Article

Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning

Axel Sepúlveda [1], Francisco Castillo [2], Carlos Palma [2] and Maria Rodriguez-Fernandez [1,*]

[1] Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile; ansepulveda@uc.cl

[2] ATCAS-Grupo CLER, Temuco 4781136, Chile; francisco.castillo@atcas.cl (F.C.); carlos.palma@grupocler.com (C.P.)

* Correspondence: marodriguezf@uc.cl

Abstract: Affect detection combined with a system that dynamically responds to a person's emotional state allows an improved user experience with computers, systems, and environments and has a wide range of applications, including entertainment and health care. Previous studies on this topic have used a variety of machine learning algorithms and inputs such as audial, visual, or physiological signals. Recently, a lot of interest has been focused on the last, as speech or video recording is impractical for some applications. Therefore, there is a need to create Human–Computer Interface Systems capable of recognizing emotional states from noninvasive and nonintrusive physiological signals. Typically, the recognition task is carried out from electroencephalogram (EEG) signals, obtaining good accuracy. However, EEGs are difficult to register without interfering with daily activities, and recent studies have shown that it is possible to use electrocardiogram (ECG) signals for this purpose. This work improves the performance of emotion recognition from ECG signals using wavelet transform for signal analysis. Features of the ECG signal are extracted from the AMIGOS database using a wavelet scattering algorithm that allows obtaining features of the signal at different time scales, which are then used as inputs for different classifiers to evaluate their performance. The results show that the proposed algorithm for extracting features and classifying the signals obtains an accuracy of 88.8% in the valence dimension, 90.2% in arousal, and 95.3% in a two-dimensional classification, which is better than the performance reported in previous studies. This algorithm is expected to be useful for classifying emotions using wearable devices.

Keywords: affective computing; emotion recognition; feature extraction; machine learning; wavelet transform

1. Introduction

Affective computing (AC) aims for the integration of human emotional states in Human–Computer Interfaces and there is an actual need to develop and improve algorithms that allow machines to recognize human emotional states with different objectives. Emotional states are subjective experiences that are commonly classified in two or more dimensions. Russell's circumplex model proposes that all emotions arise from two fundamental neurophysiological systems—one related to valence and the other to arousal; valence refers to the level of pleasantness or unpleasantness of an emotion and arousal refers to its activation or deactivation level [1] (see Figure 1). Some researchers have tried to automatically correlate the emotional state dimensions with an input signal such as speech, face image recognition, or physiological signals [2–4]. In this way, some databases have been made publicly available to establish a standard framework to compare different methods and algorithms [5,6]. This work presents a novel emotion recognition algorithm based on wavelet scattering feature extraction and supervised machine learning and shows its performance using AMIGOS: a dataset for mood, personality, and affect research on Individuals and GrOupS [7].
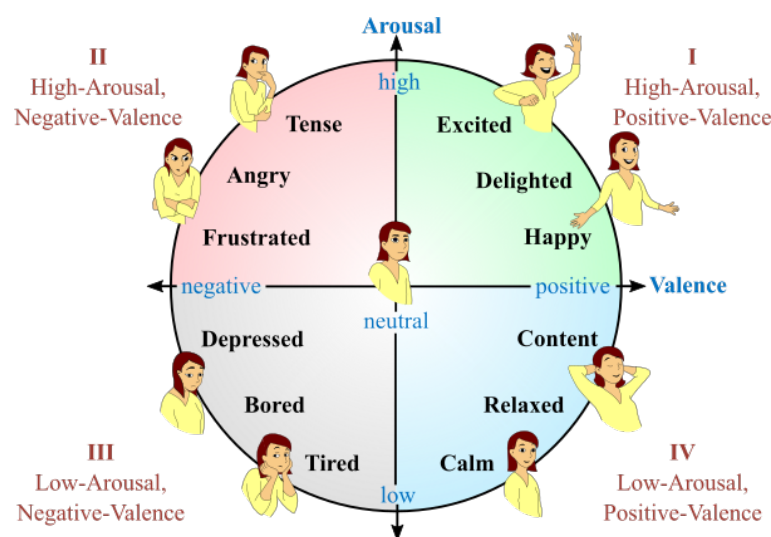
Figure 1. In Russell's circumplex model, emotions are distributed on a two-dimensional plane. The x-axis represents valence and the y-axis represents arousal [1].

Background

The computer task of emotion recognition has been studied over many years and the most used approaches include using face and body images, video recordings, and audio recordings (speech) as inputs. Most recent studies employ machine learning algorithms to perform the recognition task on an annotated dataset. Kahou et al. [8] used a multimodal deep learning approach based on convolutional neural networks (CNN), deep belief networks (DBN), and support vector machines (SVM) on audio and video datasets (AFEW) [9] used in the EmotiW Challenge [10]. Ranganathan et al. [11] fed a convolutional DBN (CDBN) with extracted regions of interest (ROI) from their published emoFBVP dataset. Fan et al. [12] combined long short-term memory (LSTM) and 3D convolutional networks (C3D). Hu et al. [13] proposed a novel method based on cascading a local enhanced motion history image (LEMHI) and a CNN-LSTM, comparing results over three different datasets (AFEW, the extended Cohn–Kanade dataset (CK+) [14], and MMI [15]).

Other types of signals, such as physiological signals, have also been explored as inputs for classification. Zheng et al. [16] recorded and extracted features from EEG and eye-tracking data to feed a support vector machine (SVM) classifier. Alhagry et al. [17] used a LSTM Recurrent Neural Network to classify emotions from the EEG signal of DEAP dataset [5], obtaining up to 85% accuracy. Balan et al. [18] proposed a comparative analysis between different machine learning and deep learning techniques, also using the DEAP dataset. Paszkiel [19] used blind signal separation for EEG signal reconstruction, allowing the identification of the source generating a given potential. However, recording audio, videos, or EEG for ambulatory or daily life application of emotion recognition is unfeasible in most cases. Thus, a less-intrusive physiological signal measurement is needed. In this way, some researchers have used less intrusive measurements such as electrocardiogram (ECG) and galvanic skin response (GSR) to classify the signals to an elicited emotion using a machine learning or deep learning classifier [4,20].

Some studies use ad hoc obtained datasets, usually with a small number of signals and volunteers. To overcome this situation, some researchers have made large, publicly available datasets with more volunteers, a mix of different signal measurements, and emotion elicitation approaches, allowing the validation and comparison of different classification algorithms in different studies. Metrics such as accuracy and F1-score were used in the comparison [21]. Some examples are DREAMER [22], HCI-Tagging [6], and AMIGOS [7]. Some recent publications use at least one of these datasets. For example, a probabilistic Bayesian deep learning algorithm has been employed to classify valence in the DREAMER (accuracy, 0.90; F1-score, 0.88) and AMIGOS (accuracy, 0.86; F1-score, 0.83) datasets [23]. HCI-Tagging data is employed in [24] to extract intrinsic mode function features from the

ECG and feed a K-Nearest Neighbors (KNN) classifier, obtaining an accuracy of 0.558 for arousal and 0.597 for valence classification. In [25], the authors use deep learning to classify ECG signals from the AMIGOS dataset, obtaining 0.76 and 0.68 F1-score for arousal and valence, respectively. In the original publication of AMIGOS [7], manually extracted time and frequency features are used, obtaining mean F1-scores of 0.545 and 0.551, respectively. In [26], mean F1-scores of 0.851 and 0.837 were obtained with a CNN self-supervised approach. Tung et al. [27] employed entropy domain features and an XGboost model for classification, obtaining mean F1-scores between 0.56 and 0.63.

Regarding model validation, one of the most used methods in this field is leave k subjects out (LkSO) cross-validation. ECG may show subject-specificity [28], so LkSO is employed to assess how the algorithms generalize to a new subject and prevent overfitting. Among LkSO, leave one subject out cross-validation (LOSO) is the most used in previous works [23,27]. Another validation scheme is 10-fold cross-validation, which is employed in [26,29,30]. Moreover, a comparison of validation schemes is performed in [24], where they find consistent results with other studies, in which arousal classification accuracy is higher than valence.

2. Materials and Methods

2.1. Dataset

A publicly available database, AMIGOS, was used in this work [7]. This database was preferred over other available databases due to its relatively high number of participants, recent publication, large number of signals and experiments, standardized data collection methods, and detailed and high-resolution annotation. This database includes EEG, ECG, galvanic skin response (GSR), and face video data from 40 participants recorded during two experiments. The first experiment (short videos) consisted of individually watching 16 short videos of less than 250 seconds each. The second experiment (long videos) consisted of watching 4 long videos of around 14 min each in groups (see Figure 2).
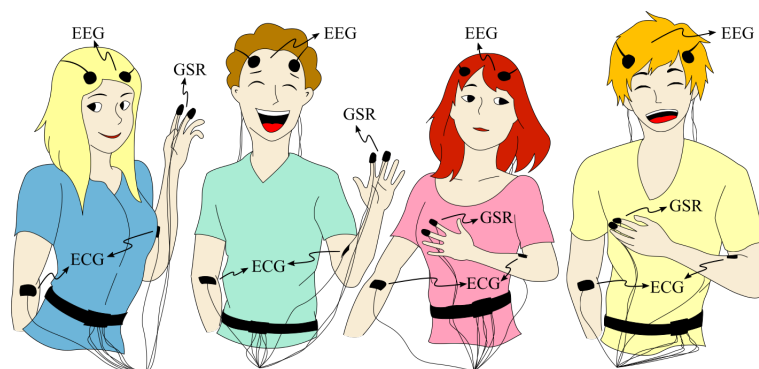

Figure 2. Participants of the AMIGOS database with EEG, ECG, and GSR recording.

Experiments were done in a lab-controlled environment to record changes in physiological signals when exposed to the video stimuli. The face video recording was subsequently externally score-annotated by experts in arousal and valence dimensions with a time resolution of 20 s (length of each annotated time window) for both short and long experiments. In addition, in the case of short experiments, a self-annotation was carried out by means of a questionnaire after each short-video presentation.

2.2. Algorithm Overview

The algorithm described in this work can be represented by the block diagram in Figure 3. Data are first preprocessed, missing data are filtered out, then the signal is band-pass filtered and segmented. Features are extracted using two different approaches: classical time and frequency features, and wavelet scattering features; later, the features are fed into different classifiers, which are evaluated with a 10-fold cross validation for their accuracy and F1-score performance.
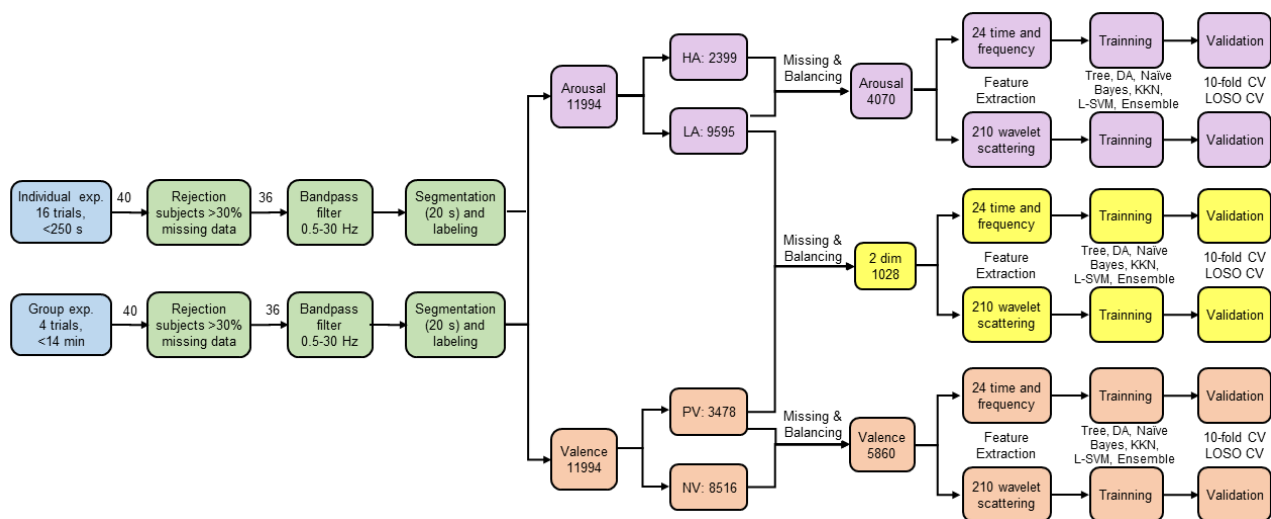
Figure 3. Block diagram of the algorithm used for classification. PV—positive valence; NV—negative valence; HA—high arousal; LA—low arousal.

2.3. Data Processing

In this work, we used the recordings of the short and the long experiments together. The database was first preprocessed in search of missing data and the subjects with more than 30% missing ECG measurements or without annotations were omitted. The ECG signal was then filtered with a band-pass Butterworth filter between 0.5 and 30 Hz. The self and external annotation scores for the valence dimension range from -1 to 1 and were converted to categorical values (positive or negative) using 0 as the threshold. Similarly, the annotations scores for the arousal dimension range from 0 to 9 and were converted to categorical values (high or low), with a threshold of 5. After the conversion, Fisher's exact test [31] was performed between self and external labels for short experiments, as there is no self-annotation for the long experiment. The test rejected the null hypothesis at the 5% significance level, so there is an association between the external and self-annotations. Based on this result, the higher time resolution of the external annotations, and the availability of labels for both experiments, the classification was performed using the external annotations as targets.

Since the external annotations were made at 20-s time windows, we divided each ECG signal into segments of that length and assigned them to their respective label. Each class (positive/negative valence and high/low arousal) contains different amounts of data. For example, from the 8610 segments of the long experiment, 29% present positive valence and 71% negative valence, and 20% present high arousal and 80% low arousal. Given this imbalance, classifying all the segments as negative valence and low arousal would give 71% and 80% accuracy in valence and arousal, respectively [32]. There are different methods to avoid the undesirable effects of an imbalanced dataset, namely, data reduction, data augmentation, undersampling, oversampling, resampling, and cross-validation [33]. In this work, a reduction of the number of data from the overrepresented classes was performed; therefore, the number of elements of the smallest group is preserved and random samples are discarded from the classes with more elements until the same amount of data is achieved in all groups.

2.4. Feature Extraction

To classify the segments in the dimensions of arousal and valence with classical machine learning algorithms, characteristics of the measured signal must be extracted. In this work, we considered classical time- and frequency-based features and compared them with wavelet scattering features, as detailed below.

2.4.1. Time Domain Features

Commonly extracted features in emotion recognition include time domain features [7]. In this study, 17 features were extracted, corresponding to the root mean square successive differences (RMSSD) in interbeat intervals (IBI); the proportion of total IBI that are longer than 20 ms and 50 ms (pNN20, pNN50); 2 Poincare coefficients [34]; and 6 statistical parameters of heart rate (HR) and heart rate variability (HRV), namely, mean, standard deviation, skewness, kurtosis of the raw signal over time, and number of times the signal value is above/below the mean $\pm$ 1 standard deviation.

2.4.2. Frequency Domain Features

Other commonly extracted features are related to the transformation of the signal into the frequency domain and the power of the signal associated to different frequency bands: very low frequency (VLF), low frequency (LF), mid frequency (MF), and high frequency (HF). Several studies indicate that there is some correlation between the ratio of LF/HF power and the sympathetic and parasympathetic activity of the nervous system [35,36]; therefore, 7 frequency domain features were extracted: VLF, LF, MF, HF, LF/HF ratio, power spectral entropy, sample entropy, and Shannon entropy [27].

2.4.3. Wavelet Scattering Features

The extraction of features in the time and frequency domains might not be enough to capture the detail and variability present in the ECG signal during emotional state changes; therefore, we employed mathematical tools based on the wavelet transform to extract more complex features [37]. The wavelet scattering algorithm subdivides each signal in a determined number of scattering windows and extracts features from them using wavelet transformations. Then, each scattered window is classified independently and the classification of the original segment is performed based on a uniform weighting (voting) of the classification of each scattered window. The wavelet scattering algorithm consists of a three-stage iterative transformation of the signal: wavelet convolution, modulation, and filtering. This architecture is similar to a Convolutional Neural Network (CNN) with the difference that the convolution filters are not learned but, instead, they are predefined wavelet functions. The scattering coefficients are intended to have low variance within a class and high variance across classes. Moreover, they are insensitive to input translations on an invariance scale and have some desirable properties such as multiscale contractions, linearization of hierarchical symmetries, and produce sparse representations of data [38–41]. The algorithm used in this work is the wavelet scattering function implemented in the MATLAB Signal Processing Toolbox. The function was set with an invariance scale of 20 s, with the default parameters for filters and wavelets, resulting in two filter banks with 8 and 1 wavelets per octave, respectively.

2.5. Dimensionality Reduction

The wavelet scattering method generates a large number of features, so it could be convenient to reduce the number of features because smaller datasets are easier to explore and make analyzing data much easier and faster for machine learning algorithms. In this work, Principal Component Analysis (PCA) was used to reduce the dimensionality by creating linear combinations of the original features and selecting subsets that capture as much of the information as possible [42]. The PCA option of the MATLAB Classification Learner App that allows selecting the number of desired new features was used. The performance of several classifiers with different numbers of PCA obtained features was later assessed.

2.6. Classification Methods

To build and evaluate the performance of different classification models, the MATLAB Classification Learner App was used, allowing quick training, testing, and validation of several classifiers. The methods used were Linear Discriminant Analysis (LDA) [43], decision trees (DT) with a maximum number of splits of 20 [44], Kernel Naive Bayes [45],

KNN with K = 10 [46], linear Support Vector Machines (SVM) [47], and Ensemble Bagged Tree classifiers [48] with default parameters for screening. The previously obtained features were used as inputs and the labels of each segment or scattered window as target. Later, the classifiers with the best performance in terms of accuracy and F1-score were tested with PCA-obtained features.

2.7. Validation Methods

Given the large amount of data, k-fold cross-validation [49] with k = 10 was used. In k-fold cross-validation, the dataset is first randomly divided into k disjoint folds with approximately the same number of samples, and then every fold in turn plays the role for testing the model induced from the other k-1 folds. However, if we want to build a model that can generalize to new subjects, k-fold cross validation can be this overly optimistic since records from the same subject are present in both training and test sets [50]. Leave-one-subject-out cross-validation (LOSO CV) prevents this by leaving out all the data from a given subject for testing while using the rest of the data for training the model and repeating the process for every subject [51]. Therefore, LOSO CV was also performed, as reported in other works [7].

3. Results

3.1. Data Processing

Participants with missing data and incomplete measurements were removed. Participant 9 was removed for having less than 70% of the total data and three more subjects (participants 8, 24, 28) for having missing annotations or no data from the long experiment.

After dividing each signal into 20-s segments, 94 times windows were obtained for each of the 36 subjects in the short experiment, leading to 3384 segments. Moreover, 246 times windows were obtained for each subject in the long experiment leading to 8610 segments. Therefore, considering both short and long videos, a total of 11,994 segments were obtained. After deleting some segments with missing data and balancing the groups for each experiment, 5860 valence (50% positive and 50% negative) and 4070 arousal segments (50% high and 50% low) were obtained for the short and long experiments together. When using both dimensions (four classes), the dataset was left with 1028 segments, equally distributed for each class (257 segments each).

3.2. Extracted Features

From each ECG segment, 24 times and frequency domain features were extracted. Additionally, the wavelet scattering algorithm set with an invariance scale of 20 s subdivides each signal into 5 scattering windows (so there are 5 times more scattered windows than original samples) and extracts a vector of 210 scattering coefficients for each of the 5 scattered windows of each signal. The coefficients from each scattered window are used as separate inputs, assuming that each scattered window inherits the label of its original signal. After classification, the outputs for each of the 5 scattered windows are combined in an equal-weighted voting to classify the original window.

3.3. Classification in One Dimension

The results below summarize the performance of classifiers with the two types of features, both for one dimension (valence or arousal) with two classes (positive and negative for valence, high and low for arousal) and for two dimensions (valence and arousal) with four classes I–IV (see Figure 1).

A first screening of classifiers was performed across each dimension using default parameters, considering time and frequency features and wavelet scattering features with scattered windows separately. Table 1 summarizes the overall performance (accuracy and F1-score) of different classifiers. From the results, it can be observed that the mean performance of the classifiers using wavelet scattering features is higher than using time and frequency features for both valence and arousal. The classifiers that perform best with

wavelet scattering are Ensemble (accuracy: 89.3% valence and 89.1% arousal) and KNN (82.7% valence and 81.2% arousal), followed by SVM, Discriminant Analysis, Decision Tree, and Naïve Bayes.

Table 1. Classifier screening in one dimension using short and long experiment data.

Classifier	Valence				Arousal			
	Time and Freq. Features		Wavelet Features		Time and Freq. Features		Wavelet Features	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Decision Tree	0.592	0.560	0.631	0.654	0.606	0.595	0.642	0.627
Discriminant Analysis	0.590	0.575	0.622	0.610	0.593	0.558	0.661	0.652
Naïve Bayes	0.526	0.381	0.602	0.614	0.577	0.509	0.618	0.615
KNN	0.566	0.603	0.827	0.834	0.572	0.601	0.812	0.821
SVM	0.597	0.556	0.629	0.612	0.584	0.499	0.667	0.654
Ensemble	0.588	0.593	0.893	0.896	0.587	0.595	0.891	0.894
Mean	0.577	0.545	0.701	0.703	0.587	0.559	0.715	0.711

After screening, the best performance classifiers (accuracy and F1-score over 0.80) with default parameters and wavelet scattering features were further analyzed. Figure 4a,b show the confusion matrix for the 10-NN classifier, for the valence and arousal dimensions, respectively. The overall accuracy is 82.7% in the valence dimension and 81.2% in the arousal dimension.
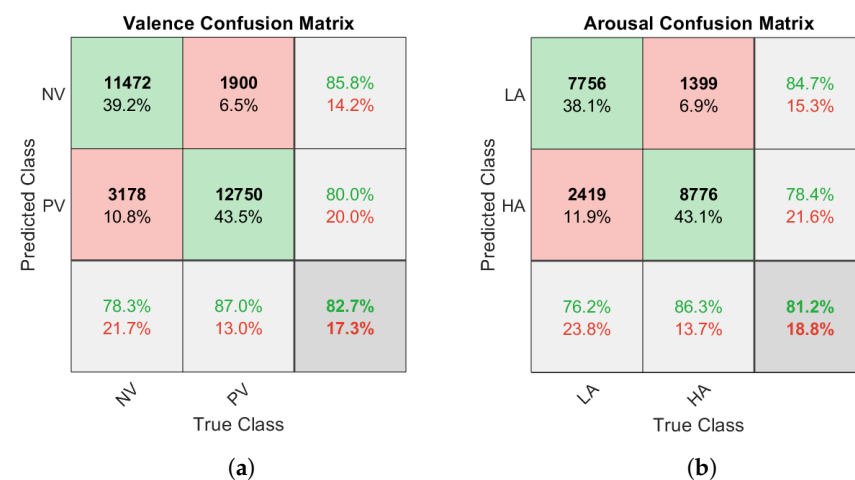

Figure 4. Confusion matrices for the 10-NN classifier in the valence (**a**) and arousal (**b**) dimensions.

Figure 5a,b show the confusion matrices for the Ensemble Bagged Tree classifier for the valence and arousal dimensions, respectively. In this case, the overall accuracy is 89.4% for the valence and 89.2% for the arousal dimension.
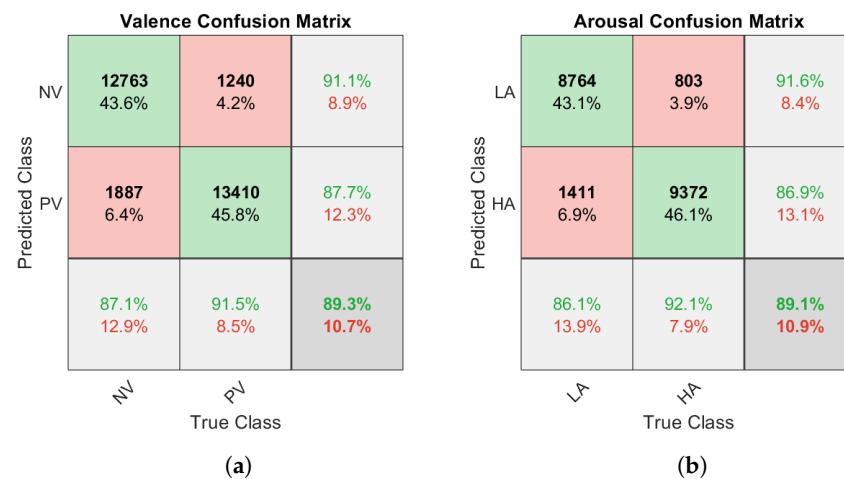
Figure 5. Confusion matrix for the Ensemble Bagged Tree classifier in the valence (**a**) and arousal (**b**) dimensions.

3.4. Classification in Two Dimensions

A similar procedure was applied to classify the signals in two dimensions (four classes) using data from short and long experiments. A total of 24 times and frequency features and 210 wavelet scattering features were considered. Table 2 presents the classifier screening result, showing that the best classifiers are Ensemble and KNN. KNN has significantly less accuracy (73.6%) than in the case of two classes (average 82.0%), while the Ensemble Bagged Tree classifier maintains a high accuracy (88.9%).

Table 2. Classifier screening, classifying in two dimensions.

Classifier	Time and Freq. Features		Wavelet Features	
	Acc.	F1	Acc.	F1
Decision Tree	0.690	0.673	0.432	0.673
Discriminant Analysis	0.361	0.357	0.527	0.357
Naïve Bayes	0.276	0.193	0.267	0.193
KKN	0.351	0.347	0.736	0.347
L-SVM	0.591	0.577	0.519	0.577
Ensemble	0.426	0.421	0.889	0.421
Mean	0.449	0.428	0.562	0.428

Additionally, PCA was used to obtain 50, 100, 150, and 200 linear combinations of the 210 original features, which were then fed to the classifiers. The results of the Ensemble classifier using PCA and the wavelet scattered features for four classes using short and long experiments are presented in Table 3. Accuracy only decreased from 89.0% to 84.4% when reducing the number of components from 210 to 50.

Table 3. PCA dimensionality reduction.

Number of Components	Accuracy
210	89.0%
200	84.5%
150	85.0%
100	85.6%
50	84.4%

Table 4 presents the detailed per-class results for precision, recall, and F1-score for the Ensemble classifier, using 210 features (PCA disabled) and classifying four classes.

Table 4. Per class performance, Ensemble Bagged Tree classifier.

	Precision	Recall	F1-Score
HA/PV	0.853	0.905	0.879
LA/PV	0.895	0.851	0.872
HA/NV	0.883	0.926	0.904
LA/NV	0.916	0.861	0.888
Mean	0.887	0.886	0.886

As the scattering algorithm produces a larger number of scattering windows for each original segment, the latter is classified based on a majority vote count of the class assigned to each of its scattered windows. This can result in higher accuracy, but it is also possible that there is no unique major vote class. This is reflected in the row "No Unique" in the predicted class. Figure 6 shows the confusion matrix and accuracy (95.3%) of original segments classification based on the predicted classes from each scattered window using the Ensemble classifier.

Figure 6. Confusion matrix using two dimensions—valence and arousal. No Unique—classifier could not decide due to a draw of votes for the scattered windows of one segment. The results were obtained using short and long experiment data, with 210 wavelet features, the Ensemble Bagged Tree classifier, and majority vote of the scattered windows of each segment.

3.5. Short Experiment Result Comparison

In order to make the results comparable with other studies, the methods proposed in this work were also used with the short experiments from the AMIGOS dataset only. The Ensemble classifier achieved an accuracy of 0.902 (arousal) and 0.904 (classifier), and the KNN classifier achieved an accuracy of 0.888 (arousal) and 0.889 (valence), using 10-fold cross-validation and the wavelet-scattering-extracted features. Table 5 summarizes the comparison of these results to other published studies using the same dataset and different classification algorithms.

Table 5. Comparison of the results from previous studies that use AMIGOS dataset.

Classifier	Validation	Arousal		Valence	
		Acc.	F1	Acc.	F1
Naive Bayes [7]	LOSO	0.69	0.545	-	0.551
Nearest Neighbors [25]	N/R	-	0.66	0.58	0.57
Linear Discriminant [25]	N/R	0.72	0.63	0.67	0.65
Linear Support Vector [25]	N/R	0.68	0.60	0.61	0.55
Multilayer Perceptron [25]	N/R	0.68	0.59	0.61	0.51
AdaBoost [25]	N/R	0.70	0.66	0.61	0.58
Random Forest [25]	N/R	0.68	0.67	0.59	0.59
DCNN [25]	N/R	0.81	0.76	0.71	0.68
CNN w/o self-supervised [26]	10 Fold	0.837	0.828	0.809	0.808
CNN self-supervised [26]	10 Fold	0.858	0.851	0.840	0.837
XGBoost [27]	LOSO	-	0.561	-	0.633
Bayesian DL [23]	10 Fold	-	-	0.90	0.86
KNN [24]	10 Fold	0.558	-	0.597	-
KNN (this work)	10 Fold	0.888	0.821	0.889	0.834
Ensemble (this work)	10 Fold	0.902	0.894	0.904	0.896

Acc—accuracy; F1—F1-score; LOSO—Leave one subject out; 10 Fold—10-fold cross validation; N/R—not reported.

In addition to k-fold validation, subject generalization of the model was assessed with LOSO CV. The Ensemble Classifier achieved an accuracy of 0.819 (arousal) and 0.837 (valence), outperforming the KNN classifier that reach an accuracy of 0.623 (arousal) and 0.586 (valence). These results indicate how well the model would perform with data from a new subject.

4. Discussion

4.1. Algorithm Performance for Valence and Arousal

Classifier screening shows that classifiers using wavelet-extracted features outperform the ones employing features extracted with traditional time and frequency metrics, both for valence and arousal classification. The increased performance can be explained by the wavelet scattering method's ability to decompose and extract features from different scales of the signal, providing time and frequency resolution, and allowing a classifier to capture the difference between classes [52]. Regarding one-dimension classification, we found slightly better performance for arousal than for valence when considering the same type of features and classifier, similar to other works [6,53]. For both arousal and valence classification, ensemble and 10-KNN classifiers presented better accuracy than the other classifiers. The good performance of the ensemble classifier might be due to its characteristics, since the selected ensemble bagged tree classifier is a boosted aggregation algorithm that creates a collection of decision trees, which are combined in order to reduce the variance of classification result. On the other side, KNN makes clusters of data from the feature vector, which may be close enough for similar classes. Ensemble classifier has previously been reported to have high accuracy for ECG arrhythmia classification [54].

4.2. Classification Performance in Two Dimensions

When using two dimensions simultaneously (four classes), the performance of the KNN classifier dropped, possibly due to the similarity or low distance between features of each class. On the other side, the Ensemble Tree classifier accuracy was similar to that achieved in one-dimension classification, possibly due to the classifier being deep enough to capture the differences between classes of the scattering features of each signal.

PCA can be used to reduce the dimensionality of the input vector, but it adds an additional step in the algorithm and tends to degrade classifier performance since reducing the number of variables of a dataset naturally comes at the expense of accuracy. In this

work, the accuracy was only reduced from 89.0% to 84.4% when using all 210 features, with only 50 components showing a reasonable trade of accuracy for simplicity.

4.3. Scattering Window Classification Performance

The classification using scattering features showed high accuracy for every scenario. This result can be explained by wavelet scattering's ability to represent time and frequency domain features at different scales with each scattered window [52], and then to be fed to a classifier that can take large feature vectors to learn and generate the output classification. Moreover, the original signal segment classification based on the majority vote of the assigned class to each scattered window may inherit the high accuracy of scattered window classification. This method yields a better classifier performance with the drawback of having a tie in some cases as the algorithm would not be able to determine the final class. However, this can be avoided by selecting an odd number of scattered windows and using only two classes.

4.4. Short Experiment Result Comparison

Regarding short experiment results, the algorithm also achieves high performance, but lower than the performance with all the datasets. This was somehow expected since the short experiment dataset represents less than 30% of the entire dataset, which may not be enough to achieve high accuracy with the applied classifiers. However, the achieved performance is higher than that reported by other studies.

4.5. LOSO Validation

Finally, subject-independent classification using leave-one-subject-out (LOSO) validation yielded lower performance than 10-fold cross-validation, dropping to 58% for KNN and 82% for Ensemble. This result shows that the Ensemble classifier is more robust and may generalize better for new subjects and samples.

5. Conclusions and Future Work

The extraction of features of the ECG signal by means of wavelet scattering has been shown to improve the performance of classic machine learning algorithms to classify emotions in the arousal and valence dimensions compared to features in the time and frequency domain. The wavelet scattering algorithm allows the ECG signal to be analyzed at different temporal space scales and simultaneously in the time and frequency domains. This allows increasing the separability and differentiability of the signals and patterns of two different classes. The method was validated using the AMIGOS database to classify emotions in two dimensions: arousal and valence. This study demonstrates a higher overall performance than previous works. The classifier comparison shows that the Ensemble Bagged Tree classifier can be a good choice for emotion classification from ECG signals.

Several systems have been proposed to visualize mind states of a human subject based on EEG signals [55]. As a future work, we intend to build a system capable of discriminating emotions in real time using ECG signals. Currently, we are in the process of building a wearable device consisting of first-layer clothing with embedded electrodes for ECG signal monitoring (https://www.auradt.com/productos/primera-capa, accessed on 5 May 2021). For increased comfort and long-term applications, we eliminated the need of conductive gels and skin preparation by integrating dry electrodes in a first-layer t-shirt developed by the local textile manufacturer ACTAS-Grupo CLER (see Figure 7).
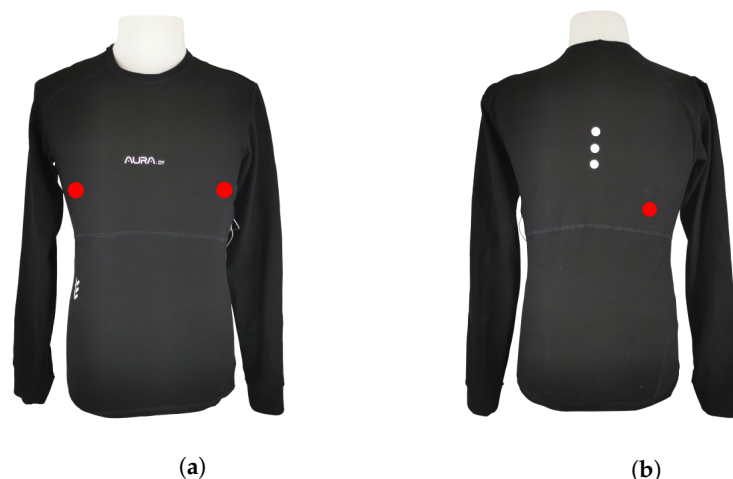
(**a**)   (**b**)

Figure 7. First-layer t-shirt with the location of the electrodes marked in red: (**a**) front view (**b**) rear view.

The prototype is capable of measuring, storing, and streaming data, and we are planning to incorporate real-time signal processing and implement the emotion recognition algorithm in the near future. The major parts and components of the system with a ground station to process and classify the ECG can be represented in the high-level block diagram shown in Figure 8.
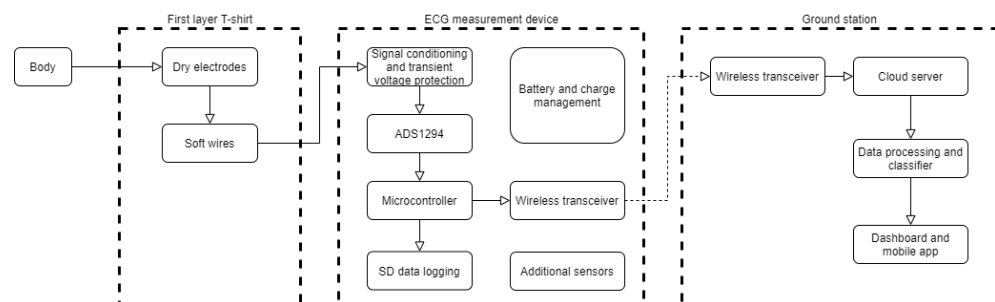

Figure 8. Monitoring system diagram.

Wavelet analysis has been successfully applied to other biopotential signals such as EMG [56], thus, the proposed methodology is also a promising alternative for EMG-based pattern recognition problems [57].

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/index.html, accessed on 6 May 2019.

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

References

1. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
2. Basu, S.; Jana, N.; Bag, A.; Mahadevappa M.; Mukherjee, J.; Kumar, S.; Guha, R. Emotion recognition based on physiological signals using valence-arousal model. In Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, India, 21–24 December 2015; pp. 50–55. [CrossRef]

3. Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A Review of Emotion Recognition Using Physiological Signals. *Sensors* **2018**, *18*, 2074. [CrossRef]

4. Goshvarpour, A.; Abbasi, A.; Goshvarpour, A. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biomed. J.* **2017**, *40*, 355–368. [CrossRef]

5. Koelstra, S.; Mühl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]

6. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A Multi-Modal Affective Database for Affect Recognition and Implicit Tagging. *Affect. Comput. IEEE Trans.* **2012**, *3*, 1. [CrossRef]

7. Miranda Correa, J.A.; Abadi, M.K.; Sebe, N.; Patras, I. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans. Affect. Comput.* **2018**. [CrossRef]

8. Kahou, S.E.; Bouthillier, X.; Lamblin, P.; Gulcehre, C.; Michalski, V.; Konda, K.; Jean, S.; Froumenty, P.; Dauphin, Y.; Boulanger-Lewandowski, N.E.A. EmoNets: Multimodal deep learning approaches for emotion recognition in video. *J. Multimodal User Interfaces* **2015**, *10*, 99–111. [CrossRef]

9. Kossaifi, J.; Tzimiropoulos, G.; Todorovic, S.; Pantic, M. AFEW-VA database for valence and arousal estimation in-the-wild. *Image Vis. Comput.* **2017**, *65*, 23–36. [CrossRef]

10. Dhall, A.; Goecke, R.; Joshi, J.; Hoey, J.; Gedeon, T. Emotiw 2016: Video and group-level emotion recognition challenges. In Proceedings of the 18th ACM international conference on multimodal interaction, Tokyo, Japan, 12–16 November 2016; pp. 427–432.

11. Ranganathan, H.; Chakraborty, S.; Panchanathan, S. Multimodal emotion recognition using deep learning architectures. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016. [CrossRef]

12. Fan, Y.; Lu, X.; Li, D.; Liu, Y. Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In Proceedings of the 18th ACM International Conference on Multimodal Interaction-ICMI 2016, Tokyo, Japan, 12–16 November 2016. [CrossRef]

13. Hu, M.; Wang, H.; Wang, X.; Yang, J.; Wang, R. Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks. *J. Vis. Commun. Image Represent.* **2019**, *59*, 176–185. [CrossRef]

14. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010, pp. 94–101.

15. Valstar, M.; Pantic, M. Induced disgust, happiness and surprise: An addition to the mmi facial expression database. In Proceedings of the 3rd International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect, Paris, France, 3–6 September 2010; p. 65.

16. Zheng, W.L.; Dong, B.N.; Lu, B.L. Multimodal emotion recognition using EEG and eye tracking data. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014. [CrossRef]

17. Alhagry, S.; Aly, A.; El-Khoribi, R. Emotion Recognition based on EEG using LSTM Recurrent Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 355–358. [CrossRef]

18. Bălan, O.; Moise, G.; Petrescu, L.; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. Emotion Classification Based on Biophysical Signals and Machine Learning Techniques. *Symmetry* **2020**, *12*, 21. [CrossRef]

19. Paszkiel, S. Characteristics of question of blind source separation using Moore-Penrose pseudoinversion for reconstruction of EEG signal. In *International Conference on Automation*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 393–400.

20. Hsu, Y.; Wang, J.; Chiang, W.; Hung, C. Automatic ECG-Based Emotion Recognition in Music Listening. *IEEE Trans. Affect. Comput.* **2020**, *11*, 85–99. [CrossRef]

21. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef]

22. Katsigiannis, S.; Ramzan, N. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 98–107. [CrossRef]

23. Harper, R.; Southern, J. A Bayesian Deep Learning Framework for End-To-End Prediction of Emotion from Heartbeat. *IEEE Trans. Affect. Comput.* **2020**. [CrossRef]

24. Ferdinando, H.; Seppänen, T.; Alasaarela, E. Comparing features from ECG pattern and HRV analysis for emotion recognition system. In Proceedings of the 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Chiang Mai, Thailand, 5–7 October 2016; pp. 1–6. [CrossRef]

25. Santamaria-Granados, L.; Munoz-Organero, M.; Ramirez-Gonzalez, G.; Abdulhay, E.; Arunkumar, N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS). *IEEE Access* **2019**, *7*, 57–67. [CrossRef]

26. Sarkar, P.; Etemad, A. Self-Supervised Learning for ECG-Based Emotion Recognition. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3217–3221.

27. Tung, K.; Liu, P.K.; Chuang, Y.C.; Wang, S.H.; Wu, A.Y. Entropy-Assisted Multi-Modal Emotion Recognition Framework Based on Physiological Signals. In Proceedings of the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 3–6 December 2018; pp. 22–26.

28. Kolodyazhniy, V.; Kreibig, S.; Gross, J.; Roth, W.; Wilhelm, F. An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology* **2011**, *48*, 908–922. [CrossRef]

29. Sarkar, P.; Etemad, A. Self-supervised ECG Representation Learning for Emotion Recognition. *IEEE Trans. Affect. Comput.* **2020**. [CrossRef]

30. Siddharth.; Jung, T.; Sejnowski, T.J. Utilizing Deep Learning Towards Multi-modal Bio-sensing and Vision-based Affective Computing. *IEEE Trans. Affect. Comput.* **2019**. [CrossRef]

31. Kim, H.Y. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor. Dent. Endod.* **2017**, *42*, 152. [CrossRef]

32. Frésard, M.E.; Erices, R.; Bravo, M.L.; Cuello, M.; Owen, G.I.; Ibáñez, C.; Rodriguez-Fernandez, M. Multi-objective optimization for personalized prediction of venous thromboembolism in ovarian cancer patients. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 1500–1508. [CrossRef]

33. Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Handling imbalanced datasets: A review. *GESTS Int. Trans. Comput. Sci. Eng.* **2005**, *30*, 25–36.

34. Piskorski, J.; Guzik, P. Filtering Poincareplots. *Comput. Methods Sci. Technol.* **2005**, *11*, 39–48. [CrossRef]

35. Luo, D.; Pan, W.; Li, Y.; Feng, K.; Liu, G. The interaction analysis between the sympathetic and parasympathetic systems in CHF by using transfer entropy method. *Entropy* **2018**, *20*, 795. [CrossRef]

36. Strüven, A.; Holzapfel, C.; Stremmel, C.; Brunner, S. Obesity, Nutrition and Heart Rate Variability. *Int. J. Mol. Sci.* **2021**, *22*, 4215. [CrossRef] [PubMed]

37. Daubechies, I.; Heil, C. Ten Lectures on Wavelets. *Comput. Phys.* **1992**, *6*, 697. [CrossRef]

38. Mallat, S. Group Invariant Scattering. *Commun. Pure Appl. Math.* **2012**, *65*, 1331–1398. [CrossRef]

39. Anden, J.; Mallat, S. Deep Scattering Spectrum. *IEEE Trans. Signal Process.* **2014**, *62*, 4114–4128. [CrossRef]

40. Bruna, J.; Mallat, S. Invariant Scattering Convolution Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1872–1886. [CrossRef]

41. Mallat, S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]

42. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013; Volume 112.

43. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* **2017**, *30*, 169–190. [CrossRef]

44. Vishwanath, M.; Jafarlou, S.; Shin, I.; Lim, M.M.; Dutt, N.; Rahmani, A.M.; Cao, H. Investigation of machine learning approaches for traumatic brain injury classification via EEG assessment in mice. *Sensors* **2020**, *20*, 2027. [CrossRef]

45. Khanna, D.; Sharma, A. Kernel-based naive bayes classifier for medical predictions. In *Intelligent Engineering Informatics*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 91–101.

46. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Cheng, D. Learning k for knn classification. *ACM Trans. Intell. Syst. Technol. (TIST)* **2017**, *8*, 1–19.

47. Ghaddar, B.; Naoum-Sawaya, J. High dimensional data classification and feature selection using support vector machines. *Eur. J. Oper. Res.* **2018**, *265*, 993–1004. [CrossRef]

48. Truong, X.L.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, X.Q.; Do, T.H.; Tien Bui, D.; Lee, S. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl. Sci.* **2018**, *8*, 1046. [CrossRef]

49. Wong, T.T.; Yeh, P.Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [CrossRef]

50. Saeb, S.; Lonini, L.; Jayaraman, A.; Mohr, D.C.; Kording, K.P. The need to approximate the use-case in clinical machine learning. *Gigascience* **2017**, *6*, gix019. [CrossRef]

51. Gholamiangonabadi, D.; Kiselov, N.; Grolinger, K. Deep Neural Networks for Human Activity Recognition With Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection. *IEEE Access* **2020**, *8*, 133982–133994. [CrossRef]

52. Liu, Z.; Yao, G.; Zhang, Q.; Zhang, J.; Zeng, X. Wavelet Scattering Transform for ECG Beat Classification. *Comput. Math. Methods Med.* **2020**, *2020*, 3215681. [CrossRef] [PubMed]

53. Ferdinando, H.; Ye, L.; Seppänen, T.; Alasaarela, E. Emotion Recognition by Heart Rate Variability. *Aust. J. Basic Appl. Sci.* **2014**, *8*, 50–55.

54. Mert, A.; Kilic, N.; Akan, A. ECG signal classification using ensemble decision tree. *J. Trends. Dev. Mach. Assoc. Technol.* **2012**, *16*, 179–182.

55. Paszkiel, S.; Hunek, W.; Shylenko, A. Project and Simulation of a Portable Device for Measuring Bioelectrical Signals from the Brain for States Consciousness Verification with Visualization on LEDs. In *International Conference on Automation*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 25–35.

56. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. Application of wavelet analysis in EMG feature extraction for pattern classification. *Meas. Sci. Rev.* **2011**, *11*, 45. [CrossRef]

57. Arozi, M.; Caesarendra, W.; Ariyanto, M.; Munadi, M.; Setiawan, J.D.; Glowacz, A. Pattern recognition of single-channel sEMG signal using PCA and ANN method to classify nine hand movements. *Symmetry* **2020**, *12*, 541. [CrossRef]