



## Article

# Machine-Learning-Based Emotion Recognition System Using EEG Signals

Rania Alhalaseh \* and Suzan Alasasfeh

Department of Computer Science, Mutah University, Karak 61710, Jordan; suzanaasasfeh@gmail.com

\* Correspondence: halaseh@mutah.edu.jo

Received: 25 September 2020; Accepted: 24 November 2020; Published: 30 November 2020



**Abstract:** Many scientific studies have been concerned with building an automatic system to recognize emotions, and building such systems usually relies on brain signals. These studies have shown that brain signals can be used to classify many emotional states. This process is considered difficult, especially since the brain's signals are not stable. Human emotions are generated as a result of reactions to different emotional states, which affect brain signals. Thus, the performance of emotion recognition systems by brain signals depends on the efficiency of the algorithms used to extract features, the feature selection algorithm, and the classification process. Recently, the study of electroencephalography (EEG) signaling has received much attention due to the availability of several standard databases, especially since brain signal recording devices have become available in the market, including wireless ones, at reasonable prices. This work aims to present an automated model for identifying emotions based on EEG signals. The proposed model focuses on creating an effective method that combines the basic stages of EEG signal handling and feature extraction. Different from previous studies, the main contribution of this work relies in using empirical mode decomposition/intrinsic mode functions (EMD/IMF) and variational mode decomposition (VMD) for signal processing purposes. Despite the fact that EMD/IMFs and VMD methods are widely used in biomedical and disease-related studies, they are not commonly utilized in emotion recognition. In other words, the methods used in the signal processing stage in this work are different from the methods used in literature. After the signal processing stage, namely in the feature extraction stage, two well-known technologies were used: entropy and Higuchi's fractal dimension (HFD). Finally, in the classification stage, four classification methods were used—naïve Bayes,  $k$ -nearest neighbor ( $k$ -NN), convolutional neural network (CNN), and decision tree (DT)—for classifying emotional states. To evaluate the performance of our proposed model, experiments were applied to a common database called DEAP based on many evaluation models, including accuracy, specificity, and sensitivity. The experiments showed the efficiency of the proposed method; a 95.20% accuracy was achieved using the CNN-based method.

**Keywords:** EEG; CNN; EMD; VMD; emotion recognition

## 1. Introduction

The brain–computer interface (BCI) is a subfield of human–computer interaction (HCI). The BCI enables the association between the human brain and electronic devices such as a computer and a mobile phone. The BCI has contributed to helping disabled people. A BCI system makes the user interact with the device, which employs EEG signals and others. The different processing steps in the BCI center focus on knowing the purposes of the brain signals and transforming them into actions [1]. BCI techniques obtain signals from a subject's brain, extract knowledge from the obtained/captured signals, and utilize this knowledge to define the purpose of the subject that might have created those

signals. EEG signals are also employed in nonmedical contexts such as entertainment education, monitoring, and games [2].

Emotions perform an essential role in human cognition, particularly in rational decision-making, perception, human interaction, and human intelligence. Affective computing has appeared to satisfy the gap in emotion, specifically in HCI, by gathering technology and emotions into HCI [3]. HCI measures the emotional status of a user by capturing emotional interactions between a human and a computer. Emotion recognition is the method of knowing a human's emotional status. Analysis of emotion recognition profits from the progress of psychology, modern neuroscience, cognitive science, and computer science [4]. In computer science, emotion recognition by computer systems aims to enhance human–machine interaction over a broad range of application areas, including clinical, industrial, military, and gaming [5].

Different approaches have been suggested for emotional recognition and can be split into two types: first, using the characteristics of emotional behavior, such as facial expression, tone of voice, and body gestures, to identify a particular emotion; second, using signals to identify emotions. The physiological activities can be registered by noninvasive sensors, often as electrical signals. These models involve skin conductivity, electrocardiogram, and EEG [6].

Emotion evaluation techniques may consist of subjective and/or objective measurements. Subjective measures can be instruments for self-reporting, such as questionnaires, adjective checklists, and pictorial tools. Objective measures can apply physiological signals such as blood pressure responses, skin responses, pupillary responses, brain waves, and heart responses. Subjective and objective methods can be used jointly to improve the accuracy and reliability of emotional state determination [7].

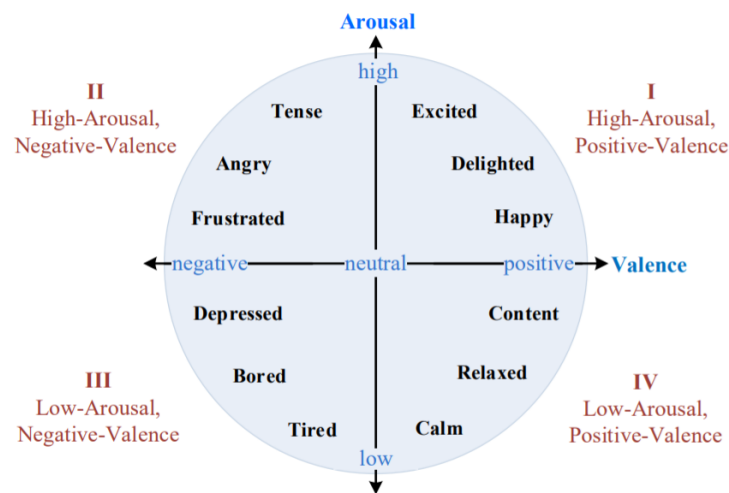
Emotion models were divided into two types: dimensional and discrete. The dimensional model describes the permanence of an emotional state. Most dimensional models combine valence and arousal. The discrete model of emotions assumes more emotions according to a particular number of emotions. Valence regards the level of pleasantness related to emotion. The range of valence represents an unpleasant state to a pleasant state. Arousal indicates the force of experience by emotion. This arousal happens along a continuous sequence and ranges from inactive (e.g., bored) to active (e.g., excited). The following points define the valence, arousal, and dominance emotion categories [8]:

- Valence: positive, happy emotions affecting a higher frontal consistency in alpha signals, and higher right parietal beta signal power, a contrast to negative emotion.
- Arousal: excitation displaying a higher beta signal power and consistency in the parietal lobe, and lower alpha signal activity.
- Dominance: the force of emotion, which is usually shown in the EEG as an addition to the beta/alpha signal activity proportion in the frontal lobe, and an increment in beta activity at the parietal lobe.

Plutchik [9] illustrates eight essential emotions: anger, fear, sadness, disgust, surprise, anticipation, acceptance, and joy. All other emotions can be created by these essential ones; for example, disappointment is a combination of surprise and sadness.

Emotions can also be classified as negative, positive, and neutral emotions. The basic positive emotions care and happiness are necessary for survival, development, and evolution. Basic negative emotions, including sadness, anger, disgust, and fear, usually operate automatically and within a short period. However, the neutral emotional show policy is not based on scientific theory or research; it is more of a theory or prescriptive model of negotiations [10].

Figure 1 shows another classification of emotions, ranging from negative to positive in the case of valence and from high to low in the case of arousal. For example, depressed, as an emotion, lies in the category of low arousal and negative valence.



**Figure 1.** Human emotions based on valence and arousal model [11].

Recognizing emotion from physiological signals primarily with EEG has obtained attention from researchers recently. EEG is the method that is most suited for signal gathering because of its high temporal resolution, safety, and ease of use. EEG has low locative resolution and is dynamic. EEG signals suffer from sensitivity produced by eye winking, eye movements, heartbeats, muscular exercises, and power line obstacles [12].

Another stimulus that is especially physiologically efficient is the activation of the brain, as many activated neurons cause electrical stimulation on the surface of the skin with EEG electrodes. The dataset also contains external records for eye activity, electromyography (EMG), galvanic skin response (GSR), pacing, blood pressure, and temperature.

An EEG is a specific kind of biological signal. It is a measure of the electrical activity of the brain, performed by positioning several electrodes across the scalp [13].

Recently, studying EEG signals has gained attention due to its availability. Today, there are new wireless EEG devices in the market that are portable, affordable, and easy to use. Studying EEG signals is an interdisciplinary approach that consists of different research areas in computer science, neuroscience, health and medical science, and biomedical engineering [14].

EEG-based emotion recognition is broadly used in entertainment, e-learning, and healthcare applications. EEG is utilized for different purposes—for example, instant messaging, online games, assisted therapy, and psychology [15].

Capturing human brain patterns is most efficient when the person is relaxed and has his/her eyes closed. Normally, they are estimated from peak to peak with a range from 0.5 to 100  $\mu$ V in amplitude, which is around 100 times below EEG signals [16].

Human brain waves have been classified according to different frequency collections: delta (0.1–4) Hz, theta (4–8) Hz, alpha (8–13) Hz, beta (13–30) Hz, and gamma (30–64) Hz [17]. Alpha can be normally noticed more easily in the posterior, and action is provoked by closing the eyes and by relaxation, by eye-opening, or by warning through any status (thinking and computation). Beta waves begin to appear at a high frequency of more than 14 Hz and reach 80 Hz during tension. Theta waves are at a frequency of (4–7) Hz, theta waves appear when normal sleep and deep meditation, and delta waves at less than (3.5) Hz occur with deep sleep and guiding meditation [16].

This work investigates human emotions based on EEG signals by applying machine learning methods to detect and classify various human emotions.

## 2. Related Work

Santamaria-Granados et al. [18] applied a deep convolution neural network on the AMIGOS dataset [19] of physiological signals (electrocardiogram and galvanic skin response). The study used advanced classic machine learning approaches to obtain the properties of physiological signals in the time, frequency, and nonlinear fields. This method accomplishes greater precision in the classification of emotional states.

Bazgir et al. [20] applied EEG signals from the DEAP dataset to recognize an emotion according to the valence/arousal model. Support Vector Machine (SVM),  $k$ -nearest neighbor ( $k$ -NN), and artificial neural network (ANN) classifiers are classified as emotional states. Further information about the DEAP dataset can be found in Section 4.1. The experiment showed a 91.3% accuracy for arousal and a 91.1% accuracy for valence in the beta frequency band using the cross-validated SVM with a radial basis function (RBF) kernel.

Alhagry et al. [21] used a deep learning approach to recognize emotion from raw EEG signals after applying long short-term memory (LSTM) to detect features from EEG signals next to the dense layer, and features were classified into low/high arousal, valence, and liking sequentially. The DEAP dataset was used to verify this method, which provided an average accuracy of 85.65%, 85.45%, and 87.99% for the arousal, valence, and liking classes, sequentially.

Mehmood et al. [22] produced EEG signals from special sensors that measured electrical activity for 21 healthy cases based on recordings from 14-channel.

The EEG signals were captured while the subjects looked at images, and four models of emotional stimuli (happy, calm, sad, or scared) were considered. The feature extraction phase used a statistical approach based on specific features for different frequency ranges. Features chosen by this statistical approach exceeded univariate and multivariate features. The optimal features were additionally prepared for emotion classification by applying SVM,  $k$ -NN, linear discriminant analysis, naïve Bayes, random forest, deep learning, and four ensembles methods. The outcomes reveal that the suggested method gave good results regarding classifying emotions.

Al-Nafjan et al. [2] used a deep neural network (DNN) to identify human emotions from EEG signals taken from the DEAP dataset. The suggested method was compared to state-of-the-art emotion detection systems using the same dataset. The study showed how EEG-based emotion recognition can be performed by applying DNNs, particularly for a large number of training datasets.

Based on the previously discussed literature, there are common and unique issues about the conducted approaches for emotion detection based on different classifiers, which can be summarized as follows. First, classifiers that are utilized in the literature are varied. As noted, most of these conducted experiments over emotion detection, in general, use different classification algorithms. Second, different emotion states are used for classification together with the selected classification algorithm. Third, most of the previously mentioned approaches used the DEAP dataset because it is applicable for the analysis of human affective states and publicly available datasets. However, the accuracy of some approaches has reached above 91.3%, the best approach being with the DEAP dataset. Moreover, the complexity of the existing approaches is high if real-time processing is implemented. Accordingly, there is a need to enhance the accuracy of emotion detection and classification and reduce the complexity of the utilized approaches. The comparison is presented in Table 1.

**Table 1.** Summary of the related work.

Ref.	Classification	Dataset	Results
Alhagry et al. [21]	Deep learning	DEAP dataset	85.65%, 85.45%, and 87.99% with arousal, valence, and liking classes
Mehmood et al. [22]	Deep learning ensembles method	IAPS <sup>1</sup>	highest average accuracy 76.62%
Li et al. [23]	Hierarchical convolutional neural networks	SEED <sup>2</sup>	beta wave high 86.2% and gamma wave high 88.2%
Bazgir et al. [20]	SVM, <i>k</i> -NN, and ANN	DEAP dataset	91.3% accuracy for arousal and 91.1% accuracy
Putra et al. [24]	Wavelet decomposition and <i>k</i> -NN	DEAP dataset	57.5% accuracy
Zangeneh Soroush et al. [25]	Machine learning	Iranian movies	90% detection
Santamaria-Granados et al. [18]	Deep convolution neural network	AMIGOS dataset	99% detection
George et al. [26]	Time–frequency domain statistical features	DEAP dataset	92.36% accuracy
Chen et al. [27]	Deep convolution neural network (CNN)	DEAP dataset	58% in valence and 3.29% in arousal

<sup>1</sup> <https://csea.phhp.ufl.edu/Media.html#topmedia>; <sup>2</sup> <http://bcmi.sjtu.edu.cn/~seed/seed.html>.

### 3. Proposed Work

Figure 2 shows the proposed framework. It shows the main steps for preprocessing stage, feature extraction, and classification. This study focuses on using different techniques for the preprocessing stage. The framework uses Empirical Mode Decomposition/Intrinsic Mode Functions (EMD/IMFs) and Variational Mode Decomposition (VMD). The EMD/IMF and VMD are widely used in biomedical and disease-related studies, but they are not commonly utilized in emotion recognition [28]. The following steps illustrate the procedures and technologies used:

- The test signals are divided into two groups. The first group, called wanted signals, consists of signals that are taken for further investigation and phases in this work. This group depends on the brain signals sensed through 32 channels. The alpha, beta, gamma, delta, and theta channels are cleaned, denoised, and filtered. The second group is called the unwanted signals. These signals are used later in cross-checking model accuracy in order to ensure the accuracy, correctness, and logic of the obtained results.
- The denoised phase involves cleaning the data using EMD/IMFs and VMD filters. This aims to remove any artifacts and noise in the signals. Using these filters in this step is to ensure that the signals are clean and ready to be processed and classified.
- The feature extraction step is to increase the accuracy of the classifiers through obtaining the most valuable features from the signals. This phase uses two types of feature extraction methods: entropy study (SE) and Higuchi's fractal dimension (HFD).
- In the classification phase, four main machine learning (ML) algorithms will be used. The algorithms are naïve Bayes, *k*-nearest neighbor (*k*-NN), convolutional neural network (CNN), and decision tree (DT). Each classifier differs in its approach. The classes will be processed and classified with the same data that have been cleaned, filtered, and featured.

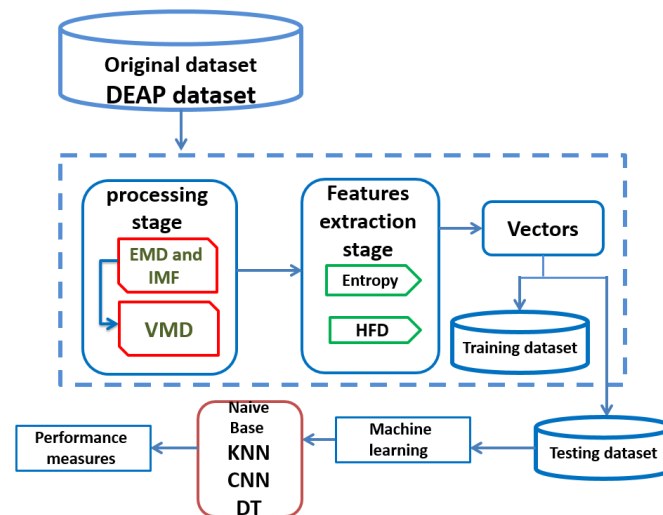


Figure 2. The proposed framework for emotion classification.

## 4. Experiment Tests

### 4.1. Dataset

DEAP (<https://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html>) is a dataset available freely on the Internet for studying human emotions using EEG signals. The DEAP dataset consists of two parts [8]:

1. The ratings from an online self-assessment where 12 one-minute extracts of music videos were each rated by 14–16 volunteers based on arousal, valence, and dominance.
2. The participant ratings, physiological recordings, and face videos of an experiment where 32 volunteers watched a subset of 40 of the above music videos. EEG and physiological signals were recorded, and each participant also rated the videos as above. For 22 participants, a frontal face video was also recorded. The duration of each video is 60 s. This specific minute was chosen because it was the one in which the emotion was stimulated.

In this work, MATLAB 2018 libraries were used for implementing the work, starting from preprocessing the data, filtering, feature extraction, and ending with classifications. Tests were carried out using an Intel Core i7 central processing unit (CPU), 16 GB RAM, and 2 GB Nvidia GeForce.

### 4.2. Dataset Cleaning, Filtering, and Feature Extraction

#### 4.2.1. Denoised Signals

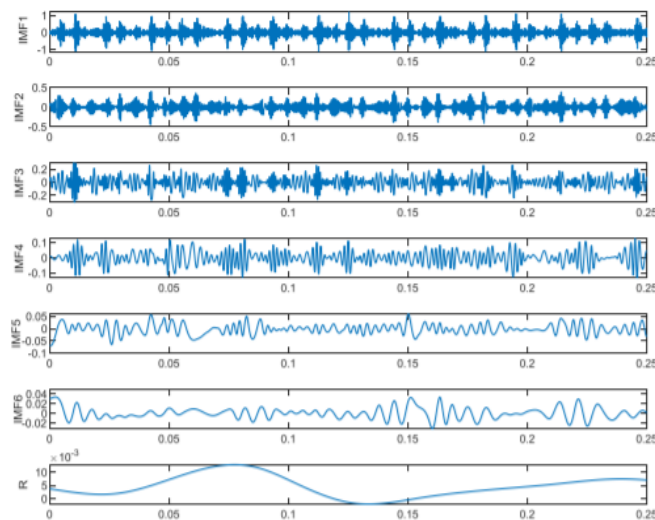
An EEG measures the electromagnetic behavior of the brain at a fairly low pressure, which also interferes with the signal reported by specific intrinsic and extrinsic components. The captured EEG signals contain various intrinsic anomalies, including the activity of the limbs, the pulse, the motion of the body, and the concentration of the mind. These anomalies and other types of artifacts, such as artificial noise and frequency components, will affect the brain function measurement. A two-stage filter (both EMD/IMF and VMD filters) method is proposed to work on cleaning the input signals.

EMD/IMF is a method proposed for the decomposition of signals of nonlinear and nonstatic signals. EMD divides the signal into a set of inherent functions (IMFs). Every IMF can be used as a sub-band signal. The EMD will then be used to decompose the substrip signal [29].

Figure 3 shows the effect of applying EMD on the used EEG signal. It will split the signals into different types of frequencies through which the system can identify high and low frequencies (min-max). For a high frequency, the filter will recognize the pattern of the wave and the general appearance and then start drawing the path and lines to minimize the sharp edges of the waves and



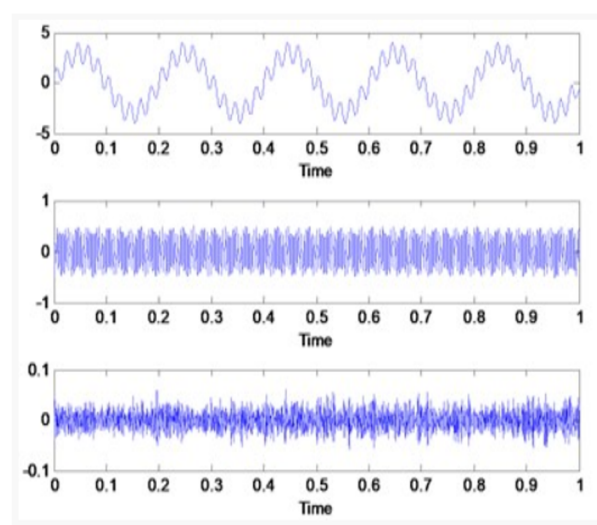
to have a general pattern of the wave. After using the smoothing filters, the wave (signal) will be more understandable and clearer for the classifiers, which will save time and performance. Generally, there is no data loss using this smoothing filter.



**Figure 3.** Empirical mode decomposition/intrinsic mode functions (EMD/IMF) function with respect to time.

On the other hand, VMD attempts to split an input message into many subsignals (modes), where each mode's bandwidth is diminished. Any mode  $k$  must therefore be compact in the middle, together with decomposition, around a pulsation. For each mode of VMD, using the Hilbert transform to measure a corresponding scan signal, the mode's frequency ranges are passed to the baseband by integrating a corresponding analytical signal at the right-center frequency and approximating the one-dimensional signal bandwidth [30].

Figure 4 shows the effect of using VMD for filtering data. The VMD is simply calculated using the Hilbert transform for each mode to obtain a unilateral spectrum of frequencies. The frequency range of modes is transferred to a determined middle frequency by combining it with an exponential. Bandwidth is determined by the demodulated signal's Gaussian smoothness. Both filters will result in a clear signal that is cleaned from artifacts, noise, and any outside effect that affected the signals during recording.



**Figure 4.** Variational mode decomposition (VMD) filter response with respect to time.

#### 4.2.2. Feature Extraction Methods

Generally, feature extraction methods aim for the most valuable information from any studied signal. This information can be either statistical or nonstatistical.

The resulting signals, after applying EMD/IMF and VMD filters, include emotional information from nonlinear measures. This study focuses on the following features: entropy and Higuchi's fractal dimension (HFD).

The complex, nonlinear, and nonstationary EEG signals are one of the challenges for EEG data recovery. The signal characteristics are not constant but are understood to be constant either for a long duration or a shorter time. In effect, various linear extraction approaches use the short-term windowing technique to follow EEG signals. However, even during mental and physical exercise, this assumption is not valid in common brain conditions. Nonstationary EEG patterns may be observed through alertness and wakefulness transitions. Several nonlinear study alternatives, such as entropy, have also been proposed because the randomness of nonlinear time series data is incorporated into the time series entropy calculation [8].

Entropy can be used to calculate the instability level of the device in brain–computer communication systems. This is a nonlinear calculation. The sum of uncertainty in a time series is quantified. Entropy indicates how much the results of each trajectory can be predicted from each other. Higher entropy implies, in the final analysis, more complex or chaotic systems. Spectral entropy has been used effectively to date in EEG feature extraction. Consequently, entropy was not used for immediate appreciation. They believe that entropy provides useful information and unique features that can also be used for the classification of individuals.

On the other hand, Higuchi's fractal dimension (HFD) is a nonlinear method, has occupied an important place in the analysis of biological signals. The use of HFD has evolved from EEG and single-neuron activity analysis to the most recent application in automated assessments of different clinical conditions. The speed, accuracy, and cost of applying the HFD method for research and medical diagnosis make it stand out from the widely used linear methods. However, only a combination of HFD with other nonlinear methods ensures reliable and accurate analysis of a wide range of neurophysiological signals [31].

#### 4.2.3. Classifiers

In the classification step, a model is developed after a feature extraction procedure with the training samples. The model is also used to determine the efficiency of the emotion classification method during the training period. The suggested solution incorporates different classification algorithms.

The  $k$ -NN algorithm is nonparametric, as defined for a particular data point, due to the heterogeneity of its neighbors.  $k$ -NN consists of two phases: defining the number of nearest neighbors and classifying the data point. This uses distance metrics such as Euclidean distance to locate the next neighbor. The teaching method chooses the closest  $k$  samples and takes a plurality vote of its sort, where  $k$  is an odd number for preventing ambiguity [32,33].

Decision tree (DT) is a structured method to construct classification models from the input dataset using a decision tree. A variety of test questions in a tree system is arranged by decision tree classifiers. Every node in a decision tree is subject to a check condition (i.e., yes or no). The evaluation cycle begins from the root node, and the test condition is added to the report input and is centered on the test results followed by the related branch [33,34]. Signal values of the EEG are included in the decision tree database, and the decision tree is structured to interpret the EEG values and outcomes as various emotional forms (positive and negative).

Naïve Bayes suggests that the existence of a chosen feature is not related to the occurrence of another feature in certain groups. This classifier assumes that the features are independent from each other with respect to the classes. Despite this simple assumption, it is considered efficient and easily



implemented. Naïve Bayes is particularly suitable for higher dimensionality. This classifier is then checked in an experiment [33].

CNN is a typical and widely used model for deep learning. Deep learning aims to automatically learn and extract multilevel feature representation from raw data. The characteristics of CNN, such as local connection, weight sharing, and downsampling operation, make it possible to effectively reduce the complexity of the network, reduce the number of training parameters, and present the advantages of strong robustness and fault tolerance, as well as being easy to train and optimize. Multiple filters or kernels were convolved with the input data in terms of vectorized EEG epochs in each convolutional layer, and these layers were designed to capture different local temporal and spatial EEG features. The output of a convolution layer from one kernel is called a feature map (FM). All the output feature maps are combined by the fully connected layers at the end of the last convolution layer [35].

## 5. Results, Discussion, and Comparison

In this work, training, validation, and testing of the data are performed. Figure 5 shows the data regarding the machine learning classifications. Three major sizes of testing and training data were used on this method to obtain the accuracy and run time. The sizes are as follows:

- 80% for the training and 20% for the testing.
- 70% for the training and 30% for the testing.
- 50% for the training and 50% for the testing.

The training phase involves splitting the data, shuffling, and random training to obtain the best accuracy rate for the different machine learning algorithms. The testing phase is the same as the training phase in order to test the model in all possibilities that the dataset presents. The following measurements are used to test the performance of each one of the used classifiers: sensitivity (SN), specificity (SP), positive predictive (PPV), and accuracy (ACC).

In this phase, the 40 signals and channels are divided between actual (32) and non-actual (8) brain signals, the latter of which is used later in the cross-check method after obtaining results from the classifiers.

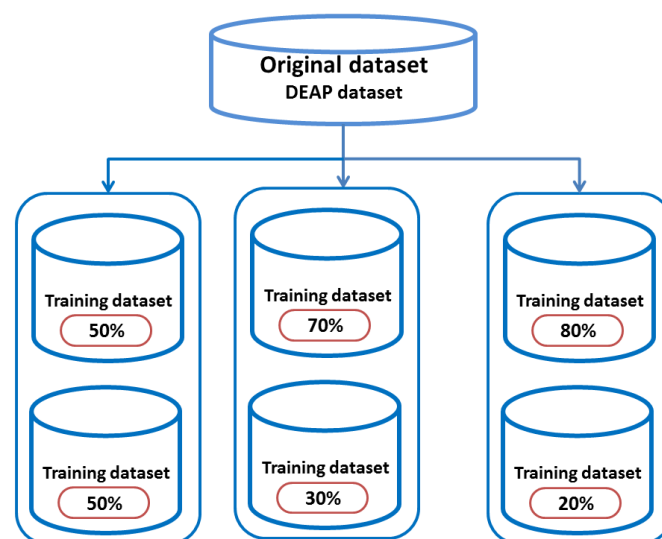


Figure 5. The different dataset split for testing the proposed model.

### 5.1. Results

The classification process is done based on two stages: training and testing. In each task, the training and testing processes are implemented in  $n$ -folds, where  $n$  is set to 10. In  $n$ -folds, the data are divided into  $n$  equal folds, and the experiments are conducted in  $n$ -rounds. In each round,  $n - 1$

folds are used for training and 1-fold for testing. Accordingly, each of the folds is used as a testing set in each round leading to tests of all the available data. The results are reported as the results of all folds.

For the emotion classification task, four subclasses are presented: happy, calm, angry, and sad. Based on these subclasses, two main classes are calculated: valence and arousal. The comparison between the classifiers is based on the size of the training and testing data as well as each classifier's run time and performance. Other researchers' results are lastly compared with the proposed model.

Table 2 presents, for each dataset, the testing and training data sizes. The first table, which contains the valence and arousal results, shows the accuracy of each section with and without the other brain signals. The mean and standard deviation are shown for all results obtained from each instance of valence and arousal, using the EEG signals alone, the other brain signals and power alone, and all signals obtained from the brain together. The next table shows the overall accuracy with other accuracy measurements for each classifier and plots the results for a visual representation of the results obtained.

**Table 2.** Mean and standard deviations for all sets of experiments.

Emotion	EEG	Peripheral	Both
Valence	Mean: 90.21% SD: 6.306	96.31% 96.09% SD: 6.186	SD: 5.367
Arousal	Mean: 95.03% SD: 9.486	98.83% 90.61% SD: 4.455	SD: 3.579

Participants experienced sadness and happiness emotions, and these were reflected in the brain signals. Furthermore, calmness and boredom emotions were experienced to a smaller degree, which indicates that the participants had stopped paying attention over time or that the videos were replayed.

## 5.2. Classifier Results

Tables 3 and 4 show the results obtained from each classifier ( $k$ -NN and CNN) whereby the parameters were chained. The overall accuracy is shown in the tables in this section.

**Table 3.** The classification accuracy using  $k$ -nearest neighbor ( $k$ -NN) based classifier with different  $k$  values.

$k$	Results
1	92.38%
3	93%
5	93%
7	86.8%

**Table 4.** The classification accuracy using convolutional neural network (CNN) based classifier with different epoch, layers, and hNodes.

Epoch	Layers	hNodes	Results	Overall Results
10	5	10	91.2%	91.4918%
20	10	20	94.9608%	95.2647%
40	20	30	90.6048%	90.8947%
100	100	50	87.12%	87.3988%

Table 3 shows the results based on different  $k$  values in order to know which  $k$  gives the best results. It is shown that the accuracy is 93% when  $k = 3$  and  $k = 5$ , whereas when  $k = 7$ , the accuracy is 86.8%. This is due to the fact that the smaller the value of  $k$  is, the more accurate the result is.

The method of preparation and testing is as follows. The whole sample was split into 10 sections, including nine training pieces and one testing portion. Each element per study was special. The other nine sections were used for preparation 10 times for the overall exercise and examination, and tests from training and testing did not always overlap. Therefore,  $k$  was set to three and five.

The best results were in  $k$ -NN, when  $k = 3$ , and in CNN, when epoch, layers, and hidden nodes (hNodes) = 20, 10, and 20, respectively. These results and parameters were selected to be set in the next tests for comparison between classifiers and to be used in our test for the experiment designed for the proposed method.

Table 4 shows the results of applying CNN with different values of epochs, layers, and hNodes. The best results were obtained when  $epoch = 20$ ,  $layer = 10$ , and  $hNode = 20$ .

Tables 5 and 6 display three instances of data splitting to train and test the DEAP and the classifiers. The results show each splitting result. A summary and discussion are presented below. These results show the performance accuracy of 80% training and 20% testing of signals.

**Table 5.** Classification results of 80% training and 20% testing of signals (valence and arousal).

Classifier	Valence			Arousal		
	Precision	Recall	F1	Precision	Recall	F1
$k$ -NN	93.09	93.18	93.32	93.60	93.95	93.35
DT	92.65	91.23	92.06	91.22	91.56	91.28
NB	91.93	92.51	91.72	92.32	92.62	92.80
CNN	95.20	95.51	94.92	95.49	95.41	95.22

**Table 6.** Classification results of 80% training and 20% testing of signals (specificity (SP), sensitivity (SN), positive predictive (PPV), and accuracy (ACC)).

Classifier	SP	SN	PPV	ACC
$k$ -NN	94.03	94.03	94.03	94.03
DT	88.50	88.50	88.43	88.50
Naïve Bayes	92.27	92.27	92.26	92.27
CNN	94.93	94.93	94.93	94.93

Tables 5 and 6 show the results of dividing the dataset into two groups: 80% and 20% for testing the accuracy of emotion classification. The obtained results were higher since the percentage of the trained data was large. The CNN classifier obtained the highest values regarding arousal and valence, shown in Table 5, which was due to the convolution layer. Regarding the decision tree and naïve Bayes classifiers, the results were close. Similar results were obtained in terms of accuracy, shown in Table 6, where the highest value was obtained with CNN and the lowest value was obtained with decision tree.

The dataset was preprocessed using the previously explained filters and feature extraction algorithms. The model ran each classifier separately. The results are based on all channels and signals, which were studied.

Tables 7 and 8 show the results of classification based on the experiment of the two datasets: 70% training and 30% testing. Again, the CNN classifier yielded the highest accuracy for both arousal and valence.

**Table 7.** Classification results of 70% training and 30% testing of signals (valence and arousal).

Classifier	Valence			Arousal		
	Precision	Recall	F1	Precision	Recall	F1
$k$ -NN	93.09	93.18	93.32	93.60	93.95	93.35
DT	92.65	91.23	92.06	91.22	91.56	91.28
NB	91.93	92.51	91.72	92.32	92.62	92.80
CNN	95.20	95.51	94.92	95.49	95.41	95.22

**Table 8.** Classification results of 70% training and 30% testing of signals (specificity (SP), sensitivity (SN), positive predictive (PPV), and accuracy (ACC)).

Classifier	SP	SN	PPV	ACC
<i>k</i> -NN	92.51	90.65	90.84	91.54
DT	90.17	91.79	91.90	91.03
NB	90.36	90.18	91.73	90.34
CNN	94.48	93.01	92.01	93.79

Finally, Tables 9 and 10 show the results of dividing the dataset into 50% and 50% for testing and training. This group yielded the lowest values because the percentages of the trained and tested dataset were equal. Nevertheless, the CNN classifier still showed the highest accuracy.

**Table 9.** Classification results of 50% training and 50% testing of signals (valence and arousal).

Classifier	Valence			Arousal		
	Precision	Recall	F1	Precision	Recall	F1
<i>k</i> -NN	90.27	90.02	90.21	90.94	90.37	90.55
DT	89.64	90.48	90.36	89.56	89.80	90.20
NB	92.05	92.00	90.75	91.32	90.55	90.88
CNN	94.09	93.65	94.41	93.58	94.34	93.55

**Table 10.** Classification results of 50% training and 50% testing of signals (SP, SN, PPV, and ACC).

Classifier	SP	SN	PPV	ACC
<i>k</i> -NN	90.45	90.45	90.45	90.45
DT	90.15	90.15	90.15	90.15
NB	991.91	91.91	91.91	91.91
CNN	94.26	94.26	94.26	94.26

In the previous results, the CNN had a better performance for each test and train size. The accuracy decreased in each run when the training size was smaller. The 80% training size achieved better results than the 50% size due to the amount of training data and the amount of test data.

In general, all classifiers could detect emotions from the DEAP dataset and could classify and process the signals.

### Comparison

A confusion matrix is a technique for summarizing the performance of a classification algorithm. Classification accuracy alone can be misleading if one has an unequal number of observations in each class or if one has more than two classes in a dataset. Calculating a confusion matrix can give one a better idea of which types of errors a classification model is making. This matrix can be used for two-class problems that are easy to understand, but it can also be easily applied to problems with three or more class values by adding more rows and columns to the confusion matrix. Table 11 shows the values of the confusion matrix for testing the correctness of the used data. For example, in the sadness cases, the percentage of correctness was 68%.

**Table 11.** Confusion matrix.

Category	Normal	Happy	Angry	Sad	Afraid	Total
<b>Normal</b>	66.3	2.5	7.0	18.2	6.0	100%
<b>happy</b>	11.9	61.4	10.1	4.1	12.5	100%
<b>angry</b>	10.6	5.2	72.2	5.6	6.3	100%
<b>sad</b>	11.8	1.0	4.7	68.3	14.3	100%
<b>afraid</b>	11.8	9.4	5.1	24.2	49.5	100%

### 5.3. Comparison with Other Model Results

Finally, Table 12 compares the proposed work with others that used the same DEAP dataset. The proposed work yielded better results, where the accuracy is 92.44%.

However, the CNN classifier yielded better results than  $k$ -NN. However, it required more time due to the number of layers and calculations.  $k$ -NN yielded similar results to CNN but in a shorter time.

**Table 12.** Comparison between the proposed model and other previous research.

Method	Classifier	Dataset	Results
Wang and Shang [36]	SVM, NN	DEAP	60.9
Choi Eun Jeong [37]	LSTM, NN	DEAP	73.05
Rodriguez et al. [38]	SVM	DEAP	81.46
Bálan O [39]	SVM, NN	DEAP	90.75
The proposed method	CNN, $k$ -NN, NB, DT	DEAP	92.44

## 6. Conclusions

The evolution in the creation of sensors and signal record devices, as well as the development of signal handling and feature extraction techniques, has increased opportunities for using signals extracted from human organs, such as brain signals or heart signals, to identify a person's condition, and thus detect psychological or pathological conditions in humans. This made the task of classifying signals required for improving the productivity of performance in the categorization of cases based on signals.

Categorizing emotions based on EEG signals could be one of the most complex applications with regard to analyzing human actions. This type of application can be defined as determining a person's emotional state, which could reflect particular problems. EEG data can be extracted using different systems or devices. In this study, a DEAP dataset was used to identify and classify human emotions.

The proposed model in this paper is based on three main steps: processing, feature extraction, and classification. In the signal processing stage, three different techniques were used, including EMD/IMF and VMD, to remove noise from the signals and clean them to obtain the best possible details from the primary EEG data. For the feature extraction method, three methods were adopted to provide the classifiers with refined data for their classification and prediction.

In the classification stage, four main classifiers were used:  $k$ -NN, decision tree, naïve Bayes, and CNN. These were used to classify and define human feelings. After applying these classifiers under different criteria, each classifier yielded different results and running times, and these results were studied. It was concluded that the CNN classifier yielded the best results in terms of model performance. The work also contains a section for comparing the results of the proposed method with the work and results of other studies, which showed that the proposed method had better results in runtime and accuracy for predicting arousal and valence, and thus human emotions in general.

There are several differences in the performance of machine learning classifiers in terms of accuracy, precision, recall, and  $F1$ -measure. Through our tests, we found that the CNN was the best in terms of accuracy. Results also showed that the results of NB and  $k$ -NN were convergent. However, CNN outperformed other methods in EEG signal categorization. When applying an  $F1$ -measure on various cases and different classifiers, CNN yielded the highest  $F1$ -measure and accuracy in all cases.

**Author Contributions:** This work is part of a master’s thesis submitted by S.A. for the fulfillment of a master’s degree in Computer Science at Mutah University, Jordan. The idea of the work was conceptualized by R.A. Methodology and validation were provided by R.A. and S.A. Software implementation and visualization were completed by S.A. Writing—original draft preparation was completed by S.A. and R.A. Writing—review and editing were completed by R.A. Project administration was conducted by R.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

1. Jerry, S.; Dean Krusienski, J.W. Brain-computer interfaces in medicine. *Mayo Clin. Proc.* **2012**, *87*, 268–79.
2. Al-Nafjan, A.; Hosny, M.; Al-Wabil, A.; Al-Ohali, Y. Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 419–425. [\[CrossRef\]](#)
3. Damasio, A.R. *Descartes’ Error. Emotion, Reason and the Human Brain*; Avon Books: New York, NY, USA, 1994.
4. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [\[CrossRef\]](#)
5. Duan, R.; Zhu, J.; Lu, B. Differential entropy feature for EEG-based emotion classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; pp. 81–84.
6. Zheng, W.; Lu, B. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [\[CrossRef\]](#)
7. Kim, M.K.; Kim, M.; Oh, E.; Kim, S.P. A review on the computational methods for emotional state estimation from the human EEG. *Comp. Math. Methods Med.* **2013**, *2013*, 573734. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. Plutchik, R. *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*; American Psychological Association: Washington, DC, USA, 2003.
10. Izard, C.E.E. Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues. *Annu. Rev. Psychol.* **2009**, *60*, 1–25. [\[CrossRef\]](#)
11. Yu, L.-C.; Lee, L.-H.; Hao, S.; Wang, J.; He, Y.; Hu, J.; Lai, K.R.; Zhang, X. Building Chinese Affective Resources in Valence-Arousal Dimensions. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 540–545.
12. Niedermeyer, E.; da Silva, F. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; LWW Doody’s All Reviewed Collection; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.
13. Sano, A.; Picard, R.W. Stress Recognition Using Wearable Sensors and Mobile Phones. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 671–676.
14. Acharya, U.R.; Hagiwara, Y.; Deshpande, S.N.; Suren, S.; Koh, J.E.W.; Oh, S.L.; Arunkumar, N.; Ciaccio, E.J.; Lim, C.M. Characterization of focal EEG signals: A review. *Future Gener. Comput. Syst.* **2019**, *91*, 290–299. [\[CrossRef\]](#)
15. Alarcão, S.M.; Fonseca, M.J. Emotions Recognition Using EEG Signals: A Survey. *IEEE Trans. Affect. Comput.* **2017**, *10*, 374–393. [\[CrossRef\]](#)
16. Teplan, M. Fundamental of EEG Measurement. *Meas. Sci. Rev.* **2002**, *2*, 1–11.
17. Konar, A.; Chakraborty, A. *Emotion Recognition: A Pattern Analysis Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
18. Santamaria-Granados, L.; Muñoz-Organero, M.; Ramírez-González, G.; Abdulhay, E.; Arunkumar, N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS). *IEEE Access* **2018**, *7*, 57–67. [\[CrossRef\]](#)
19. 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**. [\[CrossRef\]](#)



20. Bazgir, O.; Mohammadi, Z.; Habibi, S.A.H. Emotion Recognition with Machine Learning Using EEG Signals. In Proceedings of the 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME), Qom, Iran, 29–30 November 2019.
21. 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*, [CrossRef]
22. Mehmood, R.M.; Du, R.; Lee, H.J. Optimal Feature Selection and Deep Learning Ensembles Method for Emotion Recognition From Human Brain EEG Sensors. *IEEE Access* **2017**, *5*, 14797–14806. [CrossRef]
23. Li, J.; Zhang, Z.; He, H. Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition. *Cogn. Comput.* **2017**, *10*, 368–380. [CrossRef]
24. Putra, A.E.; Atmaji, C.; Ghaleb, F. EEG-Based Emotion Classification Using Wavelet Decomposition and K-Nearest Neighbor. In Proceedings of the 2018 4th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 18–19 October 2018; Volume 1, pp. 1–4.
25. Zangeneh Soroush, M.; Maghooli, K.; Setarehdan, S.K.; Nasrabadi, A.M. Emotion Classification through Nonlinear EEG Analysis Using Machine Learning Methods. *Int. Clin. Neurosci. J.* **2018**, *5*, 135–149. [CrossRef]
26. George, F.P.; Mannafee, I.; Hossain, P.; Parvez, M.Z.; Uddin, J. Recognition of emotional states using EEG signals based on time-frequency analysis and SVM classifier. *Int. J. Electr. Comput. Eng. (IJECE)* **2019**, *9*, 1012. [CrossRef]
27. Chen, J.X.; Zhang, P.W.; Mao, Z.J.; Huang, Y.F.; Jiang, D.M.; Zhang, Y.N. Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 44317–44328. [CrossRef]
28. Zhuang, N.; Zeng, Y.; Tong, L.; Zhang, C.; Zhang, H.; Yan, B. Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain. *BioMed Res. Int.* **2017**, *2017*, 531–544. [CrossRef]
29. Zeiler, A.; Faltermeier, R.; Keck, I.; Tomé, A.; Puntotnet, C.; Lang, E. Empirical Mode Decomposition—An Introduction; In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8.
30. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
31. Kesić, S.; Spasic, S. Application of Higuchi’s fractal dimension from basic to clinical neurophysiology: A review. *Comput. Methods Programs Biomed.* **2016**, *133*, 55–70. [CrossRef] [PubMed]
32. Tan, S. Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Syst. Appl.* **2005**, *28*, 667–671. [CrossRef]
33. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2013.
34. Valecha, H.; Varma, A.; Khare, I.; Sachdeva, A.; Goyal, M. Prediction of Consumer Behaviour using Random Forest Algorithm. In Proceedings of the 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, India, 2–4 November 2018; pp. 1–6.
35. Hajinoroozi, M.; Mao, Z.; Jung, T.; Lin, C.T.; Huang, Y. EEG-based prediction of driver’s cognitive performance by deep convolutional neural network. *Signal Process. Image Commun.* **2016**, *47*, 549–555. [CrossRef]
36. Wang, D.; Shang, Y. Modeling Physiological Data with Deep Belief Networks. *Int. J. Inf. Educ. Technol. (IJJET)* **2013**, *3*, 505–511.
37. Choi, E.J.; Kim, D.K. Arousal and Valence Classification Model Based on Long Short-Term Memory and DEAP Data for Mental Healthcare Management. *Healthc. Inform. Res.* **2018**, *24*, 309–316. [CrossRef]
38. Rodriguez, A.; Angel, M.; Flores, M. Classification model of arousal and valence mental states by EEG signals analysis and Brodmann correlations. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 230–238. [CrossRef]
39. 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]

**Sample Availability:** Samples of the compounds are available from the authors.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).