Machine Learning Strategies to Improve Cross-Subject and Cross-Session Generalization in EEG-Based Emotion Recognition: A Systematic Review

Graphical Abstract

Machine Learning Strategies to Improve Cross-Subject and Cross-Session Generalization in EEG-based Emotion Recognition: a Systematic Review [1]

Andrea Apicella, Pasquale Arpaia, Giovanni D'Errico, Davide Marocco, Giovanna Mastrati, Nicola Moccaldi, Roberto Prevete

Highlights

Machine Learning Strategies to Improve Cross-Subject and Cross-Session Generalization in EEG-based Emotion Recognition: a Systematic Review [2]

Andrea Apicella, Pasquale Arpaia, Giovanni D'Errico, Davide Marocco, Giovanna Mastrati, Nicola Moccaldi, Roberto Prevete

- The non-stationarity of EEG signals can lead to the Dataset Shift problem.
- Transfer learning methods improve generalizability in EEG-based emotion classification.
- Adaptive feature extraction also in combination with transfer learning are promising for generalization.

Machine Learning Strategies to Improve Cross-Subject and Cross-Session Generalization in EEG-based Emotion Recognition: a Systematic Review [1]

Andrea Apicella[a], Pasquale Arpaia[a,d], Giovanni D'Errico[b], Davide Marocco[c], Giovanna Mastrati[a], Nicola Moccaldi[a], Roberto Prevete[a]

[a]*Department of Electrical Engineering and Information Technology,University of Naples Federico II, Via Claudio, 21, Naples, 80138, Italy,*
[b]*Department of Applied Science and Technology, Polytechnic University of Turin, , Turin, 10129, , Italy*
[c]*Natural and Artificial Cognition Laboratory, University of Naples Federico II, , Naples, 80133, , Italy*
[d]*Interdepartmental Research Center in Health Management and Innovation in Healthcare (CIRMIS), University of Naples Federico II, Naples, 80138, , Italy*

Abstract

A systematic review on machine-learning strategies for improving generalization in electroencephalography-based emotion classification was realized. In particular, cross-subject and cross-session generalization was focused. In this context, the non-stationarity of electroencephalographic (EEG) signals is a critical issue and can lead to the *Dataset Shift* problem. Several architectures and methods have been proposed to address this issue, mainly based on transfer learning methods. In this review, 418 papers were retrieved from the *Scopus*, *IEEE Xplore*, and *PubMed* databases through a search query focusing on modern machine learning techniques for generalization in EEG-based emotion assessment. Among these papers, 75 were found eligible based on their relevance to the problem. Studies lacking a specific cross-subject or cross-session validation strategy, or making use of other biosignals as support were excluded. On the basis of the selected papers' analysis, a taxonomy of the studies employing Machine Learning (ML) methods was proposed, to-

[1]This work has been submitted possible publication. Copyright may be transferred without notice.

gether with a brief discussion of the different ML approaches involved. The studies with the best results in terms of average classification accuracy were identified, supporting that transfer learning methods seem to perform better than other approaches. A discussion is proposed on the impact of (i) the emotion theoretical models and (ii) psychological screening of the experimental sample on the classifier performances.

Keywords: BCI, EEG, Emotion Recognition, Machine Learning, Transfer Learning, Domain Adaptation, Systematic Review, Generalization

1. Introduction

Emotions are our internal compass and play a primary role in learning, reasoning, decision-making processes, and communication between individuals. In recent years, the interest towards emotions of the Information and Communication Technology (ICT) sector has grown tremendously, giving birth to the new field of *affective computing* aimed at monitoring and predicting emotions in order to improve human-computer interaction [1]. For instance, the introduction of *affective loops* makes it possible to implement increasingly adaptive human-machine interfaces and virtual assistants tailored to users [2]. Furthermore, the outputs of emotion monitoring systems, in the healthcare context, can be useful in the treatment of psychological disorders based on emotional deficits, in autism [3], in the improvement of wellbeing [4], and in stress containment [5].

In particular, in this context, there is a growing interest in the literature for Brain-Computer Interface (BCI) systems based on EEG signals [6]. In fact, the number of annual scientific publications indexed on *Scopus* database on the topic of EEG-based emotion recognition shows an exponential growth trend (see Fig. 1).

Over the years, EEG-based assessment of emotion has been widely employed both in non-clinical and clinical applications. Car driving [7, 8], working environment [9], neuromarketing [10, 11, 12], and entertainment [13] are the main non-clinical application fields. Regarding clinical applications, the main studies about EEG-based emotion assessment concern the measurement of sleep parameters [14], the detection of epileptic seizures [15], and the screening, intervention, and monitoring of autism spectrum disorders [16, 17].

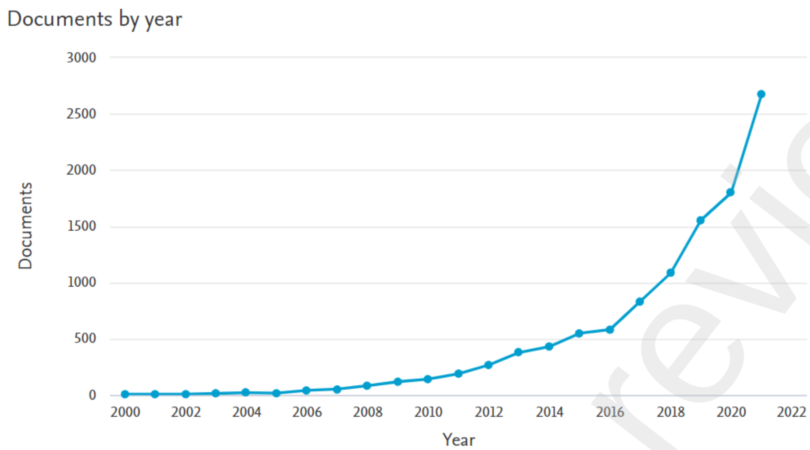

Figure 1: Scopus trend for EEG-based Emotion Recognition studies.

A critical issue underlying the processing and classification of EEG signals is their inherent variability among different subjects or different acquisition times (i.e. sessions) of the same subject, since the EEG signal is usually stochastic and stationary only for short intervals (generally ranging from a few seconds to minutes) [18, 19, 20]. More in detail, the EEG signal is not a Wide Sense Stationary signal [21]. This characteristic of non-stationarity implies a variation in the temporal and spectral characteristics of the EEG signal over time. This is an open issue in the literature leading to a loss of generalizability for classification systems across subjects (*inter-subject* task) and, for the same subject, across different sessions (*intra-subject* task) [22].

Data-driven approaches using Machine Learning (ML) are often employed at multiple levels in the EEG signal processing pipeline to pursue the classification of emotional states and their generalization across subjects and sessions.

Currently, the literature shows increasing use of modern machine learning strategies, adopting deep neural networks and transfer learning-based approaches, such as domain adaptation, domain generalization and/or hybrid methods [23]. This paper proposed a systematic review on the use of machine learning to improve generalizability capabilities in EEG-based emotion recognition systems across different subjects and sessions.

As will be discussed in detail in the next section, several surveys have been proposed in recent years, gathering and discussing the main directions of the literature on this research topic. However, to the best of our knowledge, a focus on the application of ML methods to improve the inter/intra-subjective generalization performance of EEG-based emotion recognition is missing in the literature.

The rest of the paper is organized as follows: Section 2 reviews related works, with reference to recent surveys carried out on this specific topic. Section 3 presents a theoretical background on EEG, with a first part focused on BCIs for emotion recognition and a second part on ML for emotion recognition. Section 4 presents the used search queries and the paper selection process according to the PRISMA method [24]. Section 5 presents the results of the review, proposing a taxonomy of the ML methods currently proposed in the selected papers, discussing the ML methods with respect the proposed taxonomy. A statistical analyses of the results was reported. Section 6 aims to discuss the results obtained, reporting the most promising lines of research and approaches that have emerged and highlighting possible future directions in this area. Finally, Section 7 draws conclusions.

2. Related Works

In recent years, several reviews have been conducted on generalization in EEG-based Emotion Recognition. Alarcao and Fonseca [25] focus on the generic topic of EEG-based Emotion Recognition, presenting a review of papers published in the period from 2009 to 2016. The survey appears interesting in that it focuses on the different stages of the emotion recognition process from EEG signals. Moreover, the survey proposed a criterion for assessing the quality of the papers by applying a set of well-known guidelines (Brouwer's recommendations [26]). However, there is no in-depth analysis on the issue of inter/intra-subject generalization, nor is the EEG-nonstationarity problem addressed. Other reviews [27, 28] analyse studies on the EEG-based classification methods, but without focusing on the emotion domain. Wu et al. [28], offer a non-systematic review focusing on affective BCIs (aBCIs), but without an in-depth analysis of the emotion recognition problem. The study proposed in [29] defers further investigation of the problem of EEG-inter/intra-subject variability to future works. Recently, Li and colleagues [30] published a review focusing on the topic of EEG-based emotion recognition and discussing the importance of transfer learning. While offering some interesting results, it is not a systematic review (only 18 studies were reported without PRISMA methodology to collect them).

This paper proposed a systematic literature review focused on the inter/intra-subject generalization on EEG-based emotion recognition systems and the use of modern ML-based methods as a possible solution.

3. Theoretical Background

3.1. *Emotional theories*

Over the years, different theories on emotions have been proposed but none of these has been universally accepted. Currently, many theories coexist in multiple application fields. Anyway, the *discrete theory* and *dimensional theory* are the most recurrent in literature . The discrete theory identifies universal and innate emotions [31]. Drawing on the Darwinian tradition, Ekman's theory identifies six basic emotions: anger, disgust, fear, happiness, sadness, and surprise [32]. Plutchik, identifies eight basic emotions (anger, anticipation, joy, trust, fear, surprise, sadness, and disgust) and arranges them on a wheel model [29]. In contrast, the dimensional theory represents emotions in a continuous two-dimensional (valence-arousal) or three-dimensional (valence-arousal-dominance) space. *Valence* measures levels of pleasantness (happy vs. sad) of an emotion. *Arousal* identifies degrees of excitement or motivational activation. In the three-dimensional model, the *dominance* dimension is added to valence and arousal to evaluate emotions on a scale between submission and empowerment [6].

Two brain networks underlying the valence and arousal dimensions are identified by the dimensional approach [33]. Conversely, the assumption of discrete approach is that few fundamental emotions are mediated by dedicated neural circuits. Therefore, the two theoretical approaches focus different neurophysiological phenomena with specific spatial signal features. In the framework of the emotion assessment, the specific task (i.e., discrete emotions or emotional dimensions classification) relies on the choice of the reference theory.

3.2. *BCI for Emotion Recognition*

Emotional states can be recognized through several biosignals. In particular, brain signals have received increasing attention from the scientific community. Indeed, the EEG signal is particularly effective for emotion recognition due to its high temporal resolution and non-invasiveness. The EEG signal has a frequency range between [0.01, 100.00] Hz and an amplitude varying typically within the range [-100, 100] μV. Five background rhythms are present in the EEG and can be classified into different frequency bands: delta [0.5, 4.0] Hz, theta [4, 7] Hz, alpha [8, 13] Hz, beta [14, 30] Hz, and gamma [30, 100] Hz.

The International 10-20 Positioning System is an internationally recognized method to place the electrodes on the scalp [34] for recording the EEG signal. The method allows to maintain a standardized EEG electrodes placement proportional to the scalp size and shape in order to preserve the relationship between each location and the underlying brain area. A basic requirement for obtaining a high-quality EEG and for ensuring good contact between the electrode and the skin is to use high-performance electrodes [35]. The electrode-skin contact can either be ensured by adding a conductive gel between the electrode and the skin or by increasing the contact surface that ensures electrical contact. Recently, besides wet electrodes, dry electrodes are employed for the EEG signal recording. A good signal quality and comparable performances with respect to wet electrodes are achieved using dry electrodes [36].

Besides the quality of the EEG signal, the emotion induction methods and the eliciting stimuli have a great impact on the EEG-based emotion assessment. Specifically, the emotion induction methods and the eliciting stimuli represent a crucial point for the effectiveness of the emotional elicitation. Facial and body movements, recall of past events, odors, images, film clips, and music are techniques currently used in laboratories for inducing emotions. Current literature reports that film clips, images, and music are particularly effective to elicit emotions [37, 38]. The use of images over other kind of stimuli represents a great advantage, as images are standardized stimuli. Image datasets were experimentally validated (e.g., International Affective Picture System - IAPS [38], Open Affective Standardized Image Set - OASIS [39], and Geneva Affective Picture Database - GAPED [40]). There are several publicly-available databases of EEG signals that can be used for emotion recognition (e.g., DEAP [41], SEED [42], and DREAMER [43]). Each dataset contains different physiological signals and is characterized by a well-established experimental setup (in terms of stimulus sources, emotional theory adopted, number of subjects, and psychometric metrological references). For a comprehensive description of the various available datasets, see [30].

In case of self-produced datasets, the EEG data must be carefully preprocessed in order to be used in the emotion assessment. Some steps are often helpful to achieve a successful EEG signal preprocessing: (i) line noise removal, (ii) referencing, (iii) bad channels removal, and (iv) artifacts removal (see [19] for insights).

Once the EEG signal has been pre-processed, it is usually divided into epochs, and a feature extraction process is then applied. EEG features can be categorized into three domains, namely time, frequency, and time-frequency.

- Time domain: the main features are the statistics of the signal, such as mean, variance, skewness, kurtosis, etc [44, 45, 46]. Other time-domain features are the Hjorth parameters, namely Activity, Mobility, and Complexity [47]. Good results in the recognition of emotional states can be achieved by using entropy-based features, i.e., approximate, sample, differential, and wavelet entropy [48]. Higher-order crossing (HOC), the fractal dimension, and the Non-Stationary Index (NSI) [49, 50, 51] are further time domain feature often used for the EEG analysis.

- Frequency domain: the most used feature is the power spectral density (PSD). PSD is the signal power in the unit frequency band [52]. Other representative features of different emotional states involving the PSD are: (i) logarithm, (ii) maximum, (iii) minimum and, (iv) standard deviation of the power spectrum.

- Time-frequency domain: the time-frequency analysis (TFA) allows to observe spectrum changes with time[53]. The short-time Fourier transform (STFT),the continuous wavelet transform (CWT), the discrete wavelet transform (DWT) [54], matching pursuit, and empirical mode decomposition are the most used methods to extract time-frequency features.

The number of EEG features is often very high, therefore a feature selection strategy is required [55]. Another critical point is the large number of EEG channels often used for signal acquisitions. A high number of channels can lead to high computational complexity. Therefore, the selection of the most informative EEG channels can be crucial [56].

3.3. Machine Learning for Emotion Recognition

After the EEG signal has been properly pre-processed and a suitable set of features has been extracted, the data are ready to be fed to a supervised

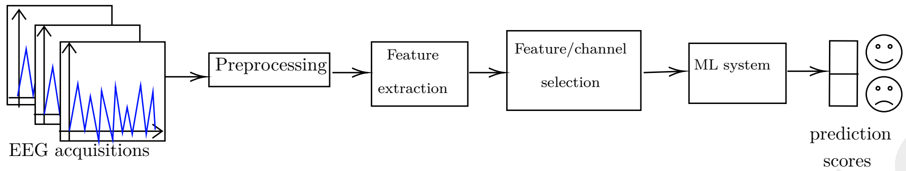

Figure 2: A pipeline of a classical ML process involving EEG signals.

ML system. The typical pipeline of a ML framework applied to an EEG emotion recognition task is reported in Fig. 2.

A large part of the current literature on Emotion Recognition proposed methods framed into Transfer Learning approach. This is because in the classical supervised ML framework a set of already labeled data has to be available. This implies that, in EEG emotion recognition tasks, a set of EEG signals recorded from one or more subjects has to be labeled with the emotion felt during the acquisition. Labeled data can then be used to train the ML system, generating a ML model able to classify the input data. Once the ML model is obtained, new unlabeled data can be fed to the ML model to estimate the corresponding emotion/class. Reserving a portion of the labeled data outside the training stage to evaluate the trained model is a good practice. These data can then be used to evaluate the final model predictions using suitable performance metrics (e.g., accuracy). However, a standard hypothesis of traditional ML methods is that all available data come from the same probability distribution, no matter if involved in the training process or not. Due to the characteristics of the EEG data, this assumption results not always verified in the EEG signal. Indeed, the EEG recordings of different subjects can be strongly different from each other, even under the same conditions [19]. Strong differences can arise also for EEG recordings acquired from the same subject but in different times/sessions, leading to low generalization performance in cross-subject/session problems. In the current literature, this problem was initially addressed by exploiting additional unlabeled data belonging to the target subject/session during the training stage (Transductive Learning approaches). However, these methods do not make any consideration about the data distributions. Indeed, the training EEG data can belong to probability distribution(s) sensibly different from the ones of the data used outside of the training stage. In ML literature, this can be considered an instance of the Dataset Shift problem [57]. Dataset Shift occurs in an experimental environment where the standard ML assumption is

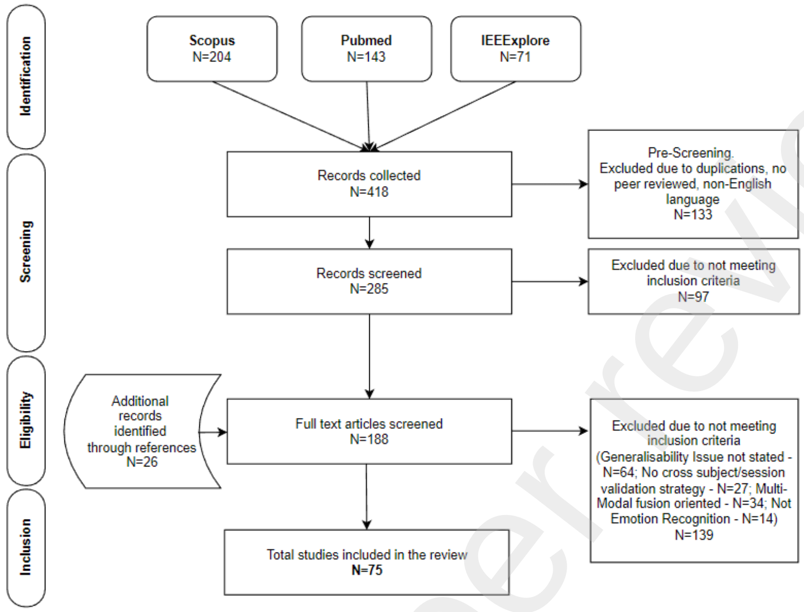

Figure 3: PRISMA flow diagram of the systematic review process.

not verified, i.e. the distributions of the training data and the data used outside the training stage may be different. The idea that data used inside and outside of training stage can belong to different probability distributions is the main hypothesis of the transfer learning approaches.

In the last years, several ML architectures and methods have been proposed to address the dataset shift problem following the base assumptions of transfer learning, and different categorizations of these methods have been reported [58, 59]. One of the first and most important review on Transfer Learning methods was proposed in [58]. However, several new strategies were proposed in the following years (e.g., Domain Generalization-based works).

4. Papers selection method

The present literature review took into account the guidelines for systematic literature reviews presented by Kitchenham ([60]). In addition, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations were adopted in order to transparently report the document extraction process ([24]). The survey was conducted covering the period between January 2010 to March 2022, using the following databases: *Scopus*, *IEEE (Institute of Electrical and Electronics Engineers)*, *Xplore*, and *PubMed*.

In accordance with the PRISMA recommendations, the review pipeline comprised four successive steps: 'Identification', 'Screening', 'Eligibility', and, finally, 'Inclusion', which considerably reduced the amount of surveyed work. For the initial identification of the articles, the following query was used in all selected data sources, taking into account titles and abstracts: EEG AND (Emotion OR Preference) AND ("Domain Adaptation" OR "Domain Generalization" OR "Transfer Learning" OR "Adversarial" OR "Transfer" OR "Cross Session" OR "Cross Subject" OR "Cross Gender" OR "Non-stationary EEG").

From the first phase, 418 articles were collected. Therefore, duplicated papers, not peer-reviewed, or not written in English were excluded from review as an initial prescreening process. For each paper that passed the screening stage, a careful examination of the full text was carried out. In a final screening, further papers were excluded according to the following exclusion criteria: (a) generalizability issue not explicitly stated, (b) absence of a cross-subject/cross-session validation strategy, (c) adoption of a 'multimodal' approach (i.e. aimed at supporting EEG-based classification with other biosignals and/or information), (d) lack of focus on emotion recognition. As a result, 75 papers remained and were included in the review analysis. The complete flow diagram of the systematic review process according to PRISMA is presented in Fig. 3.

5. Results

The analyzed works can be divided into two main big families, based on the assumption about the origin of the handled data (e.g. from a single population/domain described by the same probability distribution or from different populations/domains):

- *Classical ML approaches*: all data are assumed to belong to the same population and are described by the same probability distribution;
- *Transfer Learning (TL) approaches*: these methods rely on the hypothesis that data can belong to different populations (domains). Data with heterogeneous probability distributions can lead to the dataset shift problem, resulting in a loss of the model's generalization. In general,

the main goal of a TL method is to exploit the knowledge extracted from a *Source* domain to solve a problem in a *Target* domain. TL methods try to reduce the discrepancy between the probability distributions of the different domains.

5.1. *Classical ML approaches*

A model trained on a set of EEG data acquired from a given subject at a specific time (or during a specific session) could not work as expected in classifying EEG signal acquired from a different subject or from the same subject at different times. In other words, the model can result in poor generalization performance. To deal with this problem, several solutions based on ML approaches have been proposed over the years. One attempt to mitigate the problem was the adoption of Transductive methods. Transductive methods [61] start from the hypothesis that the unlabeled data target of the classification problem are available in the training stage. However differently from the DA methods, no assumption about the distribution of the data is made. The idea is that in several problems there is only a specific set of data (usually corresponding to the test set) to classify, and it is available at training time. Note that standard ML approaches the goal is to generalize on new unseen data and the test set is used only to validate the learned model on new, unseen data adopting the inductive learning principle [61], while in Transductive learning the goal is to correctly classify the test set only, therefore the classification problem is defined only on the test data. Transductive SVM (TSVM, [62]) is an example of a transductive method. Differently from classical SVMs that leverage only on labeled data, TSVMs exploit both labeled and unlabeled test data to find the best decision boundary between the classes. In other words, the target data is an additional set of information about the data in the training stage. One of the main drawback of TSVM is that an estimation of the number of elements of each class in the test set is needed. Progressive TSVM (PTSVM, [63]) tries to progressively solve this problem labeling the unlabeled data during the training stage. However, the only study collected in this review that explicitly uses transdutive methods is [64], where a PTSVM is used in a cross-session Emotion Recognition problem on EEG data acquired from different subjects. Instead, the greatest part of the reviewed proposals consisted in proper feature transformations and/or feature selection processes. The former wants to transform the data features to hold only the most useful information, assuming that it is shared between all the subjects/sessions, while the latter are methods to select only the most useful features from the input signals without changing them. Usually, the feature extraction/selection is one of the first step of a machine learning pipeline, where the data are transformed before being fed to the machine learning model. These methods adopt the classical Machine Learning framework, that is no knowledge of the effective data on which the model will be effectively used after the training is available in the training stage. This can be viewed as a consequence of the the starting hypothesis of the traditional ML methods stating that all the available data, no matter if used in the training process or not, come from the same probability distribution, therefore the training data are enough for the generalization purpose. In other words, the training data are enough to generalize over all the possible data.

In the EEG Emotion Recognition case, this means that a proper EEG data transformation is enough to allow a ML model to generalize well on never seen EEG data, independently from the fact the these new data belongs to a subject/session used during the training stage or not. Going deeper, in a ML problem on EEG data the feature extraction and selection process can be made considering two different aspects: i) the acquired EEG features or ii) the electrodes. In the first case, a proper transformation or selection strategy for the EEG features is made.

EEG features extraction/selection strategies

The reviewed literature proposed different works discussing if several known feature extraction methods are suitable to generalize across several Emotion Recognition datasets [65, 66]. In particular, in [65] the authors investigated the robustness of Emotion Recognition features in different experimental conditions, subjects, and datasets.

In [67, 68] Sequential Backward Selection (SBS) was applied to find a good set of features able to generalize across different subjects. To find the best subset of features, SBS decreases the number of features in an iterative way measuring, at each step, the performance on a given classifier (SVM in [67], Decision Trees in [68]). SBS method is adopted to exploit the significant differences between the classes. A Leave-One-Subject-Out (LOSO) verification strategy was employed on DEAP and SEED datasets in [67], while [68] validates its results on DEAP and self-produced data.

In [69, 70] a family of Transferable Recursive Feature Elimination (TRFE) methods are used to remove the EEG features resulting not generic for all the involved subjects. The proposed feature selector is validated using SVM classifiers on DEAP dataset, analyzing both the within-subject and the cross-subject behaviours. In [71] Cross-subject Recursive Feature Elimination (C-RFE) is exploited to rank the features in order of importance removing the ones giving a low contribution to the classification. The method is validated on SVM classifiers.

In [72], an improved version of the well-known Differential Entropy features is proposed. Differently from the classical DE which consider only the frequency domain of the data, The Dynamic Differential Entropy (DDE) features take into account both the time-domain and the frequency domain. The goal is to learn a set of common characteristics across different subjects maximizing the difference between classes and minimizing, at the same time, the difference within classes.

In [73] a latent representation of the EEG data from SEED and DEAP is learned through a Variational Auto Encoder (VAE, [74]) and then classified using a LSTM. VAEs start from the hypothesis that all the data are generated by a random process involving latent variables. A VAE is usually trained to encode the input data into a latent representation, and then mapping it to a reconstructed version of the data. [73] assumes that i) there exists learnable intrinsic features shared across several EEG signals belonging to different subjects and taking part in emotional processes, and ii) these intrinsic features can be learned and encoded in the VAE latent representations. The power of VAE to represent latent EEG factors is also investigated in [75], together with classical Auto-Encoders (AEs) and Restricted Boltzmann Machines (RBMs). Final emotion classifications are made with an LSTM, while the generalization performances are evaluated in LOSO on DEAP and SEED dataset.

In [76] the cross-subject problem is tackled using Variational Mode Decomposition (VMD) as feature extraction technique. The proposed framework is validated in an Hold-Out (HO) way, taking care that no intersection exists between subjects' data in the training and the test set. Performance are measured using a DNN as emotions classifier. Despite the encouraging results reported, no reason about why the proposed system works well in a cross-subject approach seems to be provided. In [77, 78, 79] is shown that some normalization functions usually used to preprocess the EEG data can affect the cross-subject performances.

In particular, in [77] several normalization functions were applied and evaluated following two different schemes: i) *All-subjects*, where the whole dataset was normalized, ii) *Single-subject*, where the normalization is applied individually to each subject. The All-subject schema is the most common method to normalize the entire dataset. Single-subject, instead, consider each subject individually, applying normalization on each subject. The authors empirically shown, on SEED dataset, that Single-subject Z-score performs better in a EEG emotion recognition problem respect to other normalization schemes, such as min-max normalization. On the same data, in [78] the authors apply single-subject $Z - score$ normalization after each layer of a neural network (Stratified Normalisation).

Differently, in [79] a simple transformation of the original data is proposed. It consists in transforming the original features into binary vectors, having 0 or 1 as components values if the feature is lower or higher than the median feature value, respectively. The author assumption is that this leads to a more effective reduction of the subject-dependent part of the EEG signal.

Channel selection strategies

In [80, 81] a general channels set for Emotion Recognition valid for several subject is searched exploring different channels selection strategies. To achieve EEG-based cross-session emotion recognition, the authors of [82] separate discriminative features from the noisy and redundant ones learning the importance of the EEG channels in an Emotion recognition task. The proposed strategy is evaluated on pairs of sessions chosen a-priori.

In [83] a neural network to classify emotion by EEG signals is proposed. The proposed network introduces a channel-attention layer to select the most important channels for a set of emotions. Notably, the different subjects' personalities are taken into account, grouping together subjects with similar personalities and training a different network for each group. Validation is made on the ASCERTAIN dataset. This dataset results particularly suited for this task, since it links together personality and emotional state and physiological reactions. The structure of the electrodes is taken into account and modeled as a graph.

To model structured data, graph representation methodologies resulted effective achieving significant performance in many applications, included EEG emotion signal processing [84]. Indeed, GNNs are useful to retain the spatial structure of the electrodes disposition. Usually, the graph structure is fixed and given a priori following the spatial disposition of the electrodes on the scalp. Differently, Dynamical Graph Convolutional Neural Networks

(DGCNN, [84]) and Self-Organized Graph Neural Network (SOGNN, [85]) changes the graph structure leveraging on the input brain signals, instead of relying on a predefined graph structure. The resulting graph can be processed by graph convolutional layers to extract the more suitable features and channels for emotion recognition. The features obtained are also tested in cross-subject scenarios [86].

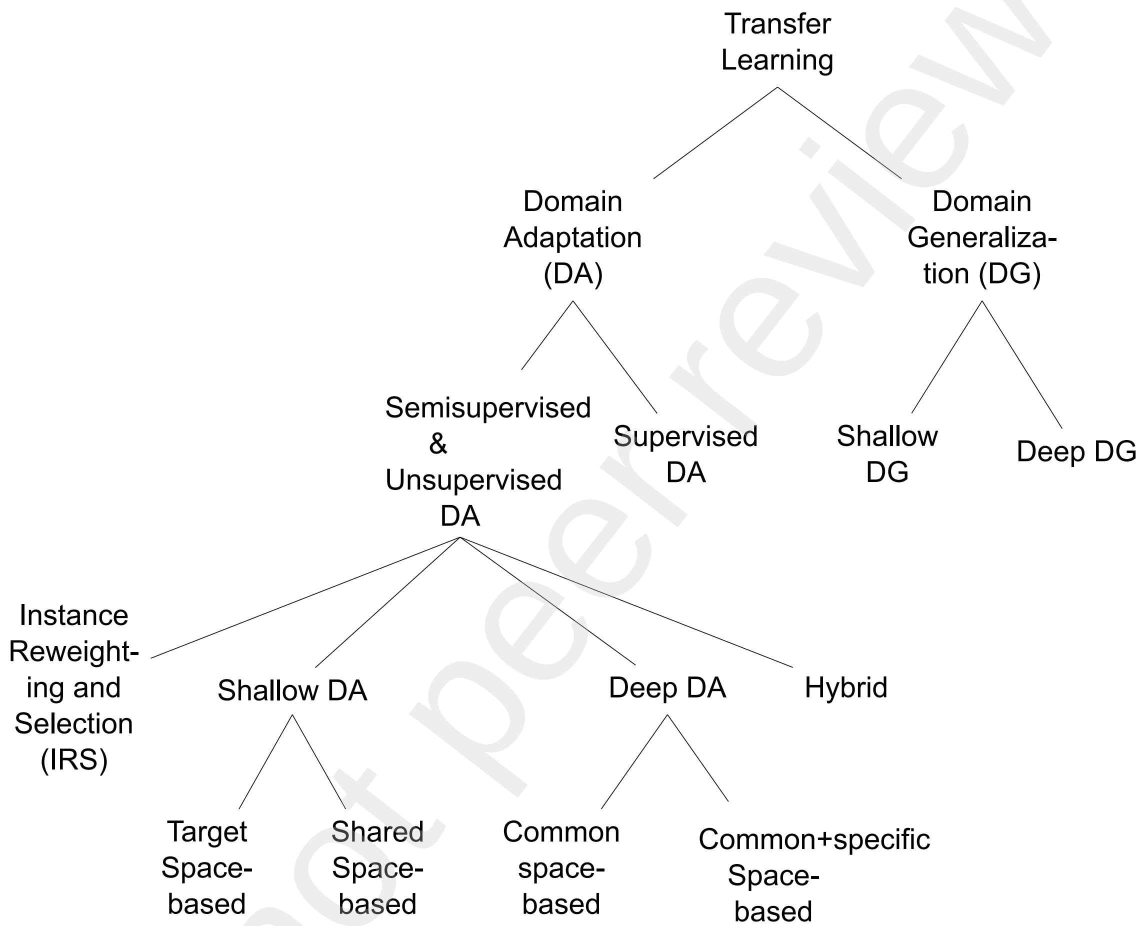

Figure 4: The proposed taxonomy of the Transfer Learning methods in EEG-based emotion recognition. Starting from the root node, the criterion ”target data used in the training phase” leads to the creation of two child nodes: Domain Adaptation (if yes) and Domain Generalization (if no). Domain Generalization in turn generates two nodes depending on the type of data transformation: learned (Deep DG) or unlearned (Shallow DG). From Domain Adaptation, the Supervised DA node is generated if all target data labels are used in the training phase; otherwise, the Semisupervised/Unsupervised DA node is generated. From the latter node, four branches distinguish the data handling strategy: (i) data is transformed using an a priori defined function (Shallow DA), (ii) the data transformation function is learned as part of the method (Deep DA), (iii) data is selected or reweighted using some strategy (Instance Reweighting and Selection), and (iv) a combination of the previous approaches is applied (Hybrid). The Shallow DA methods are divided into two leaf nodes according to the projection space of the transformation: (i) all data is projected into the same space as the Target domain (Target Space-based), (ii) a new shared space is used between the Source and Target (Shared Space-based). Similarly, for Deep DA methods, two branches lead to two leaf nodes: (i) Common Space-based if the Source and Target are projected into a new shared space, (ii) Common+Specific Space-based if Source and Destination are first projected into a single shared space, then, for each available domain, a projection is used in an ad hoc space.

In the following of this section, the reviewed papers are discussed considering the belonging family (classical ML and TL approaches). In particular, TL-based works are discussed according to the proposed taxonomy.

5.2. ***TL methods***

With respect to Classical ML, TL approaches are gaining popularity thanks to their reported better performances. In literature, current TL methods are divided in two main categories according to the use of target data in the training phase:

- *Domain Adaptation (DA) methods*: the DA general assumption is that data of the Target are available during the training of the model, together with data belonging to the Source domain(s). For instance, in the EEG Emotion Recognition, data acquired from both source and target subjects/sessions are available during the model construction.

- *Domain Generalization (DG) methods*: in these methods, data belonging to several domains are available and can be used during the training, but no data from the target domain are available during the training stage. The knowledge extracted from multiple source domains is exploited to improve the model generalization. For instance, in the EEG Emotion Recognition, labeled data acquired from different subject/session can be considered as belonging to different domains, and can be used to build a model able to generalize to a new unseen subject/session, where no data are available during the construction of model. Domain Generalization methods can be splitted in two subcategories depending on the type of data transformation: learned (Deep DG) or unlearned (Shallow DG).

Regarding the DA methods, a further division can be made, depending on the use made of the target data labels in the training phase:

- *Supervised DA* methods (also known as *PreTrained* methods, PT): these methods benefit from the availability of labeled data from the target subject/session during the training stage; Supervised DA methods adapt a model already trained on a known Source domain to work in a new Target domain, where a labeled dataset can be sampled. Since supervised DA methods usually relied on a model already trained, they are also known as PreTrained methods.

- *Unsupervised and Semisupervised DA (UDA)* methods: these methods benefit from the availability of unlabeled data coming from the target subject/session during the training stage. The method is Unsupervised if only unlabeled target data are exploited. Instead, the method is Semisupervised if further labeled target data are available.

UDA methods, in turn, can be distinguished based on the strategy for handling data:

- *Shallow DA*: data are transformed using an a priori defined function;
- *Deep DA*: the data transformation function is learned as part of the method;
- *Instance Reweighting and Selection (IRS)*: data are selected or reweighted using some strategy;
- *Hybrid*: a combination of the previous approaches is applied.

The Shallow DA methods can be further divided according to the projection space of data transformation: (i) all data are projected into the same space as the Target domain (*Target Space-based - TSB*), (ii) a new shared space is used between the Source and Target (*Shared Space-based - SSB*). Similarly, two subcategories can be identified also for Deep DA methods: (i) *Common Space-based (CS)* if the Source and Target are projected into a new shared space, (ii) *Common+Specific Space-based (CSS)* if Source and Destination are first projected into a single shared space, then, for each available domain, a projection is used in an ad hoc space.
Relying on the above considerations, a taxonomy of the TL methods used in EEG-based emotion recognition is reported in Fig. 4.

Transfer Learning methods are based on the concepts of *Domain* and *Task*. Following the survey of Pan et al. [58], a Domain can be defined as a set $D = \{F, P(X)\}$ where F is a feature space and $P(X)$ is the marginal probability distribution of a specific dataset $X = \{x_1, x_2, \ldots, x_n\} \in F$. Instead, a Task is a set $T = \{L, f\}$ where L is a label space and f is a predictive function usually learned by the data. For instance, $f(x_i)$ assigns the predicted label to $x_i \in X$. Therefore, f can be equivalently viewed as the probability of a label y given a data x, i.e. $p(y \in L|x \in X)$.

A Dataset of n points can be defined as a set $S = \{(x_i \in X, y_i \in L)\}_{i=1}^{n}$. Transfer learning wants to exploit the knowledge of a domain D_A on a task T_A to resolve the same or another task T_B on another domain D_B.
By the definition of domain, it is straightforward that two domains $D_A = \{F_A, P(X_A)\}$ and $D_B = \{F_B, P(X_B)\}$ can be considered different if they differ in the feature spaces or in the marginal probability distributions. Obviously, the same holds for two Tasks $T_A = \{L_A, f_A\}$ and $T_B = \{L_B, f_B\}$. More in details, the following cases can happen:

1. $D_A = D_B$ and $T_A = T_B$: since the Tasks and the domains are the same, this can be considered a standard ML Problem.
2. $D_A \neq D_B$: $F_A \neq F_B$ or $F_A = F_B$ and $P(X_A) \neq P(X_B)$
3. $T_A \neq T_B$: $L_A \neq L_B$ or $f_A \neq f_B$.

Due to the non-stationarity of the EEG signals between different subjects/sessions, an emotion classification problem can be viewed as a multi-domain problem where the data belonging to each subject/session are sampled from different domains. More specifically, given two different subjects A and B, a common feature space is assumed to be shared by the two domains (i.e. the EEG data representation), the conditional data distributions $P(L_A|X_A) = P(L_B|X_B)$ are assumed to be the same, but the marginal probability distributions are assumed different on the available data, i.e. $P(X_A) = P(X_B)$. Therefore, generalizing across different subjects/sessions can be viewed as reducing a discrepancy measure between several domains.

In the current literature, TL strategies can be divided into DA and DG families. These families differ mainly in which data are processed during the learning stage. DA methods start from the hypothesis that data sampled from at least two different domains are available, consisting in one or more *Source* domains and one *Target* domain. Usually, methods involving more that a single source domain are said *multi-source*. In contrast, DG methods rely on the hypothesis that $d \geq 2$ source domains together with their labeled samples are available, while any data from the Target domain is unknown. DA and DG methods are getting a great deal of attention in the scientific literature in different contexts (e.g. image classification and voice recognition), and several proposals have been made until now. One trend of the literature is to adapt DA/DG methods originally proposed for a context to another one. For example, in [87] methods to adapt DA strategies for image classification to EEG emotion classification are proposed. However, each context has its characteristics and peculiarities, making the transfer of a DA method from a task to another task not immediate. Several attempts were made by the scientific community to adapt well-established DA/DG methods in tasks involving the processing of EEG signals in the emotion recognition field.

Emotion Recognition tasks usually involve several subjects or sessions with different statistical properties, therefore they can be easily reduced to TL framework. In particular, since data belonging to different subjects/sessions can have different statistical properties due to the non-stationarity of the signal, each subject/ can be viewed as a different domain. However, it is interesting to notice that several TL strategies proposed in literature were developed assuming that the generalization problem is composed by two domains, the former for the corresponding to *all* the available labeled data (usually the training set), and the latter to the unlabeled one (usually the test set). Several recent Emotion Recognition works proposed strategies following this framework, considering all the labeled data available as belonging to the same domain, regardless to actual probability distributions they belong to. This can be viewed as strong assumption, since it implicitly assume that all subjects/sessions belong to the same probability distribution, that it is the same to consider all the data as belonging to the same subject/session. Instead, more recent works project the available data in a multi-source framework, considering each labeled subject/session as belonging to different domain.

Regarding DA methods, they can further be divided into the following subfamiles:

- *Unsupervised/semi-supervised Domain Adaptation (UDA)* methods;
- *Supervised DA*, also known as *PreTraining (PT)*, methods.

The main difference between them is that, while labeled dataset $S_{Source} = \{(x_i, y_i)\}_{i=1}^{n}$ can be sampled from the Source domain(s), only feature data points $X_{Target} = \{x_j\}_{j=1}^{m} \in F_{Target}$ can be sampled from the Target one, without knowledge (unsupervised DA) or minimal knowledge (semi-supervised DA) on their real labels.

UDA methods

We consider a method as UDA if it uses unlabeled data belonging to the Target domain during the training stage. If no extra labeled target data is available in the training stage, the method is said *Unsupervised*, otherwise if there are also labeled target data, the method is considered *Semisupervised*.

Several UDA methods relied on minimizing discrepancy measures between the Source and the Target domains. In [59] these methods are categorized into *shallow DA* and *deep DA*, where

A) *Shallow DA*: a a-priori defined function for a new data representation is given. At most the mapping parameters are learned, without affecting the starting data representation;

B) *Deep DA*: the data representation is fully learned as part of the DA strategy.

However, this categorization does not consider works leveraging on the hypothesis that not all the training data can effectively be useful for the target space. In order to avoid negative transfer, a selection of the training data may be necessary. Therefore, in this work the *Instance Reweighting and Selection (IRS)* category is added. Reviewing the literature, it results that these methods are used, in some case, together with Shallow DA and Deep DA. In the proposed taxonomy, such methods are identified as belonging to the *Hybrid* category.

In the following part of this section, reviewed studies according to the categorization above discussed are reported.

A) Shallow DA methods

Different Shallow DA strategies were proposed in literature, usually relied on one of following alternatives:

- *Target Space-Based (TSB)*: searching for a good transformation which directly maps data belonging to the Source domain to the Target domain space;

- *Shared Space-Based (SSB)*: searching for a good transformation which maps Source S and Target T data in a new shared space having minimal discrepancy between S and T.

Once all the data are projected in a common space, any supervised method can be applied for classification, as both source and target domains follow a similar distribution.

- *TSB methods:*

As TSB proposal, [88] tried to align the source space toward the target one (Subspace Alignment, SA). Rather than using the data in their original feature spaces, PCA is adopted for a more robust and compact data representation. More specifically, two PCA projection matrices Z_S and Z_T are computed for the Source and the Target domain, respectively. Therefore, a transformation matrix M able to align the source space to the target one is searched by an optimization problem, that is

$$\arg\min_{M} ||Z_S M - Z_T||_F^2.$$

This problem has a closed form solution, that is $M = Z_S^T Z_T$.

In [89] Adaptive Subspace Feature Matching (ASFM) is proposed for EEG-based Emotion Recognition. Relying on SA, ASFM takes in care that subject attention level and user fatigue can lead to mismatched marginal and conditional distributions of the data.

Differently from other DA strategies, in [90] (Multi-Subject Subspace Alignment, MSSA) the ASFM strategy is applied to each source subject individually, then the projected data are fed to different subject-specific classifiers.

Other data transformations have been investigated in the DA scenario for EEG emotion recognition, such as Robust Principal Component Analysis (RCA) [91] in [92]. RCA decomposes a set X of data as $X = L+S$, with L and S superimposed matrices, in particular the former is a low-rank matrix, the latter a sparse matrix. These matrices are computed resolving the following optimization problem:

$$\min_{L,S} ||L||_* + \lambda ||S||_1$$

where $||\cdot||_*$ is the matrix nuclear norm, $||\cdot||_1$ the l_1 norm and λ a weighting parameter. In [92] a proposal to build a Cross-Day emotion recognition model using RCA is made.

In [93] a method originally proposed for personalized handwriting recognition (Style Transfer Mapping, STM [94]) is adapted for EEG emotion recognition task to generalize across different subjects. In a nutshell, STM maps source data to target data by an affine transformation. The solution of the proposed problem is in closed form, so it can be easily computed. Few labeled target data are used to select source data, therefore it starts from the hypothesis that a small amount of labeled data is available.

- *SSB methods:*

On the other side, among the SSB methods, the Maximum Mean Discrepancy (MMD,[95]) is one of the most used discrepancy measure in DA/DG strategies. MMD was originally proposed to test if two probability distributions are different or not. Formally, the authors show that, in a Reduced Kernel Hilbert Space (RKHS), a discrepancy measure between two distributions p and q can be defined as

$$MMD(p, q) = ||\mathbb{E}_{X_S \sim p}(\phi(X_S)) - \mathbb{E}_{X_T \sim q}(\phi(X_T))||_H^2$$

where $\phi(\cdot)$ is an appropriate feature mapping. In [95] is proven that, in a RKHS, $MMD(p, q)$ is 0 if and only if the two distributions p and q are the same.

MMD can be empirical estimated as the difference between the averages of two dataset sampled from the two distributions projected in a RKHS. Therefore, considering X_S and X_T as two sets sampled from the Source and the Target domain respectively, empirical $MMD(X_S, X_T)$ can be expressed as:

$$MMD(X_S, X_T) = ||\frac{1}{|X_S|}\sum_{i=1}^{|X_S|}\phi(\mathbf{x}_S^{(i)}) - \frac{1}{|X_T|}\sum_{i=1}^{|X_T|}\phi(\mathbf{x}_T^{(i)})||_H^2$$

where $\mathbf{x}_S^{(i)}$ and $\mathbf{x}_T^{(i)}$ are elements of X_S and X_T respectively. In other words, having two samples beloning to two different distributions, the distance between the two distributions can be estimate through the distance between the averages of the samples projected in a RKHS.

Transfer Component Analysis (TCA, [96]) is one of the most used MMD-based DA method. In the proposing work, two different TCA versions were proposed: i) an *unsupervised* version, consisting in finding a data transformation such that the data variance is maximally preserved and, at the same time, the MMD distance of the domains distributions is minimized, and ii) a *supervised* one, where the dependence between training data and labels is taken into account.

A performance evaluation of the Unsupervised TCA applied on EEG data for Emotion Recognition was made in [97]. Instead of using all the available EEG data, a random selection of samples from Source domain data, letting out the data of a subject as target. Since TCA allows to project data in a reduced space, in [98] several spaces with different dimensions are evaluated on SEED dataset. Instead, in [99] TCA is tested on self-made EEG data.

In [100] through Transfer Sparse Coding (TSC) the MMD was used to find a sparse representation of image data sampled from different distribution. Sparse code representations are well-known data approximation obtained as linear combinations of elements in a set of basis functions. In a nutshell, a sparse coding method searches for a representative over-complete set of basis functions (a *dictionary*) together with an encoding that best represent the data. In its simplest form, the sparse coding problem can be expressed as

$$\min_{B,S} ||X - BS||_F^2 + \lambda \sum_{i=1}^{n} |\mathbf{s}_i|$$

where $X \in \mathbb{R}^{m \times n}$ is a matrix containing addthe n data points to approximate while $B \in \mathbb{R}^{m \times k}$ and $S \in \mathbb{R}^{k \times n}$ are the dictionary matrix and the encoding matrix respectively, with $k > m$ to ensure the over-completeness. The sparsity is induced by the second equation term on the coefficient matrix columns $\mathbf{s_i}$ and regulated through the hyperparameter $\lambda \in \mathbb{R}$. However, if X is composed of data sampled from two different domains (e.g., $X = [X_S|X_T]$) the above formalization does not take into account the differences between the marginal distributions. To deal with this problem, [100] adds a further regularization term to the objective function that considers the MMD distance between the different domains of the input data.

Similarly, PCA and Fisher criteria [101] are used together in in [102] with the aim to compute a common dictionary between source and target domain, but preserving the local information between samples together with the discriminative knowledge between the domains. This work required a little set of labeled data from the target domain during the training stage, falling in the Semi-supervised DA approaches.

While it is not specifically designed for Domain Adaptation, Kernel-PCA (KPCA,[103]) is often used in comparisons with several DA methods. In a nutshell, KPCA uses the kernel trick [104] to project the data into a kernel space followed by a PCA. A comparison between Kernel-PCA and TCA for EEG emotion recognition is reported in [97].

Proposed in [105] Subspace Alignment Auto-Encoder (SAAE) combines together auto-encoders and subspace alignment. The subspace alignment is obtained through MMD and KPCA to maximize the embedded data variance. Before the transformation, an auto-encoder trained on source and target data was employed to extract features from the data.

In [106] several shallow DA approoaches such as TCA, KPCA, TSVM are evaluated on SEED dataset in a Leave-On-Subject-Out approach, while in [107] similar methods are tested on SEED and DEAP also for Cross-Dataset generalization.

[107] exploits Maximum Independence Domain Adaptation (MIDA) [108] in the EEG Emotion Recognition case. As similar methods, MIDA projects the data into a subspace able to reduce the inter-domain discrepancy in distributions. In this case the independence between domain is computed with the Hilbert-Schmidt Indipendence Criterion (HSIC, [109]).

B) Deep DA methods

In deep DA approaches, a feature data representation transformation is embedded into the DA method, in an end-to-end way..

Deep DA methods can be further divided in:

- *Common Space (CS)*: Source and Target are projected in a new shared space;

- *Common+Specific Spaces (css)*: Source and Target are first projected in a unique shared space, then, for each available domain, a projection into an ad-hoc space is used.

- CS methods:

As CS methods, [110] (Deep Domain Confusion, DDC) proposed two identical neural networks trained together, the former classifying data from the Source domain, the latter adapting the distance between Source and Target domains using features of Target data. A combination of both the classification performance and the MMD is used as final loss to minimize.
[111] uses DDC for cross-subject EEG emotion recognition. The networks' architectures used are of type residual CNNs [112]. To be fed to CNNs, the EEG inputs are firstly transformed into Electrode-frequency Distribution Maps (EFDMs, [113]). The proposed results are validated with a LOSO approach.

The authors of [114] proposed a DA framework exploiting characteristics of a standard Convolutional Neural Network (CNN), usually composed by a sequence of convolutional layers ended by a fully-connected ones. The start hypothesis is that in a DNN the transition from general to the particular task features grows with the increasing of the network depth. Indeed, in a CNN, while the initial convolutional layers learn general features, the final fully-connected ones learn domain specific features that are not transferable.

Their proposed model (Deep Adaptation Network, DAN) deeply adapt the final fully connected layers minimizing the Multi-Kernel Maximum Mean Discrepancies (MK-MMD, [115]), a multiple kernel variant of MMD used as distribution discrepancy measurement. DAN was evaluated in EEG emotion recognition on SEED and SEED-IV in [116]. In [117] the proposed Multi-Spatial Domain Adaptation Network (MSDAN) aligns source and target domains considering the spatial relationships between the electrodes. This is done by using Graph Convolutional Layers and exploiting MMD distance in the resulting graph space. Differently from other works, [117] uses data acquired in a Virtual Reality (VR) environment to generate stimuli, and the cross-device problem is taken into account.

One of the most used deep DA strategies is the Domain Adversarial Learning, proposed in [59, 118, 119]. The authors proposed an embedded problem formulation considering both the desired task and the Source-Target discrepancy. The basic idea is to make the data distributions indistinguishable for an ad-hoc domain classifier. This can be obtained by a deep neural network model (Domain Adversarial Neural Network, DANN) that, for each input, predicts both the corresponding class and the belonging domain. In a nutshell, DANN is composed of three main components: a feature extractor, a label predictor, and a domain classifier. Therefore, a learning process searches for a feature mapping maximizing the class prediction performances and, at the same time, also maximizing the domain classification loss to make the feature distributions as similar as possible. DANN is evaluated in EEG emotion recognition task in [120] on SEED. In [121] BiDANN, a variation of the original DANN, is adopted for EEG emotion recognition, but considering the differences between the brain hemispheres. More in detail, EEG data from the two hemispheres are processed separately: two different features mapping, together with a domain discriminator, are learned for the brain hemispheres, instead of just one feature mapping as in the original DANN formulation. Difference between the hemispheres in a DA approach is not dealt only by BiDANN; for instance, BiHDM [122, 123] uses two different RNN to encode the data belonging to the two hemispheres, and a domain discriminator is used to mix up the features of the Source and the Target domain. In [124] the authors propose a new DA method which is framed in the context of deep adversarial learning approaches. In particular a temporal convolutional network is used as encoder. Interestingly, the method is successfully evaluated in both cross-subject and cross-dataset. In [125, 126] domain adversarial approaches are used together with Graph Neural Networks (GNN, [87]) as feature extractor. In particular, [125] leverages on an attention mechanism [127] to lead the learning process to focus on the more tricky areas of the feature space. Performances are evaluated on SEED dataset. Instead, [126] proposed a Node-wise Domain Adversarial Training (NodeDAT) method to regularize the learning of a GNN for better subject-independent performances. In EEG literature, domain adversarial learning addand attentional mechanisms are widely used in several other studies for EEG data recognition, for example in [128, 129, 130, 131, 132, 133]. In particular, in [130] possible differences between several brain regions are also taken into account with a proposed attention module. In [134] (ATtention-based LSTM with Domain Discriminator, ATDD-LSTM) a domain discriminator in terms of LSTMs is presented to reduce the discrepance between the distributions. An attention-based encoder-decoder focuses on emotion-related input data, helping the final classification probability estimation.

An interesting adversarial approach was also investigated in [135]. The proposed work exploits the Covariance Matrices between EEG data and Riemannian distances [136]. The work proposed a new kind of Neural Network (daSPDnet) able to retain the intrinsic geometry information of the data. However, a little set of labeled data belonging to the Target domain are required during the training process, resulting as semi-supervised DA method. A similar approach, also requiring a few of labeled target data, was proposed in [137].

In [128] Adversarial Discriminative Domain Adaptation (ADDA), a strategy to tackle the DA on an image classification task, was proposed. Differently from DANN, the ADDA basic idea consisted in building two different functions for the Source and the Target domains, represented with two different encoders E_S and E_T, respectively. E_S is trained together with a classifier C using labeled data from the Source domain. Then, through an adversarial learning procedure, E_T is trained to map the Target domain data in the same space of E_S outputs. Target data can now be classified by C, conseguently. A similar idea was adapted in EEG Emotion Recognition tasks in [138] (Wasserstein GAN Domain Adaptation, WGANDA). More in detail, two generators, the former for the Source and the latter for the Target domain, are pre-trained to output two feature vectors of the same size. These vectors are assumed to belong to the same feature space. Then, an adversarial training step based on minimizing the Wasserstein distance tunes the parameters of the generators such that the outputs match more closely as possible between them. The combined outputs are then used as input for a final classifier.

Inspired by the MMD optimization made in [105], in [129](TDANN) a two stage DA method is proposed. In the first stage, MMD is minimized training a 2D CNN equipped with adaBN [139]. To be fed to the 2D CNN and to preserve spatial information, the EEG input signals are transformed into images [140, 141]. In the second stage, a domain discriminator is used to further reduce the distance between the source and the target distributions. The method was evaluated in a LOSO cross validation framework. One of the main issue of the DANN networks is that only the feature data without any label is considered during the adversarial learning process. This type of DA methods can overlap the distributions of Source and Target domains by reducing the distance between them without any consideration on the belonging classes, resulting in a simple mixing of the samples of the two domains, leading the categories within each domain to not be distinguishable. Indeed, in DANN the decision boundary inside each domain is ignored. Differently, in [142] (Maximum Classifier Discrepancy, MCD), also the labels of the Source domain data are considered, helping to build a good task-specific decision boundaries between the classes. In particular, MCD exploits different classifiers fed with the same inputs and evaluating the discrepancy. More in detail, two classifier C_1 and C_2 having the same structure are fed with the output of a feature generator G. G can be fed with data x coming from the source or the target domain. The output of C_1 and C_2 are the labels of the input x transformed by G. Before the training step, C_1 and C_2 start from different initial states, rising two different classifiers after the training. How much the two classifiers disagree on their predictions on the same input is defined *discrepancy* by the authors. Indeed, the generator G is trained to minimize the discrepancy (that is, project source and target data in the same space), while C_1 and C_2 are trained to maximize the discrepancy (so that the two classification boundaries are far from each other). The learned generator G will be able to relocate the target domain data in the source space, but considering its most probable belonging class. Task-Specific Domain Adversarial Neural Network (T-DANN, [143]) is an MCD similar model proposed for EEG Emotion Recognition. T-DANN adapts the conditional distribution between domains and, at the same time, adapts classification boundaries between classes exploiting MCD in conjunction with a domain discriminator.

From a different point of view, [144] exploits a few-shot learning approach together with an attention mechanism to deal with the excessive alignment problem. Few-shot learning-based approaches are also used in [145] where Siamese Networks [146] are used to evaluate the similarity between samples belonging to different domains. Siamese networks were originally proposed to determine whether two different inputs belong to the same class or not. In [145] the Siamese framework is enhanced to handle different domains. However, this method require a few of labeled data belonging to the Target domain.

In [147] the authors propose a DA approach for EEG-based emotion recognition based on a Multi-source co-Adaptation framework by mining diverse Correlation Information (MACI). Notably, MACI consider each subject as belonging to a different domain. The proposed method is compared with several both standard (shallow) DA approaches and CNN-based (deep) DA approaches. Cross-subjects and cross-datasets evaluations are performed. The authors of [148] propose a novel approach which attempts to unify in an unique optimization problem two standard DA approaches, that are instance reweighting and feature matching. This novel approach is named Progressive Low-Rank Subspace Alignment (PLRSA). In particular, instance reweighting is implemented by minimizing the Maximum Mean Discrepancy (MMD) distance with TrAdaBoost algorithm, and feature matching by Transfer Component Analysis (TCA). Importantly, a tiny amount of labeled target data is used. The proposed method is evaluated in a both cross-subjects and cross-sessions scenario. The method is compared with five state-of-the-art DA methods. The results seem promising, however the time complexity is a little more expensive than related state-of-the-art methods.

In [149], Neighborhood Component Analysis (NCA, [150]) is employed to learn the Mahalanobis distance between data. Therefore, data are linearly projected into a subspace such that the classification accuracy is maximized and the dimensionality of the EEG features is reduced. The obtained features are then used with Geodesic flow kernel for Unsupervised Domain Adaptation [151]. In another direction goes Multi-source Domain Transfer Discriminative Dictionary Learning modeling (MDTDDL) [152]. In this case, dictionary learning are used to learn a joint subspace between Source and Target domains [153]. DEEP and SEED are evaluated both in Cross-Subject and Cross-Session mode.

- css methods:

Although several studies start from the hypothesis that a single common feature space is enough for DA, *Common+Specific Space (CSS)* methods go

in different direction, assuming that a single shared classifier built in a shared space still has poor performance with sessions/subjects never seen. Notably, in these studies each subject/session available is considered as a single domain, and not as a whole. Hypothetically, EEG data representations can be splitted into emotional components shared among all the subjects, and private components, specific to each subject.
Leveraging on this hypothesis, [23] builds a shared encoder and private encoders for each source subject data, with the aim to capture the subject-invariant emotional representations and private components, respectively. The learned encoders are then used to build several emotion classifiers. Finally, a classifier for a new subject is built. The parameters of these classifiers are learned exploiting the shared encoder. A fusion strategy between the classifiers' outputs is then applied to obtain the final classification result. However, the proposed framework requires few labeled target data, falling in the semisupervised DA category. Multi-source EEG-based Emotion Recognition Network (MEERNet) [154] proposed a different classifier for each different domain (subject or session), preceded by a feature extractor shared by all the domains. Final classification is made averaging between domain-specific classifiers. Similarly, [155] proposed a framework composed of a common feature extractor to map all the domains in a common subspace, a main task classifier or regressor, and private discriminators for each domain. The training is made reducing the Wasserstein distance between the marginal distribution of each source domain and target one in an adversarial way. In [156] the authors proposed a Multi Source-Marginal Distribution Adaptation (MS-MDA) algorithm for EEG emotion recognition. Also in this case, the key idea is that the final response is obtained aggregating the responses of different target-source specific classifiers, preceded by a common feature extractor. Notably, the authors explore the impact of different types of data normalization on the performance of the proposed model. MS-MDA is also compared with several standard DA methods. Similarly, the authors of [157] proposed Multi-Source and Multi-Representation Adaptation (MSMRA), an approach with many similarities with MS-MDA. Both cross-subjects and cross-sessions evaluations are performed.

Supervised DA (PreTraining) methods

In the supervised DA category, four studies were included in the review. In [158] a pretrained version of InceptionResnetV2 [159] is used as feature extractor for EEG data. The classification is made by a final net-

work layer added to the InceptionResnetV2 network. Instead, [160] exploited DenseNet121 [161] as pre-trained model to build a new architecture fed with EEG data transformed into spectrogram images.
In [113] a CNN trained on different subjects and sessions of the SEED dataset is then re-trained on a small amount of data acquired from a subject belonging to the DEAP dataset. This was made to evaluate the cross-dataset emotion recognition performances. In [162] several classifiers trained on different data belonging to different subjects and sessions are ensembled together obtaining a final classifier suitable both for cross-sessions and cross-subjects EEG emotion recognition.

IRS & Hybrid

IRS methods take into account that not all the training data can effectively be useful for the target space. Indeed, a part of the data can lead toward bad performance, therefore it can be better to remove them or to reduce their weights in the training stage. In [163] TrAdaBoost [164], a semi-supervised DA method acting the instances' weights, is used to score the source EEG data in order to avoid possible negative influence of the data during the training process. In a nutshell, a small amount of labeled target data available during the training stage helps to vote on the usefulness of each of the available source data instance. As initial step, only the source subjects data closest to the target one are selected. Similarity between subjects is computed according to the MMD similarity and fed to TrAdaBoost as auxiliary data.

In the revised literature, IRS methods are often used as an initial step of other DA methods. For instance, in [165, 166] the similarity between source and target EEG data is measured using the Pearson Correlation Coefficient and the Average Frechet Distance, respectively. In particular, in [166] EEG source data closer to the available target data are projected to a new space through TCA, together with the target one. Finally, the classification step is made by an Echo State Network (ESN, [167]).

In [168] (DMATN) source data belonging to the existing subjects are divided into several subdomains. Then, a set of subdomains is chosen as the most relevant ones for the target data. The proposed architecture combines together DAN and DANN to learn representation that are domains invariant.

DG methods

Differently from classical DA, DG assumes that data from several domains are available, but no data from the target domain is observed during the training stage. Differently from classical domain adaptation methods, data from several domains are available, but no data from the test domain is observed during the training stage [169]. DG methods can be divided as:

A) *shallow DG*: a data transformation is given a priori;

B) *deep DG*: the data rapresentation is learned as part of the DG strategy.

A) Shallow DG methods

Shallow DG methods share the same principles of shallow DA ones, building a shared space between domains, letting the input data representation unchanged. Domain Invariant Component Analysis (DICA) [169] searches for common features across several domains. Features data are transformed by a learned orthogonal transformation able to minimize the dissimilarity between a set of known domains and preserving, at the same time, the relations between data features and their real labels. The authors also provided an unsupervised DICA version which did not take care of the class labels.

In [170] Scatter Component Analysis (SCA) is proposed. The aim of the authors is to propose a method adapt both for DG and DA requirements. SCA searches for a data transformation where, at the same time, i) the source and the target domains are similar, ii) elements of the same class are similar, iii) elements of different classes are well separated, and iv) the variance of the whole data is maximized. This is made introducing *Scatter*, a measure closely related to MMD. In [171], SCA and DICA are applied and evaluated on SEED dataset.

B) Deep DG methods

On the other side, *deep DG* methods embed the data representation as part of the generalization strategy. In [172] data from similar subjects are used to train the same classifier. The similarity is computed through a clustering algorithm. This subset of similar subjects is used to train a final CNN classifier. Notably, in [173] a similar strategy is adopted, but for DA context.

[174] joined together BiDANN and Variational Autoencoder (VAE), obtaining a subject-invariant Bi-lateral Variational Domain Adversarial Neural Network (BiVDANN). VAEs are generative neural networks able to learn embedding of data in a latent space. As any classical autoencoder, a VAE is

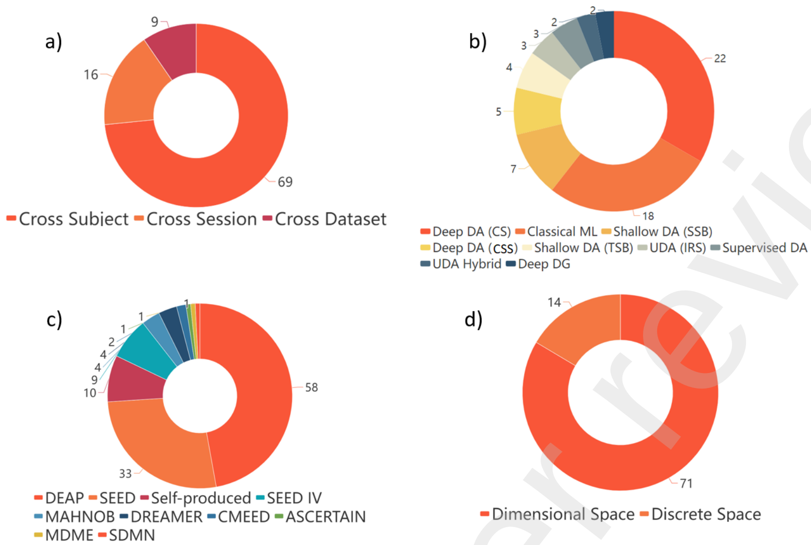

Figure 5: Pie charts for distribution of papers occurrences according to: a) generalization types, b) categories of the taxonomy, c) used datasets, d) emotional theories.

composed of an encoder network able to project data to an embedding space, and an decoder network able to reconstruct the original input from the embedding. In the proposed work, the learned features are further refined by domain adversarial training made across different subjects, with the aim to learn subject-independent features. Furthermore, to maximize cross-dataset performance, spectral topography data of the EEG signal are used as input.

The pie charts in Figure 5 show some statistics about the papers included in the survey. First of all, it is evident that almost three quarters of the studies surveyed (73.4 %) focus on a cross-subject mode of generalization, while cross-session studies account for only 17 % and only 9 % operate a cross-dataset mode of generalization. Graph b) shows the percentage distribution according to the proposed taxonomy. Looking at this graph, it is evident that the majority of generalization studies are moving towards the use of Deep DA (CS) (29.33 %), at the expense of more traditional approaches, which still retain 24 %. This is followed by Shallow DA (SSB) approaches (9.33 %), Deap DA (CSS) (6.67 %), Shallow DA (TSB) (5.33 %), Supervised DA and UDA (IRS) (4 %), and finally Deap DG and Hybrid DA (2.67 %).
The pie chart in Figure 5.c) shows the number of times each EEG dataset

is exploited in the reviewed literature. The mainly used datasets are SEED (45.8 %) and DEAP (27.5 %). This is followed by 8.3 % of studies that propose their own self-produced dataset. Among the other datasets available in the literature, only SEED IV stands out (at 7.5 %), which is interesting in that it adopts a discrete space of four emotions for classification (happy, sad, fear and neutral). Instead, each of the other datasets do not exceed 5 % (MAHNOB [175] and DREAMER [43] (3.33 %), CMEED [176] (1.7 %), ASCERTAIN [177], MDME [178] and SDMN [179] (0.8 %)).

Finally, Figure 5.d) offers an interesting statistic about the interest of the authors of the studies examined in the various perspectives of emotion representation. As already mentioned in section 1.3, the two dominant perspectives, and the only ones considered in the literature examined, are those based on categorical and dimensional models. More than 80 per cent of the works are based on a representation of emotions, and then their subsequent classification in terms of valence and arousal (and only in one case also dominance [135]).

In Tab. 1 all papers included in the review were reported, indicating if belonging to the proposed taxonomy or to classical ML methods. Moreover, for each research study as the type of generalization (cross-subject, cross-session, or cross-device), the EEG dataset, the adopted classifier (whether proposed as a personal contribution or adopted from the literature), and the validation strategy is reported. Studies in which the description of the experimental setup was not sufficiently clear, especially in terms of validation strategies, were voluntarily omitted from the table for reproducibility issues. The best performer solutions in terms of mean classification accuracy are proposed in Table 2. Only cross-subject studies were considered, as most of the studies examined. Ten classification issues were focused on. Each issue is defined by considering the number and type of classes (binary and ternary on valence and arousal, quaternary on the two-dimensional valence-arousal plane, binary and quaternary on discrete dimensions) and the adopted dataset (DEAP, SEED, other). For each issue, the best performer study in terms of accuracy was identified. Only studies reporting both mean accuracy and standard deviation were included in the performance assessment.

6. Discussion

To date, no robust electroencephalographic patterns are recognized in scientific literature for correlating with emotional states. Some studies base their results on the asymmetry of scalp activations. In general, many theories still coexist and are not statistically well founded (validated on small experimental samples) [180, 181, 182].

When aiming for a generalization goal in EEG-based Emotion Recognition, Transfer Learning methods are becoming more and more established in the literature. Domain Adaptation methods (Deep DA (CSB), Shallow DA (SSB), IRS DA, Deep DA (CSS), Shallow DA (TSB) and Supervised DA) exceed 60 % of the total surveyed studies and exhibit very high accuracy performances in the table of best performers (see Table 2). In particular, Deep DA (Common Space) is used by five best performers studies. This could be also due to the current massive use of Deep DA (Common Space) in the literature. Indeed, one third of the surveyed studies (Figure 5.d) belongs to this category.

However, a still substantial percentage of works (22.67 %) belongs to the Classical ML category. An emblematic case in this context is [85], namely the best performer in the classification issue on SEED IV with four discrete classes. This is an interesting study based on a self-organized graph construction module. This solution can be considered as a peculiar implementation of the well established adaptive filters strategy, when the generalization goal is pursued by customizing the network to the current input. Conversely, the DA strategies make the data belonging to different domains more homogeneous by means of appropriate transformations. The different impact between DA and adaptive filters approaches can be better appreciated by making a comparison between the previous study and [126]. Both studies address the problem of four-class classification on the same dataset by using a pipeline based on graphs and deep networks. In the first case, an adaptive graph is used without any DA methods, while the second study makes use of a (nonadaptive) graph approach in combination with DA techniques. Even though they use different approaches, the reported accuracy performances are comparable. This suggests how the dynamic search for feature extraction procedures represents an interesting frontier for future studies in this area, not excluding the potential of using this approach in combination with DA/DG techniques.

Several issues should be taken into account considering TL methods.

Firstly, UDA methods require the availability of unlabeled target data in the training phase. In the emotion recognition problem, this implies that EEG recordings belonging to the target session/subject are provided. This availability is not granted, especially in online applications. In alternative, initial calibration data can be acquired from the target subject, but this requires additional efforts both for the subjects and the operators.

Secondly, a large part of DA strategies uses the source data as they all belong to the same domain. This assumption can be too strong, especially if the source data are acquired in different sessions or, worse, from different subjects. This point is taken into account by multi-source DA approaches, such as Transductive Parameter Transfer (TPT) [183], or by multi-source DA approaches specific for EEG data, such as MSSA or MEERNet, where different source subjects/sessions are considered as different source domains, and by DG methods.

Thirdly, in several DA strategies labels of both the source and target domains are not considered in the alignment of the domains. This type of DA methods can lead to overlapping distributions of source and target domains, without any consideration on the belonging classes. As a consequence, the class separability can worsen. This can be a weakness of some Shallow DA methods, such as unsupervised TCA.

Specific weakness can be highlighted for each DA category of the proposed taxonomy.

Shallow DA methods require data projections between different spaces by means of handcrafted transformations. Therefore, the adopted transformations may not be suitable for the available data. Moreover, among proposed data transformations, some of them require all the data be processed together, with a large amount of memory needed.

On the other side, Deep DA methods require a greater number of parameters with respect to a shallow method. This can result in high computational complexity, such as in MACI and Adversarial learning-based methods. This can lead toward overfitting and the curse of dimensionality problems [104] if not enough data are available. However, despite their high computational load, Deep DA methods exhibit the best performance in terms of accuracy at least in four of the cases considered, as can be seen in Tab. 2.

Instance Reweighting and Selection methods select or score data to manage uncorrelation between source and target data. Therefore, a part of the data may not fully used in the training stage, making these methods strongly dependent on the score/selection function adopted. Moreover, their compu-

tational cost can be not negligible, since they re-weight the available data according to their similarity.
Supervised DA methods require models trained on data belonging to the source domain. In several tasks (such as image classification) several pre-trained model on big amount of data are freely and publicly available to the user (for example ResNet models trained on ImageNet [112]). However, it is harder to find similar models trained on EEG data for Emotion Recognition task. This can be due to the scarcity of publicly available large dataset. Moreover, also by collecting together several public dataset, the resulting model may have low performance. Indeed, being the EEG is a highly non-stationary signal, it results very susceptible to different experimental conditions. In contrast to DA approaches, the aim of Domain Generalization (DG) is to generalize over several domains. Therefore, data belonging to the target may not be required in the training stage. Instead, data from several other domains (i.e. sessions/subjects) are required. This may be impractical especially in experimental scenarios, due to the difficulties of enrolling subjects and the time required to conduct multiple acquisitions. Thus, DA approaches are currently the most proposed methods in Emotion Recognition scenarios involving TL.
Originally, the well-known TL methods were developed outside the Emotion Recognition framework. In recent years, several studies have exploited these methods in emotion recognition but focusing only on their effectiveness, without conducting an in-depth analysis of the specific contributions made by TL methods in this field. In fact, several DA pipelines consist of several steps, and a comparison of performance with and without TL methods does not identify the most effective steps. For example, in [184] it is shown that in several cases the data normalization adopted as first step of several TL pipelines has a stronger impact on the model generalization respect to the TL methods themselves. Other studies [77, 78, 79] have confirmed that simple data normalization with low computational and spatial efforts allows for interesting results in EEG-based Emotion Recognition, in some cases comparable or better than several current DA/DG approaches. In general, the merits of TL techniques are not in question, but in the future, a more in-depth analysis by scientists is needed, for example by accompanying their proposals with ablation studies.

Another point to take into account is that the proliferation of EEG acquisition devices on the market is not always coupled with consistency in terms of quality between the various devices (considering electrode type and positioning, interference shielding, and signal-to-noise ratio, amplification strategies, etc). A comparison among different studies must take into account the quality of EEG instrumentation used. The IEC 60601-2-26 standard applies to basic safety and essential performance of electroencephalographs used in a clinical environment. Among the requirements, the minimum overall signal quality for an electroencephalographic device to be considered acceptable is defined [185]. Even if IEC 60601-2-26 is a standard specifically developed for clinical purposes, it is nowadays the only available standard for EEG instrumentation quality certification. In the future, it is desirable for research to be increasingly based on certified instruments. However, an encouraging trend emerges from the most recent public datasets. Indeed, they are all based on standardized equipment: (i) Neuroelectrics Enobio 8 in the case of LUMED [158], (ii) NuAmp Neuroscan in the case of CMEED [134], and (iii) gtec.HIamp in the case of the dataset produced by [129]).

A further concern in the use of public datasets is its underlying theoretical background, often uncritically accepted by the scientists. Many studies validate the same machine learning algorithm on different datasets although the targeted psychic phenomena are radically different. Indeed, each dataset leverages on a specific theory of emotions and related experimental setup of emotion elicitation. For instance, DEAP is based on a dimensional approach and SEED IV on discrete one.

Finally, at present, the available public datasets do not adopt an established practice of psychological screening of the subjects involved. In general, studies on EEG-based emotion assessment could benefit from administering psychometric questionnaires to participants. Indeed, psychological data could help to understand individual differences in emotional response, leading to clustering of subjects [173]. Recently, unsupervised clustering based on large datasets is emerging as a promising strategy for empirical identification of personality types [186]. Meanwhile, correlations have been found between personality types and EEG patterns [187]. Moreover, prior psychological assessments allow to manage bias due to individual traits or states. The introduction of psycho-metric tests and assessments during the production of upcoming datasets could lead to a much more fruitful use of data in support of generalization.

7. Conclusion

In this work, a systematic literature review collecting papers on machine learning strategies to pursue (cross-subjects and cross-sessions) generalizability in EEG-based emotion recognition was carried out. Among the 418 articles retrieved from Scopus, IEEE (Institute of Electrical and Electronics Engineers) Xplore, and PubMed databases, 75 papers resulted eligible. Furthermore, the studies with the best results in terms of average classification accuracy were identified, and the ten best results considering as many classification problems were highlighted.

Most of the analyzed works adopted Classical ML or TL approaches to deal with the generalization problem. In particular, TL methods received a considerable attention from the scientific community, as their basic framework is particularly suited to the EEG Emotion Recognition generalization problem. In spite of their limitations (i.e., the need for target data during the training stage), today DA methods result to be particularly encouraging to handle the EEG Emotion Recognition generalization problem. DG methods aim to achieve a more generalized approach compared to DA, which relies on target data availability. However, due to the challenging nature of this approach, current DG methods generally have lower performance than DA approaches in the task of Emotion Recognition. Finally, works relying on simple ML methods combined with proper normalization strategies lead to interesting results with a low computational load. This can be due to the ability of some simple transformations to project data into spaces where shared characteristics between the domains are emphasized.

An interesting perspective based on self-organized graph construction modules emerged as peculiar strategy. This suggests how the adaptive feature extraction procedures represent an interesting frontier for future studies in this area, not excluding the potential of using this approach in combination with DA/DG techniques.

Future research on EEG-based emotion assessment could also benefit from administering psychometric questionnaires to participants in order to conduct a psychological screening of the experimental sample. This could help to understand individual differences in emotional responses, leading to clustering of subjects also taking into account the different subjects' personality.

8. Acknowledgement

This work was carried out as part of the “INTENSE” project, which was financially supported by the Italian Ministry of Enterprise and Made in Italy (Mimit). The work was also supported by European Union - FSE-REACT-EU, PON Research and Innovation 2014-2020 DM1062/2021 contract number 18-I-15350-2 and by the Ministry of University and Research, PRIN research project ”BRIO – BIAS, RISK, OPACITY in AI: design, verification and development of Trustworthy AI.”, Project no. 2020SSKZ7R. The authors also thank the PhD Grant ”AR4ClinicSur- Augmented Reality for Clinical Surgery” (INPS - National Social Security Institution – Italy). The authors would like to thank the editor and the anonymous reviewers for their stimulating and precious suggestions and comments to the review of this paper.

Acronyms

A-DNN - Adversarial Deep Neural Network
AD-TCN - Adversarial Discriminative Temporal Convolutional Network
ASFM - Adaptive Subspace Feature Matching
ASI - Add-Session-In
ATDD-LSTM - Attention-based LSTM
BiDANN - Bi-hemispheres DANN
BiHDM - Bi-Hemispheric Discrepancy Model
BiLSTM - Bidirectional LSTM
BiVDANN - Bi-lateral Variational Domain Adversarial Neural Network
CS - Common Space-based
DAN - Deep Adaptation Network
DANN - Domain Adversarial Neural Network
DASC - Domain Adaptation Subject Clustering
DASRC - Domain Adaptation Sparse Representation Classifier
DDC - Deep Domain Confusion
DECNN - Dynamic Empirical Convolutional Neural network
DGCNN - Dynamical Graph Convolutional Neural Networks
DG-DANN - Domain Generalization DANN
DResNet - Domain Residual Network
ESN - Echo State Network
GNB - Gaussian Naïve Bayes

HO - Hold Out
IRS - Instance Reweighting and Selection
LOO - Leave One Out
LSTM - Long short-term memory
MACI - Multi-Source Co-adaptation Correlation Information
MDTDDL - Multi-source Domain Transfer Discriminative Dictionary Learning modelling
MEERNet - Multi-Source EEG-based Emotion Recognition Network
MIDA - Maximum Independence Domain Adaptation
MSDAN - Multi-Spatial Domain Adaptation Network
MS-MDA - Multi Source-Marginal Distribution Adaptation
MSSA - Multi-Subject Subspace Alignment
Na - not available
NCA = Neighborhood Component Analysis
O2OSE - ONE-TO-ONE-SESSION
PLRSA - Progressive Low-Rank Subspace Alignment
PPDA - Plug-and-Play Domain Adaptation
R2G-STNN - Regional To Global Spatial-Temporal Neural Network
RCNN - Residual CNN
RF - Random Forest
RFE - Recursive Feature Elimination
RGNN - Regularized Graph Neural Network
RPCA - Robust Principal Component Analysis
SAAE - Subspace Alignment Auto Encoder
SBS - Sequential Backward Selection
SDA-FSL - Single-Source Domain Adaptive Few-Shot Learning Network
SE2SE - session-to-session
SOGNN - Self-Organized Graph Neural Network
sp - self-produced
css - Specific+Common Space-based
SSB - Shared Space-Based
STM - Style Transfer Mapping
SU2SU - subject-to-subject
SVM - Support Vector Machine
TCA - Transfer Component Analysis
TDANN - Two-Level Domain Adaptation Neural Network
TPT - Transductive Parameter Transfer
TRFE - Transferable Recursive Feature Elimination

TSB - Target Space-Based
UDA - Unsupervised/Semi-supervised Domain Adaptation
VAE - Variational Auto Encoder
WGANDA - Wasserstein Generative Adversarial Network Domain Adaptation
wMADA - Wasserstein-Distance-based Multi-Source Adversarial Domain Adaptation

Funding

References

[1] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, Affective computing and sentiment analysis, in: A practical guide to sentiment analysis, Springer, 2017, pp. 1–10.

[2] S. Saganowski, A. Dutkowiak, A. Dziadek, M. Dzieżyc, J. Komoszyńska, W. Michalska, A. Polak, M. Ujma, P. Kazienko, Emotion recognition using wearables: A systematic literature review-work-in-progress, in: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2020, pp. 1–6.

[3] H. Feng, H. M. Golshan, M. H. Mahoor, A wavelet-based approach to emotion classification using eda signals, Expert Systems with Applications 112 (2018) 77–86.

[4] M. Healy, R. Donovan, P. Walsh, H. Zheng, A machine learning emotion detection platform to support affective well being, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 2694–2700.

[5] S. Saganowski, Bringing emotion recognition out of the lab into real life: Recent advances in sensors and machine learning, Electronics 11 (3) (2022) 496.

[6] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, S. G. Yoo, Eeg-based bci emotion recognition: A survey, Sensors 20 (18) (2020) 5083.

[7] M. Choi, G. Koo, M. Seo, S. W. Kim, Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness, IEEE Transactions on Instrumentation and Measurement 67 (3) (2017) 634–645.

[8] S. Mühlbacher-Karrer, A. H. Mosa, L.-M. Faller, M. Ali, R. Hamid, H. Zangl, K. Kyamakya, A driver state detection system—combining a capacitive hand detection sensor with physiological sensors, IEEE Transactions on Instrumentation and Measurement 66 (4) (2017) 624–636.

[9] N. Millard, L. Hole, In the moodie: Using 'affective widgets' to help contact centre advisors fight stress, in: Affect and Emotion in Human-Computer Interaction, Springer, 2008, pp. 186–193.

[10] E. Gurgu, I.-A. Gurgu, R. B. M. Tonis, Neuromarketing for a better understanding of consumer needs and emotions, Independent Journal of Management & Production 11 (1) (2020) 208–235.

[11] M. Aldayel, M. Ykhlef, A. Al-Nafjan, Deep learning for eeg-based preference classification in neuromarketing, Applied Sciences 10 (4) (2020) 1525.

[12] T. Kolar, Z. Batagelj, I. Omeragić, M. Husić-Mehmedović, How moment-to-moment eeg measures enhance ad effectiveness evaluation: Peak emotions during branding moments as key indicators, Journal of Advertising Research 61 (4) (2021) 365–381.

[13] C. Jones, J. Sutherland, Acoustic emotion recognition for affective computer gaming, in: Affect and emotion in human-computer interaction, Springer, 2008, pp. 209–219.

[14] Z. Liang, M. A. Chapa Martell, Validity of consumer activity wristbands and wearable eeg for measuring overall sleep parameters and sleep structure in free-living conditions, Journal of Healthcare Informatics Research 2 (1) (2018) 152–178.

[15] D. Sopic, A. Aminifar, D. Atienza, e-glass: A wearable system for real-time detection of epileptic seizures, in: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2018, pp. 1–5.

[16] M. Taj-Eldin, C. Ryan, B. O'Flynn, P. Galvin, A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers, Sensors 18 (12) (2018) 4271.

[17] R. Francese, X. Yang, Supporting autism spectrum disorder screening and intervention with machine learning and wearables: a systematic literature review, Complex & Intelligent Systems (2021) 1–16.

[18] T. Inouye, S. Toi, Y. Matsumoto, A new segmentation method of electroencephalograms by use of akaike's information criterion, Cognitive brain research 3 (1) (1995) 33–40.

[19] C.-H. Im, Computational eeg analysis, Springer Singapore, Singapore 10 (2018) 978–981.

[20] L. Sörnmo, P. Laguna, Bioelectrical signal processing in cardiac and neurological applications, Vol. 8, Academic press, 2005.

[21] C. Cao, S. Slobounov, Application of a novel measure of eeg non-stationarity as 'shannon-entropy of the peak frequency shifting'for detecting residual abnormalities in concussed individuals, Clinical Neurophysiology 122 (7) (2011) 1314–1321.

[22] S. Nasiri, G. D. Clifford, Attentive adversarial network for large-scale sleep staging, in: Machine Learning for Healthcare Conference, PMLR, 2020, pp. 457–478.

[23] L.-M. Zhao, X. Yan, B.-L. Lu, Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, sn, 2021.

[24] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, D. Moher, The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, Journal of clinical epidemiology 62 (10) (2009) e1–e34.

[25] S. M. Alarcao, M. J. Fonseca, Emotions recognition using eeg signals: A survey, IEEE Transactions on Affective Computing 10 (3) (2017) 374–393.

[26] A.-M. Brouwer, T. O. Zander, J. B. Van Erp, J. E. Korteling, A. W. Bronkhorst, Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls, Frontiers in neuroscience 9 (2015) 136.

[27] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update, Journal of neural engineering 15 (3) (2018) 031005.

[28] D. Wu, Y. Xu, B.-L. Lu, Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016, IEEE Transactions on Cognitive and Developmental Systems 14 (1) (2020) 4–19.

[29] N. S. Suhaimi, J. Mountstephens, J. Teo, Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities, Computational intelligence and neuroscience 2020 (2020).

[30] W. Li, W. Huan, B. Hou, Y. Tian, Z. Zhang, A. Song, Can emotion be transferred?–a review on transfer learning for eeg-based emotion recognition, IEEE Transactions on Cognitive and Developmental Systems (2021).

[31] W. D. TenHouten, From primary emotions to the spectrum of affect: An evolutionary neurosociology of the emotions, in: Neuroscience and social science, Springer, 2017, pp. 141–167.

[32] P. Ekman, Basic emotions, Handbook of cognition and emotion 98 (45-60) (1999) 16.

[33] J. Posner, J. A. Russell, A. Gerber, D. Gorman, T. Colibazzi, S. Yu, Z. Wang, A. Kangarlu, H. Zhu, B. S. Peterson, The neurophysiological bases of emotion: An fmri study of the affective circumplex using emotion-denoting words, Human brain mapping 30 (3) (2009) 883–895.

[34] J. N. Acharya, A. J. Hani, J. Cheek, P. Thirumala, T. N. Tsuchida, American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature, The Neurodiagnostic Journal 56 (4) (2016) 245–252.

[35] A. J. Casson, Wearable eeg and beyond, Biomedical engineering letters 9 (1) (2019) 53–71.

[36] M. A. Lopez-Gordo, D. Sanchez-Morillo, F. P. Valle, Dry eeg electrodes, Sensors 14 (7) (2014) 12847–12870.

[37] R. Westermann, K. Spies, G. Stahl, F. W. Hesse, Relative effectiveness and validity of mood induction procedures: A meta-analysis, European Journal of social psychology 26 (4) (1996) 557–580.

[38] P. J. Lang, International affective picture system (iaps): Affective ratings of pictures and instruction manual, Technical report (2005).

[39] B. Kurdi, S. Lozano, M. R. Banaji, Introducing the open affective standardized image set (oasis), Behavior research methods 49 (2) (2017) 457–470.

[40] E. S. Dan-Glauser, K. R. Scherer, The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance, Behavior research methods 43 (2) (2011) 468.

[41] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: A database for emotion analysis; using physiological signals, IEEE transactions on affective computing 3 (1) (2011) 18–31.

[42] W.-L. Zheng, B.-L. Lu, Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks, IEEE Transactions on autonomous mental development 7 (3) (2015) 162–175.

[43] 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 22 (1) (2017) 98–107.

[44] P. Geethanjali, Y. K. Mohan, J. Sen, Time domain feature extraction and classification of eeg data for brain computer interface, in: 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2012, pp. 1136–1139.

[45] C. Vidaurre, N. Krämer, B. Blankertz, A. Schlögl, Time domain parameters as a feature for eeg-based brain–computer interfaces, Neural Networks 22 (9) (2009) 1313–1319.

[46] C. T. Yuen, W. San San, T. C. Seong, M. Rizon, Classification of human emotions from eeg signals using statistical features and neural network, International Journal of Integrated Engineering 1 (3) (2009).

[47] S.-H. Oh, Y.-R. Lee, H.-N. Kim, A novel eeg feature extraction method using hjorth parameter, International Journal of Electronics and Electrical Engineering 2 (2) (2014) 106–110.

[48] S. Phadikar, N. Sinha, R. Ghosh, A survey on feature extraction methods for eeg based emotion recognition, in: International Conference on Innovation in Modern Science and Technology, Springer, 2019, pp. 31–45.

[49] A. Patil, C. Deshmukh, A. Panat, Feature extraction of eeg for emotion recognition using hjorth features and higher order crossings, in: 2016 Conference on Advances in Signal Processing (CASP), IEEE, 2016, pp. 429–434.

[50] X. Nan, X. Jinghua, The fractal dimension of eeg as a physical measure of conscious human brain activities, Bulletin of Mathematical Biology 50 (5) (1988) 559–565.

[51] E. Kroupi, A. Yazdani, T. Ebrahimi, Eeg correlates of different emotional states elicited during watching music videos, in: International Conference on Affective Computing and Intelligent Interaction, Springer, 2011, pp. 457–466.

[52] X.-W. Wang, D. Nie, B.-L. Lu, Eeg-based emotion recognition using frequency domain features and support vector machines, in: International conference on neural information processing, Springer, 2011, pp. 734–743.

[53] Z. Zhang, Spectral and time-frequency analysis, in: EEG Signal Processing and feature extraction, Springer, 2019, pp. 89–116.

[54] D. Hernández, L. Trujillo, O. Villanueva, O. Romo-Fewell, et al., Detecting epilepsy in eeg signals using time, frequency and time-frequency

domain features, in: Computer science and engineering—theory and applications, Springer, 2018, pp. 167–182.

[55] R. Jenke, A. Peer, M. Buss, Feature extraction and selection for emotion recognition from eeg, IEEE Transactions on Affective computing 5 (3) (2014) 327–339.

[56] A. Apicella, P. Arpaia, F. Isgrò, G. Mastrati, N. Moccaldi, A survey on eeg-based solutions for emotion recognition with a low number of channels, IEEE Access (2022) 1–1doi:10.1109/ACCESS.2022.3219844.

[57] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, N. D. Lawrence, Dataset shift in machine learning, Mit Press, 2008.

[58] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (10) (2009) 1345–1359.

[59] Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, in: International conference on machine learning, PMLR, 2015, pp. 1180–1189.

[60] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele University 33 (2004) (2004) 1–26.

[61] V. Vapnik, 24 transductive inference and semi-supervised learning (2006).

[62] T. Joachims, et al., Transductive inference for text classification using support vector machines, in: Icml, Vol. 99, 1999, pp. 200–209.

[63] Y. Chen, G. Wang, S. Dong, Learning with progressive transductive support vector machine, Pattern Recognition Letters 24 (12) (2003) 1845–1855.

[64] K. Yang, G. Bao, Y. Zeng, L. Tong, J. Shu, B. Yan, Improving session-to-session transfer performance of emotion recognition using adaptive support vector machine, in: Journal of Physics: Conference Series, Vol. 1601, IOP Publishing, 2020, p. 042028.

[65] S. Rayatdoost, M. Soleymani, Cross-corpus eeg-based emotion recognition, in: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2018, pp. 1–6.

[66] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, B. Hu, Exploring eeg features in cross-subject emotion recognition, Frontiers in neuroscience 12 (2018) 162.

[67] F. Yang, X. Zhao, W. Jiang, P. Gao, G. Liu, Multi-method fusion of cross-subject emotion recognition based on high-dimensional eeg features, Frontiers in computational neuroscience 13 (2019) 53.

[68] W. Jiang, G. Liu, X. Zhao, F. Yang, Cross-subject emotion recognition with a decision tree classifier based on sequential backward selection, in: 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Vol. 1, IEEE, 2019, pp. 309–313.

[69] J. Cai, W. Chen, Z. Yin, Multiple transferable recursive feature elimination technique for emotion recognition based on eeg signals, Symmetry 11 (5) (2019) 683.

[70] Z. Yin, Y. Wang, L. Liu, W. Zhang, J. Zhang, Cross-subject eeg feature selection for emotion recognition using transfer recursive feature elimination, Frontiers in neurorobotics 11 (2017) 19.

[71] W. Zhang, Z. Yin, Eeg feature selection for emotion recognition based on cross-subject recursive feature elimination, in: 2020 39th Chinese Control Conference (CCC), IEEE, 2020, pp. 6256–6261.

[72] S. Liu, X. Wang, L. Zhao, J. Zhao, Q. Xin, S. Wang, Subject-independent emotion recognition of eeg signals based on dynamic empirical convolutional neural network, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2020).

[73] X. Li, Z. Zhao, D. Song, Y. Zhang, C. Niu, J. Zhang, J. Huo, J. Li, Variational autoencoder based latent factor decoding of multichannel eeg for emotion recognition, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019, pp. 684–687.

[74] D. P. Kingma, M. Welling, Auto-Encoding Variational Bayes, in: 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.

[75] X. Li, Z. Zhao, D. Song, Y. Zhang, J. Pan, L. Wu, J. Huo, C. Niu, D. Wang, Latent factor decoding of multi-channel eeg for emotion recognition through autoencoder-like neural networks, Frontiers in neuroscience 14 (2020) 87.

[76] P. Pandey, K. Seeja, Subject independent emotion recognition from eeg using vmd and deep learning, Journal of King Saud University-Computer and Information Sciences (2019).

[77] H. Chen, S. Sun, J. Li, R. Yu, N. Li, X. Li, B. Hu, Personal-zscore: Eliminating individual difference for eeg-based cross-subject emotion recognition, IEEE Transactions on Affective Computing (2021).

[78] J. Fernandez, N. Guttenberg, O. Witkowski, A. Pasquali, Cross-subject eeg-based emotion recognition through neural networks with stratified normalization, Frontiers in neuroscience 15 (2021) 11.

[79] M. Arevalillo-Herráez, M. Cobos, S. Roger, M. García-Pineda, Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from eeg signals, Sensors 19 (13) (2019) 2999.

[80] V. Gupta, M. D. Chopda, R. B. Pachori, Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals, IEEE Sensors Journal 19 (6) (2018) 2266–2274.

[81] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, Y. Cao, Relieff-based eeg sensor selection methods for emotion recognition, Sensors 16 (10) (2016) 1558.

[82] Y. Peng, W. Kong, F. Qin, F. Nie, J. Fang, B.-L. Lu, A. Cichocki, Self-weighted semi-supervised classification for joint eeg-based emotion recognition and affective activation patterns mining, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1–11.

[83] Z. Tian, D. Huang, S. Zhou, Z. Zhao, D. Jiang, Personality first in emotion: a deep neural network based on electroencephalogram channel attention for cross-subject emotion recognition, Royal Society open science 8 (8) (2021) 201976.

[84] T. Song, W. Zheng, P. Song, Z. Cui, Eeg emotion recognition using dynamical graph convolutional neural networks, IEEE Transactions on Affective Computing 11 (3) (2018) 532–541.

[85] J. Li, S. Li, J. Pan, F. Wang, Cross-subject eeg emotion recognition with self-organized graph neural network, Frontiers in Neuroscience (2021) 689.

[86] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S. Y. Philip, A comprehensive survey on graph neural networks, IEEE transactions on neural networks and learning systems 32 (1) (2020) 4–24.

[87] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81.

[88] B. Fernando, A. Habrard, M. Sebban, T. Tuytelaars, Unsupervised visual domain adaptation using subspace alignment, in: Proceedings of the IEEE international conference on computer vision, 2013, pp. 2960–2967.

[89] X. Chai, Q. Wang, Y. Zhao, Y. Li, D. Liu, X. Liu, O. Bai, A fast, efficient domain adaptation technique for cross-domain electroencephalography (eeg)-based emotion recognition, Sensors 17 (5) (2017) 1014.

[90] X. Chai, Q. Wang, Y. Zhao, X. Liu, D. Liu, O. Bai, Multi-subject subspace alignment for non-stationary eeg-based emotion recognition, Technology and Health Care 26 (S1) (2018) 327–335.

[91] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma, Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, Advances in neural information processing systems 22 (2009).

[92] Y.-P. Lin, Constructing a personalized cross-day eeg-based emotion-classification model using transfer learning, IEEE journal of biomedical and health informatics 24 (5) (2019) 1255–1264.

[93] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, H. He, Multisource transfer learning for cross-subject eeg emotion recognition, IEEE transactions on cybernetics 50 (7) (2019) 3281–3293.

[94] X.-Y. Zhang, C.-L. Liu, Writer adaptation with style transfer mapping, IEEE transactions on pattern analysis and machine intelligence 35 (7) (2012) 1773–1787.

[95] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola, A kernel method for the two-sample-problem, Advances in neural information processing systems 19 (2006).

[96] S. J. Pan, I. W. Tsang, J. T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE transactions on neural networks 22 (2) (2010) 199–210.

[97] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, B.-L. Lu, Transfer components between subjects for eeg-based emotion recognition, in: 2015 international conference on affective computing and intelligent interaction (ACII), IEEE, 2015, pp. 917–922.

[98] B. Xue, Z. Lv, J. Xue, Feature transfer learning in eeg-based emotion recognition, in: 2020 Chinese Automation Congress (CAC), IEEE, 2020, pp. 3608–3611.

[99] Z. He, N. Zhuang, G. Bao, Y. Zeng, B. Yan, Cross-day eeg-based emotion recognition using transfer component analysis, Electronics 11 (4) (2022) 651.

[100] M. Long, G. Ding, J. Wang, J. Sun, Y. Guo, P. S. Yu, Transfer sparse coding for robust image representation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 407–414.

[101] R. A. Fisher, Statistical methods for research workers, in: Breakthroughs in statistics, Springer, 1992, pp. 66–70.

[102] T. Ni, Y. Ni, J. Xue, S. Wang, A domain adaptation sparse representation classifier for cross-domain electroencephalogram-based emotion classification, Frontiers in Psychology (2021) 3015.

[103] B. Schölkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, in: International conference on artificial neural networks, Springer, 1997, pp. 583–588.

[104] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, Vol. 4, Springer, 2006.

[105] X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, Y. Li, Unsupervised domain adaptation techniques based on auto-encoder for non-stationary eeg-based emotion recognition, Computers in biology and medicine 79 (2016) 205–214.

[106] W.-L. Zheng, B.-L. Lu, Personalizing eeg-based affective models with transfer learning, in: Proceedings of the twenty-fifth international joint conference on artificial intelligence, 2016, pp. 2732–2738.

[107] Z. Lan, O. Sourina, L. Wang, R. Scherer, G. R. Müller-Putz, Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets, IEEE Transactions on Cognitive and Developmental Systems 11 (1) (2018) 85–94.

[108] K. Yan, L. Kou, D. Zhang, Learning domain-invariant subspace using domain features and independence maximization, IEEE transactions on cybernetics 48 (1) (2017) 288–299.

[109] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf, Measuring statistical dependence with hilbert-schmidt norms, in: Algorithmic Learning Theory: 16th International Conference, ALT 2005, Singapore, October 8-11, 2005. Proceedings 16, Springer, 2005, pp. 63–77.

[110] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, arXiv preprint arXiv:1412.3474 (2014).

[111] W. Zhang, F. Wang, Y. Jiang, Z. Xu, S. Wu, Y. Zhang, Cross-subject eeg-based emotion recognition with deep domain confusion, in: International conference on intelligent robotics and applications, Springer, 2019, pp. 558–570.

[112] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[113] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, S. Coleman, Emotion recognition with convolutional neural network and eeg-based efdms, Neuropsychologia 146 (2020) 107506.

[114] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International conference on machine learning, PMLR, 2015, pp. 97–105.

[115] X. Zhu, K.-H. Thung, E. Adeli, Y. Zhang, D. Shen, Maximum mean discrepancy based multiple kernel learning for incomplete multimodality neuroimaging data, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 72–80.

[116] H. Li, Y.-M. Jin, W.-L. Zheng, B.-L. Lu, Cross-subject emotion recognition using deep adaptation networks, in: International conference on neural information processing, Springer, 2018, pp. 403–413.

[117] F. Kuang, L. Shu, H. Hua, S. Wu, L. Zhang, X. Xu, Y. Liu, M. Jiang, Cross-subject and cross-device wearable eeg emotion recognition using frontal eeg under virtual reality scenes, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021, pp. 3630–3637.

[118] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, Domain-adversarial neural networks, arXiv preprint arXiv:1412.4446 (2014).

[119] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The journal of machine learning research 17 (1) (2016) 2096–2030.

[120] Y.-M. Jin, Y.-D. Luo, W.-L. Zheng, B.-L. Lu, Eeg-based emotion recognition using domain adaptation network, in: 2017 international conference on orange technologies (ICOT), IEEE, 2017, pp. 222–225.

[121] Y. Li, W. Zheng, Z. Cui, T. Zhang, Y. Zong, A novel neural network model based on cerebral hemispheric asymmetry for eeg emotion recognition., in: IJCAI, 2018, pp. 1561–1567.

[122] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, X. Zhou, A bi-hemisphere domain adversarial neural network model for eeg emotion recognition, IEEE Transactions on Affective Computing (2018).

[123] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, T. Song, A novel bi-hemispheric discrepancy model for eeg emotion recognition, IEEE Transactions on Cognitive and Developmental Systems 13 (2) (2020) 354–367.

[124] Z. He, Y. Zhong, J. Pan, An adversarial discriminative temporal convolutional network for eeg-based cross-domain emotion recognition, Computers in biology and medicine 141 (2022) 105048.

[125] Y. Ye, X. Zhu, Y. Li, T. Pan, W. He, Cross-subject eeg-based emotion recognition using adversarial domain adaption with attention mechanism, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 1140–1144.

[126] P. Zhong, D. Wang, C. Miao, Eeg-based emotion recognition using regularized graph neural networks, IEEE Transactions on Affective Computing (2020).

[127] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.

[128] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7167–7176.

[129] G. Bao, N. Zhuang, L. Tong, B. Yan, J. Shu, L. Wang, Y. Zeng, Z. Shen, Two-level domain adaptation neural network for eeg-based emotion recognition, Frontiers in Human Neuroscience 14 (2020).

[130] Y. Li, W. Zheng, L. Wang, Y. Zong, Z. Cui, From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition, IEEE Transactions on Affective Computing (2019).

[131] J. Li, S. Qiu, C. Du, Y. Wang, H. He, Domain adaptation for eeg emotion recognition based on latent representation similarity, IEEE

Transactions on Cognitive and Developmental Systems 12 (2) (2019) 344–353.

[132] S. Furukawa, T. Sakuma, S. Kato, Emotion recognition with domain adaptation based on few-shot eeg learning, in: 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), IEEE, 2021, pp. 1–2.

[133] S. Hwang, M. Ki, K. Hong, H. Byun, Subject-independent eeg-based emotion recognition using adversarial learning, in: 2020 8th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2020, pp. 1–4.

[134] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, H. Wang, An efficient lstm network for emotion recognition from multichannel eeg signals, IEEE Transactions on Affective Computing (2020).

[135] Y. Wang, S. Qiu, X. Ma, H. He, A prototype-based spd matrix network for domain adaptation eeg emotion recognition, Pattern Recognition 110 (2021) 107626.

[136] F. Barbaresco, Innovative tools for radar signal processing based on cartan's geometry of spd matrices & information geometry, in: 2008 IEEE Radar Conference, IEEE, 2008, pp. 1–6.

[137] Z. Li, H. Chen, M. Jin, J. Li, Reducing the calibration effort of eeg emotion recognition using domain adaptation with soft labels, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 5962–5965.

[138] Y. Luo, S.-Y. Zhang, W.-L. Zheng, B.-L. Lu, Wgan domain adaptation for eeg-based emotion recognition, in: International Conference on Neural Information Processing, Springer, 2018, pp. 275–286.

[139] Y. Li, N. Wang, J. Shi, X. Hou, J. Liu, Adaptive batch normalization for practical domain adaptation, Pattern Recognition 80 (2018) 109–117.

[140] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from eeg with deep recurrent-convolutional neural networks, arXiv preprint arXiv:1511.06448 (2015).

[141] S. Hwang, K. Hong, G. Son, H. Byun, Learning cnn features from de features for eeg-based emotion recognition, Pattern Analysis and Applications 23 (3) (2020) 1323–1335.

[142] K. Saito, K. Watanabe, Y. Ushiku, T. Harada, Maximum classifier discrepancy for unsupervised domain adaptation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3723–3732.

[143] K.-M. Ding, T. Kimura, K.-i. Fukui, M. Numao, Eeg emotion enhancement using task-specific domain adversarial neural network, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–8.

[144] R. Ning, C. P. Chen, T. Zhang, Cross-subject eeg emotion recognition using domain adaptive few-shot learning networks, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021, pp. 1468–1472.

[145] Y. Wang, J. Liu, Q. Ruan, S. Wang, C. Wang, Cross-subject eeg emotion classification based on few-label adversarial domain adaption, Expert Systems with Applications 185 (2021) 115581.

[146] G. Koch, R. Zemel, R. Salakhutdinov, et al., Siamese neural networks for one-shot image recognition, in: ICML deep learning workshop, Vol. 2, Lille, 2015, p. 0.

[147] J. Tao, Y. Dan, Multi-source co-adaptation for eeg-based emotion recognition by mining correlation information, Frontiers in Neuroscience 15 (2021) 401.

[148] J. Luo, M. Wu, Z. Wang, Y. Chen, Y. Yang, Progressive low-rank subspace alignment based on semi-supervised joint domain adaption for personalized emotion recognition, Neurocomputing 456 (2021) 312–326.

[149] Y. Hua, X. Zhong, B. Zhang, Z. Yin, J. Zhang, Manifold feature fusion with dynamical feature selection for cross-subject emotion recognition, Brain Sciences 11 (11) (2021) 1392.

[150] O. M. Kenneth, S. A. Bashir, O. A. Abisoye, A. D. Mohammed, Face morphing attack detection in the presence of post-processed image sources using neighborhood component analysis and decision tree classifier, in: International Conference on Information and Communication Technology and Applications, Springer, 2020, pp. 340–354.

[151] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012, pp. 2066–2073.

[152] X. Gu, W. Cai, M. Gao, Y. Jiang, X. Ning, P. Qian, Multi-source domain transfer discriminative dictionary learning modeling for electroencephalogram-based emotion recognition, IEEE Transactions on Computational Social Systems (2022).

[153] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, F. Bach, Supervised dictionary learning, Advances in neural information processing systems 21 (2008).

[154] H. Chen, Z. Li, M. Jin, J. Li, Meernet: Multi-source eeg-based emotion recognition network for generalization across subjects and sessions, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 6094–6097.

[155] Y. Luo, B.-L. Lu, Wasserstein-distance-based multi-source adversarial domain adaptation for emotion recognition and vigilance estimation, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021, pp. 1424–1428.

[156] H. Chen, M. Jin, Z. Li, C. Fan, J. Li, H. He, Ms-mda: Multisource marginal distribution adaptation for cross-subject and cross-session eeg emotion recognition, Frontiers in Neuroscience 15 (2021).

[157] J. Cao, X. He, C. Yang, S. Chen, Z. Li, Z. Wang, Multi-source and multi-representation adaptation for cross-domain electroencephalography emotion recognition., Frontiers in Psychology 12 (2021) 809459–809459.

[158] Y. Cimtay, E. Ekmekcioglu, Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition, Sensors 20 (7) (2020) 2034.

[159] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence, 2017.

[160] A. N. Pusarla, B. A. Singh, C. S. Tripathi, Learning densenet features from eeg based spectrograms for subject independent emotion recognition, Biomedical Signal Processing and Control 74 (2022) 103485.

[161] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[162] J. Li, H. Chen, T. Cai, Foit: Fast online instance transfer for improved eeg emotion recognition, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2020, pp. 2618–2625.

[163] X. Zhang, W. Liang, T. Ding, J. Pan, J. Shen, X. Huang, J. Gao, Individual similarity guided transfer modeling for eeg-based emotion recognition, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019, pp. 1156–1161.

[164] W. Dai, Q. Yang, G.-R. Xue, Y. Yu, Boosting for transfer learning, in: Proceedings of the 24th International Conference on Machine Learning, ICML '07, Association for Computing Machinery, New York, NY, USA, 2007, p. 193–200. doi:10.1145/1273496.1273521.
URL https://doi.org/10.1145/1273496.1273521

[165] Y.-P. Lin, P.-K. Jao, Y.-H. Yang, Improving cross-day eeg-based emotion classification using robust principal component analysis, Frontiers in computational neuroscience 11 (2017) 64.

[166] J. Zhou, S. Chu, X. Li, F. Xiao, L. Sun, An eeg emotion recognition method based on transfer learning and echo state network for hilcps, Microprocessors and Microsystems (2020) 103381.

[167] M. C. Ozturk, D. Xu, J. C. Principe, Analysis and design of echo state networks, Neural computation 19 (1) (2007) 111–138.

[168] F. Wang, W. Zhang, Z. Xu, J. Ping, H. Chu, A deep multi-source adaptation transfer network for cross-subject electroencephalogram emotion recognition, Neural Computing and Applications (2021) 1–13.

[169] K. Muandet, D. Balduzzi, B. Schölkopf, Domain generalization via invariant feature representation, in: International Conference on Machine Learning, PMLR, 2013, pp. 10–18.

[170] M. Ghifary, D. Balduzzi, W. B. Kleijn, M. Zhang, Scatter component analysis: A unified framework for domain adaptation and domain generalization, IEEE transactions on pattern analysis and machine intelligence 39 (7) (2016) 1414–1430.

[171] B.-Q. Ma, H. Li, W.-L. Zheng, B.-L. Lu, Reducing the subject variability of eeg signals with adversarial domain generalization, in: International Conference on Neural Information Processing, Springer, 2019, pp. 30–42.

[172] H. A. Gonzalez, J. Yoo, I. M. Elfadel, Eeg-based emotion detection using unsupervised transfer learning, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 694–697.

[173] J. Liu, X. Shen, S. Song, D. Zhang, Domain adaptation for cross-subject emotion recognition by subject clustering, in: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2021, pp. 904–908.

[174] J. L. Hagad, T. Kimura, K.-i. Fukui, M. Numao, Learning subject-generalized topographical eeg embeddings using deep variational autoencoders and domain-adversarial regularization, Sensors 21 (5) (2021) 1792.

[175] M. Soleymani, M. Pantic, T. Pun, Multimodal emotion recognition in response to videos, IEEE transactions on affective computing 3 (2) (2011) 211–223.

[176] G. Zhao, Y. Zhang, Y. Ge, Frontal eeg asymmetry and middle line power difference in discrete emotions, Frontiers in behavioral neuroscience 12 (2018) 225.

[177] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, N. Sebe, Ascertain: Emotion and personality recognition using commercial sensors, IEEE Transactions on Affective Computing 9 (2) (2016) 147–160.

[178] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, K.-R. Müller, Towards adaptive classification for bci, Journal of neural engineering 3 (1) (2006) R13.

[179] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, Eeg-based emotion recognition in music listening, IEEE Transactions on Biomedical Engineering 57 (7) (2010) 1798–1806.

[180] H. A. Demaree, D. E. Everhart, E. A. Youngstrom, D. W. Harrison, Brain lateralization of emotional processing: historical roots and a future incorporating "dominance", Behavioral and cognitive neuroscience reviews 4 (1) (2005) 3–20.

[181] J. A. Coan, J. J. Allen, The state and trait nature of frontal eeg asymmetry in emotion. (2003).

[182] R. J. Davidson, Hemispheric asymmetry and emotion, Approaches to emotion 2 (1984) 39–57.

[183] E. Sangineto, G. Zen, E. Ricci, N. Sebe, We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 357–366.

[184] A. Apicella, F. Isgrò, A. Pollastro, R. Prevete, On the effects of data normalisation for domain adaptation on eeg data (2022). doi:10.48550/ARXIV.2210.01081.

[185] J. C. Goldsack, A. Coravos, J. P. Bakker, B. Bent, A. V. Dowling, C. Fitzer-Attas, A. Godfrey, J. G. Godino, N. Gujar, E. Izmailova, et al., Verification, analytical validation, and clinical validation (v3): the foundation of determining fit-for-purpose for biometric monitoring technologies (biomets), npj digital Medicine 3 (1) (2020) 1–15.

[186] M. Gerlach, B. Farb, W. Revelle, L. A. Nunes Amaral, A robust data-driven approach identifies four personality types across four large data sets, Nature human behaviour 2 (10) (2018) 735–742.

[187] W. Li, X. Hu, X. Long, L. Tang, J. Chen, F. Wang, D. Zhang, Eeg responses to emotional videos can quantitatively predict big-five personality traits, Neurocomputing 415 (2020) 368–381.

Table 1: Reviewed studies on generalization strategies for emotion recognition. Datasets used, classifiers, evaluation strategy, and type of generalization (i.e. intersubjects, cross sessions and cross datasets) are presented for each entry in the table. (sp = self produced, nl = not labeled; for the other abbreviations see section 8).

Classifier Category	Study	Dataset	Classifier	Evaluation Strategy	Cross Subject	Cross Session	Cross Dataset
CLASSICAL ML	[70]	DEAP	TRFE	LOO	X		
	[65]	DEAP, MAHNOB, sp	RF	LOO	X		X
	[66]	DEAP, SEED	SVM	LOO	X		
	[84]	SEED, DREAMER	DGCNN	LOO	X		
	[67]	DEAP, SEED	SVM	LOO	X		
	[68]	DEAP, sp	SBS	LOO	X		
	[69]	DEAP	TRFE	LOO	X		
	[73]	DEAP, SEED	VAE-LSTM	LOO	X		
	[76]	DEAP	SVM	LOO	X		
	[79]	DEAP, MAHNOB, DREAMER	SVM	LOO	X		
	[71]	DEAP, MAHNOB	RFE	LOO	X		
	[72]	SEED	DECNN	LOO	X		
	[75]	DEAP, SEED	VAE-LSTM	LOO	X		
	[77]	SEED	SVM	LOO	X		
	[78]	SEED	SVM	LOO	X		
	[83]	ASCERTAIN	BiLSTM	LOO	X		
	[85]	SEED, SEED IV	SOGNN	LOO	X		
	[64]	SEED, sp	PTSVM	LOO		X	
UDA (IRS)	[163]	DEAP	SVM	LOO	X		
	[149]	DEAP	NCA	LOO	X		
	[168]	SEED	DMATN	LOO	X		
SHALLOW DA (TSB)	[89]	SEED	ASFM	LOO	X	X	
	[90]	SEED	MSSA	LOO	X		
	[92]	MDME, SDMN	RPCA	ASI		X	
	[93]	SEED	STM	LSO	X		
SHALLOW DA (SSB)	[97]	SEED	TCA	LOO	X		
	[105]	SEED	SAAE	SU2SU, SE2SE, LOO	X	X	
	[106]	SEED	TPT	LOO	X		
	[107]	DEAP, SEED	MIDA	LOO	X		X
	[98]	SEED	TCA	LOO	X		
	[102]	DEAP, SEED	DASRC	LOO		X	X
	[99]	sp	TCA	HO		X	
DEEP DA (CS)	[120]	SEED	DANN	LOO	X		
	[116]	SEED, SEED IV	DAN	LOO	X		
	[121]	SEED	BiDANN	LOO	X		
	[138]	DEAP	WGANDA	LOO	X		
	[111]	SEED	DDC	LOO	X		
	[130]	SEED	R2G-STNN	LOO	X		
	[131]	DEAP, SEED	nl	LOO, SU2SU, O2OSE	X	X	
	[123]	SEED, SEED IV, MPED	BiHDM	LOO	X		
	[126]	SEED, SEED IV	RGNN	LOO	X		
	[129]	SEED, sp	TDANN	LOO	X	X	
	[133]	SEED, CMEED	A-DNN	LOO	X		
	[134]	DEAP, SEED, CMEED	ATDD-LSTM	LOO	X	X	
	[117]	sp	MSDAN	LOO	X		
	[132]	SEED	nl	LOO	X		
	[137]	SEED	nl	LOO	X		
	[143]	SEED	TDANN	SU2SU	X		
	[144]	DEAP, SEED	SDA-FSL	LOSO	X		X
	[147]	DEAP, SEED	MACI	LOO	X		X
	[148]	DEAP, SEED	PLRSA	LOO, SU2SU	X	X	
	[124]	DEAP, DREAMER	AD-TCN	LOO	X		X
	[122]	SEED	BiHDM	LOO	X		
	[152]	DEAP, SEED	MDTDDL	LOO	X		X

Classifier Category	Study	Dataset	Classifier	Evaluation Strategy	Cross Subject	Cross Session	Cross Dataset
DEEP DA (CSS)	[23]	SEED	PPDA	LOO	X		
	[154]	SEED, SEED IV	MEERNet	LOO	X	X	
	[173]	DEAP, sp	DASC	LOO	X		
	[156]	SEED	MS-MDA	LOO	X	X	
	[155]	SEED	wMADA	LOO	X		
UDA (HYBRID)	[165]	sp	GNB	ASI		X	
	[166]	DEAP	ESN	LOO	X		
SUPERVISED DA	[158]	DEAP, SEED, sp	nl	LOO	X		X
	[113]	DEAP, SEED	RCNN	LOO	X		X
	[160]	DEAP, SEED	Densenet	LOO	X		
DEEP DG	[171]	SEED	DG-DANN, DResNet	LOO	X		
	[174]	DEAP, SEED	BiVDANN	LOO	X		

Table 2: The most representative studies according to their classification accuracy, categorised by EEG dataset (SEED, DEAP, others), by number and type of classes considered. DIM = Dimensional; DIS = Discrete; sp = self-produced.

Proposed category	Study	Dataset	Reference Theory	#Classes	Accuracy
SUPERVISED DA	[158]	DEAP	DIM (VAL)	#2 (LV/HV)	72.81 ± 5.07
		Other	DIM (VAL)	#2 (LV/HV)	81.80 ± 10.92
UDA (IRS)	[166]	DEAP	DIM (VAL)	#3 (LV/MV/HV)	68.06 ± 10.93
DEEP DA (CSS)	[173]	DEAP	DIM (VAL-ARO)	#2 (LV/HV) #2 (LA/HA)	73.90 ± 13.50 (VAL) 68.80 ± 11.20 (ARO)
DEEP DA (CS)	[131]	DEAP	DIM (VAL-ARO)	#4 (LALV-HALV-LAHV-HAHV)	62.66 ± 10.45
DEEP DA (CS)	[144]	SEED	DIM (VAL)	#2 (LV/HV)	97.66 ± 14.46
DEEP DA (CS)	[134]	SEED	DIM (VAL)	#3 (LV/MV/HV)	90.92 ± 1.05
		Other	DIM (VAL-ARO)	#2 (LV/HV) #2 (LA/HA)	94.21 ± 5.88 (VAL) 88.03 ± 6.32 (ARO)
DEEP DA (CS)	[129]	Other	DIS (HAPPY, SAD FEAR, ANGER)	#2 (JOY/SADNESS) #2 (JOY/ANGER) #2 (JOY/FEAR)	83.79 ± 1.55 (JOY/SAD) 84.13 ± 1.37 (JOY/ANGER) 81.72 ± 1.30 (JOY/FEAR)
CLASSICAL ML	[85]	Other	DIS (HAPPY, SAD FEAR, NEUTRAL)	#4 (HAPPY/SAD/ FEAR/NEUTRAL)	75.27 ± 8.19