2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9630535

Unsupervised Machine Learning Methods for Artifact Removal in Electrodermal Activity

Sandya Subramanian, *Student Member, IEEE,* Bryan Tseng, Riccardo Barbieri, *Senior Member, IEEE,* and Emery N Brown, *Fellow, IEEE*

***Abstract*— Artifact detection and removal is a crucial step in all data preprocessing pipelines for physiological time series data, especially when collected outside of controlled experimental settings. The fact that such artifact is often readily identifiable by eye suggests that unsupervised machine learning algorithms may be a promising option that do not require manually labeled training datasets. Existing methods are often heuristic-based, not generalizable, or developed for controlled experimental settings with less artifact. In this study, we test the ability of three such unsupervised learning algorithms, isolation forests, 1-class support vector machine, and K-nearest neighbor distance, to remove heavy cautery-related artifact from electrodermal activity (EDA) data collected while six subjects underwent surgery. We first defined 12 features for each half-second window as inputs to the unsupervised learning methods. For each subject, we compared the best performing unsupervised learning method to four other existing methods for EDA artifact removal. For all six subjects, the unsupervised learning method was the only one successful at fully removing the artifact. This approach can easily be expanded to other modalities of physiological data in complex settings.**

***Clinical Relevance*— Robust artifact detection methods allow for the use of diverse physiological data even in complex clinical settings to inform diagnostic and therapeutic decisions.**

I. Introduction

As it becomes common to collect physiological time series data in increasingly complex naturalistic scenarios outside of controlled experimental settings, the data are more susceptible to unanticipated and uncontrollable sources of artifact and interference. Most of this artifact can be clearly discerned by eye with minimal training. However, attempting to automate what can be seen by eye, such as by thresholding signal value or signal derivative, fails to be robust across datasets and subjects. Supervised learning, in which machine learning models learn to separate signal from artifact using labeled training datasets, has been very powerful in a number of other settings [1]. However, supervised learning is labor-intensive and impractical in the case of artifact detection, since each small time increment of data must be manually labeled for a large number of training datasets.

Nevertheless, the clear visual detectability of artifact in most physiological time series data indicates that artifactual data are fundamentally different in nature from true signal. Unsupervised machine learning methods offer an alternative that satisfies time and manual labor constraints since labeled training sets are not required, and they allow for the learning of more complex patterns not explicitly codified [2].

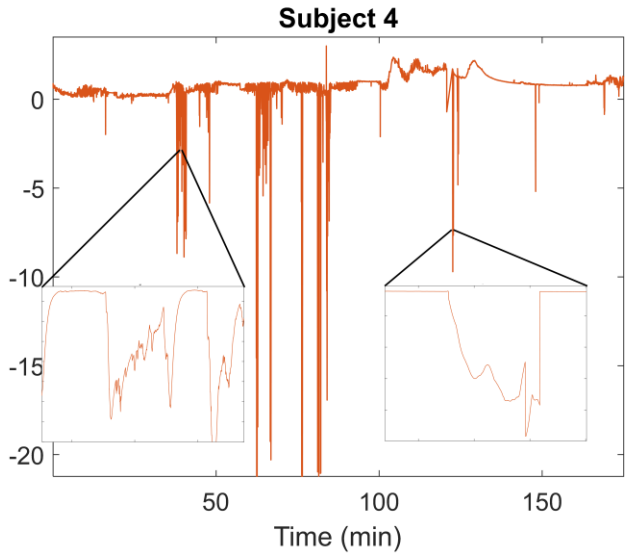

Figure 1. Raw EDA data for Subject 4 showing zoom-ins of different artifacts, where there is also true EDA data mixed in.

In this paper, we demonstrate the efficacy of unsupervised learning methods for artifact detection in electrodermal activity (EDA) datasets [3] using just 12 well-defined features. These data were collected continuously during lower abdominal surgery in 6 human subjects and were especially susceptible to motion and surgical cautery-related artifacts that cause large and clearly visible deflections in the data. While these artifactual deflections are clearly visible, to complicate matters, there are periods of intact but shifted EDA between large deflections. In addition, the beginning and end of each deflection is not clearly demarcated. Finally, the magnitude, sharpness, and direction of deflections varies across subjects and datasets. Figure 1 shows an example dataset with specific characteristics of artifacts labeled.

*This work was supported by a grant from the Picower Institute for Learning and Memory, Cambridge, MA and the National Science Foundation Graduate Research Fellowship.

SS is with the Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (phone: 616-406-6087; e-mail: sandya@mit.edu).

BT is with the Picower Institute of Learning and Memory, Cambridge, MA (bryan@neurostat.mit.edu).

RB is with the Department of Electronics, Informatics and Engineering, Politecnico di Milano, Milano, Italy. He is also with MGH and Massachusetts Institute of Technology, MA (riccardo.barbieri@polimi.it)

ENB is with the MGH Department of Anesthesia, Critical Care, and Pain Medicine, MIT Department of Brain and Cognitive Sciences, MIT Institute for Medical Engineering and Science, and Picower Institute for Learning and Memory, Cambridge, MA (enb@neurostat.mit.edu).

Existing methods for artifact removal from EDA are limited and specific to the datasets on which they were published [4-8]. These datasets were usually collected in fully controlled or semi-controlled experimental settings. None have the degree of artifact as in this study. Unsurprisingly, these existing methods are insufficient to completely remove the artifact when used on these data. In this study, no supervised learning methods were included to emphasize practicality of use and robustness across research settings.

In this study, we tested three unsupervised machine learning methods for artifact detection: isolation forest [9], K-nearest neighbor (KNN) distance [10], and 1-class support vector machine (SVM) [11]. We defined a set of 12 features for each half-second window as inputs to all of the unsupervised learning methods. These 12 features were taken from features used by existing methods and also adding features codifying what is visibly different by eye. We also compared the performance to the existing methods. We found that the unsupervised machine learning algorithms, using the features we defined, were successfully able to remove heavy artifact from EDA data across all 6 subjects, while none of the existing methods did. In ***Methods***, we discuss the details of our datasets, the features we defined, and our implementation of the unsupervised learning algorithms. In ***Results***, we show the EDA datasets before and after artifact removal using all of the methods included in the study (including existing methods). Finally, in ***Discussion*** and ***Conclusion***, we explain the implications of this study and our plans for future work.

II. Methods

A. Data

In this study, we use EDA data recorded from six subjects (2 female), collected under protocol approved by the Massachusetts General Hospital (MGH) Human Research Committee. All subjects were undergoing laparoscopic urologic or gynecologic surgery at MGH. The EDA data were recorded from two digits of each subject's left hand at 256 Hz using the Thought Technology Neurofeedback System [12], starting from before induction of anesthesia to just after extubation. Figure 1 shows an example of the raw data from one subject. The main sources of artifact were movement at the beginning and end, including positioning, and use of surgical cautery. Each instance of turning cautery on or off caused a visible deflection in the data. All data were analyzed using Matlab 2020b.

B. Features and Unsupervised Learning Methods

The 12 features that were defined with guidance from literature are listed in Table 1. These features were computed for each 0.5 second window (128 samples) for each dataset to match the timescale of individual artifacts. These feature vectors were then fed as inputs into three unsupervised learning methods.

KNN distance computes the average distance between each feature vector and the K nearest feature vectors in the data set [10], in this case using Euclidean distance and K of 50. Artifactual data is hypothesized to have a greater KNN distance than normal data [10]. 1-class SVM is similar to regular SVM; however, it is trained on data that is all labeled as a single class representing 'normal' data and tested on data that may also contain anomalies, assumed to be rare [11]. In this case, the 1-class SVM was trained on 90% of the data, excluding the 10% with the greatest KNN distance. Isolation forest is similar to random forest; however, each feature vector is scored based on the average length of the path to isolate it as a leaf in a forest of decision trees. Artifactual data is hypothesized to have shorter path lengths than normal data [9]. In this case, each isolation forest consisted of 100 decision trees, and the isolation scores were computed as the median of 10 forests.

TABLE I. Features

	Feature Description
1	Standard deviation of signal
2	Difference between max and min of signal
3	Mean of first derivative
4	Median of first derivative
5	Standard deviation of first derivative
6	Min of first derivative
7	Max of first derivative
8	Mean of level 4 Haar wavelet coefficients
9	Median of level 4 Haar wavelet coefficients
10	Standard deviation of level 4 Haar wavelet coefficients
11	Min of level 4 Haar wavelet coefficients
12	Max of level 4 Haar wavelet coefficients

All three unsupervised learning methods yielded scores for each window of data quantifying how anomalous that segment of data was (the isolation forest scores – IF scores – were made negative to match the directionality of the others). The last step of the process was to threshold the scores for each subject to determine artifact. The process used to select these thresholds relied on the insight that as the threshold is increased on each dataset, the portions of data that are labeled artifact decrease non-continuously, in discrete jumps. To take advantage of this, the skewness and kurtosis (3rd and 4th moments) of the inter-artifact interval distribution was computed across thresholds, since the inter-artifact interval distribution will become more skewed as the proportion artifact decreases. The thresholds at which local maxima in skewness and kurtosis occurred (large change in labeled artifact) were tested. Using a binary search method within this set of thresholds streamlined the threshold determination process to checking at most ~5 thresholds for each unsupervised learning method per subject. The final threshold was chosen based on visual inspection to ensure removal of artifact.

After identifying and removing the artifact, the gaps were filled using linear interpolation to result in continuous data. Any 'islands' of data that were shifted upward or downward due to artifactual deflection were translated downward or upward to match with the linearly interpolated mean of the data at that time. Finally, we compared our method to three other existing methods: variational mode decomposition [4,5], wavelet decomposition [6,7], and simple hardcoding of heuristic rules based on thresholding the derivative of the data.

III. Results

Table II summarizes the results from all three unsupervised methods for all six subjects. For each subject and each method, the proportion of artifact from 0 to 1 and the maximum contiguous length of artifact are given. The best method, by smallest proportion of artifact removed (removing the least excess signal) and shortest contiguous length of artifact, is in

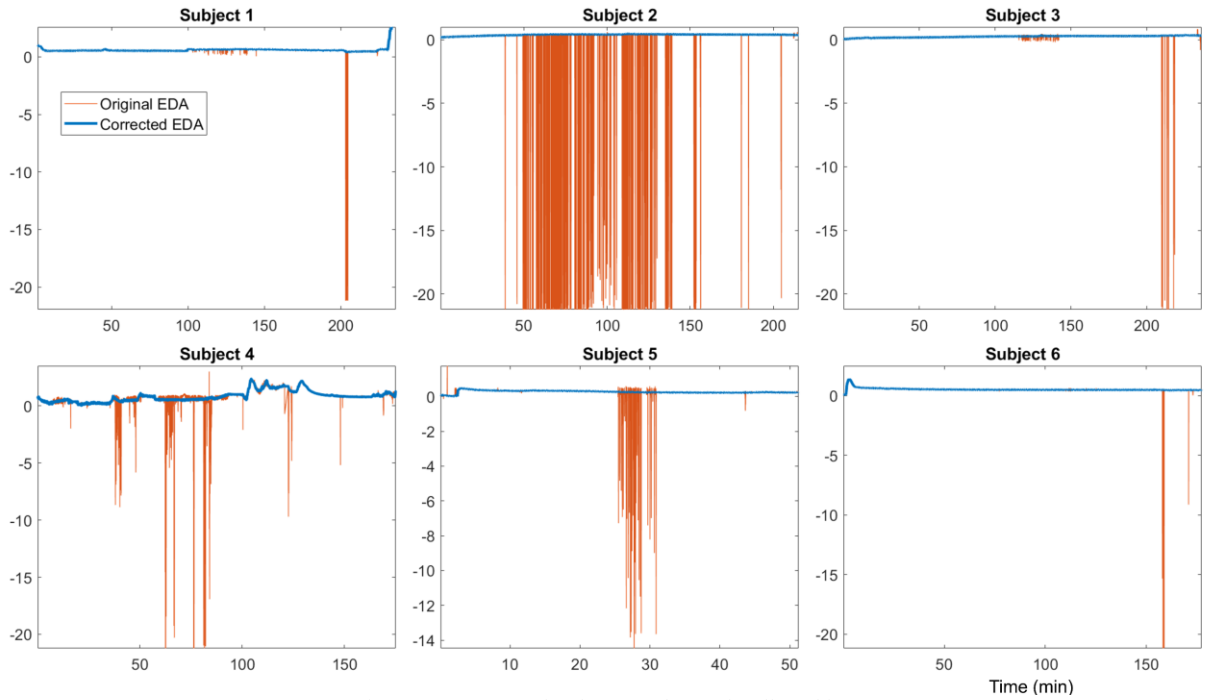

Figure 2. Uncorrected and corrected EDA for all 6 subjects.

bold for each subject. Isolation forest was the best method for 3 of the 6 subjects, KNN distance and 1-class SVM each for one subject, and all three methods were identical for one subject. Across all of the subjects, the proportions of artifact ranged from under 1% to just above 10% and the longest contiguous artifact from 6 seconds to 106 seconds.

TABLE II. SUMMARY OF RESULTS

Subj	Proportion artifact Max contiguous length of artifact (sec)		
	Isolation Forest	*KNN Distance*	*1-class SVM*
1	0.0876 17.5039	0.0729 17.5039	**0.0321**[a] **17.5039**
2	**0.1030** **26.0039**	0.1169 26.5039	0.1314 26.5039
3	**0.0361** **28.5039**	0.0398 28.5039	0.0422 28.5039
4	**0.0191** **105.5659**	**0.0191** **105.5659**	**0.0191** **105.5659**
5	0.1000 14.5039	**0.0974** **14.5039**	0.1062 14.5039
6	**0.0062** **6.0039**	0.0225 6.0039	0.0149 13.5039

a. Best model for each subject is in bold

Fig 2 shows the uncorrected and final corrected EDA data for all 6 subjects. The degree of artifact varied across subjects, but we were able to remove the artifact in all cases. Fig 3 shows an example of using the kurtosis of the inter-artifact interval distribution to select the optimal threshold for Subjects 2 and 4. Local maxima of the kurtosis, shown in Fig 3, were tested, and the highlighted values were selected as the final thresholds based on visual inspection of the corrected EDA data. Finally, Fig 4 shows a comparison between our method and several existing methods for Subjects 2 and 4. Of the methods compared, both variational mode decomposition and wavelet decomposition were ineffective, thresholding the EDA signal at 0 or hardcoding heuristic rules to threshold the derivative of the EDA signal were partially effective, and only our method was fully effective to remove the artifact.

IV. DISCUSSION

In this study, we used unsupervised machine learning methods combined with a set of 12 features we defined to remove heavy cautery and movement-related artifact from EDA data collected during surgery from 6 subjects. We compared three unsupervised learning methods (isolation forest, KNN distance, and 1-class SVM) with other existing methods (variational mode decomposition, wavelet decomposition, and hardcoded heuristic rules). For all 6 subjects, the unsupervised learning methods were the only ones successful at fully removing the artifact. We selected the best unsupervised learning method for each subject based on minimizing the amount of excess EDA signal removed along with artifact.

This method is notable because it did not require manual labeling of a training data set and yet was able to remove heavy artifact. In addition, our method allowed for preservation of as much true EDA signal as possible, even when it was interspersed in between sections of artifact. Even in cases of visibly intense artifact, the actual proportion of data detected as artifact was around 10% or less; any thresholding-based method would likely have removed a much larger proportion of the data, including true EDA signal. Most existing methods also employ decomposition

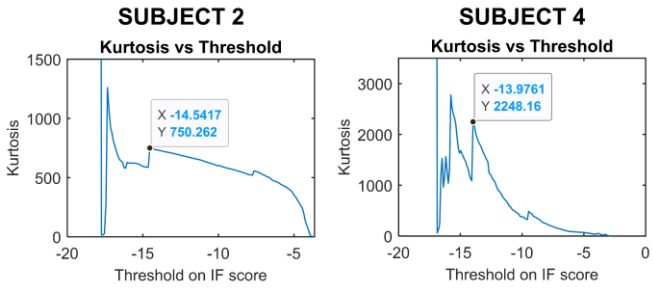

Figure 3. Use of kurtosis of inter-artifact interval distribution to select thresholds for Subjects 2 and 4. IF score refers to isolation forest score.

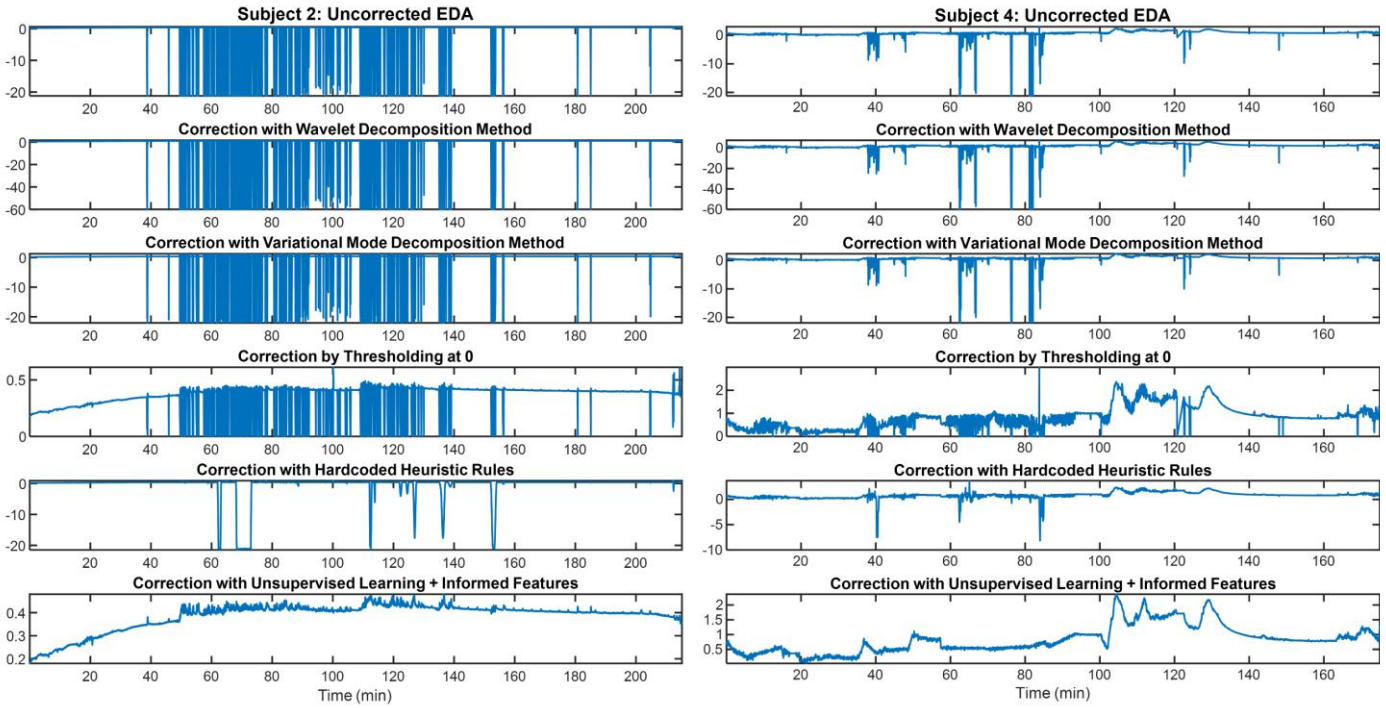

Figure 4. Comparison between different methods for Subjects 2 and 4.

algorithms that can affect the entire signal, including regions of signal that clearly have no artifact. Our method, in contrast, leaves non-artifact regions of the raw data unchanged. Most of this artifact was removed in very short segments, indicated by the fact that the longest continuous artifact was under 30 seconds for 5 out of the 6 subjects and under 20 seconds for 3 out of the 6 subjects.

EDA is typically analyzed in terms of two components that operate at different timescales and have complementary information, the tonic and phasic components [3]. The tonic component drifts slowly over the course of tens of seconds to minutes [3]; which can easily be interpolated when sections of artifact are short. For phasic EDA, short durations of data (< 30 seconds) are unlikely to contain more than a few pulses [3]. Furthermore, dynamic methods can compute mean and standard deviation pulse rate over time even if a few pulses are missing in short segments; they can account for this missing data in the estimate of uncertainty [13].

Finally, our method used only 12 features for each window, and many of these features overlapped with those used by existing methods. However, our method allowed AI to "learn" the differences between artifact and signal for each dataset. These 12 features were informed by knowledge of the physiology of EDA data. Our method can easily be expanded to similar classes of "easily visible" artifact in other modalities of physiological data, such as ECG and EEG. Custom feature definition can be informed by the knowledge of the physiology and types of artifact in those data.

ACKNOWLEDGMENT

S.S. thanks the Department of Anesthesia at MGH, the operating room staff, and Drs. Douglas Dahl, Marcela Del Carmen, and Annekathryn Goodman for their support in carrying out this study.

REFERENCES

[1] G. Biagetti, P. Crippa, L. Falaschetti, G. Tanoni, C. Turchetti, "A comparative study of machine learning algorithms for physiological signal classification," *Procedia Computer Science*, vol. 126, pp. 1977-1984, 2018.

[2] M. Goldstein, S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," *PLoS ONE,* vol. 11, no. 4, pp. e0152173, Apr 2016.

[3] W. Boucsein, *Electrodermal Activity*. New York, NY: Springer, 2012.

[4] C. Qi, R.T. Faghih, "Detection of Autonomic Sympathetic Arousal from Electrodermal Activity," PowerPoint presentation (unpublished).

[5] K. Dragomiretskiy, D. Zosso, "Variational Mode Decomposition," *IEEE Trans. Sig. Proc.*, vol. 62, no. 3, pp. 531-544, Feb. 2014.

[6] W. Chen, N. Jacques, S. Taylor, A. Sano, S. Fedor, R. W. Picard, "Wavelet-Based Motion Artifact Removal for Electrodermal Activity," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 6223-6226, Aug 2015.

[7] S. Taylor, N. Jacques, W. Chen, S. Fedor, A. Sano, R. Picard, "Automatic Identification of Artifacts in Electrodermal Activity Data," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 1934-1937, Aug 2015.

[8] Y. Zhang, "Unsupervised Motion Artifact Detection in Wrist-Measured Electrodermal Activity Data," M.S. thesis, Dept. Electric. Eng., Univ. of Toledo, Toledo, Ohio, 2017.

[9] F. T. Liu, K. M. Ting, Z-H. Zhou, "Isolation Forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, pp. 413-422, 2008.

[10] L-Y. Hu, M-W. Huang, S-W. Ke, C-F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *SpringerPlus*, vol. 5, no. 1304, Aug 2016.

[11] L. M. Manevitz, M. Yousef, "One-Class SVMs for Document Classification," *Journal of Machine Learning Research*, vol. 2, pp. 139-154, 2001.

[12] "Neurofeedback Expert System", Thought Technology Ltd, https://thoughttechnology.com/neurofeedback-expert-system/, accessed 1/6/21.

[13] R. Barbieri, E. C. Matten, A. A. Alabi, E. N. Brown, "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 288, no. 1, pp. H424–435, Jan 2005.