Improving Emotion Recognition Systems by Exploiting the Spatial Information of EEG Sensors

Guido Gagliardi, Antonio Luca Alfeo, Vincenzo Catrambone, Diego Candia-Rivera, Mario G. C. A. Cimino, Gaetano Valenza

This is a preprint. Please cite using:

```
@article{improving2023,
    author={Gagliardi, Guido and Alfeo, Antonio Luca and Catrambone, Vincenzo and Candia-Rivera, Diego and Cimino, Mario G. C. A. an
    title={Improving Emotion Recognition Systems by Exploiting the Spatial Information of EEG Sensors},
    journal={IEEE Access},
    year={2023},
    volume={11},
    pages={39544-39554},
    publisher={Institute of Electrical and Electronics Engineers (IEEE)},
    doi={10.1109/access.2023.3268233},
    issn={2169-3536},
}
```

Guido Gagliardi, Antonio Luca Alfeo, Vincenzo Catrambone, Diego Candia-Rivera, Mario G. C. A. Cimino, Gaetano Valenza. "Improving Emotion Recognition Systems by Exploiting the Spatial Information of EEG Sensors" IEEE Access 11 (2023): 39544-39554.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 March 2023, accepted 13 April 2023, date of publication 19 April 2023, date of current version 26 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3268233

RESEARCH ARTICLE

Improving Emotion Recognition Systems by Exploiting the Spatial Information of EEG Sensors

GUIDO GAGLIARDI[1,2], (Graduate Student Member, IEEE), ANTONIO LUCA ALFEO[1,3], VINCENZO CATRAMBONE[1,3], DIEGO CANDIA-RIVERA[1,3], MARIO G. C. A. CIMINO[1,3], (Member, IEEE), AND GAETANO VALENZA[1,3]

[1]Department of Information Engineering, University of Pisa, 56126 Pisa, Italy
[2]Department of Electrical Engineering, KU Leuven, 3000 Leuven, Belgium
[3]Bioengineering and Robotics Research Center E. Piaggio, School of Engineering, University of Pisa, 56126 Pisa, Italy

Corresponding author: Guido Gagliardi (guido.gagliardi@phd.unipi.it)

This work was supported in part by the Italian Ministry of Education and Research (MIUR) in the framework of the FoReLab Project (Departments of Excellence) research partially funded by the Piano Nazionale di Ripresa e Resilienza (PNRR)–M4C2; in part by Partenariato Esteso PE00000013—"FAIR—Future Artificial Intelligence Research"—Spoke 1 "Human-Centered AI," funded by the European Commission through NextGeneration EU Programme; in part by the European Commission H2020 Framework Program of the Project "EXPERIENCE" under Grant 101017727; and in part by the European Research Council through the European Community's 7th Framework Programme (FP7/2007-2013)/European Research Council (ERC) Starting Grant (203143).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of the University of Pisa.

ABSTRACT Electroencephalography (EEG)-based emotion recognition is gaining increasing importance due to its potential applications in various scientific fields, ranging from psychophysiology to neuromarketing. A number of approaches have been proposed that use machine learning (ML) technology to achieve high recognition performance, which relies on engineering features from brain activity dynamics. Since ML performance can be improved by utilizing 2D feature representation that exploits the spatial relationships among the features, here we propose a novel input representation that involves re-arranging EEG features as an image that reflects the top view of the subject's scalp. This approach enables emotion recognition through image-based ML methods such as pre-trained deep neural networks or "trained-from-scratch" convolutional neural networks. We have employed both of these techniques in our study to demonstrate the effectiveness of our proposed input representation. We also compare the recognition performance of these methods against state-of-the-art tabular data analysis approaches, which do not utilize the spatial relationships between the sensors. We test our proposed approach using two publicly available benchmark datasets for EEG-based emotion recognition tasks, namely DEAP and MAHNOB-HCI. Our results show that the "trained-from-scratch" convolutional neural network outperforms the best approaches in the literature, achieving 97.8% and 98.3% accuracy in valence and arousal classification on MAHNOB-HCI, and 91% and 90.4% on DEAP, respectively.

INDEX TERMS Convolutional neural networks, electroencephalography, emotion recognition, spatial information representation.

I. INTRODUCTION

Affective computing is a broad research field that investigates emotional and mental states through the analysis of physiological signals or other sources of information, such as videos, images, or sounds.

In this field, emotion recognition is becoming increasingly important due to the many applications in which it is involved. The exploitation of physiological data for emotion recognition may be motivated by several factors, such as their

The associate editor coordinating the review of this manuscript and approving it for publication was Ludovico Minati.

psycho-physiological correlates, or the ability of intelligent systems to analyze them and potentially identify patterns associated with affective disorders (e.g., anxiety and depression). Additionally, there has been an increasing number of easy-to-use, non-invasive, portable devices capable of gathering robust physiological data [1], [2]. Emotions can be identified as discrete regions in a multidimensional space, whose main dimensions according to the circumplex model of affect [3], [4] are valence (positive to negative feelings) and arousal (sleepy to excited), or as a series of discrete basic emotions such as the Ekman model [5], that identifies six basic emotions, i.e. anger, disgust, fear, happiness, sadness and surprise, or the Plutchik's model [6], that proposed a wheel of eight emotions: joy, trust, fear, surprise, sadness, disgust, anger and anticipation; all these categorical emotions can combined to define more detailed perception.

In the context of emotion recognition tasks based on non-invasive physiological data, electroencephalography (EEG) is one of the most commonly used signals [7] due to its good compromise between temporal and spatial resolution [8]. It is also widely used thanks to the number of non-invasive, low-cost, and easy-to-operate wearable devices on the market [9]. EEG is usually sampled by placing a group of sensors or channels on the patient's scalp, arranged according to a standard scheme, such as the international standard pattern 10-20 [10]. The EEG signals are usually analyzed in the frequency and time domains, and EEG channels are mostly handled as independent time series, meaning that the spatial relationship among EEG sensor dynamics is mostly neglected [11].

It has been reported that the majority of studies (89.4%) performing EEG-driven emotion recognition extract features from the frequency domain and employ methods such as Short-time Fourier Transform or Discrete Fourier Transform (25.4%), Power Spectral Density (PSD) (22.2%), and Wavelet Transform (19.1%) [12]. Once computed, these data are generally converted into tabular shape, where a number of location-specific features are extracted from each classification instance. Eventually, the tabular features are processed by some machine learning (ML) model capable of recognizing the corresponding class label associated with emotional correlates. Most of the emotion recognition tasks presented in the literature rely on ML algorithms such as Support Vector Machines (SVM) (59%), K-Nearest Neighbors (KNN) (14%), Multilayer Perceptron Networks (MLP) (6.63%), linear discriminant analysis (LDA) (6.3%), and quadratic discriminant analysis (QDA) (3.2%) [12]. The operations of model training and testing are performed either separately, using data from a single subject (i.e., a subject-dependent framework), or using data from multiple subjects (i.e., a subject-independent framework) [12].

Recent studies have shown that the spatial information integrated into the EEG arrangements can be used to improve EEG classification performances [13]. To include the spatial domain information, the features obtained by processing each EEG channel can be considered as spatial points displayed in a tri- or bi- dimensional space, considering the sensors' placement [12], and then processed through spatial filters [13] or image processing tools [14].

As a consequence, arranging EEG features as an image would allow employing specific image-based ML approaches like Convolutional Neural Networks (CNNs) [14]. CNN is considered one of the most widely used ML techniques, especially in image-related applications; CNNs can learn new representations from images and have shown substantial performance improvement in various ML applications [15]. Usually, several neurons' layers of different natures are intermixed in a CNN architecture, with the last one performing the actual classification task.

The arrangement of CNN layers plays a fundamental role in the designing and training of new architectures, thus allowing increasing algorithmic performance. With the correct architecture, CNNs have demonstrated the capability to handle and generalize big data exploiting high-computational resources. Indeed, there are several publicly available state-of-the-art CNN-based architectures trained with huge datasets [16], such as the well-known AlexNet [17], DenseNet [18], or MobileNetV2 [19]. These pre-trained architectures can be easily imported into a new system, quickly fine-tuned if necessary, and then used in a new task with the benefit of previously learned knowledge. This is very useful for all the applications, like the ones based on physiological data, in which it is difficult to get a sufficient amount of data to correctly train the CNN [15]. The effectiveness of these models in an affective computing scenario for EEG-based emotion classification is also suggested by the fact that some of the most accurate architectures proposed in recent years used an image rearrangement of the EEG to allow the usage of CNN, without considering the spatial relations of the electrodes [20], [21].

To summarize, EEG-based emotion recognition is an extremely important task, which has mainly been tackled using standard signal processing techniques in the time and frequency domains; spatial information has been overlooked so far, despite its known importance in ML-based applications.

To this extent, the main contribution of this study is a novel approach to exploit the spatial relationship among EEG sensors through image-based machine learning algorithms. Specifically, the proposed method involves a rearrangement of EEG features into images, which are subsequently fed to pre-trained neural networks and a novel CNN to extract features and classify emotions. To evaluate the proposed approach, we compared the performance of several pre-trained neural networks on image classification tasks in emotion recognition with different state-of-the-art algorithms for the classification of EEG tabular features. Furthermore, a novel ad-hoc CNN is here proposed, trained from scratch to process the new input. The experimental results show that the proposed approach outperforms the state-of-the-art algorithms in subject-dependent valence- and arousal-emotion classification.

The paper is organized as follows: In section II, we detail the experimental setup, dataset description, feature extraction, and the rearrangement of EEG features into images. Additionally, we provide an explanation of the machine learning architectures tested, including pre-trained neural networks and the proposed CNN. In section III, we report on the experimental results and the comparison between the multiple machine learning models employed. In section IV, we discuss the achieved results in light of the associated literature, and in section V we illustrate the proposed approach's main strengths and limitations, concluding with possible future developments.

II. MATERIALS AND METHODS

This section provides a detailed description of the experimental pipeline employed in this study, which is graphically depicted in Figure 1 using a block diagram that describes the overall architecture. The experimental pipeline involves four blocks, with the first two blocks, characterized by black dashed lines, involving the sampling and windowing of EEG signals from 32 sensors into 8-second epochs. From each epoch, EEG bands and features are extracted and represented in their tabular form. The last two blocks, characterized by red dashed lines, represent the main contribution of our work, which involves the rearrangement of EEG tabular features into a new image-like format that exploits spatial information. This is achieved by mapping the tabular features to a two-dimensional grid that resembles an image. This new representation of EEG signals is then used as input for an image-based machine learning algorithm for emotion classification.

A. EXPERIMENTAL DATASET

For this study, two publicly available datasets were used, both of which involved physiological signal collection from healthy volunteers undergoing emotional video elicitation. Emotion perception was evaluated through the well-known circumplex model of affect [3], consisting in a bi-dimensional space: arousal associated with the strength of the feeling, and valence associated with the pleasantness of the feeling. Both variables were quantified through a 0-9 Likert-type scale. As reported in the papers in which the datasets were presented, prior to the experiment, each participant signed a consent form.

In this study, trials of both datasets and for both variables (i.e., arousal and valence) were split into two classes: high and low. Arousal and valence labels were assigned separately for all the videos analyzed according to what was expressed by each subject individually. Consequently, regarding arousal, trials were separated into high-arousal (HA, arousal $>= 5.5$) evoking a strong emotional response, and low-arousal (LA, arousal $<= 4.5$), evoking a weak elicitation. Regarding valence, trials were separated into high-valence (HV, valence $>= 5.5$) evoking a pleasant response, and low-valence (LV, valence $<= 4.5$), evoking unpleasant elicitation. Numerical thresholds were selected to exclude elements related to neutral responses (i.e., with arousal or valence in the interval $[4.5, 5.5]$) from the experimental set, thus preventing the deep learning model from encountering ambiguity in the boundary between low and high classes, and enhancing class separability.

1) THE DEAP DATASET

The dataset consisted of 32 healthy participants (age range, 19–27 yo; 16 females) [22]. A number of physiological signals were gathered, and, in this study, 32-channel EEG sampled at $512Hz$ was considered. It is available at *https://www.eecs.qmul.ac.uk/mmv/datasets/deap/*.

The experimental protocol consisted of 40 emotional video trials from famous music videos. After an initial $2min$ resting state, $60sec$ emotional videos, with different levels of arousal and valence, were presented. Extensive details can be found at [22].

2) THE MAHNOB–HCI DATASET

The dataset consisted of 27 healthy participants (age range, 19–40 years; 15 females) [23]. Different trials might involve a different number of volunteers, ranging from 25 to 27, because of missing or bad-quality signals. A number of physiological signals were gathered, and, in this study, 32-channel EEG sampled at $256Hz$ was considered. It is available at *https://mahnob-db.eu/hci-tagging/*.

The experimental protocol consisted of 20 emotional video trials from famous movies. After an initial $30sec$ resting state, emotional videos of varying lengths (between 35 and $177sec$), with different levels of arousal and valence, were presented. Extensive details can be found at [23].

B. EEG PROCESSING AND FEATURE EXTRACTION

The EEG processing procedure was implemented to obtain artefact-free signals to compute the EEG spectrogram to be used in the classification task. The processing procedure comprised frequency filtering, large artefacts rejection, eye movements and cardiac-field artefact removal, interpolation of contaminated channels, and average re-referencing [24]. These steps were implemented in MATLAB R2018b (MathWorks) using the Fieldtrip Toolbox [25]. An extensive description of the preprocessing procedure applied can be found in [26].

The EEG power spectral density (PSD) was extracted through Welch's method with a Hanning window. A sliding time window $2sec$ long and with 50% of overlap was employed, and PSD time series were integrated within four frequency bands, namely: $theta : \theta \in (4 - 8]Hz$,$alpha : \alpha \in (8 - 12]Hz$, $beta : \beta \in (12 - 30]Hz$, and $gamma : \gamma \in [30 - 45]Hz$.

For each time segment corresponding to a classification instance (i.e., $8sec$ length, with $2sec$ of overlap), a set of EEG channels (i.e., 32) and frequency bands (i.e., 4) were considered and five features were derived: total PSD; the three Hjort parameters (i.e., activity, mobility, and

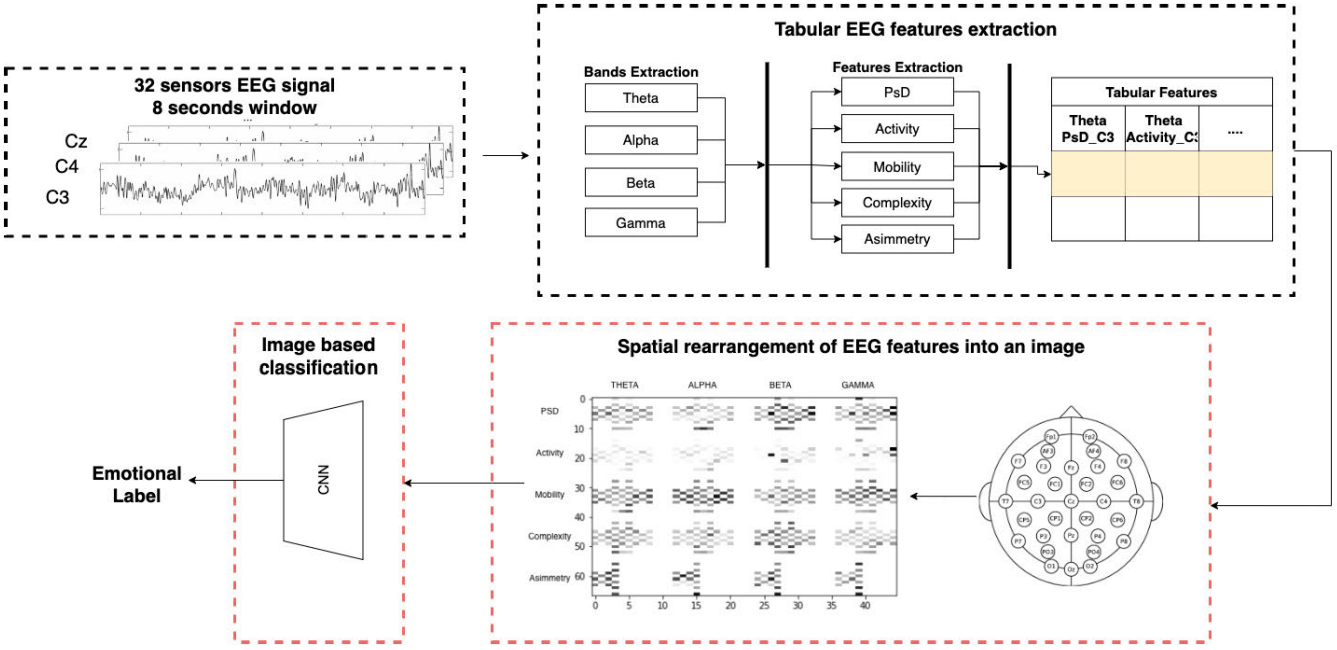


FIGURE 1. Block diagram of the experimental pipeline. In the first block, the acquired 32-channel EEG time series are divided into 8-second windows; in the second, EEG features are extracted from each frequency band, then rearranged into images according to the proposed scheme, and then classified using an image-based classification approach, e.g. CNN.

complexity [27]); and asymmetry, as the difference between PSD in channels symmetric with respect to the vertical cerebral axis. Summarizing, each instance was represented by 640 features (32 channels × 4 frequency bands × 5 features).

C. EEG FEATURES SPATIAL ARRANGEMENT

The following EEG-features spatial arrangement scheme has been designed to allow image-based ML algorithms to exploit the spatial proximity among electrodes. The main idea is to build an *image-block* arranging the EEG features obtained via the electrodes considering their spatial proximity and then aggregating the image associated with each feature following their proximity in the frequency domain.

To incorporate spatial information, a single image-block was constructed using a single feature and frequency band (e.g., the α band of the PSD), and the corresponding 1×32 channel vector was transformed into a $2D$-image of size 11×9 (Fig. 2.b) that depicts a top view of the subject's scalp, where each element corresponds to a specific sensor position on the scalp. The matrix elements (Fig. 2.b) represent a $2D$-map of the 10-20 EEG sensor international scheme. Non-0 elements correspond to positions occupied by sensors, while 0 elements represent empty positions.

This spatial rearrangement provides clear information to the classification system about the spatial arrangement of the input features, enabling the exploitation of electrical patterns localized in different brain regions. The neuroscience literature [28] suggests that such information should enhance emotion recognition performance, since human emotions are closely linked to specific brain regions.

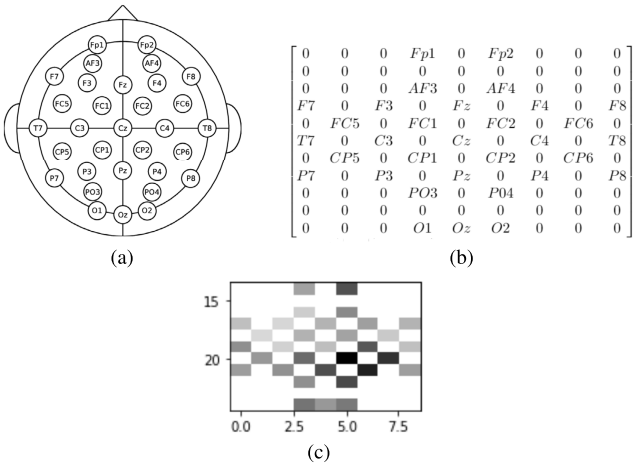


FIGURE 2. Spatial localization of the 32-channel EEG in the 10-20 standard schema (a); Spatial rearrangement in an 11 × 9 matrix (b); interpreted as a grayscale image (i.e., an image-block) (c).

Figure 2 provides an example of an image-block. The resulting 11×9 matrix can be easily transformed into a grayscale image, as shown in Fig. 2.c. Assuming that all EEG values are greater than or equal to zero, only the pixels corresponding to electrode positions are non-white in the image. The darker the pixel, the higher the feature value.

At this stage, we obtain 20 $2D$-matrices of size 11×9 for each instance. To obtain the final representation of the EEG sample, we reshape the $20 \times 11 \times 9$ matrices into a new $2D$-matrix by placing the 11×9 matrices side by side in the same column for those extracted from the same frequency band and

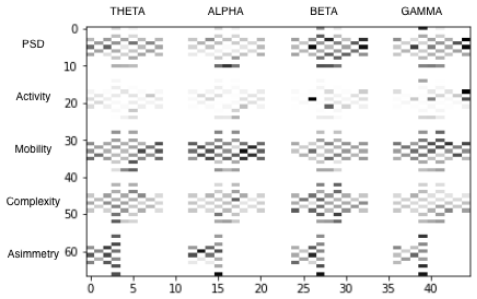


FIGURE 3. Exemplary image obtained exploiting spatial and frequency proximity.

in the same row for those belonging to the same feature type (Fig. 3). Thus, in each image-block, features are arranged in rows (i.e., PSD, Activity, Mobility, Complexity, Asymmetry), and frequency bands are arranged in columns (i.e., θ, α, β, and γ).

To realistically represent the spatial proximity between electrodes, EEG features obtained from a single sensor placed on one side of the scalp (e.g., right) and a particular frequency band (e.g., β) should not be processed together with features from the opposite side of the scalp (e.g., left) or those obtained from a different frequency band (e.g., γ). To achieve this, the convolutional operations should not process different image-blocks at the same time, not even partially. To achieve this separation, each image-block is separated both horizontally and vertically by a number of white pixels equal to the width of the filter used in convolutional operations. If we consider a filter width equal to 3, the size of the image obtained by rearranging all EEG features would be 67×45 pixels.

D. MACHINE LEARNING MODELS

Firstly, we detail the state-of-the-art image-based approaches implemented in our experiments. Then, we include nine classical ML approaches for tabular data, i.e., with no spatial information embedded in the inputs. These approaches have been named tabular-features-based. All of the models were developed to perform two distinct binary classifications: the first distinguishing between high-arousal (HA) and low-arousal (LA) trials, and the second disentangling between high-valence (HV) and low-valence (LV) trials.

1) IMAGE-BASED ALGORITHMS

Seven different approaches were implemented to evaluate the image-based classification performance using the procedure described in section II-C. Initially, we tested several pre-trained architectures on the ImageNet [16] dataset, including MobileNetV2 [19], DenseNet121 [18], ResNet152V2 [29], ResNet50V2 [29], VGG16 [30], and VGG19 [30]. Furthermore, we developed a Convolutional Neural Network (CNN) [31] architecture from scratch to solve the image-classification task.

To fine-tune the pre-trained neural networks for the emotion classification datasets, we removed the top classification layers and added multiple fully-connected layers with the *relu* activation function. During training, these newly added layers were set as trainable, while the remainder of the architecture was frozen. Each pre-trained neural network had a similar number of trainable parameters in the added layers. The last fully connected layer of each architecture comprised two neurons with the *softmax* activation function to address the binary classification problem. The activation of these neurons is mutually exclusive, meaning that the only possible outcomes are either 0, 1 or 1, 0.

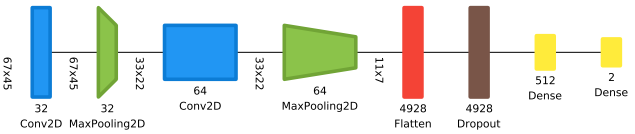


FIGURE 4. Visual representation of the CNN model trained from scratch. Architecture plot provided by Net2Vis [33].

The other image-based classification approach implemented is based on a CNN. In particular, CNN has been widely and successively adopted in several image classification tasks, even in clinical scenarios [32]. A classic approach to solve an image classification problem is to train a CNN from scratch, meaning first defining convolutional layers, able to extract features from input images, and then adding dense fully connected layers to perform the final classification step; all neurons are usually randomly initialized. Unlike the pre-trained architectures, CNN was trained as a whole with the arousal/valence classification input.

The implemented CNN architecture (Fig. 4), consists of two convolutional layers, with depths of 32 and 64, respectively, followed by two Max-Pooling layers. As mentioned above, the size of the convolutional filters was set to 3×3. Subsequently, a flattening layer was inserted to prepare the data to be classified by the final two dense fully connected layers, of 512 and 2 neurons, respectively. The last level has 2 neurons due to the number of classes to classify, i.e. the last layer is composed of a 2 neurons dense layer with *SoftMax* activation function, which performs a binary classification task of the label arousal or valence, so the final output of the network is a tuple $\{1, 0\}$ in case of predicting low arousal or valence, or $\{0, 1\}$ in case of predicting high arousal or valence. To prevent overfitting during training, a dropout layer has been added after the flattening layer.

2) TABULAR-FEATURES-BASED ALGORITHMS

The Support Vector Machine (SVM) is the most commonly used feature-based ML model in affective computing and emotion recognition tasks. Other widely used approaches include, but are not limited to, K-Nearest Neighbours (KNN), Random Forest (RF), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and MultiLayer Perceptron network (MLP) [34]. In this study, all of the aforementioned algorithms were implemented. Additionally, other ensemble and boosting-based algorithms, such as Extremely Randomized Trees (ET), AdaBoost, and GradientBoosting,

TABLE 1. The number of trainable and total parameters of the neural network models used for the experiments.

	Trainable	Total
MLP	2.438.138	2.438.138
DenseNet	2.624.002	9.661.506
MobileNetV2	2.689.052	4.947.036
ResNet50	2.458.208	26.023.008
ResNet152	2.458.208	58.331.648
VGG16	2.423.176	22.447.560
VGG19	2.423.176	17.137.864
CNN	2.543.490	2.543.490

were considered to ensure completeness of the results, as well as their generally good performance on tabular data.

The SVM algorithm transforms the original feature space into a higher-dimensional space using a kernel function. Then, it identifies support vectors to maximize the separation (margin) between the classes [35]. The algorithm uses these support vectors to construct hyperplanes that separate the two classes in a high-dimensional space [35]. ET and RF are tree-based ensemble methods that use a recursive feature selection procedure through decision trees until a minimum subset of data corresponding to a class is identified. The main difference between them is the selection of cut points to split nodes: RF performs an optimization procedure to split the input, while ET does it randomly to achieve convergence in a shorter period of time [36]. KNN is a lazy learner algorithm that stores the entire training input and then performs a classification strategy, assigning to each sample in the test set the majority class of its K nearest neighbour samples in the training set. AdaBoost and GradientBoosting are two boosting-based classification methods that rely on decision tree ensembling. The AdaBoost algorithm attempts to minimize the loss function related to the classification error and was designed for binary classification problems. Gradient Boosting, on the other hand, is used to optimize differentiable loss functions and can be used for both classification and regression. MLP, implemented as feed-forward neural networks, is characterized by fast operation, ease of implementation, and smaller training set requirements [37]. MLP performs a mapping between classes and input data through a generally non-linear function whose parameters (or neurons) are set during training. This makes MLP a very effective and adaptable approach to various classification problems. The MLP employed in this study was a 7-layer neural network. The number of neurons per layer was chosen to feature comparable computational complexity (i.e., the number of trainable parameters) with the other pre-trained image-based architectures. Table 1 lists the number of trainable parameters for all of the NN-based methods implemented in this study.

3) IMPLEMENTATION DETAILS

The classification was performed in a subject-dependent framework, i.e., in each experiment, the samples belonging to a single subject were considered. A 10-fold cross-validation (CV) was applied to each classification task, providing the value of the CV average accuracy as the reference accuracy for each subject. Finally, the results from all the subjects are aggregated, presenting the average and standard deviation among all of the reference accuracy.

All NN-based models employ the same hyper-parameters: *categorical cross-entropy* loss function, consistent with the classes encoding in one hot encoding; *Adam* optimizer; the batch size equal to 16; and early stopping to managing the number of training epochs, managed with patience set to 8. All the features-based algorithms have been tested with different parameters setup using nested cross-validation in order to find the optimal parameters.

III. EXPERIMENTAL RESULTS

A. IMAGE-BASED VS TABULAR-FEATURE-BASED COMPARISON

Firstly, the accuracy of the image-based ML approaches is compared with the tabular-features-based ones; then the best performing model is chosen and its main hyperparameters' space is explored to compare the obtained classification performance with the state-of-the-art on the same dataset. Table 2 summarizes the classification performance of the implemented models in terms of average accuracy, f1 score, precision, recall, and area under the receiver operating characteristic curve (AUROC) in the HA-LA, and HV-LV binary classification tasks.

Based on the results obtained from the MAHNOB-HCI, it can be observed that the ensemble-based algorithms and the boosting-based algorithms demonstrate similar results. However, in comparison to other features-based approaches, they exhibit the best recognition performance in both the HA-LA and the HV-LV classification task. Notably, the RF classifier achieves an accuracy of 72.43% $\pm$ 13.24% in HA-LA and 73.03% $\pm$ 8.77% in HV-LV, while GradientBoosting attains an f1 score of 69.07% $\pm$ 11.91% in HA-LA and 65.55% $\pm$ 8.62% in HV-LV. In terms of AUROC score, Gradient Boosting shows superior performances in the arousal classification task, achieving 76.19% $\pm$ 11.61%, whereas RF has the highest performance in valence classification, achieving 78.40% $\pm$ 9.75%. Overall, it can be concluded that RF has better performance over tabular-features-based algorithms, but it suffers more from class imbalance in comparison to GradientBoosting.

Similar observations can be made for the DEAP dataset, where ET demonstrates better accuracy performances, achieving 64.17% $\pm$ 7.48% in HA-LA and 61.09% $\pm$ 8.84% in HV-LV, while GradientBoosting attains better f1 scores, achieving 55.88%$\pm$14.51% in HA-LA and 55.69%$\pm$11.81% in HV-LV.

The proposed image-based approaches show superior performance when compared to the tabular-features-based algorithms, with CNN achieving the highest results in both the arousal and valence classification tasks for both the MAHNOB-HCI and DEAP datasets. Specifically, on the

TABLE 2. Performance results for the tabular-features-based algorithms and image-based algorithms for the subject-dependent HA-LA, and HV-LV classification tasks on MAHNOB-HCI and DEAP dataset.

		MAHNOB									
		arousal					valence				
		accuracy	f1	precision	recall	AUROC	accuracy	f1	precision	recall	AUROC
features based	AdaBoost	70.41%	66.75%	69.82%	68.13%	72.54%	70.02%	62.43%	65.56%	64.60%	73.24%
	GradientBoosting	72.20%	**69.07%**	71.14%	71.16%	**76.19%**	72.67%	**65.55%**	70.72%	**67.43%**	77.97%
	ET	72.10%	65.46%	72.58%	65.43%	75.96%	72.85%	60.12%	**74.99%**	57.41%	78.29%
	RF	**72.43%**	67.53%	**72.71%**	**68.44%**	75.82%	**73.03%**	62.34%	72.54%	61.67%	**78.40%**
	KNN	54.89%	51.65%	57.53%	60.05%	64.11%	56.29%	54.93%	53.37%	68.21%	65.54%
	SVM	67.71%	62.58%	67.60%	63.14%	70.10%	66.98%	55.44%	64.77%	54.08%	69.61%
	LDA	64.40%	60.50%	61.80%	62.33%	64.69%	61.73%	55.66%	54.50%	59.72%	62.46%
	QDA	62.67%	36.01%	29.26%	47.62%	50.04%	60.47%	21.10%	16.40%	29.63%	50.00%
	MLP	66.98%	61.59%	65.42%	63.04%	69.67%	66.74%	56.61%	60.99%	58.88%	69.44%
image based	**CNN**	**96.77%**	**96.20%**	**96.33%**	**96.30%**	**99.26%**	**97.42%**	**97.21%**	**97.49%**	**97.05%**	**99.49%**
	DenseNet121	84.82%	78.75%	79.72%	79.45%	89.12%	82.51%	75.94%	75.79%	77.42%	87.22%
	MobileNetV2	84.92%	78.20%	78.32%	79.56%	87.49%	83.41%	77.77%	78.25%	79.09%	87.78%
	ResNet152V2	82.68%	75.40%	75.20%	77.31%	85.59%	82.77%	77.79%	78.12%	78.75%	87.93%
	ResNet50V2	83.34%	76.28%	76.52%	77.70%	85.89%	83.73%	77.39%	77.52%	78.31%	88.23%
	VGG16	75.72%	65.23%	64.50%	68.93%	78.48%	74.28%	64.06%	64.16%	66.62%	79.54%
	VGG19	73.23%	59.87%	59.90%	63.70%	74.30%	70.85%	60.72%	61.73%	63.47%	76.67%

		DEAP									
		arousal					valence				
		accuracy	f1	precision	recall	AUROC	accuracy	f1	precision	recall	AUROC
features based	AdaBoost	60.14%	54.13%	55.72%	54.19%	60.21%	58.53%	54.28%	55.16%	55.10%	59.09%
	GradientBoosting	62.49%	**55.88%**	58.37%	56.06%	63.17%	60.26%	**55.69%**	57.32%	56.24%	**61.89%**
	ET	**64.17%**	52.39%	57.53%	52.50%	63.33%	**61.09%**	53.10%	**58.58%**	51.61%	60.89%
	RF	63.97%	54.05%	**58.52%**	53.70%	**64.03%**	60.92%	54.35%	58.55%	53.27%	61.78%
	KNN	45.50%	44.11%	50.98%	**56.49%**	55.65%	47.92%	48.52%	49.78%	**59.81%**	54.44%
	SVM	61.32%	54.99%	56.39%	55.40%	60.18%	58.96%	54.14%	55.16%	54.45%	59.25%
	LDA	57.77%	53.51%	53.34%	55.14%	57.19%	56.39%	53.20%	53.28%	54.31%	56.97%
	QDA	61.62%	33.97%	28.30%	43.66%	50.14%	57.88%	32.26%	27.77%	40.60%	50.07%
	MLP	60.18%	52.70%	53.38%	54.72%	60.07%	58.25%	52.79%	53.17%	55.14%	58.36%
image based	**CNN**	**88.68%**	**89.53%**	**89.25%**	**90.11%**	**94.48%**	**88.03%**	**89.28%**	**89.31%**	**89.47%**	**94.45%**
	DenseNet121	76.82%	75.71%	74.13%	78.56%	77.28%	72.86%	74.84%	72.95%	78.34%	77.03%
	MobileNetV2	77.23%	77.25%	75.58%	79.84%	78.94%	72.93%	75.19%	73.18%	78.70%	77.84%
	ResNet152V2	75.81%	75.62%	73.83%	78.71%	76.66%	72.42%	74.69%	72.41%	78.53%	77.18%
	ResNet50V2	76.63%	77.07%	75.55%	79.71%	77.62%	72.60%	74.43%	72.48%	78.02%	77.48%
	VGG16	70.98%	67.88%	64.73%	74.06%	66.73%	66.15%	67.75%	64.06%	74.28%	66.82%
	VGG19	70.54%	67.72%	64.36%	74.89%	64.78%	63.86%	65.34%	60.34%	73.94%	63.23%

MAHNOB-HCI dataset, CNN attains an accuracy of 96.77%±2.24% and 97.42%±2.01%, which is significantly higher than the best results obtained by tabular-features-based algorithms, which were 72.43% and 73.03%, respectively. The difference in accuracy between the two methods is 24.34% and 24.39%, highlighting the impact of the proposed input rearrangement. Similarly, on the DEAP dataset, CNN attains an accuracy of 88.68% ± 3.59% and 88.03% ± 3.12%, with a difference of 24.51% and 26.94% when compared to the tabular-features-based approaches. In this case, the classification of valence level benefits more from the proposed input representation than the classification of arousal.

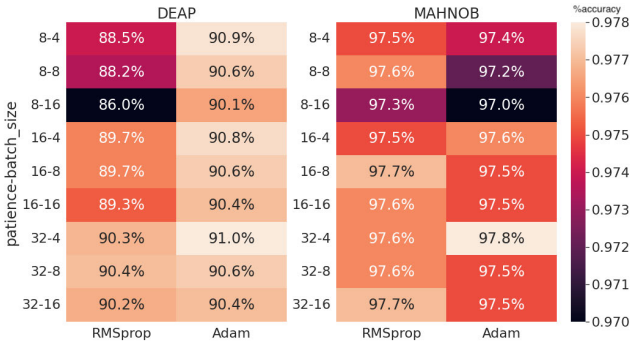
Regarding the other image-based methods, including DenseNet121, MobileNetV2, ResNet, and VGG, it can be observed that all of them achieve similar performances in both datasets compared to the ones achieved by CNN, with MobileNetV2 demonstrating higher performance in all the classification tasks when compared to the others. This is particularly interesting, as MobileNetV2 has a smaller number of total parameters than the other methods, indicating its potential for efficient image-based classification.

In summary, our results confirm that the proposed input representation effectively incorporates valuable information for the emotion classification task. This leads to improved classification performance compared to any other feature-based approaches that were tested.

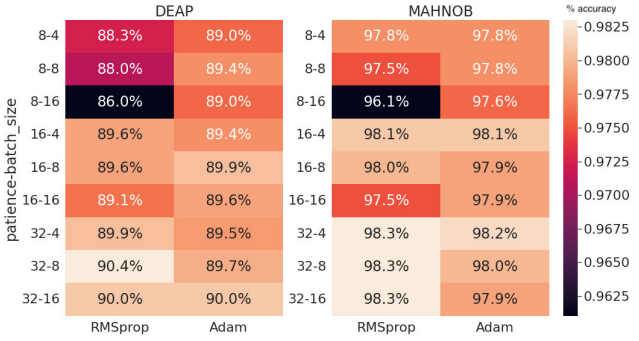
B. HYPER-PARAMETRIZATION

At this point, the best-performing approach (CNN) is investigated further by exploring the space of its hyper-parameters: batch size, early stopping patience, and optimizer. For each of these parameters, different settings were tested: Results of the configurations implemented are summarized in Fig. 5. *Adam* optimizer outperforms *RMSprop* for the HA-LA classification problem, whereas the opposite happens for the HV-LV classification.

Regarding the other two hyper-parameters, patience and batch size, a trend can be seen whereby increasing patience and decreasing batch size improve results up to the configura-



(a) arousal



(b) valence

FIGURE 5. Heatmap showing the hyperparametrization results in terms of % accuracy of the proposed CNN varying the patience, batch size and optimizer for the two selected datasets, DEAP and MAHNOB-HCI, and the two binary classification tasks, arousal and valence.

tion of patience 32 and batch size 4, with which 91% accuracy on DEAP, and 97.8% on MAHNOB-HCI are reached for the HA-LA discrimination task. The same pattern can be seen in the HV-LV classification results, where the best results are obtained with the *RMSprop* optimizer and patience equal to 32, while varying the batch size appears to have no effect.

C. COMPARISON WITH THE STATE-OF-THE-ART

The performance of the CNN with the best hyper-parameterization is compared with the current state-of-the-art, as shown in Table 3. The approach proposed in this study outperforms the state-of-the-art on both MAHNOB-HCI and DEAP datasets, with the exception of the DEAP HV-LV recognition task, where [38] and [39] achieved comparable performance. Zhang et al. [40] propose a hierarchical fusion convolutional neural network to integrate information coming from different modalities, i.e., EEG and other physiological signals, to classify emotions. However, in their study, the authors neglected the EEG spatial information, resulting in lower performances compared to this study, which only considered EEG as the information source. Piho and Tjahjadi [41] achieved good recognition performance, thanks to a non-trivial human-driven processing procedure. Specifically, the authors proposed a feature extraction and selection process in two separate steps via a trial and error process. However, this procedure has to be repeated for each new dataset, as it is not possible to determine the optimal subset of features and channels in advance. Therefore, the optimal subset of extracted features-selected features must be determined through a complex search, as explained by Piho and Tjahjadi [41].

The fact that the approach presented in this study outperforms the one in [41] may suggest that the EEG spatial information, which was not taken into consideration in [41], is actually relevant for the classification problem. The proposed approach employs a small CNN with only two convolutional layers, which automatically performs all the tasks of feature selection, feature learning, and classification, thus being trainable for each dataset in a simple manner.

The approach introduced by Lin et al. [20] relies on an end-to-end fine-tuning of a big CNN such as *AlexNet*, which was pre-trained on the Imagenet dataset. However, the main difference with the proposed approach is that the EEG-derived grayscale image does not consider spatial information. Furthermore, six different EEG images are built from each EEG signal, one for each band frequency, and fed to the network separately. Salama et al. [21] propose an image-based model that exploits a three-dimensional CNN (3D-CNN). This representation is similar to the one proposed by Lin et al. [20], but it also considers the time-domain feature representation. Instead, Yin et al. [38] propose a different approach based on graph-CNN and long short-term memory -NN. An interesting aspect of their study is the use of graph-CNN to model EEG inter-channel relations, which should have the added value of exploiting deeper information than the simple spatial localization of the electrodes.

Our approach outperforms the one proposed by Yin et al. [38] for the arousal-classification problem (0.4% higher) and has comparable performance on the valence-levels classification problem. However, since they did not provide a standard deviation for their results, it is impossible to determine which approach actually offers better performance.

Zhang and colleagues propose a new approach in [39] that is based on heterogeneous convolutional neural networks and multimodal factorized bilinear pooling. This approach constructs a neural network ensemble to classify emotions from a multimodal input that includes EEG and other physiological signals. Compared to our study, Zhang et al. achieved lower performance on the MAHNOB-HCI dataset, and comparable results, which were slightly higher, on DEAP. However, it is worth noting that the comparison with our approach may not be entirely fair, since the authors of [39] did not consider the spatial relationship between the EEG sensors and also included other information sources.

IV. DISCUSSION

In this study, we propose a novel approach for emotion recognition tasks by rearranging EEG-based dynamical features as images, resulting in a new input representation. The approach

TABLE 3. Comparison between the performances achieved by the current state-of-the-art approaches in the subject-dependent HA-LA and HV-LV classification task and the proposed method. * Multimodal architectures based on EEG and other physiological signal.

	MAHNOB-HCI	
	arousal	valence
*Zhang et al. [40]	88.28%	89%
*Zhang et al. [39]	90.37%	90.50%
Piho et al. [41]	94%	94.6%
Our approach	**97.8%±1.8%**	**98.3%±1.9%**
	DEAP	
	arousal	valence
*Zhang et al. [40]	83.28%	84.71%
Lin et al. [20]	87.3%	85.5%
Salama et al. [21]	88.5%	87.4%
Piho et al. [41]	89.8%	89.6%
Yin et al. [38]	90.6%	90.4%
*Zhang et al. [39]	93.22%	90.46%
Our approach	**91%±4.1%**	**90.4%±3.3%**

* Multimodal architectures based on EEG and other physiological signal

focuses on converting traditional EEG feature-based classification problems into image-based ones, enabling algorithms to exploit the spatial information associated with electrode placement. The proposed input representation allows for the use of well-established image recognition approaches such as pre-trained deep neural networks and convolutional neural networks, which can use spatially informed inputs to solve the emotion recognition task.

The proposed approach has been evaluated using two benchmark publicly-available datasets, namely DEAP and MAHNOB-HCI, These datasets consist of data collected from healthy participants with different age ranges and experimental protocols, involving various emotional stimuli, resting states, and labelling procedures. Despite the differences between the datasets, the proposed approach demonstrates high performance on both, indicating its potential for generalization. However, it is worth noting that both datasets share the same EEG sensor arrangement, with 32 channels placed according to the international 10-20 system. Table 2 demonstrates the superior performance of image-based approaches over tabular-feature-based ones in discriminating affective states. Among these approaches, ET, RF, and GradientBoosting (see Tab. 2) are the most effective due to their utilization of powerful decision tree ensembles and recursive feature selection algorithms [36], enabling them to exploit the informative nature of the dataset. While ET and RF result in better overall performance, GradientBoosting is more adept at handling class imbalance, as indicated by its superior F1 score. All image-based approaches outperformed feature-based ones, underscoring the competitive advantages offered by arranging EEG features into an image representation and using image-based classification models.

The effectiveness of the proposed approach is further illustrated by the maximum accuracy scores attained by an image-based method. Specifically, a simple CNN with two convolutional layers trained from scratch achieved 96.77% and 97.42% accuracy in arousal and valence detection, respectively, for MAHNOB-HCI, and 88.68% and 88.03% accuracy in arousal and valence detection, respectively, for DEAP. Moreover, as shown in Table 2, the proposed CNN outperformed other models in terms of F1 score, AUROC, precision, and recall, successfully addressing the class imbalance problem. With further hyper-parametrization, the simple model achieved even higher accuracy rates, with 97.8% and 98.3% accuracy in arousal and valence detection, respectively, for MAHNOB-HCI, and 91% and 90.4% accuracy in arousal and valence detection, respectively, for DEAP (see Table 3). The proposed approach outperforms the majority of existing approaches in the literature, particularly those that do not rely on EEG spatial information [20], [21], [40], [41].

The limitations of the current study are rooted in the fact that the proposed approach is solely based on a rearrangement of EEG features. This is in contrast to the results achieved by Zhang et al. [39], which emphasize the importance of utilizing multimodal physiological data for emotion recognition. Therefore, future developments could potentially integrate brain-heart interplay features to combine the EEG dynamics with cardiovascular data [42], or adopt a neural network-based fusion strategy [39] to merge different signals while preserving the spatial features' arrangement.

V. CONCLUSION

In conclusion, our study confirms the significance of spatial information in EEG analysis and recommends its inclusion in EEG-based emotion recognition tasks. Furthermore, transitioning from feature-based to image-based classification would enable explainability algorithms to provide physiologically plausible insights into the informative features that contribute the most to a given classification [43]. These explainable artificial intelligence approaches are crucial for clinicians and technicians to validate the algorithm's outcomes and enhance the decision-making process.

Given the promising results of this study, future research should concentrate on subject-independent frameworks, as well as emotion recognition tasks that enable a finer sampling of arousal and valence space. Additionally, an interesting research direction could be to integrate the two approaches to leverage all the relations between the EEG electrodes, including the ones that can be extracted with graphs [44], such as the correlation between channels' activity, and the spatial ones, that can be simply provided with an image feature rearrangement. Finally, future research will explore the impact of different EEG sensor arrangements and densities on emotion recognition tasks.

ACKNOWLEDGMENT

Portion of the research in this paper uses the MAHNOB Database collected by Prof. Pantic and the iBUG Group

at Imperial College London, and in part collected in collaboration with Prof. Pun and his team of University of Geneva, in the scope of MAHNOB Project.

REFERENCES

[1] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2003.

[2] A. L. Alfeo, M. G. C. A. Cimino, and G. Vaglini, "Measuring physical activity of older adults via smartwatch and stigmergic receptive fields," in *Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods*, 2017, pp. 724–730.

[3] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Develop. Psychopathol.*, vol. 17, no. 3, pp. 715–734, 2005.

[4] P. J. Lang, "The emotion probe: Studies of motivation and attention," *Amer. Psychologist*, vol. 50, no. 5, pp. 372–385, May 1995.

[5] P. Ekman, *Basic Emotions*. Hoboken, NJ, USA: Wiley, 1999, ch. 3, pp. 45–60.

[6] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[7] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, Feb. 2019, Art. no. 011001.

[8] V. Catrambone, G. Averta, M. Bianchi, and G. Valenza, "Toward brain–heart computer interfaces: A study on the classification of upper limb movements using multisystem directional estimates," *J. Neural Eng.*, vol. 18, no. 4, Apr. 2021, Art. no. 046002.

[9] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–19, Sep. 2020.

[10] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, Apr. 1987.

[11] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[12] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2019.

[13] S. Lemm, B. Blankertz, G. Curio, and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[14] Y. Li, H. Yang, J. Li, D. Chen, and M. Du, "EEG-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-CAM," *Neurocomputing*, vol. 415, pp. 225–233, Nov. 2020.

[15] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, pp. 5455–5516, Apr. 2020.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, 2012, pp. 1–7.

[18] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4700–4708.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[20] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal," in *Proc. Int. Conf. Image Graph.* Cham, Switzerland: Springer, 2017, pp. 385–394.

[21] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, Jan. 2018.

[22] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jun. 2012.

[23] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Aug. 2012.

[24] D. Candia-Rivera, V. Catrambone, and G. Valenza, "The role of electroencephalography electrical reference in the assessment of functional brain–heart interplay: From methodology to user guidelines," *J. Neurosci. Methods*, vol. 360, Aug. 2021, Art. no. 109269.

[25] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–9, Jan. 2011.

[26] D. Candia-Rivera, V. Catrambone, J. F. Thayer, C. Gentili, and G. Valenza, "Cardiac sympathetic-vagal activity initiates a functional brain–body response to emotional arousal," *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 21, May 2022, Art. no. e2119599119.

[27] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[28] J. C. Britton, K. L. Phan, S. F. Taylor, R. C. Welsh, K. C. Berridge, and I. Liberzon, "Neural correlates of social and nonsocial emotions: An fMRI study," *NeuroImage*, vol. 31, no. 1, pp. 397–409, May 2006.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," 2016, *arXiv:1603.05027*.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[31] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.

[32] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proc. 13th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, 2014, pp. 844–848.

[33] A. Bauerle, C. van Onzenoodt, and T. Ropinski, "Net2 Vis–A visual grammar for automatically generating publication-tailored CNN architecture visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 6, pp. 2980–2991, Jun. 2021.

[34] P. J. Bota, C. Wang, A. L. N. Fred, and H. P. Da Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," *IEEE Access*, vol. 7, pp. 140990–141020, 2019.

[35] V. N. Vapnick, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.

[36] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[37] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Exp. Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, Sep. 2011.

[38] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106954.

[39] Y. Zhang, C. Cheng, S. Wang, and T. Xia, "Emotion recognition using heterogeneous convolutional neural networks combined with multimodal factorized bilinear pooling," *Biomed. Signal Process. Control*, vol. 77, Aug. 2022, Art. no. 103877.

[40] Y. Zhang, C. Cheng, and Y. Zhang, "Multimodal emotion recognition using a hierarchical fusion convolutional neural network," *IEEE Access*, vol. 9, pp. 7943–7951, 2021.

[41] L. Piho and T. Tjahjadi, "A mutual information based adaptive windowing of informative EEG for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 722–735, Oct. 2020.

[42] V. Catrambone, R. Barbieri, H. Wendt, P. Abry, and G. Valenza, "Functional brain–heart interplay extends to the multifractal domain," *Philos. Trans. Roy. Soc. A*, vol. 379, no. 2212, Dec. 2021, Art. no. 20200260.

[43] A. L. Alfeo, M. G. C. A. Cimino, and G. Gagliardi, "Concept-wise granular computing for explainable artificial intelligence," *Granular Comput.*, pp. 1–12, Dec. 2022.

[44] T. Azevedo, L. Passamonti, P. Lio, and N. Toschi, "A deep spatiotemporal graph learning architecture for brain connectivity analysis," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1120–1123.

GUIDO GAGLIARDI (Graduate Student Member, IEEE) was born in Pisa, Italy, in 1996. He received the B.S. degree in computer engineering and the M.S. degree (cum laude) in artificial intelligence and data engineering from the University of Pisa, Italy, in 2019 and 2021, respectively. He is currently pursuing the Ph.D. degree with the University of Pisa, the University of Florence, and the University of Siena (in the International Ph.D. Program in Smart Computing) and with the Doctoral School of the Faculty of Engineering Science, KU Leuven, Belgium. His research interest includes explainable artificial intelligence design to physiological signal analysis.

ANTONIO LUCA ALFEO was born in Taranto, Italy, in 1987. He received the B.S. and M.S. degrees in computer engineering from the University of Pisa, Italy, and the Ph.D. degree from the International Ph.D. Program in Smart Computing (University of Pisa, University of Florence, and University of Siena), in 2019. In 2018, he was a Visiting Student with the MIT Media Laboratory, where he studied different swarm intelligence solutions to analyze collective behaviors in smart cities with Prof. Alex Sandy Pentland. From 2019 to 2021, he was a Postdoctoral Research Fellow with the Department of Information Engineering, University of Pisa, where he studied different deep learning approaches for the optimization of maintenance processes in the field of Industry 4.0. Since 2022, he has been an Assistant Professor with the Department of Information Engineering and a fellow of the Bioengineering and Robotics Research Center E. Piaggio, University of Pisa. His research interest includes the design of machine learning pipelines to analyze physiological and behavioral data via explainable artificial intelligence and deep representation learning. He is the coauthor of many international scientific contributions in these fields published in peer-reviewed international journals and conference proceedings.

VINCENZO CATRAMBONE is an Assistant Professor with the Department of Information Engineering and a fellow with the Neuro-Cardiovascular Intelligence Laboratory, Bioengineering and Robotics Research Centre E. Piaggio, University of Pisa. In the past few years, he has been a Visiting Researcher with École Normale Supérieure de Lyon, France, and the Brain Imaging Centre, Maastricht University, The Netherlands. His research interests include statistical and nonlinear biomedical signal and image processing, cardiovascular and neural modeling, and physiologically interpretable artificial intelligence systems. Applications of his research include the assessment of brain–heart interactions in physiological and pathological conditions, brain–computer interfaces, affective computing, assessment of mood and mental/neurological disorders, and neurorehabilitation. He is the author of several international scientific contributions in these fields published in peer-reviewed international journals and conference proceedings. He is involved in several international research projects.

DIEGO CANDIA-RIVERA was born in Los Ángeles, Chile, in 1992. He received the dual B.S. degree in biotechnology and electrical engineering and the Electrical Engineering degree with a specialization in computational intelligence from the University of Chile, in 2014 and 2016, respectively, and the Ph.D. degree in information engineering from the University of Pisa, Italy, in 2022. From 2017 to 2019, he was a Research Engineer with École Normale Supérieure de Paris and the Paris Brain Institute. Since 2019, he has been with the Research Center E. Piaggio, University of Pisa. His research interests include the study of brain–heart interactions through generative models of brain dynamics and machine learning to understand their role in cognition and consciousness. He uncovered that brain–heart interactions reflect the presence and levels of consciousness in severely brain-damaged patients—this work was awarded by the Society of Neuroscience, USA, in 2022.

MARIO G. C. A. CIMINO (Member, IEEE) is an Associate Professor with the Department of Information Engineering, University of Pisa. He is also a Research Associate with the Institute for Informatics and Telematics (IIT) and the Institute of Information Science and Technologies (ISTI), Italian National Research Agency (CNR). He teaches software systems engineering, process mining and intelligence, and advanced programming. He is the coauthor of about 90 scientific publications in international journals and conference proceedings. His research interests include information systems and artificial intelligence. He is a Co-Founder of the ''Machine Learning and Process Intelligence'' Initiative with the Department of Information Engineering. He is a member of the IEEE Computational Intelligence Society (CIS) and ACM. He is the Vice-Chair of the IEEE CIS Task Force Intelligent Agents. He is an Associate Editor of the *Journal of Granular Computing* (Springer) and the *Journal of Ambient Intelligence and Humanized Computing* (Springer).

GAETANO VALENZA received the M.Eng. and Ph.D. degrees from the University of Pisa, Pisa, Italy. He is currently an Associate Professor with the Department of Information Engineering, University of Pisa, and the Head of the Neuro-Cardiovascular Intelligence Laboratory. His research interests include statistical and nonlinear biomedical signal and image processing, cardiovascular and neural modeling, physiologically interpretable artificial intelligence systems, and wearable systems for physiological monitoring. Applications of his research include the assessment of autonomic nervous system activity on cardiovascular control, brain-heart interactions, affective computing, and the assessment of mood and mental/neurological disorders. He is the author of more than 250 international scientific contributions in these fields published in peer-reviewed international journals, conference proceedings, books, and book chapters.

Open Access funding provided by 'Università di Pisa' within the CRUI CARE Agreement