Stacking Ensemble Learning Approaches Applied to Emotional State Classification

Luiza Scapinello Aquino
Electrical Engineering Graduate Program (PPGEE)
Federal University of Parana (UFPR)
Curitiba, Brazil
luiza.scapinello@ufpr.br

Leandro dos Santos Coelho
Electrical Engineering Graduate Program (PPGEE)
Federal University of Parana (UFPR) and Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR)
Curitiba, Brazil
leandro.coelho@pucpr.br

Abstract— **The ability to automatically recognize human emotions is a wide field of research that supports both psychology and engineering, through the improvement of human-computer interface systems, and psychiatry, assisting in diagnosis, and treatment decision-making of mental diseases. To deal with big amounts of information, artificial intelligence (AI) approaches are often adopted, such as machine learning (ML), which enables computers to learn and adapt to new situations automatically. This paper aims to compare some ML approaches in classifying the human emotional states from electroencephalogram examination signals through software implementation. The dataset used in this study, known as DEAP (Database for Emotion Analysis using Physiological Signals), comes from a set of exams available for the public by a researcher group from four European Universities. A comparison between the methods decision tree (DT), support vector machine (SVM), and convolutional neural networks (CNN) was made in terms of performance metrics. Moreover, unlike other works focused on these concepts, combinations of these architectures were used in this study to find an accurate result, through the stacking ensemble learning approaches applying in human emotion classifications. Therefore, eight classification methods were applied, three of which have no history of use in this application previously. K-fold and cross-validation were used in order to estimate the capacity of the algorithm of generalization. The best results were obtained with the ensemble of the three base methods, with an accuracy of 0.946. The combination of CNN with SVM was the second best, which obtained a score of 0.922 in the accuracy metric.**

Keywords— machine learning, stacking, ensemble learning, emotion classification, electroencephalography.

I. Introduction

Emotion is a conscious experience of a mental state and an affective reaction to some stimulus of a subjective experience [1]. Emotions may be characterized by a functional mental activity at a certain level of contentment or discontent. In [2] emotion can also be defined as a sudden problem, transient agitation caused by an acute cognitive experience appropriate to this state of emotional arousal. They are essential for the daily communication of human beings, as they are a psychophysiological process that affects the behavior of individuals in certain situations. They also reach the responses of different biological systems, such as facial expressions, voice, and activities of the nervous and endocrine systems [1].

According to [3], it should be easy to determine which emotion a person experiences based on intuition and behavior analysis. However, there is scientific evidence to suggest that measuring a person's emotional state is one of the most complex fields of study. Therefore, scientific studies of emotion with particular emphasis on affective science and neuroscience are important. There is a hypothesis that emotions are generated in the brain, through the neural synapses, which are electrical impulses that could be measured by the equipment. Consequently, the brain is of fundamental importance to human interrelationships, since it processes all the senses stimulated and enables the communication between individuals.

A test adopted in neuroscience for the capture of brain synapses is the electroencephalogram (EEG). The EEG is a non-invasive technique that records the electrical activity and patterns present in the brain, thus analyzing the spontaneous electrical activity from the brain. Electric patterns, also called brainwaves, are produced by the millions of nerve cells that make up the brain. Therefore, to perform the examination small electrodes are positioned on the person's scalp, and EEG signals are essentially recordings collected of the electrical activities of the brain waves, which in turn are amplified and recorded, will be read by a trained neurologist.

As noted by [4] it is possible to recognize, with a satisfactory accuracy metric, human emotions through EEG signals. Recent advances in Machine Learning (ML) have enabled the development of techniques to detect and recognize human emotions. The objective of this paper is to apply and compare the ML approaches, including deep learning and ensemble learning methods, such as decision tree, support vector machine (SVM), convolutional neural networks (CNNs), and stacking, to the classification of human emotions from EEG signal.

Artificial intelligence (AI) can be defined as a machine's ability to perform functions done by humans, which require intelligence. AI is one of the most effective and modern techniques used in the modern healthcare area. One of the successful uses of AI today to support decision-making in medical and psychiatric diagnostics is ML. ML mainly integrates a set of specialized methods in solving problems that usually cannot be easily modeled, as is the case with EEG monitoring systems.

ML can use samples of known data to learn from them and then leverage that knowledge to ascertain unknown data or make decisions based on past experiences [5]. Thus, from a finite set of initial data, the machine performs multiple statistical calculations to recognize the patterns and simulate the behavior of the observed data, being able to then judge the result that was previously unknown. ML is mostly used in pattern recognition, grouping, and feedback of states in the environment [6].

If a computer performs tasks after training and validation periods, it does not currently qualify as error rate and execution error, so ML depends on pattern detection and data classification. Thus, ML has been used to analyze and identify patterns in large and complex datasets, faster and more precisely than previously, being used in researches involving brain and human behavior [6]. Various approaches were adopted for pattern recognition tasks in the EEG dataset such as decision trees (DT) [7] and SVM [8].

Due to the constant need for more complex applications for more advanced and promising paradigms, deep learning (DL) techniques have emerged. The concept of DL is focused on replicating the human learning process, like the functioning of the human brain, with the presence of small units that perform activities similar to neurons [9]. The commonly used DL tool today is CNN, which is inspired by the human visual system (visual cortex). CNN has the predominant characteristic of invariance, allowing neurons to recognize elementary features such as edges and corners [10].

The main contribution of this paper is the development of more efficient ensemble learning approaches to classify human emotion, based on EEGs available by the Database for Emotion Analysis using Physiological Signals (DEAP) [11]. Eight ML algorithms were evaluated in terms of classification performance metrics. The main objective is to find which models will better predict the desired output. Instead of choosing only one classification technique, ensemble learning models were applied, combining different independent base models to generate an optimal predictive model. The combination of multiple models was done through stacked generalization. Furthermore, three ensemble learning techniques used in this study were not found in the literature [12] for this type of application.

The remainder of this paper is structured as follows. In Section II are presented the fundaments of EEG signal processing and emotions. Afterward, in Section III methods and the dataset are described, Section IV exhibits the results, and finally, this study is concluded in Section V.

II. Fundamentals of EEG Signal Processing And Emotions

The electroencephalogram is a record of brain bioelectric activities, corresponding to the flow of information that is processed in the cerebral cortex [13]. To detect these signals, electrodes positioned on the patient's scalp measure the potential difference between two points in the brain or between one point and the earth.

The international 10-20 system specifies the position of the electrodes on the patient's ensuring the reproducibility of the exam. The name refers to the distances between the electrodes, which are always 10% or 20% of the frontal or lateral diameter of the human skull. Thus, regardless of the age or cranial development of the person, the position of the electrodes remains constant [14].

Some facts interfere with the perfect execution of the exam, such as the patient's voluntary and involuntary movements and poor calibration of the device. For the specific purpose of emotion recognition using EEG signals, the second biggest challenge is the variation and personal induction of emotions, as different people demonstrate different emotional responses [4].

The brain's bioelectric activity is classified according to the electrical frequency as a function of the length of each frequency range and the region of signal origin [15]. The frequencies captured from the cortex are related to the person's state of behavior, attention levels, sleep, wakefulness, concentration, and complex cognitive processes. The four most common signal types are illustrated in Fig. 1.

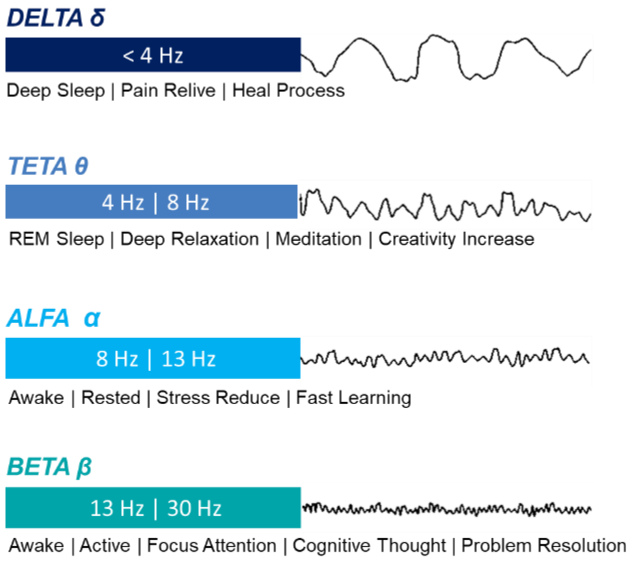

Fig. 1. Brainwaves of a healthy adult.

A. Emotion Modeling

There are three main models for representing a person's emotional state: the discrete model, the dimensional model, and the hybrid model or combinational model. The main difference between them refers to the number of emotions represented [16]. The first model was proposed by [17], consisting of six basic emotions: joy, sadness, surprise, anger, disgust, and fear. These six emotions were chosen by analyzing the facial expressions of people from different cultures, forming the so-called - universal expressions – since that they are recurrent and identical worldwide, regardless of culture. Thus, the discrete model is more common for the recognition of emotions through facial expressions.

The dimensional model shows a greater number of emotions. Besides the basic emotions, it also ranks the correlations between them. Russell [18] proposed a circumplex model of emotion, consisting of a two-dimensional plane with the axes defined as arousal versus valence, punctuated from zero to nine with continuous values, which can be observed in Fig. 2. The (emotional) arousal axis implies that the person is relaxed when the value is lower and stimulated if the value is higher, whereas the (emotional) valence axis indicates an unpleasant to a pleasant state.

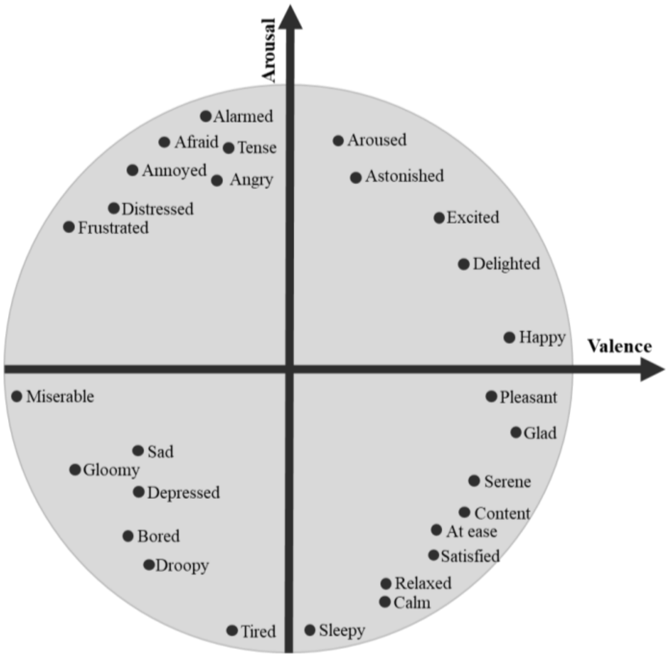

Fig. 2. Russell's circumplex model.

The third model, called the hybrid model, is a coupling of the two previous models. According to Christie and Friedman [19], this model is more effective and consistent when it comes to evaluating emotions with physiological signals (such as the EEG), as it involves both the autonomic nervous system and the self-rated states of individuals. This model is similar to Russell's circumplex model [18], but with the inclusion of discrete emotional states, *e.g.* one state in each quadrant defined by the application.

Physiologically, expressions of emotions are directly related to the limbic system, the brain region located below the cortex, on the medial surface of the brain [20]. Since that this region is not in the first layer of the brain, biopotential registration by the EEG does not yield good results in terms of well recognizable waves. Nevertheless, according to [19], the frontal and prefrontal lobes located in the cerebral cortex, even without being part of the traditional limbic system, taking part an important role in the expression of emotional states and this is due to the thalamus and amygdala connections (section also included in the limbic system) and with other subcortical regions.

III. Related Works

In [12], the authors used a decision tree model to extract data from EEG signals for the recognition of the emotional state and obtained an accuracy of 75.3%. Using the random forest, the average percentage of correct answers was 76.3%. The study [21] obtained an accuracy of 91.5% for SVM application and 85.8% with CNN, also [21] applied the combination of CNN and SVM in which resulted in 85.1% accuracy. In [22], an ensemble learning model with SVM and decision tree was applied to detect individuals' emotional state and stress, and an accuracy rate of 84.6% was obtained. In [23], EEG signals were classified using an adaptive multilayer generalized learning vector quantization algorithm with an accuracy of 62.98%, while with SVM the resulted was 55.77% and with random forest was 54.54%. Comparisons between different evolutionary computation algorithms were proposed in [24] for feature selection in an EEG-based emotion classification model. Such algorithms were applied to improve the performance of classifiers. The proposed method of that research obtained 67.47% of accuracy. More recently, [25] reached a value of 92.19% using the CNN structure of GoogleNet. In [25] two combinations of CNN+SVM were tested, transforming the DEAP database in a new model of topographic and holographic feature maps, with an accuracy of 76.5% and 77.7% respectively.

The methods of stacking ensemble with the combination of CNN + DT, SVM + DT, and CNN + SVM + DT have few records in the current literature for application in human emotional classification, and for this reason, it is not possible to perform a comparative analysis of those models with results obtained in this study.

IV. Data and Methodology

In this section, the DEAP database is explained, from which kind of participants the exam held, to how the data is collected. Moreover, the methodology used in this research is exposed.

A. *Description of the DEAP Database*

The DEAP is a publicly available database composed of physiological signals acquired from participants central nervous system via a 32-channel EEG (BioSemi Active Two) and from the galvanic skin response (GSR) peripheral nervous system whose purpose is to detect the sweat level on the skin, the range of respiration and the surface temperature of the participant's skin. In addition, the database includes an electrocardiogram (ECG), measurement of thumb blood volume measured by a plethysmograph, electromyogram (EMG) to record facial movements when the individual demonstrates a reaction, trapezius to record the possible movement of the head, and electrooculogram (EOG) that records the rate at which the participant blinks [11].

The EEG sampling rate was 512 Hz, however before pre-processing the recorded signal, the sampling rate was downsampled to 128 Hz. The pre-processing phase was achieved by implementing a bandpass filter between the frequencies of 4− 45 Hz, which encompasses the brainwaves θ, α, and β. In this research, nevertheless, the brainwaves required were only the α and β ones, therefore another bandpass filter was needed to remove the 4 – 8 Hz and 30 – 45 Hz frequencies.

To stimulate the emotions in participants, they were encouraged with music and videos. These videos were selected in two distinct steps, firstly the responsible team chooses 120 videos, half of them being chosen with the help of the last.fm music website, which allows users to search their listening habits to receive recommendations. The second half of the videos were manually selected by the search group. Then, were extracted one minute of a segment from each video, which was assisted by volunteers who subjectively assessed their emotions on discrete scales for valence, arousal (emotional), and dominance.

The study group subsequently selected forty videos that obtained the strongest ratings from the volunteers, using Russell's Circumplex model, with still a slight variation to maximize the strength of the emotions elicited, moreover ten videos were selected in each quadrant, in order to guarantee a more balanced response.

The second phase of the DEAP survey involved only the forty most relevant videos and 32 healthy individuals, half male and half female, aged from 19 to 37 years. The experiment allows the recording of the physiological signals of the volunteers who are watching the forty videos. After each video, the participant performs a self-assessment using the self-assessment manikins (SAM) technique, which indicates the level of valence, arousal or dominance, and their relevance through a visual classification choice, as shown in Fig. 4. The dummies are displayed on-screen with numbers from 1 to 9 for the participant to scale their level of self-assessment.

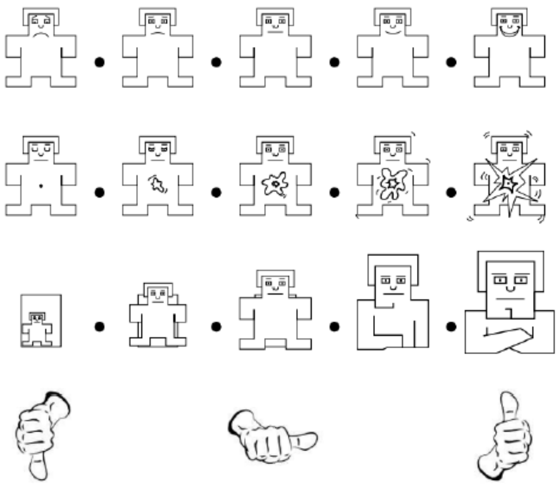

Fig. 3. Images used for self-assessment. From the top: valence SAM, arousal SAM, dominance SAM and acceptance [11].

According to [11], the valence scale ranges from unhappy to joyful, the arousal (emotional) scale ranges from calm or bored to stimulated or lively and the dominance scale ranges from submissive or uncontrolled to dominant or in control. Acceptance has only three stages, dislike, indifferent, and like.

The main reason why this database was chosen, was because of its reliability, and data pre-processing. The final product of the DEAP database used was the forty one-minute long EEG samples for each of the 32 participants, $40 \times 32 = 1280$ instances available. Which were then separated by α and β bands

Most databases for emotional classification or recognition are based on images or videos, such as the AffectNet (database of facial Affect from the InterNet) [27], Google Facial Expression Comparison Dataset [28], and EMOTIC (EMOTions In Context) [29]. Even though the DEAP is based on the response of a relatively small number of participants, other well-known datasets based on EEG signals may have a smaller number of candidates, such as SEED (SJTU Emotion EEG Dataset) [30], DREAMER [31], and IDEA (Intellect database for emotion analysis using EEG signal) [32], which have been performed on 15, 23, and 14 subjects, respectively. The one with the most volunteers in its trials is the database Ascertain (databaASe for impliCit pERsonaliTy and Affect recognitIoN) [33], with 58 applicants.

B. *The Proposed Machine Learning Approach*

The EEG is a test that generates a large amount of data on brain information and electrical activity in time-series samples. Therefore, its analysis by classical methods turns into something also complicated, resulting in the need to use more powerful computational tools. Such tools assist in the extraction of relevant characteristics and the diagnosis and classification of these data [34]. In this study, we compare the supervised learning algorithms DT, SVM, and CNNs for data classification. These are the most used ML techniques, alongside with K-nearest neighbor and artificial neural networks, in this approach, reason why they were chosen to be ensembled.

The dynamics of a Machine Learning algorithm can be divided into three distinct stages. The first involves training with the input data, then an evaluation model of these data is created, called cross-validation. The third and final step is the test, which verifies learning using unknown data to verify the performance of the algorithm.

Decision Tree: This technique is made up of hierarchies, as shown in Fig. 5. It starts with the root node, which has predictive attributes or variables, and proceeds to the inner nodes, which have decision variables at the level intermediate until it reaches the terminal nodes, named leaves. These nodes have decision variable values to classify the instance [35].

The nodes are connected by branches, which have the attribute values of the decision variables. The internal nodes decide which branch will follow the execution until it reaches the terminal node, doing so through a logic test based on sample characteristics [6]. Figure 4 represents a simple DT structure

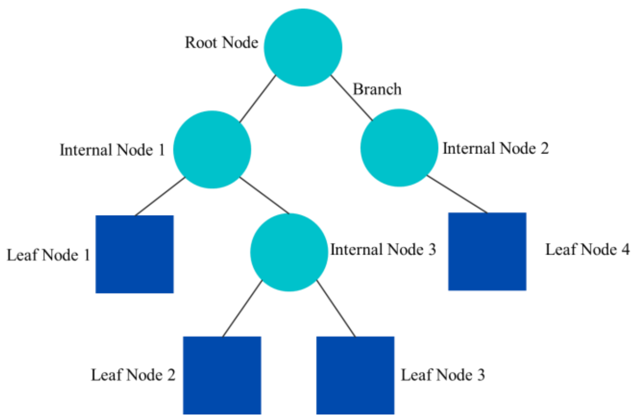

Fig. 4. Decision tree structure.

Support Vector Machine (SVM): It is a supervised machine learning technique used for the classification and regression of linear and nonlinear data [36]. For classification, SVM is based on a binary classifier, that is, given a set of labeled training examples, each information belonging to one of two categories, a model is produced that predicts the correct category of test examples. This model is the representation of the training set as points for a larger dimensional space, mapped with examples from each category and separated by a hyperplane [37].

To select the ideal SVM, the algorithm uses the maximum margin concept. The ideal hyperplane has the best generalizability when it comes to unknown data [38]. When data cannot be separated linearly, the size of the characteristic space is increased to linearize the problem. Meantime, when performing this mapping, the basic operations in SVM become computationally costly. To bypass this problem, Kernels are used, which are mathematical devices of known functions with lower computational cost [38]. In this research, the kernel model used was represented by a radial basis function (RBF).

Convolutional Neural Network (CNN): CNNs have their functioning inspired by the human visual cortex, which works with cortical neurons responding to stimuli in restricted regions in the visual field [39]. A simple CNN is a sequence of layers, each of which transforms one volume of activations into another through a differential function. The three main types of layers in a CNN are: convolution, pooling, and fully connected. The first type is composed of several neurons responsible for applying a filter to a specific part of the data.

Pooling layers combine the outputs of previous layer neurons into a single neuron in the next layer to reduce data size. Fully connected layers connect all neurons in one previous layer to all neurons in another layer [40]. The structure of CNN is illustrated in Fig. 5.

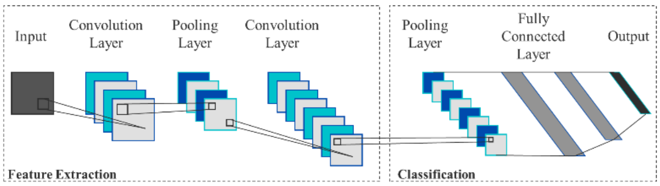

Fig. 5. CNN structure.

The architecture for the chosen CNN contains convolution filters of 3x3 units and 22 depth layers in the network.

Ensemble learning: An ensemble of classifiers combines different generated ML models to create a new predictive model [41]. When the models are generated, they are combined by a grade-level or decision-level committee to make the final decision in each test sample. In the first method, all subjects were treated equally and combined by the average. Zhou et al. [42] and Zhang et al. [43] proposed methods, in which a weight is distributed to each component, indicating the contribution of the corresponding individual in the final decision. Kuncheva [44] developed the majority voting strategy, also known as the decision ensemble.

Breiman [45] proposed a method called bagging, which is the application of the bootstrap process in a high variance machine learning algorithm. Based on this method, Breiman [46] proposed in 2001 the Random Forest, in which a decision tree is considered as the base model and each node is separated by several subsets of randomly selected characteristics.

The ensemble model used in this paper was the stack or stacking method. The focus of stacking is to learn several different learning models and combine them. Thus, the output predictions are based on the multiple predictions returned by these chosen models [47]. Fig. 6 represents the structure of a stacking algorithm.

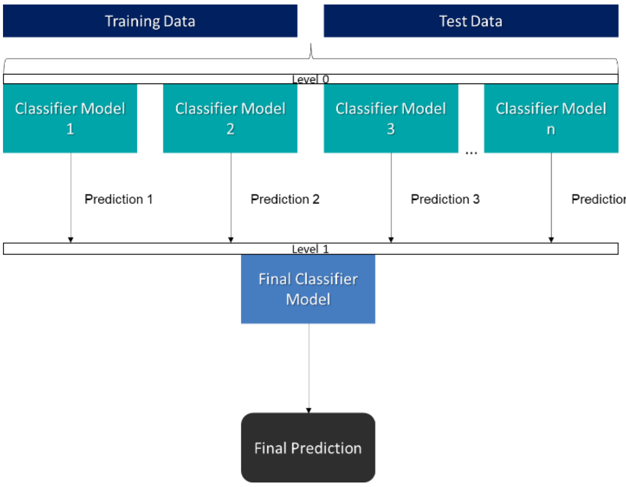

Fig. 6. Stacking learning algorithm architecture.

In Fig. 6, it is possible to comprehend that a stacking learning method involves combining the predictions from multiple ML models on the same training and test dataset. On Level 0, there are the base models, whose predictions are compiled. Level 1, however, represents the meta-model (or final classifier model), which learns how to best merge the predictions of the Level 0 models. Thus, the outputs from the base models are used as inputs for the meta-model [48].

C. *Performance Metrics*

There are different methods to evaluate a supervised learning model. In this study, the most relevant relate binary classification of problems. For example, if a model classifies an event between anomalous (Positive) and non-anomalous (Negative), then there can be cases where the model correctly predicts a positive or negative event and may call them True Positive (TP) and True Negative (TN), respectively. If the model is misclassified, there can be cases of False Positive (FP) or False Negative (FN).

Commonly used metrics include Recall, which describes the fraction of the relevant instances that are found, calculated by

$$Recall = \frac{TP}{TP+FN}. \quad (1)$$

If the rating reaches a Recall value of 1.0 for any class, it means that each item in that class was correctly labeled, but no information on how many items were incorrectly tabulated. To find out this information there is the metric Precision given by

$$Precision = \frac{TP}{TP+FP}. \quad (2)$$

The combination of these two metrics forms the F1-Score, which is better if it is necessary to achieve a balance between

Precision/Recall and if there is an unbalanced distribution of classes [49]. F1-Score is given by

$$F_1 = 2 * \frac{Precision*Recall}{Precision+Recall}. \quad (3)$$

Another metric used is Accuracy, known as the proportion of true results in all results, calculated by

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad (4)$$

D. Classification Approach

The algorithm was developed in Python. As opposed to feedforward neural networks, instead of being trained through error backpropagation method, the artificial neural network is constructed iteratively, one layer at a time, using Wolpert's stack generalization (or stacking) [24], a meta-learning method.

Meta-learning is a learning process for classifiers. Training takes place with two or more stages, unlike with standard learners (only one stage). In this case, firstly, the base classifiers are trained and after the meta-classifier, i.e., the second stage. The base classifiers produce their classifications in the forecasting phase, and then the meta-classifier makes the final classification according to the base classifiers [50].

The stacking meta-combination method trains first-level learners using the original training data set. It generates a new data set to train the second-level learners. First-level outputs are considered input resources and the original labels are still considered labels for the new training data [25]. Unlike bagging and boosting, stacking can combine models of different types.

Algorithm 1 shows a pseudo-code for the stacking procedure. In that figure, D is a dataset with m elements, where x represents the attribute values of each i^{th} instance and y is the class value. In the first step, the algorithm learns h, which is the output of each base-level classifier, and T represents the number of classifiers. Then, a new dataset is created to allow meta-classifier learning. The output from the algorithm is H, which is the ensemble classifier's output.

Algorithm: Stacking

Input: Data set $D = \{x_i, y_i\}_{i=1}^{m}$ *% training data*

Step 1: First-level learning
  for $t = 1$ to T:
    learn h_t based on D
  end
Step 2: Second-level learning *% creating new data set*
  for $i = 1$ to m:
    $D_h = \{x_i', y_i\},\ considering\ \ x_i' = \{h_1(x_i), \ldots, h_T(x_i)\}$
  end
Step 3: Meta-classifier learning *% from new data set D_h*
  Learn H based on D_h

Output: $H(x) = h'(h_1(\boldsymbol{x}), \ldots, h_T(\boldsymbol{x}))$

Algorithm 1: Stacking learning.

Improvements in the results with stacking occur mostly when diversity among the models can be seen, because models with different principles of generalization are inclined to generate different results. Stacking advantages include a better capacity of generalization, when compared to the individual models, and flexibility to adapt to different tasks, however, on the other hand, it has a higher computational cost and a more challenging results interpretability [51].

V. Results and Discussion

The results are presented according to the performance metrics applied in each algorithm. Table I shows the average of the metrics of each model studied after being tested individually ten times, as well as the standard deviation information.

Analyzing the metrics, especially the accuracy, it is notable that the stacking ensemble methods between CNN + SVM and CNN + SVM + Decision Tree (DT) have the best results when compared with the other methods. Besides that, both have better performance along with the further metrics, meaning that the method is also more concise in results repeatability when exposed to 10-fold cross-validation.

TABLE I. Algorithms' performance metrics.

Classifier	Accuracy	Recall	Precision	F1-Score
CNN	0.859 ± 0.020	0.708 ± 0.012	0.705 ± 0.025	0.692 ± 0.016
SVM	0.789 ± 0.013	0.695 ± 0.017	0.695 ± 0.010	0.695 ± 0.024
DT	0.805 ± 0.079	0.613 ± 0.089	0.695 ± 0.075	0.651 ± 0.082
CNN + SVM	0.922 ± 0.053	0.813 ± 0.056	0.813 ± 0.049	0.813 ± 0.050
CNN + DT	0.824 ± 0.026	0.779 ± 0.021	0.808 ± 0.030	0.793 ± 0.027
SVM + DT	0.843 ± 0.037	0.728 ± 0.038	0.643 ± 0.041	0.683 ± 0.027
CNN + SVM + DT	**0.946 ± 0.017**	**0.893 ± 0.022**	**0.866 ± 0015**	**0.879 ± 0.010**
Random Forest	0.872 ± 0081	0.779 ± 0.088	0.781 ± 0.079	0.780 ± 0.074

Other methods with remarkable performance were Random Forest, SVM + DT, and DT. Random Forest and ensemble of SVM + DT have high metrics average. However, the standard deviation was one of the largest between the methods. Thus, these methods can be considered inferior to those previously discussed, even with a similar average. The least performing method was SVM, nevertheless, its standard deviation was the smallest of all methods, showing more concise and uniform results.

Table II shows the accuracy comparison between the results of this study and several data found in the literature. Some of these articles also used other databases, in addition to DEAP.

TABLE II. Literature results. Comparison to existing models

Classifier	Authors	Literature (%)	Present study (%)	Difference (%)
CNN	Garg and Verma [24]	**92.2**		-6.3
	Martin et al. [52]	78.9	85.9	0
	Shu et al. [20]	85.8		0.1
SVM	Masruroh et al. [22]	55.8	78.9	23.1

Classifier	Authors	Literature (%)	Present study (%)	Difference (%)
	Ozel et al. [11]	75.5		3.4
	Shu et al. [20]	**91.5**		-12.6
DT	Ozel et al. [11]	75.3	**80.5**	5.2
CNN + SVM	Shu et al. [20]	85.1	**92.2**	7.1
	Topic and Russo [25]	76.5		15.7
		77.7		14.5
SVM + DT	Panicker and Gayathri [21]	**84.6**	84.3	-0.3
Random Forest	Ozel et al. [11]	76.3	**87.2**	10.9
CNN + DT	-	-	**82.4**	-
CNN + SVM + DT	-	-	**94.6**	-

The values presented for the ensemble classifiers of CNN + DT, CNN + SVM + DT were close to those found in the literature using other models. Furthermore, the second method showed higher accuracy than the other models found. Since that, there are no records of such methods for the recognition of the emotional state from EEG signals, and with results above the values described in the literature, the use of such learning techniques is characterized as the main contribution of this research.

The learning curves of the ensemble classifiers proposed in this study are shown in Fig. 8, where the values are obtained behind 12 epochs and exhibit the training score in terms of the accuracy metric.

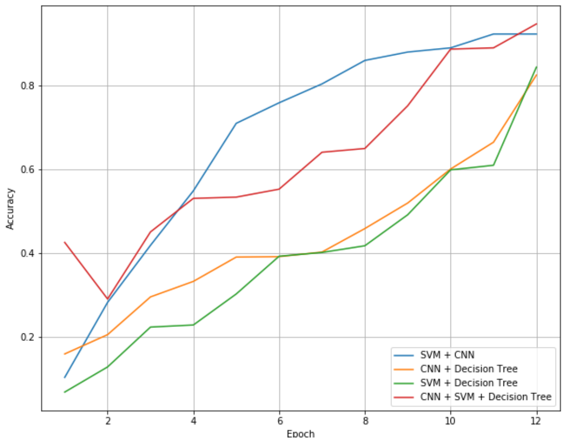

Fig. 7. Convergence curves of training score from ensemble models

VI. Conclusion and Future Research

In this work, eight machine learning classifier models (CNN, SVM, DT, and their ensembles) were tested to identify and classify emotions from EEG signals. These techniques used the public database DEAP as an evaluation set, which contains preprocessed EEG data, as well as the references for classification.

The experimental results obtained in the performance metrics among the proposed techniques were analyzed according to their means and standard deviation. The best results were obtained with a combination of CNN and SVM models and a stacking ensemble learning method combining SVM, CNN, and decision tree, which achieved slightly greater accuracy than that found in the current literature.

The decision tree and random forest models are among the most accurately tested models in this research. The SVM model was the one with the lowest performance, however, it is close to the one already found in scientific articles, and it was the method with the lowest standard deviation.

Future works could achieve better percentages of accuracy, testing other machine learning techniques such as k-nearest neighbors (k-NN) and boosting ensemble methods, which were not in focus here, as well as testing other databases, besides DEAP.

Acknowledgment

The authors would like to thank the National Council of Scientific and Technologic Development of Brazil - CNPq (Grants number: 307958/2019-1-PQ and 404659/2016-0-Univ), PRONEX `Fundação Araucária' 042/2018

References

[1] K. S.Bhagat, P. M. Mahajan, and G. P. Waghulade, “Emotion detection using EEG signal analysis,” International Journal of Electronics Communication and Computer Technology (IJECCT), vol. 5, n. 2, pp. 834–839, 2015.

[2] M. Cabanac, “What is emotion?,” Behavioural Processes, vol. 60, n. 2, pp. 69–83, 2002.

[3] Iris B. Mauss and M. D. Robinson, “Measures of emotion: A review,” Cognition and Emotion, vol. 23, n. 2, pp. 209-237, 2009.

[4] A. Kumar and W. N. Dattatray, “Emotion detection using EEG signal analysis,” International Journal & Magazine of Engineering, Technology, Management and Research, vol. 3, n. 9, pp. 103-111, 2016.

[5] A. T. Sohaib, and S. Qureshi, "An empirical study of Machine Learning techniques for classifying emotional states from EEG data". Master Thesis — School of Engineering Blekinge Institute of Technology, Karlskrona, Sweden, 2012.

[6] R. T. Sousa, “Avaliação de classificadores na classificação de radiografias de tórax para o diagnóstico de pneumonia infantil,” M. S. Thesis, Computer Science, Federal University of Goiás, Goiânia, GO, Brazil, 2013 (in Portuguese). Accessed on: Sep 10, 2019. [Online]. Available: https://repositorio.bc.ufg.br/tede/handle/tede/3356

[7] S. Shalev-Shwartz and S. Ben-David, “Understanding machine learning: From theory to algorithms,” Cambridge University Press, Cambridge, United Kingdom, 2014.

[8] S. Chourasia, “Survey paper on improved methods of ID3 decision tree classification,” International Journal of Scientific and Research Publications, vol. 3, n. 12, pp. 1-4, 2013.

[9] T. Evgeniou, M. Pontil, and T. Poggio, “Regularization networks and support vector machines,” Advances in Computational Mathematics, vol. 13, n. 1, pp. 1–50, 2011.

[10] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, n. 7553, pp. 436-444, 2015.

[11] S. Koelstra *et al*., “DEAP: A Database for emotion analysis using physiological signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, 2012.

[12] P. Ozel, A. Akan, and B. Yilmaz, “Biomedical Signal Processing and Control Synchrosqueezing transform based feature extraction from EEG signals for emotional state prediction,” Biomedic Signal Process Control, vol. 52, pp. 152-161, 2019.

[13] S. J. Luck, “An introduction to the event-related potential technique,” 2nd ed., Cambridge: The MIT Press, Cambridge, Massachusetts, EUA, 2014.

[14] E. Niedermeyer and F. H. L. Silva, “Electroencephalography: basic principles, clinical applications, and related fields,” Lippincott Williams & Wilkins, 2005.

[15] M. Menezes, "Neuroarte: interface ciberartística baseada em tecnologia eletroencefalográfica," M. S. Thesis, University of Brasília, Brasília, DF, Brazil, 2014 (in Portuguese). Accessed on: Oct 12, 2019. [Online]. Available: http://repositorio.unb.br/handle/10482/17281

[16] H. Xu, "Towards automated recognition of human emotions using EEG," M. S. Thesis, Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, 2012. Accessed on: Sep 18, 2019. [Online]. Available: https://pdfs.semanticscholar.org/8b53/94b071cedabd9e097370ee741ccfac70db67.pdf

[17] P. Ekman et al., "Universals and cultural differences in the judgments of facial expressions of emotion," Journal of Personality and Social Psychology, vol. 53, n. 4, pp. 712, 1987.

[18] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, n. 6, pp. 1161-1178, 1980.

[19] I. C. Christie and B. H. Friedman, "Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach," International Journal of Psychophysiology, vol. 51, n. 2, pp. 143–153, 2004.

[20] D. Purves et al., White, "Neuroscience-4," Sunderland, Mass.: Sinauer. xvii, vol. 857, pp. 944, 2007

[21] L. Shu et al., "A review of emotion recognition using physiological signals," Sensors, vol. 18, n. 7, pp. 2074, 2018.

[22] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology-based mental stress detection systems," Biocybernetics and Biomedical Engineering, vol. 39, n. 2, pp. 444-469, 2019.

[23] A. H. Masruroh, E. M. Imah, and E. Rahmawati, "Classification of emotional state based on EEG signal using AMGLVQ," Procedia Computer Science, vol. 157, pp. 552-559, 2019.

[24] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," Expert Systems with Applications, vol. 93, pp. 143-155, 2018.

[25] D. Garg, and G. K. Verma. "Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet-based deep learning framework." Procedia Computer Science, vol. 171, pp. 857-867, 2020.

[26] A. Topic, and M. Russo. "Emotion recognition based on EEG feature maps through deep learning network." Engineering Science and Technology, an International Journal, In Press, 2021.

[27] A. Mollahosseini, B. Hasani, and Mahoor, M. H. "Affectnet: A database for facial expression, valence, and arousal computing in the wild". IEEE Transactions on Affective Computing, vol. 10, n. 1, pp. 18-31, 2017.

[28] Google. Google facial expression comparison dataset. Available at: https://research.google/tools/datasets/google-facial-expression/. Accessed on August 14, 2021.

[29] R. Kosti, J.M. Álvarez, A. Recasens, and A. Lapedriza, "Emotion Recognition in Context", Computer Vision and Pattern Recognition (CVPR), 2017.

[30] W.-L. Zheng, and B.-L. Lu. "A multimodal approach to estimating vigilance using EEG and forehead EOG." Journal of neural engineering vol. 14, n. 2, article number: 026017, 2017.

[31] S. Katsigiannis, N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 98-107, 2018.

[32] V. M. Joshi, and R. B. Ghongade. "IDEA: Intellect database for emotion analysis using EEG signal." Journal of King Saud University-Computer and Information Sciences. In Press, 2020.

[33] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and personality recognition using commercial sensors". IEEE Transactions on Affective Computing, vol. 9, n. 2, pp. 147-160, 2016.

[34] A. X. Stewart, A. Nuthmann, and G. Sanguinetti, "Single-trial classification of EEG in a visual object task using ICA and machine learning," Journal of Neuroscience Methods, vol. 228, pp. 1-14, 2014.

[35] A. R. C. Medeiros, Y. B. de Araujo, R. P. de T. Vianna, and R. M. de Moraes, "Modelo de suporte à decisão aplicado à identificação de indivíduos não aderentes ao tratamento anti-hipertensivo," Saúde em Debate, Centro Brasileiro de Estudos de Saúde, vol. 38, n. 100, pp. 104–118, 2014 (in Portuguese).

[36] C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, vol. 2, n. 2, pp. 121-167, 1998.

[37] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 1055-1064, 1999.

[38] N. Cristianini and B. Scholkopf, "Support vector machines and kernel methods: the new generation of learning machine," AI Magazine, v. 23, n. 3, p. 31-31, 2002.

[39] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," The Handbook of Brain Theory and Neural Networks, vol. 3361, n. 10, article number 1995, 1995.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25 pp. 1097-1105, 2012.

[41] T. G. Dietterich, "Ensemble methods in machine learning," International workshop on multiple classifier systems. Springer, Heidelberg, Germany, pp. 1-15, 2000.

[42] Z. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," Artificial Intelligence, vol. 137, n. 1-2, pp. 239-263, 2002.

[43] L. Zhang and W. Zhou, "Sparse ensembles using weighted combination methods based on linear programming," Pattern Recognition, vol. 44, n. 1, pp. 97-106, 2011.

[44] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, n. 2, pp. 281-286, 2002.

[45] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, n. 2, pp. 123–140, 1996.

[46] L. Breiman, "Random forests". Machine learning, vol. 45, n. 1, p. 5-32, 2001.

[47] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992.

[48] Z. H. Zhou, "Combination methods", in Ensemble methods: Foundations and algorithms, Chapman and Hall/CRC, 2012, ch. 4, pp. 83-86.

[49] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall, and F-score, with implication for evaluation," European Conference on Information Retrieval. Springer, Heidelberg, Germany, 2005, pp. 345-359.

[50] L. Rokach, "Ensemble Classification", In Pattern Classification Using Ensemble Methods, World Scientific Publishing Co., 2010, Chapter 3, pp. 82-85.

[51] M. H. D. M. Ribeiro and L. dos S. Coelho, "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series". Applied Soft Computing, v. 86, article number. 105837, 2020.

[52] N. Martin, J. M. Diverrez, S. Em, N. Pallamin, and M. Ragot, "How to recognize emotions without signal processing: an application of convolutional neural network to physiological signals," Neuroergonomics, pp. 191-194, 2019.