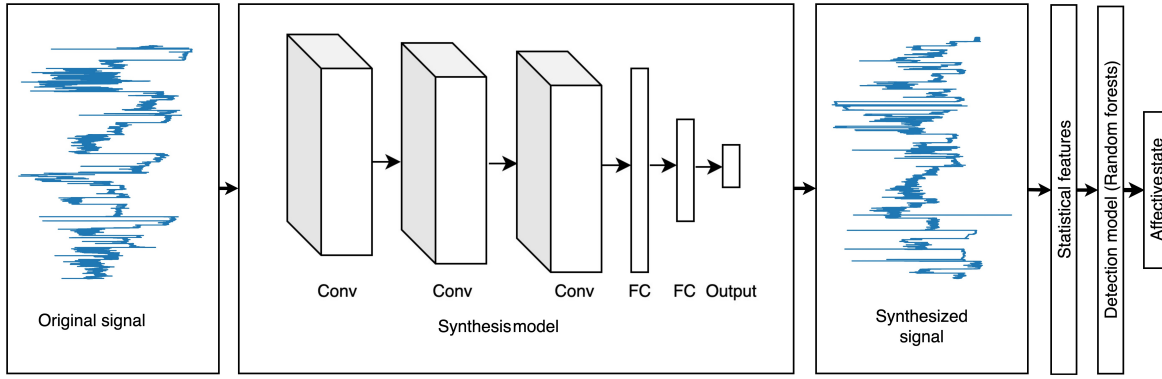


Fig. 1. Overview of proposed Approach. The original signal is used to train a CNN to predict signal value in time step $t + 1$ from time step t . Features are then extracted from the synthesized signal to predict the affective state. Here, Conv means convolution and FC means fully connected layer.

A. One dimensional CNN for data synthesis

We propose a simple, yet effective approach to synthesize physiological and motion data. We have applied 1DCNN [9] to synthesize the signal, given its ability to capture spatiotemporal relationships in data. We have used signal values from t number of time steps (prior and current) to train the 1DCNN to predict signal value in the next (future) time step $t + 1$. Our 1DCNN consists of three convolution layers, two fully connected layers (64, 32), two dropout layers (0.5, 0.5), and an output layer. We have used linear as activation function in the output layer and relu as activation function in the rest of the layers. Convolutional layers contain variable number of filters (128, 64, 64) and fixed kernel (3×3) size, as well as padding. We have used Adam as the optimization algorithm, given Adam's ability for faster convergence, to train the model with learning rate and number of epochs set to 0.01 and 500, respectively. See Fig. 1 for an overview of the proposed approach.

B. Random forest for affective state prediction from synthetic signals

In the prediction phase, we calculated 10 statistical features [13] from each synthetic signal from a given range of time steps. More specifically, we computed the following 10 features: *mean, maximum, minimum, index of maximum values, standard deviation, kurtosis, skewness, interquartile range, energy, and mean absolute deviation*. We then use the newly created feature vector to train a random forest classifier [3] to predict stress, amusement, and meditation. In our experiments, we tuned the random forest performing grid search over the parameters (number of trees = {5, 10, 20, 40, 80} and maximum depth of trees = {50, 100, 200, 400, 1000}).

III. EXPERIMENTAL DESIGN AND RESULTS

A. Dataset

WESAD [15] dataset is collected from 15 graduate students in which 12 of them are male and 3 are female, with an average age of $27.5(\pm 2.4)$. Physiological and motion data are captured from the chest and wrist of the participants using RespiBAN and Empatica E4 devices. In total, WESAD dataset

contains 10 modalities (chest-worn modalities are sampled at 700Hz) including acceleration (ACC) X, Y, Z, electrocardiogram (ECG), electrodermal activity (EDA), electromyography (EMG), temperature and respiration. The data collection procedure consists of four states: neutral (induce a neutral affective state; this phase is 20 minutes long), amusement (induce a happy state; this phase is 6.5 minutes long), stress (the participant performed two tasks: public speaking and mental mathematics solving following Trier Social Stress Test [11] to mimic the stress behavior. Stress phase is 10 minutes long), and meditation (breathing exercise to induce neutral state after the performance of amusement and stress states. This phase is 7 minutes long). The duration of all experiments (including 15 subjects) is 11 hours. Note that, in this work, we only used the modalities captured via the chest-worn device since wrist-based modalities could be neglected in the presence of chest-based modalities [15].

B. Evaluation

To evaluate our model, we first synthesize data and then use the synthesized data to detect affective state ahead of time. We performed 3-fold subject independent cross-validation in both signal synthesis and affective states prediction. We used the concordance correlation coefficient (CCC) [16] score as an evaluation metric to measure the effectiveness of the synthesis model given that CCC score (range: $[-1, 1]$, higher is better) takes care of both correlation and accuracy. On the other hand, we used precision, recall, and F1-score to measure the performance of affect prediction.

C. Data Synthesis

We preprocessed the dataset before performing synthesis. More precisely, we applied Savitz Golay filter [14] to smooth the signal. Then, we used a factor n to re-sample the dataset by taking the average of n consecutive values for a given signal. Finally, we performed min-max normalization in between $[0, 1]$ for each signal for better generalization of the neural network. To validate the effectiveness of the synthesis approach, we experimented with different durations of time to synthesize the signals. As in time $T1$, $T2$, and $T3$ settings,

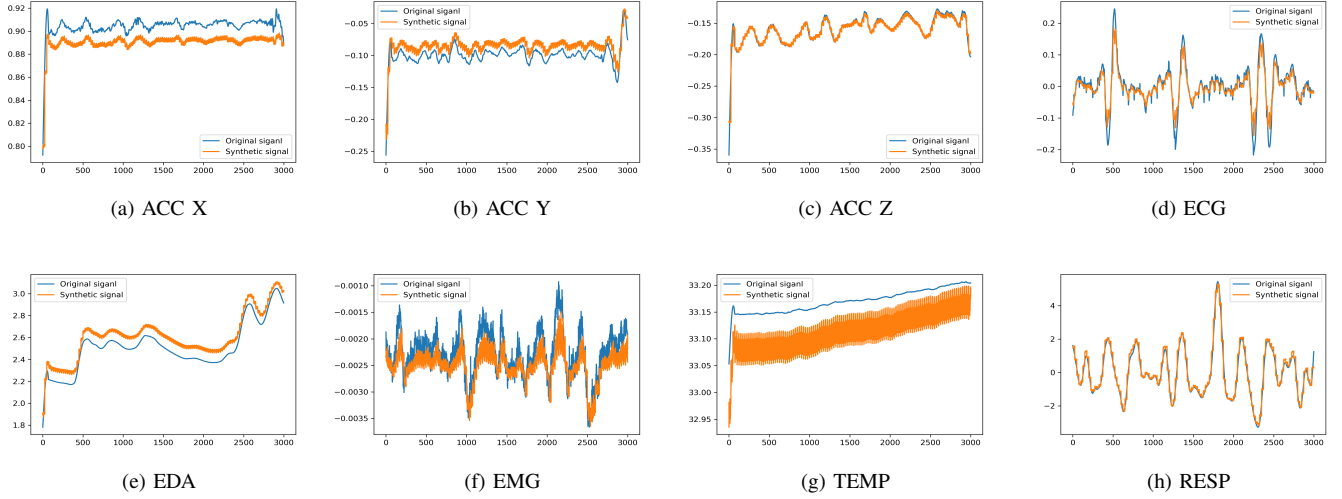


Fig. 2. Original and synthetic signals for subject 4 of WESAD dataset during stress phase. Each plot represents 3 seconds worth of signal. Here, X-axis and Y-axis represent time step and signal value in that order. *NOTE: best viewed in color.*

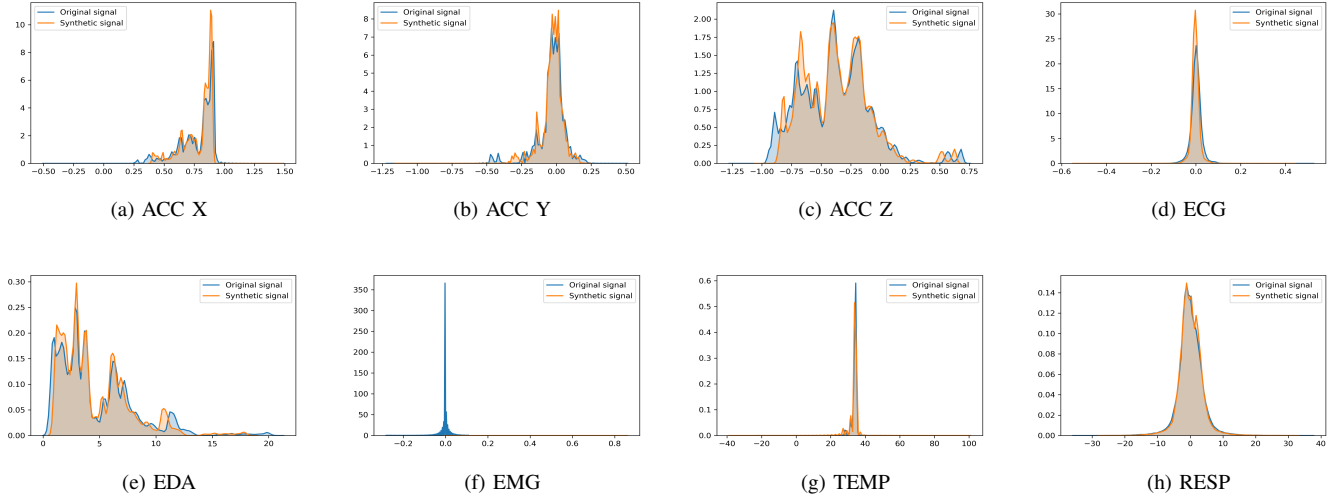


Fig. 3. Distribution of original and synthetic signals of entire WESAD dataset. *NOTE: best viewed in color.*

TABLE I
SYNTHESIS RESULTS ON WESAD DATASET USING PROPOSED 1DCNN.

Signal	CCC score		
	Time T1	Time T2	Time T3
ACC X	0.9909 \pm 0.0054	0.9856 \pm 0.0119	0.9899 \pm 0.0014
ACC Y	0.9675 \pm 0.0272	0.9711 \pm 0.0081	0.7882 \pm 0.2552
ACC Z	0.9957 \pm 0.003	0.992 \pm 0.0026	0.9943 \pm 0.0024
ECG	0.8104 \pm 0.0327	0.5001 \pm 0.0922	0.3995 \pm 0.0602
EMG	0.4381 \pm 0.6313	0.5837 \pm 0.0561	0.3407 \pm 0.4757
EDA	0.9928 \pm 0.0012	0.9964 \pm 0.0009	0.9927 \pm 0.0036
TEMP	0.95 \pm 0.0268	0.9401 \pm 0.0437	0.7405 \pm 0.1785
RESP	0.9895 \pm 0.0009	0.9697 \pm 0.0065	0.9563 \pm 0.0094

we used 0.25, 1, 2 second(s) worth of data as feature vector

to synthesize 0.01, 0.04, and 0.1 second(s) worth of data, respectively.

As it can be seen from Table I, in $T1$ $T2$, $T3$ settings, the proposed model synthesized 7/8, 6/8, and 6/8 signals with high CCC scores, respectively. In the case of EMG synthesis, we obtained a high CCC score for some subjects while our model failed to generalize well on some other subjects. This result can be partially explained by the personalized influence of subjects, since the representation of affect states including stress, has variability among people [15]. From Figures 2 and 3, we can observe that our model was able to learn the distribution of all signals with high margin except the EMG signal. We can see from Figure 2 that the EMG signal moves back and forth rapidly from one time step to the next; that could cause challenges when predicting EMG signal.

TABLE II
CLASSIFICATION RESULTS ON BOTH ORIGINAL AND SYNTHETIC WESAD DATASET USING RANDOM FOREST.

Detection Setting	Original data			Synthetic data		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Stress vs Baseline	0.71 \pm 0.06	0.68 \pm 0.10	0.69 \pm 0.08	0.71 \pm 0.10	0.66 \pm 0.11	0.68 \pm 0.10
Stress vs Amusement	0.54 \pm 0.11	0.61 \pm 0.06	0.57 \pm 0.10	0.55 \pm 0.15	0.56 \pm 0.07	0.55 \pm 0.11
Stress vs Meditation	0.72 \pm 0.05	0.71 \pm 0.06	0.71 \pm 0.06	0.71 \pm 0.02	0.7 \pm 0.02	0.71 \pm 0.02
Stress vs Rest	0.76 \pm 0.03	0.7 \pm 0.06	0.73 \pm 0.04	0.74 \pm 0.03	0.67 \pm 0.07	0.7 \pm 0.04

D. Prediction of Synthesized Data

To predict the affective state, we calculated 10 statistical features from n consecutive steps (in our experiment, we set $n = 25$) of all synthetic signal types (i.e. multimodal), resulting in 80 features (10 features \times 8 signal types). For affective state prediction, we used signal synthesized in $T1$ setting. To provide a baseline comparison with the original signal, we also calculated the same 80 features from the raw signal dataset. To test the effectiveness of the proposed synthesis model for stress, amusement, and meditation detection, we performed both one vs one and one vs rest classification using a random forest classifier. It can be observed in Table II that the results we obtained from synthetic data are comparable to original data. Some of the differences in accuracy can partially be explained by the fact that our model synthesized EMG poorly for some subjects, and also we lost some precision in the data during synthesis (See Figures 2 and 3). Moreover, Table II suggests that the prediction of stress from amusement is difficult, which can be partially explained due to the subtlety of amusement among subjects [15] and the difficulty of valence detection from physiological signals.

IV. CONCLUSION AND DISCUSSION

We have presented a method for synthesizing physiological and motion signals for predicting stress and meditation. We have shown the efficacy of our method by comparing the synthesized signals to the original ground truth data. We have shown that the proposed method can accurately synthesize these signals across varying time steps (Table I). We have also shown that the accuracy of the synthesized signal for predicting stress, amusement and mediation is comparable to the classification accuracy of the original ground truth signal.

Although these results are promising, showing a first step towards the prediction of stress, amusement and mediation states, there are open challenges in this paradigm. The end goal of the model is to synthesize data over longer periods of time to predict the affective state, therefore, the problem can be formulated as a dynamic regression problem in which we predict the signal value of time $t + 1$ using previous time t . After making the prediction at time step $t + 1$, we will incorporate the predicted (i.e. synthesized) data $f(X_{t+1})$ in our feature vector to use the synthesized data to make prediction (i.e. synthesize data) at next time step $t + 2$ and so on. Therefore, after making t predictions, our feature vector will only contain the synthetic signal for future predictions. If

the model incorporates some degree of error (e.g. noise, drift), then error propagation must be handled. This error propagation will introduce the out of distribution challenge in a dynamic setting. It has been shown that the out of distribution problem can degrade the performance of machine learning models [1].

We will also explore the online learning [2] paradigm to synthesize data in this dynamic setting. Along with this paradigm, we will also explore continuous learning [6] in which we will first train the model on raw data to build an initial synthesis model that will be improved over time using reinforcement learning.

ACKNOWLEDGMENT

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

REFERENCES

- [1] M. Alcorn et al. Strike (with) a pose: Neur. net. are easily fooled by str. poses of familiar objects. *arXiv preprint arXiv:1811.11553*, 2018.
- [2] T. Anderson. *The theory and practice of online learning*. Athabasca University Press, 2008.
- [3] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [4] Y. Cho et al. Deepbreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In *ACII*, pages 456–463, 2017.
- [5] F. Coppola and D. Spector. Natural stress relief meditation as a tool for reducing anxiety and increasing self-actualization. *Social Behavior and Personality: an international journal*, 37(3):307–311, 2009.
- [6] S. Gu, T. Lillicrap, I. Sutskever, and S. Levine. Cont. deep q-learning with model-based acceleration. In *ICML*, pages 2829–2838, 2016.
- [7] B. Hwang et al. Deep ecgnet: An optimal deep learning framework for monitoring mental stress using ultra short-term ecg signals. *TELEMEDICINE and e-HEALTH*, 24(10):753–772, 2018.
- [8] H. Jebelli, M. Khalili, and S. Lee. Mobile eeg-based workers stress recognition by applying deep neural network. In *Advances in Info. and Computing in Civil and Construction Eng.*, pages 173–180. 2019.
- [9] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A conv.neural net. for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2014.
- [10] P. Karthikeyan, M. Murugappan, and S. Yaacob. Detection of human stress using short-term ecg and hrv signals. *Journal of Mechanics in Medicine and Biology*, 13(02), 2013.
- [11] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer. The trier social stress test—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.
- [12] C.-Y. Liao, R.-C. Chen, and S.-K. Tai. Emotion stress detection using eeg signal and deep learning technologies. In *International Conference on Applied System Invention*, pages 90–93, 2018.
- [13] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita. Transition-aware human activity recognition using smartphones. *Neuro-computing*, 171:754–767, 2016.
- [14] R. W. Schafer et al. What is a savitzky-golay filter. *IEEE Signal processing magazine*, 28(4):111–117, 2011.
- [15] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 400–408. ACM, 2018.
- [16] T. J. Steichen and N. J. Cox. A note on the concordance correlation coefficient. *The Stata Journal*, 2(2):183–189, 2002.