Feature Extraction for Stress Detection in Electrodermal Activity

Erika Lutin[1,2,a], Ryuga Hashimoto[3,b], Walter De Raedt[2,c] and Chris Van Hoof[1,2,4,d]
[1]*Electrical Engineering-ESAT, KU Leuven, Kasteelpark Arenberg 10, Heverlee, Belgium*
[2]*Imec, Kapeldreef 75, Heverlee, Belgium*
[3]*Dept. of Mechanical and Intelligent Systems Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Japan*
[4]*Imec-Nl, OnePlanet Research Center, Stippeneng 2, Wageningen, The Netherlands*

Keywords: Electrodermal Activity (EDA), Feature Extraction, Stress.

Abstract: Electrodermal activity (EDA) is a sensitive measure for changes in the sympathetic system, reflecting emotional and cognitive states such as stress. There is, however, inconsistency in the recommendations on which features to extract. In this study, we brought together different feature extraction methods: trough-to-peak features, decomposition-based features, frequency features and time-frequency features. Regarding the decomposition analysis, three different applications were used: Ledalab, cvxEDA and sparsEDA. A total of forty-seven features was extracted from a previously collected dataset. This dataset included twenty participants performing three different stress tasks. A Support Vector Machine (SVM) classifier was built in a Leave-One-Subject-Out Cross Validation (LOOCV) set-up with feature selection within the LOOCV loop. Three features were consistently selected over all participants: 1) the number of responses in the driver function generated by Ledalab and 2) by sparsEDA and 3) a time-frequency feature, previously described as TVSymp. The classifier obtained an accuracy of 88.52%, a sensitivity of 72.50% and a specificity of 93.65%. This research shows that EDA can be successfully used in stress detection, without the addition of any other physiological signals. The classifier, built with the most recent feature extraction methods in literature, was found to outperform previous classification attempts.

1 INTRODUCTION

Prolonged exposure to psychological stress in daily life has been associated to various diseases such as depression and cardiovascular disease (Cohen et al., 2007). To prevent these adverse consequences, there is a strong need for correct quantification of personal stress (Epel et al., 2018). The body's response to stress can be quantified by measuring activation of the autonomic nervous system (ANS). The ANS consists of the sympathetic, parasympathetic and enteric branches. The main function of the sympathetic nervous system (SNS) is to initiate a rapid response in case of a dangerous or threatening situation, the so-called "fight or flight" response (Lovallo, 2005). As part of this response, the sweat glands become activated (Critchley, 2002), which causes a change in the electrical properties of the skin: the skin becomes more conductive. This change in electrical properties is referred to as Skin Conductance Response (SCR) or Galvanic Skin Response (GSR) (Boucsein, 2012). Many researchers have demonstrated a high correlation between Electrodermal activity (EDA) and cognitive and emotional processes (Critchley, 2002). However, in comparison to other physiological signals, such as an Electrocardiogram (ECG) (Camm et al., 1996), there are few guidelines on which features to extract. Recent reviews indicate multiple approaches but do not recommend a specific one (Topoglu et al., 2019; Posada-Quintero & Chon, 2020).

Electrodermal activity generally consists of two components. The first component is the tonic component, referred to as Skin Conductance Level (SCL). It varies slowly and changes only slightly within tens of seconds to minutes. The second

[a] https://orcid.org/0000-0002-9254-7374
[b] https://orcid.org/0000-0002-3189-1235
[c] https://orcid.org/0000-0002-7117-7976
[d] https://orcid.org/0000-0002-4645-3326

Lutin, E., Hashimoto, R., De Raedt, W. and Van Hoof, C.
Feature Extraction for Stress Detection in Electrodermal Activity.
DOI: 10.5220/0010244601770185
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 4: BIOSIGNALS*, pages 177-185
ISBN: 978-989-758-490-9

component is the phasic component, also referred to as the Skin Conductance Response (SCR). The phasic component represents a rapid change with a very short time to onset, usually between 1-5 seconds after the onset of a stimulus. If a response occurs in the absence of a stimulus, it is referred to as a non-specific SCR (NS.SCR) (Boucsein, 2012).

Regarding feature extraction in EDA, research has been focussed on the extraction of time domain features. Within the time domain, the focus has been on the evaluation of single SCRs, rather than patterns of multiple SCRs. Every SCR shows a characteristic course, which can be parameterized by features such as latency time, rise time, peak amplitude and recovery time (Boucsein, 2012). However, the evaluation of a single response becomes difficult in case of overlapping responses or superimposition. In the last decade, new analysis applications have been developed, which can handle superimposed SCRs (Benedek & Kaernbach, 2010b; Ghaderyan & Abbasi, 2016; Greco et al., 2016; Hernando-Gallego et al., 2018). These applications model EDA as a result of discrete bursts of sudomotor nerve activity, mathematically referred to as the driver function. By deconvolving the EDA signal into the driver function, and then convolving the peaks in the driver function with an SCR shape, these applications decompose an EDA signal in pure SCRs, independently from SCL and previous SCRs. Multiple features can be easily extracted from both the resulting SCL and SCR signal (Boucsein, 2012; Topoglu et al., 2019). Although analysis of EDA in the frequency domain has been described as less relevant (Boucsein, 2012), Posada et al. (2016a & 2016b) proposed two new frequency-related features, including one time-variant feature, which can be used to detect stress.

The purpose of the current study is to extract both traditionally studied features and recently developed features and discuss their performance with respect to stress detection.

2 METHODS

2.1 Data Collection

EDA data was collected in a previous study by Smets et al., (2016). The dataset consisted of twenty healthy participants (ten males and ten females, mean age = 40 years ± 10 years), who did not suffer from any mental or physical disease. The participants were asked to perform three stress tasks, during which EDA was recorded at the fingertip with the NeXus 10 – MK II (sampling rate = 32 Hz). Each of the three stress tasks was carried out for two minutes. The first task was the Stroop color-word test (Van Der Elst et al., 2006). During this task, color words are displayed in a different color as the words represent. For example, when the word *red* is presented in blue, the participant must answer the printed color (blue in this example). The second task was an arithmetic test in which the participant had to countdown from 1081 by subtracting 7 in a serial manner. The final task included a stress talk, in which participants were asked to talk about past stressful or emotionally negative events. In addition to these stress tasks, participants were instructed to count from zero to hundred out loud to control for the physiological response caused by vocalization. The counting task was performed before the Stroop color-word test and after the stress talk. All tasks were separated by a resting period of two minutes. The experiment was conducted in a quiet and controlled laboratory.

2.2 Feature Extraction

In this study, multiple methods for features extraction were performed: trough-to-peak features, decomposition-based features, frequency features and time-frequency features. Whereas time domain analysis is usually performed in small windows (about 10s) around a single Skin Conductance Response (SCR) (Lim et al., 1997), frequency domain analysis with inclusion of slow changing responses requires longer processing windows. Therefore, a window of 64s was selected. The windows were acquired with 32s overlap from the start point of each task to an integer multiple of 64 seconds. Pre-processing, i.e. filtering and downsampling, differed among the different feature sets, as it was performed according to the pre-processing procedures in the original research describing the feature set to be extracted.

2.2.1 Trough-to-Peak Features

In the first feature extraction method, SCR onset and latency were estimated based on the trough-to-peak analysis (Boucsein, 2012). Regarding this analysis, the procedure by Healey (2000) was adopted. The data was first pre-processed using a low-pass filter with a cut-off of 4 Hz. Thereafter, a threshold was applied on the derivative of the filtered EDA signal. Crossing the threshold indicated a new response if this happened more than one second away from other responses. For every new response, the onset (trough) and peak were determined as the zero-crossings of the derivative preceding and following the response respectively. Given the onset and peak values, four features could be extracted: the number of SCRs, the summed magnitude of the SCRs, the summed durations of the SCRs and the summed area under the SCRs.

2.2.2 Decomposition-based Features

The second set of features was extracted following the decomposition of the EDA signal into its tonic, phasic and driver components. Three different MATLAB applications were used: Ledalab (Benedek & Kaernbach, 2010b), cvxEDA (Greco et al., 2016) and sparsEDA (Hernando-Gallego et al., 2018).

The first application, Ledalab, provides two analysis methods to calculate the components: Discrete deconvolution Analysis (DDA) (Benedek & Kaernbach, 2010a) and Continuous Deconvolution Analysis (CDA) (Benedek & Kaernbach, 2010b). Whereas DDA applies a strictly nonnegative deconvolution, CDA merely tries to minimize negativity and therefore facilitates a more robust analysis. In this study, CDA was chosen as analysis strategy. The second application, cvxEDA, performs a nonnegative deconvolution by solving a convex optimization approach problem (Greco et al., 2016). The application parameters were defined as follows: $\tau\_0= 4.0$, $\tau\_1= 0.7$, $\alpha = 0.0008$, $\gamma = 0.01$. Lastly, sparsEDA also performs a non-negative deconvolution, though with specific focus on the sparse nature of the driver components (Hernando-Gallego et al., 2018). The application parameters were defined as follows: $\varepsilon = 0.0001$, $K\_max = 40$ iterations, $N_{min} = 5/4$ fs = 10 samples, $\rho = 0.025$.

These applications were selected above others as they do not require an a priori specification of events, whereas PsPM, another application by Bach et al., (2011) relies on information of the presented stimuli (Kelsey et al., 2018).

Following the pre-processing procedures described in Hernando-Gallego et al., (2018), the EDA data was first downsampled to 8 Hz. All three decomposition methods were applied on the complete signal, resulting in a continuous tonic, phasic and driver component for every MATLAB application. The sparsity of the driver components in both Ledalab and cvxEDA were increased post-extraction by applying thresholds of 0.1 and 0.01 respectively. From each of the components, statistical features, i.e. the mean, maximum, minimum and standard deviation, were extracted in windows of 64 seconds, resulting in a set of twelve features per application. Four additional features were extracted regarding the driver component: firstly, the number of responses, i.e. the number of non-zero elements after thresholding and secondly, the number of (inactive) intervals between responses, together with their mean and maximum length.

2.2.3 Frequency-based Features

The third set of features was derived from the work of Posada-Quintero et al. (2016a), who expressed a specific interest in sympathetic tone. The sympathetic tone has been described by time domain features such as the SCL and the NS.SCRs. These time domain features are, however, highly variable between persons. Therefore, Posada et al. proposed a new frequency-domain approach. Starting from the low frequency (LF) range for heart rate variability, they tested whether stressful tasks resulted in a spectral peak between 0.045 Hz and 0.15 Hz. Their results indicated a broader range of 0.045Hz to 0.25 Hz. Based on these results, they described a new feature called EDASymp, which represents the spectral power in this specific frequency band, calculated in windows of two minutes using Welch's periodogram (Blackman window of 128 points) with 50% data overlap. EDASympn is the normalized adaptation in which the power of the frequency band is divided by the total power. In this study, we examined the occurrence of a spectral peak during stress tasks and calculated continuous features based on EDASymp and EDASympn.

Preprocessing was performed in accordance with Posada-Quintero et al., (2016a). First, an 8th order Chebyshev Type I low-pass filter (0.8 Hz) was applied, followed by down-sampling to 2 Hz and finally, an 8th order Butterworth high-pass filter (0.01 Hz).

The spectral peak was examined, per task, in the following frequency ranges: VLF = 0-0.045 Hz, LF = 0.045-0.15 Hz, HF1 = 0.15-0.25 Hz, HF2 = 0.25-0.4 Hz and VHF = 0.4-0.5 Hz (Posada-Quintero et al., 2016a). The PSD was obtained via Welch's periodogram as described above.

The continuous frequency domain features, relative to EDASymp and EDASympn, were calculated for every 64 seconds window within the tasks. A Blackman window of 128 samples was applied to each window. The PSD was obtained via the Fast Fourier Transform (FFT).

2.2.4 Time-frequency-based Features

The final feature was also derived from Posada-Quintero et al. In addition to invariant frequency-domain analysis, Posada-Quintero et al., (2016b) proposed a time-variant index of sympathetic tone, called TVSymp. The time-frequency representation (TFR) of EDA was computed using the variable frequency complex demodulation (VFCDM), a time-frequency spectral (TFS) analysis technique that provides accurate amplitude estimates and one of the highest time-frequency resolutions (Wang et al., 2006). The components comprising the frequency power in the range from 0.08 to 0.24 Hz were used to compute TVSymp.

In this study, the raw EDA was first downsampled from 32Hz to 2Hz and thereafter high-pass filtered using a 2nd-order Butterworth filter (0.01 Hz). TVSymp was calculated by adapting the procedures described in Wang et al., (2006) and Posada-Quintero et al., (2016b). The resulting TVSymp was averaged per window of 64 seconds.

2.3 Classification

A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was used to classify rest (baseline, relaxing and counting) and stress (Stroop color-word test, arithmetic task, and stress talk). All features were standardized before training the classifier. As the sample size was rather limited, Leave-one-subject-out cross-validation (LOOCV) was used to evaluate the performan54ce of the classifier. Feature selection was performed within the LOOCV loop resulting in 20 sets of selected features. All features were ranked according to their correlation (point-biserial correlation) with the binary target variable (Hall, 1999) indicating the occurrence of either a resting task or a stress task. Only features with a correlation over 0.5 were retained. The remaining features were compared based on their correlation among each other. If features had a correlation over 0.85, only the feature with the highest target correlation was retained. Accuracy, sensitivity, specificity, F1-score and precision were calculated and averaged over all 20 cross-validation folds.

3 RESULTS

3.1 Feature Extraction

Figure 1 displays the tonic component of one participant as computed by the three applications: Ledalab, cvxEDA and sparsEDA. Similarly, the phasic components are shown in Figure 2, and the driver components in Figure 3. The light-yellow areas represent the counting tasks, the light-red areas represent the stress tasks. Figure 2 only displays the phasic components as computed by Ledalab and cvxEDA, since SparsEDA does not provide tools for the calculation of the phasic component. Ledalab and cvxEDA show a comparable decomposition, whereas sparsEDA illustrates a different approach. The driver function by sparsEDA is much sparser than the ones by Ledalab and cvxEDA. The tonic component, based on this sparse driver function, however, deviates from the true tonic component as it crosses the raw data

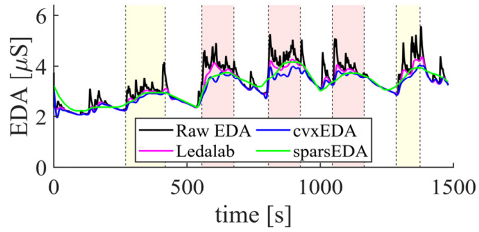

Figure 1: Tonic component of one participant. Light yellow areas indicate a counting task, light red areas a stress task.

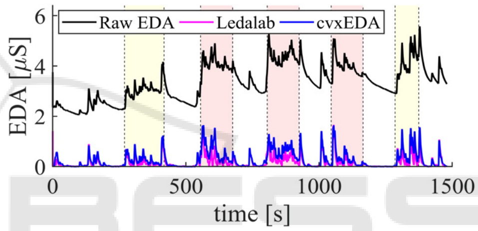

Figure 2: Phasic component of one participant. Light yellow areas indicate a counting task, light red areas a stress task.

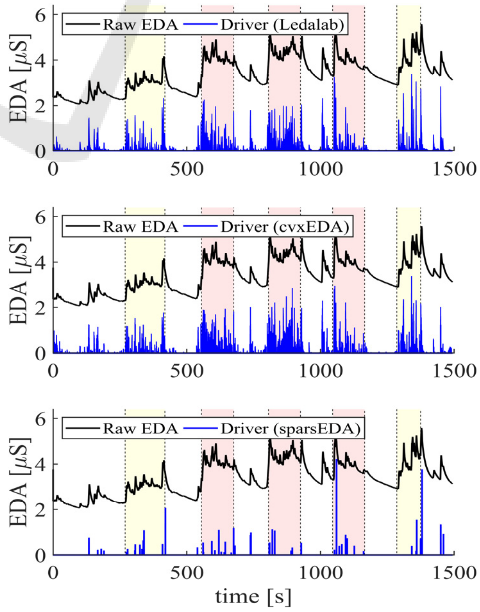

Figure 3: Driver component of one participant. Light yellow areas indicate a counting task, light red areas a stress task.

Table 1 summarizes the average calculation time for each application on the same machine, together with the standard deviation. The results show that sparsEDA has the lowest computation time, and Ledalab the highest.

Table 1: Computation time per application.

Application	Time
Ledalab	11.16 ±3.28
cvxEDA	2.40 ±0.45
sparsEDA	0.53 ±0.13

The results on the presence of a stress-related spectral peak are reported in Table 2. Five frequency ranges were examined: VLF (0-0.045 Hz), LF (0.045-0.15 Hz), HF1 (0.15-0.25 Hz), HF2(0.25-0.4 Hz) and VHF (0.4-0.5 Hz). The distribution of spectral energy is presented per task. Table 3 shows the values of the adapted TVSymp per task. The Stroop color-word test resulted on average in the highest value.

3.2 Classification

Table 4 presents the model performance of the SVM classifier. Table 5 shows the features selected in the LOOCV procedure.

In the group of selected features, features of both Ledalab and sparsEDA appeared. Post-classification analysis showed that limiting the decomposition features to features originating from either Ledalab or sparsEDA, lowered sensitivity but increased specifity further (SVM-spars; accuracy = 88.65%, sensitivity = 68.33%, specificity = 95.10%, F1 = 71.36%, SVM-Led; accuracy = 88.40%, sensitivity = 70.83%, specificity = 94.10%, F1=72.22%). When the same model was built with decomposition features originating only from cvxEDA, both sensitivity and specificity were lowered (SVM-cvx; accuracy = 86.67%, sensitivity = 69.17%, specificity = 93.27%, F1=70.21%)

Table 5: Features selected in LOOCV classification.

Feature	Votes
TVSymp (~ Posada-Quintero et al.)	20
Number of responses - Driver Ledalab	20
Number of responses - Driver sparsEDA	20
Summed duration of SCRs (~ Healey et al.)	4

4 DISCUSSION

The purpose of the current study was to extract both traditionally studied features and recently developed features and discuss their behavior as well as their performance in a simple classifier.

The first feature extraction method corresponded to the traditionally applied trough-to-peak analysis. In this analysis, SCRs are simply extracted based on zero-crossings of the derivative, while maintaining a minimal spacing between SCRs of 1 second. In four folds of the LOOCV, the summed duration of the SCRs was retained as an important feature. This finding agrees with the wide use of these features and

Table 2: Percentage of energy within the five frequency ranges presented per task.

Range	Baseline (%)	Relax (%)	Count (%)	Stroop (%)	Arithm. (%)	Talk (%)
VLF	70.27 ± 26.23	83.48 ± 17.04	59.32 ± 20.33	53.50 ± 20.82	53.59 ± 19.23	61.27 ± 19.10
LF	31.30 ± 23.09	17.68 ± 14.35	39.03 ±15.31	36.41 ± 15.45	42.06 ± 16.25	40.01 ± 16.04
HF1	5.43 ± 6.90	4.12 ± 5.80	10.42 ± 6.95	18.46 ± 15.17	13.27 ± 8.71	8.42 ± 7.11
HF2	0.66 ± 1.28	0.42 ± 0.76	1.33 ± 1.56	1.69 ± 1.85	1.80 ± 1.70	1.11 ± 0.92
VHF	0.11 ± 0.19	0.12 ± 0.24	0.49 ± 0.73	0.85 ± 1.10	0.61 ± 0.72	0.39 ± 0.48

Table 3: TVSymp results per task.

	Baseline	Relax	Count	Stroop	Arithmetic	Talk
TVSymp	0.59 ± 0.51	0.54 ± 0.16	1.03 ± 0.34	1.69 ± 0.46	1.62 ±0.46	1.51 ± 0.51

Table 4: Model performance.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)	Precision (%)
SVM	88.52 ± 10.99	72.50 ± 30.24	93.65 ± 9.60	72.84 ± 28.96	81.20 ± 26.37

the high accuracy (96%) in driver stress detection originally obtained by Healey (2000), but is contrary to the findings of Shukla et al., (2019), who reported low weighted occurence of these features in comparison to others in an emotion recognition task.

Although this method is easy to implement and could be used in a real-time setting, it is limited in the analysis of successive SCRs. They will most likely appear as one response (Benedek & Kaernbach, 2010b), which will lead to an overall underestimation of the number of responses.

To overcome the limitation of superimposed SCRs, we included the decomposition of the EDA signal as a second feature extraction method. We compared three decomposition methods. SparsEDA gave rise to the sparsest driver, which was claimed by the original authors to improve interpretability while reducing computation cost at the same time (Hernando-Gallego et al., 2018). This is, however, at the cost of an accurate tonic component as can be seen in Figure 1. Moreover, Amin & Faghih (2019) reported sparsEDA as oversparsifying the driver function. Ledalab and cvxEDA both gave rise to a more accurate tonic component, though at a much slower rate. In this study, it was argued that the sparsity and thus the interpretability of the driver functions of Ledalab and cvxEDA could easily be improved by introducing a threshold. Figure 3 illustrates that, nevertheless, sparsEDA remained the sparsest. Two decomposition-related features were retained in the LOOCV. Surprisingly, these were the number of driver responses of both sparsEDA and Ledalab. This suggests there is information in both sparsity and continuity. Limiting the classification to one single decomposition method showed that the sparisity of sparsEDA results in a higher specificity as it decreases the presence of drivers in resting periods, while the continuity of Ledalab results in a more robust sensitivity as it increases the presence of drivers in stress periods. Ledalab might be interchanged with cvxEDA to reduce time, though at cost of accuracy.

The third feature extraction method was linked to the frequency domain. We first tried to replicate the results of Posada-Quintero et al., (2016a), who investigated the spectral energy in multiple frequency ranges during four conditions: baseline, postural stimulation, a cold pressor and a Stroop color-word test. Only the latter corresponded to the protocol of the current study. Table 2 shows for all tasks the distribution in spectral energy. In this study, participants were fully at rest during the relaxation task and not during the baseline. Therefore, the results of relaxing resemble the results of the baseline in Posada-Quintero et al., (2016a) (VLF between 79.2% and 87.3%, LF between 8.1% and 14.6%, HF1 between 1.2% and 3.7%) more strongly. The distribution of spectral energy within the Stroop color-word test is also comparable to the one of Posada-Quintero et al., (2016a). The results show again the same trend (VLF 51.6%, LF 32.9%, HF1 10.7% in Posada-Quintero et al., 2016a). As described by Posada-Quintero et al., (2016a), during a stressor, the spectral energy in the VLF frequency range goes down, whereas in the other ranges the energy goes up. Remarkably, this trend is present in a comparable magnitude within the counting tasks as within the stress talk task. As the LF and HF1 ranges were confirmed as most interesting ranges, EDASymp and EDASympn were calculated according to Posada-Quintero et al., (2016a). They were not retained as important features during the feature selection.

The final feature to be extracted was an adaptation of TVSymp, a new time-frequency-based feature. Again, the presented results (Table 3) show the same trend as the data of Posada-Quintero et al. (2016b), i.e. the values of TVSymp go up in case of a stressor. The absolute numbers are comparable as well (baseline: ~ 0.2-0.5, stressor: ~ 1.4-1.6) . TVSymp was confirmed to be a highly sensitive index for stress detection in EDA, as it was selected in all folds of the cross validation. This in line with the conclusion of Ghaderyan & Abbasi (2016), who reported time-frequency domain features as highly performant in a mental workload classification task. Whereas in TVSymp the time-frequency representation is obtained by VFCDM, Ghaderyan & Abbasi (2016) obtained it via wavelets. In addition to time-frequency features, Ghaderyan & Abbasi (2016) extracted decomposition features via cepstral analysis. These were found to perform equally well as the wavelet features. The latter is contrary to the work of Shukla et al. (2019) who reported that cepstrum-based features outperformed wavelet-related features.

Classification using an SVM resulted in 88.52% accuracy, 72.50% sensitivity and 93.65% specificity. These results closely resemble the performance of an earlier machine learning exercise on this dataset which gave 72.0% sensitivity and 93.4% specificity (Smets et al., 2016) while this exercise included heart rate features on top of EDA features.

Previous studies have focused largely on datasets including multiple physiological signals. Only few studies have used EDA as a sole predictor for stress. Kurniawan et al., (2013) obtained 80.72% accuracy in classifying the Stroop color-word test from recovery with statistical features and trough-to-peak features. Liu & Du (2018) got an average accuracy of 81.82% in a three-level stress detection task on driving stress with statistical features. While the results of Kurniawan et al. (2013) and Liu & Du (2018) present a lower accuracy as reported in this study, Zangróniz et al., (2017) achieved an accuracy, sensitivity and specificity as high as 89.18%, 93.90% and 85.36% respectively using statistical features (directly derived from the EDA signal, i.e. without decomposition) and morphological features. The best performing parameter in a single-parameter classifier was the spectral power in bandwidth 0.2 Hz to 0.3 Hz, this parameter outperformed the parameter based on the spectral power in bandwidth 0.1 Hz to 0.2 Hz. This result differs from the bands suggested by Posada-Quintero et al., (2016a), which were confirmed in this study. Nevertheless, EDASymp was not retained in any of the feature selection folds.

A possible explanation for the difference in performance might be the experimental design of Zangróniz et al., (2017), as it was rather different from the one described in this paper. In the study by Zangróniz et al., (2017) calmness and distress were elicited by pictures of the International Affective Picture System (IAPS). This design might have favoured a well-balanced dataset without interference from vocalization, which is known to effect EDA (Levenson, 2014). Grimley et al., (2019) demonstrated that 36-78% of stress responses involving vocalizations are solely attributed to vocalizations. This is also apparent from the results presented in this study: the counting task shows highly increased EDA (Figures 1 – 3, Tables 2-3), while it is assumed to be free of stress. Since daily life measurements will often include vocalizations it is important to take this into account when building a classifier.

There are some limitations to this research. Firstly, the current study included twenty participants, which is rather limited for the multitude of features examined in this work. In future work, EDA data of more participants should be included.

Secondly, the current study included EDA data collected at the fingertip in a controlled environment. However, for future purposes, it would be more relevant to include data collected at the wrist in an ambulatory setting. This type of data would require more intensive pre-processing related to artefact removal.

Lastly, signal transformations such as deconvolution and VFCDM were performed prior to windowing, on the complete signal. This is consistent with prior research in which these transformations were performed on complete tasks or experiments (Bobade & Vani, 2020; Murugappan et al., 2020; Posada-Quintero & Bolkhovsky, 2019; Posada-Quintero et al., 2018). However, for future purposes such as continuous or real-time feature extraction, the effect of performing these transformations within windows as short as 64 seconds should be explored. The application of windows prior to transformation will abrupt ongoing responses which might introduce noise into the feature extraction. In light of this potential noise introduction, different window sizes should be explored.

5 CONCLUSION

The aim of the present study was to assess different feature extraction methods for stress detection in EDA. An SVM classifier was built in a Leave-One-Subject-Out Cross Validation (LOOCV) set-up with feature selection within the LOOCV loop. Decomposition-derived features and time-frequency features were found to be most relevant. The resulting classifier obtained an accuracy of 88.52%, a sensitivity of 72.50% and a specificity of 93.65%. Therefore, by including novel features, we could outperform an earlier classification attempt.

The research shows that EDA can be successfully used as a sole predictor for stress when using the most recent features in literature. In future research, the presented work should be repeated in a dataset collected in ambulatory settings. In addition, a continuous mode of feature extraction should be envisioned, which requires signal transformations such as decomposition to be performed in windows instead of complete signals. Related to this, different window sizes need to be explored.

ACKNOWLEDGEMENTS

This work was supported by a PhD fellowship from the Research Foundation - Flanders (FWO) awarded to EL (1SB4719N).

REFERENCES

Amin, M. R., & Faghih, R. T. (2019). Sparse deconvolution of electrodermal activity via continuous-time system identification. IEEE Transactions on Biomedical Engineering, 66(9), 2585–2595.

Bach, D. R., Daunizeau, J., Kuelzow, N., Friston, K. J., & Dolan, R. J. (2011). Dynamic causal modelling of spontaneous fluctuations in skin conductance.

Psychophysiology, 48(2), 252–257. https://doi.org/10.1111/j.1469-8986.2010.01052.x

Benedek, M., & Kaernbach, C. (2010a). Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology, 47(4), 647–658.

Benedek, M., & Kaernbach, C. (2010b). A continuous measure of phasic electrodermal activity. Journal of Neuroscience Methods, 190(1), 80–91.

Bobade, P., & Vani, M. (2020). Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 51–57). IEEE.

Boucsein, W. (2012). Electrodermal activity. Springer Science & Business Media.

Camm, A. J., Malik, M., Bigger, J. T., Breithardt, G., Cerutti, S., Cohen, R. J., … Kleiger, R. E. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology.

Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological Stress and Disease. JAMA, 298(14), 1685–1687. https://doi.org/10.1001/jama.298.14.1685

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. Neuroscientist, 8(2), 132–142. https://doi.org/10.1177/107385840200800209

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Mendes, W. B. (2018). More than a feeling: A unified view of stress measurement for population science. Frontiers in Neuroendocrinology, 49(March), 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Ghaderyan, P., & Abbasi, A. (2016). An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations. International Journal of Psychophysiology, 110, 91–101.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. IEEE Transactions on Biomedical Engineering, 63(4), 797–804.

Grimley, S. J., Ko, C. M., Morrell, H. E. R., Grace, F., Bañuelos, M. S., Bautista, B. R., … Gurning, J. (2018). The Need for a Neutral Speaking Period in Psychosocial Stress Testing. Journal of Psychophysiology.

Hall, M. A. (1999). Correlation-based feature selection for machine learning.

Healey, J. A. (2000). Wearable and automotive systems for affect recognition from physiology. Massachusetts Institute of Technology.

Hernando-Gallego, F., Luengo, D., & Artés-Rodríguez, A. (2017). Feature extraction of galvanic skin responses by nonnegative sparse deconvolution. IEEE Journal of Biomedical and Health Informatics, 22(5), 1385–1394.

Kelsey, M., Akcakaya, M., Kleckner, I. R., Palumbo, R. V., Barrett, L. F., Quigley, K. S., & Goodwin, M. S. (2018). Applications of sparse recovery and dictionary learning to enhance analysis of ambulatory electrodermal activity data. Biomedical Signal Processing and Control, 40, 58–70.

Kurniawan, H., Maslov, A. V, & Pechenizkiy, M. (2013). Stress detection from speech and galvanic skin response signals. In Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (pp. 209–214). IEEE.

Levenson, R. W. (2014). The autonomic nervous system and emotion. Emotion Review, 6(2), 100–112. https://doi.org/10.1177/1754073913512003

Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. International Journal of Psychophysiology, 25(2), 97–109. https://doi.org/https://doi.org/10.1016/S0167-8760(96)00713-1

Liu, Y., & Du, S. (2018). Psychological stress level detection based on electrodermal activity. Behavioural Brain Research, 341, 50–53.

Lovallo, W. R. (2015). Stress and Health: Biological and Psychological Interactions. SAGE Publications. Retrieved from https://books.google.be/books?id=kXtZDwAAQBAJ

Murugappan, R., Bosco, J. J., Eswaran, K., Vijay, P., & Vijayaraghavan, V. (2020). User Independent Human Stress Detection. In 2020 IEEE 10th International Conference on Intelligent Systems (IS) (pp. 490–497). IEEE.

Posada-Quintero, H. F., & Bolkhovsky, J. B. (2019). Machine learning models for the identification of cognitive tasks using autonomic reactions from heart rate variability and electrodermal activity. Behavioral Sciences, 9(4), 45.

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in electrodermal activity data collection and signal processing: A systematic review. Sensors, 20(2), 479.

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., & Chon, K. H. (2016a). Power spectral density analysis of electrodermal activity for sympathetic function assessment. Annals of Biomedical Engineering, 44(10), 3124–3135.

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., & Chon, K. H. (2018). Electrodermal activity is sensitive to cognitive stress under water. Frontiers in Physiology, 8, 1128.

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, Á. D., & Chon, K. H. (2016b). Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 311(3), R582–R591.

Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G. C., & Puig, D. (2019). Feature extraction and selection for emotion recognition from electrodermal activity. IEEE Transactions on Affective Computing.

Smets, E., Casale, P., Großekathöfer, U., Lamichhane, B., De Raedt, W., Bogaerts, K., … Van Hoof, C. (2015). Comparison of machine learning techniques for psychophysiological stress detection. In International Symposium on Pervasive Computing Paradigms for Mental Health (pp. 13–22). Springer.

Topoglu, Y., Watson, J., Suri, R., & Ayaz, H. (2019). Electrodermal Activity in Ambulatory Settings: A Narrative Review of Literature. In International Conference on Applied Human Factors and Ergonomics (pp. 91–102). Springer.

Van Der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006). The stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment, 13(1), 62–79. https://doi.org/10.1177/1073191105283427

Wang, H., Siu, K., Ju, K., & Chon, K. H. (2006). A High Resolution Approach to Estimating Time-Frequency Spectra and Their Amplitudes. Annals of Biomedical Engineering, 34(2), 326–338. https://doi.org/10.1007/s10439-005-9035-y

Zangróniz, R., Martínez-Rodrigo, A., Pastor, J. M., López, M. T., & Fernández-Caballero, A. (2017). Electrodermal activity sensor for Classification of Calm / Distress Condition, 1–14. https://doi.org/10.3390/s17102324.