

Automatic Artifact Recognition and Correction for Electrodermal Activity in Uncontrolled Environments

Jose Llanes-Jurado (✉ jllajur@i3b.upv.es)

Universitat Politècnica de València

Lucía Amalia Carrasco-Ribelles

Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)

Mariano Alcañiz

Universitat Politècnica de València

Javier Marín-Morales

Universitat Politècnica de València

Research Article

Keywords: Artifact correction, deep learning, electrodermal activity, galvanic skin response, machine learning, phasic component, skin conductance

Posted Date: July 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-717360/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Automatic artifact recognition and correction for electrodermal activity in uncontrolled environments

Jose Llanes-Jurado^{1,*}, Lucía A. Carrasco-Ribelles², Mariano Alcañiz¹, and Javier Marín-Morales¹

¹Universitat Politècnica de València, Spain, Instituto de Investigación e Innovación en Bioingeniería (i3B), Valencia, 46022, Spain

²Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, 08007, Spain

*jllajur@i3b.upv.es

ABSTRACT

Scholars are increasingly using electrodermal activity (EDA) to assess cognitive-emotional states in laboratory environments, while recent applications have recorded EDA in uncontrolled settings, such as daily-life and virtual reality (VR) contexts, in which users can freely walk and move their hands. However, these records can be affected by major artifacts stemming from movements that can obscure valuable information. Previous work has analyzed signal correction methods to improve the quality of the signal or proposed artifact recognition models based on time windows. Despite these efforts, the correction of EDA signals in uncontrolled environments is still limited, and no existing research has used a signal manually corrected by an expert as a benchmark. This work investigates different machine learning and deep learning architectures, including support vector machines, recurrent neural networks (RNNs), and convolutional neural networks, for the automatic artifact recognition of EDA signals. The data from 44 subjects during an immersive VR task were collected and cleaned by two experts as ground truth. The best model, which used an RNN fed with the raw signal, recognized 72% of the artifacts and had an accuracy of 87%. An automatic correction was performed on the detected artifacts through a combination of linear interpolation and a high degree polynomial. The evaluation of this correction showed that the automatically and manually corrected signals did not present differences in terms of phasic components, while both showed differences to the raw signal. This work provides a tool to automatically correct artifacts of EDA signals which can be used in uncontrolled conditions, allowing for the development of intelligent systems based on EDA monitoring without human intervention.

Keywords: Artifact correction, deep learning, electrodermal activity, galvanic skin response, machine learning, phasic component, skin conductance.

Introduction

Electrodermal Activity (EDA) is a non-stationary signal that refers to the electrical potential which comes from the sweat glands on the surface of the skin¹⁻³. Sudomotor nerves originates EDA signals as a manifestation of the eccrine sweat glands activity. These glands are innervated by the sympathetic branch of the autonomic nervous system (ANS). Sweat glands located on the palmar surface evolved to increase grip and enhance sensitivity, and are responsive to psychological significant stimuli³. EDA represents a quantitative functional measure of sudomotor activity and due to that, an objective assessment of arousal⁴. Its shape shows a slow decrease over time when the subject is at rest and increase rapidly when an arousal stimulus is introduced^{2,3}. EDA signal can be decomposed into two different and non-redundant components named phasic and tonic¹⁻⁵. Both components have different temporal dynamics and relationships with the triggering stimuli. Phasic component is the decomposition of the rapid movements of the signal, called skin conductance response (SCR), and it finds applications in different research fields. The spike density is manifested through certain amplitude in the SCR signal facing a stimulus. Tonic component represents the slowest variations of the original ANS dynamics. This component is referred as the subject's basal conductance¹⁻³. In clinical analysis, SCR is used for the assessment of pain, stress, schizophrenia and peripheral neuropathy^{4,5}. In other fields as neuroscience and psychology, the phasic component is used for the assessment of the subject's arousal levels³. The obtention of the decomposition between phasic and tonic has been mainly achieved through the use of continuous decomposition analysis (CDA) algorithm¹⁻⁵. However, novel algorithms as cvxEDA⁶ show different methodologies to obtain this decomposition.

EDA signal can be measured using monitoring devices. They provide valuable information in an inexpensive and straightforward manner and constitute a compelling source of data in many scientific research fields, particularly psychology and health-related studies⁷. For example, Anusha et al.⁸ used EDA signals to assess subjects' stress in emulated real-life job

scenarios, while Zangróniz et al.⁹ studied EDA to distinguish between stressful or calm conditions. Liu et al.¹⁰ also analyzed the stress levels of a subject through this signal. Other works related to mental illness have utilized EDA signals. For instance, Greco et al.¹¹ found statistical evidence concerning healthy patients and patients with bipolar disorder through features of EDA signals. Moreover, Perugia et al.¹² discovered significant correlations between EDA signals and engagement in dementia patients.

The majority of the previous work collected the signals in lab environments¹³ where the subjects are usually seated and are often asked not to move the hand to which the electrodes are attached. However, recent applications have recorded EDA in environments where the users can freely walk and move their hands, such as daily-life settings and virtual reality (VR) environments. First, many wearable devices have been developed to enable the possibility of acquiring EDA signals in a daily-life scenario². Malathi et al.¹⁴ proposed a wearable EDA sensor to detect drowsiness in drivers. Leite et al.¹⁵ analyzed the affective state of children in users' everyday settings while interacting with robots, and Kim and Fesenmaier¹⁶ measured travelers' emotions in real time with EDA during a four-day visit. Second, VR has been used to simulate environments wherein the subject can freely move and interact, creating the sensation of being in the real world¹⁷. VR can display different scenarios to evoke emotions or cognitive processes within the subject^{18,19}. It has been used for many case studies, such as social adaptation in social phobia contexts, the reduction of anxiety and pain, rehabilitation, and neurological diagnosis²⁰⁻²⁵. EDA has been used in VR experiments to examine sudomotor activity and arousal levels to assess anxiety and stress^{24,26}, conduct emotional assessments²⁷ and diagnose autism²⁵.

However, one of the most significant issues concerning the use of EDA signals in daily-life and VR environments is the subject's movement during data collection. Although these technologies can offer an accurate environment to record subjects' responses, they are uncontrolled environments that could affect EDA records. Most movements can cause interferences in the contact between the skin and the recording electrodes, producing major artifacts in EDA recordings^{2,28}. Shukla et al.¹³ suggested that the presence of artifacts in EDA signals may conceal the existence of important correlations between the signal and the subject's stress levels.

Most of the experimentations that use EDA signals manually remove these major artifacts since there is no established methodology to automatically remove artifacts and correct EDA signals. Artifacts can be manually corrected using different software, such as Ledalab²⁹ or SCRalyze³⁰. Nevertheless, this is a time-consuming and tedious task that many researchers are forced to undertake for the correct analysis of the signal^{2,25}. Furthermore, manual correction can add a subjective human bias. Another central limitation of this methodology is that it cannot be applied in real time or in a short time period without human intervention, which is critical in the development of automatic, intelligent wearables whereby evaluations must be performed without human intervention. Examples of such mechanisms include adaptive VR therapies that use EDA to evaluate the stress reaction of a subject inside a virtual environment or intelligent systems to diagnose autism spectrum disorder²⁵. Since studies are increasingly collecting EDA data, an automatic algorithm for the detection and correction of artifacts in EDA signals appears to be essential for future research².

Previous work

Several works have studied how to correct EDA signals automatically. Most contributions come from the signal processing field, which has proposed using low-pass filters or exponential smoothing for artifact correction³¹. However, these approaches can modify certain segments of an EDA trace, affecting genuine physiological responses and thereby creating more artifacts^{13,28}. Other studies have used wavelet stationary transform models to remove the artifacts in EDA signals automatically. Chen et al.³² modeled the wavelet coefficients using a Gaussian mixture distribution. This model requires the estimation of three parameters using the expectation-maximization algorithm. The breakthrough work of Greco et al.⁶ studied an automatic model called cvxEDA that decomposes the EDA signal in linear terms of tonic and phasic components and a Gaussian noise that represents the white noise of the signal. This algorithm facilitates the direct decomposition of the EDA signal into two main components while the noise term of the signal is removed. This model is based on Bayesian statistics and convex optimization. Greco et al.⁶ demonstrated a better performance of cvxEDA in the finer discrimination of arousal levels than another decomposition algorithm, CDA. In addition, cvxEDA has been used in other experiments^{33,34} due to its low computational cost and efficiency. Finally, Shukla et al.¹³ proposed a transformation based on stationary wavelet transform models. This method also used a zero-mean Laplace distribution to model the wavelet coefficients and only needed the estimation of one parameter. Nevertheless, although the filtering methods of the previous works showed improvements in signal quality, the researchers only compared their findings against the raw signal and did not compare their results with an EDA signal corrected manually by an expert as a ground truth. Moreover, they did not analyze the implications of artifact recognition in the phasic component of the signal by comparing them with the phasic component of a manually corrected signal as a ground truth.

Meanwhile, some studies have examined EDA artifact recognition. Kleckner et al.³⁵, for instance, explored how to recognize EDA artifacts using a model founded on four rules. This set of rules is based on the minimum and maximum range of an EDA signal or its variation over time. Nonetheless, scant research has focused on automatic artifact detection in EDA signals

using machine learning (ML) methodologies. The work of Taylor et al.²⁸ detected motion artifacts in 5s EDA segments at a sampling frequency of 8Hz. The researchers used different features extracted from the raw EDA signal as statistical variables, such as mean, maximum and minimum value, and wavelet coefficients, to classify between artifact or non-artifact. The test achieved an accuracy of 0.96 using a support vector machine (SVM) model. However, the proportion of artifacts was not reported, which biases the metrics reported from the model. The same methodology and objective were followed by Gashi et al.³⁶. The dataset used in the latter work was more extensive than that of other experiments, achieving a total of 107.56h between 13 different participants. The collected data included an ambulatory EDA signal with a sampling frequency 32Hz, which was subsequently resampled to 8Hz. Finally a true positive ratio (TPR) of 0.98 was achieved in validation. There were, however, some limitations to this study. First, the recognition of artifacts was in a time interval of 5s segments; second, there was a lack of evaluation of artifacts in the whole signal and the provision of a final corrected signal. Furthermore, the final dataset documented 48.96% artifacts, which is dissimilar to its initial unbalanced percentage of artifacts (17%). A different approach can be found in Zhang et al.'s work³⁷, which analyzed the use of unsupervised learning to identify these artifacts from the raw signal, achieving competitive results compared with supervised learning.

Thus, previous artifact recognition works involve some limitations since researchers did not test the recognition models over a whole, continuous EDA signal. As they do not provide a final, clean signal, the results cannot be compared with the signal corrected manually by an expert as a ground truth. In addition, they do not present a precise characterization of the motion artifacts on the signals, including the total number, duration, or percentage of the signal affected, which is especially important to assess the level of noise in the works in daily-life or VR environments with major artifacts.

The most prominent existing studies on EDA signal filtering and artifact recognition are summarized in Table. 1. The table also presents the total signal time used in each of the experiments.

Table 1. Summary description of each study related to the automatic correction or recognition of artifacts methodology in EDA signal, compared with the characteristics of the proposed work.

cvxEDA ⁶	Bayesian statistics and convex optimization	Signal correction	Static breath task and simulated data	Raw signal	5.1 h (30 subjects)
Shukla et al. ¹³	Wavelet transformation	Signal correction	Driving	Raw signal	20.11 h (15 subjects)
Chen et al. ³²	Wavelet transformation	Signal correction	Physical, cognitive and emotional tasks	Raw signal	81.5 h (32 subjects)
Taylor et al. ²⁸	Support vector machine	Artifact recognition	Physical, cognitive and emotional tasks	Expert labelling of 5s window	2.17 h (32 subjects)
Zhang et al. ³⁷	<i>k</i> -Nearest neighbour	Artifact recognition	Static task in laboratory and real-life walking	Expert labelling of 5s window	23 h (21 subjects)

Concerning the model used to recognize artifacts in EDA, no previous work has explored the use of deep learning (DL) algorithms, which have been applied in recent years to other physiological signals, such as an electrocardiogram (ECG) or electroencephalogram (EEG). Models such as U-Net³⁸, ResNet³⁹ or recurrent neural networks (RNNs) have shown good performances and versatility in different types of health-related problems. For example, Kyathanahally et al.⁴⁰ used two DL models for ghost artifact correction in EEG spectrograms. The first model classified whether the spectrogram had a spurious echo in it or not. The second model conducted a regression over the spectrogram to correct the artifact. This work shows promising results for the detection of ghost artifacts in EEG. With regard to ECG signals, the work of Bento et. al.⁴¹ utilized two different DL classification models to study the classification of atrial fibrillation. Finally, RNN was used by Antczak⁴² to automatically denoise ECG signals. Various models were tested, including re-trained ones, and results improved for the pre-trained models with artificial ECG signals.

Objectives

This work has developed an automatic recognition and correction algorithm for EDA signals, providing an artifact-free corrected signal to be used in highly noisy environments. We examined three different approaches from the DL field that have been previously unexplored for artifact recognition in EDA signals. We compared these approaches with one previous, state-of-the-art ML method. A total of 76.46h hours of EDA signal recordings were collected from 44 participants in an uncontrolled virtual environment wherein the participants had to undergo different tasks that required bodily and hand movements. Two experts

corrected the signals, generating an artifact-free signal to be used as a ground truth. The labels obtained from the manual correction were used to train and test the artifact recognition models. Subsequently, an automatic correction was performed in the artifacts detected by the mechanism selected as the best recognition model. Finally, to measure the quality of the automatic correction, we evaluated the phasic components, in pairs, between the automatic correction, the manual correction, and the original raw signal using two different decomposition algorithms: CDA and cvxEDA. The main objective of this work was to automatically recognize and correct EDA artifacts and achieve an automatically corrected signal that did not show differences in terms of phasic components compared to the manual correction done by the experts.

Materials and Methods

Participants

A group of 44 volunteers (13 females and 31 males) was recruited to participate in the experiment. The mean age of the group was 37.52 (SD = 8.38). Inclusion criteria were: age between 18 and 50 years, Spanish nationality, and not having any previous VR experience. Prior to the subject's participation, they received documentary information on the study and gave their informed consent for their involvement. All methods and experimental protocols were performed accomplishing the guidelines of the ethics committee of the Universitat Politècnica de València. The experimental protocol was approved by the ethics committee of the Universitat Politècnica de València (P4_18_06_19).

Data collection

The objective of the VR study was to induce stress on the subject through daily situations at work that were simulated in a virtual environment. The participants had to perform different tasks in the virtual scenario to achieve this objective. First, subjects were placed in an office setting and talked to a virtual avatar about issues related to work and personal life. Second, the subjects were moved to another scenario: a meeting with five virtual avatars where they had to actively participate in the decisions made (see Fig. 1). In all the settings that required conversations with the avatars, the subjects could choose between four options that were displayed on the lower part of the screen. Finally, the participants played three different minigames. The first minigame involved extinguishing a fire as fast as possible from some virtual trees. The second minigame entailed reorganizing a pipe to allow the water to flow through in the minimum possible time. In the last minigame, the subjects had to complete a maze while simultaneously solving simple arithmetic equations as a parallel task. The faster the subjects solved both problems, the higher their score. In all three minigames, the participants had to move both of their hands to complete the games. As such, the EDA signal became noisier in the minigames section due to the induced stress and the subjects' rapid movements.



Figure 1. Second part of the VR scenario used, from the subject's perspective.

The subjects performed the VR scenario with a HTC Vive Pro-eye head mounted display, with 1440×1600 pixels per eye, a field of view of 110° , working at 90Hz refresh rate. EDA data was recorded using a Shimmer3 together with Consensus software⁴³, with a sampling frequency of 128Hz. In total, 44 EDA signals were collected. The average experiment duration was $104 \pm 8min$, having a total time of 74.46h of signals.

Manual artifact correction

The procedure to obtain the manual signal correction was as follows. The expert cleaned the signal using Ledalab software²⁹. This software allowed the expert to visualize the complete EDA signal and indicate, directly in the signal, in which sample the artifact started and ended. The expert then performed an automatic interpolation above the signal, correcting the parts of it that were artifacts according to the criteria. Ledalab recorded the samples that had been corrected, thereby collecting the artifact samples. These data were subsequently used as labels to perform a binary classification problem divided into “artifact” or “non-artifact” samples.

The correction procedure of the signals followed by two different experts was the following. Every expert corrected 22 signals, having 17 of them for train and 5 for test, randomly selected. This division ended up with a total of 34 signals for training and 10 signals for the test set. From the labels of each corrected signal, a descriptive artifact analysis table was done. Afterwards, signals with less than 1% of artifacts were removed from the training set, ending up with a set of 30 signals for train and 10 for test evaluation.

Artifact recognition models

There is no currently accepted protocol for the evaluation of EDA artifact recognition algorithms. The characterization of the algorithm in terms of accuracy or TPR appears to be the only way to evaluate and compare different recognition models. This work proposes four different classification algorithms that use ML and DL models. The first of the used algorithms replicates the methodology applied by Taylor et al.²⁸. In this case, the labels for each second were obtained according to the percentage of the signal labeled as an artifact in the temporal segment. If there were more than 50% artifacts in one second of the signal, the sample was labeled as an artifact and as a non-artifact otherwise. Moreover, this work proposes three novel methodologies, all of which share the same processed target. However, in this case, the temporal segment chosen was 0.5s seconds to increase the precision of the detection. Therefore, each signal was divided into 0.5s segments, and the percentage of artifacts in each segment was obtained. If this percentage was higher than 50%, the target segment was labeled an artifact segment; if not, it was considered a non-artifact segment.

Once all the models were trained, a test-evaluation of the models was conducted, which collected the mean values of different metrics such as accuracy, kappa, TPR, and true negative ratio (TNR). The model with the highest kappa and TPR was considered the best classifier. Once the best model was selected, we administered post-processing of the labeling provided by the automatic recognition model. This process involved re-labeling the signal segment that existed between two artifacts as an artifact if they were less than a certain time threshold, with the aim to merge two artifacts that were very close. The time threshold used was fixed to 2s. Subsequently, the ML/DL performance metrics were computed, and an additional metric was implemented: the percentage of artifacts detected with a certain overlap. This metric was used because artifacts are not particular and punctual points; conversely, artifacts are a set of points with a time duration. Therefore, a metric that contemplates the percentage of the detected artifact within a certain threshold is more appropriate for this problem than metrics such as accuracy or kappa that study the coincidence sample by the sample between the target and the prediction. The percentage of artifacts detected with a certain overlap measures the percentage of overlap between the artifacts predicted and the target ones. If this percentage was above a certain percentage threshold, it was counted as a correct detection.

Feature extraction and classical ML

The first proposed method extracts different features directly from the raw EDA, divided into subsamples of 1s length, following the methodology of Taylor et al.²⁸. Every subsample is therefore a temporal segment from the participants' signals. The total percentage of artifacts in the train set is 15.05%.

Several features were extracted from each raw signal. The first type of features obtained statistics such as the minimum, maximum, or mean of the temporal segment. The first type of statistics was also obtained for the first and the second derivative. Other types of features, such as the Shannon entropy or the energy of the segment, were also collected. In addition, the Haar discrete wavelet transform was used to obtain a second set of features. The Haar window was used due to its ability to detect edges and sharp changes from a signal⁴⁴. This type of window computes the degree of relation between subsequent points in the signal. The obtained coefficients were for the scales 64Hz, 32Hz and 16Hz. Finally, the last type of feature calculated was the maximum frequency of the segment signal obtained through its fast Fourier transform (FFT). The final set of studied features is reflected in Table 2. A total of 51 features were obtained, and the initial time of the computed segment was also added as a feature.

The ML models studied, to achieve the artifact recognition, were a Support Vector Machine (SVM), Random Forest Classifier (RFC) and Logistic Regression (LGC). The model that achieved a higher kappa score in test evaluation is selected as the best. A pre-processing was done on the feature dataset, removing all the features which value was constant (4 variables) and the features that had a high correlation (spearman coefficient > 0.95) with other from the dataset (23 variables). A final dataset with 24 features was obtained. The input data of the model was also processed with two normalization processes that are first standard scaler and afterwards, min-max normalization. Finally, a parameter tuning was done with a group cross validation of

Table 2. Computed features for the input signal and for its first and second derivative.

Source	Feature
Raw signal and filtered signal	Mean, median, min, max, std., amplitude, percentage of points above median. Energy, Shannon entropy, squeezeiness, kurtosis, number of pics.
1st derivative and 2nd derivative for raw and filtered signal	Absolute max, min, absolute mean, median, std.
Wavelet coefficients of raw signal (Haar window)	Max, mean, standard deviation, median, number of non zeros
Raw signal fourier transform	Maximum frequency
Raw signal	Initial time of the segment

10 folds and 4 repetitions. This type of cross validation method was selected to ensure that the samples that belong to the same subject are not simultaneously present in both the train and validation split.

Raw signal classification and RNN

The following model studies the detection of an artifact in the last 0.5s of a 5s signal segment. A schematic approach of the timeline for this data processing is shown in Fig. 2. The percentage of artifact in the training set was 12.60%. No filter or pre-processing was applied to the raw signal but a scaling between 0 and 1 for each 5s segment.

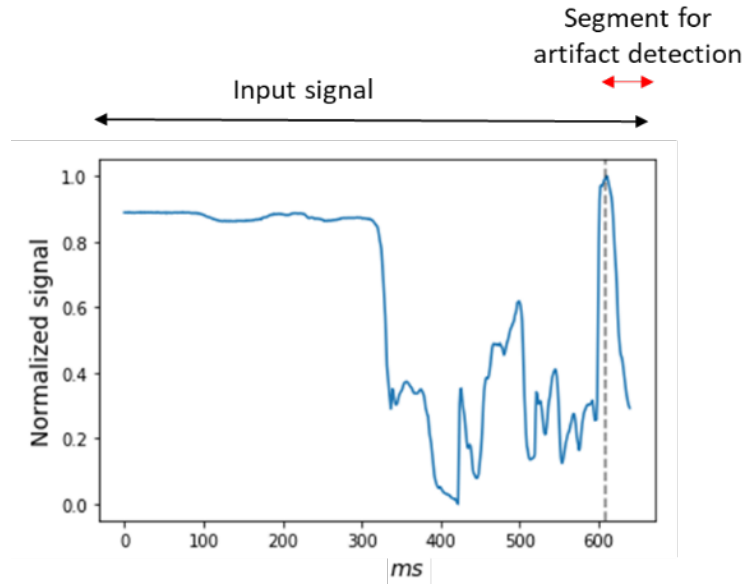


Figure 2. Schematic input and output processing for the *raw signal classification and RNN* model.

The main purpose of this model was to learn from the temporal evolution of the signal. For this reason, a set of long-short time memory (LSTM) layers were used. Moreover, due to the high amount of data in a segment (640 points), a set of convolutional network (CN) layers followed by max-pooling operation were used before the LSTM layers, see Fig. 3. These max-pooling layers reduce the amount of data whereas the CN layers decompose the segments through different filters to generate a number of features from the segment signal. The model architecture was the following. It starts with two LSTM layers of 16 neurons that returned the hidden state output for each input time step. After that, the network had four convolutional levels. Every convolutional level had three convolutional layers with a batch-normalization operation after each convolution. Each level ended up with a dropout value of 0.05 and a max-pooling operation with a pool size of 2. The number of filters was 32, 64, 128 and 256, per each level, with a kernel size of 5. Finally, the model had two fully connected layers of 256 and 16 layers, ending up with a final fully connected layer of one single perceptron and sigmoid activation function. The model was trained with *rmsprop* optimizer with a learning rate of 5×10^{-5} and a batch size of 16. Due to the unbalancement, the cost function to train the model was the DSC. The model finished its training with an early stopping threshold of 30 epochs.

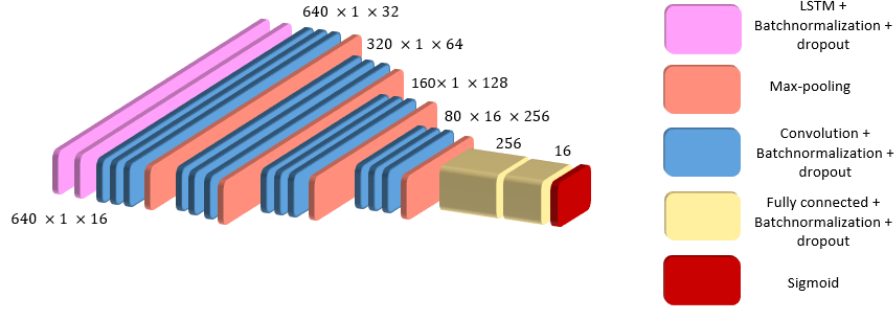


Figure 3. Scheme of the model architecture used for the *raw signal classification and RNN*.

Spectrogram classification and segmentation

The final suggested approach is the study of the recognition of artifacts through spectrogram classification and segmentation. To obtain a spectrogram representation of this signal, we used FFT. The FFT of this signal achieved a frequency resolution of 64 samples as its frequency was 128Hz. In order to obtain the squared matrix, the time segments of each signal were divided into segments of 32s. A matrix representation with the dimensions 64×64 was obtained. In these representations, the vertical axis represents the frequencies in Hz, while the horizontal axis shows the temporal information in seconds. The spectrograms were obtained with a 50% overlap. Two consecutive models were used for the temporal segmentation of artifacts. The first model is a spectrogram classification model that classifies whether a spectrogram has an artifact on it or not. The second model is a spectrogram segmentation model that creates a temporal segmentation inside the spectrogram to find the artifacts. This second model only studied the spectrograms classified as containing an artifact by the first spectrogram classification model.

The labeling procedure explained at the beginning of this section culminated, in the spectrogram classification and segmentation case, in a target image where the artifacts were represented as vertical lines. These target images correspond to the target labels for the spectrogram segmentation model. The percentage of artifacts in each spectrogram was then computed. The spectrogram would be classified as having an artifact if this percentage was higher than 0%; otherwise, the spectrogram would be classified as clean. This binarization was used as the labels for the spectrogram classification model. Finally, the total percentage of artifacts in the training set of the classification spectrogram model was 45.38%. The training set of the spectrogram segmentation model was 39.80%. Figure 4 shows a scheme from a raw signal to the automatic recognition of the artifacts for the same signal.

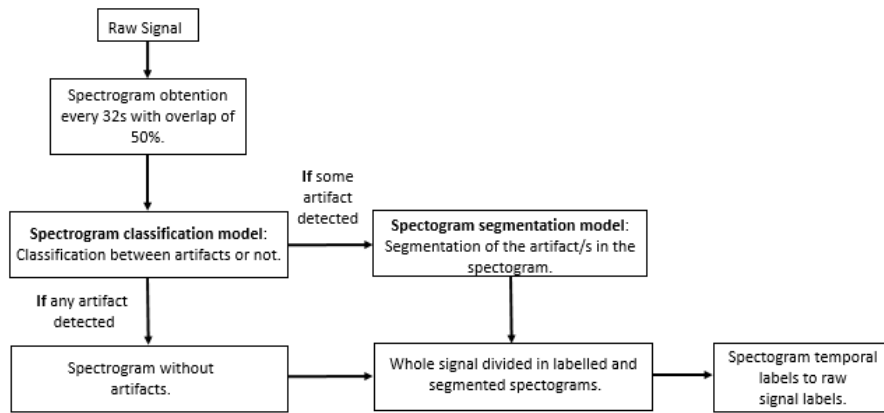


Figure 4. Scheme of the followed methodology for spectrogram detection and classification.

The data introduced in both models was a set of min-max normalized spectrograms with dimension of 64×64 . The target data was a binary label for each spectrogram for the classification model, while for the spectrogram segmentation model they were binary vertically segmented images. An example is shown in Fig. 5.

Both classification models used convolution layers as the main operation of the model. The first used model, the classification model, was a set of convolutional layers in order to perform an image artifact recognition problem. The architecture of the

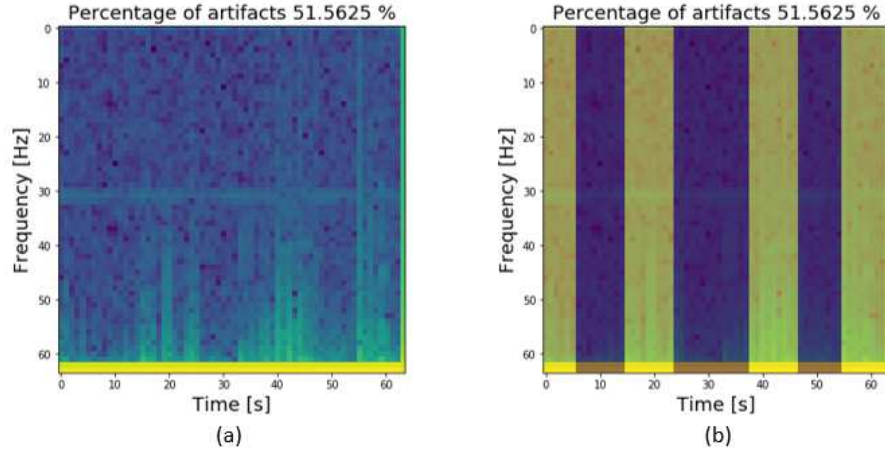


Figure 5. Spectrogram segmented according to the target labels. The image (a) corresponds to the spectrogram of a raw signal segment that contents artifacts. Image (b) corresponds to its target, which has the artifacts indicated by yellow vertical segments.

classification model had four convolutional levels, from 16 to 128 filters for each level, see Fig. 6. The fully connected layers in the last two levels of the model had a dropout rate of 0.5 value. The cost function of the model was binary cross-entropy. On the other hand, the segmentation model follows a U-Net architecture, see Fig. 6. This model had a maximum number of filters of 256. The kernel size for all convolutional layers was set to 5 and the dropout rate of the convolutional levels to 0.05. The cost function of the model was the mean value between DSC and binary cross-entropy. The learning rate for both models was 1×10^{-4} using Adam optimizer with a batch size for both models of 4. Also, both models finished their training by an early stopping threshold of 30 epochs.

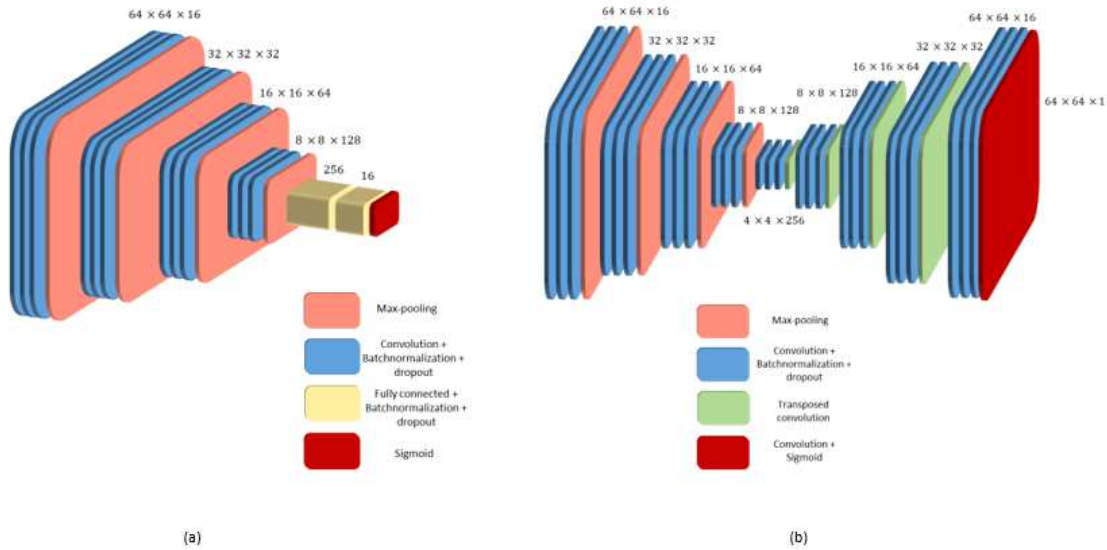


Figure 6. Scheme of the architecture of the two models used for the *Spectrogram classification and segmentation*. Image (a) shows the artifact classification image. Image (b) shows the model that achieves the segmentation of the artifacts in the spectrogram.

To increase the size of the training dataset and also to achieve a higher generalization and robustness of the models, both were trained using a data augmentation technique⁴⁵. Two different types of transformation were implemented to this extent. The first one was the definition of random vertical or horizontal lines equal to zero that hide, randomly, certain pixels in the

spectrogram. The minimum and maximum threshold number of hidden pixels was 256 and 1024 respectively. The second transformation was the translation of the spectrogram image through a random vertical and horizontal pixel distance. The minimum and maximum threshold distance defined was 4 and 16 pixels. All the images in the dataset suffered both types of transformation, increasing the size of the dataset three times.

Artifact correction

After the artifact recognition among the signal, a regression method was developed to correct the detected artifacts through the samples of the signals labeled as artifacts. This automatic correction process combined two interpolation methods. The first was a linear interpolation between the beginning and the end of the artifact. The second method implemented was the obtention of a polynomial with a degree of 8. The first and last samples of the artifact were taken to obtain this polynomial, and six additional internal and evenly spaced samples were considered. Both methods brought forth a set of points for each sample labeled as an artifact. Finally, both techniques were averaged by each point of the artifact to combine the corrections performed using a linear and a non-linear approach. This methodology partially reproduced that used in the Ledalab software, whereby manual correction of the artifacts can be done either by linear interpolation or spline interpolation. The method used in this work combined both approaches. It had a linear fit to capture the tendency of the artifact segment and a polynomial estimation to adjust the interpolation to the non-linearity of the EDA signal. Subsequently, a simple moving average (SMA) of eight samples was implemented. The SMA was applied from 0.125s seconds before the beginning of the corrected artifact until 0.125s seconds after the end of the artifact to smooth the joint between the corrected artifact segment and the original EDA signal.

Different metrics were computed to evaluate the quality of the automatic and manual correction and to relate the raw signal, the automatically corrected signal, and the signal corrected by the expert in terms of the phasic components. This is because phasic components assess the sympathetic activity; the central meaning of EDA can be found in their peaks. The computation of different features from this component is essential for the correct evaluation of various mental disorders or illnesses⁵. To probe the robustness of the proposed methodology, we obtained the phasic components using two different approaches: CDA, through the use of the Ledapy library, and the cvxEDA algorithm. The automatic correction performed was compared through these two algorithms that can obtain the phasic component of an EDA signal. The computed metrics were the root mean square error (RMSE), mean absolute error (MAE), cross-correlation, and the difference in the area under the curve (DAUC). This similarity was also evaluated through a one-way ANOVA test. Every phasic component of the three different signals—raw signal, manual cleaned, and automatic cleaned—were segmented by temporal intervals of 5min. The mean value of every 5min segment was computed. Every phasic component would show a different distribution that was compared using a one-way ANOVA test. The *p*-value of the statistical test, compared by pairs, indicated any statistical difference between the automatically corrected and manually cleaned signals or between the automatically corrected and the raw signals. This set of metrics would show if the phasic component of the automatic correction were closer to the manual correction than to the raw signal.

Results

Signal and artifact description

The descriptive analysis of the artifacts identified was divided into train (30 signals), test (10 signals), and all the signals together (44 signals). A set of metrics was obtained that were subsequently compared using a one-way ANOVA test between test and train signals. The statistics were first computed per participant and then averaged among all the signals. Each metric shows a statistical difference (*p*-value < 0.05) between train and test. Table 3 demonstrates a significant imbalance in the percentage of artifacts, with a mean artifact percentage of $10.63 \pm 11.59\%$.

Table 3. Descriptive features for the artifacts extracted from all the signals. Results are shown as mean and standard deviation.

	Artifact duration (s)	Number of artifacts	Artifact percentage (%)	First artifact (s)	Minimum artifact duration (s)	Time between artifacts (s)
Train	5.24 ± 3.72	124.40 ± 95.21	10.92 ± 11.97	88.71 ± 182.02	0.93 ± 0.52	$103.50_{pm}140.30$
Test	5.14 ± 3.01	182.30 ± 86.71	12.81 ± 10.57	45.65 ± 22.36	0.73 ± 0.55	48.47 ± 44.01
<i>p</i> -value	0.94	0.11	0.67	0.47	0.31	0.24
All	5.32 ± 3.46	129.49 ± 99.16	10.63 ± 11.59	76.72 ± 153.75	1.00 ± 0.68	141.24 ± 261.39

Artifact recognition

The metrics obtained by each of the three different methodologies are shown in Table. 4. The results were evaluated on the test set, computing every metric in each signal and afterwards averaging them with for all the signals. According with the results of Table. 4, The raw signal classification and RNN and the spectrogram classification and segmentation models are the ones with highest TPR (0.63). Finally, the raw signal classification and RNN model is considered as the best recognition model due to its higher kappa value (0.48).

Table 4. Evaluation on the test set of the different proposed approaches. Results are shown as mean and standard deviation.

Model	Artifact ratio	Accuracy	Kappa	DSC	AUC	TPR	TNR
<i>Feature extraction and classical ML</i>	0.16 ± 0.18	0.90 ± 0.10	0.34 ± 0.10	0.40 ± 0.16	0.64 ± 0.06	0.28 ± 0.12	0.99 ± 0.01
<i>Raw signal classification and RNN</i>	0.13 ± 0.11	0.88 ± 0.09	0.49 ± 0.08	0.57 ± 0.07	0.76 ± 0.06	0.65 ± 0.16	0.89 ± 0.17
<i>Spectrogram classification and segmentation</i>	0.12 ± 0.12	0.87 ± 0.10	0.42 ± 0.09	0.50 ± 0.11	0.75 ± 0.06	0.63 ± 0.17	0.87 ± 0.15

After the obtention of the raw signal classification and RNN labels, these labels were post-processed to accurate the artifact recognition process using the fixed temporal threshold in 2s. Table. 5 shows an increment in the results for the raw signal classification and RNN model in metrics as Kappa, DSC, AUC or TPR. Indeed, the increase of TPR and kappa achieves a final value of 0.72 and 0.50 respectively.

Table 5. Evaluation of the raw signal classification and RNN model predictions after artifact recognition post-processing. Results are shown as mean and standard deviation

Model	Artifact ratio	Accuracy	Kappa	DSC	AUC	TPR	TNR
<i>Raw signal classification and RNN post-processing</i>	0.13 ± 0.12	0.87 ± 0.10	0.50 ± 0.10	0.58 ± 0.10	0.79 ± 0.06	0.72 ± 0.13	0.86 ± 0.18

To evaluate the accuracy of the recognized artifacts and its percentage of overlap against the ground-truth, the metric percentage of artifacts detected with certain overlap was computed. Fig. 7 shows a decrease in the percentage of detected artifacts according with the overlap ratio threshold. The raw signal classification and RNN model achieves a percentage of artifacts detected of $78.39 \pm 12.14\%$ for an artifact overlap threshold of 50%. Indeed, with an overlap threshold of 20% the percentage of detected artifacts increase to $85.15 \pm 9.93\%$. Moreover, the number of detected artifacts was also computed. Before the post-processing, the number of detected artifacts was 501.20 ± 201.30 . After the post-processing, the number of detected artifacts decreases to 231.10 ± 79.50 where the real number of artifacts is 182.30 ± 86.70 .

Artifact correction

After the interpolation of each artifact segment, an automatically cleaned EDA signal was obtained. In the supplementary materials, there is a set of test signals automatically corrected by the exposed algorithm. Fig. 8 shows the final interpolation result for a raw signal segment after the automatic correction process.

We obtained the phasic components for the test set signals, the raw signal, the manual, and the automatic correction to evaluate the similarity between the automatic and manual corrections. Table. 6 compares the phasic components of the automatically corrected and the manually corrected signals as well as the phasic component of the automatically corrected signal and the original raw data. This table shows a lower value in RMSE, MAE, and DAUC metrics between the phasic components from the manual and clean signals for both decomposition algorithms (CDA and cvxEDA). The one-way ANOVA test did not indicate a statistical difference (p -value > 0.05) between both signals for both algorithms. Meanwhile, a statistical difference (p -value < 0.05) is shown to compare the phasic components of the automatic and manual signals with the phasic components of the raw signal.

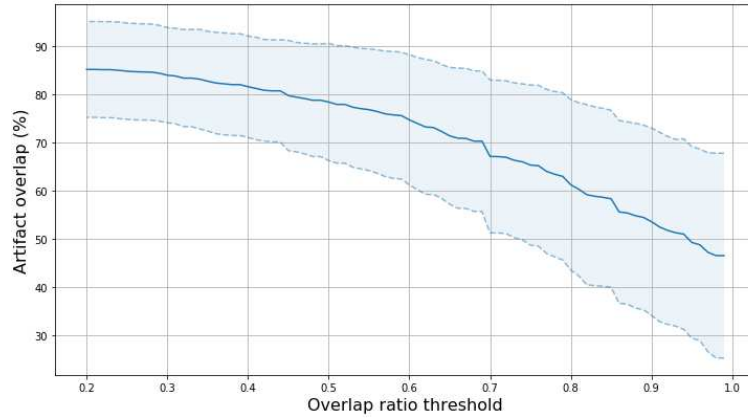


Figure 7. Scheme of the pre-processing of the signal to obtain the inputs and the targets for the Feature extraction and RNN model.

Table 6. Averaged metrics by signal. Statistical values of the phasic component for the comparison pairs of the automatic correction, the manually cleaned and the original raw signal. Results are shown as mean and standard deviation.

Algorithm	Phasic component	RMSE	MAE	Cross correlation	DAUC	<i>p</i> -value
CDA	Automatic and manual	0.146 ± 0.096	0.054 ± 0.033	0.772 ± 0.229	0.194 ± 0.184	0.427
	Automatic and raw signal	0.171 ± 0.108	0.068 ± 0.071	0.743 ± 0.216	0.246 ± 0.247	< 0.001
	Manual and raw signal	0.153 ± 0.102	0.064 ± 0.055	0.795 ± 0.186	0.377 ± 0.616	0.012
cvxEDA	Automatic and manual	0.339 ± 0.256	0.078 ± 0.039	0.633 ± 0.235	0.236 ± 0.168	0.246
	Automatic and raw signal	0.929 ± 0.786	0.272 ± 0.437	0.609 ± 0.207	0.478 ± 0.230	< 0.001
	Manual and raw signal	0.835 ± 0.809	0.255 ± 0.423	0.682 ± 0.278	0.317 ± 0.311	< 0.001

Fig. 9 shows two different boxplots of the mean values by 5 minutes segments of the phasic component, for all the different signals by each phasic obtention method, CDA and cvxEDA. Both plots shows a higher similarity in the distribution of the automatic phasic and the manual clean phasic than any of both with the raw signal phasic component.

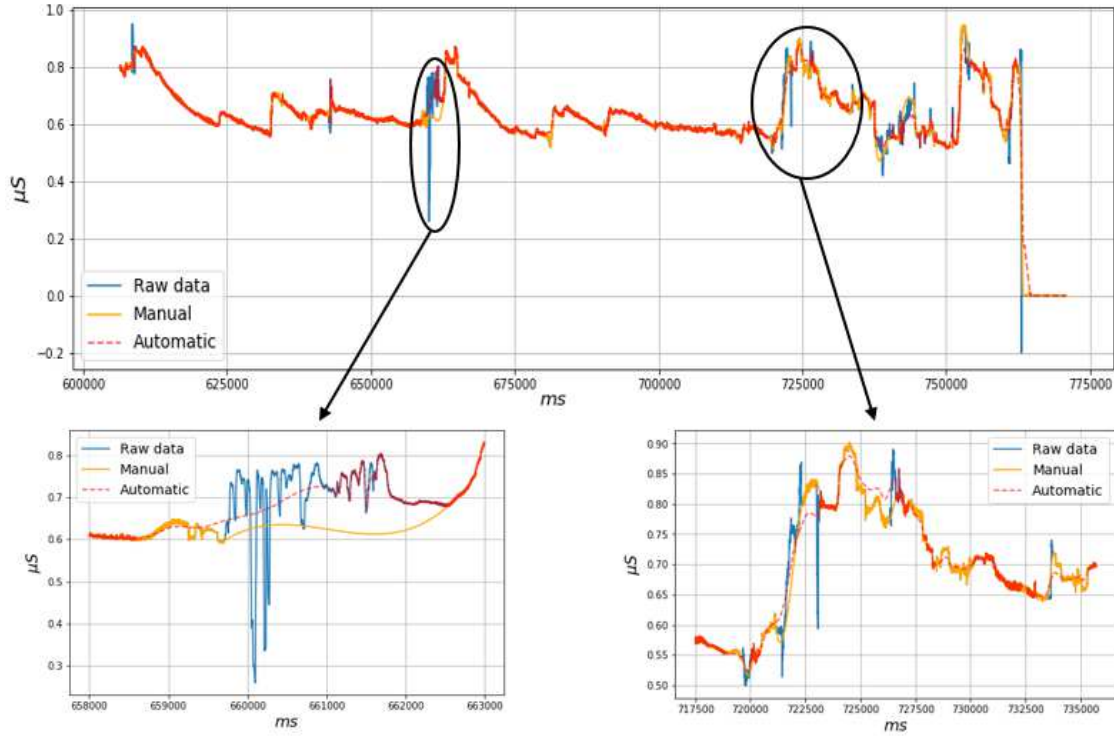


Figure 8. Automatic correction of a certain segment of an EDA signal. The blue signal is the raw signal of the segment. The orange one is the manual correction done by one of the experts, and the red line is the automatic correction done by the artifact recognition and correction algorithm.

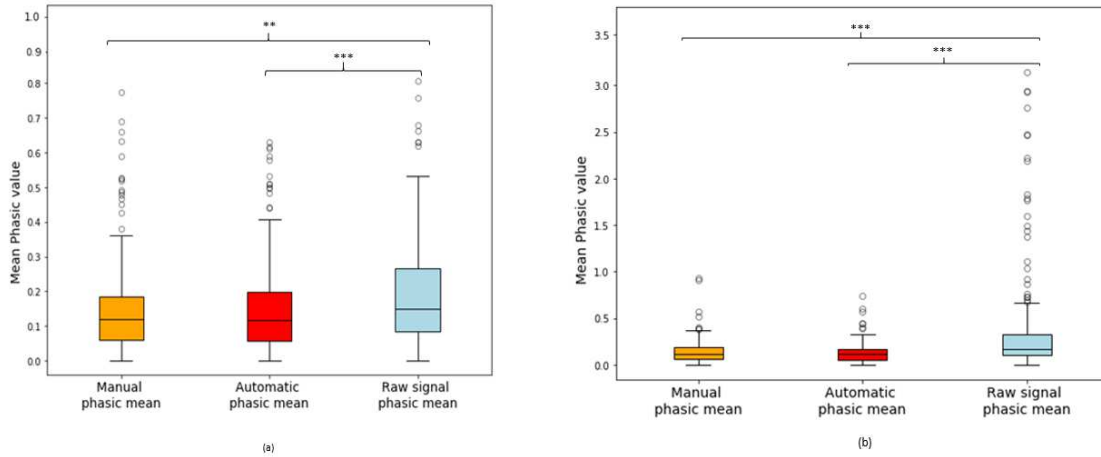


Figure 9. Box plot of the different phasic values distributions. Image (a) show the comparison using CDA decomposition method. Image (b) shows the results of the cvxEDA algorithm.

Discussion

The objective of this work was to automatically recognize and correct an EDA signal collected in an uncontrolled VR scenario and thus to provide an artifact-corrected signal. The work applied two new approaches using DL algorithms: RNN applied to a sequential feature extraction, RNN applied to the raw signal, and a CNN applied to the spectrogram. The work of Taylor et al.²⁸ was used as a benchmark method.

The present study addresses three main, novel points. First, we analyzed EDA artifact recognition in an uncontrolled

immersive VR scenario where the different tasks performed by the subjects required hand or body movements. This type of experiment implies the appearance of more artifacts than in experiments where the participant has to be at rest and can be considered a noisy environment. Second, the model provided a final corrected signal during a continuous assessment, in contrast to previous artifact recognition works, which divided the signal into 5-second time windows, recognized if each segment had or did not have an artifact, and did not offer a corrected signal. Third, the artifact-free signals obtained by the best method were statistically compared against a set of signals manually corrected by two experts in terms of phasic components using two methods: CDA and cvxEDA. While previous works compared the ANS assessment achieved with their proposals to that achieved with a raw signal, the benchmark used in this work was a signal cleaned by experts. Thus, a correction without significant differences to one that would be made by the experts has been achieved. As such, this study constitutes a breakthrough in the validation of recognition and correction algorithms in EDA signals.

The feature extraction and classical ML methodology, also used in other works^{28,36,37}, achieved the lowest kappa (0.34) and TPR (0.28). Meanwhile, the results obtained were worse than those obtained in previous studies^{28,36,37}. There are two potential reasons for this discrepancy. First, the type of cleaning undertaken by the experts in the present work is different from that used in other experimentations, complicating comparisons between the results of this work and other studies' findings. Second, the imbalance of the current dataset (10.63%) is higher than the one used in prior experiments (48.96% in the work of Gashi et. al.³⁶), which could bias the performance and the results of the ML models.

The last model is the spectrogram classification and segmentation approach, which achieved a kappa of 0.42, a TPR of 0.63, and an accuracy of 0.87. These results are lower than those from the raw signal classification and RNN model without post-processing, which is justified due to the type of model used based on convolution operation. CNNs are not yet optimized for the study of spectrogram classification or segmentation. This is due to the non-locality information that a spectrogram provides, while CNNs base their knowledge on the local information of the image. More specialized models such as spectral-CNN could be implemented in future research to study the artifact detection problem in EDA. Finally, Table. 3 justifies the time threshold used in this model. The spectrogram classification and segmentation model recognized artifacts in time intervals of 0.5 seconds according to the temporal resolution set in the spectrogram. This temporal resolution is also below the minimum mean time artifact duration reported in Table. 3.

The last model is the spectrogram classification and segmentation which achieves a kappa of 0.42, TPR of 0.63 and an accuracy of 0.87. This results are lower than the ones from the raw signal classification and RNN results without post-processing. This is justified due to the type of used model based in convolution operation. CNN are not optimized yet for the study of spectrogram classification or segmentation. This is due to the non-locality information that a spectrogram provides, while CNNs base their knowledge in the local information of the image. More specialized models such as Spectral-CNN could be implemented in future researches to study the artifact detection problem in EDA. Finally, Table. 3 also justifies the time threshold used in this model. The spectrogram classification and segmentation model recognize artifacts in time intervals of 0.5s according with the temporal resolution setted in the spectrogram. This temporal resolution is also below to the minimum mean time artifact duration reported in Table. 3.

In order to approximate the recognition of the model's artifacts, a post-processing technique was performed over the labels of the best artifact recognition model, the raw signal classification and RNN based on kappa. This post-processing helped enhance the detection of artifacts that can be seen through metrics as kappa and TPR that improved to final values of 0.50 and 0.72, respectively, decreasing the accuracy value to 0.87. The percentage of artifacts detected with certain overlap showed a high value of artifact recognition, achieving a value of 78.39% and an overlap threshold ratio of 0.5. Moreover, due to the post-processing, more realistic results were achieved through metrics as the number of detected artifacts (231.10) was closer to the real one (182.30); see Table. 3. This technique is a compelling means of achieving a much more accurate recognition of artifacts in EDA. Table. 3 shows that the mean time between two artifacts was 141.24s, which is much higher than the time threshold used for the post-processing, set to 2s. This time threshold was selected to be conservative with the post-processing of the model's label.

Finally, the complete model was evaluated, including recognition and signal correction. The one-way ANOVA test (Table 6) reflected no statistical difference (p -value > 0.05) between the phasic component from the manually cleaned signal and the phasic component from the automatically corrected signal for CDA or cvxEDA algorithms. Furthermore, the type of correction performed was robust to the type of decomposition analysis applied, showing similar results for CDA and cvxEDA. Meanwhile, there was a statistically significant difference (p -value < 0.05) between the phasic component from the raw signal and the phasic component from the automatically and manually corrected signal. This indicates that the automatically cleaned signal decreased the level of noise from the artifacts of the raw signal (see Fig. 9). Other metrics, such as RMSE, MAE, and DAUC, also showed that the phasic component from the automatic correction was closer to the phasic component from the manually corrected signal than to the phasic component from the raw signal. Therefore, the results suggest that the automatic correction simulates the manual correction accurately, independently of the decomposition algorithm used, and can be used in combination with CDA or cvxEDA. Finally, these results support the main objective of this work: the obtention of an automatic recognition

and correction algorithm for EDA artifacts in an environment that requires subjects' hand and body movements.

The correction algorithm used in this work has been designed similarly to the type of correction allowed by Ledalab software. While the obtained results fulfilled the initial objective, the type of automatic correction could be complemented or replaced by other correction methodologies, such as those suggested by Chen et al.³² or Shukla et. al.¹³, who used wavelet transformation, low-pass filters³¹, or even cvxEDA⁶ algorithm. All these methodologies are complementary to the work shown and could enrich the correction made by the proposed algorithm. Indeed, future research could study the automatic correction of artifacts through the use of DL regressor models.

Lastly, the present study provides a foundation for future studies to address several points. This study presents an imbalance between the number of men and women. A more balanced sample would enrich the dataset and the generality of the obtained findings. The inclusion of more experts in the labeling process could improve and generalize the artifact detection methodology, thereby improving the models' results. In addition, future research needs to evaluate the model in other types of environment and tasks since the movements performed can modify the form of the artifacts. The validation of the presented methodology for EDA signals collected during these tasks would strengthen it and demonstrate its applicability to contexts in which the use of this signal is common. Moreover, the methodology proposed in this work should be validated in real-world experimentations. The procedure presented has not been tested for the data recording of different EDA devices or those with lower frequencies than that of this work. All the used methodologies could be studied for different sampling frequencies to review their performance and generalization to other case studies. Continuous advances in the ML and DL field would facilitate the exploration of more efficient and precise correction models for EDA signals. Future experiments could research the development of fine-tuned architectures of the different models to improve their classification metrics. Alternative techniques, such as generative-adversarial networks (GANs) or reinforcement learning (RL), are other interesting and promising alternatives to the models shown in this work.

Conclusion

These results show that the correction of EDA signals in scenarios that require hand and body movements can be achieved automatically. The obtained automatic artifact-free signals do not show statistical differences in terms of phasic components to the signal corrected manually by an expert. Meanwhile, the methodology proposed in this work offers a corrected signal that can be used in combination with either CDA or cvxEDA algorithms. This work presents the first automatic recognition and correction of EDA artifacts through DL techniques. This automatization will support the use of EDA signals due to its ability to save time and costs for researchers in future experiments related to health, neuroscience, and psychology in uncontrolled environments such as immersive VR or daily-life activities. Furthermore, the automatic correction algorithm does not require human intervention, allowing for the possibility to develop automated wearable systems based on artificial intelligence for health monitoring. Finally, it enables the opportunity to correct EDA signals in real time, allowing for the correct use of this signal for adaptive therapies or live diagnosis. The EDA correction algorithm could be used through the following link <https://i3b.webs.upv.es/webs/gsrartifactcorrection/>.

Acknowledgements

The research leading to these results has received partial funding from the European Commission (Project HELIOS H2020-825585) and from the Universitat Politècnica de València (PAID-10-20). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Boucsein, W. *Electrodermal activity* (Springer Science+Business Media, LLC, New York, 2012).
2. Posada-Quintero, H. & Chon, K. Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors* **20**, 479, DOI: [10.3390/s20020479](https://doi.org/10.3390/s20020479) (2020).
3. Greco, A., Valenza, G. & Scilingo, E. *Advances in Electrodermal Activity Processing with Applications for Mental Health* (2016).
4. Ellaway, P., Kuppuswamy, A., Nicotra, A. & Mathias, C. Sweat production and the sympathetic skin response: improving the clinical assessment of autonomic function. *Auton. neuroscience : basic amp; clinical* **155**, 109—114, DOI: [10.1016/j.autneu.2010.01.008](https://doi.org/10.1016/j.autneu.2010.01.008) (2010).
5. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80 – 91, DOI: <https://doi.org/10.1016/j.jneumeth.2010.04.028> (2010).

6. Greco, A., Valenza, G., Lanata, A., Scilingo, E. P. & Citi, L. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Trans. Biomed. Eng.* **63**, 797–804 (2016).
7. Dawson, M., Schell, A. & Fillion, D. The electrodermal system. DOI: [10.1017/CBO9780511546396.007](https://doi.org/10.1017/CBO9780511546396.007) (2000).
8. Anusha, A. S., Jose, J., Preejith, S. P., Jayaraj, J. & Mohanasankar, S. Physiological signal based work stress detection using unobtrusive sensors. *Biomed. Phys. & Eng. Express* **4**, 065001, DOI: [10.1088/2057-1976/aadb4](https://doi.org/10.1088/2057-1976/aadb4) (2018).
9. Zangróniz, R., Martínez Rodrigo, A., Pastor García, J. M., López Bonal, M. & Fernández-Caballero, A. Electrodermal activity sensor for classification of calm/distress condition. *Sensors* **17**, 2324, DOI: [10.3390/s17102324](https://doi.org/10.3390/s17102324) (2017).
10. Liu, Y. & Du, S. Psychological stress level detection based on electrodermal activity. *Behav. Brain Res.* **341**, 50–53 (2018).
11. Greco, A., Valenza, G., Lanatà, A., Rota, G. & Scilingo, E. Electrodermal activity in bipolar patients during affective elicitation. *IEEE journal biomedical health informatics* **18**, 1865–1873, DOI: [10.1109/JBHI.2014.2300940](https://doi.org/10.1109/JBHI.2014.2300940) (2014).
12. Perugia, G. *et al.* In *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, 1248–1254, DOI: [10.1109/ROMAN.2017.8172464](https://doi.org/10.1109/ROMAN.2017.8172464) (Institute of Electrical and Electronics Engineers, United States, 2017). 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2017), RO-MAN 2017 ; Conference date: 28-08-2017 Through 01-09-2017.
13. Shukla, J., Barreda-Ángeles, M., Oliver, J. & Puig, D. Efficient wavelet-based artifact removal for electrodermal activity in real-world applications. *Biomed. Signal Process. Control.* **42**, 45 – 52, DOI: <https://doi.org/10.1016/j.bspc.2018.01.009> (2018).
14. Malathi, D., Jayaseeli, J. D., Madhuri, S. & Senthilkumar, K. Electrodermal activity based wearable device for drowsy drivers. *J. Physics: Conf. Ser.* **1000**, 012048, DOI: [10.1088/1742-6596/1000/1/012048](https://doi.org/10.1088/1742-6596/1000/1/012048) (2018).
15. Leite, I., Henriques, R., Martinho, C. & Paiva, A. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, HRI '13*, 41–48 (IEEE Press, 2013).
16. Kim, J. J. & Fesenmaier, D. R. Measuring emotions in real time: Implications for tourism experience design. *J. Travel. Res.* **54**, 419–429, DOI: [10.1177/0047287514550100](https://doi.org/10.1177/0047287514550100) (2015). <https://doi.org/10.1177/0047287514550100>.
17. Chicchi Giglioli, I., Pravettoni, G., Sutil, L., Parra, E. & Alcañiz Raya, M. A novel integrating virtual reality approach for the assessment of the attachment behavioral system. *Front. Psychol.* **8**, DOI: [10.3389/fpsyg.2017.00959](https://doi.org/10.3389/fpsyg.2017.00959) (2017).
18. Raya, M. A., Baños, R., Botella, C. & Rey, B. The emma project: Emotions as a determinant of presence. *PsychNology J.* **1**, 141–150 (2003).
19. Marín-Morales, J. *et al.* Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PLOS ONE* **14**, 1–24, DOI: [10.1371/journal.pone.0223881](https://doi.org/10.1371/journal.pone.0223881) (2019).
20. Tarrant, J., Viczko, J. & Cope, H. Virtual reality for anxiety reduction demonstrated by quantitative eeg: A pilot study. *Front. Psychol.* **9**, 1280, DOI: [10.3389/fpsyg.2018.01280](https://doi.org/10.3389/fpsyg.2018.01280) (2018).
21. Matijević, V. *et al.* Virtual reality in rehabilitation and therapy. *Acta clinica Croat.* **52**, 453–7 (2013).
22. Li, A., Montañó, Z., Chen, V. J. & Gold, J. Virtual reality and pain management: current trends and future directions. *Pain management* **12**, 147–157 (2011).
23. Bekele, E., Bian, D., Peterman, J., Park, S. & Sarkar, N. Design of a virtual reality system for affect analysis in facial expressions (vr-saafe); application to schizophrenia. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **25**, 739–749, DOI: [10.1109/TNSRE.2016.2591556](https://doi.org/10.1109/TNSRE.2016.2591556) (2017).
24. Maskeliunas, R., Šalkevicius, J., Damaševičius, R., Maskeliunas, R. & Laukienė, I. Anxiety level recognition for virtual reality therapy system using physiological signals. *Electronics* **8**, 1039 (2019).
25. Alcañiz Raya, M. *et al.* Application of supervised machine learning for behavioral biomarkers of autism spectrum disorder based on electrodermal activity and virtual reality. *Front. Hum. Neurosci.* **14**, 90, DOI: [10.3389/fnhum.2020.00090](https://doi.org/10.3389/fnhum.2020.00090) (2020).
26. Kritikos, J., Tzannetos, G., Zoitaki, C., Pouloupoulou, S. & Koutsouris, D. Anxiety detection from electrodermal activity sensor with movement interaction during virtual reality simulation. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 571–576, DOI: [10.1109/NER.2019.8717170](https://doi.org/10.1109/NER.2019.8717170) (2019).
27. Salgado, D. *et al.* A qoe assessment method based on eda, heart rate and eeg of a virtual reality assistive technology system. 517–520, DOI: [10.1145/3204949.3208118](https://doi.org/10.1145/3204949.3208118) (2018).
28. Taylor, S. *et al.* Automatic identification of artifacts in electrodermal activity data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1934–1937 (2015).

29. Ledalab. <http://www.ledalab.de/>. Accessed: 2020-01-13.
30. Bach, D. R. A head-to-head comparison of scralyze and ledalab, two model-based methods for skin conductance analysis. *Biol. Psychol.* **103**, 63 – 68, DOI: <https://doi.org/10.1016/j.biopsycho.2014.08.006> (2014).
31. Hernandez, J., Morris, R. R. & Picard, R. W. Call center stress recognition with person-specific models. In D’Mello, S., Graesser, A., Schuller, B. & Martin, J.-C. (eds.) *Affective Computing and Intelligent Interaction*, 125–134 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011).
32. Chen, W. *et al.* Wavelet-based motion artifact removal for electrodermal activity. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6223–6226, DOI: [10.1109/EMBC.2015.7319814](https://doi.org/10.1109/EMBC.2015.7319814) (2015).
33. Ganapathy, N., Veeranki, Y. R. & Swaminathan, R. Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features. *Expert. Syst. with Appl.* **159**, 113571, DOI: <https://doi.org/10.1016/j.eswa.2020.113571> (2020).
34. Can, Y., Chalabianloo, N., Ekiz, D. & Ersoy, C. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors* **19**, DOI: [10.3390/s19081849](https://doi.org/10.3390/s19081849) (2019).
35. Kleckner, I. R. *et al.* Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. *IEEE transactions on bio-medical engineering* **65**, 1460—1467, DOI: [10.1109/tbme.2017.2758643](https://doi.org/10.1109/tbme.2017.2758643) (2018).
36. Gashi, S. *et al.* Detection of artifacts in ambulatory electrodermal activity data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4**, DOI: [10.1145/3397316](https://doi.org/10.1145/3397316) (2020).
37. Zhang, Y., Haghdan, M. & Xu, K. S. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers, ISWC ’17*, 54–57, DOI: [10.1145/3123021.3123054](https://doi.org/10.1145/3123021.3123054) (Association for Computing Machinery, New York, NY, USA, 2017).
38. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597* (2015). [1505.04597](https://arxiv.org/abs/1505.04597).
39. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90) (2016).
40. Kyathanahally, S. P., Döring, A. & Kreis, R. Deep learning approaches for detection and removal of ghosting artifacts in mr spectroscopy. *Magn. Reson. Medicine* **80**, 851–863, DOI: <https://doi.org/10.1002/mrm.27096> (2018).
41. Bento, N., Belo, D. & Gamboa, H. Ecg biometrics using spectrograms and deep neural networks. DOI: [10.18178/ijmlc.2020.10.2.929](https://doi.org/10.18178/ijmlc.2020.10.2.929) (2020).
42. Antczak, K. Deep recurrent neural networks for ECG signal denoising. *CoRR abs/1807.11551* (2018). [1807.11551](https://arxiv.org/abs/1807.11551).
43. Shimmer,gsr + expansion board user guide.realtimetechnologies. <http://www.shimmersensing.com>. Accessed: 2020-01-30.
44. Xu, Y., Weaver, J., Jr, H. & Lu, J. Wavelet transform domain filters: A spatially selective noise filtration technique. *Image Process. IEEE Transactions on* **3**, 747 – 758, DOI: [10.1109/83.336245](https://doi.org/10.1109/83.336245) (1994).
45. Ghosh, S., Das, N., Das, I. & Maulik, U. Understanding deep learning techniques for image segmentation. *ACM Comput. Surv.* **52**, DOI: [10.1145/3329784](https://doi.org/10.1145/3329784) (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [EDAalgorithmsupplementarymaterials.pdf](#)
- [Signalexample.zip](#)