

Review

Machine Learning for Stress Detection from Electrodermal Activity: A Scoping Review

Roberto Sánchez-Reolid^{1,2}, **María T. López**^{1,2} and **Antonio Fernández-Caballero**^{1,2,3,*}¹ Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, 02071 Albacete, Spain² Instituto de Investigación en Informática de Albacete, Universidad de Castilla-La Mancha, 02071 Albacete, Spain³ CIBERSAM (Biomedical Research Networking Center in Mental Health), 28016 Madrid, Spain

* Correspondence: antonio.fdez@uclm.es; Tel.: +34-967-599200 (A.F.-C.)

Abstract: Early detection of stress can prevent us from suffering from a long-term illness such as depression and anxiety. This article presents a scoping review of stress detection based on electrodermal activity (EDA) and machine learning (ML). From an initial set of 395 articles searched in six scientific databases, 58 were finally selected according to various criteria established. The scoping review has made it possible to analyse all the steps to which the EDA signals are subjected: acquisition, preprocessing, processing and feature extraction. Finally, all the ML techniques applied to the features of this signal have been studied for stress detection. It has been found that support vector machines and artificial neural networks stand out within the supervised learning methods given their high performance values. On the contrary, it has been evidenced that unsupervised learning is not very common in the detection of stress through EDA. This scoping review concludes that the use of EDA for the detection of arousal variation (and stress detection) is widely spread, with very good results in its prediction with the ML methods found during this review.

Keywords: Electrodermal activity; Stress detection; Machine learning; Scoping review

1. Introduction

Stress usually produces a mental state of tiredness that may result in physical and mental disorders as a major depression, insomnia or generalised anxiety [1–4]. From an emotional point of view, calm and stress situations cause a variation in arousal. Arousal is the increase of activation or deactivation produced in the brain in the presence of a stimulus [5]. Stress detection is a very current topic in many clinical and educational areas. For this reason, there is a growing interest in developing methods that make automatic detection possible. In addition, in order to detect emotions in all places and at all times, the technologies used are focusing on the adoption of wearable devices and the implementation of machine learning (ML) techniques.

These technologies usually work with the physiological conditions of the human body [6,7]. The acquisition, processing and monitoring of physiological variables allow the creation of a map of the physical, mental and cognitive state of a subject [8,9]. There are numerous physiological variables that are used for stress detection. In particular, we will focus on the analysis of electrodermal activity (EDA), as it has demonstrated to be very effective in the estimation of arousal. However, we are not only interested in the topic of stress detection, but we will focus on the different ML methods used to detect stress from EDA. For this reason, a scoping review is conducted that focuses on the use of EDA for stress detection through ML techniques.

The remainder of the article is structured as follows. Section 2 provides a short explanation about the methodology followed to do the review, pointing out the most important steps. Section 3 introduces a brief summary about the status of the topic addressed in the review. Section 4 describes the most relevant results and provides a discussion about the most important studies found. Finally, Section 5 offers the conclusions of this work.

2. Review Strategy

A scoping review aims to quickly map out the concepts in a given area [10]. Such a review allows to rely on the main sources and evidence available. It should be done especially when an area is complex or has not been subject to a thorough and complex examination before. This type of review has been widely used in certain research areas to systematically find their central body. As a result of work conducted in a scoping process, it has become possible to systematically review and classify a considerable volume of peer-reviewed literature in a specific field [10,11].

In this article, we have undertaken a scoping review that connects the concepts of bio-signals and ML techniques for the detection of physiological states in people. Within this more generic field, the specific scoping review shown in the present article is to study all the previous works dedicated to the detection of human stress using any type of ML technique on collected electrodermal activity (EDA) signals.

2.1. Search Strategy and Study Selection

A total of six scientific databases have been selected for the most comprehensive search possible on ML and EDA. The selected databases have been ScienceDirect, Scopus, IEEE Xplore, PubMed, ACM Digital Library and Web of Science. Articles labelled as “electrodermal activity”, “bio-signals”, “machine learning”, “deep learning”, “stress detection”, “support vector machine”, “artificial neural network” and “deep neural network”, among others, have been searched. As mentioned above, the examination has been carried out since when the records are kept in each of the databases until June 30, 2020.

395 articles were returned from the initial searches of the above databases. Table 1 shows the terms searched for each of the databases. In order to have as small a set of search terms as possible without losing the scope of the review, the queries have been refined by successive searches. This has allowed us to keep a reasonably manageable number of keywords without losing the perspective and focus of the scoping review.

Table 1. Search terms in different databases.

Databases	Searching Terms	Papers
Scopus	TITLE-ABS-KEY ((electrodermal activity OR Deep Learning) AND AND (LIMIT-TO (DOCTYPE , “ar”) AND (LIMIT-TO (SUBJAREA , “COMP” OR LIMIT-TO (SUBJAREA , “ENGI”) OR LIMIT-TO (SUBJAREA , “MEDI”) OR LIMIT-TO (SUBJAREA , “NEUR”))	46
IEEE Xplore	(Electrodermal Activity) AND (Machine Learning OR Deep Learning)	52
PubMed	(“electrodermal activity”[All Fields] AND (“machine learning”[All Fields]))	30
ScienceDirect	“Electrodermal Activity” AND “Machine Learning” OR “Deep Learning”	63
ACM Digital Library	[All: electrodermal activity] AND [All: machine learning] AND [All: deep learning] AND [All: Bio-signal] AND [All: stress] AND [Publication Date: (01/01/2000 TO 05/31/2020)]	97
Web of Science	“electrodermal activity” AND “machine learning”	107
Raw Total		395

2.2. Inclusion and Exclusion Criteria

A series of inclusion and exclusion criteria have been established to help us refine the results of the review and arrive at the final number of selected articles. The criteria used to include or exclude each of the publications are listed below:

- Only one copy of each of the duplicate papers initially collected in the different databases should be kept.
- Publications that evaluate the design and performance of new ML-based methods and algorithms should be included.

- Excessively general revisions must be excluded. Only reviews focused on the search topic should be included.
- Articles with few participants or with a small or non-representative dataset should be excluded.
- Articles that do not include any type of system or method for the detection of emotional state based on ML techniques will be excluded.

2.3. Study Selection

Figure 1 details the scheme followed to obtain the final selection of the articles through the scoping review. This search process has carried out in the stages of *identification*, *screening* and *inclusion* of the articles based on the given criteria. Within the identification phase are the direct searches in the various databases. A total of 395 papers have been found, of which 46 papers have been detected in Scopus, 52 in IEEE Xplore, 63 in ScienceDirect, 30 in PubMed, 97 in ACM Digital Library, and 107 in Web of Science. In the *screening* phase, the papers have been selected and eliminated according to the inclusion and exclusion criteria mentioned above. 89 duplicates have been removed from the various databases. Moreover, 144 articles have been removed as they were noticed not to be within the scope of the review after reading their abstract. The criterion here has been to only select papers using EDA signals, individually or in conjunction with other signals, and employing ML techniques. Finally, in the last stage (*inclusion*), from the remaining 162 articles a total of 104 articles have been removed after a thorough reading of the complete content. This way, 58 articles have been studied in depth in this scoping review.

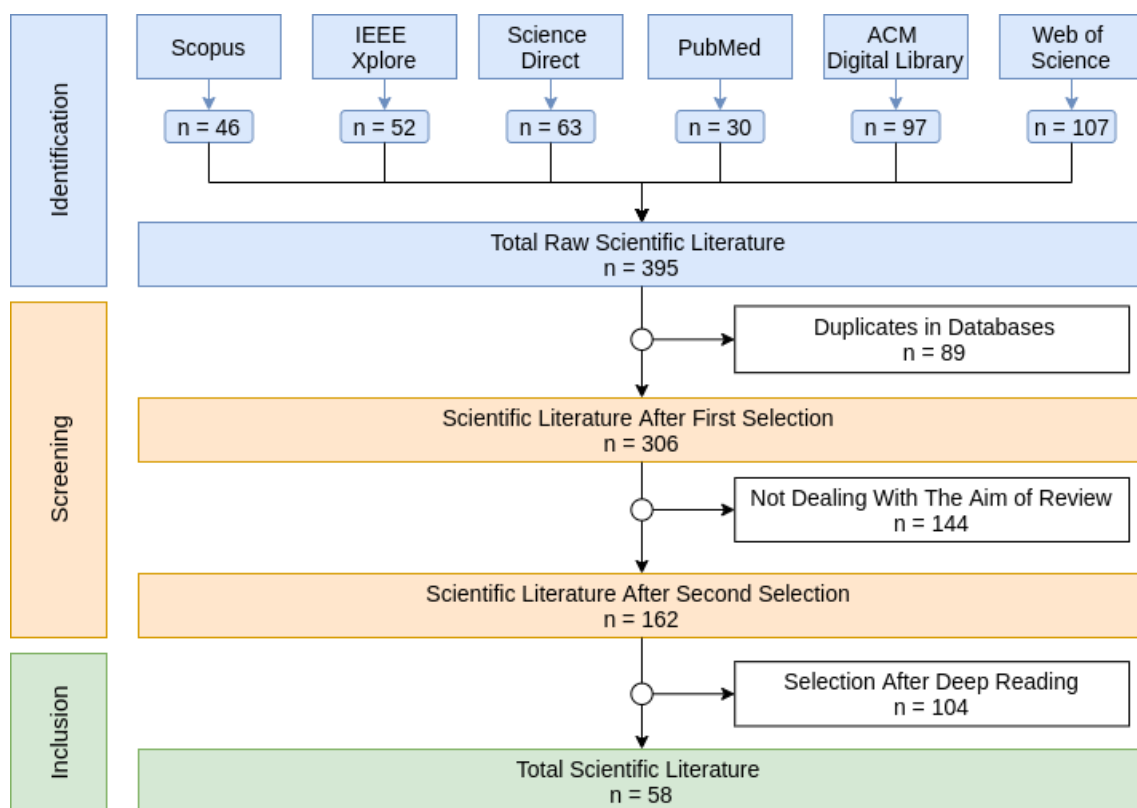


Figure 1. Search strategy.

2.4. Paper Classification Categories

The analysis of literature studied in this scoping review has identified different criteria to organise the articles and their approaches. The papers selected in the search can be classified into three main categories, as can be seen in Figure 2.

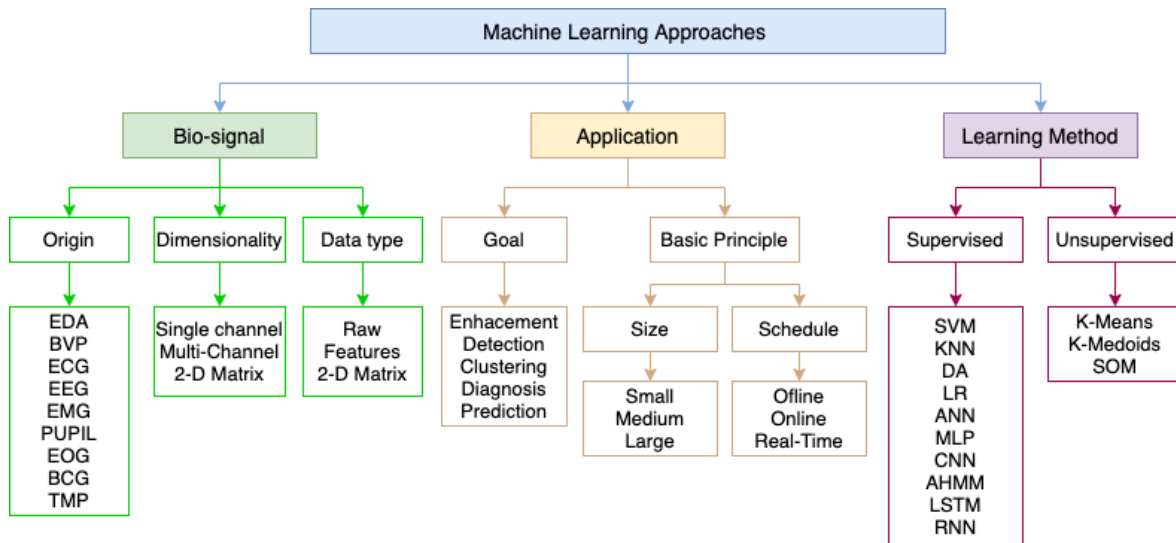


Figure 2. Paper grouping.

The first category, *Bio-signal*, is grounded on the different bio-markers used for the detection of calmness/stress. The specific bio-signals used to detect calm/stress can be easily identified within this group. The dimensionality of the data source can also be determined, i.e. whether a single source or multiple indicators are used for detection. In addition, the type of data used for detection can be identified, differentiating between raw data, processed data and two-dimensional matrices. This category will be extensively covered in Sections 3.3 and 3.4.

The second category, *Learning Method*, bases on the use and implication of different learning methods for the detection. The majority of the papers analysed root their learning capacity on supervised classification algorithms, while the use of unsupervised classifiers is minor. This second category will be expanded in Section 3.7.

The last category, *Application*, focuses on the creation of applications that employ different types of classifiers for a specific use. It centres on the goals that are to be achieved, focusing on the creation of applications for the detection, grouping, diagnosis and future prediction of emotional or calm/stress states. Other basic classification principles are whether the application works with a large or small number of participants and signals, and whether the system is intended to be used offline or in real-time. The category on applications is not covered in depth in this article, as it is somehow out of the objectives posed.

Table 2 has also been added, listing all the documents used in this scoping review, grouped according to the classification formulated. Naturally, there are repeated documents between several columns in the table. This is because there is no unambiguous separation in most of them. For example, the creation of an application requires ML methods and bio-signals to be processed.

Table 2. Paper classification by group.

Machine Learning Approaches		
Bio-signal	Application	Learning Method
[12] [13] [14] [15] [16] [17] [18] [19]	[13] [20] [21] [22] [23] [24]	[12] [15] [16] [25] [26] [27]
[28] [27] [29] [30] [31] [32] [33]	[34] [35] [36] [37] [38] [39]	[24] [35] [36] [40] [37] [41]
[42] [43] [44] [45] [46] [47] [36]	[48] [49] [50] [51] [52] [53]	[54] [48] [55] [50] [15] [46]
[49] [20] [38] [56] [48] [57] [58]	[59] [28]	[18] [60] [61] [17] [51] [62]
[22] [52] [63] [64] [34] [65] [66]		[57] [41] [28] [67] [19] [68]
[69] [23] [68] [25] [70] [71] [72]		[29] [73] [31] [32]
[50] [67] [60] [74] [39] [75] [62]		
[76] [77] [21] [37] [41] [40]		
[78] [54] [55]		

2.5. Data Analysis

After the final selection of the papers, different measurement indicators have been studied as part of the scoping review. Clearly, with the increase of the varied devices of use and also with the growth of the popularity and the easiness of use of the various ML techniques, the number of articles in the studied field has increased significantly. Evidently, one determinant analysis to know the importance of the topics studied is the scientific production of articles over time. As can be seen in Figure 3 the growth is almost exponential ($R^2 = 0.80$). Presumably, and given that the search in year 2020 has been limited to half the year, the production of scientific literature will increase and surpass that in previous years.

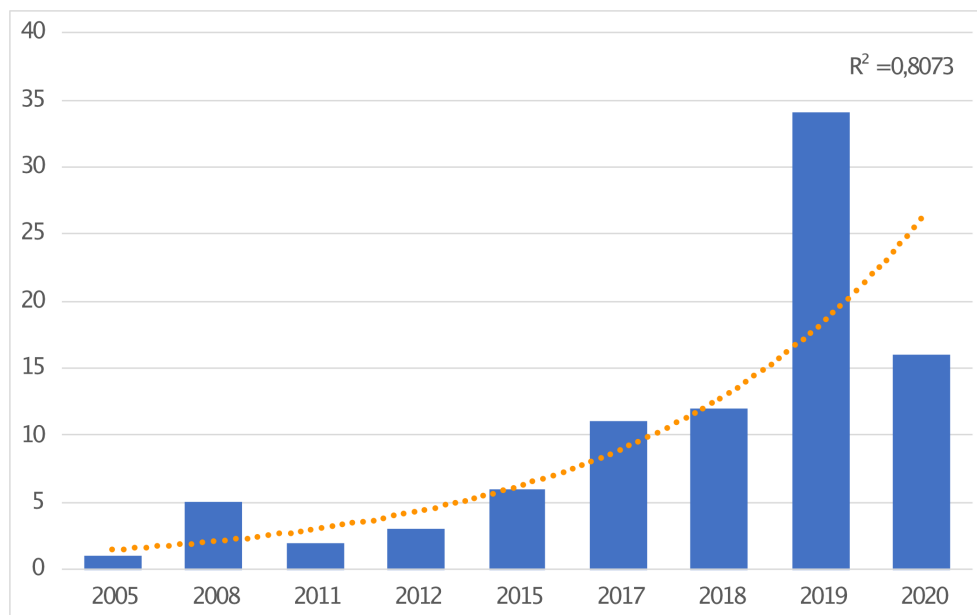


Figure 3. Scientific production according the scoping review of physiological signal and ML per year.

3. Methods in Machine Learning for Stress Detection

The human body can be seen as an electromechanical system composed of perceptive, affective and cognitive processes. The inherent dynamic changes allow different measurements to be made on various bio-signals. These time-changing signals allow to establish with enough precision the physical, psychological and cognitive state of the human being [79,80]. Most biological signals involve electrical activity and conductivity, as well as changes in flow, temperature, volume, pressure, sound and acceleration [76,81–83].

There are many physiological variables that can be collected from the human body. The most common are the following. (a) The electrocardiogram (ECG) measures any change in the heartbeat and its pattern of beating makes it possible to diagnose various coronary diseases. (b) Electromyography (EMG) monitors changes in neuromuscular activity. (c) Blood volume pressure (BVP) measures changes in blood volume, which affects blood pressure by changing the cardiac output. (d) Electrooculography (EOG) allows monitoring of eye movements. (e) Pupillometry or pupillometry (PUP) is based mainly on the measurement of the pupil diameters under basal conditions and after applying different stimuli. (f) Electroencephalography (EEG) measures the variation of electrical signals produced in different areas of the brain. (g) The inter-breath (IBR) measures the rate of breathing. (h) Acceleration (Acc) monitors body movements. (i) Skin temperature (TMP) is used to quantify temperature changes. (j) The electrodermal activity (EDA) is used to check the state of activation (arousal) and, consequently, one important component of the emotional state of the person. In Table 3 the main characteristics and properties of these bio-markers can be appreciated.

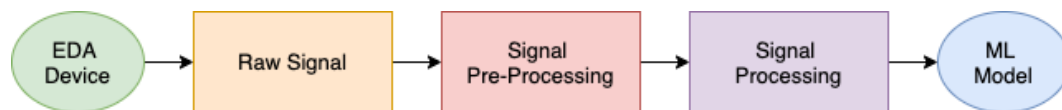
Table 3. Bio-signals and properties

Signal	Abb.	Channels	Sig. Freq. (Hz)	Rec. Freq.Hz	Amplitude Level
Electrocardiogram	ECG	1–12	0.05–150	250–1K	0.1–5
Electromyography	EMG	1–32	25–5K	512–10K	0.1–100
Blood Volume Pressure	BVP	1	0.25–40	5–500	-10–10
Electrooculography	EOG	2	0–100	1–100	50–3.5K
Pupulography	PUP	2	120	240	–
Electroencephalography	EEG	1–128	128–2K	128–2K	1- 150mV
Inter-Breath	IBR	1	1–20	1–20	-0.05–0.05
Acceleration	Acc	3	20-2k	20-2k	-1– 1v
Skin Temperature	TMP	1	1–200	2 - 50k	-50–50
Electrodermal Activity	EDA	1	1–16	16–128	0–100 μ S

Correct monitoring of these signals provides us with a map of the physical, cognitive and emotional state of a person. It must be considered that the process that has taken place since the bio-signals are known and until it has been made possible to use them depend to a great extent on the technology available. The evolution of the technology is influenced by the historical use of “handmade” signals obtained by the criteria of an expert technician. The specialists were in charge of processing the signals and transforming them into data used by the detection system. Obviously, this procedure has clear limitations: (a) it is an inefficient process, since all the weight falls on the efficiency of the professional and, (b) if the expert does not have much experience, the signals may present deficiencies and cause the system to not operate correctly. Currently, this procedure can be performed without almost any expert intervention. ML contributes to automation and analysis based on biomedical signals [84].

3.1. Acquisition Model

The acquisition of the signal is one of the most important processes in using EDA (or any other bio-signal). Most authors related in this review, agree that a good acquisition process is crucial for the subsequent recognition system to work properly. Most acquisition systems have as their main scheme the model shown in Figure 4. As usual, these systems have three distinct phases. The first phase is the acquisition of the raw signal. The signal originates from the device that measures the physiological variable. The next phase of the treatment of the signal is preprocessing. This stage is designed to eliminate from the signal all the defects that have interfered during the acquisition process. As part of this operation, artefacts are removed and the signal is filtered, making it softer and eliminating possible noise. Finally, the next phase consists of signal processing. In this stage, as a rule, a series of characteristics of the signal are obtained that will allow it to be used later by means of a classifier.

**Figure 4.** Usual acquisition model.

3.2. Datasets and Experimental Designs

According to the outcomes of our scoping review, the authors always choose between two different procedures to acquire the raw signals. The first one is to create an experimental design to obtain the EDA signal that will be processed later. As a rule, this phase is done as shown in Figure 5. A first step is to start the experiment, then the baseline recording of the input data is started. Next, the person is subjected to a sensory stimulus, most commonly visual and auditory, and the individual’s reaction is recorded. The process is labelled and repeated as many times as necessary.

A second alternative procedure to the previous one uses several datasets already validated by the scientific community. These datasets were also recorded in the same manner, usually containing a

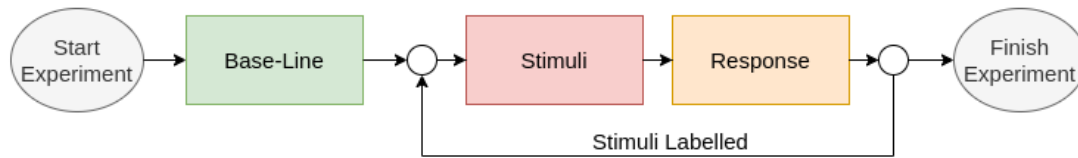


Figure 5. Flowchart of the experimental design during raw signal acquisition.

number of other physiological signals recorded in addition to the EDA signal for use in multi-class classifiers. The most common datasets used for EDA analysis are MAHNOB [85], DEAP [86], BioVid [87] and UT Dallas Database [88].

3.3. Signal preprocessing

As mentioned above, at this stage the signal is preprocessed for cleaning. Preprocessing procedure is used to clean, adapt and prepare the obtained signals for further processing. This process is also critical as many authors agree that the effectiveness of a classification system begins at this stage. As a rule, this process comprises two different steps, the detection and elimination of artefacts, and the filtering of noise.

3.3.1. Artefact detection and removal, and noise filtering

An artefact is defined as a distortion, addition or error in a signal in an uncontrolled manner. An artefact may be due to changes in pressure or disconnection of the electrodes that capture the signal during the acquisition of the EDA signal. This type of motion artefact (MA_t) could degrade the signal very quickly and, as a result, make the signal unusable [63]. With the MA_t it is really important to know if it is better to just detect it or, on the contrary, suppress it. Artefacts are suppressed by deflecting the signal through the application of various softening filters [89,90]. This procedure causes a loss of information in the EDA signal in most cases.

Moreover, the detection of the MA_t consists of identifying each of the signal segments in which the artefact appears and thus being able to eliminate it in later phases. Often this type of removal must be done manually. The human operator must be expert in the application of the filters and remove the MA_t without losing information. Fortunately, there are studies that automatically detect and remove artefacts in EDA [63,91].

Closely associated with the artefact detection and/or removal process is noise reduction or elimination. Noise in a signal can be defined as an unwanted disturbance in an electrical signal. In signal acquisition systems, noise is an unwanted random error or disturbance of a useful information signal such as an unwanted or disturbing sum of energy from natural and sometimes artificial sources. In the EDA signal, due to the slow evolution of the same, the noise that should concern us is high frequency [83]. Therefore, the EDA signals are filtered to remove artefacts and noise recorded during the acquisition period. Generally, two different types of filters are used; firstly, a low pass filter with a 4 Hz cutoff frequency, and secondly, a Gaussian filter to attenuate the signal, artefacts and noise.

3.4. Signal Processing

The next step in the process that the signal goes through is the processing phase. In this stage, the EDA signal (already preprocessed) is treated to obtain its features. These features will be in charge of quantifying in a unique way each of the obtained segments. This process of treatment of the EDA signal is known as *deconvolution process*. Although some of the authors studied do not use this process, convolutional analysis is very common because this method allows to standardise and treat a large group of signals from different participants [17,92].

3.4.1. EDA Deconvolution

As already stated in this revision, the processing of EDA signals consists of different phases. Typically, these phases are preprocessing, filtering, artefact removal and discrete deconvolution. The measurement of EDA signals are usually made on the palms of the hand, the fingertips or the wrist. These measurements are composed of the convolution of two signals: a first signal, which varies slowly, called the skin conductance level (SCL), and a second signal that varies rapidly, the skin conductance response (SCR). The SCL signal establishes the base level of the signal while the SCR is closely related to the activity of the sweat motor system which, at the same time, is closely associated with the parasympathetic nervous system [93].

The deconvolution procedure consists of separating the SCR signal from the SCL one. This approach makes it possible to minimise any effect that race, sex, and age might have on the SC signal. Figure 6 illustrates the process. As you can see, the SCR “driver” is the one that can be used to detect the level of excitation of the individual.

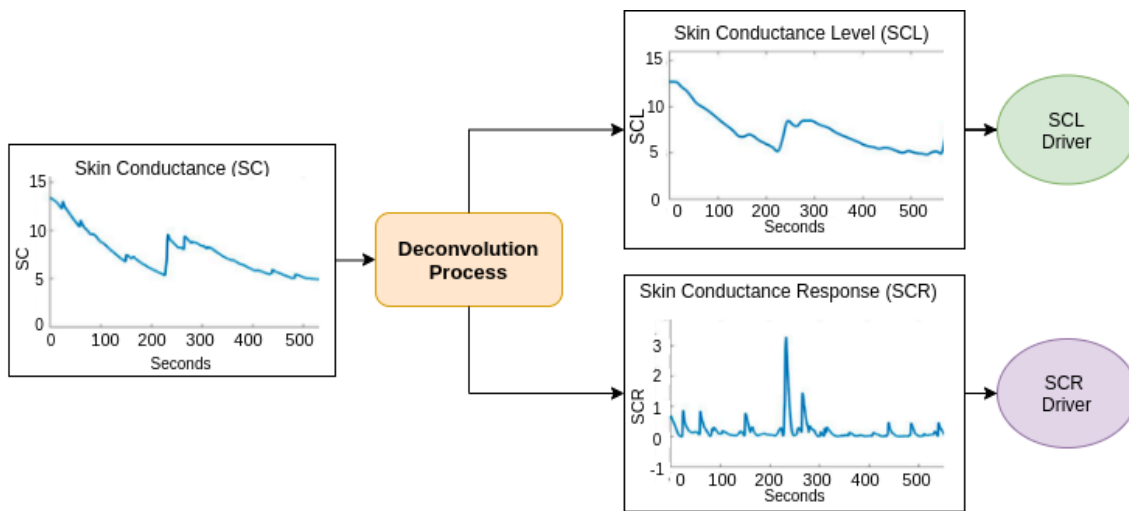


Figure 6. Flowchart of the deconvolution process

Mathematically, the sudomotor nerve function may be considered a driver that has a train of impulses that evolve over time. This response is embedded in the SCR and SCL signal. The outcome is presented by a convolution (*) of the driver with the impulse-response function (IRF), describing the impulse response flowing through time as shown in equation (1).

$$SC = SC_{Driver} * IRF \quad (1)$$

The SC signal is formed by the SCL and SCR signals, as displayed in equation (2).

$$SC = SCL + SCR \quad (2)$$

$$SC = (SCL_{Driver} + SCR_{Driver}) * IRF \quad (3)$$

Thus, by deconvolution of equation (3), the tonic signal driver is obtained as:

$$SCR_{Driver} = \frac{SC}{IRF} - SCL_{Driver} \quad (4)$$

After this stage, the signals obtained can be used in the feature extraction process.

3.4.2. Feature Extraction

Many of the bibliographic references consulted in this survey coincide in using different features to quantify the signal. Four main groups of features can be distinguished: *time domain features*, which refers to all the variables that can be defined in terms of time; *frequency domain features*, which refers to all the parameters that can be defined in terms of frequency or based on them; *statistical features* defined as variables that belong to the statistical field; and, *morphological features* that quantify the shape of the signal. Table 4 details several features that use to characterise the different segments of the SC, SCL and SCR. It should be noted that these features are used to characterise the signals more accurately. It is a good practice to use the best features that are most suited in relation to their contrasting performance.

Table 4. Features obtained in the process

Domain	Features
Time	Mean, SD, EDL, D1, D2, D1M, D2M, D1SD, D2SD, SRT, SFT, RM, RRSTD, DCRM, DCRSD, RM, PHVM, PHVSD, RRSTD, DCRM, DCRSD, STM, STSD, STRMS, STRMSSD, STRMSOV, EDL, EDR, CMax, CMin, SWE, DR, RMS, PMRMSR, RSSL, P, PLoc, PPT, pNN50
Morphological	NO, EC, EP, EPC, EN, AL, IN, AP, RMS, IL, EL
Statistical	M, Var, MedVal, p-Val, AKAIKE, LOG-LIKE, FCM, FVCM, KU, SKU, MO, COVMAT
Frequency	SP, SSP, MSSP, SSPMed, NSSCRs, FFT

In the time domain, the following features are commonly used: mean amplitude (Mean); amplitude standard deviation (SD), the SD first and second derivative (D1, D2), the SD means (D1M, D2M) and their standard deviations (D1SD and D2SD) [29]; sum rise time (SRT), sum fall time (SFT), rise rate mean (RM), rise rate standard deviation (RRSTD); decay rate mean (DCRM), decay rate standard deviation (DCRSD); phasic value mean (PHVM), phasic value standard deviation (PHVSD); startle time mean (STM), startle time standard deviation (STSD), startle RMS mean (STRMS), startle RMS standard deviation (STRMSSD); startle RMS overall (STRMSOV); electrodermal level (EDL), electrodermal response (EDR); cumulative maximum (CMax), cumulative minimum (CMin); smallest window elements (SWE); dynamic range (DR); root-mean square level (RMS), peak-magnitude-to-RMS ratio (PMRMSR); root-sum-of-squares level (RSSL); peak (P), peak location (PLoc), peak to peak time (PPT), peaks intervals differs 50ms (pNN50) [24,40,41,53,65,75].

According to the morphology of the signal we can find these different features: epoch-capacity (EC) is a relation between the number of epochs and the total number of them; epoch-peak (EP); epoch peak counter (EPC) is a number of epochs in all times; entropy (EN) [18]. On the other hand, there are features that result from different measurements such as arc length (AL), integral area (IN), normalised mean power (AP), root mean square (RMS), perimeter to area ratio (IL) and energy to perimeter ratio (EL) [29]. These parameters are due to the need to understand the morphological differences in the shape of the SCR_{Driver} . Not only is it a matter of studying signal peaks, but changes in the general morphology of the signals are of interest.

As far as statistical parameters are concerned, let us highlight: mean value (M), variance (Var), median value (MedVal), p-value (p-Val), Akaike information criterion (AKAIKE), Log-likelihood (LOG-LIKE), covariance matrix (COVMAT), transition probabilities lag (TPL), number of observations (NO), switching betas (beta-Numb), number of estimated parameters (STP), standard error coefficient (SCE), smoothed probabilities of regimes (SPR), conditional standard deviation (CSTD), four central moment (FCM), five central moments (FVCM), kurtosis (KU), skewness (SKU) and momentum (MO) [15,41].

Finally, in the frequency domain, the following parameters can usually found: sum spectral components (SSP), spectral power (SP), mean spectral components (MSSP), median spectral components (SSPMed), frequency non-specific of skin conductance response (NSSCRs), fast Fourier transform (FFT) for bandwidths F1 (0.1, 0.2), F2 (0.2, 0.3) and F3 (0.3, 0.4) [2,15,29,41].

3.5. Machine Learning for Stress Detection

As a general rule, these kinds of signal-based experiments yield a large amount of data that has to be processed further. It is necessary to use techniques that help us to classify this enormous amount of data. Most authors prefer ML-based techniques as opposed to other techniques based on pure statistical analysis (ANOVA, MANOVA and Student's t-test). For this reason, a comprehension of existing ML models, their main characteristics and methods of evaluation, and their most relevant results is essential. Therefore, next we will study the metrics used, the different classification methods employed, as well as the most relevant results described in each of the publications reviewed.

3.5.1. Evaluation Metrics

According to the literature studied, the process of stress and calm detection is limited to a binary classification problem. The different metrics used to measure performance have been identified. They are presented below, but to understand what they mean, it is necessary to introduce the terms true positive (TP), true negative (TN), false positive (FP) and false negative (FN). On the one hand, TP and TN are used to indicate whether an element of a set has been classified as corresponding to its category. On the other hand, it is possible to define FP and FN when an element belongs to a class completely contrary to that in which it was classified.

The metrics used to assess the performance of the detection are shown as follows:

- The **accuracy (ACC)** is the simplest metric one can imagine. It is defined as the number of correct predictions (TP and TN) divided by the total number of predictions (TP , TN , FP and FN).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- The **precision (P)** refers to the ratio of successful positive predictions (TP). It is measured as the number of true positives (TP) divided by the total number of positive cases.

$$P = \frac{TP}{TP + FP} \quad (6)$$

- The **recall (R)** is defined as the proportion of positive cases captured. Recall describes the sensitivity of the model for identifying the positive class. It is computed as the number of true positives divided by the sum of true positives and false negatives.

$$R = \frac{TP}{TP + FN} \quad (7)$$

- The **specificity (S)**, also named **False Negative Rate (FNR)**, is defined as the ratio between the true negatives and the total number of true negatives and false positives (negatives that have been labelled as positive). This implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. This proportion could also be called a false positive rate. The sum of specificity and false positive rate would always be 1.

$$S = \frac{TN}{TN + FP} \quad (8)$$

- The **F1-score**, also called F-measure, is a measurement of the accuracy of a test. It is well defined as the harmonic mean between accuracy and recall. It is applied as a statistical assessment of performance.

$$F1\text{-score} = \frac{2 \times P \times R}{P + R} \times 100 \quad (9)$$

- The **area under the curve (AUC)** or **receiver operating characteristics (ROC)** curve is a performance measurement for classification problems at various thresholds settings. ROC

is a probability curve and AUC represents a degree or measure of separability. It tells how much a model is capable of distinguishing between classes. The higher the AUC the better the model is at predicting. Also, the higher the ROC the better the model is at distinguishing between classes. This chart is constructed by plotting sensibility in the y-axis and specificity in the x-axis.

- The **confusion matrix (CM)**, also known as an error matrix, is a specific table disposition that allows to visualise the performance of an algorithm. Every row of the matrix is an instance of an expected class, while every column is an instance of an actual class (or vice versa).
- The **kappa-coefficient** is a measure of how closely the instances classified by the machine learning classifier match the data labelled as ground truth, controlling the accuracy of a random classifier as measured by the expected accuracy.

3.6. Classification Methods

In the several studies analysed from this scoping review, different classification methods have been found. These methods are based on different parameters. In first place, there is a *direct classification* vs a *hierarchical classification*. Furthermore, if considering the duration of the classification, there is *long-term* vs *short-term*. Finally, we can distinguish between *supervised* vs and *unsupervised learning* methods. Another aspect that must be considered is that ML models have some limitations due to the large number of parameters handled. Consequently, it is necessary to know how to implement methods that will help us reduce the number of redundant or non-relevant parameters. Dimensionality reduction techniques are becoming significant in the areas of ML, data mining and bio-informatics.

Therefore, one of the main objectives of classification is to eliminate redundant and interdependent features, reducing the dataset to a smaller dimensional space. The number of dependent variables must be reduced, but the bulk of the data must be maintained. In the literature related to signal processing methods, the feature reduction methodologies detailed next are usually used. Principal component analysis (PCA) is a standard statistical data analysis, which attempts to explain observable signals as a linear mixture of the orthogonal principal component that optimises the variance between the different components. Linear discriminant analysis (LDA) is typically used to reduce the dimensionality by maximising the space between the different classes. Finally, independent component analysis (ICA) is an analysis and data processing strategy that attempts to recover unobservable signals or sources of monitored mixtures only under the assumption of mutual independence between them.

The use of these feature reduction techniques allows the computational cost to be much lower, since the resulting classifier is simpler and only attends to the important features of the signal. Many of the papers studied in this overview use these techniques and the results are really good compared to others who do not use these techniques. Below is an explanation of the different methods used.

3.6.1. Direct vs Hierarchical Classification

In many of the articles analysed in this review it is possible to find these two types of classification: direct and hierarchical. A *direct classification* consists in classifying the emotional state of the person in a direct way considering one or more physiological variables. This process is quite common when dealing with simple emotional models or classifiers that only seek to find a state, as in our case with the use of EDA for stress detection (increase or decrease of arousal). On the other hand, when a *hierarchical classification* is proposed, there are two distinct stages. In the first stage, the emotional state is established (positive or negative) and in a second stage, the more complex emotional state is classified [15].

3.6.2. Long-term vs Short-term Affective State Classification

Another aspect to consider is whether a classification of the emotional state should consider the duration of the experiment, as well as the evolution of the signals over time. The first aspect to be pointed out is the need for a classifier that works quickly and is consistently robust over a long period of time. In this sense, such a classification can be defined as *short-term* and *long-term*. The former is

aimed at almost instantaneously determining results, while the latter is oriented towards long-term applications. In the context of stress detection, a long-term classification is usually recommended [29].

3.7. Supervised vs Unsupervised Learning

Within the different learning methodologies, there are (apart from reinforcement learning and stochastic learning) other two main groups, namely *supervised* and *unsupervised* learning [94]. From the perspective of classification, these are opposite approaches. The first one emphasises and uses the different labelled datasets to be able to make the predictions accordingly. The second option uses unlabelled datasets to analyse whether there are large groups of data that reflect a situation (i.e. stress detection or emotional state detection).

3.7.1. Supervised Learning Methods

As mentioned above, supervised learning techniques are based on training a classifier from a dataset that is already labelled. Once the system has learned to identify the different patterns, the classifier is able to effectively distinguish between the different classes. In our particular, it has to distinguish between calm and stress states.

There is a wide range of classifiers with supervised learning. Below you can find the most commonly used classifiers, as well as a small summary of their internal operation. The following are worth mentioning: *support vector machines*, *auto-hidden Markov models*, *discriminatory analysis*, *decision trees*, *naive Bayes*, *logistic regression*, *k-nearest neighbours*, *artificial neural networks*, *long short-term memory* and *recurrent neural networks*:

- **Support vector machines** (SVMs) were originally conceived by Vapnik in 1995 to solve a dual classification challenge [95]. The algorithm has been evolving to solve the multi-classification problem [96]. This algorithm seeks a hyperplane that optimally divides elements of different classes. The hyperplane is the one with the minimum distance (margin) to the closest points to itself. Depending on the number of features and their dimensions, sometimes a transformation in vector space is required to achieve the best possible separation. The search for the separation hyperplane in the transformed spaces, usually of very high dimension, is based on the kernel function. From the point of view of emotion detection by EDA it is one of the most used algorithms, since it responds very well to the different features obtained during the acquisition process. The most used kernels are *linear* [24,65,71,72]; *quadratic* [35,65,75]; *polynomial* [65,71,75]; *Gaussian* [65,71]; and *radial* [17,19,26–28,33–36,38–41,44,47,49,53,55,58,60,63,64,67,69,72,74,78,97].
- An **auto-hidden Markov model** (AHMM) is defined as a discrete-time Markov chain whose states are hidden. In each state there is an observed output in accordance with a state-dependent distribution. In an auto-regressive HMM (AR-HMM) the outputs to be monitored are generated by an auto-regressive model with coefficients that depend on the current state of the HMM [15,16]. This algorithm have been widely used for classification purposes using three different criteria: (a) the sequence of the observations and the number of different HMMs determine the most likely HMM from which they are derived; (b) for each sequence of events that came from a particular HMM, determine the most likely sequence of events that gave rise to these observations; (c) for all sequences of observations identify the most likely parameters of the HMM that gave rise to the sequence [15,16]. The different approaches are used to determine the status of each person from the EDA signals.
- **Discriminant analysis** (DA) groups those methods based on Bayes theorem to estimate probabilities. A DA makes predictions based on the probability that a new set of input data will belong to each class. The class that has the highest probability is considered the output class. It also considers the possibility of each class and the probability of the data belonging to each class. There are many classifiers based on DA, being the most common for the detection of stress in EDA: *linear discriminant analysis* (LDA) [20,53]; *quadratic discriminant analysis* (QDA) [13,17,44,68,71] and *Gaussian discriminant analysis* (GDA) [65].

- **Decision trees** (DTs) are described as algorithms that perform repeated splits in the dataset to provide maximum data separation. The resulting structure is similar to a tree [98]. The most frequently used criterion is information gain, This implies that the entropy reduction caused by dataset division is maximised in every split. Within this type of classifier the most used are *tree medium*, *regression tree* [13,18,27,55,64,68] and other ensemble methods like *random forest* and *bagged tree* [18,75].
- **Naive Bayes** is defined as a type of probabilistic classifier that aims to process and categorise data. The operation of this classifier is simple. It is essentially a technique for assigning probability theory to classify data. Naive Bayes classification algorithms utilise the Bayes theorem. The central idea is that the probability of an event may be adjusted as new data are entered. This classifier is not a single algorithm, but a family of automatic learning algorithms that makes use of statistical independence. These are relatively easy to write and perform more smoothly than the most complex algorithms. A Bayes classifier is defined as naive if all the properties of a particular data item are independent of each other. This classifier fits well for calm/stress detection as it has two classes. During this study it has been found that the most used naive Bayes methods in our scoping review are *naive-Bayes-Gaussian* [18,27,34,44,64] and *naive-Bayes-Gaussian with PCA* [18,64].
- **Logistic regression** (LR) is a classification algorithm used to predict the probabilities of a targeted variable. The objective or dependent variable is dichotomous, meaning that there may only be two possible classes. The dependent variable is binary and has data encoded as 1 (success) or 0 (failure). According to the references found, different papers have been published where this methodology is used as *logistic regression* [13,19,28,63] and a variant called *zero-regression* [28].
- A **K-nearest neighbours** (KNN) classifier is different from all the other methods considered here, as this algorithm uses the dataset directly for classification and does not build a model first [99]. The benefit KNN have over other algorithms is the possibility for the neighbours to explain the result of the categorisation. This explanation may provide an advantage in areas where black-box models are inappropriate. Within the different configurations that have been found are *KNN-Fuzzy* [75], *KNN-Fine* [75], *KNN-Cubic* [20,75], *KNN-Medium* [13,16,19,27,34,40,41,53,55,60] and *KNN-Weighted* [63].
- **Artificial neural networks** (ANNs) are widely used computer models in prediction. ANNs consist of a set of units, called artificial neurons, connected together to transmit signals. The input information passes through the neural network (where it undergoes various operations) producing output values. Each neuron is connected to others through links. As a general rule, they have one or more layers of interconnection, called hidden layers, which are responsible for connecting the neurons to each other. Between each of the neurons, a gate called the activation function which allows each neuron to be activated or deactivated according to a certain activation function (i.e. sigmoid, step function, hyperbolic tangent, linear, rectified linear(ReLU), leaky-ReLU, among others) is usually added. It should be noted that there are many topologies that can be used for the processing of the obtained features, such as *feed-forward NN* [41], *multi-layer perceptron with back-propagation* (MLP) [13,25,38,63,64,68,72], *Bayesian probabilistic NN* (BPNN) [34], *probabilistic NN* [64], *1-dimension convolutional NN* (1D-CNN) [20,41], and, finally, *convolutional NN* (CNN) [17,34,35,67,69,78].
- **Long short-term memory** (LSTM) and **recurrent neural networks** (RNNs). A RNN is a generalised version of the neural network which has an internal memory. On the other hand, a LSTM is a customised model of recurrent neural networks which facilitates the memory of past data. The LSTM is very adequate to classify, process and predict time series given time intervals of unknown duration [100]. Due to the properties of this type of network for predicting signals over time, many authors use them to work with physiological signals. In this scoping review they have been used as *LSTM* [62] and ensemble-based methods like *CNN+LSTM* [62] and adaptive neuro-fuzzy inference system (*ANFIS-based short-term*) [53].

3.7.2. Unsupervised Learning Methods

The second group of learning methods addressed is unsupervised learning [101]. This type of methods is based on learning by using an unlabelled dataset. The model obtained is automatically adapted to the observations. The model is mainly created through the use of clustering methods. According to the literature found in the scoping review the following unsupervised methods have been used:

- **K-means** is a clustering method, aimed at splitting an unlabelled dataset of n observations into k groups in which every single observation belongs to the group whose mean value is the closest. The algorithm explores a predefined number of clusters in an unlabelled multi-dimensional dataset, and concludes with an easy interpretation of how an optimised cluster might be formulated. Primarily the approach would be a two-step process. First, the centre of the cluster is the arithmetic mean of all data points associated with the cluster. Second, each point is contiguous to its cluster centre compared to other cluster centres. The fundamental idea is to find the clusters that cover the greatest number of elements while minimising inertia [40].
- **K-medoids** is a grouping approach for the partitioning of a dataset into k groups or k -clusters, each group being represented by one of the group data points. These points are called cluster medoids. The term medoid describes an item within a cluster for which the average difference between it and all other members of the cluster is minimal. It corresponds to the most central point of the cluster. The K-medoid is a strong option to the K-media cluster. This implies that the algorithm is less sensitive to noise and outlines [40].
- A **self-organising map** (SOM) is a type of ANN that is formed by the use of unsupervised learning to generate a low-dimensional map, typically two-dimensional [102]. SOMs are different from ANNs in applying competitive learning unlike the error corrective learning (such as gradient descending feedback), and in the way that it uses a neighbourhood function to maintain the topological properties of the input space. In the selected literature we have found the use of SOMs for the detection of arousal [40,44].

4. Results and Discussion

This section presents the different results obtained during this scoping review. As it has been commented throughout the paper, different analyses of the data obtained are carried out in this type of review. Our analysis focuses on determining which are the best classifiers (supervised and unsupervised) for calm/stress detection. Complementary, an analysis has been carried out to evaluate if the EDA as a single variable or in combination with other physiological signals is a estimator estimator of arousal (stress) variation.

4.1. Bio-markers Used in the Papers

One of the considerations taken during this study was to analyse the number of articles that only use EDA to make the different classifications, as well as those in which, in addition to the EDA signal, other biological markers are used to reinforce the classification results. As can be seen in Table 5, the publications have been grouped according to the number of variables used. Note that the EDA variable is present in all eligible documents. As already mentioned throughout the article, this is because EDA is a very good estimator of the activation level due to its connection with the parasympathetic nervous system through the sweat-motor system. In the works found, a minimum of 5 participants and a maximum of 260 have been counted, having used other variables besides EDA like BVP, TMP, EEG, EOG, EMG, Acc, PUP and IBR.

A total of 20 papers have only used EDA signals for stress detection [17–20,24,27–33,36,42,43,45–47,49,53,73,103]. The authors have focused on developing methods or evaluating ML models based in EDA and its features. The use of deconvolution methods were emphasised to obtain the different features of the EDA signals.

Table 5. Bio-markers used in the papers

Papers	Year	Parameters	Part.	Evaluation	Annotations
Anusha et al. [103]	2017	EDA	12	ML Met.	Stressors in EDA
Sadeghi et al. [18]	2020	EDA	41	ML Met.	Empatica E4
Susam et al. [24]	2018	EDA	34	ML Met.	
Rizwan et al. [19]	2020	EDA	5	ML Met.	
Garcia-Faura et al. [28]	2019	EDA	14	ML Met.	
Al-Machot et al. [27]	2018	EDA	30	SAMS + ML	MAHNOB dataset
Ganapathy et al. [17]	2020	EDA	32	ML Met.	Convolutional Analysis
Sanchez-Reolid [29]	2020	EDA	147	ML Met.	D-SVM based
Greco et al. [30–33,73]	2014–19	EDA	18–32	ML Met.	
Hernando-Gallego [42]	2018	EDA	100	SparseEDA	Deconvolution method
Hwang et al. [43]	2017	EDA	17	ML Met.	
Jang et al. [44]	2015	EDA	40	ML Met.	
Kelsey et al. [45]	2018	EDA	73	ML Met.	
Khalaf et al. [46]	2020	EDA	260	ML Met.	Clustering features maps
Momin et al. [47]	2019	EDA	ND	ML Met.	Oriented Tasks
Taylor et al. [36]	2015	EDA	100	ML Met.	Wavelet transformation
Liu et al. [49]	2019	EDA	21	ML Met.	Kappa coefficients
Ding et al. [20]	2020	EDA	35	ANOVA + ML	
Patlar-Akbulut et al. [15]	2020	EDA BVP	30	LDA + Acc.	Exploratory res.
Zontone et al. [38]	2020	EDA BVP	18	ML Met.	
Posada-Quintero [12]	2019	EDA BVP	16	ML Met.	Multi-class classif.
Alhamid et al. [56]	2013	EDA BVP	–	ML Met.	
Cho et al. [48]	2017	EDA BVP	12	ML Met.	Unsupervised Learning
Singh et al. [57]	2013	EDA BVP	19	ML Met.	NN topologies
Pinto et al. [58]	2019	EDA BVP	23	ML Met.	Multi-class classifier
Posada-Quintero [22]	2019	EDA BVP	70	ML Met.	
Kukolja et al. [52]	2014	EDA BVP	14	ML Met.	
Zhang et al. [63]	2017	EDA BVP TMP	87	ML Met.	
Zhao et al. [64]	2018	EDA BVP TMP	32	ML Met.	PCA analysis
Ali et al. [34]	2018	EDA BVP TMP	30	ML Met.	MAHNOB dataset
Jebelli et al. [65]	2019	EDA BVP TMP	10	ML Met.	Stress in workers
Kleckner et al. [66]	2018	EDA TMP	20	ML Met.	
Chowdhury et al. [69]	2019	EDA BVP TMP	22	F-score + ML	
Regalia et al. [23]	2019	EDA BVP ACC TMP	69	ML Met.	Empatica E4 based
Sabeti et al. [68]	2019	EDA BVP ACC TMP	20	LUCKK	
Thiam et al. [25]	2019	EDA BVP EMG	87	ML Met.	BioVid Database
Batista et al. [70]	2019	EDA BVP EMG	15	ML Met.	
Liu et al. [16]	2019	EDA BVP EMG	17	Accuracy	Markov-Chain Based
Setz et al. [71]	2010	EDA EMG	33	ML Met.	Stress cognitive
Al-Machot et al. [72]	2019	EDA BVP EMG IBR	30	SAM + ML	MAHNOB dataset
Hadi et al. [50]	2019	EDA BVP IBR EMG	59	ML Met.	SVM-RBF best perf.
Katsis et al. [67]	2008	EDA BVP IBR EMG	20	ML Met.	Automatic method
Khezri et al. [60]	2015	EDA BVP IBR EMG	20	ML Met.	
Kim et al. [74]	2018	EDA BVP EEG	30	ML Met.	
Siddarth et al. [39]	2020	EDA BVP EEG	12	ML Met.	LSTM model
Cavallo et al. [75]	2019	EDA BVP EEG	34	ML Met.	Multi-class model
Rastgoo et al. [62]	2019	EDA ECG	6	ML Met.	LSTM model
Zhang et al. [76]	2017	EDA ACC	87	ML Met.	Unsupervised ML
Sandeghi et al. [18]	2019	EDA BVP ACC	20	ML Met.	
Alam et al. [77]	2016	EDA BVP ACC	–	ML Met.	
Alcañiz-Raya [21]	2020	EDA BVP PUP	52	ANOVA + ML	
Gjoreski et al. [37]	2020	EDA EOG PUPIL	68	ML Met.	
Anderson et al. [55]	2017	EDA BVP EOG	41	ML Met.	Multi-class classifier
Bianco et al. [41]	2019	EDA BVP IBR	68	ML Met.	Deep classifier
Fiorini et al. [40]	2020	EDA BVP IBR	50	SAM + ML	
Katsis et al. [78]	2011	EDA BVP IBR	5	ML Met.	Multi-class classification
Martinez et al. [54]	2019	EDA BVP IBR	18	ML Met.	Expert system

Another variable that helps to determine different emotional states in the participants is BVP, which combined with EDA gives very good results in the prediction as well [12,15,20,22,38,48,56–58].

Table 5 shows additional variables. The following major cluster (besides EDA and BVP) includes TMP. These articles focus on the integration of this variable for stress detection. On the other hand, when adding the EMG signal, the results are slightly improved. This may be due to the fact that this physiological signal comes from the central nervous system and complements itself very well with EDA. Another variable used for stress measurement is EEG mixed with EDA. This type of signal is widely used individually and provides good results in stress detection. Nonetheless, EEG requires very expensive and precise devices and quite specific knowledge to set up the acquisition of the signals. Finally, IBR also provides additional information to improve the classifiers, but without achieving great improvements.

4.2. Supervised Learning Methods

A large number of the papers found use supervised learning methods (see Table 6). Below are the results obtained for each of the ML algorithms used.

4.2.1. Support Vector Machines

SVMs are the most widely used classification methods for detecting stress. Throughout the present survey, results have been obtained for each of the existing configurations. SVMs with linear, quadratic, cubic, polynomial, Gaussian, radial and radial with PCA kernels have been proposed for stress detection.

In the case of the linear kernel, the average accuracy in the articles studied is 55.64%. This value can be considered bad, but, as a rule, linear classifiers do not work very well with this type of physiological variable. On the other hand, for the quadratic kernel we have results of 85.26%. This type of kernel already shows very good results for EDA. In this sense, the cubic kernel also works very well, giving an accuracy of 82.86%, which is still in the same range as the quadratic kernel. The worst classifier is the SVM with polynomial kernel. In most works accuracy is applied it is very poor with an average of 46.56%. For a medium Gaussian kernel we have that the papers that apply it have a performance of 76.33%.

As the most used kernel in literature, the radial obtains average results of 75.34%. This result achieves an acceptable performance, because other estimators such as the ROC curve or the sensitivity and specificity values are very high, in many cases approaching 1 (maximum achievable level). This makes us think that it is a more robust classification than the rest, and it is normal that it is used frequently. In addition, it should be noted that these classifiers present values higher than 90% at the prediction level, only comparable with the performance of the different topologies and configurations of ANNs [41]. Finally, when a feature reduction analysis (PCA) is applied to the previous approach, the average result of the classification is 82.24%.

4.2.2. Auto-Hidden Markov Models

There are two types of algorithms used within the Markov chains. The first one, the auto-hidden Markov chains have an associated result of 88.6% with a PCA and non-PCA approach. On the other hand, if we use a more classic approach, the standard Markov chains, we have a value of 68.7% taking into account the baseline, while for an approach where the baseline is not considered, the accuracy increases to 79.83%.

4.2.3. Discriminant Analysis

In the case of discriminant analysis, the mean value of the prediction for the linear and quadratic configurations is 73.54% and 64.81%, respectively. As can be seen, a higher order configuration worsens the results. On the contrary, if we apply a feature reduction algorithm to the linear discriminant, the results obtained are 71.09%. On the other hand, for a Gaussian type configuration, the results obtained

Table 6. Supervised Learning Results

Type	Config	Papers	Performance	Annotations
SVM	Linear	[65] [71] [24] [72]	55.64(16.88) ¹	
SVM	Quadratic	[35] [75] [65]	85.26(4.10) ¹	
SVM	Cubic	[75] [65] [71]	82.86(1.40) ¹	
SVM	Polynomial	[72] [13]	46.56(14.19) ¹	
SVM	Medium-Gaussian	[65] [71]	76.33(17.95) ¹	
SVM	Radial (RBF)	[72] [41] [53] [64] [63] [19] [36] [58] [49] [40] [28] [44] [78] [60] [47] [34] [38] [17] [35] [33] [67] [19] [27] [55] [39] [69] [74] [97] [26]	75.34(2,53) ¹	
SVM	Radial + PCA	[75]	82,4(0) ¹	
Markov-Model	Auto-Hidden	[15]	88,6(0) ¹	With LDA + Acc.
Markov-Model	Auto-Hidden	[15]	86,6(0) ¹	Without LDA +Acc.
Markov-Model	Markov-Chain	[16]	68,74(7.85) ¹	With Baseline +Acc.
Markov-Model	Markov-Chain	[16]	79,83(5,67) ¹	Without Baseline +Acc.
Discriminant	LDA	[20] [53]	73.54(10.66) ¹	Standard Config.
Discriminant	PCA + LDA	[53]	71,09(0) ¹	PCA analysis
Discriminant	QDA	[68] [13] [17] [44] [71]	64,81(12,43) ¹	
Discriminant	GDA	[65]	71(0) ¹	Gaussian DA
Tree	Regression	[64]	73.30(2.99) ²	Matlab app + ML Met.
Tree	Medium	[18] [68] [27] [13] [55]	76.35(6.02) ²	Matlab app + ML Met.
Tree	Bagged	[75] [18]	78.35(7.56) ²	Matlab app + ML Met.
Tree	Random Forest	[69] [74] [84] [78] [47] [63] [64]	77.85(10.21) ²	Matlab app + ML Met.
Naïve-Bayes	Gaussian	[44] [34] [27] [18] [64]	49.46(14.16) ¹	Standard configuration
Naïve-Bayes	Gaussian	[64] [18]	70.8(0.53) ¹	PCA analysis
Logistic Reg.	Standard	[19] [28] [13] [63]	75,01(11,36) ²	Standard Configuration
Logistic Reg.	ZeroR	[28]	57,54(0) ²	Zero Regression
KNN	Fuzzy	[75]	78,65(11,24) ²	Matlab app + ML Met.
KNN	Fine	[75]	82,8(0) ²	Matlab app + ML Met.
KNN	cubic	[75] [20]	76,85(0,21) ²	ML Met.
KNN	Medium	[16] [41] [53] [19] [13] [27] [55] [40] [60] [34]	79,04(8,82) ²	ML Met.
KNN	Weighted	[63]	76,53(8,64) ²	ML Met.
ANN	1D-CNN	[41] [20]	85,57(8.91) ³	Convolutional-NN
ANN	1D-CNN-E	[41]	92.2(1.36) ³	Convolutional ensemble
ANN	Adaboost	[41]	99,69(0) ¹	Adaboost Method
ANN	MLP	[68] [72] [13] [38]	81.50(5.46) ²	Feed-Forward Propagation
ANN	3-NN	[41]	95.02(6.34) ²	
ANN	5-NN	[41]	98.81(0) ²	
ANN	MLP-BP	[25] [63] [64]	79.13(9.04) ¹	Back-Propagation
ANN	LUCCK	[57]	89.23(0) ²	Concave and Convex Kernel
CNN	MLP-BP	[34] [69] [17] [35]	73.13(9.59) ²	NN based.
CNN	MLP-NF	[67] [78]	80.5(5.38) ²	Neuro-fuzzy inference
BPNN	Bayes	[34]	89.38(0) ³	Cellular Neural-Network
PNN	Probabilistic	[64]	71,31(0) ³	Probabilistic NN
LSTM	CNN + LSTM	[62]	79.13(2.47) ³	Ensemble CNN + LSTM
LSTM	LSTM	[62]	81.4(0) ³	Sklearn Standard Config.
ANN	ANFIS	[53]	95(0) ²	ANFIS-Based Short-Term

Note: 1 = accuracy; 2 = precision; 3 = F1-score.

are 71%. These results suggest that the only method that can be used with acceptable results is the configuration of the linear discriminant. This is largely due to the internal workings of the classifier, as well as its ability to remove those features that do not provide relevant information. In papers where feature removal is performed, such as in the case of LDA, something similar occurs as will be seen below.

4.2.4. Decision Trees

Other classifiers commonly used for EDA (and other bio-signals) processing are decision trees. This type of classifiers works by creating smaller and smaller sub-parts to classify an element into a particular class. This method is very effective when we have a model with multiple features. In most of the articles selected in the scoping review, the implementation of this type of algorithm has used the Matlab library called "*App learner*" with standard configurations (Gini criterion). The average results obtained for the regression trees are 73.30%. Another configuration used in this type of tree is the "medium" configuration, which has given a result of 70.8%. If we now look at the ensemble methods, the results for random forest are 77.85%. On the other hand, if we use the "bagged trees" ensemble method, the precision of the classifier is 78.35%. This configuration is widely used for classification as it gives better results than many other ensemble methods.

4.2.5. Naive Bayes

As for the Bayes classifier, the results obtained of 49.46% for the Gaussian configuration are quite poor. This is because this type of classifier assumes that there is independence in the variables (which is not the case for EDA signals). You can see how the results improve considerably up to 70.8% when this method is combined with a PCA.

4.2.6. Logistic Regression

The average precision for the standard configuration are 75.01% for the classifier based on logistic regression, while they are 57.54% for a zero-regression configuration. It seems that if you change the configuration to one that seems better, the results do not improve. In this case they even get noticeably worse. This type of classifier is not widely used with biological signals, compared to others found in this study; so the results are in line with expectations.

4.2.7. K-Nearest Neighbours

The KNN is one of the most widely used classifiers for physiological classification (also widely used for EDA). This type of classifier performs a grouping of data without making a model first. There are many configurations that can be used in this classifier. The most widely used according to the literature reviewed are the *KNN-Medium*. This type of configuration uses a not very large cluster size. Having a size like this makes it more immune to noise produced by outliers. In our scoping review study, these classifiers have an average performance of 79.04%. Apart from this, the approaches with *KNN-Weighted* show a precision of 76.53%. On the other hand, for the cubic configuration the classifier has almost the same level of prediction as the weighted configuration with an average precision of 76.85%. *KNN-Fuzzy*, which uses fuzzy logic to operate, yields a precision of 78.65%. Finally, for the configuration that uses a smaller cluster size, *KNN-Fine*, although it is more sensitive to noise, we have very good results of 82.8%.

4.2.8. Artificial Neural Networks

From the point of view of the use of artificial neural networks for the detection of stress through EDA, it must be recognised that there are many configurations, topologies and types. Therefore, on many occasions a configuration that works fine in certain cases simply obtains normal results (unremarkable) in other studies. The studied authors do agree that ANNs provide very good results in many different configurations. Thus, a convolutional analysis with a 1-D-CNN configuration obtains a F1-score of 85.57%, and a 1-D-CNN-E network assembly achieves a F1-score of 92.2%.

For ANNs there are many configurations in the literature, the results of which are as follows. Adaboost achieves an accuracy of 99.69%. The 3 and 5-layers configurations provide a precision of 95.02% and 98.81%, respectively. For multi-layer perceptron in the feed-forward configuration, the accuracy drops to 81.50%. Similarly, for the same perceptron, but with the back-propagation

algorithm, the classification accuracy is 79.13%. As for the implementation of new methods in ANNs, there is a configuration that uses the LUCCK method (concave and convex kernel) with a result of 89.23%, in line with those obtained previously. Also the use of cellular networks has brought results of 89.38%. However, for the rest of the configurations, the results do not improve with respect to the previous ones, being a precision of 73.13% and an accuracy of 80.5% on CNN (MLP-BP) and CNN-neuro-fuzzy (MLP-NF), respectively.

4.2.9. Long Short-Term Memory and Recurrent Neural Networks Results.

Another great field that has opened up in signal processing, detection and classification is networks that have memory. As is the case of ANN and its variant the LSTM. From the point of view of the different configurations, this type of network may be used alone or also by assembly method in other configurations. For a LSTM network the F1-score is 81.4%, while with the CNN+LSTM obtains a F1-score of 79.13% and the ANFIS configuration variant gets 95%. Although there is little literature on this type of classifier, it should be regarded as a good alternative when it is desired to use a dataset in the time domain based on the processed skin conductance response (SCR).

4.3. Unsupervised Learning

In terms of unsupervised learning there is very little literature on activation detection or analysis of emotional states (see Table 7). Below are the most used methods detected throughout this review and their most important results. We believe that the lack of development of these techniques is not due to the fact that they are more difficult to implement than the supervised learning methods mentioned above. The reason is that most experiments are directed at data collection and labelling, so it is much easier with an already labelled dataset. It should be noted that since there are not many studies of this type, they cannot be compared with others and we have to take these results with caution.

Table 7. Unsupervised Learning Method Results.

Type	Config.	Papers	Precision	Annotations
K-means	Standard	[40]	77.5(2.12) ²	Standard configuration
K-medoids	Standard	[40]	75.5(2.12) ²	Standard configuration
SOM	Standard	[40] [44]	77.5(0.5) ²	Bi-dimensional Map

One of the unsupervised learning algorithms used is K-means. This algorithm has a precision of 77.5%. In this case, as an advantage we can point out that it is very easy to implement and usually performs well with large data sets. But on the contrary, it is very sensitive to the number of clusters created. In many cases, you have to look for this number of clusters by hand and then let the system learn. It is quite sensitive to noise, so outliers can be a determining factor in achieving a successful detection. In order to minimise the effects of noise produced in outlier data on a dataset, the K-medoids approach has been tested. The result of 75.5% precision is at the same level as those obtained for K-means. Finally, as an alternative method to the previous methods, within the unsupervised learning techniques, there are the methods based on Self-Organising Maps. These variants of ANNs are widely used because they allow us to generate bi-dimensional maps like in convolutional-ANN analysis for later classification. In this case, the results obtained for this classifier are at the same level as the previous ones (77.5%).

5. Conclusions

This paper has presented a scoping review of the use of physiological signals for stress detection, focusing on the use of electrodermal activity (EDA) and various machine learning (ML) techniques. An initial number of 395 documents has been considered, from which 58 have been selected to carry out in depth the scoping review process. These articles have allowed to provide a global perspective

on a specific topic such as the use of EDA, individually or in conjunction with other variables, for the detection of stress (variation of arousal) using ML techniques.

Several important issues have emerged during this study that should be considered by researchers interested in signal acquisition in general and EDA processing in particular. The first point to note is that the classification process must be kept in mind from the moment the signals are obtained (acquisition process). Without a robust acquisition process the signals become useless for a correct later classification. All operations carried out in later stages will be useless if the first step of the whole process is not carried out correctly.

In addition, most of the authors studied in this scoping review emphasised that this process is not exempt from dealing with interference, artefacts or noise in the signals. A correct application of the different filters during the preprocessing stage will be crucial for the following phases. When working with EDA signals, all the articles surveyed highlighted that the signals must go through a deconvolution process to homogenise and normalise them. The standardisation process allows to use a dataset that has a large amount of data coming from people without being affected by their race, sex or age. In fact, those studies in which there was no deconvolution process have been discarded, due to the poor results obtained by any classifier.

Once the signals have been preprocessed, the next very important step is to obtain the different features that allow us to quantify them. This will be done in a further process which is the classification. At this point, most authors agree to use features from different domains, generally the *temporal domain* and the *frequency domain*. There are also approximations that analyse the shape of the signal (*morphological*) and those that analyse the signal statistically (*statistical features*). Nobody agrees on the number of variables or the minimum number of functions that should be used. The general approach is to use different types and fit the model using LDA, PCA, or ICA analysis to perform dimensionality reduction.

Furthermore, during the review, two well differentiated pathways for the estimation of the participant's emotional state have been found. A first approach that aims to use only the EDA for detection, while, the second approach bets for use EDA signal mixing with other physiological signals such as BVP, ECG, EMG, among others. Both methods have their advantages and disadvantages. One advantage of using EDA individually is the possibility to use very small non-invasive devices with great autonomy. Another advantage is that the results using only EDA are quite good. A disadvantage found is that the use of this variable alone limits us to the problem of stress (arousal variation) detection. This is because the response is limited by the parasympathetic nervous system (sudomotor nervous system) response. On the contrary, using more physiological signals has the advantage that it enables to monitor several systems and, as a consequence, several types of responses provide a better mapping of the physical, psychological and cognitive state of a subject. Nonetheless, a disadvantage is that the use of different signals makes the system more complex, more difficult to maintain and it has a higher computational cost when using a classifier.

If we now consider classification methods based on ML techniques, it is clear that there are two main groups of methods, those based on supervised and unsupervised learning. Throughout the literature, most authors are committed to the use of techniques based on supervised learning. This is largely due to the fact that the experiments and datasets are labelled for each of the states. For this reason, few articles use unsupervised techniques.

Among the supervised learning methods, SVMs and many of the ANN topologies report better classification results than the rest, followed closely by KNN algorithms. This is because in most cases these algorithms work very well when facing many features of the signals (EDA and others). For SVMs those implementing quadratic, cubic and radial kernels outperform with accuracy of 85.26%, 82.86% and 82.4%, respectively. ANNs, on the other hand, highlight for their accuracy in different configurations, especially ANN-Adaboost with 99%, and different configurations of the MLP with 95% and 98% for the 3-layer and 5-layer, respectively. Moreover, from the perspective of unsupervised

learning, the different results are at a similar level. K-means, K-medoids and SOMs are acceptable classifiers showing around 75% accuracy.

As a final conclusion we can say that the individual use of EDA for the detection of the excitation variation is very widespread, having achieved very good results in its prediction. Furthermore, its use in combination with other physiological signals and with the help of robust and novel ML techniques has increased over time. For this reason, emotion detection is being integrated in a non-invasive manner into user-centred devices, and at the same time, the robustness and precision of to date's systems and applications has been complemented. This use allows us to continuously monitor the emotional state of the subject in a continuous manner. Although it is not a perfect detection model, combined with other variables, it is possible to work towards the design of increasingly useful and interesting applications in health and education. Specifically, stress detection will be gaining an important place in manifold areas.

Author Contributions: Conceptualisation, A.F.-C.; methodology, R.S.-R.; validation, A.F.-C. and M.T.L.; writing—original draft preparation, R.S.-R.; writing—review and editing, A.F.-C. and M.T.L.; funding acquisition, A.F.-C.

Acknowledgments: This work has been partially supported by Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) / European Regional Development Fund (FEDER, UE) under EQC2019-006063-P and DPI2016-80894-R grants, and by CIBERSAM of the Instituto de Salud Carlos III. Roberto Sánchez-Reolid holds BES-2017-081958 scholarship from Spanish Ministerio de Educación y Formación Profesional.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	Accuracy
AHMM	Auto-hidden Markov Model
AI	Artificial Intelligence
AKAIKE	AKAIKE Information Criterion
AL	arc length
ANNs	Artificial Neural Networks
AP	normalised mean power
AR-HMM	Auto-Regressive Hidden-Markov Model
AUC	Area Under Curve
BioVid	BioVid Heat Pain Database
BPNN	Bayesian Probabilistic Neural Network
BVP	Blood Volume Pressure
CMax	Cumulative Maximum
CMin	Cumulative Minimum
CNN	Convolutional Neural Network
COVMAT	Covariance Matrix
D1	First Derivative
D1M	First Derivative Mean
D1SD	First Derivative Standard Deviation
D2	Second Derivative
D2M	Second Derivative Mean
D2SD	Second Derivative Standard Deviation
DA	Discriminant Analysis
DCRM	Decay Rate Mean
DCRSD	Decay Rate Standard Deviation
DEAP	Database for Emotion Analysis using Physiological Signal
DR	Dynamic Range
DTs	Decision Trees
EC	Epoch-Capacity
ECG	Electrocardiogram
EDA	Electrodermal Activity
EDL	electrodermal Level
EDL	Electrodermal Level

EDR electrodermal Response
EEG Electroencephalography
EL Energy to Perimeter Ratio
EMG Electromyography
EN Entropy
EOG Electroculography
EP Epoch-Peak
EPC Epoch Peak Counter
FCM Four Central Moment
FFT Fast Fourier Transform
FVCM Five Central Moment
GDA Gaussian Discriminant Analysis
IBR Inter-Breath
IL Perimeter to Area Ratio
IN Integral Area
IRF Impulse Response Function
KNN K-nearest Neighbours
KU Kurtosis
LDA Linear Discriminant Analysis
LOG-LIKE Log-likelihood

LR	Logistic Regression
LSTM	Long Short-Term Memory
MAHNOB	Multi-modal Data base for Affect Recognition
MAt	Motion Artefact
Mean	Mean
Median-Val	Median Value
MedVal	Median Value
ML	Machine Learning
MLP	Multilayer Perceptron
MLT	Machine Learning Techniques
MO	Momentum
MSSP	Mean Spectral Components
NO	Number of Observation
NSSCRs	Frequency Non-Specific of Skin Conductance Response
P	Peak
p-Val	p-Value
PHVM	Phasic Value Mean
PHVSD	Phasic Value Standard Deviation
PLoc	Peak Location
PMRMSR	Peak-Magnitude-to-RMS Ratio
pNN50	Peaks Intervals Differs 50ms
PPT	Peak to Peak Time
PUP	Pupilometry
QDA	Quadratic Discriminant Analysis
RM	Rise Rate Mean
RMS	Root-mean Square Level
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
RRSTD	Rise Rate Standard Deviation
RSSL	Root Sum of Squares Level
SC	Skin Conductance
SCL	Skin conductance Level
SCR	Skin Conductance Response
SD	Standard Deviation
SFT	Sum Fall Time
SKU	Skewness
SOM	Self-Organising Maps
SP	Spectral Power
SRT	Sum Rise Time
SSP	Sum Spectral Power
SSPMed	Median Spectral Power Components
STM	Startle Time Mean
STRMS	Startle Time Mean
STRMSOV	Startle RMS Overall
STRMSSD	Startle RMS Standard Deviation
STSD	Startle Time Standard Deviation (
SVM	Support Vector Machine
SWE	Smallest Window Elements
TMP	Temperature
Var	Variance

References

1. Mingardi, M.; Pluchino, P.; Bacchin, D.; Rossato, C.; Gamberini, L. Assessment of Implicit and Explicit Measures of Mental Workload in Working Situations: Implications for Industry 4.0. *Applied Sciences* **2020**, *10*, 6416.
2. Castillo, J.C.; Castro-González, a.; Fernández-Caballero, A.; Latorre, J.M.; Pastor, J.M.; Fernández-Sotos, A.; Salichs, M.A. Software architecture for smart emotion recognition and regulation of the ageing adult. *Cognitive Computation* **2016**, *8*, 357–367.
3. Castillo, J.C.; Fernández-Caballero, A.; Castro-González, Á.; Salichs, M.A.; López, M.T. A Framework for Recognizing and Regulating Emotions in the Elderly. *Ambient Assisted Living and Daily Activities*; Pecchia, L.; Chen, L.L.; Nugent, C.; Bravo, J., Eds.; Springer International Publishing: Cham, 2014; pp. 320–327.
4. Le Fevre, M.; Matheny, J.; Kolt, G.S. Eustress, distress, and interpretation in occupational stress. *Journal of Managerial Psychology* **2003**, *18*, 726–744.
5. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology* **2005**, *17*, 715.
6. Said, S.; Karar, A.S.; Beyrouthy, T.; Alkork, S.; Nait-ali, A. Biometrics Verification Modality Using Multi-Channel sEMG Wearable Bracelet. *Applied Sciences* **2020**, *10*, 6960.
7. Vrtana, D.; Krizanova, A.; Skorvagova, E.; Valaskova, K. Exploring the Affective Level in Adolescents in Relation to Advertising with a Selected Emotional Appeal. *Sustainability* **2020**, *12*, 8287.
8. Alfaras, M.; Primett, W.; Umair, M.; Windlin, C.; Karpashevich, P.; Chalabianloo, N.; Bowie, D.; Sas, C.; Sanches, P.; Höök, K.; others. Biosensing and Actuation—Platforms Coupling Body Input-Output Modalities for Affective Technologies. *Sensors* **2020**, *20*, 5968.
9. Thammasan, N.; Stuldreher, I.V.; Schreuders, E.; Giletta, M.; Brouwer, A.M. A Usability Study of Physiological Measurement in School Using Wearable Sensors. *Sensors* **2020**, *20*, 5380.
10. Arksey, H.; O'Malley, L. Scoping studies: towards a methodological framework. *International journal of social research methodology* **2005**, *8*, 19–32.
11. Peters, M.D.; Godfrey, C.M.; Khalil, H.; McInerney, P.; Parker, D.; Soares, C.B. Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare* **2015**, *13*, 141–146.
12. Posada-Quintero, H.F.; Dimitrov, T.; Moutran, A.; Park, S.; Chon, K.H. Analysis of Reproducibility of Noninvasive Measures of Sympathetic Autonomic Control Based on Electrodermal Activity and Heart Rate Variability. *IEEE Access* **2019**, *7*, 22523–22531. doi:10.1109/access.2019.2899485.
13. Can, Y.S.; Chalabianloo, N.; Ekiz, D.; Ersoy, C. Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. *Sensors* **2019**, *19*, 1849. doi:10.3390/s19081849.
14. Wormwood, J.B.; Khan, Z.; Siegel, E.; Lynn, S.K.; Dy, J.; Barrett, L.F.; Quigley, K.S. Physiological indices of challenge and threat: A data-driven investigation of autonomic nervous system reactivity during an active coping stressor task. *Psychophysiology* **2019**, *56*. doi:10.1111/psyp.13454.
15. Akbulut, F.P.; Perros, H.G.; Shahzad, M. Bimodal affect recognition based on autoregressive hidden Markov models from physiological signals. *Computer Methods and Programs in Biomedicine* **2020**, *195*, 105571. doi:10.1016/j.cmpb.2020.105571.
16. Liu, Y.; Jiang, C. Recognition of Shooter's Emotions Under Stress Based on Affective Computing. *IEEE Access* **2019**, *7*, 62338–62343. doi:10.1109/access.2019.2916147.
17. Ganapathy, N.; Veeranki, Y.R.; Swaminathan, R. Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features. *Expert Systems with Applications* **2020**, *159*, 113571. doi:10.1016/j.eswa.2020.113571.
18. Sadeghi, R.; Banerjee, T.; Hughes, J.C.; Lawhorne, L.W. Sleep quality prediction in caregivers using physiological signals. *Computers in Biology and Medicine* **2019**, *110*, 276–288. doi:10.1016/j.combiomed.2019.05.010.
19. Rizwan, A.; Ali, N.A.; Zoha, A.; Ozturk, M.; Alomainy, A.; Imran, M.A.; Abbasi, Q.H. Non-Invasive Hydration Level Estimation in Human Body Using Galvanic Skin Response. *IEEE Sensors Journal* **2020**, *20*, 4891–4900. doi:10.1109/jsen.2020.2965892.

20. Ding, Y.; Cao, Y.; Duffy, V.G.; Wang, Y.; Zhang, X. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. *Ergonomics* **2020**, *63*, 896–908. doi:10.1080/00140139.2020.1759699.
21. Alcañiz-Raya, M.; Chicchi-Giglioli, I.A.; Marín-Morales, J.; Higuera-Trujillo, J.L.; Olmos, E.; Minissi, M.E.; Teruel-Garcia, G.; Sirera, M.; Abad, L. Application of Supervised Machine Learning for Behavioral Biomarkers of Autism Spectrum Disorder Based on Electrodermal Activity and Virtual Reality. *Frontiers in Human Neuroscience* **2020**, *14*. doi:10.3389/fnhum.2020.00090.
22. Posada-Quintero, H.F.; Reljin, N.; Moutran, A.; Georgopolis, D.; Lee, E.C.H.; Giersch, G.E.W.; Casa, D.J.; Chon, K.H. Mild Dehydration Identification Using Machine Learning to Assess Autonomic Responses to Cognitive Stress. *Nutrients* **2019**, *12*, 42. doi:10.3390/nu12010042.
23. Regalia, G.; Onorati, F.; Lai, M.; Caborni, C.; Picard, R.W. Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands. *Epilepsy Research* **2019**, *153*, 79–82. doi:10.1016/j.eplepsyres.2019.02.007.
24. Susam, B.T.; Akcakaya, M.; Nezamfar, H.; Diaz, D.; Xu, X.; de Sa, V.R.; Craig, K.D.; Huang, J.S.; Goodwin, M.S. Automated Pain Assessment using Electrodermal Activity Data and Machine Learning. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018. doi:10.1109/embc.2018.8512389.
25. Thiam, P.; Bellmann, P.; Kestler, H.A.; Schwenker, F. Exploring Deep Physiological Models for Nociceptive Pain Recognition. *Sensors* **2019**, *19*, 4503. doi:10.3390/s19204503.
26. Posada-Quintero, H.F.; Bolkhovsky, J.B. Machine learning models for the identification of cognitive tasks using autonomic reactions from heart rate variability and electrodermal activity. *Behavioral Sciences* **2019**, *9*, 45.
27. Machot, F.A.; Ali, M.; Ranasinghe, S.; Mosa, A.H.; Kyandoghere, K. Improving Subject-independent Human Emotion Recognition Using Electrodermal Activity Sensors for Active and Assisted Living. Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference. ACM, 2018. doi:10.1145/3197768.3201523.
28. García-Faura, Á.; Hernández-García, A.; Fernández-Martínez, F.; de María, F.D.; San-Segundo, R. Emotion and attention: Audiovisual models for group-level skin response recognition in short movies. *Web Intelligence* **2019**, *17*, 29–40. doi:10.3233/web-190398.
29. Sánchez-Reolid, R.; Martínez-Rodrigo, A.; López, M.T.; Fernández-Caballero, A. Deep Support Vector Machines for the Identification of Stress Condition from Electrodermal Activity. *International Journal of Neural Systems* **2020**, p. 2050031.
30. Greco, A.; Valenza, G.; Lanata, A.; Rota, G.; Scilingo, E.P. Electrodermal Activity in Bipolar Patients during Affective Elicitation. *IEEE Journal of Biomedical and Health Informatics* **2014**, *18*, 1865–1873. doi:10.1109/jbhi.2014.2300940.
31. Greco, A.; Valenza, G.; Lanata, A.; Scilingo, E.; Citi, L. cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering* **2016**, pp. 1–1. doi:10.1109/tbme.2015.2474131.
32. Greco, A.; Valenza, G.; Nardelli, M.; Bianchi, M.; Citi, L.; Scilingo, E.P. Force-Velocity Assessment of Caress-Like Stimuli Through the Electrodermal Activity Processing: Advantages of a Convex Optimization Approach. *IEEE Transactions on Human-Machine Systems* **2016**, pp. 1–10. doi:10.1109/thms.2016.2586478.
33. Greco, A.; Valenza, G.; Citi, L.; Scilingo, E.P. Arousal and Valence Recognition of Affective Sounds Based on Electrodermal Activity. *IEEE Sensors Journal* **2017**, *17*, 716–725. doi:10.1109/jsen.2016.2623677.
34. Ali, M.; Machot, F.; Mosa, A.; Jdeed, M.; Machot, E.; Kyamakya, K. A Globally Generalized Emotion Recognition System Involving Different Physiological Signals. *Sensors* **2018**, *18*, 1905. doi:10.3390/s18061905.
35. Ghaderyan, P.; Abbasi, A. An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations. *International Journal of Psychophysiology* **2016**, *110*, 91–101. doi:10.1016/j.ijpsycho.2016.10.013.
36. Taylor, S.; Jaques, N.; Chen, W.; Fedor, S.; Sano, A.; Picard, R. Automatic identification of artifacts in electrodermal activity data. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015. doi:10.1109/embc.2015.7318762.

37. Gjoreski, M.; Gams, M.Z.; Lustrek, M.; Genc, P.; Garbas, J.U.; Hassan, T. Machine Learning and End-to-End Deep Learning for Monitoring Driver Distractions From Physiological and Visual Signals. *IEEE Access* **2020**, *8*, 70590–70603. doi:10.1109/access.2020.2986810.
38. Zontone, P.; Affanni, A.; Bernardini, R.; Piras, A.; RINALDO, R.; Formaggia, F.; Minen, D.; Minen, M.; Savorgnan, C. Car Driver's Sympathetic Reaction Detection through Electrodermal Activity and Electrocardiogram Measurements. *IEEE Transactions on Biomedical Engineering* **2020**, pp. 1–1. doi:10.1109/tbme.2020.2987168.
39. Siddharth, S.; Trivedi, M.M. On Assessing Driver Awareness of Situational Criticalities: Multi-modal Bio-Sensing and Vision-Based Analysis, Evaluations, and Insights. *Brain Sciences* **2020**, *10*, 46. doi:10.3390/brainsci10010046.
40. Fiorini, L.; Mancioffi, G.; Semeraro, F.; Fujita, H.; Cavallo, F. Unsupervised emotional state classification through physiological parameters for social robotics applications. *Knowledge-Based Systems* **2020**, *190*, 105217. doi:10.1016/j.knosys.2019.105217.
41. Bianco, S.; Napoletano, P. Biometric Recognition Using Multimodal Physiological Signals. *IEEE Access* **2019**, *7*, 83581–83588. doi:10.1109/access.2019.2923856.
42. Hernando-Gallego, F.; Luengo, D.; Artes-Rodriguez, A. Feature Extraction of Galvanic Skin Responses by Nonnegative Sparse Deconvolution. *IEEE Journal of Biomedical and Health Informatics* **2018**, *22*, 1385–1394. doi:10.1109/jbhi.2017.2780252.
43. Hwang, S.H.; Park, K.S.; Seo, S.; Yoon, H.N.; Jung, D.W.; Baek, H.J.; Cho, J.; Choi, J.W.; Lee, Y.J.; Jeong, D.U. Sleep Period Time Estimation Based on Electrodermal Activity. *IEEE Journal of Biomedical and Health Informatics* **2017**, *21*, 115–122. doi:10.1109/jbhi.2015.2490480.
44. Jang, E.H.; Park, B.J.; Park, M.S.; Kim, S.H.; Sohn, J.H. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *Journal of Physiological Anthropology* **2015**, *34*. doi:10.1186/s40101-015-0063-5.
45. Kelsey, M.; Akcakaya, M.; Kleckner, I.R.; Palumbo, R.V.; Barrett, L.F.; Quigley, K.S.; Goodwin, M.S. Applications of sparse recovery and dictionary learning to enhance analysis of ambulatory electrodermal activity data. *Biomedical Signal Processing and Control* **2018**, *40*, 58–70. doi:10.1016/j.bspc.2017.08.024.
46. Khalaf, A.; Nabian, M.; Fan, M.; Yin, Y.; Wormwood, J.; Siegel, E.; Quigley, K.S.; Barrett, L.F.; Akcakaya, M.; Chou, C.A.; Ostadabbas, S. Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat. *Expert Systems with Applications* **2020**, *140*, 112890. doi:10.1016/j.eswa.2019.112890.
47. Momin, A.; Sanyal, S. Analysis of Electrodermal Activity Signal Collected During Visual Attention Oriented Tasks. *IEEE Access* **2019**, *7*, 88186–88195. doi:10.1109/access.2019.2925933.
48. Cho, D.; Ham, J.; Oh, J.; Park, J.; Kim, S.; Lee, N.K.; Lee, B. Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine. *Sensors* **2017**, *17*, 2435. doi:10.3390/s17102435.
49. Liu, C.; Conn, K.; Sarkar, N.; Stone, W. Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *International Journal of Human-Computer Studies* **2008**, *66*, 662–677. doi:10.1016/j.ijhsc.2008.04.003.
50. Hadi, W.; El-Khalili, N.; AlNashashibi, M.; Issa, G.; AlBanna, A.A. Application of data mining algorithms for improving stress prediction of automobile drivers: A case study in Jordan. *Computers in Biology and Medicine* **2019**, *114*, 103474. doi:10.1016/j.combiomed.2019.103474.
51. Harley, J.M.; Bouchet, F.; Hussain, M.S.; Azevedo, R.; Calvo, R. A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior* **2015**, *48*, 615–625. doi:10.1016/j.chb.2015.02.013.
52. Kukolja, D.; Popović, S.; Horvat, M.; Kovač, B.; Čosić, K. Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications. *International Journal of Human-Computer Studies* **2014**, *72*, 717–727. doi:10.1016/j.ijhcs.2014.05.006.
53. S., A.A.; P., S.; V., S.; S., S.K.; A., S.; Akl, T.J.; P., P.S.; Sivaprakasam, M. Electrodermal Activity Based Pre-surgery Stress Detection Using a Wrist Wearable. *IEEE Journal of Biomedical and Health Informatics* **2020**, *24*, 92–100. doi:10.1109/jbhi.2019.2893222.

54. Martinez, R.; Salazar-Ramirez, A.; Arruti, A.; Irigoyen, E.; Martin, J.I.; Muguerza, J. A Self-Paced Relaxation Response Detection System Based on Galvanic Skin Response Analysis. *IEEE Access* **2019**, *7*, 43730–43741. doi:10.1109/access.2019.2908445.
55. Anderson, A.; Hsiao, T.; Metsis, V. Classification of Emotional Arousal During Multimedia Exposure. Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2017. doi:10.1145/3056540.3064956.
56. Alhamid, M.F.; Rawashdeh, M.; Osman, H.A.; Saddik, A.E. Leveraging biosignal and collaborative filtering for context-aware recommendation. Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare - MIIRH '13. ACM Press, 2013. doi:10.1145/2505323.2505332.
57. Singh, R.R.; Conjeti, S.; Banerjee, R. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control* **2013**, *8*, 740–754. doi:10.1016/j.bspc.2013.06.014.
58. Pinto, J.; Fred, A.; da Silva, H.P. Biosignal-Based Multimodal Emotion Recognition in a Valence-Arousal Affective Framework Applied to Immersive Video Visualization. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019. doi:10.1109/embc.2019.8857852.
59. Betti, S.; Lova, R.M.; Rovini, E.; Acerbi, G.; Santarelli, L.; Cabiati, M.; Ry, S.D.; Cavallo, F. Evaluation of an Integrated System of Wearable Physiological Sensors for Stress Monitoring in Working Environments by Using Biological Markers. *IEEE Transactions on Biomedical Engineering* **2018**, *65*, 1748–1758. doi:10.1109/tbme.2017.2764507.
60. Khezri, M.; Firoozabadi, M.; Sharafat, A.R. Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. *Computer Methods and Programs in Biomedicine* **2015**, *122*, 149–164. doi:10.1016/j.cmpb.2015.07.006.
61. Maldonado, S.; López, J.; Jimenez-Molina, A.; Lira, H. Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment. *Expert Systems with Applications* **2020**, *143*, 112988. doi:10.1016/j.eswa.2019.112988.
62. Rastgoo, M.N.; Nakisa, B.; Maire, F.; Rakotonirainy, A.; Chandran, V. Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications* **2019**, *138*, 112793. doi:10.1016/j.eswa.2019.07.010.
63. Zhang, Y.; Haghdan, M.; Xu, K.S. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. Proceedings of the 2017 ACM International Symposium on Wearable Computers - ISWC '17. ACM Press, 2017. doi:10.1145/3123021.3123054.
64. Zhao, B.; Wang, Z.; Yu, Z.; Guo, B. EmotionSense: Emotion Recognition Based on Wearable Wristband. 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI). IEEE, 2018. doi:10.1109/smartworld.2018.00091.
65. Jebelli, H.; Choi, B.; Lee, S. Application of Wearable Biosensors to Construction Sites. II: Assessing Workers' Physical Demand. *Journal of Construction Engineering and Management* **2019**, *145*, 04019080. doi:10.1061/(asce)co.1943-7862.0001710.
66. Kleckner, I.R.; Jones, R.M.; Wilder-Smith, O.; Wormwood, J.B.; Akcakaya, M.; Quigley, K.S.; Lord, C.; Goodwin, M.S. Simple, Transparent, and Flexible Automated Quality Assessment Procedures for Ambulatory Electrodermal Activity Data. *IEEE Transactions on Biomedical Engineering* **2018**, *65*, 1460–1467. doi:10.1109/tbme.2017.2758643.
67. Katsis, C.; Katertsidis, N.; Ganiatsas, G.; Fotiadis, D. Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **2008**, *38*, 502–512. doi:10.1109/tsmca.2008.918624.
68. Sabeti, E.; Gryak, J.; Derksen, H.; Biwer, C.; Ansari, S.; Isenstein, H.; Kratz, A.; Najarian, K. Learning Using Concave and Convex Kernels: Applications in Predicting Quality of Sleep and Level of Fatigue in Fibromyalgia. *Entropy* **2019**, *21*, 442. doi:10.3390/e21050442.

69. Chowdhury, A.K.; Tjondronegoro, D.; Chandran, V.; Zhang, J.; Trost, S.G. Prediction of Relative Physical Activity Intensity Using Multimodal Sensing of Physiological Data. *Sensors* **2019**, *19*, 4509. doi:10.3390/s19204509.
70. Batista, D.; da Silva, H.P.; Fred, A.; Moreira, C.; Reis, M.; Ferreira, H.A. Benchmarking of the BITalino biomedical toolkit against an established gold standard. *Healthcare Technology Letters* **2019**, *6*, 32–36. doi:10.1049/htl.2018.5037.
71. Setz, C.; Amrich, B.; Schumm, J.; Marca, R.L.; Troster, G.; Ehlert, U. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* **2010**, *14*, 410–417. doi:10.1109/titb.2009.2036164.
72. Machot, F.A.; Elmachot, A.; Ali, M.; Machot, E.A.; Kyamakya, K. A Deep-Learning Model for Subject-Independent Human Emotion Recognition Using Electrodermal Activity Sensors. *Sensors* **2019**, *19*, 1659. doi:10.3390/s19071659.
73. Greco, A.; Marzi, C.; Lanata, A.; Scilingo, E.P.; Vanello, N. Combining Electrodermal Activity and Speech Analysis towards a more Accurate Emotion Recognition System. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019. doi:10.1109/embc.2019.8857745.
74. Kim, A.Y.; Jang, E.H.; Kim, S.; Choi, K.W.; Jeon, H.J.; Yu, H.Y.; Byun, S. Automatic detection of major depressive disorder using electrodermal activity. *Scientific Reports* **2018**, *8*. doi:10.1038/s41598-018-35147-3.
75. Cavallo, F.; Semeraro, F.; Mancioffi, G.; Betti, S.; Fiorini, L. Mood classification through physiological parameters. *Journal of Ambient Intelligence and Humanized Computing* **2019**. doi:10.1007/s12652-019-01595-6.
76. Zhang, Q.; Chen, X.; Zhan, Q.; Yang, T.; Xia, S. Respiration-based emotion recognition with deep learning. *Computers in Industry* **2017**, *92*, 84–90.
77. Alam, M.A.U.; Roy, N.; Holmes, S.; Gangopadhyay, A.; Galik, E. Automated Functional and Behavioral Health Assessment of Older Adults with Dementia. 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, 2016. doi:10.1109/chase.2016.16.
78. Katsis, C.D.; Katertsidis, N.S.; Fotiadis, D.I. An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomedical Signal Processing and Control* **2011**, *6*, 261–268. doi:10.1016/j.bspc.2010.12.001.
79. Critchley, H.D.; Garfinkel, S.N. The influence of physiological signals on cognition. *Current Opinion in Behavioral Sciences* **2018**, *19*, 13–18.
80. Degoulet, P.; Fieschi, M. *Introduction to clinical informatics*; Springer Science & Business Media, 2012.
81. Peng, L.; Hou, Z.; Chen, Y.; Wang, W.; Tong, L.; Li, P. Combined use of sEMG and accelerometer in hand motion classification considering forearm rotation. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013, pp. 4227–4230.
82. Karthick, P.; Ramakrishnan, S. Surface electromyography based muscle fatigue progression analysis using modified B distribution time–frequency features. *Biomedical signal processing and control* **2016**, *26*, 42–51.
83. Karenbach, C. Ledalab—A Software Package for the Analysis of Phasic Electrodermal Activity. Technical report, Institut für Allgemeine Psychologie, 2005.
84. Ganapathy, N.; Swaminathan, R.; Deserno, T. Deep Learning on 1-D Biosignals: a Taxonomy-based Survey. *Yearbook of Medical Informatics* **2018**, *27*, 098–109. doi:10.1055/s-0038-1667083.
85. Lichtenauer, J.; Soleymani, M. Mahnob-Hci-Tagging Database, 2011.
86. Fortin, F.A.; De Rainville, F.M.; Gardner, M.A.; Parizeau, M.; Gagné, C. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research* **2012**, *13*, 2171–2175.
87. Walter, S.; Gruss, S.; Ehleiter, H.; Tan, J.; Traue, H.C.; Werner, P.; Al-Hamadi, A.; Crawcour, S.; Andrade, A.O.; da Silva, G.M. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. 2013 IEEE international conference on cybernetics (CYBCO). IEEE, 2013, pp. 128–131.
88. Birjandtalab, J.; Cogan, D.; Pouyan, M.B.; Nourani, M. A non-EEG biosignals dataset for assessment and visualization of neurological status. 2016 IEEE International Workshop on Signal Processing Systems (SiPS). IEEE, 2016, pp. 110–114.

89. Chen, W.; Jaques, N.; Taylor, S.; Sano, A.; Fedor, S.; Picard, R.W. Wavelet-based motion artifact removal for electrodermal activity. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 6223–6226.
90. Hernandez, J.; Morris, R.R.; Picard, R.W. Call center stress recognition with person-specific models. International Conference on Affective Computing and Intelligent Interaction. Springer, 2011, pp. 125–134.
91. Taylor, S.; Jaques, N.; Chen, W.; Fedor, S.; Sano, A.; Picard, R. Automatic identification of artifacts in electrodermal activity data. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 1934–1937.
92. Amin, M.R.; Faghih, R.T. Sparse Deconvolution of Electrodermal Activity via Continuous-Time System Identification. *IEEE Transactions on Biomedical Engineering* **2019**, *66*, 2585–2595. doi:10.1109/tbme.2019.2892352.
93. Bartolomé-Tomás, A.; Sánchez-Reolid, R.; Latorre, A.F.S.J.M.; Fernández-Caballero, A. Arousal Detection in Elderly People from Electrodermal Activity Using Musical Stimuli. *Sensors* **2020**, *20*, 4788.
94. Sathya, R.; Abraham, A. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence* **2013**, *2*, 34–38.
95. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297.
96. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. Advances in Neural Information Processing Systems. ACM, 1997, pp. 155–161.
97. Setz, C.; Arnrich, B.; Schumm, J.; La Marca, R.; Tröster, G.; Ehlert, U. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine* **2009**, *14*, 410–417.
98. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and regression trees. Belmont, CA: Wadsworth. *International Group* **1984**, *432*, 151–166.
99. Dasarathy, B. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques IEEE Computer Society Press. *Silver Spring, MD* **1991**.
100. Sherstinsky, A. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena* **2020**, *404*, 132306.
101. Flinton, G.; Sejnowski, T. Unsupervised learning and map formation: Foundations of neural computation, 1999.
102. Kohonen, T. The self-organizing map. *Proceedings of the IEEE* **1990**, *78*, 1464–1480.
103. Anusha, A.; Joy, J.; Preejith, S.; Joseph, J.; Sivaprakasam, M. Differential effects of physical and psychological stressors on electrodermal activity. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 4549–4552.