

Article

A Comprehensive Evaluation of Features and Simple Machine Learning Algorithms for Electroencephalographic-Based Emotion Recognition

Mayra Álvarez-Jiménez *, Tania Calle-Jimenez and Myriam Hernández-Álvarez 

Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Ladrón de Guevara E11-25 y Andalucía, Edificio de Sistemas, Quito 170525, Ecuador; tania.calle@epn.edu.ec (T.C.-J.); myriam.hernandez@epn.edu.ec (M.H.-Á.)

* Correspondence: mayra.alvarez@epn.edu.ec

Abstract: The study of electroencephalographic (EEG) signals has gained popularity in recent years because they are unlikely to intentionally fake brain activity. However, the reliability of the results is still subject to various noise sources and potential inaccuracies inherent to the acquisition process. Analyzing these signals involves three main processes: feature extraction, feature selection, and classification. The present study extensively evaluates feature sets across domains and their impact on emotion recognition. Feature selection improves results across the different domains. Additionally, hybrid models combining features from various domains offer a superior performance when applying the public DEAP dataset for emotion classification using EEG signals. Time, frequency, time–frequency, and spatial domain attributes and their combinations were analyzed. The effectiveness of the input vectors for the classifiers was validated using SVM, KNN, and ANN, which are simple classification algorithms selected for their widespread use and better performance in the state of the art. The use of simple machine learning algorithms makes the findings particularly valuable for real-time emotion recognition applications where the computational resources and processing time are often limited. After the analysis stage, feature vector combinations were proposed to identify emotions in four quadrants of the valence–arousal representation space using the DEAP dataset. This research achieved a classification accuracy of 96% using hybrid features in the four domains and the ANN classifier. A lower computational cost was obtained in the frequency domain.

Keywords: DEAP dataset; feature extraction; feature selection; frequency domain; time domain; time–frequency domain; location or spatial domain; emotion recognition; EEG signals



Citation: Álvarez-Jiménez, M.; Calle-Jimenez, T.; Hernández-Álvarez, M. A Comprehensive Evaluation of Features and Simple Machine Learning Algorithms for Electroencephalographic-Based Emotion Recognition. *Appl. Sci.* **2024**, *14*, 2228. <https://doi.org/10.3390/app14062228>

Academic Editors: Tomislav Stipančić and Duska Rosenberg

Received: 24 September 2023

Revised: 30 November 2023

Accepted: 6 December 2023

Published: 7 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain–computer interfaces (BCIs) provide a means to develop interactions between humans and computers [1]. Within BCIs, affective BCIs (aBCIs) are identified, and aim to detect users’ emotional states using electroencephalographic (EEG) signals [2]. BCI devices record brain responses using various invasive and non-invasive acquisition techniques [3]. EEG monitored via external BCIs is non-invasive because it does not damage tissues. The use of BCIs for monitoring EEG is one of the preferred methods for emotion recognition due to its speed, low cost [2], and strong connection with emotional states [4]. Therefore, emotions can be recognized through the classification of incoming EEG signals.

Brain signal studies have gained importance in recent years, because they are unlikely to intentionally fake brain activity [5]. However, the reliability of the results is still subject to various noise sources and potential inaccuracies inherent to the acquisition process. In general, the implementation of EEG-based emotion recognition has a wide range of potential applications in entertainment, therapy, security, business, and education.

Emotion recognition primarily relies on combining features and models obtained from classifiers [6]. Thus, evaluating different types of characteristics is essential and necessary

for developing more accurate emotion recognition systems. However, the challenge lies in determining which features are suitable for ensuring acceptable accuracy in terms of emotion recognition. There is no consensus on the EEG features for emotion detection to improve classification accuracy. Consequently, EEG emotion recognition has become an active topic in affective computing.

In this work, we review different feature extraction methods in the time domain, frequency domain, time–frequency, and location (spatial) domains to obtain input vectors for classifiers such as the support vector machine (SVM), k-nearest neighbor (k-NN), and artificial neural networks (ANNs). The data to be analyzed come from the public access dataset DEAP [7,8], which correlate emotional states with EEG signals using audio and images as stimuli, acquiring their data through BCI devices.

Based on the EEG data from the DEAP dataset, sixteen feature vectors are constructed, some within a specific domain and others using multi-domain characteristics. These features are input to simple machine learning classifiers (SVM, k-NN, ANN). The resulting models are evaluated to define attributes that provide more information for classification and, thus, yield a better performance.

The rest of the article is structured as follows: in Section 2, several related works found in the literature are presented. Section 3 explains the materials and methods employed in this research, including emotion models, DEAP dataset description, feature extraction in time, frequency, time–frequency, and location (spatial) domains, and classifiers. Section 4 provides details of the implementation and evaluation of the emotion recognition system in the four quadrants of valence–arousal (VA). Section 5 presents a discussion emerging from the results of the present study and various works in the literature. Finally, Section 6 outlines the conclusions and directions for future work.

2. Related Work

Several studies on emotion recognition have been found in the literature, addressing classification based on the two-dimensional model categorizing valence/arousal (VA) in terms of the categories derived from EEG signals from the DEAP dataset.

Some studies in the literature begin with the classification of the categories of high/low valence (HV and LV) and high/low arousal (HA and LA) for emotion recognition. For example, Nawaz et al. studied emotion recognition by implementing features such as entropy, statistics, fractal dimension, power functions, and wavelet, which correspond to time, frequency, and time–frequency, respectively. The authors achieved average accuracy results in the valence dimension of 75.06% with the k-NN classifier and 77.62% with the SVM classifier. In comparison, in arousal, they obtained average results of 74.71% and 78.96% for k-NN and SVM, respectively. The authors conclude that the time domain statistical features better distinguish between low and high valence and arousal levels [9]. The authors of [10] propose selecting locally robust EEG features (LRFS) for emotion recognition. In feature extraction, they employ methods such as power ratio and power difference, statistical techniques like variance, kurtosis, and skewness, and variables like the zero-crossing rate and Shannon entropy; using these methods, they obtained a total of 364 EEG features. The authors achieved average accuracy results in Arousal of 60.05%, 65.10%, and 61.85%. The average accuracy in valence was 65.13%, 67.97%, and 65.60% using the k-NN, SVM, and ANN classifiers, respectively. The authors implement features in time and frequency domains for feature extraction without establishing which features improve classifier performance. In [11], the authors propose emotion recognition by implementing the spectral power method for individual channels and spectral power asymmetry among the 12 pairs of electrodes for feature extraction from the five frequency bands (delta, theta, alpha, beta, and gamma) of the EEG signals. These methods correspond to the frequency domain. The authors obtained average accuracy values of 65.0% for Arousal and 63.3% for Valence using the SVM classifier with a radial basis function (RBF) kernel. Using the naive Bayes classifier, they achieved an average accuracy of 65.6% in arousal and 68.0% in valence. However, their approach does not determine which feature extraction methods better

classify emotional states. In [12], the authors propose feature extraction from EEG signals by implementing statistical procedures (median, standard deviation, kurtosis), Hjorth parameters, the spectral density of frequency bands, fractal dimension, and wavelets.

Additionally, they incorporate feature selection using the minimum redundancy maximum relevance (mRMR) method for emotional state recognition in the VA space. The authors report accuracy results of 60.72% in arousal and 62.39% in valence with the SVM classifier, using an RBF kernel. The authors implement feature extraction methods in time, frequency, and time–frequency without concluding which features yield the best results.

Other authors perform emotion recognition based on the four quadrants of the two-dimensional model; that is, they classify the quadrants of the VA model into high arousal–high valence (HAHV), high arousal–low valence (HALV), low arousal–low valence (LALV), and low arousal–high valence (LAHV), which implies greater complexity, given the number of classes to classify. For example, the authors of [13] propose exploring multi-domain features to identify the attributes that contribute to ranking a higher number of emotion classes. The authors implement feature extraction methods, such as Hjorth parameters and entropy, Fourier transform, RASM, DASM, and discrete wavelet transform and differential entropy. The authors achieved an average accuracy of 65.72% with time-domain features using the SVM classifier with a polynomial kernel. However, they explore a minimal set of feature extraction methods. The authors of [14] propose recognizing four emotion classes in VA, using feature extraction methods in the time and frequency domain and the mRMR method to select significant features. The authors implement statistical methods (mean, standard deviation, mean of absolute values of the first and second difference of the raw and processed signal), discrete wavelet transform, and the wavelet energy and wavelet entropy estimation. They use 15 EEG channels and the SVM classifier, obtaining accuracy percentages of 52.1% HAHV, 49.1% HALV, 49.6% LAHV, and 48.3% LALV, with an overall accuracy of 49.7%. However, the authors do not conclude which features provide better results.

On the other hand, some authors use public access datasets such as DEAP for emotion classification. For example, the authors of [15], for emotion recognition, apply features such as power spectral density (PSD), differential entropy (DE), differential asymmetry (DASM), and rational asymmetry (RASM). They use the mean accuracy rate (mAR) metric, obtaining average results in VA space of 44.07% and 59.09% using SVM and ANN classifiers for the DEAP dataset. The authors of this study use frequency domain methods for feature analysis and individual participants.

Li et al. propose the recognition of three (positive, negative, neutral) and four (VA space) affective states using PSD, DE, and Fourier transform features, obtaining accuracy results of 62.54% in DEAP using the ANN classifier [16].

Several works mentioned in the literature aim to achieve emotion recognition using different feature extraction methods in the time, frequency, time–frequency, and location (spatial) domains; however, their goal is not to determine which features or combination of them provide better results. Very few studies focus on determining which feature extraction methods from the different domains give better results in classifying emotional states. Therefore, there is still no agreement on the EEG features that provide better results to improve classification accuracy.

One of the key contributions of this work lies in studying the recognition of emotional states defined in the four quadrants of the two-dimensional (VA) model, exploring a broader set of feature extraction methods in time, frequency, time–frequency, and location (spatial) domains to obtain input vectors for the SVM, k-NN, and ANN classifiers.

The data to be analyzed come from the public access DEAP dataset that correlates emotional states with EEG signals using audio and image stimuli that acquire data based on BCI. Sixteen input vectors were created in order for the classifiers to evaluate which feature extraction method provides more information and produces a better performance.

3. Materials and Methods

In this section, the proposed emotion recognition process, the methods for feature extraction, and the classification algorithms used in this work are mentioned, as well as general information on the DEAP [7] dataset.

3.1. Emotion Recognition Process

Figure 1 represents the structure of an EEG-based BCI system for emotion recognition [8].

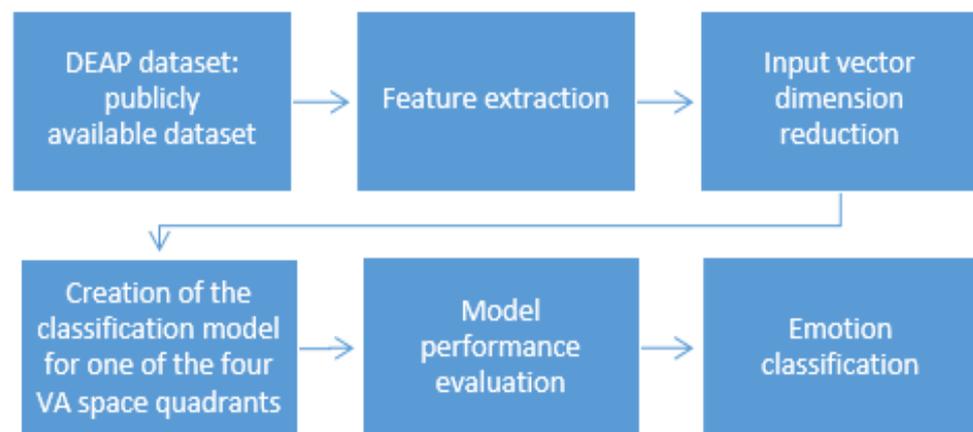


Figure 1. Steps in the proposed emotion recognition process.

This work used preprocessed signals from the DEAP dataset previously labeled with emotions extracted from subjects who self-assessed their emotional states. Subsequently, feature extraction was performed to relate the EEG signals to the different emotional states; the input vector to the classifier contains various combinations of features. Dimension-reduction techniques were tested using a correlation matrix to eliminate attributes that have redundant information that could add noise to the classification. The models for classifying the four quadrants of the VA space were created using simple machine-learning-based classification algorithms. Model evaluation was obtained using accuracy as the primary metric.

3.2. Data Acquisition

We have considered the public dataset commonly used for emotion analysis: the EEG signal set of the DEAP dataset.

DEAP is a dataset containing multiple physiological signals with emotional evaluations, including EEG signals from 32 subjects (16 men and 16 women) while watching 40 one-minute video clips with different emotional tendencies.

The EEG signals are recorded at a sampling frequency of 512 Hz to activate 32 active electrodes according to the international 10–20 system [17]. Each participant evaluates their valence, arousal, dominance, and liking levels using the Self-Assessment Manikin (SAM) [18]. Each data file (s01.dat–s32.dat) contains two matrices:

- Data matrix: $40 \times 40 \times 8064$, where the first 40 elements represent the total number of videos, the second 40 represent the collection of signals from the total number of channels (32 EEG), and 8064 are the experimental data based on the video and sampling sequence (63×128). The first 3 s correspond to the reference data obtained before the experiment. The last 60 s are the information recorded during the experiment.
- Label matrix: 40×4 , where the number of videos used is represented by 40, and 4 is the number of labels describing the affective dimensions: valence, arousal, dominance, and liking, with scores ranging from 1.0 to 9.0 according to the SAM (Self-Assessment Manikin) scale.

According to the two-dimensional emotion model (see Figure 2), the emotions labeled in each video can be classified into 4 quadrants based on the valence–arousal space. Valence is on the x -axis, and arousal is on the y -axis. Therefore, there are four quadrants: high arousal–high valence (HAHV), low arousal–high valence (LAHV), high arousal–low valence (HALV), and low arousal–low valence (LALV). Arousal scales range from passive to active, and valence ranges from negative to positive.

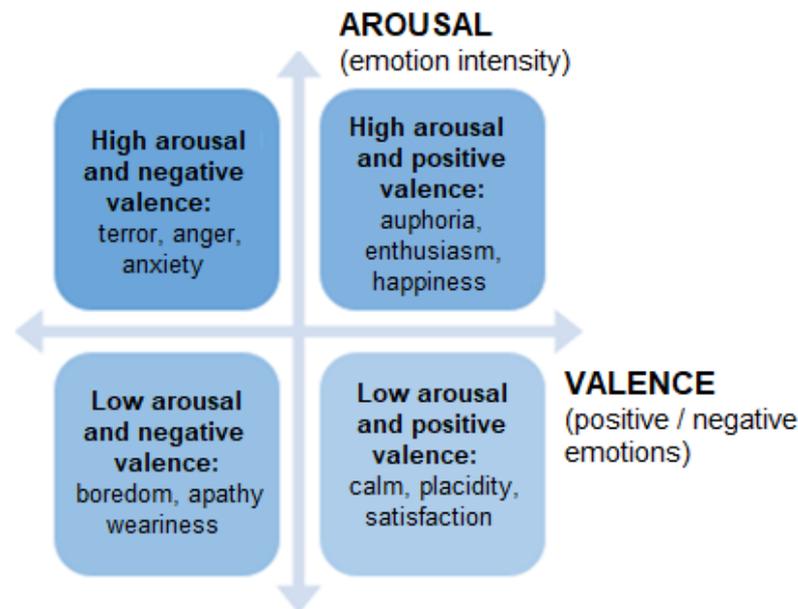


Figure 2. Bidimensional model representation.

In this bi-dimensional representation, the emotions are described using valence, i.e., positive or negative feelings. Arousal refers to the degree of physical or mental activation or energy. High arousal is the sensation that generates high energy, and low arousal is the sensation of low energy. For instance, terror, stress, and anger are negative emotions that produce high arousal. Sadness, depression, lethargy, and fatigue are states with negative valence and low arousal. Euphoria, happiness, and enthusiasm are positive-valence and high-arousal emotions. Calm, placidity, and satisfaction are positive-valence and low-arousal feelings.

In the present work, we evaluate valence and arousal levels from preprocessed EEG signal recordings of the 32 channels at a downsampled sampling frequency of 128 Hz. Each trial contains 8064 data points for the sampling frequency, resulting in a $40 \times 32 \times 8064$ structure corresponding to the number of trials per participant, the number of channels, and the stored signal data for each channel.

The emotions that could be differentiated according to their ubication in a quadrant in the VA space are, for example, terror, anger, anxiety/euphoria, enthusiasm, happiness/boredom, apathy, sadness/tranquility, placidity, and satisfaction. These emotions correspond to various combinations of valence and arousal levels within the VA space, allowing for an understanding of a person’s emotional state.

3.3. Feature Extraction Methods

The feature extraction stage is of particular importance since the quality of the data obtained will directly affect the accuracy of emotion classification. That is, finding informative features of the EEG signals can improve the discrimination capacity between emotions with limited dimensionality [19]. Various common feature extraction methods for BCI systems are based on analyzing signals in the time, frequency, time–frequency, and location (spatial) domains [20].

3.3.1. Time Domain

Most EEG signal acquisition equipment supports time-domain series. These methods mainly consider the geometric characteristics of the EEG signals, leading to a lower loss of information. The time-domain attributes provide valuable information about the statistical patterns in the EEG data.

The time-domain methods included in this work are statistical functions such as the maximum, mean, standard deviation, variance, skewness, kurtosis, and mean of absolute values of a normalized first and second difference. Entropy functions include Shannon entropy, approximate entropy, sample entropy, permutation entropy, and other functions like energy, average power, root mean square, line length, Higuchi fractal dimension, Petrosian fractal dimension, Hjorth parameters, zero crossing, and higher-order crossing.

3.3.2. Frequency Domain

Frequency domain methods transform the EEG signals from the time domain to the frequency domain for analysis. The acquired spectrum is typically decomposed into five sub-bands, delta, theta, alpha, beta, and gamma [4], for feature extraction and the subsequent analysis of EEG signals. The functions implemented from this domain in the present work are the spectral entropy, power ratio, power spectral density, and fast Fourier transform. This domain provides insights into the power spectral density or energy distribution across various frequency bands. Each frequency band is associated with different cognitive states and emotional processes.

3.3.3. Time–Frequency Domain

Time–frequency domain analysis methods have localized analysis capabilities in both the time and frequency domains simultaneously. The functions used in this work are discrete wavelet transform and wavelet entropy. The feature extraction in this domain is powerful, capturing the transient behavior of EEG signals and the evolution of frequency components over time.

3.3.4. Location Domain

Location-domain features would be obtained from voltage measurements at various locations on the scalp and their differences or ratios among left and right sites. Differential asymmetry (DASM) and rational asymmetry (RASM) may capture some aspects of the asymmetry between the two sides. These features are based on the premise that certain cognitive and emotional processes are lateralized, meaning that they are more dominant in one hemisphere of the brain than the other.

Table 1 shows the mathematical expressions and the parameters of the feature extraction methods in the time, frequency, time–frequency, and location domains implemented in this research.

Table 1. Mathematical expressions and a brief description of the feature extraction methods.

| Method | Mathematical Expression | Description |
|---|--|--|
| Time domain | | |
| Maximum (Max) | $(X_i \in C)$ (1) | X_i is the maximum element of C if any other component of that set is less than or equal to X_i . |
| Mean | $mean = \frac{1}{N} \sum_{i=1}^N X_i$ (2) | |
| Standard deviation (Stddev) | $stddev = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - mean)^2}$ (3) | |
| Variance (Var) | $var = \frac{1}{N} \sum_{i=1}^N (X_i - mean)^2$ (4) | In Equations (2)–(6), X_i represent the data (EEG signal), where $i = 1, \dots, N$, and N is the total number of samples (experiments). |
| Skewness | $skewness = \frac{\sum_{i=1}^N (X_i - mean)^3 / N}{(stddev)^3}$ (5) | |
| Kurtosis | $kurtosis = \frac{\sum_{i=1}^N (X_i - mean)^4 / N}{(stddev)^4}$ (6) | |
| Mean of absolute values of the first difference of normalized (AFD_N) | $AFD_N = \frac{1}{N-1} \sum_{i=1}^{N-1} X(i+1) - X(i) $ (7) | In Equations (7) and (8), X_i represents the data (EEG signal), where $i = 1, \dots, N$, and N is the total number of samples (experiments). |
| The mean of absolute values of the second difference of normalized (ASD_N ()) | $ASD_N = \frac{1}{N-2} \sum_{i=1}^{N-2} X(i+2) - X(i) $ (8) | |
| Shannon entropy (ShEn) | $ShEn = - \sum_{i=1}^N P_i(X) \log P_i(X)$ (9) | X represents the EEG signal, P_i is the occurrence probability for the values in X , and N is the total number of experiments. |
| Approximate entropy (ApEn) | $ApEn(X, m, r) = \frac{1}{N-m+1} \sum_{i=0}^{N-m} \log C_i^m(r) - \frac{1}{N-m} \sum_{i=0}^{N-m-1} \log C_i^{m+1}(r)$ (10) | X represents the EEG signal, m is the size of the vector, r is the tolerance value, N is the total number of experiments, and C_i^m is the vector self-similarity. |
| Sample entropy (SampEn) | $SampEn(X, m, r) = \log \mathcal{O}^m(r) - \mathcal{O}^{m+1}(r)$ where $\mathcal{O}^m(r) = \sum_{j=0}^{N-m} \sum_{i=0}^{N-m} \theta(r - \ u[i] - u[j]\ _\infty)$ (11) | Self-similarity of the pairs of vectors $u[i]$ and $u[j]$ with a tolerance of r . If the signals are self-similar, then $\mathcal{O}^m(r)$ is high. |

Table 1. Cont.

| Method | Mathematical Expression | Description |
|-----------------------------------|---|--|
| Permutation entropy (PerEn) | $PE(X) = \sum_{k=1}^{m!} p(\pi_k) \log p(\pi_k)$ (12) where $p(\pi_k) = \frac{1}{N-m+1} \sum_{i=0}^{N-m} f(u[i], \pi_k)$ | $f(u[i], \pi_k) = 1$ when $u[i]$ and π_k have the same pattern, else is 0, and the pattern is defined by the order of $u[i]$ that corresponds to each element. The occurrence probability of each pattern category is $p(\pi_k)$. |
| Energy (Eng) | $Eng = \sum_{n=-\infty}^{\infty} X(n) ^2$ (13) | $X(n) \rightarrow 0$ as $n \rightarrow \pm\infty$, where $X(n)$ is the EEG signal. |
| Average power (Avg) | $Avg = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N}^N X(n) ^2$ (14) | N is the number of samples taken for the computation, and $X(n)$ is the EEG signal. |
| Root mean square (EMS) | $RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N X^2(n)}$ (15) | N is the number of samples taken for the computation, and $X(n)$ is the EEG signal. |
| Line length (LinLen) | $LL = \sum_{i=1}^N X[i-1] - X[i] $ (16) | X is the EEG signal, N is the number of samples in the signal, and i is the data index. |
| Petrosian fractal dimension (PFD) | $PFD = \frac{\log(m)}{\log(m) + \log\left(\frac{m}{m+0.4N_{\Delta}}\right)}$ (17) | The length of the signal is m , and N_{Δ} is the number of pairs of segments that are not similar in the binary sequence. |
| Higuchi fractal dimension (HFD) | $HDF = -\left(\lim_{k \rightarrow \infty} \frac{\log(L(k))}{\log(k)}\right)$ where $L_m(k) = \frac{1}{\tau} \left(\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} x[m+i.k] - x[m+(i-1)k] \right) \frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k}$ (18) and $L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k)$ | N is the total number of signals, and $y = \frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k}$ is the normalization correction factor. The average length of k sequences with the same interval as the signal length $L(k)$ corresponds to the interval k . |
| Zero crossing (ZeCr) | $ZeCr(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} sgn[x_i(n)] - sgn[x_i(-1)] $ (19) where $sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0 \end{cases} \quad i = 1, \dots,$ | The total count of samples present in a block of the EEG signal is represented by W_L , $x(n)$ represents the input signal, and sgn is the sign function. |

Table 1. Cont.

| Method | Mathematical Expression | Description |
|----------------------------------|--|---|
| Higher-order crossing (HOC) | $HOC_k = ZeCr\{S_k(Z_t)\}$ (20) | $ZeCr$ is the estimation of the number of zero crossings. Z_t represents the finite zero-mean series data, ∇ is the high-pass filter, and S_k is the high-pass filter sequence. |
| Hjorth parameters (HPs) | Activity = $var(x(t))$ (21) | The activity is defined as the variance of the input signal represented by $x(t)$, $var(x'(t))$ represents the variance of the first derivative of $x(t)$, $var(x'(t))$ represents the variance of the signal, and y mobility($x(t)$) is the mobility of the first derivative of $x(t)$. |
| | Mobility = $\sqrt{\frac{var(x'(t))}{var(x(t))}}$ (22) | |
| | Complexity = $\frac{Mobility(x'(t))}{Mobility(x(t))}$ (23) | |
| Frequency domain | | |
| Power ratio (PR) | $PR1 = \frac{DeltaPower+ThetaPower}{AlphaPower+BetaPower}$ (24) | It determines power ratios between the current and background epoch in the same frequency range to compare their power levels. Each power ratio is used with different applications for state-of-mind recognition. |
| | $TBR = \frac{ThetaPower}{BetaPower}$ (25) | |
| | $PR2 = \frac{AlphaPower+ThetaPower}{DeltaPower+BetaPower}$ (26) | |
| | $BAR = \frac{BetaPower}{AlphaPower}$ (27) | |
| Spectral entropy (SE) | $SE(x) = -\sum_{i=1}^m p(s_i)\log_2 p(s_i)$ where $P(m) = \frac{S(m)}{\sum_i S(i)}$ (28) | m is the number of values for the EEG signal. The denominator \log_2 represents the maximum uniformly distributed noise, and $S = \{S_i, \dots, S_m\}$ is a function of the occurrence probability $P = \{p(S_i), \dots, p(S_m)\}$. |
| Power spectral density (PSD) | $PSD = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) ^2 dt$ (29) | $x(t)$ represents the given EEG signal by the following time average, where T is centered at some arbitrary point $t = t_0$. |
| Fast Fourier transform (FFT) | $FFT[k] = \sum_{n=0}^{N-1} x[n].W_N^{kn}$ (30) | $W_N^{kn} = e^{-j\frac{2\pi}{N}}$, N is the number of samples in the signal, and $x(n)$ represents the input EEG signal, with $k = 0, 1, \dots, N - 1$. |
| Time–frequency domain | | |
| Discrete wavelet transform (DWT) | $\gamma(t) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^a}} \psi\left(\frac{t-b*2^a}{2^a}\right) dt$ (31) | $\gamma(t) = DWT$ of any signal in the time domain $x(t)$; $\psi(t) = wavelet$; a and b are the scale parameter and the shift parameter. |

Table 1. Cont.

| Method | Mathematical Expression | Description |
|-----------------------------------|--|---|
| Wavelet entropy (WEnt) | $T(l) = -\frac{E(l)}{\sum_{m=1}^M E(m)} \log\left(\frac{E(l)}{\sum_{m=1}^M E(m)}\right)$ | (32) E is the wavelet energy based on a value l , l is the level of wavelet entropy, $E(l)$ is the square of the vector elements, and m is the number of the wavelet decomposition. |
| Location domain (LD) | | |
| Rational asymmetry (RASM) | $\text{RASM} = \frac{P_{left}}{P_{right}}$ | (33) P_{left} and P_{right} represent the power of the electrodes in the left and right hemispheres of the brain. |
| and differential asymmetry (DASM) | $\text{DASM} = P_{left} - P_{right}$ | (34) |

3.4. Dimensionality Reduction

Extracted features may not be correlated with emotional states and can lead to degradation in classifier performance. Therefore, dimensionality reduction could help increase the speed and stability of the classifier. This study used a correlation matrix approach to reduce dimensions.

Correlation matrix is a technique to detect correlations and eliminate redundancy from the original data. The correlation matrix method for feature selection consists of analyzing pairwise correlations between dataset characteristics. It is a simple but effective way to understand the relationships between features and identify which ones might contain redundant information and can be eliminated without loss of knowledge. The correlation matrix method preserves the features, making the results easily interpretable [21].

The goal is to identify the linear combinations that best represent the variables X_1, \dots, X_p . Let (Z_1, Z_2, \dots, Z_M) $M < p$ be $M < p$ linear combinations of the original p variables, using

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j \quad (32)$$

where ϕ_{jm} are the loadings of the principal components. Each loading vector of length p defines the direction in space along which the variance of the data is maximized, and X_j represents the EEG signal data.

3.5. Classification Algorithms

After extracting the feature vectors and implementing dimensionality reduction, we classified emotions according to the VA space. There are several classifiers for the automatic identification of feelings; below, we will mention some of the most commonly used.

3.5.1. Support Vector Machine (SVM)

This algorithm projects the input space to a higher-dimensional space to separate nonlinear data. The fundamental feature of SVM is that the separation margin of the data becomes as wide as possible [22].

SVM employs kernel methods, which can be of various types, such as linear, polynomial, and Gaussian. SVM aims to choose an optimal separating hyperplane that maximizes the distance between two data points of different classes [22].

The SVM machine learning method is chosen due to its high generalization and classification ability. A training set (x_j, y_j) , $1 \leq j \leq N$, x_j denotes the feature vectors extracted from the EEG signals, y_j denote the corresponding emotion labels, and N is the number of data. The SVM decision function is calculated as follows:

$$f(x) = \sum_i^N \alpha_i y_i k(S_i, x) + b \quad (33)$$

where x is the input vector (in this case, the feature vector extracted from EEG signals), k is the kernel function, S_i denotes support vectors, α_i are the weights, and b is the bias.

3.5.2. k-Nearest-Neighbor (k-NN)

k-NN is a non-parametric estimation method that implements refinement where the feature environment is high resolution in regions with dense training and low resolution in variance. This algorithm assigns labels to previously unsampled points, usually with lower efficiency as the data size increases [23].

Classification with k-NN is based on the constant k defined by the user, where the new case will be assigned to the most common class among its k -nearest neighbors measured using a distance metric such as Euclidean, Manhattan, Minkowski, or Hamming.

Most k-NN classifiers use the Euclidean metric to measure differences between examples represented as vector inputs [24]. The Euclidean distance is defined as:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n W_r (a_r(x_i) - a_r(x_j))^2} \quad (34)$$

where an example is defined as the vector $x = (a_1, a_2, a_3, \dots, a_n)$, n is the dimensionality of the input vector (number of attributes of an example), a_r is the r th attribute of the example, W_r is the weight of the r th attribute— r is from 1 to n —and $d(x_i, x_j)$ represents the smaller of the two most similar examples [25]. The class label assigned to a test example is determined via the majority vote of its k -nearest neighbors using:

$$y(d_i) = \underset{k}{\operatorname{argmax}} \sum_{x_j \in kNN} y(x_j, C_k) \quad (35)$$

where d_i is a test example, x_j is one of its k -nearest neighbors in the training set, and $y(x_j, C_k)$ indicates whether x_j belongs to class C_k .

3.5.3. Artificial Neural Networks (ANNs)

An ANN is based on biological neural systems with nonlinearity, adaptability, responsiveness, and fault-tolerance characteristics. The inputs in ANNs are called interconnected neurons, and work together to search for the best configuration by modifying the network weights to solve a problem.

A neural network receives as input a set of $i = 1, \dots, n$ patterns in the form of vectors of dimension p ; each input vector is processed through the neurons of the I hidden layers according to the connections between the nodes. Each node contains an activation function f , which obtains the node output value through a weighted sum of the node inputs and an additional bias value [22].

In a feedforward ANN, there are three layers: input, output, and hidden. The input layer buffers the distribution of the input signals x_n ($n = 1, 2, 3, \dots$) and sends them to the neurons in the hidden layer. The neurons in the hidden layer aggregate the input signals, x_n , after weighting them with their respective connectivity strengths, W_{nl} , which are the inputs of the layer. Finally, the output Y is calculated as follows.

$$y_l = f \left(\sum_{n=1}^n W_{nl} x_n \right) \quad (36)$$

The number of neurons is I . The activation function can have different functional forms, such as sigmoidal, radial, linear, hyperbolic tangent, etc. Using a similar approach, the outputs of the neurons are calculated.

3.6. Assessment Performance

The performance of a classifier can be evaluated by calculating the number of correctly recognized class examples (TPs), the number of correctly identified samples that do not belong to the class (TNs), the samples that were assigned to the wrong category (FPs), and the examples that were recognized as belonging to the wrong class (FNs) [26].

The performance evaluation metric used in this work is accuracy, obtained from the values presented in the confusion matrix.

3.6.1. Accuracy

Accuracy is a measure of a classifier's class-wise effectiveness, defined as the ratio of correctly classified instances to the total number of instances, as shown in Equation (37), where N is the number of class labels.

$$\text{Accuracy} = \frac{\sum_{i=1}^N \frac{TP+TN}{TP+FN+FP+TN}}{N} \quad (37)$$

3.6.2. Confusion Matrix

The confusion matrix is used to measure the performance of the class problem for a dataset. Table 2 shows an example of a confusion matrix.

Table 2. Confusion matrix example.

| Labels | Prediction | |
|--------------|--------------|--------------|
| | Negative (0) | Positive (1) |
| Negative (0) | TN | FP |
| Positive (1) | FN | TP |

Nomenclature: true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs).

The elements on the right diagonal, TP and TN, classify the instances correctly. In contrast, FP and FN classify the instances incorrectly, where

- total instances = correct instances + incorrect instances;
- correctly classified instance = TP + TN;
- incorrectly classified instance = FP + FN.

4. Implementation and Results

The DEAP dataset's characteristics are presented in Table 3. It consists of 32.dat files, composed of two parts: the data and the emotion labels classified on a numeric scale from 1 to 9. Regarding the labels, we are only interested in the first two of the four available variables, corresponding to valence and arousal.

Table 3. DEAP dataset characteristics.

| Description | DEAP Dataset |
|--------------------|--|
| EEG devise | Biosemi ActiveTwo |
| Number of channels | 32 for EEG, 8 physiological signals |
| Sampling | Original 512 Hz, downsampled samples of 128 Hz |
| Number of subjects | 32 |
| Stimulus | 40 musical videos (one minute each) |
| Emotions | Valence, arousal, dominance, and liking (scale from 1 to 9, familiarity from 1 to 5) |

The one-hot encoding process was applied to convert the numerical variables to categorical data, proposing three criteria: median, mean, and fixed value. The first criterion, "Median", consists of finding the median value between the valence–arousal (VA) labels independently and categorizing these labels with values of 0 and 1, accordingly. Label 1 is assigned to values above the median, and label 0 to values below the median. The second criterion, "Mean", initially looks for the average value among the VA data independently. Once the average value is identified, the labels are standardized so that values equal to or greater than the average are labeled 1, and values below the average are labeled 0. Finally, the third criterion, "Greater than 5", proposes assigning the label 1 to all data greater than or equal to 5, and 0 to data less than 5.

Figure 3 shows the percentages of each categorization according to the proposed criteria. The “Median” criterion has 28% of classes labeled HAHV, 22% LAHV, 25.2% HALV, and 24.8% LALV. In the “Mean” criterion, 26.6% of categories correspond to the HAHV label, 23.7% to LAHV, 20% to HALV, and 29.6% to LALV. The difference between the “Median” and “Mean” criteria ranges from 1.7% to 5.2%. However, when comparing the results with the third criterion, “Greater than”, we have 35.8% in HAHV, 23.1% in LAHV, 20.8% in HALV, and 20.3% in LALV, representing a significant difference of 7.8% and 9.2% in the HAHV label.

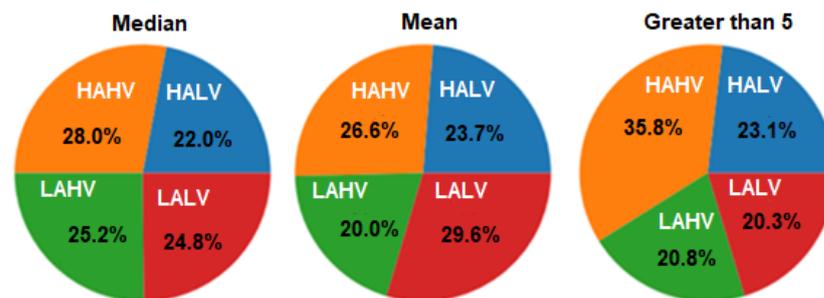


Figure 3. Criteria for categorizing emotional quadrants. High arousal–high valence (HAHV), high arousal–low valence (HALV), low arousal–low valence (LALV), and low arousal–high valence (LAHV).

Based on these results, several initial tests were carried out for the one-hot encoding process via assigning 80% and 20% to the train and test sets [27]. It was identified that the best option for the process was using the “Median” criterion. It separated the data so that there were samples of emotions from each VA quadrant both in the train and test sets.

4.1. Selection of Feature Extraction Methods and Classification Algorithms

The experiments to be carried out were based on the results of a literature review where a total of 25 studies were analyzed, addressing the following questions:

- Q1. What methods are used for feature extraction in the time, frequency, and time–frequency domains for emotion recognition?
- Q2. Which classification algorithms yield better results in emotion classification?

4.1.1. Feature Extraction Methods

In response to question Q1, regarding the most commonly used feature extraction methods from the different domains, the results are shown in Figures 4–7.

Figure 4 shows the feature extraction methods in the time domain found in our literature review vs. the number of papers that used them. Statistical methods are the most used, followed by ShEn, HFD, SampEn, and HP methods.

In Figure 5, the statistical procedures found in the SLR are shown. In total, 26% of articles employ Stddev as a statistical method, 19% use the mean method, and 17% use its AFSD_N variants. Meanwhile, 12% corresponds to the kurtosis method, and 7% to the skewness method.

Based on the frequency of use of methods in the time domain, the most commonly used approaches were selected: Shannon entropy, sample entropy, Higuchi fractal dimension, Hjorth parameters, and statistical methods. Some of the selected statistical methods were the standard deviation, mean, mean absolute deviation, and kurtosis, corresponding to percentages greater than 12%.

Additionally, specific methods in the time domain with lower percentages were considered, as they present good results in the evaluation metrics of classification algorithms. Among the techniques included in the time domain are two variants of entropy (approximate and permutation), the Petrosian fractal dimension method, energy, root mean square,

line length, average power, zero crossing, higher-order crossing, and the statistical techniques: variance, skewness, and maximum.

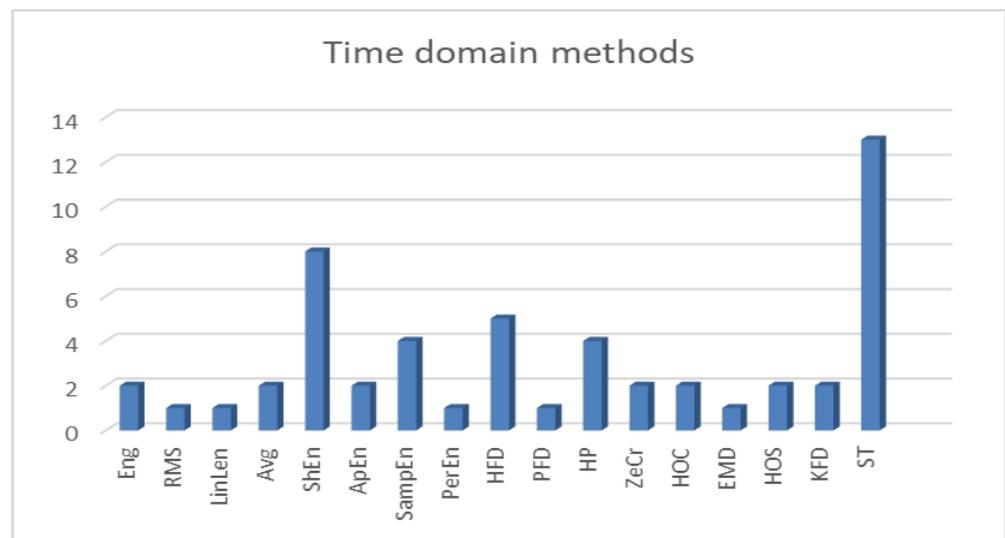


Figure 4. Feature extraction methods in the time domain. Energy (Eng), root mean square (RMS), line length (LinLen), average power (Avg), Shannon entropy (ShEn), approximate entropy (ApEn), sample entropy (SampEn), permutation entropy (PerEn), Higuchi fractal dimension (HFD), Petrosian fractal dimension (PFD), Hjorth parameters (HP), zero crossing (ZeCr), higher-order crossing (HOC), empirical mode decomposition (EMD), higher-order spectral (HOS), Katz’s fractal dimension (KFD), statistics (ST).

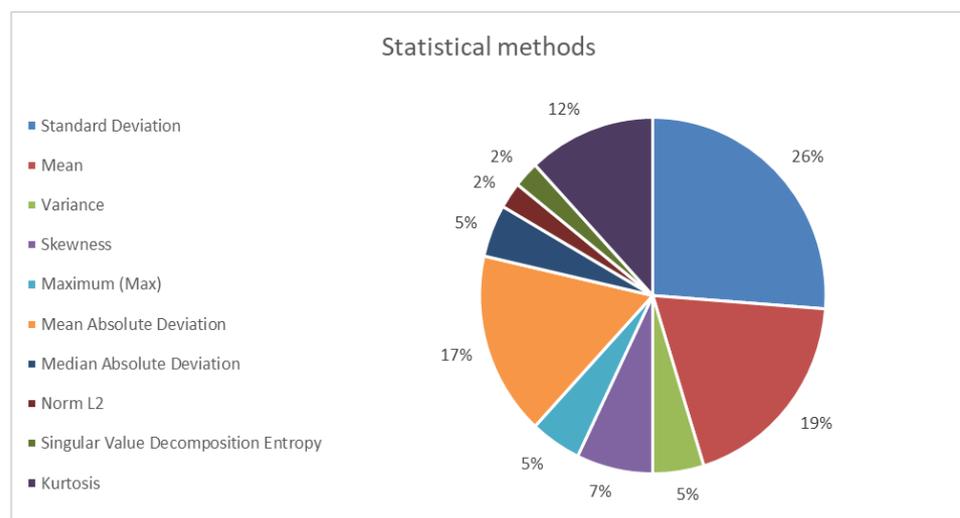


Figure 5. Statistical feature extraction methods in the time domain.

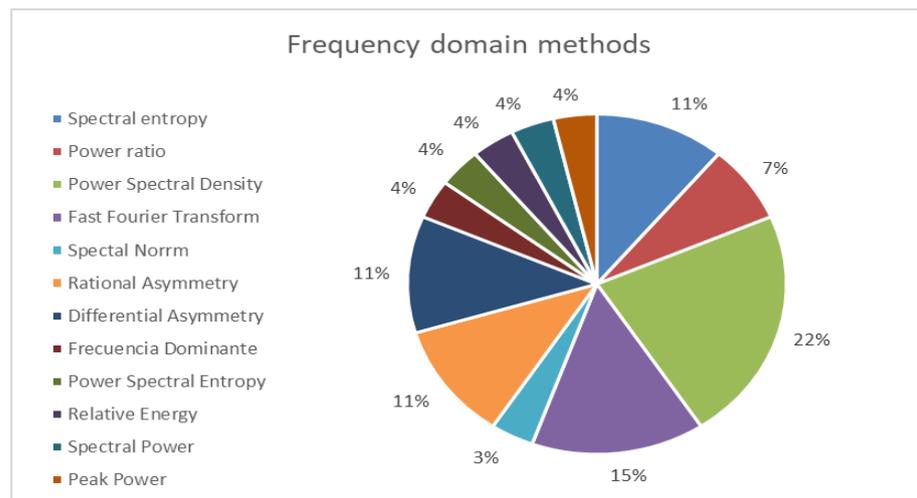


Figure 6. Feature extraction methods in the frequency domain.

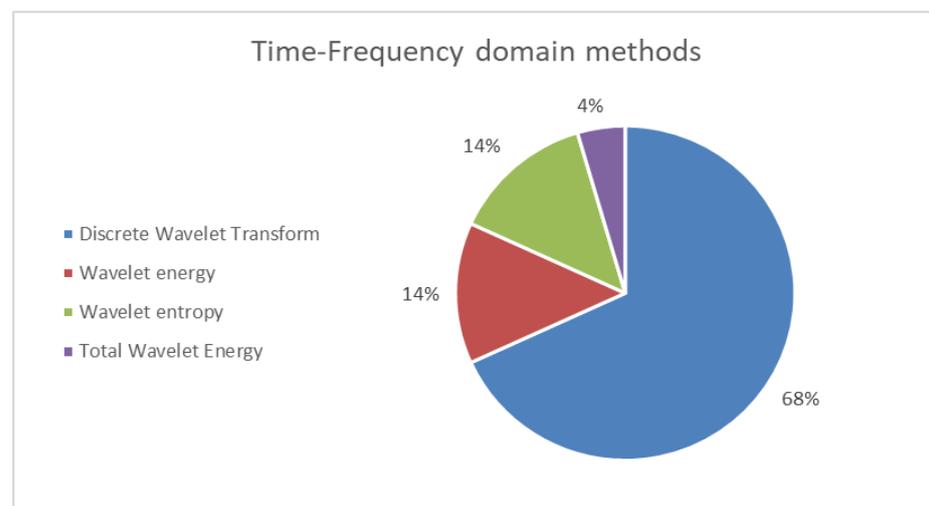


Figure 7. Feature extraction methods in the time–frequency domain.

Figure 6 shows the frequency-domain methods found in the SLR. We found that 22% of articles employ the PSD method for feature extraction, 15% use the FFT method or one of its variations, and 11% use the SE, RASM, and DASM methods, followed by the PR method with 7%.

The selected methods in the frequency domain are power spectral density, fast Fourier transform, rational asymmetry, differential asymmetry, and spectral entropy, which correspond to percentages greater than or equal to 11% of the frequency of use. Additionally, the power ratio method, which corresponds to 7%, was selected due to its good results in the evaluation metrics of classification algorithms.

Finally, Figure 7, feature extraction methods in the time–frequency domain, shows the feature extraction methods in the time–frequency domain found in the SLR. In total, 68% of articles use the DWT method, followed by the WEng and Went methods with 14% each.

The selected techniques in the time–frequency domain are discrete wavelet transform and wavelet entropy, which correspond to the highest percentages of frequency of use. The analysis involved the creation of four vectors with features from the time, frequency, and time–frequency domains, which were used for the experiments in Section 4.3.

Spatial features in EEG (electroencephalography) refer to the attributes and information that can be derived from the distribution and variation of EEG signal amplitudes across different electrode sites on the scalp. In the DEAP dataset, EEG electrodes are placed

according to the International 10–20 system, a standardized method to locate the sensors on the scalp in a specific pattern based on anatomical landmarks of the head. We used the algorithms DASM and RASM, described in Table 1, to extract these spatial features.

4.1.2. Classification Algorithms

In response to question Q2, regarding which algorithms are used for emotion classification, Figure 8 shows the results in the considered papers of the SLR. SVM achieves better results in 32% of the articles; 23% of articles obtained better results with ANN, and 18% with the k-NN classifier, followed by the NB and RF classifiers, with 13.5% each.

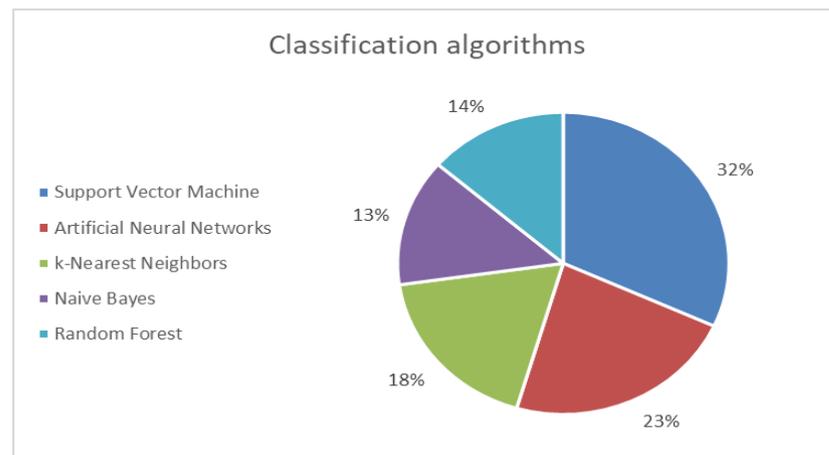


Figure 8. Classification algorithms.

In the papers considered in our SLR, the percentage of the frequency and performance of the algorithms used in the research was calculated, based on which the SVM, k-NN, and ANN algorithms were selected to be implemented in this work because they have percentages equal to or greater than 18%. The parameters of the classification algorithms are discussed in Section 4.3.

4.2. Feature Extraction Times

The feature extraction methods presented in Table 1 were grouped into statistical features, additional features (energy, root mean square, line length, average power, Higuchi fractal dimension, Petrosian fractal dimension, Hjorth, zero crossing, and higher-order crossing), frequency-domain features, time–frequency, and location (spatial)-domain features. The calculation of the approximate time for entropy variants is performed individually. This division was made because the number of characteristics in the time domain is numerous, and the entropy variants have a high computational cost.

The time it takes to generate the features was calculated via averaging three executions of our feature package for a given signal for a single data point, i.e., a specific user, with 1 and 32 channels.

Table 4 summarizes the approximate time in seconds it takes to extract features in the DEAP dataset. The “Additional” features group has a shorter extraction time despite being composed of several algorithms (Petrosian fractal dimension method, energy, root mean square, line length, average power, zero crossing, higher-order crossing, and the statistical techniques: variance, skewness, and maximum). Next are the time–frequency feature group and statistical feature group.

On the other hand, frequency-domain features have a longer execution time than the other features, followed by the approximate and sample entropy variants.

Table 4. Execution time for feature extraction in the DEAP dataset.

| Features | 1 Channel (s) | 1 Trial 32 Channels (s) | 40 Trials 32 Channels (s) |
|------------------------|---------------|----------------------------|------------------------------|
| Statistical | 0.055000 | 1.760 | 70.40 |
| Additional | 0.220964 | 5.974 | 23.935 |
| Frequency domain | 1.114581 | 35.667 | 1426.68 |
| Frequency–time domain | 0.047963 | 1.535 | 61.40 |
| Shannon entropy method | 0.272020 | 8.705 | 348.20 |
| Approximate entropy | 3.471318 | 111.082 | 4443.28 |
| Sampling entropy | 3.368673 | 107.798 | 4311.92 |
| Permutation entropy | 0.061990 | 1.984 | 79.36 |

4.3. Emotion Classification and Performance Evaluation

One of the objectives of this work is to use different feature vectors to detect and classify emotional states, employing different classification algorithms to evaluate which feature extraction methods provide more information and produce classifiers with a better performance.

The data were split into 80% and 20% for training and testing. Then, the data were associated with the previously extracted features to construct the different input vectors for the classifiers.

The classification algorithms used are SVM, k-NN, and ANN. An RBF kernel was used in the SVM classifier, and “auto” was set as the gamma parameter. In the k-NN classifier, the value of $k = 5$ was used. These two classifiers were implemented using different functions offered by the sklearn library. The ANN classifier was implemented, assigning 1000 epochs, three layers given the number of quadrants to classify, of which the first two layers are activated with RELU with 64 neurons each layer, and the last output layer has four outputs, its activation function is softmax and, in the compilation, the parameter “categorical_crossentropy” and optimizer “Adam” [16] are used.

The feature vectors for the DEAP dataset were composed of features from one-domain and multi-domain, original and no-correlation-selected attributes, with the following setups:

- Vector with 21 characteristics in the time domain.
- Vector with selected nine features in the time domain.
- Vector with 11 characteristics in the frequency domain that includes 5 bands for delta, theta, alpha, beta, and gamma plus PSD ranges.
- Vector with selected seven attributes in the frequency domain.
- Vector with two characteristics in the time–frequency domain.
- Vector with 32 features (DASM and RASM) for the location domain.
- Vector with 16 attributes (DASM) chosen in the location domain.
- Nine vectors corresponding to the combination of selected multi-domain attributes: time, frequency; time, time–frequency; time, location; frequency, time–frequency; frequency, location; time, frequency, time–frequency; frequency, time–frequency, location; time, frequency, time–frequency, location.

We applied cross-validation, given that we are working with features, not raw data. Therefore, the models are simple enough, so that we had a manageable computational load. In Table 5, we present the average accuracy performance using 5-fold cross-validation.

Table 5. Accuracy results using the DEAP dataset in the valence–arousal space.

| Domain | Number of Features | Accuracy | | |
|--|---|----------|------|------|
| | | SVM | KNN | ANN |
| Time domain | 21 | 0.70 | 0.71 | 0.76 |
| | Selected 9 | 0.75 | 0.78 | 0.80 |
| Frequency domain | 11 (included 5 bands) | 0.80 | 0.77 | 0.82 |
| | 7 (included 5 bands) | 0.81 | 0.81 | 0.88 |
| Time–frequency domain | 2 | 0.86 | 0.83 | 0.90 |
| Location domain | 16 pairs for 32 channels for DASM and 16 pairs for 32 channels for RASM | 0.60 | 0.62 | 0.75 |
| | Selected 16 pairs for 32 channels for DASM | 0.64 | 0.64 | 0.81 |
| Hybrid (combination of selected features in the domains) | 9 time + 7 frequency | 0.87 | 0.82 | 0.95 |
| | 9 time + 2 time–frequency | 0.78 | 0.82 | 0.89 |
| | 9 time + 16 location | 0.70 | 0.74 | 0.86 |
| | 3 frequency + 2 time–Frequency | 0.81 | 0.80 | 0.90 |
| | 3 frequency + 16 location | 0.74 | 0.7 | 0.92 |
| | 9 time + 7 frequency + 2 time–frequency | 0.81 | 0.83 | 0.94 |
| | 9 time + 7 frequency + 16 location | 0.80 | 0.80 | 0.93 |
| | 7 frequency + 2 time–frequency + 16 location | 0.75 | 0.76 | 0.91 |
| | 9 time + 7 frequency + 2 time–frequency + 16 location | 0.82 | 0.83 | 0.96 |

We used StratifiedKFold from `sklearn.model_selection` for the cross-validation process to ensure that each fold was representative of the entire dataset. We modified the training and evaluation loop to iterate over the folds. For each fold, we split the dataset into training and testing sets. Then, we trained and evaluated all models within each fold. Finally, we kept track of the performance metrics for each fold to calculate the average performance after all folds were assessed. In our case, the accuracy results were similar.

Results Using Data from DEAP

It was not necessary to apply methods for balancing data between classes. The input vectors to the classifiers were reduced in dimensionality using a correlation matrix to identify features with redundant information that could add noise to the classifier and result in lower accuracy scores. No-correlation selected the following time features as non-redundant: mean, standard deviation, variance, Shannon entropy, energy, Higuchi, higher-order crossing, Hjorth mobility, and Hjorth complexity. In the same way, in the frequency domain, the selected features were the PSD of the five frequency bands: delta (0.5 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz), beta (13 to 30 Hz), and gamma (over 30 Hz), plus the power ratio, power spectral density for all range of frequencies, and fast Fourier transform. The selected features were the same as the originals in the time–frequency domain. In the location domains, the chosen features were only DASM; the RASM was eliminated for being correlated to the first ones. It is worth noting that these features were calculated between the pairs of signals located on each side of the scalp (16 pairs for 32 channels).

Table 5 presents an exhaustive evaluation of feature extraction methods and machine learning algorithms for emotion recognition in the valence–arousal space using the DEAP dataset. We explore time, frequency, time–frequency, and location (spatial)-domain features that measure asymmetry between the right and left brain regions. The performance of three machine learning algorithms, SVM, KNN, and ANN, are reported for each feature set and their combinations to monitor performance.

In the time domain with the 21 features described in Table 1, the results for accuracy were 0.70 for SVM, 0.71 for KNN, and 0.76 for ANN. A correlation matrix was applied to eliminate characteristics with redundant information that could add noise to the classification. When using the selected nine features in the time domain, the performance increased substantially, showing accuracies of 0.75, 0.78, and 0.80 for SVM, KNN, and ANN, respectively.

Using all the described features in the frequency domain (Table 1), the results were 0.80 for SVM, 0.77 for KNN, and 0.82 for ANN. Applying feature selection, we obtained seven relevant features and, again, the accuracy values improved to 0.81, 0.81, and 0.88 for SVM, KNN, and ANN. This last algorithm showed the best performance.

In the time–frequency domain, the original two features were used and yielded accuracies of 0.86 for SVM, 0.83 for KNN, and 0.90 for ANN.

The location domain was calculated using the asymmetries on the right and left of the brain (DASM and RASM). With the original features, the accuracies were 0.60 for SVM, 0.62 for KNN, and 0.75 for ANN. Using a correlation matrix, the selected characteristics were DASM (16 pairs), and ANN's performance shot up to 0.81.

The tests with hybrid models that were conducted combining features from multiple domains yielded the following best-performing models: nine time plus seven frequency features: ANN 0.95; nine time plus seven frequency plus two time–frequency features: ANN 0.94; nine time plus seven frequency plus two time–frequency + sixteen location features: ANN 0.96. The highest-performing model used a combination of nine attributes from the time domain, seven from the frequency domain, two from the time–frequency domain, and 16 from the location domain, achieving an accuracy of 0.96 when using ANN. This suggests that combining selected features from different domains can result in a superior performance. SVM and KNN showed a lower performance than ANN for all the hybrid vectors.

Moreover, features from the time–frequency domain consistently yielded a high accuracy across all machine learning models, emphasizing their importance. The location-domain features initially performed poorly but showed significant improvement when combined with features from other domains, highlighting the effectiveness of hybrid models.

5. Discussion

In this study, emotion recognition was performed based on the two-dimensional valence/arousal (VA) model, which divides emotions into four groups: high arousal–high valence (HAHV), high arousal–low valence (HALV), low arousal–low valence (LALV), and low arousal–high valence (LAHV). We used various features in different domains, such as time, frequency, time–frequency, and location. The features, classifiers, and evaluation metrics were selected according to their frequency of use and results found in the RSL. Sixteen input vectors were created for the classifiers, seven were collected for features in each domain (the original, and that selected via applying a correlation matrix), and nine were composed of combinations of features chosen from multi-domains.

Then, the present study extensively evaluated feature sets across domains and their impact on emotion recognition. Feature selection improved results across the different domains. Additionally, hybrid models combining domain features offered a superior performance. An ANN using an input vector of nine time, seven frequency, two time–frequency, and sixteen location features yields the best results, with a maximum accuracy of 0.96.

There are multiple proposed methods for EEG-based emotion classification. Still, they have limitations, including the number of emotional states considered. Generally, strategies dealing with more classes and participants tend to decrease their performance. According to [8], the average accuracy for emotion recognition using four quadrants in the two-dimensional valence/arousal space is 76.68%. Therefore, our results are better than those reported in the literature. An important part of our contribution is the anal-

ysis of the participation of multi-domain features in the classification performance for emotion recognition.

Our study uses the DEAP dataset for EEG-based emotion recognition. It is important to recognize the role of self-labeling in the data acquisition process and its implications for emotion classification. Participants in the DEAP study used the Self-Assessment Manikin (SAM) to label their emotions. This method relies on visual mannequins to facilitate the self-assessment of emotional responses elicited by various video stimuli. This is an intuitive and accessible approach for participants to convey their emotional states. However, the self-labeling of emotions still introduces a degree of subjectivity into the dataset. This subjectivity arises from individual differences in emotional perception and expression. Individuals may interpret and respond to the same stimuli differently due to personal experiences, cultural backgrounds, and even momentary psychological states. As a result, self-labeled emotional states may not always align consistently with the physiological signals captured by the EEG. This aspect is recognized with the conviction of the potential of machine learning techniques to account for this variability in obtaining their recognition models.

6. Conclusions

This study presents a comprehensive evaluation of various feature extraction methods and machine learning algorithms—namely, support vector machine (SVM), K-nearest neighbor (KNN), and an artificial neural network (ANN)—for emotion recognition. The study uses the DEAP dataset and explores feature extraction in the time, frequency, time–frequency, and location (spatial) domains. The following key points emerged:

- Effectiveness of feature selection: applying feature selection techniques like a correlation matrix to eliminate redundant characteristics generally improved the model's performance across all domains and algorithms, emphasizing the importance of targeted feature selection.
- The superiority of ANN: ANN consistently outperformed the other machine learning models, particularly in scenarios where features were selected carefully.
- Importance of time–frequency features: features from the time–frequency domain consistently yielded high accuracies across all machine learning algorithms, underlining their relevance for emotion recognition tasks.
- Role of hybrid models: combining features from multiple domains led to the highest-performing models. In particular, a hybrid model employing a mixture of nine time-domain features, seven frequency-domain features, two time–frequency-domain features, and sixteen location-domain features achieved an accuracy of 0.96 with ANN.
- Improvement in location domain: Initially, the location-domain features performed poorly compared to other domains. However, when combined with features of multiple domains, the performance significantly increased, showing the importance of the information provided for spatial features.
- The present study uses relatively simple and computationally inexpensive machine-learning algorithms like SVM, KNN, and ANN. Despite their simplicity, these algorithms could achieve high levels of accuracy. This makes the findings particularly valuable for real-time emotion recognition applications where computational resources and processing time are often limited.

One of the key contributions is the exhaustive evaluation of feature extraction methods across multiple domains, providing a robust understanding of what works best for emotion recognition in the valence–arousal space. The study identifies optimized subsets of features in each domain that reduce the computational burden and increase accuracy, aiding in efficient and effective emotion classification. Our research demonstrates that hybrid models, combining selected features from different domains, can achieve superior performance, laying the groundwork for future research. Using a well-acknowledged DEAP dataset, the study provides empirical evidence supporting the efficacy of different machine learning algorithms and feature extraction methods for emotion recognition. It offers crucial insights

into the potential of ANN in emotion recognition tasks. It establishes the importance of feature selection, setting the stage for future research to build upon these findings.

An additional contribution is the use of less-computationally intensive machine learning algorithms, which adds a dimension of practicality to this study. The algorithms used are scalable and capable of running on systems with limited computational power, making the findings directly applicable to real-time emotion recognition systems. This particularly benefits wearable devices or mobile applications where rapid processing is paramount.

This study sets a new benchmark by demonstrating that a high accuracy in emotion recognition can be achieved without resorting to complex or computationally expensive machine learning algorithms. It shows that effective feature selection and an intelligent combination of features can deliver exceptional performance, even when the algorithmic backdrop is simple.

7. Future Work

Finally, regarding future work, we propose studying emotion recognition defined in the three-dimensional (VAD) model, considering the axes corresponding to valence, arousal, and dominance for a broader recognition of specific emotions. Another possibility is to address the regression approach instead of classification, as valence and arousal values are continuous, allowing the prediction of the user's sentiment through the EEG signal rather than classifying it.

We plan to continue exploring feature selection using genetic algorithms that could provide a more dynamic and adaptive approach to feature selection. We could uncover complex, nonlinear relationships between variables that are not evident with traditional methods. This approach can lead to the choice of more efficient feature sets that further improve the accuracy and generalization of EEG-based emotion classification models.

Our work also paves the way for future research that explores new avenues in emotion recognition using EEG signals and BCI devices as personalized emotion recognition: individual differences in emotional responses could be considered to develop personalized emotion recognition models. This approach could potentially improve the performance of classifiers and provide more accurate emotion recognition for each user.

The findings of this study could be applied to various real-world scenarios, such as developing more effective human–computer interaction systems, enhancing mental health interventions, or creating more engaging and adaptive virtual reality experiences.

Author Contributions: Methodology, M.Á.-J. and M.H.-Á.; Validation, T.C.-J.; Investigation, M.Á.-J. and M.H.-Á.; Resources, M.Á.-J. and M.H.-Á.; Funding acquisition, M.Á.-J. and M.H.-Á.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [<https://www.eecs.qmul.ac.uk/mmv/datasets/deap/>] (accessed on 23 September 2023).

Acknowledgments: The authors would like to acknowledge VIIV (Vicerrectorado de Investigación, Innovación y Vinculación) of the Escuela Politécnica Nacional for making the publication of this article possible.

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

References

1. Bablani, A.; Edla, D.R.; Tripathi, D.; Cheruku, R. Survey on Brain-Computer Interface: An Emerging Computational Intelligence Paradigm. *ACM Comput. Surv.* **2019**, *52*, 1–32. [[CrossRef](#)]
2. Alarco, S.M.; Fonseca, M.J. Emotions recognition using EEG signals: A survey. *IEEE Trans. Affect. Comput.* **2019**, *10*, 374–393. [[CrossRef](#)]
3. Daly, I. Affective Brain-Computer Interfacing and Methods for Affective State Detection. In *Brain-Computer Interfaces Handbook: Technological and Theoretical Advances*; CRC Press; Taylor & Francis Group: Boca Raton, FL, USA, 2018; pp. 147–164. [[CrossRef](#)]
4. Bhatti, A.M.; Majid, M.; Anwar, S.M.; Khan, B. Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Human Behav.* **2016**, *65*, 267–275. [[CrossRef](#)]

5. Chanel, G.; Kronegg, J.; Grandjean, D.; Pun, T. Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. In *International Workshop on Multimedia Content Representation, Classification and Security*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4105, pp. 530–537. [CrossRef]
6. Zheng, W. Multichannel EEG-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis. *IEEE Trans. Cogn. Dev. Syst.* **2017**, *9*, 281–290. [CrossRef]
7. Koelstra, S.; Muhl, 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. 2012. Available online: <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/> (accessed on 11 June 2020).
8. Torres, E.P.; Torres, E.A.; Hernández-Álvarez, M.; Yoo, S.G. EEG-based BCI emotion recognition: A survey. *Sensors* **2020**, *20*, 5083. [CrossRef]
9. Nawaz, R.; Cheah, K.H.; Nisar, H.; Yap, V.V. Comparison of different feature extraction methods for EEG-based emotion recognition. *Biocybern. Biomed. Eng.* **2020**, *40*, 910–926. [CrossRef]
10. Yin, Z.; Liu, L.; Chen, J.; Zhao, B.; Wang, Y. Locally robust EEG feature selection for individual-independent emotion recognition. *Expert Syst. Appl.* **2020**, *162*, 113768. [CrossRef]
11. Arnau-González, P.; Arevalillo-Herráez, M.; Ramzan, N. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. *Neurocomputing* **2017**, *244*, 81–89. [CrossRef]
12. Atkinson, J.; Campos, D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl.* **2016**, *47*, 35–41. [CrossRef]
13. Khateeb, M.; Anwar, S.M.; Alnowami, M. Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset. *IEEE Access* **2021**, *9*, 12134–12142. [CrossRef]
14. Zubair, M.; Yoon, C. EEG based classification of human emotions using discrete wavelet transform. *Lect. Notes Electr. Eng.* **2017**, *450*, 21–28. [CrossRef]
15. Zheng, W.-L.; Zhu, J.-Y.; Lu, B.-L. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Trans. Affect. Comput.* **2019**, *10*, 417–429. [CrossRef]
16. Li, J.; Qiu, S.; Du, C.; Wang, Y.; He, H. Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *12*, 344–353. [CrossRef]
17. Jurcak, V.; Tsuzuki, D.; Dan, I. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *Neuroimage* **2007**, *34*, 1600–1611. [CrossRef] [PubMed]
18. Chen, J.-M.; Chang, P.-C.; Liang, K.-W. Speech Emotion Recognition Based on Joint Self-Assessment Manikins and Emotion Labels. In *2019 IEEE International Symposium on Multimedia (ISM)*; IEEE: New York, NY, USA, 2019; pp. 327–330. [CrossRef]
19. Wang, J.; Wang, M. Review of the emotional feature extraction and classification using EEG signals. *Cogn. Robot.* **2021**, *1*, 29–40. [CrossRef]
20. Khosla, A.; Khandnor, P.; Chand, T. A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybern. Biomed. Eng.* **2020**, *40*, 649–690. [CrossRef]
21. Kurita, T. Principal Component Analysis (PCA). *Comput. Vis.* **2020**, 1–4. [CrossRef]
22. Santa Cruz, O.; del Mar Ramírez, L.; Trujillo-Romero, F. Técnicas de aprendizaje automático aplicadas a electroencefalogramas. *Res. Comput. Sci.* **2016**, *113*, 53–65. [CrossRef]
23. Doma, V.; Pirouz, M. A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *J. Big Data* **2020**, *7*, 1–21. [CrossRef]
24. Sun, S.; Huang, R. An adaptive k-nearest neighbor algorithm. In *Proceedings of the 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, China, 10–12 August 2010; Volume 1, pp. 91–94. [CrossRef]
25. Dudani, S.A. The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *SMC-6*, 325–327. [CrossRef]
26. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
27. Moon, S.-E.; Chen, C.-J.; Hsieh, C.-J.; Wang, J.-L.; Lee, J.-S. Emotional EEG classification using connectivity features and convolutional neural networks. *Neural Netw.* **2020**, *132*, 96–107. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.