Multi-task Learning for Stress Recognition

Annonymous Authors

ABSTRACT

Different people experiencing the same stressor can have different psycho-physiological reactions. Such differences happen even while reporting the same affective states. This mismatch might cause a drop in the performance of affect recognition models. On the other hand, this same psycho-physiological uniqueness has been successfully utilized to develop biometric recognition systems based on wearable sensing data. All this raises the potential for exploiting the benefits of the psycho-physiological uniqueness observed in biometric recognition systems to improve stress affect machine learning models. In this paper, we investigate the joint learning of a model for stress recognition and user identification to improve the performance of the first task. The joint model was learned using Multi-task Learning (MTL) neural network. Using Leave-One-Subject-Out evaluation procedure, we showed that MTL brought an improvement of 6.5 percentage points in F1-score compared to single-task models (from 86.1 to 92.6 F1-score). The MTL approach did not bring increased complexity to the overall processing pipeline. Instead, it showed how user identification – typically disregarded in machine learning approaches for affect recognition – could be utilized as an additional source of information to improve the performance of ML models, even in scenarios where the models are tested on newly presented users.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

User Identification, Stress Recognition, Machine Learning, Wearable Sensors

ACM Reference Format:

Annonymous Authors. 2018. Multi-task Learning for Stress Recognition. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the last few decades, technology has advanced significantly, allowing machines to move well beyond the initial desktop revolution, gathering a special role in everyday life of humans. Initially, many people shared a single machine. Subsequently, the concept of a 1:1 relationship between machines and humans significantly impacted the way people used computer systems. This kind of human-machine relationship changed in the last decade when a 1:M relationship was established, where each user had many computers. From such a relationship, a new computing paradigm took shape: the Ubiquitous Computing era [10]. Computing can be defined as the interaction with artifacts and environments that intertwine with communication and processing capabilities [39, 40]. The foundation of ubiquitous computing is to empower the human component to act and interact in such a system and to further enhance the human experience inside of it. For an effective human-machine interaction, it is pivotal to put together machines able to recognize human affect.

Human affect is part of natural everyday communication. Affect can be expressed in various ways, for example, verbally or through facial expressions and hand gestures. Research shows that affective states can have a crucial role in the decision-making process [24], being strongly related to tendencies to perform actions. These capabilities and the ability to understand or feel and react to affective states of other human beings have been linked to the so-called *"emotional intelligence"* [17]. Including affect in the human-machines communication is defined as "affective computing" and *affect recognition* is an important problem in affective computing. One of the applications of affect recognition systems is to forge machines that are capable of detecting humans under stress conditions, with the goal of improving the interaction between humans and machines [26], [25].

The affective states of human beings can be recognized using sensing devices. For example, the polygraph, commonly referred to as the *lie detector*, was the first tool with the capabilities of estimating the true internal state of a human being [37]. In the affective computing field, various approaches have been established exploiting audio-visual expressions or the analysis of physiological signals and body movements [3]. Regarding sensor technologies, though, the commonly utilized sensors, for example, the electromyography (EMG) sensors, were lab-specialized devices that were bulky and expensive. These devices were especially intrusive with respect to the end-users. This did not allow extensive usages of such instrumentation in everyday settings [1]. The situation completely changed with the breakthroughs in hardware and software engineering, which led to the availability, first for professionals and academic purposes and then for the public market adoption, of wearable devices. The major contribution of wearable devices equipped with multiple sensors is the capability of capturing behavioral and physiological traits in a non-invasive manner.

In the research community, there are still open challenges that need to be tackled in order to make stress recognition ready for real-life deployment. A major issue is the uniqueness and thus the interpersonal variability of the physiological responses of different humans to stimuli [22]. Different people experiencing the same stressor, or even reporting the same affective state, can have different psycho-physiological reactions. This mismatch might cause a drop in the performance of stress recognition models. On the other hand, this psycho-physiological uniqueness can and has been used

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

recently to develop wearable biometric recognition systems with encouraging results [12, 28].

The main goal of this article is to investigate whether we can exploit the benefits of the psycho-physiological uniqueness observed in biometric recognition systems to improve affect recognition machine learning models. To do so, we use Multi-task learning (MTL) to jointly recognize users' stress levels and identity. The MTL paradigm could allow exploiting the commonalities and differences across these two tasks to improve the performance of the main task. A model able to jointly recognize user identity and affective state would enable the delivery of appropriate and personalized intervention strategies. MTL for jointly recognizing users and performing other classification tasks has been previously studied in other domains, with activity recognition from wearable accelerometers being the closes one to our work [31, 8]. To the best of our knowledge, this is the first study exploring the conjunction of physiological data from wearable sensors and user identification for stress recognition via MTL.

2 BACKGROUND

2.1 Stress recognition

Stress is considered a process by which a stimulus elicits an emotional, behavioral and/or physiological response, which is conditioned by an individual's personal, biological and cultural context [19]. The system that generates a response to a stress process, is called autonomic nervous system (ANS). This system is the primary mechanism in control of the fight-or-flight response, during which a mixture of hormones (like adrenaline) are released, leading, for example, to increased heart rate and sweating rate. These physiological changes prepare the organism for a faster response to the stress stimulus [20]. The human body can overreact to stressors that are not life-threatening (work/academic pressure, family difficulties, etc.). Phobias are good examples of how the fight-or-flight response might be falsely triggered in the face of a perceived threat.

The ANS is composed of two main subsystems, the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). The SNS is ultimately responsible for the fight-or-flight response [27]. It innervates all vascular smooth muscle and sweat glands. The stimulation of the SNS results in stimulation of sweat glands, accelerated heart rate, and increased blood pressure, among other effects. The PNS is responsible for the body's rest and digestion response when the body is relaxed, resting, or feeding. It undoes the damage caused by the SNS activation during a stressful situation. Stimulation of the PNS results in decreased heart rate and blood pressure, among other effects. The interaction between PNS and SNS can be quantified by analyzing changes in the related physiological signals, e.g., via the Heart Rate Variability. Additionally, the Electrodermal Activity (EDA) is used to measure changes in the SNS subsystem [15, 16].

2.2 User recognition

The task of recognizing users can be seen as a pattern recognition problem in which a user provides a set of physiological and/or behavioral characteristics in order to match a previously registered signature that allows the system to verify the subject itself [5]. This task takes advantage of the fact that humans have natural traits that are inherently unique for each individual. Typical applications involve the identification of users by means of controlling the access to resources such as transportation vehicles, computer systems, or a building. In such systems, the users need to actively present themselves to the system in order to be recognized.

2.3 Multi-task Learning

Multi-task learning (MTL) [6] is a specific *machine learning* (ML) paradigm in which a model is jointly trained for multiple tasks (the types of prediction or inference based on the problem at hand), instead of training on a single task, as in the more traditional ML paradigm of single-task learning. The architectural structure of such models is composed of shared layers and task-specific layers. The shared layers allow the model to be trained on multiple tasks, and thus to share common information about the two tasks. This common information should help the model to generalize better compared to single-task models. To build such a model, an inner relationship between the tasks should be present [41].

MTL models are typically performed via deep learning (DL) models [23]. MTL models allow the DL network to learn more generalized representations by learning different (but related) tasks jointly, resulting in a performance boost and a more robust model ready to be deployed in the wild. The effectiveness of MTL models can be seen in the fact that it utilizes more data from different learning tasks than single-task learning. Thanks to the larger volume of data, MTL models can obtain improved knowledge sharing among the different tasks, thus resulting in better performance and a lower risk of overfitting each of the learning tasks. Some fields in which MTL has been used successfully are natural language processing (NPL) [9], computer vision [42] and wearable computing [30], where the model architecture proposed showed that the shared network architecture allowed the models to be more effective.

3 DATASET

To investigate the impact of joint learning of stress recognition and user identification model using multi-task learning, we use the publicly available *WEarable Stress and Affect Detection (WESAD)* dataset [34]. *WESAD* includes data from 15 subjects (12 males and 3 females) recorded in a laboratory setting, each of the subjects involved in the study experienced four conditions: *Baseline* – reading task; *Amusement* – watching funny videos; *Stress* – exposure to the TSST [21]; and *Meditation*, de-excite procedure for bringing back the subjects to a more neutral affective state. The Baseline condition was recorded for 20 minutes, Amusement for 392 seconds ($\sim$ 6.5 minutes), Stress for 10 minutes, and lastly Meditation for 7 minutes. The WESAD dataset contains motion and physiological data recorded from a wrist and chest-worn device. For this particular article, the focus was on the *EDA* and *BVP* gathered via the wrist-worn device, an Empatica E4 [14]. The Empatica E4 records EDA and BVP at 4Hz and 64Hz sampling frequencies respectively.

4 METHOD

Fig. 1 shows the ML pipeline we designed. The pipeline consists of five major steps: *data acquisition*, *pre-processing*, *segmentation*, *feature extraction* and *feature cleaning*. More details of each step are presented in following sections.

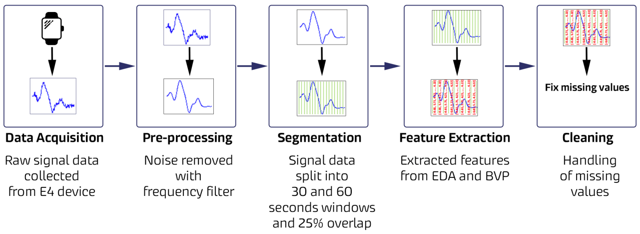

Figure 1: Architectural Overview of the *Data* pipeline

4.1 Pre-processing and segmentation

Physiological signals collected using wearable devices can be affected by artifacts that can hamper the reliability of the results. A typical source of artifacts is temporary misplacement of the device and motion artifacts. In order to reduce the presence of such artifacts, we applied a filtering procedure to the raw EDA and BVP data. Specifically, as a standard practice in the literature [36], we used a first-order Butterworth low-pass filter with a cutoff of 5Hz for BVP and 0.4Hz for EDA.

We then segmented the filtered sensor data using windows with a duration of 60 seconds and a 25% overlap. We then used the derived segments to extract the features.

4.2 Feature extraction

We extracted a total of 77 features from EDA and BVP used in the literature [32], to provide a high-level representation of the signals for the tasks of interest: *stress recognition* and *user identification*. We used the derived features in input to the machine learning models.
EDA features The EDA signal consists of two main components, *phasic* and *tonic* [2]. The phasic component is a fast-changing signal related to the momentary stimulus, while the tonic component is slow-changing [2]. The EDA is characterized by peaks also known as *skin conductance responses (SCRs)* [2]. We decomposed the EDA using the *cvxEDA* algorithm presented in [18], which uses methods of convex optimization to describe the activity of the autonomic nervous system in response to strong affective stimulation. We extracted a total of 50 features presented in the literature from the EDA and its components using the publicly available tool *EDA-Explorer* [11]. A summary of the features is presented in Table 1.
BVP features We derived 27 features from BVP to characterize cardiovascular activity. We extracted features in the time and frequency domain using the publicly available *HeartPy* tool [38] as well as statistical features (e.g., mean, standard deviation), as a common practice in the literature [13]. The list of the BVP features and their description is presented in Table 2.

4.3 Imputation and normalization

During the feature extraction procedure, it is possible that an accurate computation of some features may not be possible due to missing, noisy or invalid data originating from the starting dataset. For some features, the number of missing values amounted to 47.51% of the total of the available data points.

We imputed the missing values using an iterative approach based on multivariate regression. This approach involves defining a regression model (Bayesian ridge regressor) to predict each of the missing features as a function of all the other features. This method has been reported to produce more accurate estimations compared to the use of constants or statistical values [4].

We then normalized the features to reduce the data-related variability between subjects inside the dataset. The normalization is used to scale the numerical values between a known range. This allows the ML models to more quickly learn the optimal parameters for each input. In this paper, we used a Min-Max normalization procedure on a *per-feature basis*, scaling the extracted features values between 0 and 1.

4.4 Classification Pipeline

In this work, we experimented with a MTL strategy implemented using DNNs to jointly solve the stress recognition and user identification tasks. We compare the performance of this model with a single-task strategy and shallow classifiers. We describe below the models and the evaluation procedure we used.
Shallow classifiers. We considered two widely used [16, 34] machine learning algorithms: *Random Forest (RF)* and *Extreme Gradient Boosting (XGBoost)* as baselines. We chose these algorithms because they do not require extensive hyperparameter tuning and are robust against noisy data.
Multi-task learning (MTL) The MTL configuration is built using a Feed-forward Deep Neural Network (FDNN). In the MTL architecture, shown in figure 2, the tasks of stress recognition and user identification are unified into a single training procedure. The FDNN consists of a fully connected layer with 100 hidden units, a normalization layer, and a dropout layer. The initial fully connected layer has the role of a size balancer, meaning that it has to bring both inputs from AR and UR down to a commonly recognized size so that later they can be fed into a shared layer. The normalization layer is used as a model-optimization technique where the output of all the neurons is normalized [33]. In addition, it allows for a higher learning rate value to be used, thus improving the performance at which the network can learn [7]. The dropout layer, configured with a dropout rate of 0.2, is utilized to reduce the risk of model overfitting. In the MTL configuration, a shared dense layer is used to unify the outputs of the two branches. We used the *ReLU* activation function in the fully connected layers. We trained the MTL models with a learning rate of 10^{-3}, batch size of 15, and 50 epochs. The implementation of the MTL model is publicly available and can be accessed at [29].

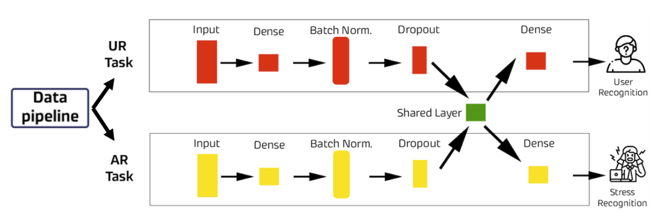

Figure 2: Architectural Overview of the *MTL* pipeline

Table 1: EDA features

Features	Explanation
$peaks$	# of peaks
$rise\_time$	avg rise time of the peaks
max_deriv	avg. value of the max derivative
amp	avg. amplitude of the peaks
$decay\_time$	average decay time of the peaks
SCR_width	average width of the peak for Skin Conductance Response
auc	average area under the curve for the peak
$median$	value separating the higher half from the lower half
$mean$	arithmetic mean
std	standard deviation
var	statistical variance
$slope$	slope of the regression line
min	minimum value of the signal
max	maximum value of the signal
$fdmean$	mean value of the gradient on the EDA signal
$fdstd$	standard deviation value of the gradient on the signal
$drange$	ratio between the largest and smallest values

Table 2: BVP features

Features	Explanation
bpm	beats per minute
ibi	interbeat interval
$sdnn$	std. dev. of intervals between adjacent beats
$sdsd$	std. dev. of differences in adjacent R-R intervals
$rmssd$	root mean sq. of differences in adjacent R-R intervals
$pnn20$	ratio of differences in R-R intervals > 20ms
$pnn50$	ratio of differences in R-R intervals > 50ms
hr_mad	median absolute deviation
$sd1, sd2, S, \frac{sd1}{sd2}$	Poincaré analysis
$breathing\_rate$	breaths per minute
lf	low frequency component
hf	high frequency component
lf/hf	low/high frequency ratio
$median$	value separating the higher half from the lower half
$mean$	arithmetic mean
std	standard deviation
var	statistical variance
$slope$	slope of the regression line
min	minimum value of the signal
max	maximum value of the signal
$fdmean$	mean value of the gradient on the BVP signal
$fdstd$	standard deviation value of the gradient on the signal
$drange$	ratio between the largest and smallest values

Evaluation procedure. We evaluated all the models using the *Leave-One-Subject-Out (LOSO)* validation procedure and the F_1-score metric with the *macro* average strategy, in which the metrics for each label are calculated and their unweighted mean is found [35]. The *LOSO* is an iterative approach in which the model is trained using the data of all the participants except one that is used for testing. The *LOSO* is a user-independent procedure suitable for determining the robustness of the model on data of unseen users.

5 RESULTS

5.1 Stress Recognition

In Table 3 the results of the LOSO evaluation are shown. The first two columns (WESAD-RF and WESAD-AB) are results from the *Random Forest* and *Ada Boost* classifier from [34], whilst the rest of the columns represent the results achieved by the methods implemented in this study. The XGBoost algorithm performs better with respect to the RF methodology, for what concerns this study. One main difference between the related-work results and the results in this work is that WESAD-RF and WESAD-AB utilized data from a wrist device and from an ECG device, whereas the methods implemented in this work use only the data from the wrist device. This may be the reason why the WESAD-RF scored 15.5 percentage points more than the RF implemented in this work achieved. On the other hand, the XGB model (using only wrist data) performed similarly to the related work models, which indicates that the pre-processing pipeline and the extracted features are informative enough to replicate the results of the related work. Furthermore, from Table 3 it can be seen that the MTL approach outperformed the best related-work model and the best baseline models implemented in this study as well.

To investigate the MTL model behavior for various test users, we present the learning curves from the LOSO evaluation in Figure 3. The x-axis presents the training epoch, the y-axis presents the test accuracy for the given epoch, the color-coded full lines represent the test accuracy for various test users, and the dashed black line represents the average test accuracy. From the figure, it can be seen that there are two test users for which the MTL model performed quite worse compared to the average performance. For the rest of the users, the learning curves are quite stable, which signifies that the MTL model is also stable.

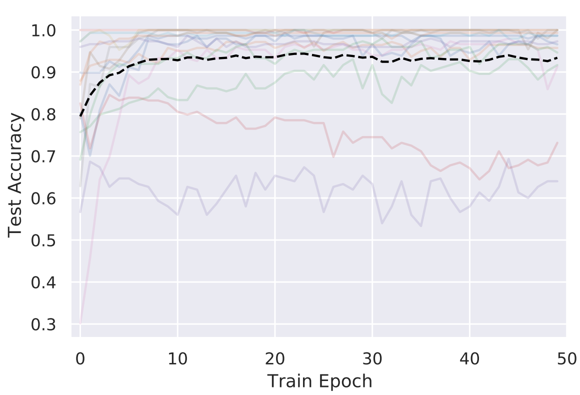

Figure 3: Learning curves for the MTL architecture

5.2 User-recognition

To evaluate the accuracy of the user-identification branch of the MTL network, we need data of the same user both in the train and in the test set, which is not plausible using LOSO evaluation. For that reason, we implemented an additional split of the data. To avoid temporal overlap between the train and the test data, we took the first 5 minutes and the last 2.5 minutes of both stress and no-stress data. Experiments were performed using LOSO approach again, but one subset of the test user was added to the train data (i.e., the first 5 minutes), and the other subset (i.e., the last 2.5 minutes) was kept for testing. The results of these experiments are presented in Table 4. While the MTL performed worse than the single-task models in this evaluation, the results still indicate that all of the models learn to recognize users to some extent, given that the accuracy of a majority classifier (a classifier that always predicts the most frequent user-id) would be 6.67% (1 out of 15 users).

Table 3: F1-score (%) MTL vs. Single task vs. Related work

WESAD-RF[34]	WESAD-AB[34]	RF	XGB	MTL
86.1	85.8	70.6	85.6	92.6

Table 4: F1-score (%) MTL vs. Single task

RF	XGB	MTL
52.0	40.0	29.7

REFERENCES

[1] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas. 2014. Window size impact in human activity recognition. *Sensors*, 14, 4, 6474–6499. DOI: 10.3390/s140406474.
[2] Wolfram Boucsein. 2012. *Electrodermal Activity*. Springer Science & Business Media.
[3] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *Int. J. Hum.-Comput. Stud.*, 59, 1–2, (July 2003), 119–155. DOI: 10.1016/S1071-5819(03)00018-1.
[4] Andriy Burkov. 2019. *The Hundred-page Machine Learning Book*. Vol. 1. Andriy Burkov Canada.
[5] C. Camara, P. Peris-Lopez, and J. E. Tapiador. 2015. Human identification using compressed ecg signals. *J Med Syst*, 39, 11, 149. DOI: 10.1007/s10916-015-0323-2.
[6] Rich Caruana. 1997. Multitask Learning. *Machine Learning*, 28. DOI: 10.1023/A:1007379606734.
[7] Guangyong Chen, Pengfei Chen, Yujun Shi, Chang-Yu Hsieh, Benben Liao, and Shengyu Zhang. 2019. Rethinking the usage of batch normalization and dropout in the training of deep neural networks. *CoRR*, abs/1905.05928. http://arxiv.org/abs/1905.05928.
[8] Ling Chen, Yi Zhang, and Liangying Peng. 2020. Metier: a deep multi-task learning based activity and user recognition model using wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4, 1, 1–18.
[9] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: deep neural networks with multitask learning. In *Natural Language Processing: Deep Neural Networks with Multitask Learning* (ICML '08). Association for Computing Machinery, Helsinki, Finland, 160–167. ISBN: 9781605582054. DOI: 10.1145/1390156.1390177.
[10] David Dearman and Jeffery S. Pierce. 2008. It's on my other computer! computing with multiple devices. ISBN: 9781605580111. DOI: 10.1145/1357054.1357177.
[11] 2021. EDA Explorer - A tool for the analysis of Electrodermal Activity (EDA) data. https://eda-explorer.media.mit.edu/. [Online; accessed 19-July-2021]. (2021).
[12] Deniz Ekiz, Yekta Said Can, Yagmur Ceren Dardagan, and Cem Ersoy. 2020. Can a smartband be used for continuous implicit authentication in real life. *IEEE Access*, 8, 59402–59411. DOI: 10.1109/ACCESS.2020.2982852.
[13] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. 1996. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93, 5, 1043–1065.
[14] Empatica. 2021. Empatica E4 wristband - Real-time physiological data streaming and visualization. https://www.empatica.com/research/e4/. [Online; accessed 7-July-2021]. (2021).

[15] Shadi Ghiasi, Alberto Greco, Riccardo Barbieri, Enzo Pasquale Scilingo, and Gaetano Valenza. 2020. Assessing autonomic function from electrodermal activity and heart rate variability during cold-pressor test and emotional challenge. *Scientific Reports*, 10, 1, (Mar. 2020), 5406. DOI: 10.1038/s41598-020-62225-2.

[16] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2017. Monitoring stress with a wrist device using context. *Journal of biomedical informatics*, 73, 159–170.

[17] Daniel Goleman. 1995. *Emotional Intelligence: Why It Can Matter More Than IQ*. Random House Publishing Group.

[18] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2015. Cvxeda: a convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63, 4, 797–804. DOI: 10.1109/TBME.2015.2474131.

[19] G. D. James H. G. Ice. 2007. *Measuring Stress in Humans: A practical Guide for the field*. Cambridge university press.

[20] 1989. *Autonomic nervous system. Human Physiology*. Springer Berlin Heidelberg, Berlin, Heidelberg, 333–370. ISBN: 978-3-642-73831-9. DOI: 10.1007/978-3-642-73831-9_16.

[21] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer. 1993. NeuropsychobiologyThe 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28, 1-2, 76–81.

[22] S. D. Kreibig. 2010. Biol PsycholAutonomic nervous system activity in emotion: a review. *Biol Psychol*, 84, 3, (July 2010), 394–421.

[23] Y. LeCun. 2020. The deep learning – applied math connection. https://bytes.usc.edu/cs585/m20_dBdSmLdM/extras/docs/LeCun_MDS20.pdf. [Online; accessed 21-June-2021]. (2020).

[24] J. S. Lerner, Y. Li, P. Valdesolo, and K. S. Kassam. 2015. Emotion and decision making. *Annu Rev Psychol*, 66, (Jan. 2015), 799–823.

[25] Daniel McDuff and Mary Czerwinski. 2018. Designing emotionally sentient agents. *Commun. ACM*, 61, 12, 74–83. DOI: 10.1145/3186591.

[26] Jose A. Miranda, Manuel F. Canabal, Jose M. Lanza-Gutiérrez, Marta Portela García, and Celia López-Ongil. 2019. Toward fear detection using affect recognition. In *2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS)*, 1–4. DOI: 10.1109/DCIS201949030.2019.8959852.

[27] Frank R. Noyes and Sue D. Barber-Westin. 2017. 40 - diagnosis and treatment of complex regional pain syndrome. In *Noyes' Knee Disorders: Surgery, Rehabilitation, Clinical Outcomes (Second Edition)*. (Second Edition ed.). Frank R. Noyes and Sue D. Barber-Westin, (Eds.) Elsevier, 1122–1160. ISBN: 978-0-323-32903-3. DOI: https://doi.org/10.1016/B978-0-323-32903-3.00040-8.

[28] Emanuela Piciucco, Elena Di Lascio, Emanuele Maiorana, Silvia Santini, and Patrizio Campisi. 2021. Biometric recognition using wearable devices in real-life settings. *Pattern Recognition Letters*, 146, 260–266.

[29] Alessandro Pogliaghi. 2021. Multi-task Learning for the Joint Recognition of User Identity and Affective States Using Physiological Signals. https://www.gitlab.com/tatoalo/mtl_affect_recognition_user_identity. [Online]. (2021).

[30] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3, 2, (June 2019), 1–30. DOI: 10.1145/3328932.

[31] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3, 2, 1–30.

[32] Stanisław Saganowski, Anna Dutkowiak, Adam Dziadek, Maciej Dzieżyc, Joanna Komoszyńska, Weronika Michalska, Adam Polak, Michał Ujma, and Przemysław Kazienko. 2020. Emotion recognition using wearables: a systematic literature review - work-in-progress. In DOI: 10.1109/PerComWorkshops48775.2020.9156096.

[33] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. 2018. How does batch normalization help optimization? DOI: 10.48550/ARXIV.1805.11604.

[34] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (ICMI '18). Association for Computing Machinery, Boulder, CO, USA, 400–408. ISBN: 9781450356923. DOI: 10.1145/3242969.3242985.

[35] Scikit-learn. 2022. Scikit-learn metrics - F1 Score. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html#sklearn.metrics.f1_score. [Online; accessed 12-March-2022]. (2022).

[36] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2010. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine*, 14, 2, 410–417. DOI: 10.1109/TITB.2009.2036164.

[37] Office of Technology Assessment U.S. Congress. 1983. *Scientific Validity of Polygraph Testing: A Research Review and Evaluation—A Technical Memorandum*. U.S. Congress.

[38] Paul van Gent, Haneen Farah, Nicole van Nes, and Bart van Arem. 2019. HeartPy: A Novel Heart Rate Algorithm for the Analysis of Noisy Signals. *Transportation research part F: traffic psychology and behaviour*, 66, 368–378.

[39] Mark Weiser. 1993. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36, 7, (July 1993), 75–84. DOI: 10.1145/159544.159617.

[40] Mark Weiser. 1999. The computer for the 21^{st} century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3, 3, (July 1999), 3–11. DOI: 10.1145/329124.329126.

[41] Yu Zhang and Qiang Yang. 2021. A survey on multi-task learning. (2021). arXiv: 1707.08114 [cs.LG].

[42] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2014. Facial landmark detection by deep multi-task learning. In *Computer Vision – ECCV 2014*. David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, (Eds.) Springer International Publishing, Cham, 94–108. ISBN: 978-3-319-10599-4.