Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2022 doi:10.20944/preprints202203.0403.v1

Article

Behavioral Change Prediction from Physiological Signals Using Deep Learned Features

Giovanni Diraco [1*], Pietro Siciliano [2], and Alessandro Leone [3]

[1] National Research Council of Italy, IMM—Institute for Microelectronics and Microsystems, 73100 Lecce, Italy; giovanni.diraco@cnr.it
[2] National Research Council of Italy, IMM—Institute for Microelectronics and Microsystems, 73100 Lecce, Italy; pietro.siciliano@le.imm.cnr.it
[3] National Research Council of Italy, IMM—Institute for Microelectronics and Microsystems, 73100 Lecce, Italy, alessandro.leone@cnr.it
* Correspondence: giovanni.diraco@cnr.it; Tel.: +39 832 422 531

Abstract: Predicting change from multivariate time series has relevant applications ranging from medical to engineering fields. Multisensory stimulation therapy in patients with dementia aims to change the patient's behavioral state. For example, patients who exhibit a baseline of agitation may be paced to change their behavioral state to relaxed. This study aims to predict changes in behavioral state from the analysis of the physiological and neurovegetative parameters to support the therapist during the stimulation session. In order to extract valuable indicators for predicting changes, both handcrafted and learned features were evaluated and compared. The handcrafted features were defined starting from the CATCH22 feature collection, while the learned ones were extracted using a Temporal Convolutional Network, and the behavioral state was predicted through Bidirectional Long Short-Term Memory Auto-Encoder, operating jointly. From the comparison with the state-of-the-art, the learned features-based approach exhibits superior performance with accuracy rates of up to 99.42% with a time window of 70 seconds and up to 98.44% with a time window of 10 seconds.

Keywords: behavioral change prediction; learned features; deep feature learning; handcrafted features; bidirectional long-short term memory; autoencoders; temporal convolutional neural network; clinical decision support system; multisensory stimulation therapy; physiological signals.

1. Introduction

In recent years, the detection and prediction of changes in time series data obtained from observations of a monitored system has become a relevant research topic in various fields [1, 2, 3]. In particular, change-point detection has attracted considerable interest in medical and neurological fields, where the accurate determination of changes in physiological parameters is particularly critical [4, 5].

Furthermore, change prediction is also important in supporting clinical decisions regarding the delivery of therapeutic interventions, such as in the case of the multisensory stimulation in dementia which was investigated in the MS-Lab (Multi Sensorial Stimulation Lab) project [6]. In MS-Lab, aiming to improve the efficacy of multisensory stimulation, a series of miniaturized non-invasive sensors located on the patient's body were used to measure various neurovegetative parameters in real-time. A Clinical Decision Support System (CDSS) was specially designed to find predictive patterns in multivariate neurovegetative time series (i.e., that anticipate behavioral reactions induced by therapy stimuli), and thus to provide useful hints to the therapist in selecting the most effective stimulation.

As it is well known, behavioral and non-behavioral reactions are induced by endogenous and exogenous stimuli. Behavioral reactions, such as expressing aggressiveness, facial emotions, etc., can be inhibited by voluntary control to some extent. On the other hand, non-behavioral reactions, as neurovegetative manifestations, are not under the influence of the cerebral cortex and thus are very difficult (if not impossible at all) to control voluntarily [7, 8]. Furthermore, neurovegetative responses (i.e., physiological parameters), in terms of ergotropic and trophotropic reactions, can be considered as anticipatory of some behavioral reactions such as activation and relaxation [9, 10, 11].

The involuntary behavior is related to autonomic nervous system (ANS) functions such as sympathetic and parasympathetic activities: stressful or relaxing situations cause dynamic changes in ANS. More specifically, the sympathetic nervous system (SNS) dominates during stressful event, whereas the parasympathetic nervous system (PNS) dominates during resting behavior [12]. These concepts have been exploited in several studies to investigate symptoms of stress, e.g., agitation, anger, fear and frustration, by measuring physiological neurovegetative parameters, since many of them are regulated by SNS and PNS, such as heart rate (HR), heart rate variability (HRV), respiration rate (RR), respiration amplitude (RA), galvanic skin response (GSR), blood pressure (BP), and so on [13, 14]. In particular, various studies have shown [15, 16] that skin temperature (ST) and GSR are indicator of stress level, i.e., high levels of stress are related to low levels of skin temperature due to contraction of blood vessels, and low levels of skin resistance due to an increase of the body moisture.

Consequently, physiological neurovegetative parameters can be investigated as candidate indicators able to anticipate the patient's behavioral state underwent stimulation therapy to support decisions about the choice of the better stimulation to apply [6]. In that context, it is overwhelming important to detect early changes in physiological parameters suitable to predict incipient changes in patient's behavioral state. The problem can be posed in terms of multivariate time series of physiological and neurovegetative parameters, and it involves identifying suitable features to highlight changes.

The watershed between methods for identifying changes in time series is undoubtedly represented by how features are obtained, which can be handcrafted or learned [17]. Most of the features reported in the literature are designed manually, i.e., handcrafted, paying attention to peculiar characteristics of the physiological parameters under consideration. The design of handcrafted features often requires finding a compromise between accuracy (ACC) and computational efficiency.

Healey and Picard [18] investigated the applicability of physiological signals from Electrocardiogram (ECG), Electromyogram (EMG), GSR, and RA (i.e., the rib cage expansion) to determine driver's stress levels in a real-world scenario. The authors reported that three stress levels could be recognized with an overall ACC of 97.4% using statistical handcrafted features extracted from 5-minute data segments.

Handcrafted feature design is often associated with data fusion when dealing with multiple sensors, that is, the problem of how to integrate them to achieve better analysis results [19]. Zhang et al. [20] proposed a Bayesian network for the hierarchical merging of multi-sensor data, which differs from conventional approaches that integrate features like a flat layer. Downstream of a two-stage process for selecting statistical features, the authors suggested an approach capable of autonomously learning the Bayesian network structure. The authors conducted the experiments using two public domain datasets for stress detection, including EMG, GSR, HR, RA, and BP signals, and so obtaining an ACC of 90.53%.

Among the various physiological and neurovegetative signals, the HRV analysis (i.e., R–R interval calculated from ECG peaks) effectively reflects the ANS regulation of cardiac activity [21]. Specifically, the high-frequency power of the HRV, i.e., from 0.15 Hz to 0.40 Hz, is associated with the PNS activity, while the low-frequency power, i.e., in the 0.04-0.15 Hz band, is an indicator associated with the activity of the SNS. Wang et al. [22] investigated the use of HRV to distinguish physiological conditions under three different stress levels. The authors proposed a statistical feature selection algorithm based on a k-nearest neighbor classifier capable of exhibiting 88.28% ACC on a public domain dataset for assessing stress while driving.

In order to develop a prototype system for mental stress assessment, Chiang [23] combined various approaches, such as Single Value Decomposition (SVD), fuzzy theory, and associative Petri nets, to extract and analyze HRV from the ECG signal. The author extracted 12 features, both time-domain (statistical) and frequency-domain (power spectrum), from which selected the nine most representative ones by using the information gain method. The reported results, obtained on a public domain dataset, showed an ACC of 95.10%.

Chen et al. [24] developed an automatic system to detect driving-related stress levels based on multi-channel physiological signals. Various features were extracted by the authors using wavelet decomposition, time, and spectral analysis, and combining Sparse Bayesian Learning (SBL) and Principal Component Analysis (PCA) techniques to select the optimal feature set. Finally, they used Kernel-based classifiers to improve stress detection ACC, which on a publicly available dataset was 89.70%.

Zhang et al. [25] investigated the feasibility of recognizing different stress levels from heterogeneous data of a physiological type (such as ECG, EMG, and GSR) and of a behavioral type, i.e., by using the reaction time. The authors employed visual and acoustic stressors to elicit stress in the subjects during the experiment, reporting a stress detection ACC of 92.36%.

Numerous methods have been devised to convert time series of any complexity into feature sets that can represent the dynamic characteristics of the original time series [26]. The selection of features relevant to the problem under consideration was typically made manually, without a quantitative comparison between the various candidate features. Nevertheless, this handcrafted process left uncertainty about the optimality of the selected ensemble. For this reason, data-driven methods have recently been proposed that can make systematic comparisons between a large number of time series features. One of these approaches has been operationalized in the form of a Matlab® toolbox with the name hctsa (highly comparative time-series analysis) [27].

Similarly, Lubba et al. [28] have developed a data-driven framework, called CATCH22 (22 CAnonical Time-series CHaracteristics), capable of distilling the most representative features from thousands of candidates extracted from a set of 93 different time series classification problems, including scientific, industrial, financial, and medical applications. The authors implemented their framework in C language while providing wrappers for python, R, and Matlab®.

The handcrafted features are obtained through a labor-intensive engineering process based on experience and a priori knowledge, marking the limits of current machine learning algorithms, unable to extract all the juice contained in the data. In order to expand the applicability of machine learning algorithms, it is highly desirable to automate the feature extraction process, making the algorithms less dependent on feature engineering.

Feature learning, also known as representation learning, or end-to-end learning, has recently established itself in the habit of Deep Neural Networks (DNNs). Indeed, initially used to solve complex image classification and object recognition problems, DNNs have also proved helpful for extracting features regardless of the specific classification/regression problem on hand [29]. Furthermore, the deep feature learning process is intimately connected with unsupervised learning. In fact, the learning of features does not require labeled samples since the aspects relevant to the prediction problem under consideration (i.e., classification or regression) are somehow incorporated in the distribution of the input data. This is particularly true under the manifold hypothesis, according to which the elements of application interest are always associated with low dimensional regions (i.e., manifolds) included in the original data space [30].

According to the "greedy layer-wise unsupervised pre-training" paradigm [31], the feature hierarchy is learned one level at a time through unsupervised learning of the transformation that connects one level to the next. In doing so, a layer of weights is added to the deep neural network at each iteration. Feature learning was applied by Wang and Guo [32] to the problem of recognizing driver's stress states. The authors proposed a two-stage model. Initially, the features were learned by deep learning based on Multi-layer Perception (MLP) Auto-Encoders (AEs) from physiological signals of ECG, GSC, HR, HRV, and

 doi:10.20944/preprints202203.0403.v1

RA; subsequently, the stress states were recognized using the AdaBoost classifier. The reported results showed 90.09% ACC on a publicly available dataset.

Time series prediction, previously based mainly on analytical models, such as Auto-Regressive Integrated Moving Average (ARIMA) [33], has recently been increasingly improved by deep learning models. In particular, the adaptation of the MLP to deal with time series is represented by the Recurrent Neural Network (RNN), whose main characteristic is that the output of the hidden layer at the present instant is transferred to the hidden layer of the following time instant to preserve the time dependence of the data. However, in the presence of long-time dependencies, this transfer becomes difficult, raising the problem known as vanishing gradient in the back-propagation calculation [34]. To overcome such drawbacks, Hochreiter and Schmidhuber [35] proposed the Long Short-Term Memory (LSTM), in which the internal structure of the hidden layers is more complicated by the presence of blocks equipped with forget gate, input gate, and output gate. The distinctive aspect is that the memory cell state crosses the entire chain to selectively add or remove information through the intervention of the three gates.

Sundaresan et al. [36] proposed an LSTM-based DNN to classify mental stress from EEG scalp signals in young people with ASD. The results showed that mental stress states could be accurately assessed in adolescents with and without ASD, and adolescents with varying baseline levels of anxiety. The effectiveness of LSTM has been demonstrated, in particular, for anomaly detection in time series with remarkable results [37, 38, 39]. The anomaly detection, in this case, is based on the application of the reconstruction error used as an anomaly score. An AE structure is often used to compress and reconstruct multi-dimensional input starting from non-anomalous training data. Indeed, AE cannot correctly reconstruct never-before-seen patterns of anomalous data, unlike previously-seen patterns of non-anomalous data.

CNNs represent the most prominent example of deep learning exploitation for feature extraction, initially applied mainly to solve computer vision problems such as object recognition and classification [40, 41, 42], later they were also used for processing of physiological multivariate time series [43, 44]. However, CNNs are not born to manage temporal dependencies; therefore, to fill this gap, Bai et al. [45] proposed the Temporal Convolutional Network (TCN), transposing the time-dependency problem from the RNN domain to the CNN one. TCNs proved superior to LSTMs in various application fields [46, 47], also resulting in computationally more efficiency [48].

The present study aims to establish a comparison between handcrafted and learned features in predicting behavioral changes from physiological signals during multisensory stimulation therapy in Dementia. The deep learning framework put together the benefits of TCN in feature extraction and of Bidirectional Long Short-Term Memory (BLSTM) in change detection, resulting in increased computational efficiency and better detection performance. The change prediction of the patient's behavioral status supports therapists in decision-making on selecting suitable stimulations. The study was carried out as part of the MS-Lab project [6], and the computational framework developed was integrated into the CDSS, currently undergoing clinical trials.

2. Materials and Methods

In this section the algorithmic framework is presented, focusing on the behavioral change prediction from physiological multivariate time series.

2.1. Experimental Setup and Data Acquisitions

The Experimental protocol adopted in this study was approved by the Ethics

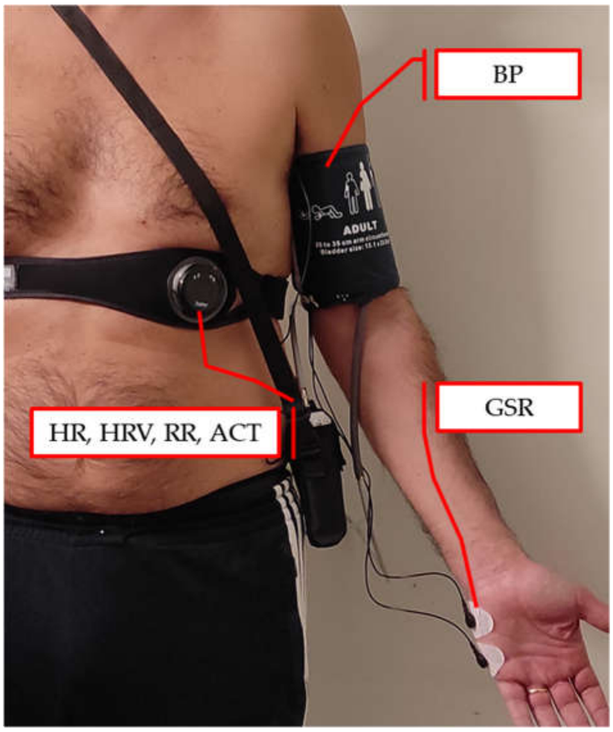

Figure 1. Multi-sensor setup for the acquisition of physiological and neurovegetative parameters.

Committee of the University of Salento (Lecce, Italy). The physiological signals of HR, RR, HRV, and Activity Level (ACT) were measured through chest strap Zephyr™ BH3 [49], worn by each subject during the stimulation session. The physiological signal GSR was measured by using the MINDFIELD® eSense [50] device provided with two electrodes attached to the outer side of the left-hand palm. The BP were measured with a wearable cuff-based device manufactured by GIMA® [51]. All sensors mentioned above are shown in Figure 1. The physiological signals were collected in two different datasets, namely Dataset 1 (DS1) and Dataset 2 (DS2). In addition, a further dataset DS2' was also obtained starting from DS2 by suppressing the GSR signal. The main characteristics of the two datasets are summarized in Table 1 and described in detail below.

In the case of DS1, four patients were recruited from the nursing center of Casa Amata Srl (Taviano, Lecce, Italy) based on their degree of dementia severity, assessed through the Mini-Mental Statement Examination (MMSE) with a score lower than 10 points [6]. Four behavioral states were considered: Active (AC), Agitated (AG), Apathetic (AP), and Relaxed (RE). Initially, each patient underwent a neurological examination to establish the underlying behavioral state, i.e., AG or AP behavior. Two patients were enrolled with an underlying AG type clinical condition, and the other two with an AP underlying clinical condition.

Then, during the therapy session, each patient was subjected to a multisensory stimulation lasting 7-13 minutes, after an initial period of equal duration in the absence of stimulation used as a baseline. The type of stimulation was chosen based on the patient's underlying behavioral status and preferences. For example, in the case of a patient with AG behavior, stimulation will be selected to relax, i.e., to change the behavioral state from AG to RE. The stimulations used consisted mainly of exposure to video clips according to each patient's preferences (e.g., dances and sounds of local folk, rock music bands, relaxing light colors and sounds, etc.).

HR, HRV, RR, and ACT parameters were acquired during the session. Instead, the BP and GSR parameters were not acquired as the measurement systems (i.e., electrodes attached to the palm and a cuff in the arm) were not adequate for the clinical conditions of some patients (especially for those with AG behavior). The therapist manually annotated the behavioral states manifested by each patient.

Regarding the DS2 dataset, to further validate the algorithmic framework with also additional physiological signals, i.e., the GSR and BP, the dataset was collected by involving five healthy volunteers. The participants were exposed to different stimulation scenarios to elicit the four behavioral states AC, AG, RE, and AP.

Specifically, for the AC and AG behaviors, the participants were asked to watch short video clips selected from the FilmStim database [52] to elicit specific emotions as follows: AC – amusement, sadness, tenderness; AG – anger, fear, disgust. In the case of the AP behavioral state, the elicited emotion was boredom by exposing the volunteers to the repetitive task of performing orally simple arithmetic operations displayed on the screen.

Finally, the RE behavior was simulated using the relaxation VR application developed by TRIPP, Inc. [53] for the VR viewer Oculus Quest 2 [54]. The exposure to stimulations lasted from 15 to 60 minutes, and the same volunteer signaled the beginning of the new behavioral state. Then, from the beginning of the new behavioral state, the neurovegetative signals were extracted for a duration equal to the dataset DS1, i.e., about 7-13 minutes.

Table 1. Overview of the collected datasets.

Dataset	Involved subjects	Included signals	Behavioral states
DS1	4 (patients)	HR, RR, HRV, ACT	AC, AG, AP, RE
DS2	5 (healthy volunteers)	HR,RR,HRV,BP,GSR,ACT	AC, AG, AP, RE
DS2′	5 (healthy volunteers)	HR, RR, HRV, ACT	AC, AG, AP, RE

2.2. Data Preprocessing

All signals provided by the BH3 [49] and eSense [50] devices were sufficiently clean, so filtering was unnecessary. The BH3 device provided the HR, HRV, RR, and ACT signals at one sample per second, whereas the eSense device sampled the GSR signal with a sample rate of 5 Hz, so it was necessary to down-sample the GSR signal at 1 Hz.

In order to have balanced datasets, the duration of the acquired stimulations was standardized to 7 minutes each. Physiological signals were treated as multivariate time series by extracting sliding windows lasting 10 to 70 seconds at a 1-second step. Thus, the processed time-series samples varied from 351 (window duration of 70 s) to 411 (window duration of 10 s) for each behavioral state, getting a total amount of samples per subject ranging from 1364 to 1604 in the case of DS1, and from 1023 to 1203 in the case of DS2.

In order to better evaluate the handcrafted features compared to the learned ones, the statistical features defined in the previous study [55] have been extended by extracting the 22 features suggested by Lubba et al. [28]. This collection of features, called by the authors "CAnonical Time-series CHaracteristics" (CATCH22) and summarized in Table 2, is the result of a selection among 5000 candidate features, carried out by evaluating the classification performance of 93 different time series.

Let d_S be the stimulation duration (in seconds), and let $c_W \in \{10, 15, 20, 30, 40, 50, 60, 70\}$ be the window duration (in seconds), the number of sampled windows is given by $n_W = d_S - c_W$, and the feature extraction process can be defined as the following map function

$$CATCH22:\ ts_V \in \mathbb{R}^{4\,n_W \times n_C\,c_W} \rightarrow ts_F \in \mathbb{R}^{4\,n_W \times n_C\,n_F} \tag{1}$$

where $n_C \in \{4, 6\}$ is the number of collected physiological signals, $n_F = 22$ is the number of extracted features, ts_V is the time-series matrix, and ts_F is the corresponding feature matrix.

Table 2. Time-series features suggested by Lubba et al. [28] and included in CATCH22 collection.

Type	Description
Distribution	• Mode of z-scored distribution: 5-bin histogram. • Mode of z-scored distribution: 10-bin histogram.
Simple temporal statistics	• Longest period of consecutive values above the mean. • Time intervals between successive extreme events above the mean. • Time intervals between successive extreme events below the mean.
Linear autocorrelation	• First $1/e$ crossing of autocorrelation function. • First minimum of autocorrelation function. • Total power in lowest fifth of frequencies in the Fourier power spectrum. • Centroid of the Fourier power spectrum. • Mean error from a rolling 3-sample mean forecasting.
Nonlinear autocorrelation	• Time-reversibility statistic, $\langle (x_{t+1} - x_t)^3 \rangle_t$. • Automutual information, $m = 2, \tau = 5$. • First minimum of the automutual information function.
Successive differences	• Proportion of successive differences exceeding 0.04σ. • Longest period of successive incremental decreases. • Shannon entropy of two successive letters in equiprobable 3-letter symbolization. • Change in correlation length after iterative differencing. • Exponential fit to successive distances in 2-d embedding space.
Fluctuation Analysis	• Proportion of slower timescale fluctuations that scale with DFA (50% sampling). • Proportion of slower timescale fluctuations that scale with linearly rescaled range fits.
Others	• Trace of covariance of transition matrix between symbols in 3-letter alphabet. • Periodicity measure of Wang et al. [56].

2.3. One-Class Support Vector Machine

The behavioral change detection can be posed as a one-class classification problem where the target (or normal) class corresponds to the behavioral state observed prior to the stimulation administration, also called baseline state. As is well known, the one-class classification is characterized by a sufficiently large number of samples belonging to the target class, while the samples belonging to classes not of interest (outliers) are absent or very few. Such condition is naturally satisfied by the application under consideration, i.e., behavioral-state change detection, since it is the therapist that determines the initiation of the multisensory stimulation. Thus, the classifier training occurs during the baseline state, and it ends when a stimulation is applied. The classic formulation of the One-Class Support Vector Machine (OCSVM) provides for using hyperplanes to isolate the target class samples from outliers that are assumed to fall on the plane through the origin [57]. Hence, the OCSVM algorithm maps data points of the feature space (ts_F) into the Kernel space to separate them with maximum margin, assigning the value +1 to points of the target class and -1 to the other points.

Let $u \in \mathbb{R}^N$ be the normal vector of the hyperplane separating the target class from the origin, let $z_i \in \mathbb{R}^N$ be the *i-th* row of ts_F $(i = 1, \dots, M)$, let $\xi \in \mathbb{R}^M$ be slack variables that penalize the outliers, let $\rho \in \mathbb{R}$ be the maximum separation distance of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2022 doi:10.20944/preprints202203.0403.v1

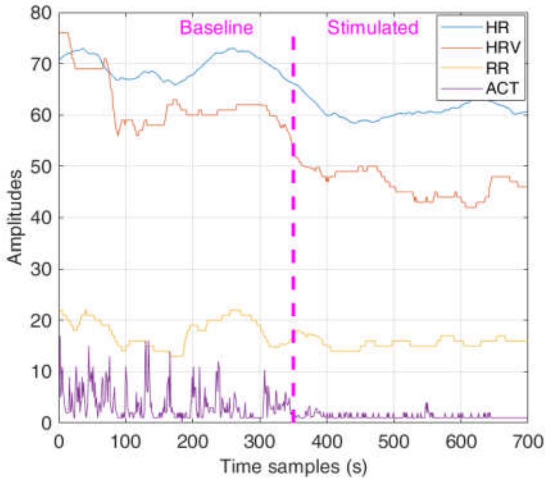

Figure 2. Physiological signals in baseline (AG) and stimulated (RE) behavioral states.

hyperplane from the origin, and let $\nu \in]0,1]$ be the upper bound on the percentage of outliers, hence the normal vector u is given by solving the following maximization problem

$$\max_{u,\xi,\rho} \frac{1}{2}\|u\|^2 + \frac{1}{\nu M}\sum_{i=1}^{M}\xi_i - \rho$$
$$\text{subject to } \forall i = 1, \dots, M\colon\ u \cdot z_i^T \geq \rho - \xi_i,\ \xi_i > 0. \tag{2}$$

Considering physiological signals measured during a multisensory therapy session, as shown in Figure 2, the baseline physiological signals are used to train the OCSVM model. After that, during the stimulation phase, changes in physiological parameter can be detected through model testing, predicting a Behavioral State Change (BSC), e.g., from AG to RE in the example reported in the previously mentioned figure.

2.4. Bidirectional Long Short-Term Memory Autoencoders

Let $x_i \in \mathbb{R}^N$ be an input time-series data (i.e., the i-*th* row of ts_V), let $W_k, R_k \in \mathbb{R}^{M\times N}$ $(\forall k \in \{\mathrm{I}, \iota, f, o\})$ be weight matrices, let $b_k \in \mathbb{R}^M$ $(\forall k \in \{\mathrm{I}, \iota, f, o\})$ be bias vectors, a LSTM memory cell at time step i is defined by its input I_i, its state c_i, its output O_i and its gates ι_i, f_i, o_i (input gate, forget gate, and output gate, respectively), and hence its transition equations are given as follows

$$\begin{aligned}
\mathrm{I}_i &= h(W_\mathrm{I} x_i + R_\mathrm{I} O_{i-1} + b_\mathrm{I}) \\
\iota_i &= \sigma(W_\iota x_i + R_\iota O_{i-1} + b_\iota + p_\iota \circ c_{i-1}) \\
f_i &= \sigma\big(W_f x_i + R_f O_{i-1} + b_f + p_f \circ c_{i-1}\big) \\
o_i &= \sigma(W_o x_i + R_o O_{i-1} + b_o + p_o \circ c_i) \\
c_i &= \iota_i \circ \mathrm{I}_i + f_i \circ c_{i-1} \\
O_i &= o_i \circ h(c_i)
\end{aligned} \tag{3}$$

where (p_ι, p_f, p_o) are three peephole connections scaling the gates with the cell state, $\sigma(\cdot)$ is the sigmoid activation function, $h(\cdot)$ is the hyperbolic tangent activation function, and $\circ$ denotes the Hadamard product.

The structure of the LSTM memory cell described above allows the network to access long time-series sequences in both backward and forward directions (within the same time window). The general structure of such a BLSTM network is shown in Figure 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2022 doi:10.20944/preprints202203.0403.v1

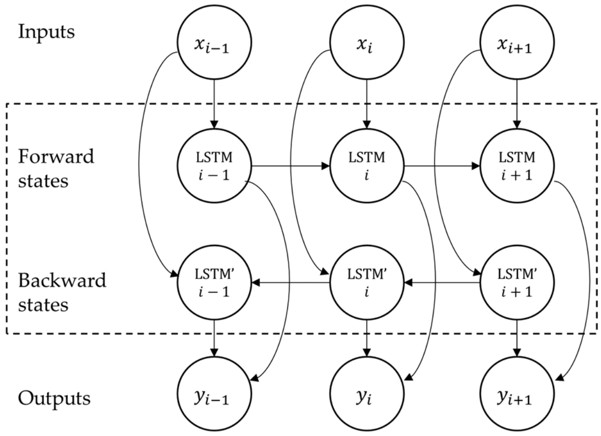

Figure 3. Diagram of a BLSTM layer.

In this paper, the BSC prediction using feature learning is accomplished via AE, in which both encoder and decoder networks are based on BLSTM. An AE is an unsupervised neural network consisting essentially of an input layer, an encoder neural network, a decoder neural network, and an output layer. Once compressed by the encoding, the data provided as input are represented in the so-called latent space. Then, the decoder decompresses such a latent representation trying to reconstruct the input data into output.

More specifically, let $z_i \in \mathbb{R}^N$ (with $i = 1, \ldots, M$) be the time series provided as input to the AE network, let $E(z_i) \in \mathbb{R}^{N'}$(with $N' < N$) be the encoded representation provided by the encoder network, let $\hat{z}_i = D\big(E(z_i)\big) \in \mathbb{R}^N$ be the reconstructed input provided by the decoder, the AE training consists in minimizing the reconstruction error

$$\mathrm{RE}(z, \hat{z}) = \frac{1}{2}\sum_{i=1}^{M} \| z_i - \hat{z}_i \|^2 \tag{4}$$

which is backpropagated through the network to update the weights.

The effectiveness of the AE in learning features lies in constraining the latent space to be smaller than the input ($N' < N$), which forcing the neural network to learn the most salient features of the time series data ts_V.

The network parameters of the BLSTM-AE architecture, whose overview is shown in Figure 4, are optimized using the genetic approach presented by Diraco et al. in [58]. For this purpose, a variable number of blocks is considered ranging from 3 to 5, two external and one more internal, each block consisting of BLSTM, fully-connected, Rectified linear unit (ReLU) and dropout layers, where the last two layers are optional. At the end of the optimization process, the obtained architecture is compound of three blocks, of which the first and last consist of only the BLSTM (B_1 and B_3) and fully-connected (F_1 and F_3) layers, while the central one includes all layers (B_2, F_2, R_2, D_2). Regarding the network parameters, the number of hidden units N_k, the output dimensions F_k, and the dropping out probability D_2 are provided in Table 3.

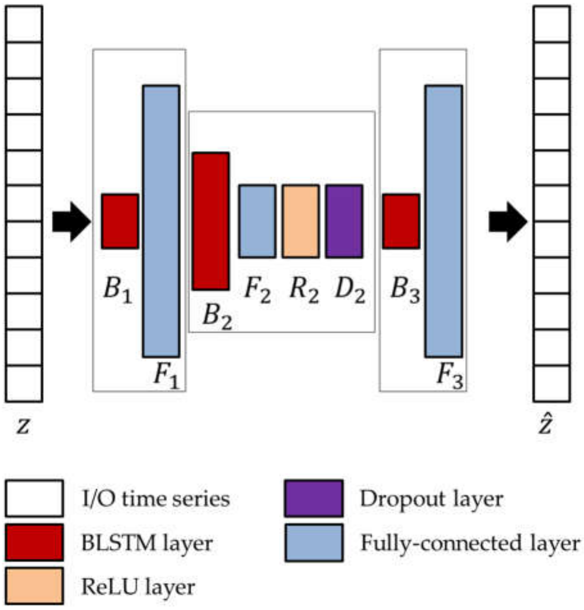

Figure 4. Architecture of the BLSTM-AE network.

Table 3. Optimized parameters of the network architecture shown in Figure 4.

Network parameters	Optimized values
B_1, F_1	16, 500
B_2, F_2, D_2	256, 50, 0.7810
B_3, F_3	16, 500

2.5. Temporal Convolutional Network

In order to increase the representational power of learned features also in situations of temporal dependencies that go beyond a single observation window, the use of a supervised pretrained network based on temporal convolution (i.e., TCN) was investigated in combination with the unsupervised Bidirectional Long Short-Term Memory Autoencoder (BLSTMAE) network previously described. In addition, to not jeopardize the basic unsupervised structure of the BLSTMAE approach, the pre-training of the TCN was conducted on the DS2 dataset, which was different from the one used for validation, i.e., the DS1 dataset.

TCN networks [45] are convolutional networks specifically designed to process time series, similar to LSTM networks but with even better performance. The main feature of TCN networks is implementing a dilated causal convolution, i.e., it only involves values temporally prior to the current one. This allows the network to capture long-term patterns, increasing the receptive field without resorting to pooling and, thus, avoiding loss of resolution [59].

Given the input sequence $x \in \mathbb{R}^N$, the dilation factor d, the convolutional kernel g of size $S \in \mathbb{N}$ (with $N > S > d$), thus the Dilated Causal Convolution (DCC) with dilation factor d at the time instant i is defined as follows

$$\mathrm{DCC_d}(x, g)(i) = \sum_{j=0}^{S-1} g(j) x_{i-dj} \quad (5)$$

that for $d = 1$ corresponds to the classical convolution. By exponentially increasing the dilation factor at each layer, it is possible to obtain a wider receptive field. In this way,

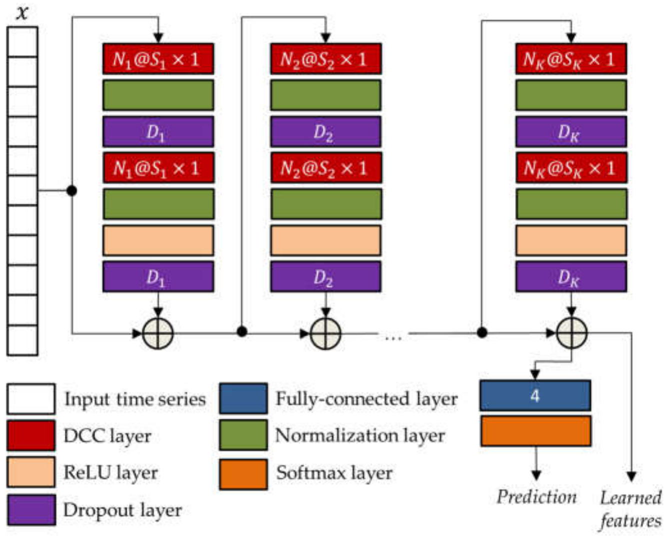

Figure 5. The general TCN architecture with K residual blocks.

considering a total amount of K layers, the size D of the receptive field of the network is given by

$$D = (S - 1)(2^K - 1) + 1. \quad (6)$$

The general TCN architecture, provided in Figure 5, has a modular structure based on K residual blocks, each including two DCC with equal dilation factor, depth and size. Suck blocks are characterized by residual connections that, as suggested by He et at. [60], improve performance of deep architectures by adding the block input to its output.

As for the BLSTMAE network, also in the case of the TCN network the parameters have been optimized using the genetic approach presented in [58]. The corresponding optimized parameters, i.e., the numbers of convolutional filters N_k, the filter sizes S_k, and the drop out percentages D_k, are reported in Table 4.

Table 4. Optimized parameters of the network architecture shown in Figure 5.

Network parameters	Optimized values
K	5
N_1, S_1, D_1	256, 8, 0.6116
N_2, S_2, D_2	256, 6, 0.6391
N_3, S_3, D_3	256, 19, 0.0438
N_4, S_4, D_4	256, 8, 0.6323
N_5, S_5, D_5	256, 7, 0.5121

2.6. Joint Temporal Convolutional Network And Bidirectional Long Short-Term Memory Autoencoders

As already mentioned, the two networks TCN and BLSTMAE are put together in order to increase the representation power of learned features. In the BLSTMAE TCN joint architecture, the TCN network plays the role of feature extraction, while the BLSTMAE network plays the role of detecting BSC. As shown in Figure 6, the TCN is pre-trained by using time-series data from DS2, whereas DS1 time-series data are used solely for testing. This distinction allows to preserve unsupervised operation during the testing phase.

It is important to note that the TCN pretraining is done by simulating the behavioral states AC, AG, RE, and AP, involving healthy volunteers, i.e., whose physiological parameters are collected in the DS2′ dataset. In the testing phase, instead, the joined

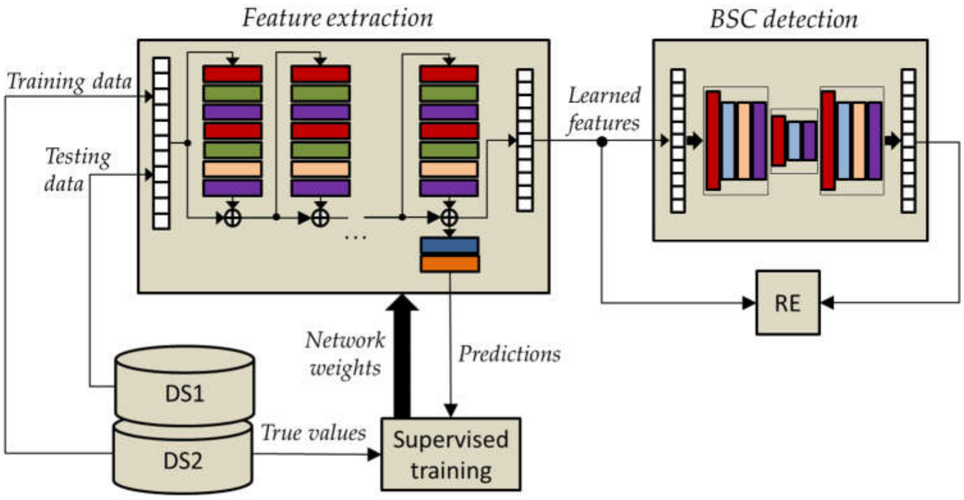

Figure 6. General overview of the joined architecture including the TCN and BLSTM-AE networks.

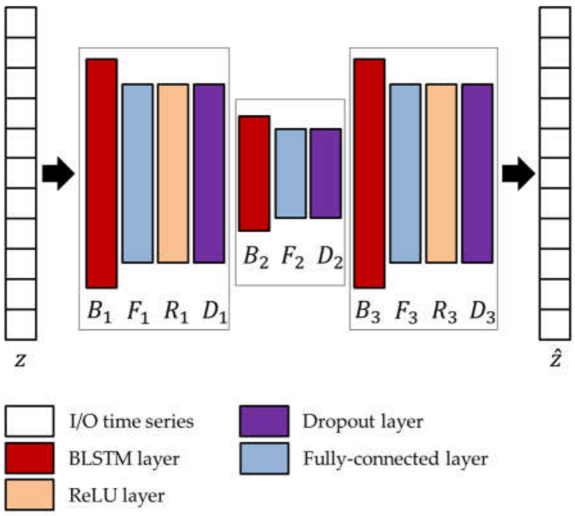

Figure 7. Architecture of the BLSTM-AE network adopted in conjunction with the TCN.

networks operate in unsupervised mode since activations (i.e., learned features) extracted from the pre-trained TCN are supplied as input to the BLSTMAE, which operates naturally in an unsupervised manner, and then the RE is estimated comparing learned features and reconstructed ones by using Eq. 4.

In the joint architecture, the parameters of the BLSTMAE network are optimized again on the basis of the activations extracted from the TCN network, and by following the approach presented in Diraco et al. [58]. The optimized architecture is provided in Figure 7, and optimized network parameters, i.e., number of hidden units B_k, output size F_k, and dropping out probability D_k, are reported in Table 5.

Table 5. Optimized parameters of the network architecture shown in Figure 7.

Network parameters	Optimized values
B_1, F_1, D_1	256, 200, 0.0083
B_2, F_2, D_2	128, 100, 0.2875
B_3, F_3, D_3	256, 200, 0.0095

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2022 doi:10.20944/preprints202203.0403.v1

3. Results

A performance comparison of the three approaches is provided in Table 6. As can be seen, although generally the ACC percentages decrease with the window (WD) durations, this trend is much less pronounced in the case of the BLSTMAE TCN approach. The OCSVM approach based on handcrafted features has been evaluated in correspondence to windows with duration equal to or greater than 15s, since for windows of shorter duration not all the features of the CATCH22 collection are defined.

The ACC is defined in terms of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) as follows

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad . \tag{6}$$

TP, TN, FP, and FN refer to change predictions, and more specifically, TP are the changed states correctly predicted as changed, TN are unchanged states correctly predicted as unchanged, FP are unchanged states wrongly predicted as changed, and FN are changed states wrongly predicted as unchanged.

All BSCs are considered from the baseline behavioral state to a different behavioral state manifested after the stimulation. For example, the AG-AC change indicates the transition from the baseline state of AG to the stimulated state of AC after the administration of sensory stimulation. The Receiver Operating Characteristic (ROC) curves of the BSP achieved with the evaluated approaches are reported from Figure 8 to Figure 13, providing both values of ACC and Area Ander the Curve (AUC).

Table 6. Average ACC percentages of the three approaches at the varying of window durations.

WD (seconds):		70s	60s	50s	40s	30s	20s	15s	10s
Method	**Dataset**								
OCSVM	DS1	95.90	94.19	91.98	87.80	85.36	82.28	79.03	NA
	DS2	98.24	97.66	96.51	93.26	91.31	88.79	85.68	NA
	DS2′	97.69	97.39	96.24	92.77	91.06	88.61	85.55	NA
BLSTMAE	DS1	85.98	85.02	83.14	81.26	80.97	80.46	79.37	74.97
	DS2	91.01	87.78	86.82	86.61	85.69	85.14	84.93	83.92
	DS2′	86.36	86.09	85.93	85.75	84.17	83.75	83.05	82.55
BLSTMAE TCN	DS1	99.25	98.92	98.91	98.72	98.71	98.69	98.68	98.44
	DS2	99.42	99.28	99.18	99.11	99.04	99.02	98.92	98.38
	DS2′	99.21	99.13	99.03	98.94	98.89	98.81	98.59	98.42

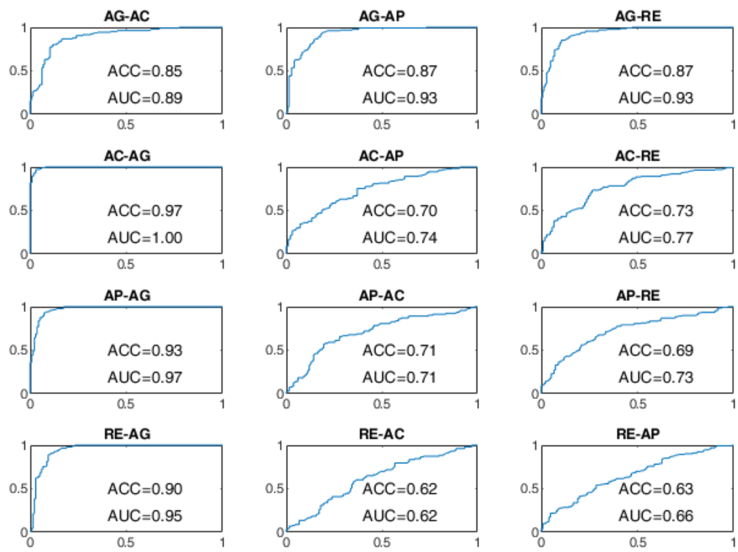

Figure 8. ROC curves of the OCSVM approach on the DS1 dataset for WD=15s.

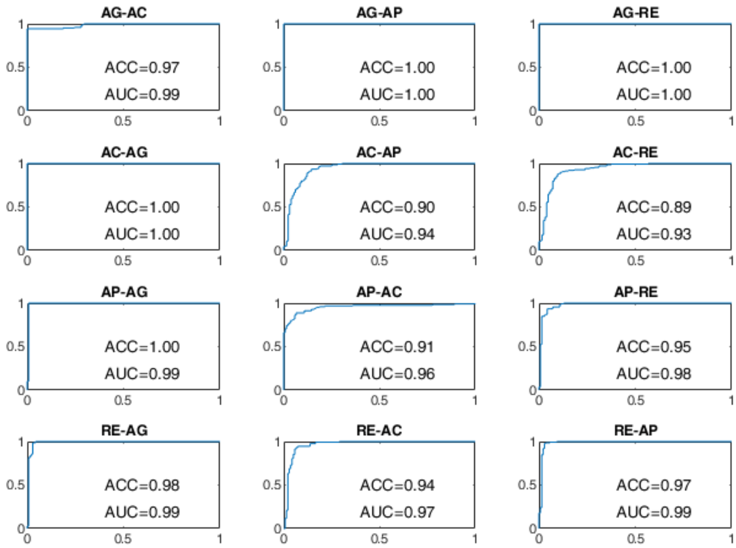

Figure 9. ROC curves of the OCSVM approach on the DS1 dataset for WD=70s.

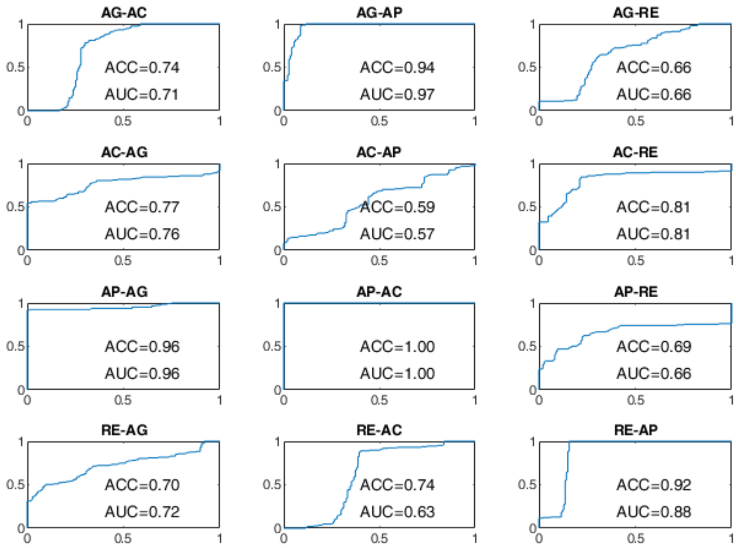

Figure 10. ROC curves of the BLSTMAE approach on the DS1 dataset for WD=15s.

The ROC curves of the OCSVM approach are shown in Figure 8 and Figure 9. With WD=15s, the worst performance was found in correspondence to changes in behavioral status AC-AP, AC-RE, AP-AC, AP-RE, RE-AC, RE-AP, with ACC less than 80%. In the cases of RE-AC and RE-AP, the ACC was lower than 70%. In all other changes, the ACC was greater than 90%. Performance has improved significantly with WD = 70s. Almost all state changes exhibited ACC greater than 95%, except AC-AP and AC-RE settled at 94% and 93%, respectively.

As regards the BLSTMAE approach, whose ROC curves are provided in Figure 10 e Figure 11, in the case of WD=15s, the worst performance was found for the state changes AG-RE, AC-AP, AP-RE, and RE-AC, with lower ACC values to 70%. The best performances, on the other hand, were found for changes of state AG-AP, AP-AG, and AP-AC. In the case of WD = 70s, the performance is improved but not that much. The BCP ACC of AG-AP and RE-AC changes even worsened.

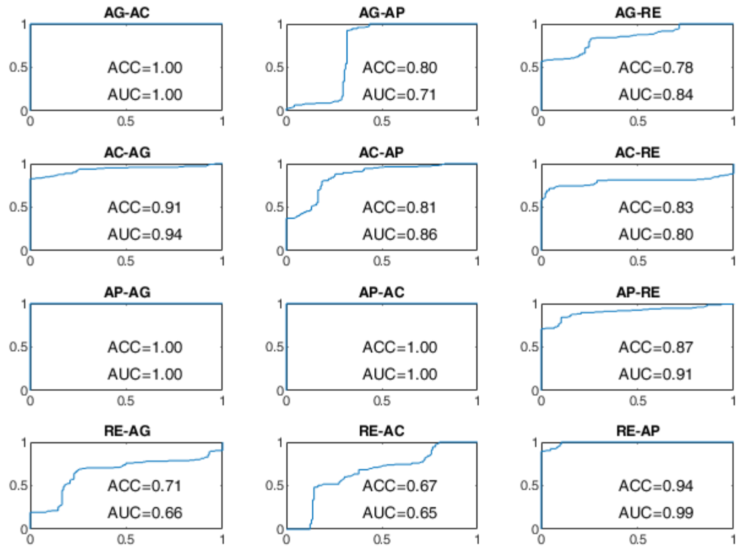

Figure 11. ROC curves of the BLSTMAE approach on the DS1 dataset for WD=70s.

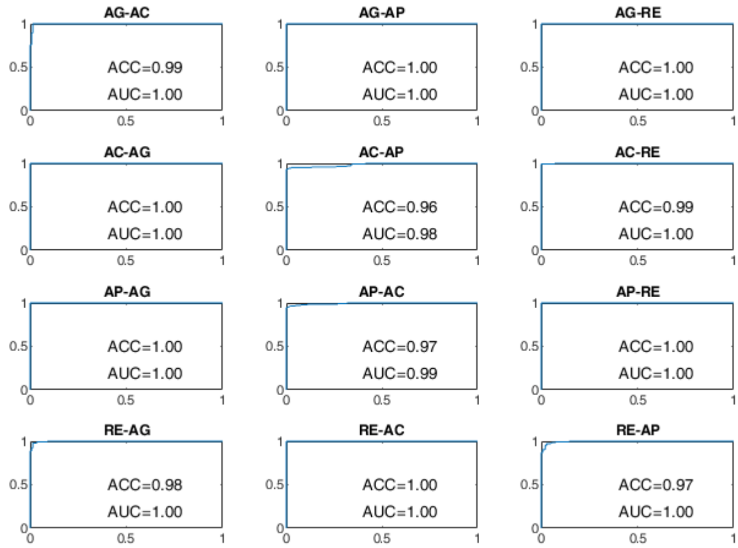

Figure 12. ROC curves of the TCN BLSTMAE approach on the DS1 dataset for WD=15s.

Finally, the ROC curves of the BLSTMAE TCN approach are provided in Figure 12 and Figure 13. With this approach, even for WD = 15s the ACC of the BCP is higher than 96% for all state changes and it is better than all other approaches including the cases where WD is equal to 70s.

In this study, all presented network architectures were implemented and evaluated using the MathWorks® Deep Learning Toolbox (v 14.2, R2021a, MathWorks Inc., Natick, Massachusetts, United States) [61]; whereas, the genetic optimizations were performed using the MathWorks® Optimization Toolbox (v 9.1, R2021a, MathWorks Inc., Natick, Massachusetts, United States) [62].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 March 2022 doi:10.20944/preprints202203.0403.v1

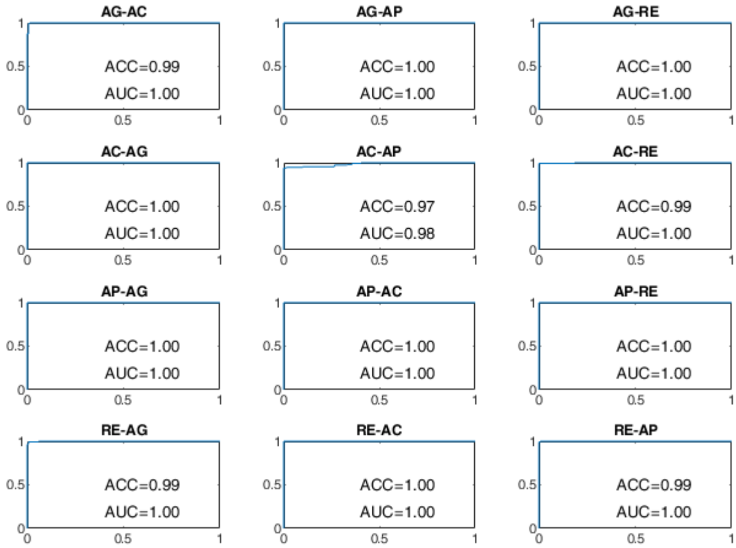

Figure 13. ROC curves of the TCN BLSTMAE approach on the DS1 dataset for WD=70s.

For each observation window from 10 to 70 seconds, the experimentation was conducted through ten-fold cross-validation on the total number of samples (i.e., ranging from 12,636 samples for 70-second windows to 14,796 samples for 10-second windows).

The OCSVM approach was evaluated (trained and tested) on a computer system with CPU Intel® Core™ i7-8565U at 2.00GHz. Both optimization, training and testing of the BLSTMAE and TCN networks were performed on a computer system equipped with CPU Intel® Core™ i7-5820K at 3.30GHz, and GPU NVIDIA GeForce® GTX TITAN X. And finally, optimization, training and testing of the joined network architecture BLSTMAE TCN were executed on a computer system based on CPU Intel® Core™ i9-10900K 3.70GHz, and GPU NVIDIA GeForce RTX™ 2060.

All network were trained from scratch using the Adam solver [63] with gradient decay factor 0.9 and initial learning rate 0.001. The number of epochs, instead, was different for the three networks, with 1000 epochs for BLSTM, 500 epochs for TCN, and 2000 epochs for BLSTMAE TCN. The genetic optimization of network hyperparameters was the process that took the most time, taking 35 days for BLSTM, 18 days for TCN, and 76 days for BLSTMAE TCN.

4. Discussion

This study developed a new approach to predicting behavioral changes from neurovegetative parameters using learned features. The proposed approach is based on deep feature learning using a pre-trained TCN. In particular, the pre-training process is based on non-field data, i.e., data not acquired by patients but prepared in a laboratory. This type of pre-training offers the advantage of making the system semi-supervised since patient data are only required during the stimulation therapy to predict behavioral changes, carried out via the BLSTMAE network.

The deep learning feature approach was then compared with the traditional approach based on handcrafted features, obtained through CATCH22, i.e., 22 features for each considered neurovegetative/physiological parameter (88 features for DS1 and DS2', and 110 features for DS2) and classified via OCSVM. The comparison between handcrafted and learned features shows a higher processing time in the handcrafted case. This is due to the greater computational complexity of the CATCH22 extraction framework than the lighter TCN (testing phase) and the need to employ a wider observation window to obtain the same performance.

On the other hand, the handcrafted CATCH22 extraction process, although it may appear to be an automatic process, it is not entirely so. In order to optimize the extraction process, it is recommended to make a feature selection to identify the most suitable set of features to exclude the least performing ones [64].

Furthermore, the learned feature approach based on joined BLSTMAE and TCN exhibits shorter reaction times and better performance than OCSVM (handcrafted features based on catch22) and BLSTMAE (learned feature) alone. As reported in Figure 6, the performances exhibited by the joint use of BLSTMAE and TCN with a window of only 10 seconds can be bought with those provided by OCSVM and by BLSTMAE alone with windows of 70 seconds. The architectures of the BLSTMAE and TCN networks have been specifically optimized to work together using the algorithm suggested in [58]. The performance was significantly lower without joint optimization, i.e., using the individually optimized BLSTMAE network.

The experimentation was conducted with six (HR, RR, HRV, BP, GSR, ACT) and four (HR, RR, HRV, ACT) signals, demonstrating that more signals bring performance benefits. Furthermore, the results with four signals on patients and volunteers were quite comparable, confirming that the volunteers' simulation of the behavioral states was carried out in a sufficiently realistic way.

Transitions involving the AC and AP states were more difficult to discriminate in the presence of four signals, particularly in the OCSVM and BLSTMAE cases. In the OCSVM case (handcrafted features), AC-AP, AP-AC, and AC-RE transitions were more problematic (Figure 9). Instead, in the case of BLSTMAE (learned features), the transitions AG-AP, AG-RE, RE-AG, AC-AP, AP-AC, AC-RE, RE-AC, and AP-RE, were more problematic (Figure 11). This may be partly because the AC state does not always manifest itself with body movements but often results in a state of sustained attention, making it difficult to discriminate from the AP state in the absence of the GSR neurovegetative signal.

The results achieved in the present study were compared with the state of the art. Given the absence of studies in the literature on predicting behavioral states such as those considered in this study, the comparison was conducted by considering studies aimed at detecting different stress levels. To this end, the scales of stress levels and behavioral states have been placed side by side considering four levels: Level -1, Level 0, Level 1, and Level 2. Level 0 indicates a total absence of stress, i.e., RE state. Level 1 corresponds to the healthy level of stress, eustress, which corresponds to the AC state considered in this study. Level 2, on the other hand, corresponds to excessive levels of stress, which also lead to a state of AG. In order to include the AP state, the negative level marked with -1 was introduced to indicate a state of no response, close to drowsiness, considered in some studies for the detection of stress and drowsiness while driving cars [22, 24]. The comparison of the results achieved with the state-of-the-art is shown in Table 7. In the case of learned features, the ACC performance obtained in this study with 10-second windows with both four and six signals exceeds the state-of-the-art. In the case of handcrafted features, however, the performances exceed the state-of-the-art only in the presence of six signals with a window of 70 seconds. However, with shorter windows, the performance remains comparable to the state-of-the-art.

The main limitation of this study concerns the small number of patients involved, which, however, was extended by involving additional volunteers. A larger clinical trial involving more patients with dementia is currently underway.

Table 7. Comparison of the achieved results with the state of the art.

Authors	**Signals**	**Features**	ACC **(%)**
Healey and Picard [18]	ECG, EMG, GSR, RA	Handcrafted	97.40
Zhang et al. [20]	EMG, GSR, HR, RA, BP	Handcrafted	90.53
Wang et al. [22]	HRV	Handcrafted	88.28
Chiang [23]	ECG, HRV	Handcrafted	95.10
Chen et al. [24]	ECG, EMG, GSR, RA	Handcrafted	89.70
Zhang et al. [25]	ECG, EMG, GSR	Handcrafted	92.36
Wang and Guo [32]	ECG, GSC, HR, HRV, RA	Learned	90.09
This study	HR, RR, HRV, ACT	Handcrafted	79.03 / 95.90
This study	HR,RR,HRV,BP,GSR,ACT	Handcrafted	85.68 / 98.24
This study	HR, RR, HRV, ACT	Learned	98.44 / 99.25
This study	HR,RR,HRV,BP,GSR,ACT	Learned	98.38 / 99.42

5. Conclusions

The contribution of this study is threefold: 1) a new approach for BSP based on BLSTMAE TCN deep feature learning has been presented; 2) the feature learning (i.e., TCN) and change detection (BLSTMAE) architectures have been set up in order to operate jointly in an optimized way; 3) the proposed framework has been validated on four patients with dementia and five volunteers, using two datasets consisting of four (three neurophysiological and one of activity) and six signals (five neurophysiological and one of activity).

Although conducted on a small number of subjects, the validation demonstrated the feasibility of the BSP, which was subsequently incorporated into a CDSS within the MS-Lab project to support therapists during the administration of multisensory stimulation therapy.

Ongoing and future activities are focused on the clinical trial of the CDSS on a more statistically significant number of subjects with dementia. The data collected during the experimentation will be used to evaluate further the effectiveness of the learned features to detect valuable indicators for predicting the patient's behavioral state during therapy.

Supplementary Materials: N.A.

Author Contributions: Conceptualization, G.D. and A.L.; methodology, G.D.; software, G.D.; validation, G.D; formal analysis, G.D.; investigation, G.D.; resources, G.D. and A.L.; data curation, G.D.; writing—original draft preparation, G.D.; writing—review and editing, G.D.; visualization, G.D.; supervision, A.L.; project administration, A.L. and P.S.; funding acquisition, A.L. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out within the project PON "Si-Robotics" (ARS_01_01120) funded by MUR - Italian Ministry for University and Research.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of University of Salento (Lecce, Italy).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.
Data Availability Statement: N.A.

Acknowledgments: The authors would like to thank colleagues of Casa Amata Srl (Taviano, Italy) and eResult (Cesena, Italy) for their support.

Conflicts of Interest: The authors declare no conflict of interest.

 doi:10.20944/preprints202203.0403.v1

Appendix A

Nomenclature

AC	Active
ACC	Accuracy
ACT	Activity Level
AE	Autoencoder
AG	Agitated
ANS	Autonomic Nervous System
AP	Apathetic
ARIMA	Auto-Regressive Integrated Moving Average
AUC	Area Ander the Curve
BLSTM	Bidirectional Long Short-Term Memory
BLSTMAE	Bidirectional Long Short-Term Memory Autoencoder
BP	Blood Pressure
BSC	Behavioral State Change
CATCH22	CAnonical Time-series CHaracteristics 22
CDSS	Clinical Decision Support System
DCC	Dilated Causal Convolution
DNN	Deep Neural Network
DS1	Dataset 1
DS2	Dataset 2
ECG	Electrocardiogram
EMG	Electromyogram
FN	False Negative
FP	False Positive
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
LSTM	Long Short-Term Memory
MLP	Multi-layer Perception
MMSE	Mini-Mental Statement Examination
MS-Lab	Multi Sensorial Stimulation Lab
OCSVM	One-Class Support Vector Machine
PCA	Principal Component Analysis
PNS	Parasympathetic Nervous System
RA	Respiration Amplitude
RE	Relaxed
ReLU	Rectified linear unit
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RR	Respiration Rate
SBL	Sparse Bayesian Learning
SNS	Sympathetic Nervous System
SVD	Single Value Decomposition
TCN	Temporal Convolutional Network
TN	True Negative
TP	True Positive
WD	Window

References

1. Bermejo, U.; Almeida, A.: Bilbao-Jayo, A.; Azkune, G. (2021). Embedding-based real-time change point detection with application to activity segmentation in smart home time series data. Expert Systems with Applications, 185, 115641.

2. Rios, R. T.; Lolli, F.; Xie, L.; Xie, Y.; Kurtis, K. E. (2021). Screening candidate supplementary cementitious materials under standard and accelerated curing through time-series surface resistivity measurements and change-point detection. Cement and Concrete Research, 148, 106538.
3. Vivancos, J. L., Buswell, R. A., Cosar-Jorda, P., & Aparicio-Fernandez, C. (2020). The application of quality control charts for identifying changes in time-series home energy data. Energy and Buildings, 215, 109841.
4. Luo, J.; Agboola, F.; Grant, E.; Masters, C. L.; Albert, M. S.; Johnson, S. C.; Xiong, C. (2020). Sequence of Alzheimer disease biomarker changes in cognitively normal adults: a cross-sectional study. Neurology, 95(23), e3104-e3116.
5. Chen, G., Lu, G., Shang, W., Xie, Z. (2019). Automated change-point detection of EEG signals based on structural time-series analysis. IEEE Access, 7, 180168-180180.
6. Boccadamo, A.; Luca, A. D.; Palamà, L.; Sancarlo, D.; Leone, A.; Diraco, G.; Siciliano, P.; Toma, A.; Maffia, M.; De Riccardis, L.; D'Alessandro, L.; Borrelli, G.; Pistoia, M. Multi Sensorial Stimulation Lab: A New Approach for Severe Dementia. In Italian Forum of Ambient Assisted Living, Springer, Cham, 2019, pp. 65-81.
7. Berntson, G. G., Cacioppo, J. T., Quigley, K. S. (**1991**). Autonomic determinism: the modes of autonomic control, the doctrine of autonomic space, and the laws of auto-nomic constraint. Psychological review, 98(4), pp. 459.
8. Beauchaine, T. (**2001**). Vagal tone, development, and Gray's motivational theory: To-ward an integrated model of autonomic nervous system functioning in psychopatholo-gy. Development and psychopathology, 13(2), pp. 183-214.
9. Goldstein, S., Naglieri, J. A., Princiotta, D., Otero, T. M. (**2014**). Introduction: A his-tory of executive functioning as a theoretical and clinical construct. In Handbook of executive functioning (pp. 3-12). Springer, New York, NY.
10. McEwen, B. S. (**2008**). Central effects of stress hormones in health and disease: Under-standing the protective and damaging effects of stress and stress mediators. European journal of pharmacology, 583(2-3), pp. 174-185.
11. Cacioppo, J. T., Tassinary, L. G., Berntson, G. G. (**2007**). Handbook of psychophysi-ology. Cambridge University Press, New York, NY.
12. Jänig, W., McLachlan E.M. (**1992**). Characteristics of function-specific pathways in the sympathetic nervous system Trends in Neurosciences, 15, pp. 475-481.
13. Sakr, G. E., Elhajj, I. H., Wejinya, U. C. (**2009**). Multi level SVM for subject independent agitation detection. In 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, IEEE, pp. 538-543.
14. Lisetti, C.L., and Nasoz, F. (**2004**). Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP Journal on Applied Signal Processing, pp. 1672-1687.
15. Andreas Kistler, Charles Mariauzoulsb, Klaus von Berlepscha. (**1998**). Fingertip temperature as an indicator for sympathetic responses. International Journal of Psychophysiology Vol.29, pp. 35-41.
16. Fernandes, A., Helawar, R., Lokesh, R., Tari, T., & Shahapurkar, A. V. (**2014**). Determination of stress using blood pressure and galvanic skin response. In 2014 International Conference on Communication and Network Technologies, IEEE, pp. 165-168.
17. Längkvist, M.; Karlsson, L.; Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters, 42, 11-24.
18. Healey, J. A.; Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems, 6(2), 156-166.
19. Mitchell, H. B. (2007). Multi-sensor data fusion: an introduction. Springer Science & Business Media.
20. Zhang, L.; Tamminedi, T.; Ganguli, A.; Yosiphon, G.; Yadegar, J. (2010). Hierarchical multiple sensor fusion using structurally learned Bayesian network. In Wireless Health 2010 (pp. 174-183).
21. Malik, M.; Bigger, J. T.; Camm, A. J.; Kleiger, R. E.; Malliani, A.; Moss, A. J.; Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European heart journal, 17(3), 354-381.
22. Wang, J. S., Lin, C. W., & Yang, Y. T. C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. Neurocomputing, 116, 136-143.
23. Chiang, H. S. (2015). Ecg-based mental stress assessment using fuzzy computing and associative petri net. Journal of Medical and Biological Engineering, 35(6), 833-844.
24. Chen, L. L.; Zhao, Y.; Ye, P. F.; Zhang, J.; Zou, J. Z. (2017). Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Systems with Applications, 85, 279-291.
25. Zhang, B.; Morère, Y.; Sieler, L.; Langlet, C.; Bolmont, B., Bourhis, G. (2017). Reaction time and physiological signals for stress recognition. Biomedical Signal Processing and Control, 38, 100-107.
26. Dong, G.; Liu, H. (Eds.). (2018). Feature engineering for machine learning and data analytics. CRC Press.
27. Fulcher, B. D.; Jones, N. S. (2017). hctsa: A computational framework for automated time-series phenotyping using massive feature extraction. Cell systems, 5(5), 527-531.
28. Lubba, C. H.; Sethi, S. S.; Knaute, P.; Schultz, S. R.; Fulcher, B. D.; Jones, N. S. catch22: Canonical time-series characteristics. Data Mining and Knowledge Discovery, 33(6), 2019, pp. 1821-1852.
29. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521 (7553), 436-444.
30. Bengio, Y. (2009). Learning deep architectures for AI. Foundations and Trends in Machine Learning, 2(1), 1–127.
31. Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural networks: Tricks of the trade (pp. 437-478). Springer, Berlin, Heidelberg.
32. Wang, K.; Guo, P. (2020). An ensemble classification model with unsupervised representation learning for driving stress recognition using physiological signals. IEEE transactions on intelligent transportation systems, 22(6), 3303-3315.

33. Guin, A. (2006, September). Travel time prediction using a seasonal autoregressive integrated moving average time series model. In 2006 IEEE Intelligent Transportation Systems Conference (pp. 493-498). IEEE.
34. Bengio, Y.; Simard, P.; Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2), 157-166.
35. Hochreiter, S.; Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
36. Sundaresan, A.; Penchina, B.; Cheong, S.; Grace, V.; Valero-Cabré, A.; Martel, A. (2021). Evaluating deep learning EEG-based mental stress classification in adolescents with autism for breathing entrainment BCI. Brain Informatics, 8(1), 1-12.
37. Park, D.; Hoshi, Y.; Kemp, C. C. (2018). A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters, 3(3), 1544-1551.
38. Fernando, T.; Denman, S.; Sridharan, S.; Fookes, C. (2018). Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection. Neural networks, 108, 466-478.
39. Ergen, T.; Kozat, S. S. (2019). Unsupervised anomaly detection with LSTM neural networks. IEEE transactions on neural networks and learning systems, 31(8), 3127-3141.
40. Dhillon, A.; Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 9(2), 85-112.
41. Radovic, M.; Adarkwa, O.; Wang, Q. (2017). Object recognition in aerial images using convolutional neural networks. Journal of Imaging, 3(2), 21.
42. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.
43. Cecotti, H.; Graser, A. (2010). Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE transactions on pattern analysis and machine intelligence, 33(3), 433-445.
44. Hajinoroozi, M.; Mao, Z.; Jung, T. P.; Lin, C. T.; Huang, Y. (2016). EEG-based prediction of driver's cognitive performance by deep convolutional neural network. Signal Processing: Image Communication, 47, 549-555.
45. Bai, S.; Kolter, J. Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. ArXiv Preprint, 2018, arXiv:1803.01271.
46. Khalili, E.; Asl, B. M. (2021). Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG. Computer Methods and Programs in Biomedicine, 204, 106063.
47. Gao, Z.; Wang, X.; Yang, Y.; Mu, C.; Cai, Q.; Dang, W.; Zuo, S. (2019). EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation. IEEE transactions on neural networks and learning systems, 30(9), 2755-2763.
48. Wu, P.; Sun, J.; Chang, X.; Zhang, W.; Arcucci, R.; Guo, Y.; Pain, C.C. (2020). Data-driven reduced order model with temporal convolutional neural network. Computer Methods in Applied Mechanics and Engineering, 360, 112766.
49. Zephyr Technology (2012). BioHarness 3.0 User Manual. Annapolis, USA. Available online: https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf (last accessed on 29 March 2022).
50. MINDFIELD® BIOSYSTEMS. (2022). MINDFIELD® eSense Skin Response - User manual Version 5.6.1 | 08.02.2022. Gronau, Germany. Available online: https://www.mindfield.de/pdf/phocadownload/eSense/English/eSense_Skin_Response_Manual_EN.pdf (last accessed on 29 March 2022).
51. GIMA S.p.a. GIMA ABPM PULSE RATE MONITOR WITH BLUETOOTH - User manual. Gessate, Italy. Available online: https://www.gimaitaly.com/DocumentiGIMA/Manuali/EN/M35112EN.pdf (last accessed on 29 March 2022).
52. Schaefer, A.; Nils, F.; Sanchez, X.; Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and emotion, 24(7), 1153-1172.
53. Schrempf, M.C.; Petzold, J.Q.; Vachon, H.; Petersen, M.A.; Gutschon, J.; Wolf, S.; Anthuber, M. (2021). Virtual reality-based relaxation for enhancement of perioperative well-being and quality of life: protocol for a randomised pilot trial. BMJ open, 11(4), e044193.
54. TRIPP(R), INC. (2022) TRIPP for Oculus Quest 2. Available online: https://www.oculus.com/experiences/quest/2173576192720129/ (last accessed on 29 March 2022).
55. Diraco, G.; Leone, A.; Siciliano, P. Clinical Decision Support System for Multisensory Stimulation Therapy in Dementia: A Preliminary Study. Accepted for publication in Lecture Notes in Electrical Engineering, 2022, Springer.
56. Wang, X.; Wirth, A; Wang, L. Structure-based statistical features and multivariate time series clustering. In: Proceedings—IEEE international conference on data mining, ICDM, 2007, pp 351–360. ISSN 15504786.
57. Alam, S.; Sonbhadra, S. K.; Agarwal, S.; Nagabhushan, P. One-class support vector classifiers: A survey. Knowledge-Based Systems, 2020, 196, 105754.
58. Diraco, G.; Siciliano, P.; Leone, A. Remaining Useful Life Prediction from 3D Scan Data with Genetically Optimized Convolutional Neural Networks. Sensors, 2021, 21(20), 6772.
59. Lara-Benítez, P.; Carranza-García, M.; Luna-Romera, J. M.; Riquelme, J. C. Temporal convolutional networks applied to energy-related time series forecasting. Applied Sciences, 2020, 10(7), 2322.
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
61. Beale, M.; Hagan, M.; Demuth, H. Deep Learning Toolbox™ Reference. MATLAB (r) R2021a. The MathWorks, Inc. 2021. Available online: https://it.mathworks.com/help/pdf_doc/deeplearning/nnet_ref.pdf (accessed on 4 March 2022).
62. Optimization Toolbox™ User's Guide. MATLAB (r) R2021a. The MathWorks, Inc. Available online: https://it.mathworks.com/help/pdf_doc/optim/optim.pdf (accessed on 4 March 2022).

63. Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
64. Pouromran, F.; Radhakrishnan, S.; Kamarthi, S. (2021). Exploration of physiological sensors, features, and machine learning models for pain intensity estimation. Plos one, 16(7), e0254108.