Artificial Intelligence Review
https://doi.org/10.1007/s10462-021-09986-y

Diagnosis and prognosis of mental disorders by means of EEG and deep learning: a systematic mapping study

Manuel J. Rivera[1] · Miguel A. Teruel[1] · Alejandro Maté[1] · Juan Trujillo[1]

Accepted: 8 March 2021

Abstract

Electroencephalography (EEG) is used in the diagnosis and prognosis of mental disorders because it provides brain biomarkers. However, only highly trained doctors can interpret EEG signals due to its complexity. Machine learning has been successfully trained with EEG signals for classifying mental disorders, but a time consuming and disorder-dependant feature engineering (FE) and subsampling process is required over raw EEG data. Deep Learning (DL) is positioned as a prominent research field to process EEG data because (i) it features automated FE by taking advantage of raw EEG signals improving results and (ii) it can be trained over the vast amount of data generated by EEG. In this work, a systematic mapping study has been performed with 46 carefully selected primary studies. Our goals were (i) to provide a clear view of which are the most prominent study topics in diagnosis and prognosis of mental disorders by using EEG with DL, and (ii) to give some recommendations for future works. Some results are: epilepsy was the predominant mental disorder present in around half of the studies, convolutional neural networks also appear in approximate 50% of the works. The main conclusions are (i) processing EEG with DL to detect mental disorders is a promising research field and (ii) to objectively compare performance between studies: public datasets, intra-subject validation, and standard metrics should be used. Additionally, we suggest to pay more attention to ease the reproducibility, and to use (when possible) an available framework to explain the results of the created DL models.

Keywords Deep learning · Diagnosis · Electroencephalogram · systematic mapping study · Mental disorder · Prognosis

1 Introduction

There are more than 150 recognised mental disorders Association (2013), some of them having high prevalence in the population. Indeed, it is estimated that 38.2% of the European population suffer some mental disorder with a 12-month prevalence, that is 164.8 million people Wittchen

✉ Miguel A. Teruel
materuel@dlsi.ua.es

[1] Lucentia Research Group - Department of Software and Computing Systems, University of Alicante, Carretera de San Vicente del Raspeig, s/n, San Vicente del Raspeig, Alicante 03690, Spain

Published online: 27 March 2021

et al. (2011). The four most common mental disorders are anxiety, insomnia, major depression and somatoform (diseases characterized by various discomforts that afflict the patient but which cannot be explained by the existence of an organic lesion). Therefore, an early and accurate diagnostic of mental disorders could improve the quality of life of patients. Regarding such diagnosis, one of the most commonly used techniques is Electroencephalogram (EEG), which provides information from patient's brain that can aid to develop a trustworthy diagnostic.

EEG is performed by using a noninvasive device that captures the electrical brain activity produced by its upper layers. It consists of an array of electrodes which are placed over the scalp of the patient. EEG is widely used in the diagnosis and prognosis of mental disorders, such as epilepsy, which is currently a very active research field Li et al. (2019b), Sharma and Pachori (2017), Zhang and Chen (2016), Sharma (2018). EEG provides objective biomarkers from the patient's brain Olbrich and Arns (2013), Jeste et al. (2015), Olbrich et al. (2016), Mcloughlin et al. (2014). These biomarkers can be recorded because of the high temporal resolution of EEG, which for some devices can reach up to 5000Hz (see Fig. 11). This high-sample frequency provides a continuous lecture of the brain function. Besides, other techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), although provide a full spatial brain map, only capture a static brain image. Furthermore, EEG is a relatively cheap technology since devices cost from in 700 Euros for *Emotiv* wearable devices and rise up to 20000 Euros in the case of a *g.tec* system Zerafa et al. (2018). EEG also enables ambulatory diagnostic, thus reducing the costs of the health system Dash et al. (2012), Faulkner et al. (2012).

Traditionally, the interpretation of the EEG records is performed by doctors who have been trained for hundred of hours. Manually analyzing EEG records is a time-consuming task since the length of the recorded data ranges may vary from hours to days. In order to try to automate the process, Artificial Intelligence (AI) techniques have been proposed to efficiently analyzing these signals, thus minimizing human intervention. The more typical AI task is a classification, where an algorithm is trained to distinguish whether a patient has a mental disorder or not. The main advantages of using AI are (i) to provide a more objective diagnosis and prognosis avoiding the cognitive bias Croskerry (2003) because AI over EEG only use biometric signals and (ii) to enable us to join EEG signals with other biometrics data captured from the patient such as Electrocardiogram (ECG) and Eye Movement(EM), among others, trying to increase the detection accuracy. Machine learning (ML) is a subclass of AI techniques which has been proposed to automatize the processing of these signals Podgorelec (2012), Acharya et al. (2012). However, a highly manual and time-consuming Feature Engineering (FE) process, including channel selection Baig et al. (2020), is still needed for traditional ML algorithms to achieve reasonable accuracy levels. To construct the input features for a ML algorithm, it is necessary to manually compound features that make explicit the information that is contained in the high-complexity raw EEG data Sun and Zhou (2014). Moreover, the FE stage is task-dependent, so this handcrafted process needs to be adjusted to each mental illness.

Deep Learning (DL) is, in turn, subset of ML algorithms that extend the concept of the Neural Networks (NNs) by incorporating more than one hidden layer. DL has successfully been used with a wide range of applications such as computer vision, natural language processing and signal processing. A key feature of DL algorithms is that using their hidden layers, they develop an automatic FE, saving considerable human time and effort although at the cost of the capability to explain the FE carried out. DL is positioning as an alternative to process EEG data because this automation of the FE stage minimizes human intervention. It also can improve the classification accuracy since DL can deal with the raw EEG data by avoiding the information loss produced in an hand-engineered FE. Because of the previously stated reasons, we are only focusing on DL techniques. The number of studies where DL is used to process EEG data to diagnose or prognose a mental disorder has been increasing over the time (see Fig. 2). This fact indicates that it is an open and promising research field. We consider that a wide number of studies can be performed in this field since (i) the high number of mental disorders, DL can process EEG biomarkers to try to automate a more objective diagnosis or prognosis, (ii) DL is an AI field in constant evolution where continuously new algorithms are presented which outperform the previous ones. Therefore, when a new algorithm is released, a new research opportunity is open.

Some previous secondary studies regarding EEG and DL have already been performed Merlin et al. (2020), Craik et al. (2019), Roy et al. (2019). However, to the best of our knowledge, and as we will explain throughout the whole paper, there is still a need for continuing with this research topic (i) by focusing in the diagnosis or prognosis of mental disorders because of the high number of people that presents this type of disorders and (ii) to provide a clear mapping between mental disorders and DL algorithms for detecting the main research areas and research gaps. These are the main reasons why, in this paper, we carry out a Systematic Mapping Study (SMS) for providing which mental disorders and DL techniques have been used for diagnosing and prognosing by means of EEG and DL. Therefore, we are discarding those works not forcing on diagnosis or prognosis. Additionally, in this scientific secondary study we present a mapping of mental disorders with DL techniques and additional biomarkers used. In this mapping, we expose both the main research areas and the existing research gaps in diagnosis and prognosis of mental disorders by means of EEG and DL. In other words, what is already done and what could be done is this field. To carry out this task, the selected primary works were mapped by using previously defined categories according to the research questions proposed in Sect. 4.1. Therefore, to the best of our knowledge, this is the first detailed systematic review work that analyzes and classifies all the current proposals that consider applying Deep Learning techniques on EEG for the diagnosis and prognosis of different mental disorders.

The main purpose of this work is to provide a clear view of which are the most prominent study topics in diagnosis and prognosis of mental disorders by using EEG with DL, as well as to expose some issues we found. For each issue, we provide insights and directions on how to potentially address them. In this manner, we aim to contribute to make genuine progress in the diagnosis and prognosis of mental disorders by using EEG and DL. We strongly believe that this work will be a useful starting point to save time in finding which are the most adequate DL techniques and their best results in the diagnosis and prognosis of mental disorders. Moreover, offering a clear view of the existing issues and potential lines of research will help to foster new ideas and solutions.

The organization of the paper is as follows: in Sect. 2 EEG and DL concepts are briefly presented. Related works and motivation of this work can be found in Sect. 3. The methodology we followed to carry out this study, the research questions, databases used to retrieve papers, screening of studies, the keywording extraction and the mapping results are shown in Sect. 4. In Sect. 5 a discussion of the main insights is presented. Finally, in Sect. 6, shows the conclusions obtained from the study.

2 EEG and deep learning

In the following, EEG and Deep Learning (DL) will be introduced for the sake of the understandably of the rest of this manuscript. Hence, Sect. 2.1 will present EGG and after that, Sect. 2.2 will do so with regard to DL.

2.1 EEG

EEG devices consist of an array of electrodes which are placed in the scalp to capture the electrical signals produced not only by a single neuron but also in a brain region. These signals can be used as biomarkers to diagnose and prognose a wide range of mental illnesses. With regards to the position of the electrodes, the international 10-20 system is the most used one Jurcak et al. (2007) which typically consists in 19 electrodes with 2 additional ones located near the ears Mecarelli (2019). Moreover, high-density electrode systems such as 10-10 system with 81 channels and the 10-5 system with up to 345 electrodes are available. The 10-5 system has not yet been accepted by the American Clinical Neurophysiology Society or by the International Federation of Clinical Neurophysiology Mecarelli (2019). Using an internationally recognized system fosters the study reproducibility.

EEG provides a high temporal resolution because of the high sample rate frequency that can be used in the data acquisition. For analyzing the EEG captured data, it is usually split into five frequency bands, namely, delta < 4 Hz, theta 4–8 Hz, alpha 8–13 Hz, beta 13–30 Hz and gamma > 30 Hz Hu and Zhang (2019). Some advantages that EEG provides over MRI and PET are the low cost price, the non-invasive nature and the possibility of performing an ambulatory diagnostic. This enables the study of the brain activity continuously, over hours or even days. In contrast, because of the sensibility of the EEG system, patient's muscle movements, eye blinks and heartbeats as well as the 50 or 60 Hz of the power lines are recorded too, thus producing interference. Hence, the signal needs to be cleaned of noise in a pre-processing stage, prior to their analysis.

2.2 Deep learning

Neural Networks (NNs) are a type of ML algorithm which is composed by layers where artificial neurons or units are placed. NNs have three kind of layers called input, hidden and output. In the input layer, the input data is placed. Next, the hidden layer takes the input data and then a computation is done to end up transferring the result to the output layer where the final result is returned by the network. Neurons between layers are connected, thus having each of these connections a different *weight*. Neurons from the hidden and the output layers take their inputs and weights and carry out a dot product which result is passed to an *activation function*. Usually, this activation function is nonlinear. The inference in NNs is done by a process called *Forward-Propagation* where the information flows from the input to the output layers. For training the NNs, that is, adjusting the parameters or weights of each unit to produce the expected output, the *Back-Propagation* algorithm is used Rumelhart et al. (1986). In this algorithm, the calculated error in the output layer is "propagated" through the networks to adjust the weights of the connections. The amount added or subtracted for each weight is calculated with the *gradient descent* algorithm which use the direction obtained with the partial derivative of the nonlinear activation function over each weight. In this process a nonlinear boundary decision is "learned".

Deep Learning (DL) involves all NNs architectures which employ a more than one hidden layer. There are several DL architectures although the most commons are: Multilayer Perceptron (MLP), convolutional neural networks (CNN), Recurrent Neural Networks (RNN), Autoencoders (AE), and generative models Shrestha and Mahmood (2019). A more detailed description of all these kinds of DL architectures is provided in Sect. 4.4.

In recent years DL has been widely applied to computer vision, natural language processing and signal processing tasks, achieving groundbreaking results in these fields. In healthcare, DL has become a relevant AI technique Miotto et al. (2017), Domingues et al. (2019), Murtaza et al. (2019). Also, in recent years, DL has become a promising research topic to process EEG records (see Fig. 2). The main advantage of processing EEG data with DL, is that, in contrast with traditional ML techniques, it can treat raw EEG data because it performs an automatic feature engineering (FE) or feature extraction. An automatic FE can extract new information from the raw data which can improve the classification result where performing hand-crafted FE is not feasible. In this sense, CNNs can process raw EEG data benefiting from spatial information, by applying convolutions over the input EEG data in various manner: 1D convolutions to process channels in a isolated form or 2D and 3D convolutions to process neighboring channels together Wei et al. (2018), Phang et al. (2020), Mumtaz and Qayyum (2019). Raw EEG data, as sequence data, can also be processed by exploiting its temporal component. RNNs can benefit from this information by using feedback connections capable of learning from this sequence data Warrick et al. (2019). Machine learning techniques such as Decision Trees, KNN, or SVM are not capable of use raw EEG data due to (i) high dimensionality: e.g. an EEG that involves 19 channels with a sample rate of 256 Hz generates 292 k records per minute, that is, 292 k raw input features. As a result, this becomes a problem for ML algorithms because (i) the data volume translates into significant processing time, (ii) can translate into an overfitting problem, and (iii) they can not take advantage of the spatio-temporal information because there is not possible to inform the algorithm that a set of features are spatially together or form a sequence. Even when a huge amount of data is not available, DL is still a feasible method by using pre-trained networks. This technique is called *transfer learning*Yosinski et al. (2014). For more details see Sect. 5.

3 Related works

Some previous secondary studies has been done by reviewing the application of the AI field to process EEG data. Concerning to the use of ML, Hosseini et al. (2020) reviewed ML methods made for EEG analysis with bioengineering applications. They presented that all main ML algorithms that have been used for EEG classification tasks, such as emotion recognition, measure mental workload, sleep scoring and mental disorder among others. As stated in the introduction section, Feature Engineering is an important step prior to using ML over EEG data, Noor et al. (2020) reviewed which EEG-extracted features are more useful for a favourable or unfavorable outcome prediction made by ML algorithms.

Regarding reviews in EEG with DL, Merlin et al. (2020) has focused in which DL techniques were used for EEG signal applications; Craik et al. (2019) analysed which DL algorithms have been used for EEG classification tasks. Additionally, they (i) showed how the input data is presented to the algorithm: images, calculated features and signal values, and (ii) explore the existence of specifics DL algorithms for a specific type of task. Finally in Roy et al. (2019) a systematic review is done for DL-based EEG analysis where they discuss the studies from various aspects: data, preprocessing methodology, DL design choices, results and reproducibility of the experiments.

To the best of our knowledge, this research is the first Systematic Mapping Study (SMS) done in diagnosis and prognosis of mental disorders by means of EEG and DL. We have analysed four main aspects in the selected studies: (i) which mental disorders have been diagnosed and prognosed, (ii) DL techniques used, (iii) additional used biomarkers and (iv) datasets used. Additionally, we present two mappings: (i) mapping with mental disorders, DL techniques and additional used biomarkers and (ii) mapping with the nature of the disorder detection (diagnosis/prognosis), mental disorders and DL techniques. With those mapping, we will show what is already done and what could be done is this field.

With the millions of people who are affected with a mental disorder in our mind, we have done this mapping to motivate to the research community to investigate in this field of study by clearly showing which are the most active topics as well as exposing some future research directions. In the same manner, we provide some guidelines for future studies.

4 Methodology

A Systematic Mapping Study (SMS) is a secondary study which reviews primary research with the aim of identifying research literature gaps in the field of interest Petersen et al. (2008), Budgen (2008), Kitchenham et al. (2011). Other secondary studies can be used to carry out reviews such as systematic literature review Kitchenham and Charters (2007) but we have chosen SMS because of the clearer visual summary that this kind of study provides.

This work follows the SMS methodology proposed in Petersen et al. (2008). However, we have applied a slightly different approach in order to provide more details and deeper understanding. Indeed, instead of reading only the papers' abstract and conclusion, a full read has been done because the shortage of works in our field of interest. The stages of the SMS are:

1. *Definition of Research Questions*. These are the questions that guide the work.
2. *Conduct Search*. Identification of the primary studies in databases through querying.
3. *Screening of Papers*. A inclusion and exclusion criteria are used to keep relevant papers according the work target.
4. *Keywording of full text*. A set of keywords and concepts is extracted of each paper while reading. These concepts could be put together into categories. Finally all keywords must be grouped. This enables us to relate all the works giving a global view and to put each one in context with regards to the rest. Keywording in the methodology proposed by Petersen et al. (2008) is done by reading abstracts, but we have carried out with a full read because the low number of works present in out field of interest.
5. *Data Extraction and Mapping of Studies*. Frequency analysis of the keywords enables us to identify which categories has been exploited in the past on the topic and hence identify gaps and future research directions.

4.1 Definition of research questions

As we have to collect works which use EEG data with DL to detect or predict illness in patients, we propose the following four research questions:

- RQ1: Which mental disorders have already been diagnosed and prognosed by means of EEG and DL?
- RQ2: Which DL techniques have already been applied for diagnosing or prognosing mental disorders by using EEG data as input?

- RQ3: Which other biometric data is combined with EEG?
- RQ4: Which is the source of the datasets?

4.2 Conducted search

The following terms were used to assemble the set of works which will be used in order to answer our research questions:

- EEG
- DL
- diagnosis
- prognosis

These terms were combined to assemble the query *EEG and DL and (diagnosis or prognosis)*. The search was carried out in five well-known public databases, namely *Scopus*, *Web of Science*, *IEEE Xplore*, *ScienceDirect* and *ACM Digital Library*. In Table 1, it is shown the composed queries for each database according to each search engine syntax. These queries were carried out by using the following text fields: title, abstract and keywords. A total of 341 works were obtained. In the next subsection we defined an inclusion and exclusion criteria to obtain only relevant works.

4.3 Screening of papers

Once the initial collection of works has been compiled, it is necessary to establish inclusion and exclusion criteria to keep only those works related to our research questions. Table 2 shows our defined criteria. Within the inclusion criteria, *i1* is relevant because only few papers were found before 2016. That is because such year could be considered as the

Table 1 Queries made for each database

Database	Query
Scopus	TITLE-ABS-KEY ((*eeg* OR electroencephalogra*) AND (diagnos* OR prognos*) AND "deep learning")
Web of Science	TS=((*eeg* OR electroencephalograp*) AND (diagnos* OR prognos*) AND "deep learning")
IEEE Xplore	("All Metadata":*eeg* OR electroencephalograp*) AND ("All Metadata":diagnos* OR prognos*) AND ("All Metadata":"deep learning")
ScienceDirect[a]	((eeg OR electroencephalography) AND (diagnosis OR prognosis) AND "deep learning")
ACM	Title:(*eeg* OR electroencephalograp*) OR Abstract:(*eeg* OR electroencephalograp*) OR Keyword:(*eeg* OR electroencephalograp*)) AND (Title:(diagnos* OR prognos) OR Abstract:(diagnos* OR prognos) OR Keyword:(diagnos* OR prognos)) AND (Title:("deep learning") OR Abstract:("deep learning") OR Keyword:("deep learning")

[a]Wildcard characters are not supported in this database

Table 2 Inclusion and exclusion criteria used to keep relevant papers

Criteria			
Inclusion		Exclusion	
i1	Date of publish between 2016 to 2020-05-28	*e1*	No EEG signal used
i2	Journal paper	*e2*	No mental disorder associated
i3	English-written works	*e3*	No deep learning
i4	Primary study		

the upswing of DL (mainly, due to the introduction of GPU parallel processing) and we considered that the algorithms and techniques used prior to 2016 were not relevant for our goal. The criterion *i2* was selected to keep only the most relevant works and *i4* because the followed methodology advised so Petersen et al. (2008). Relative to the exclusion criteria, *e1* ensure that we selected works where EEG is used in standalone form or combined with other biometric signals. Hence *e1* criteria is related to all of our four RQs showed in Sect. 4.1. The *e2* is used due to RQ1, we only want to keep works where mental disorders are treated. Finally *e3* is applied because of RQ2, and since we stick to field of artificial intelligence, that is, DL.

As shown in Fig. 1, after removing duplicate works by using DOI, inclusion and exclusion criteria from Table 2 were applied. The criteria *i4*, *e1*, *e2* and *e3* were carried out

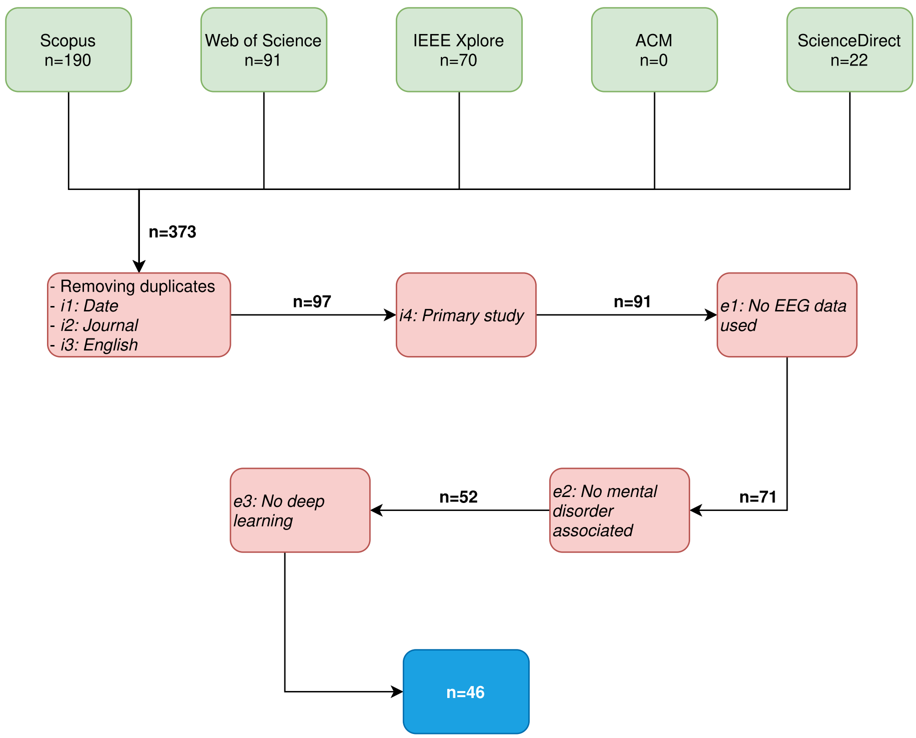


Fig. 1 Procedure to select works for mapping

manually after reading the title and abstract of the 97 candidate works. At the end of this process a total of 46 studies were carefully selected to accomplish this work. For a comprehensive list with the selected papers and their extracted features, see "Appendix 2".

4.4 Keywording of full text

Following the methodology, prior to answering the RQs (see Sect. 4.1), it is fundamental to define for each RQ a set of keywords which are extracted by reading the papers. In our analysis, we firstly made a full lecture for each of the 46 selected papers where we extracted all the keywords that were considered as relevant to answer the RQs. Because of the number of keywords and the high granularity of them, we then defined groups of related keywords called *categories*. Finally each category is associated with the related RQ. The discovered keywords and categories are as follows:

RQ1.- After a full read of the selected works, we have collected a set of 23 keywords (see Table 3) which we will use to answer the RQ. Each of these keywords correspond to a mental disorder. A brief explanation of the gathered mental disorders is presented in the following: *Attention Deficit Hyperactivity Disorder (ADHD)* is a neuropsychiatric disorder characterized by inattentively, hyperactivity and impulsive behaviour which has high prevalence in children and adolescents Kieling and Rohde (2012); *Autism Spectrum Disorder* is a developmental disorder that affects communication and social interaction by

Table 3 Keywords and categories extracted for the RQ1

Keyword	Category
ADHD	ADHD
Autism spectrum disorder	Autism
Coma	Coma
Creutzfeldt-Jakob disease	Dementia
Alzheimer’s Disease	
Dementia with Lewy bodies	
Rapidly Progressive Dementia	
Dementia	
Unipolar Depression	Depression
Major Depressive Disorder	
Mild Depression	
Depression	
Fast Ripples	Epilepsy
Idiopathic Generalized Epilepsy	
Epileptic Seizure Detection	
Epileptic Seizure Prediction	
Parkinson’s Disease	Parkinson’s Disease
Schizophrenia	Schizophrenia
Insomnia	Sleep disorder
Sleep disorder	
Apnea	
Limb Movements	
Sleep Arousals	

the presence of repetitive patterns and behaviours Johnson et al. (2007); *Coma* is a prolonged state of unconsciousness where a person is alive but unable to move or interact with his or her environment Jonas et al. (2019); *Creutzfeldt-Jakob disease* is a fatal and rapid degenerative brain disorder Zerr et al. (2009) whose first symptoms are poor coordination, memory problems or impaired vision; *Dementia* is the progressive loss of cognitive functions that avoid to develop everyday activities; *Alzheimer's Disease* is the most common cause of dementia type Ferri et al. (2005); *Dementia with Lewy bodies* is a common form of dementia characterized leading in thinking problems and fluctuations in the level of consciousness Weisman and McKeith (2007); *Rapidly Progressive Dementia* whose speed of patient's deterioration is over months, weeks or days Geschwind et al. (2008); *Depression* is a serious mood disorder that affect negatively in sleeping, eating or working Organization (2001); *Fast Ripples* are high frequency signals that can be recorded by EEG and has been studied as biomarker of Epilepsy Bernardo et al. (2018); *Epilepsy* causes frequent seizures which is a abnormal electrical activity Fisher et al. (2005); *Parkinson's Disease* is a progressive neurological disorder that cause not voluntary and uncontrollable movements in the body Jankovic (2008); *Schizophrenia* is a mental disorder that interrupts the normal thinking, speech, feelings and behaviours Andreasen (1982); *Insomnia* is a sleep disorder that can make it difficult to fall asleep Roth (2007); *Apnea* is a sleep disorder where the patient has pauses in breathing Young et al. (2002); *Limb Movements* makes the patient moves limbs involuntarily and during sleep Picchietti and Winkelman (2005); *Sleep Arousals* are short interruptions in the sleep Halász (2004). From 23 keywords, we determined 10 mental disorder categories (see Table 3). Collapsing specific mental disorders into a category will enable us to obtain a high-level vision of the spectrum of disorders that have been diagnosed and prognosed with EEG and DL.

RQ2.- These 18 keywords relative to DL algorithms were extracted (see Table 4): Multilayer Perceptron (*MLP*) is composed by one input layer, one or more hidden layers

Table 4 Keywords and categories extracted for the RQ2

Keyword	Category
Autoencoder	Autoencoder
Stacked Autoencoder	
DAE	
MDAE	
C-LSTM	C-LSTM
CAE	CAE
CNN	CNN
Fast R-CNN	
Convolutional VAE	Convolutional VAE
GRU	GRU
Bidirectional LSTM	LSTM
LSTM	
MLP	MLP
RCNN	RCNN
RBM	RBM
RNN	RNN
SNN	SNN
WGAN	WGAN

and one output layer Rumelhart et al. (1986). If the network has more than one hidden layer then can be called Deep Neural Network (DNN); Convolutional Neural Network *(CNN)* uses the convolution operation in one or more layers where a convolution kernel is applied to the input data to produce a feature map. These networks can learn spatial features from the input data LeCun (1998); Fast Region-based Convolutional Network *(Fast R-CNN)* develops object detection with high accuracy and less training time consumption because of the low number of parameters Girshick (2015); Recurrent Neural Network *(RNN)* is a DNN which includes feedback connections capable to learn from sequence data Rumelhart et al. (1986); Long-Short Term Memory *(LSTM)* is a RNN that uses LSTM cells for solving the vanish gradient problem, performing better, taking less time to train and detecting long-term dependencies Hochreiter and Schmidhuber (1997); Gated Recurrent Unit *(GRU)* is a simplification of the LSTM cell Cho et al. (2014); *Bidirectional LSTM* not only preserve information of the input sequence from the past (backward) but also from the future (forward) Graves and Schmidhuber (2005); *RCNN* and *C-LSTM* are hybrid architectures called Recurrent Convolutional Neural Network and Convolutional Long-Short Term Memory respectively, which are composed by a CNN model followed by a RNN model, these networks can learn spatio-temporal features from the input data Zhou et al. (2015); *Autoencoder* is an neural network capable to create a deterministic low dimensional representation of the input vector Bourlard and Kamp (1988); *Stacked Autoencoder* is an Autoencoder with more than one hidden layer Hinton and Salakhutdinov (2006); Denoising Autoencoder *(DAE)* differs that is trained with corrupted inputs, the result is a learned low dimensional representation more robust to noise Vincent et al. (2010); Multimodal Denoising Autoencoder *(MDAE)* is capable to learn a join representation from the input data that come from multiple modalities Poirson and Idrees (2013); Convolutional Autoencoder *(CAE)* can learn a low dimensional latent representation preserving spatial locality of the input data Masci et al. (2011); Convolutional Variational Autoencoder *(Convolutional VAE)* is a probabilistic Autoencoder that can generate new instances similar to the training set Kingma and Welling (2014) preserving spatial locality as well; Restricted Boltzmann Machines (*RBM*) is a type of generative model that learns a probability distribution from its inputs and has not connections between visible or between hidden units Smolensky (1986); Wasserstein Generative Adversarial Network (*WGAN*) is a generative model that involves two neural networks called *generator*, that produce fake noise samples from training data, and *discriminator*, that have to differentiate if the sample is real or a fake made by the generator, and uses Wasserstein distance for measuring model performance in order to avoid the vanishing gradient problem Arjovsky et al. (2017); Siamese Neural Network *SNN* are two identical networks where comparing their encoding outputs for two given inputs, is returned the degree of similarity of these two inputs Bromley et al. (1994). Once all keywords were put together, 13 categories were defined (see Table 4). That are: *Autoencoder*, *C-LSTM*, *CAE*, *CNN*, *Convolutional VAE*, *GRU*, *LSTM*, *MLP*, *RCNN*, *RBM*, *RNN*, *SNN* and *WGAN*. We only collapsed those keywords which, as far as we known, are DNN related ones.

RQ3.- We have found eight keywords corresponding to different signals that were obtained in conjunction with EEG (see Table 5): *Abdomen belt*, *Airflow* and *Chest belt* capture respiratory signals; Electrocardiogram (*ECG*) measure the electrical signals generated by the heart; Eye movement (*EM*) data can be used to detect eye blinks and clean EEG signals; Electromyography (*EMG*) measure the electrical activity produced by skeletal muscles; Electrooculography (*EOG*) capture the eye movements measuring

Table 5 Keywords and categories extracted for the RQ3

Keyword category
Abdomen belt
Airflow
Chest belt
ECG
EM
EMG
EOG
SaO2

Table 6 Keywords and categories extracted for the RQ4

Keyword category
Ad-hoc
Bonn University Andrzejak et al. (2001)
CHB-MIT Shoeb (2009)
Freiburg University University of Freiburg (2003)
Lomonosov Moscow State University Gorbachevskaya and Borisov (2002)
Kaggle[a]
Other
PhysioNet/CinC 2018 Challenge[b]
Predict[c]
Sleep Heart Health Study (SHHS) Dean et al. (2016)
Temple University Obeid and Picone (2016)
UCI[d]

[a] https://www.kaggle.com/

[b] https://physionet.org/content/challenge-2018/1.0.0/

[c] http://predict.cs.unm.edu/downloads.php

[d] https://archive.ics.uci.edu/ml/index.php

the standing corneal-retinal potential; Oxygen Saturation (*SaO2*) measure by blood analysis.

RQ4.- Thirteen keywords were defined for RQ4 which compounds the same categories (see Table 6). Category *Ad-hoc* refers to a datasets made purposely for the work and *Other* refers to the use of third-party EEG data coming from other studies.

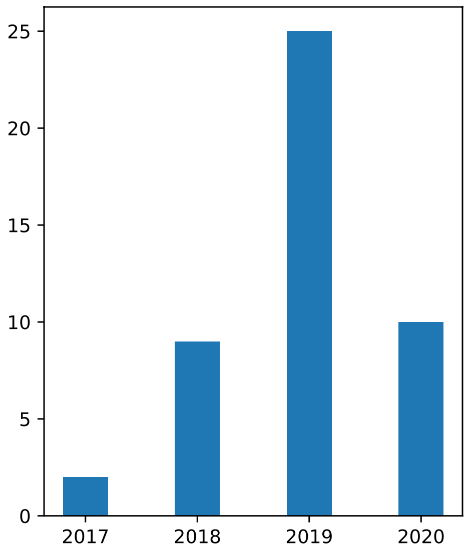
In the next section, we will answer the RQs by doing frequency analysis of the related categories and also, we will provide additional information extracted from the full read done of each work.

4.5 Data extraction and mapping of studies

For answering the proposed RQs in Sect. 4.1, we carried out a frequency analysis of the categories defined above. It is worth noticing that a paper can involve more than one category for an attribute, e.g. there are papers where more than a DL algorithm, additional biometric data or dataset was used. In the same manner, some studies do not provide information for some analyzed attributes such as sampling frequency or the number of EEG channels, among others. Thus, in the following analysis, the reader will notice how the sum of papers may be higher or lower (because the study do not provide information), that the number of selected papers (n = 46). In "Appendix 2" is provided the extracted information from the selected studies allowing the reproducibility of our results. In overall, the number of the published works growth each year (see Fig. 2) being China, USA and Malaysia the countries where more studies are produced (see Fig. 3). The frequency analysis for the RQs are as follows:

RQ1.- Nine mental disorders have been diagnosed or prognosed with DL (see Fig. 4a). The predominant disorder found was epilepsy, representing the 47.83% (22 papers) of the total of studies, followed by depression with a 15.22% (7 papers) and schizophrenia with 8.7% (4 papers). The next disorders correspond to sleep disorder, dementia (3 papers) and ADHD with a 6.52% (3 papers) for each one, Coma with 4.35% (2 papers) and finally Parkinson's disease and autism appearing in Ruffini et al. (2019), Ali et al. (2020) respectively. Concerning to the performance achieved by classifying each mental disorder, we have extracted the accuracy for those studies which made a more-reliable intra-subject validation, that is, the test set only contains patients not used in the training set. In the same manner, there is not possible to carry out an objective performance comparison between studies because (i) there is not a reference dataset for each mental disorder and (ii) there is not a common set of metrics used in the studies to measure the performance (see Sect. 5 for more details). Hence, the following information is provided as a guideline only. The median accuracy by mental disorders is as follows: sleep disorder 90.1%, schizophrenia 88.13%, epilepsy 87.84%, comma 87.04%, depression 82.74%, Parkinson's disease 81% and ADHD 83% (see Fig. 5a). For dementia and autism, the studies carry out inter-subject validation. On the other hand, we have classified the studies in two classes (see Fig. 4b): diagnosis and prognosis. Most

Fig. 2 Number of studies published per year until *2020-05-28*



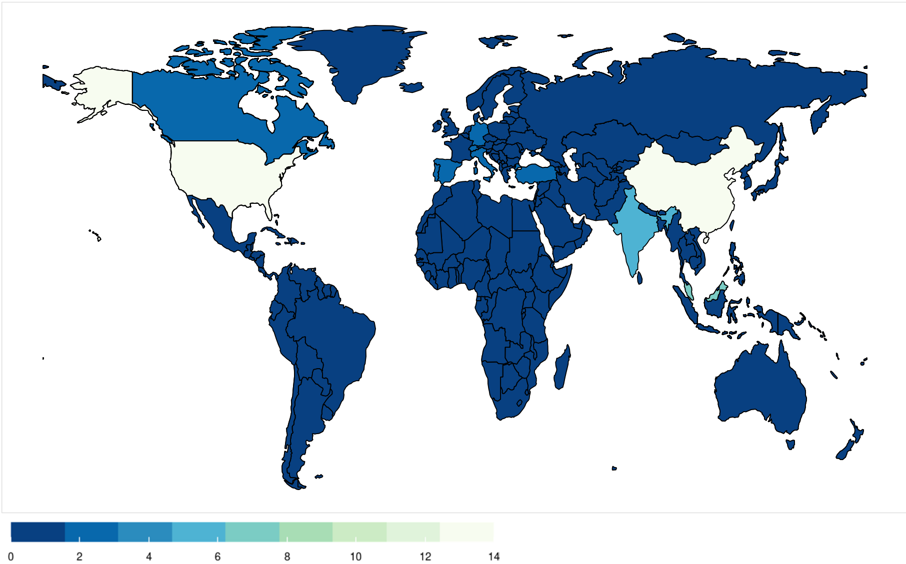


Fig. 3 Frequency of the countries for all appearing authors in the selected papers

of the studies perform diagnosis of the mental disorders with a 69.64% (39 papers) with 23.21% for prognosis (13 papers). It is worth noting that there are 6 studies Thara et al. (2019), Liang et al. (2019), Wei et al. (2018), Acharya et al. (2018b), Hussein et al. (2019), Khan et al. (2018) where we considered that both, diagnosis and prognosis, was carried out.

RQ2.- CNN is the most widely used DL algorithm (see Fig. 5b) with a 48.58% (28 papers), following by C-LSTM with a 10.34% (6 papers), LSTM with a 8.62% (5 papers) and AE with a 5.17% (3 papers). After, we can find RNN, MLP, GRU, CAE and Bidirectional LSTM with a 3.45% (2 papers) for each one. The DL algorithms with less frequency are WGAN, SNN, RBM, RCNN, DAE and Convolutional VAE used in Wei et al. (2019), Calhas et al. (2020), Bi and HaiboWang (2019), Biswal et al. (2018), Yuan et al. (2019), Abdelhameed and Bayoumi (2019) respectively. It is clear that Convolutional and Recurrent networks are the most used techniques, with the hybridization between these two models, that is, C-LSTM standing in third place. Because of epilepsy is the most frequent mental disorder, we have repeated this analysis after removing studies which involve this mental disorder. We noticed how the distribution of DL algorithms remain similar. Consequently, our analysis is not biased. For those studies that provide details, some characteristics of their DL algorithms were extracted:

- The mean of the number of layers is 8.5 with a standard deviation of 4.81 (see Fig. 6a).
- The activation functions used are: softmax appears in the 45.28% (24 papers), second is Rectified Linear Unit (ReLU) with a 30.19% (16 papers) and third is sigmoid with a 9.43% (5 papers). Next are hyperbolic tangent (Tanh) and Leaky ReLU present in a 5.66% (3 papers) for each one. Finally Exponential Linear Unit (ELU) and linear activation functions were only used in Vahid et al. (2019) (see Fig. 6b).

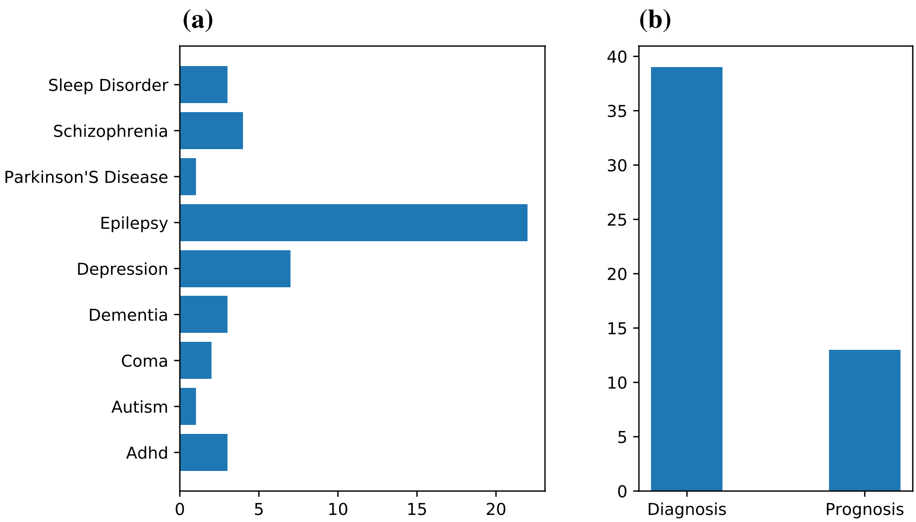


Fig. 4 **a** Frequency of defined mental disorders categories in selected works and **b** frequency of nature of the disorder detection in selected works

- Dropout to avoid overfitting is used in 52.17% (22 papers) and Batch Normalization, which avoids the vanish and explode gradient problems, is less used, thus appearing in the 19.57% (9 papers) (see Fig. 6c).
- In term of optimizers, Adam is the most widely used with a 78.79% (26 papers) following by Stochastic Gradient Descent (SGD), RMSprop and Adadelta with a 6.06% (2 papers) for each one. Finally Nadam only appears in Chen et al. (2019) (see Fig. 6d).

As for the type to the metrics used to measure the performance (see Fig. 7

a), the four most used are accuracy with a 29.13% (37 papers), sensitivity with a 23.62% (30 papers), specificity with a 17.32% (22 papers) and ROC-AUC with a 11.02% (14 papers). The rest of the metrics are as follows: precision with a 7.87% (10 papers), F1-score with a 5.51% (7 papers), false positive rate with a 3.15% (4 papers), precision-recall curve (PRC) with a 1.57% (2 papers) and false negative rate with 0.79% (1 paper). On the other hand, 39.13%, that is, 18 papers have measured the performance of the proposed DL algorithm with a intra-subject validation (see Fig. 7b).

RQ3.- As we explained above, we have found eight additional signals processed together with EEG signals (see Table 6) in only four different papers (8.7%). EMG is the most frequent one which has been used in Wei et al. (2019), Biswal et al. (2018), Warrick et al. (2019). Abdomen belt, Airflow, Chest Belt and SaO2 were used in Biswal et al. (2018), Warrick et al. (2019). ECG and EOG appear in Wei et al. (2019), Warrick et al. (2019). Finally, EM was used in Zhu et al. (2019) (Fig. 8).

RQ4.- There are 10 public EEG datasets present in the 50.98% of the studies. Ad-hoc datasets which were built for the paper purpose or reused from the author's previous study are employed in the 43.14% (see Fig. 9a). A third case named *Other* refers to the use of third-party EEG data coming from other studies which constitutes the 5.88%. In five works Abdelhameed and Bayoumi (2019), Wen and Zhang (2018), Khan et al. (2018), Biswal et al. (2018), Fürbass et al. (2020) two datasets were used. In the Fig. 9b, it is showed the frequency in the use of datasets by their source (ad-hoc, public or other) and mental

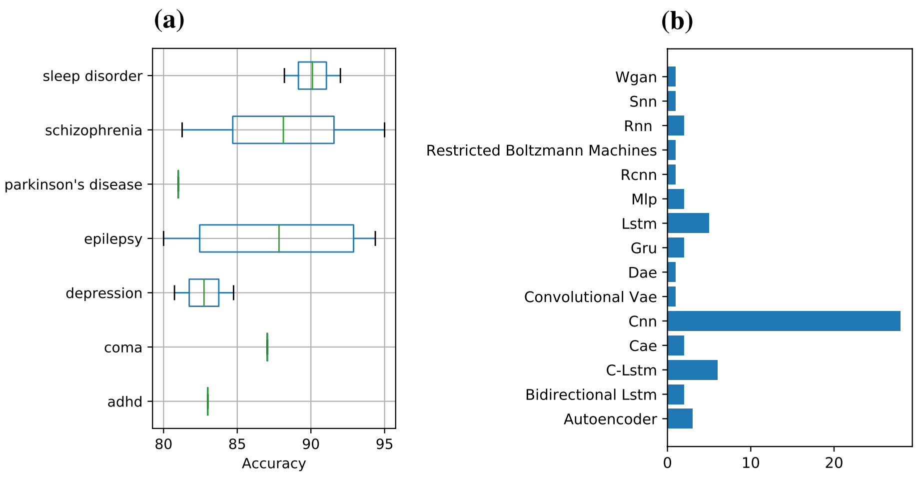


Fig. 5 **a** Accuracy achieved by mental disorder (only studies with intra-subject validation) and **b** frequency of DL techniques present in the studies

disorder. Only in epilepsy, schizophrenia, sleep disorder and depression there are studies where a public dataset has been used. Moreover, in the first three mental disorders, the number of studies carried out with public datasets is greater or equal to those performed with ad-hoc datasets. On the other hand, regarding to the number of channels used to record EEG signals, 19 is the most common number with a 21.05% (8 papers) and 256 Hz is the most used sample rate with a 35% (14 papers). See Figs. 10 and 11.

Finally, as proposed in the methodology Petersen et al. (2008), we have carried out two visualizations to map the RQ1, RQ2 and RQ3. We consider that doing these mapping where RQs are combined, enables us to detect gaps where new research works could be carried out. In the following, the relationship among RQ1 with RQ2 and RQ3 will be presented:

- *Mapping of mental disorders with DL (mapping of the RQ1 and RQ2)*. Epilepsy is the mental disorder where most of the DL techniques has been used: CNN, RNN, C-LSTM, Bidirectional LSTM, CAE, Convolutional VAE, DAE, GRU, LSTM and WGAN. In second place is dementia with four algorithms: CNN, AE, MLP, RBM. With three DL techniques are sleep disorder (MLP, Bidirectional LSTM and RCNN), Parkinson's disease (CNN, GRU, LSTM), schizophrenia (CNN, GRU, LSTM) and depression (CNN, AE, C-LSTM). In ADHD only CNN and RNN have been studied. Finally in coma and autism CNN is the only technique employed. See Fig. 12.
- *Mapping of Mental disorders and other biometric data (mapping of the RQ1 and RQ3)*. In sleep disorder EEG data has been combined with ECG, EMG, EOG, Abdomen belt, Airflow, Chest belt and SaO2 signals, that is, 7 out 8 additional biometric data categories identified. In epilepsy, ECG, EMG and EOG has been used. At last with depression only appears with EM. See Fig. 12.
- *Mapping of mental disorder with the nature of the disorder detection (mapping of the RQ1)*. Diagnosis has been carried out for ADHD, autism, depression, schizophrenia, and

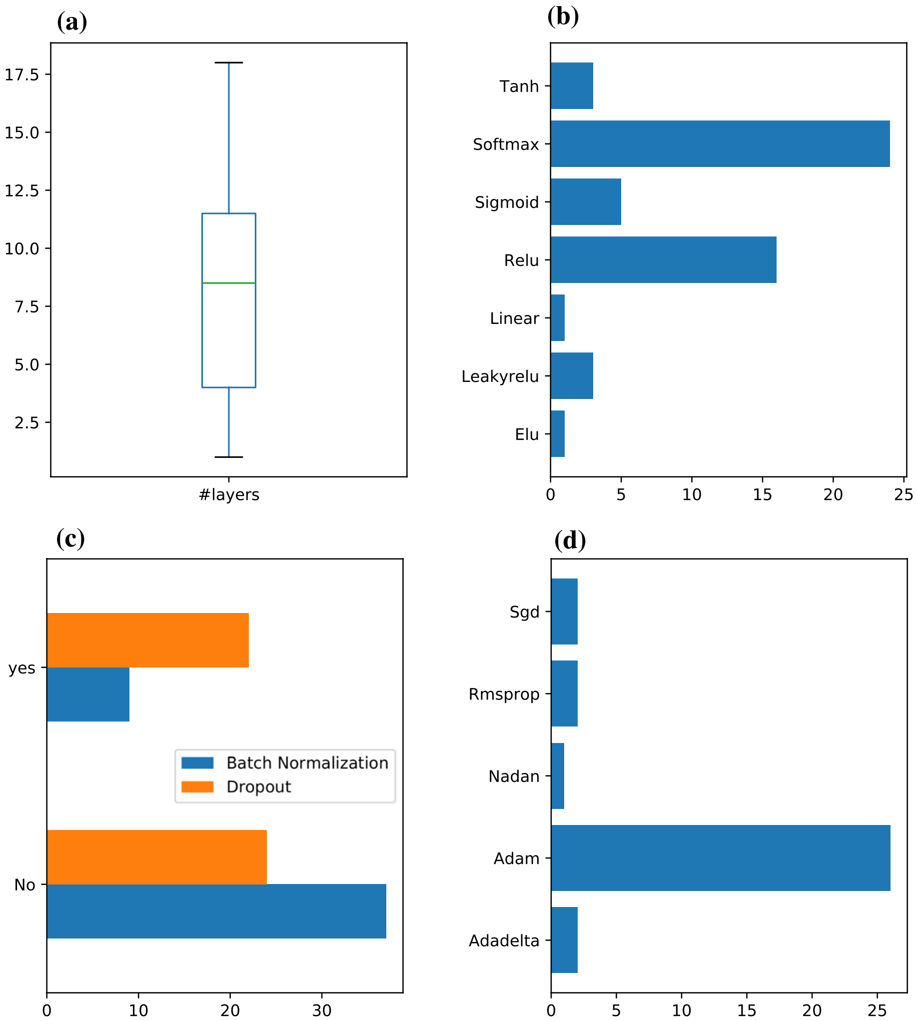


Fig. 6 **a** number of layers used for the DL algorithms, **b** frequency of the activation functions, **c** frequency of use of Batch Normalization and Dropout and **d** frequency of the optimizers

sleep disorder. Prognosis has been done for coma and Parkinson's disease. With dementia and epilepsy, both diagnosis and prognosis, has been carried out. See Fig. 13.

- *Mapping of DL with the nature of the disorder detection (mapping of the RQ1 and RQ2).* For diagnosis, only GRU cells has not been used. In prognosis there has been used AE, Bidirectional LSTM, C-LSTM, CNN, LSTM, RNN and GRU. See Fig. 13.

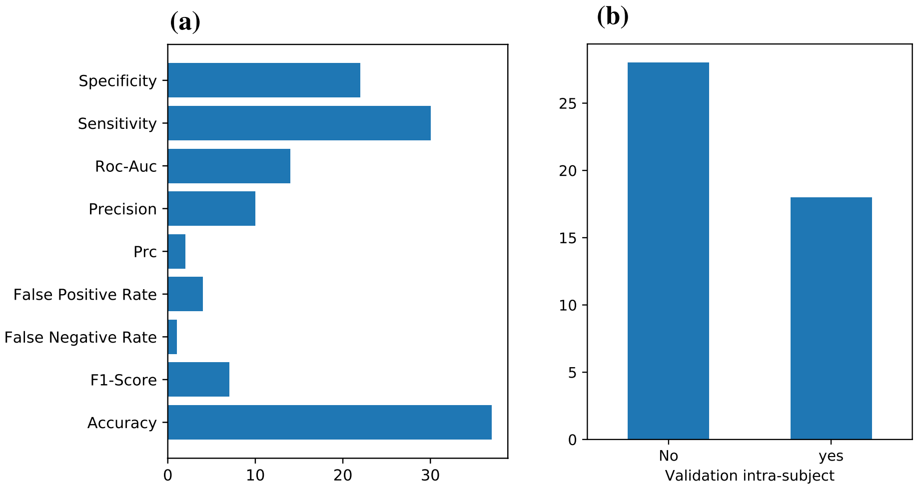


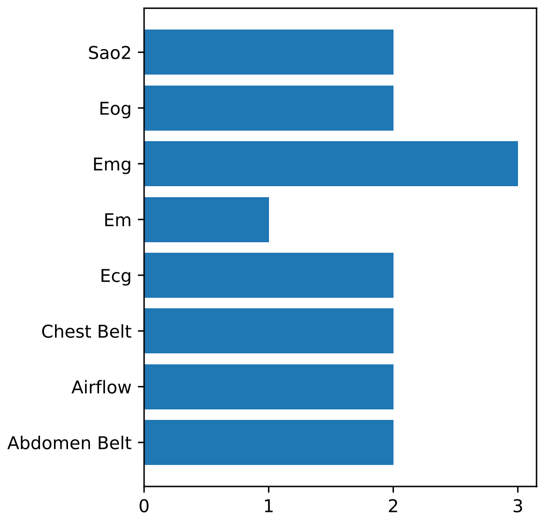
Fig. 7 **a** Frequency of the metrics used in the studies and **b** amount of studies where intra-subject validation is carried out

5 Discussion

In this section, we will discuss interesting insights discovered in this review and expose some issues we found out for those we will provide some insights in order to address them.

Related to mental disorders and DL techniques (see Fig. 12), we can say that, without taking account epilepsy which is where most studies have focused, the rest of mental disorders present several gaps regarding to DL techniques application. Thus, these gaps indicate unexplored fields where new researches could be carried out, that is, processing EEG data with DL to diagnose or prognose mental disorders. In this sense, CNN is

Fig. 8 Frequency of additional biosignals extracted while EEG recording



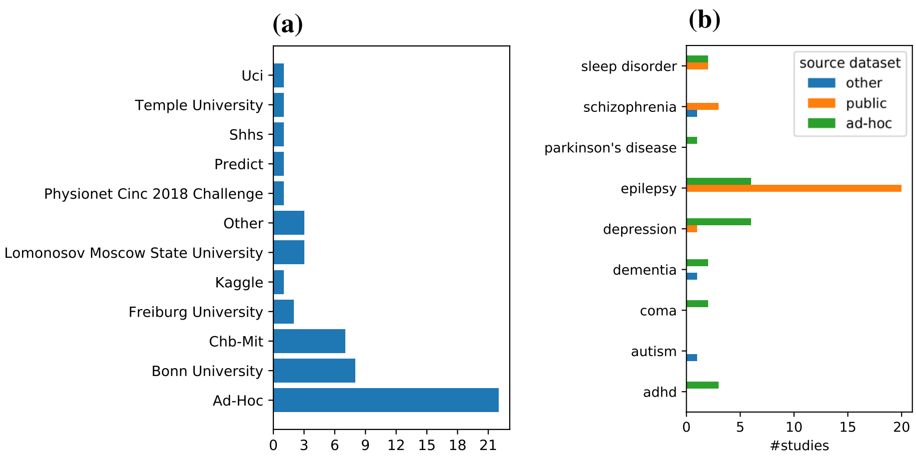


Fig. 9 **a** Frequency of the used EEG datasets and **b** frequency in the use of datasets by their source and mental disorder

the most used DL architecture, while the rest of the algorithms are much less used. To the best of our knowledge, CNNs can extract spatial features from the input data but not features that take into account the sequentiality present in those data, i.e. EEG data. We consider that more efforts should be done with regards to the use of recurrent DL architectures to take advantage of the temporal component, i.e. using LSTM or GRU cells. Furthermore in Liu et al. (2020) the hybrid architecture C-LSTM was compared with CNN and LSTM for seizure detection, obtaining C-LSTM the best performance. In our opinion this result should motivate to use C-LSTM in the new works. With this in mind, and since the *no free lunch theorem* establishes that there is no best algorithm to every problem, it should motivate us to use more than one algorithm when a new research work is done because we need (i) to achieve the best possible performance and (ii) to compare the performance of these algorithms in the diagnosis or prognosis of a specific illnesses.

When using CNNs, an important choice is how to sort EEG channels because adjacent channels will be processed together due to 2D convolution operation. Usually, EEG data is presented to CNNs as a 2D array, where the X-axis is the time and the Y-axis represents the EEG channels (see Fig. 14). To carry out a 2D convolution, a kernel with size *nxm* needs to be defined, where *n* is the number of consecutive “pieces” of time taken into account, and *m* is the number of channels which are convoluted together. Hence, the order of the channels will condition the accuracy of the network. A common solution is to sort the channels regarding their correlations. In our opinion, it may not be a good solution because channels which are not correlated with another would be treated in an isolated manner and would not be put together with another channel with no correlation. To avoid this issue, we propose the following method: firstly split the EEG record (2D array) in several equal-sized 2D arrays (called records). Secondly, build a 3D array by appending all those arrays where the dimensions are X-time, Y-channels,

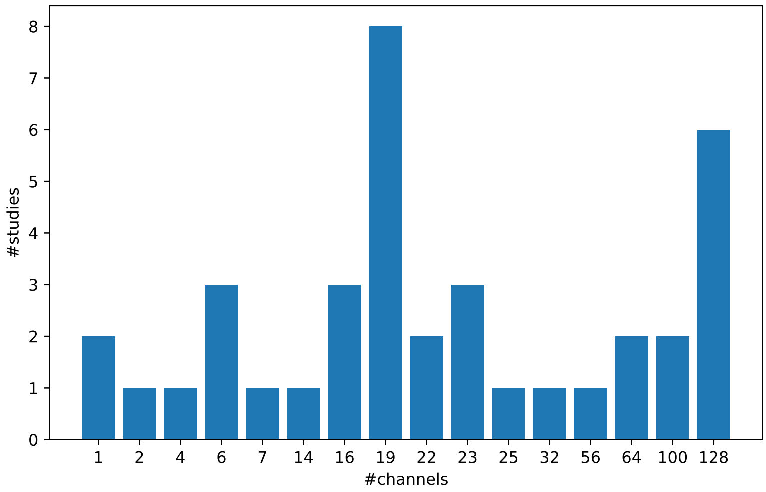


Fig. 10 Frequency of number of channels used to record EEG signals

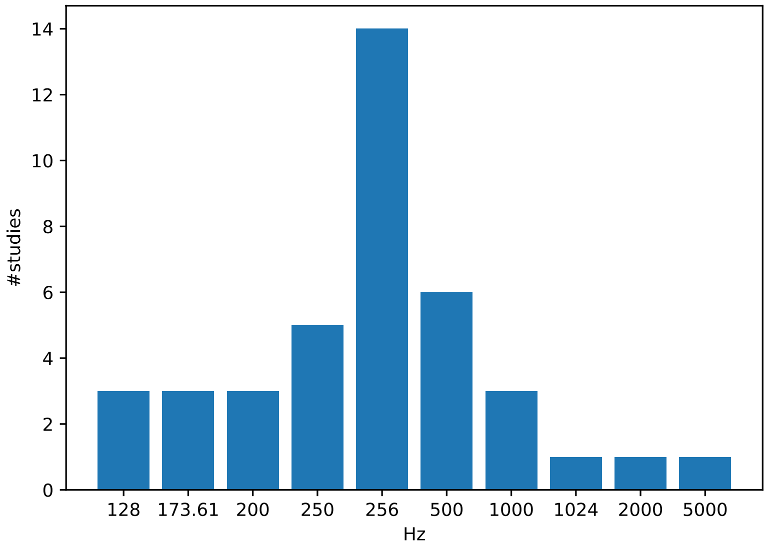


Fig. 11 Frequency of the sampling rate used to record EEG signals

Z-record. Finally, compound a 3D kernel to carry out 3D convolutions to process every channel in an isolated manner. In Fig. 15 we show this process.

A major advantage of the use of DL is the possibility of using pre-trained networks to carry out a totally new but related task. This can save computing time and enables us to perform the training stage with less data. This concept is called *transfer learning* Yosinski et al. (2014). In our opinion, using transfer learning to process EEG data should be used when the number of EEG records for training is low, e.g. due to the lack of patients to accomplish EEG records. An example of application could be training a network to classify patients with sleep disorders using available public datasets (see

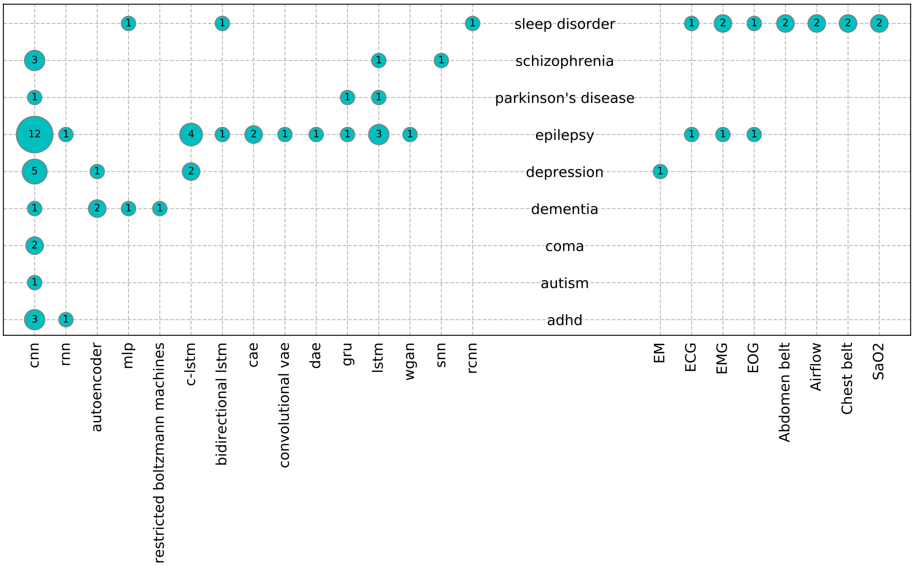


Fig. 12 Mapping with the frequency of studies combining categories from RQ1 with RQ2 and RQ3

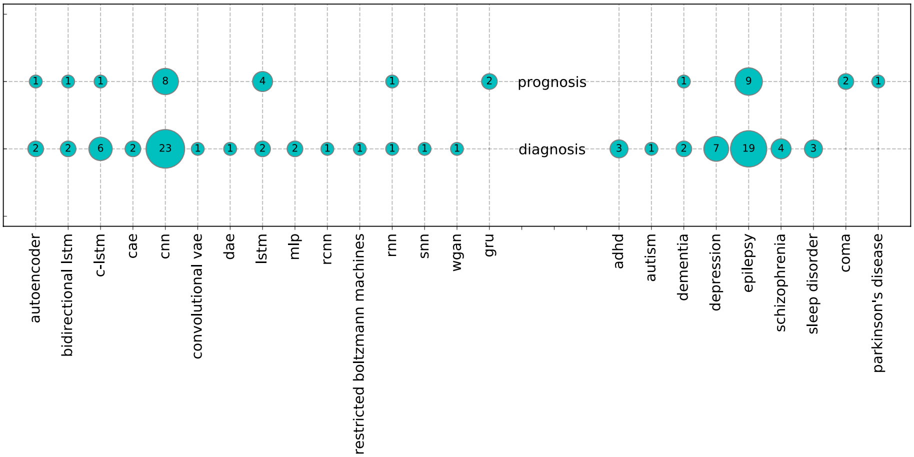


Fig. 13 Mapping with the frequency of studies combining the nature of the disorder detection (diagnosis, prognosis) with RQ1 and RQ2

Table 6), and then training a classifier to diagnosis another mental disorder, e.g. depression, by using transfer learning over this pre-trained network.

Considering the mental disorders which have been diagnosed or prognosed with DL (see Fig. 4a), and according to Wittchen et al. (2011) where *the most prevalent 12 month disorders in Europe 2010* were presented, we have identified a set of mental disorders with high prevalence which are not been diagnosed or prognosed by means of EEG and DL. Those mental disorders are: i) anxiety disorders which is the most prevalent 12-month disorder in Europe affecting to 61.5M of persons, ii) alcohol dependence 14.6M, iii) post traumatic stress disorder 7.7M, iv) personality disorders 6.3M, v) mental retardation 4.2M,

vii) obsessive compulsive disorder 2.9M, viii) conduct disorder 2.1M, ix) eating disorders 1.5M and x) cannabis dependence 1.4M. Because the amount of people affected and the promising results employing DL with EGG, efforts in diagnosis and prognosis of these mental disorders by means EEG and DL should be done. For a complete reference list of mental disorders see Association (2013).

Regarding to recording and processing other biometrical data alongside EEG, we can say that this is an open field, as we showed in our mapping (see Fig. 12), only the 8.7% of the studies have used at least one additional kind of biometric data. In the same manner, only in the 33% of the found mental disorders (Sleep Disorder, Epilepsy and Depression), EEG has been combined with at least another additional biomarker. In this sense, Zhu et al. (2019) showed promising results by joining EEG and EM to increase the performance of the classifier for mild depression recognition. In addition to the biometric signals used in the selected studies, there are previous works where other different signals were combined with EEG: in Dupuy et al. (2014) EEG and electrodermal activity (EDA) was employed to explain ADHD symptoms. In Gallagher et al. (2008), they used EEG with near-infrared spectroscopy at the same time to locate the ictal onset zone in a epileptic patient. It is important to notice that, as Zhu et al. (2019) mention, if some of these biometric signals are recorded, an accurate synchronization between these signals and EEG data is essential for simultaneous analysis.

DL are blackbox algorithms, that is, their predictions are not self-explanatory. Giving an explanation to the outcome can increase the confidence in their predictions by showing that EEG electrodes and their amplitude are relevant to classify a mental disorder. This is known as the *inverse problem*. Providing that explanations could help doctors to trust in DL-based tools. In Shahin et al. (2017), it was used the Gradient-Weighted Class Activation Mapping (Grad-CAM) to identify which EEG features have more importance for the network to predict a favourable outcome. To explain the CNNs predictions SHAP Lundberg and Lee (2017) can be used, which is a game-theory-based approach aimed at explaining ML algorithms. In their library Lundberg et al. (2016) two methods for explaining CNNs are available. For a complete review of explainable methods for DL refer to Xie et al. (2020).

The validation stage enables us to know the performance of the proposed DL framework. As we showed, the only 39.13% of the works realized an intra-subject validation (see

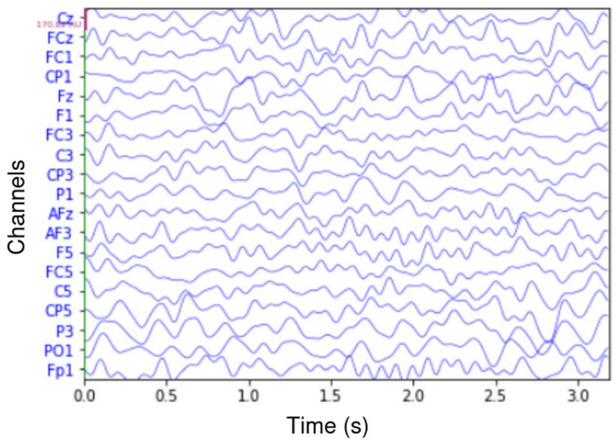


Fig. 14 Example of EEG record with 19 channels

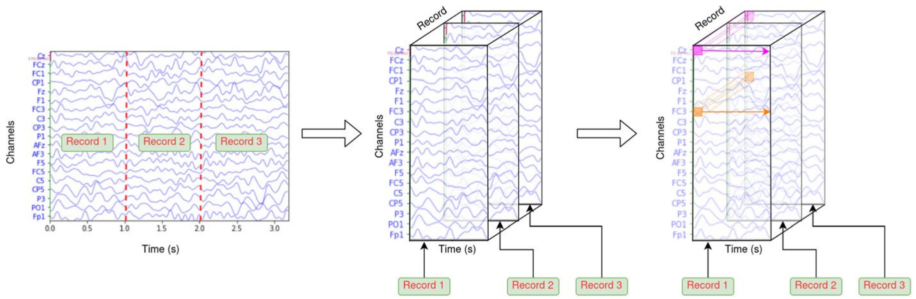


Fig. 15 Preprocessing EEG data to carry out 3D convolutions over every channel in an isolated manner. Pink and orange cuboids are two examples of 3D kernels convolutions

Fig. 7b). This type of validation offers a realistic performance of the algorithm because the test set is composed by different never-seen subjects, which were not used for training. In contrast, an inter-subject validation offers an artificial high level of performance because it do not show how the algorithm will generalize its outcomes for a new patient. Hence, we encourage the use of intra-subject validation. Furthermore, it is important to emphasise the significance of being able to objectively compare the classification results among studies. This comparison enables us to know which are the algorithms that achieve the best performance by mental disorder. Unfortunately, only a preliminary comparison of results has been showed because of (i) there is not a common set of metrics used in the studies to measure the performance and (ii) there is not a reference dataset by mental disorder. In this respect, only in 4 out 9 mental disorders (epilepsy, schizophrenia, sleep disorder and depression) there is available a public dataset (see Fig. 9b). It is worth noting that, these 4 mental disorders are the most common across the studies (see Fig. 4a). Nevertheless, for future works (i) only intra-subject validation should be performed, (ii) a reference dataset should be created by mental disorder and (iii) a standard set of metrics should be used for objectively measure the performance.

With regards to the reproducibility of the results, we have found out that the majority of the studies do not share their source code, thus making their outcome non-reproducible. Besides, some of the works neither offer a good detailed explanation of the DL architecture nor presents the set-up process for the EEG data acquisition process, such as the layout used to place the electrodes, sample frequency, or subject characteristics. In the same manner, the captured EEG data rarely is shared but as we noticed, sometimes privacy restrictions do not allow researchers to make it publicly available. We consider that a detailed explanation of the selected patients, capturing EEG data process, preprocessing (if required), DL architecture and the type of training and validation used should be provided. Additionally, source code and data ideally should be shared. All of these pieces of information would allow the progress beyond the state of the art in this field progress.

Finally, we have to mention that a limitation of our study is that it is focused on diagnosis or prognosis of mental disorders, hence the results presented in our work could not be generalized to other fields where EEG is used.

6 Conclusions

There is a growing interest in using EEG to train DL algorithms for diagnosis and prognosis of mental disorders because of the promising results observed and the continuous improvements achieved by using DL techniques. In this work, a Systematic Mapping Study is presented for clearly knowing what is already done and what could be done within this field. Four RQs were proposed in Sect. 4.1 and four public databases were queried to retrieve relevant works. 46 primary studies were finally selected and after a full read, there were extracted relevant keywords which were grouped in categories to answer the RQs.

The main results are as follows: (i) nine mental disorders categories which have been diagnosed or prognosed by using EEG data and DL, being Epilepsy the most common one, present in 47.83% of the works, (ii) 13 categories of DL algorithms, being CNN the most frequent appearing in 48.28% of the papers, (iii) eight additional biomarkers were combined with EEG data in 8.7% of the studies and (iv) in 50.98% of the works, it were used a public EEG dataset.

Additional discoveries were done by combining RQs: (i) epilepsy is the mental disorder where more DL techniques have been tried out and (ii) additional biomarkers have only been employed in sleep disorder, epilepsy and depression. Finally in the discussion section, insights and recommendations were done: (i) without taking account epilepsy, for the rest of mental disorders there are many gaps in the use of DL techniques. For future works, we recommend to focus in the use of LSTM and GRU cells, as well as the hybrid C-LSTM networks because of the continuous nature of the EEG signal, (ii) some of the most prevalent mental disorders such as anxiety, alcohol dependence, post traumatic stress, personality disorders, mental retardation, obsessive compulsive disorder, conduct disorder, eating disorders and cannabis dependence have not been diagnosed or prognosed by using DL and EEG, (iii) combining additional biomarkers with EEG is a promising approach to improve the classification results, yet more efforts should be done in this sense. Other recommendations were also provided: (i) explaining the DL model results can be achieved for some DL techniques by using of available frameworks. It can help in the adoption of DL techniques by providing the most relevant EEG channels for the model predictions, (ii) we encourage to perform intra-subject validation for getting a more realistic performance of the algorithms, thus avoiding artificial high accuracy, (iii) to objectively compare performance between studies, it is important to define a reference dataset by mental disorder, as well as the use of a standard set of metrics for measuring the DL performance, and (iv) we suggest paying more attention to ease the reproducibility of their results by providing more details of the selected patients, capturing EEG data process, preprocessing (if needed), DL architecture and the type of training and validation used as well as sharing both the source code and the recorded data in an anonymous manner.

As a limitation, our study is focused on diagnosis or prognosis of mental disorders. Hence the results presented in our work could not be generalized to other fields where EEG is used.

Appendix A: List of acronyms

See Table 7.

Table 7 List of acronyms

Text	Acronym
Attention Deficit Hyperactivity Disorder	ADHD
Autoencoder	AE
Artificial Intelligence	AI
Convolutional Long-Short Term Memory	C-LSTM
Convolutional Autoencoder	CAE
Convolutional Neural Network	CNN
Convolutional Variational Autoencoder	Convolutional VAE
Denoising Autoencoder	DAE
Deep Learning	DL
Deep Neural Network	DNN
Electrocardiogram	ECG
Electrodermal Activity	EDA
Electroencephalogram	EEG
Exponential Linear Unit	ELU
Eye Movement	EM
Electromyography	EMG
Electrooculography	EOG
Feature Engineering	FE
Fast Region-based Convolutional Network	Fast R-CNN
Gated Recurrent Unit	GRU
Gradient-Weighted Class Activation Mapping	Grad-CAM
Long-Short Term Memory	LSTM
Multimodal Denoising Autoencoder	MDAE
Multilayer Perceptron	MLP
Magnetic Resonance Imaging	MRI
Neural Network	NN
Positron Emission Tomography	PET
Restricted Boltzmann Machines	RBM
Recurrent Convolutional Neural Network	RCNN
Recurrent Neural Network	RNN
Area Under the Receiver Operator Characteristic	ROC AUC
Research Question	RQ
Rectified Linear Unit	ReLU
Stochastic Gradient Descent	SGD
Systematic Mapping Study	SMS
Siamese Neural Network	SNN
Oxygen Saturation	SaO2
Hyperbolic Tangent	Tanh
Wasserstein Generative Adversarial Network	WGAN

Appendix B: Information about extracted papers

See Tables 8, 9, 10

Table 8 Information extracted from papers (part 1)

Reference	Mental Disorder	Type of disorder detection	Source dataset	Additional biomarkers
Liu et al. (2020)	Epilepsy	Diagnosis	UCI	–
Ali et al. (2020)	Autism	Diagnosis	Other	–
Mumtaz and Qayyum (2019)	Depression	Diagnosis	Ad-hoc	–
Abdelhameed and Bayoumi (2019)	Epilepsy	Diagnosis	Bonn University CHB-MIT	–
Thara et al. (2019)	Epilepsy	diagnosis prognosis	Bonn University	–
Li et al. (2019a)	Depression	Diagnosis	Ad-hoc	–
Ay et al. (2019)	Depression	Diagnosis	Ad-hoc	–
Bi and HaiboWang (2019)	Dementia	Diagnosis	Other	–
Dominic et al. (2019)	Depression	Diagnosis	Predict	–
Baloglu and Yildirim (2019)	Epilepsy	Diagnosis	Bonn University	–
Zhu et al. (2019)	Depression	Diagnosis	Ad-hoc	EM
Ruffini et al. (2019)	Parkinson'S Disease	Prognosis	Ad-hoc	–
Liang et al. (2019)	Epilepsy	diagnosis prognosis	CHB-MIT	–
Clarke et al. (2019)	Epilepsy	Diagnosis	Ad-hoc	–
Emami et al. (2019)	Epilepsy	Diagnosis	Ad-hoc	–
Liu et al. (2019)	Epilepsy	Diagnosis	Freiburg University	–
Wei et al. (2018)	Epilepsy	diagnosis prognosis	Ad-hoc	–
Acharya et al. (2018b)	Epilepsy	diagnosis prognosis	Bonn University	–
Tsiouris et al. (2018)	Epilepsy	Prognosis	CHB-MIT	–
Acharya et al. (2018a)	Depression	Diagnosis	Ad-hoc	–
Sun et al. (2018)	Epilepsy	Prognosis	Kaggle	–

Table 8 (continued)

Reference	Mental Disorder	Type of disorder detection	Source dataset	Additional biomarkers
Shahin et al. (2017)	Sleep Disorder	Diagnosis	Ad-hoc	–
Morabito et al. (2017)	Dementia	Prognosis	Ad-hoc	–
Ieracitano et al. (2020)	Dementia	Diagnosis	Ad-hoc	–
Tjepkema-Cloostermans et al. (2019)	Coma	Prognosis	Ad-hoc	–
Oh et al. (2019)	Schizophrenia	Diagnosis	Other	–
Vahid et al. (2019)	Adhd	Diagnosis	Ad-hoc	–
Bernardo et al. (2018)	Epilepsy	Prognosis	Ad-hoc	–
Wei et al. (2019)	Epilepsy	Diagnosis	CHB-MIT	EOG EMG ECG
Wen and Zhang (2018)	Epilepsy	Diagnosis	Bonn University CHB-MIT	–
Hussein et al. (2019)	Epilepsy	diagnosis prognosis	Bonn University	–
Khan et al. (2018)	Epilepsy	diagnosis prognosis	Ad-hoc CHB-MIT	–
Yuan et al. (2019)	Epilepsy	Diagnosis	CHB-MIT	–
Biswal et al. (2018)	Sleep Disorder	Diagnosis	Ad-hoc SHHS	Chest belt Abdomen belt SaO2 Airflow EMG
Chen et al. (2019)	Adhd	Diagnosis	Ad-hoc	–
Türk and Özerdem (2019)	Epilepsy	Diagnosis	Bonn University	–
Jonas et al. (2019)	Coma	Prognosis	Ad-hoc	–

Table 8 (continued)

Reference	Mental Disorder	Type of disorder detection	Source dataset	Additional biomarkers
Fürbass et al. (2020)	Epilepsy	Diagnosis	Ad-hoc Temple University	–
Phang et al. (2020)	Schizophrenia	Diagnosis	Lomonosov Moscow State University	–
Calhas et al. (2020)	Schizophrenia	Diagnosis	Lomonosov Moscow State University	–
Dubreuil-Vall et al. (2020)	Adhd	Diagnosis	Ad-hoc	–
Li et al. (2020a)	Depression	Diagnosis	Ad-hoc	–
Li et al. (2020b)	Epilepsy	Diagnosis	Freiburg University	–
Sahu et al. (2020)	Epilepsy	Diagnosis	Bonn University	–
Naira et al. (2019)	Schizophrenia	Diagnosis	Lomonosov Moscow State University	–
Warrick et al. (2019)	Sleep Disorder	Diagnosis	PhysioNet CinC 2018 Challenge	EOG EMG ECG Chest belt Abdomen belt SaO2 Airflow

Table 9 Information extracted from papers (part 2)

Reference	Sampling rate EEG data acquisition (Hz)	# Electrodes	DL technique	# Layers	Batch normalization	Activation functions
Liu et al. (2020)	173.61	128	C-LSTM	5	Yes	relu sigmoid tanh softmax
Ali et al. (2020)	256	16	CNN	6	Yes	relu softmax
Mumtaz and Qayyum (2019)	256	19	CNN C-LSTM	11 14	No	Sigmoid
Abdelhameed and Bayoumi (2019)	173.61 256	1	Convolutional VAE	8	Yes	relu sigmoid softmax
Thara et al. (2019)	–	–	Bidirectional LSTM	4	No	sigmoid tanh softmax
Li et al. (2019a)	250	128	CNN	7 8	No	relu softmax
Ay et al. (2019)	256	1	C-LSTM	–	No	relu softmax
Bi and HaiboWang (2019)	500	64	CNN RBM	3	No	–
Dominic et al. (2019)	500	64	CNN	11	No	–
Baloglu and Yildirim (2019)	173.61	100	C-LSTM	9	No	softmax relu
Zhu et al. (2019)	250	128	multimodal denoising autoencoder	–	No	–

Table 9 (continued)

Reference	Sampling rate EEG data acquisition (Hz)	# Electrodes	DL technique	# Layers	Batch normalization	Activation functions
Ruffini et al. (2019)	–	14	CNN LSTM GRU	–	No	relu softmax
Liang et al. (2019)	256	–	C-LSTM	18	No	Softmax
Clarke et al. (2019)	256	32	CNN	10	No	–
Emami et al. (2019)	1000 500	19	CNN	16	No	–
Liu et al. (2019)	256	128	S-transform and CNN	15	Yes	softmax relu
Wei et al. (2018)	500	22	CNN	9	No	relu softmax
Acharya et al. (2018b)	–	100	CNN	13	No	relu softmax
Tsiouris et al. (2018)	256	–	LSTM	2	No	relu softmax
Acharya et al. (2018a)	256	4	CNN	13	No	leakyrelu softmax
Sun et al. (2018)	5000	–	CNN LSTM GRU RNN	–	No	–
Shahin et al. (2017)	200	6	MLP	–	No	Softmax
Morabito et al. (2017)	–	19	stacked autoencoder	2	No	Sigmoid

Table 9 (continued)

Reference	Sampling rate EEG data acquisition (Hz)	# Electrodes	DL technique	# Layers	Batch normalization	Activation functions
Ieracitano et al. (2020)	1024	19	MLP autoencoder	1	No	Softmax
Tjepkema-Cloostermans et al. (2019)	–	19	CNN	–	No	relu softmax
Oh et al. (2019)	250	19	CNN	11	Yes	Leakyrelu
Vahid et al. (2019)	500	56	CNN	4	Yes	linear elu softmax
Bernardo et al. (2018)	2000	–	CNN	–	No	–
Wei et al. (2019)	256	23	CNN WGAN	–	No	relu softmax tanh leakyrelu
Wen and Zhang (2018)	256	23	Convolutional autoencoder	–	No	–
Hussein et al. (2019)	–	–	LSTM	–	No	Softmax
Khan et al. (2018)	256 256	22	CNN	–	No	–
Yuan et al. (2019)	256	23	denoising autoencoder (DAE) convolutional autoencoder	–	Yes	Relu
Biswal et al. (2018)	200	6 2	RCNN (Recurrent Convolutional Neural Network)	–	No	–
Chen et al. (2019)	1000	128	CNN	–	No	Softmax
Türk and Özerdem (2019)	–	–	CNN	–	No	–

Table 9 (continued)

Reference	Sampling rate EEG data acquisition (Hz)	# Electrodes	DL technique	# Layers	Batch normalization	Activation functions
Jonas et al. (2019)	250 1000	19	CNN	–	No	–
Fürbass et al. (2020)	–	19 25	Fast Region-based Convolutional Network (Fast R-CNN)	–	No	–
Phang et al. (2020)	128	16	CNN LSTM	–	Yes	softmax relu
Calhas et al. (2020)	128	–	Siamese neural network (SNN)	–	No	–
Dubreuil-Vall et al. (2020)	500	7	CNN RNN	4	No	–
Li et al. (2020a)	250	128	CNN	–	No	softmax relu
Li et al. (2020b)	–	–	C-LSTM	–	No	–
Sahu et al. (2020)	–	–	CNN	–	No	–
Naira et al. (2019)	128	16	CNN	–	No	Softmax
Warrick et al. (2019)	200	6	Bidirectional LSTM	–	Yes	–

Table 10 Information extracted from papers (part 3)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Liu et al. (2020)	Yes	Adam	Sensitivity Precision F1-Score	No	–	2020	Italy Germany China Sweden
Ali et al. (2020)	Yes	Adam	Accuracy	No	–	2020	Malaysia
Mumtaz and Qayyum (2019)	Yes	adam	Accuracy Precision Sensitivity F1-Score	No	–	2019	Pakistan France
Abdelhameed and Bayoumi (2019)	No	Adam RMSProp	Accuracy Sensitivity Specificity Precision F1-Score	No	–	2019	USA
Thara et al. (2019)	No	Adam	Accuracy Precision Sensitivity F1-Score Roc-Auc	No	–	2019	India USA
Li et al. (2019a)	Yes	Adam	Accuracy	Yes	84.75	2019	China
Ay et al. (2019)	Yes	Adam	Accuracy Precision Sensitivity Specificity	No	–	2019	Malaysia India Singapore Turkey
Bi and HaiboWang (2019)	No	Adam	Accuracy Roc-Auc	No	–	2019	China

Table 10 (continued)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Dominic et al. (2019)	Yes	Adam	Accuracy Sensitivity Specificity	No	–	2019	India
Baloglu and Yildirim (2019)	Yes	Adam	Accuracy	No	–	2019	Turkey
Zhu et al. (2019)	No	–	–	No	–	2019	China
Ruffini et al. (2019)	Yes	–	Roc-Auc Accuracy	Yes	81.0	2019	Canada Spain
Liang et al. (2019)	Yes	Adam	Accuracy Sensitivity Specificity	Yes	92.4	2019	China
Clarke et al. (2019)	No	Adam	Sensitivity False Positive Rate Precision	Yes	–	2019	Australia
Emami et al. (2019)	No	Adam	Sensitivity Specificity Roc-Auc	Yes	–	2019	Japan
Liu et al. (2019)	Yes	Adam	Accuracy Sensitivity Specificity Roc-Auc	No	–	2019	China

Table 10 (continued)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Wei et al. (2018)	Yes	Adam	Accuracy Sensitivity Specificity False Negative Rate False Positive Rate	No	–	2018	China
Acharya et al. (2018b)	No	–	Accuracy Sensitivity Specificity	No	–	2018	Malaysia Singapore USA
Tsiouris et al. (2018)	Yes	Adam	Accuracy Sensitivity Specificity	No	–	2018	Greece
Acharya et al. (2018a)	Yes	Adam	Accuracy Sensitivity Specificity	No	–	2018	Malaysia India Singapore USA
Sun et al. (2018)	No	–	Roc-Auc	No	–	2018	China
Shahin et al. (2017)	No	–	Accuracy Sensitivity Precision	Yes	92.0	2017	Qatar Germany
Morabito et al. (2017)	No	–	Accuracy Sensitivity Specificity	No	–	2017	Italy

Table 10 (continued)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Ieracitano et al. (2020)	No	–	Precision Sensitivity F1-Score	No	–	2020	Italy UK
Tjepkema-Cloostermans et al. (2019)	No	Adam	Accuracy Sensitivity False Positive Rate Roc-Auc	Yes	–	2019	Portugal Netherlands
Oh et al. (2019)	Yes	Adam	Accuracy Sensitivity Precision Specificity	Yes	81.26	2019	Malaysia Singapore USA
Vahid et al. (2019)	Yes	Adam	Accuracy	Yes	83.0	2019	Germany
Bernardo et al. (2018)	No	–	Sensitivity Specificity Roc-Auc False Positive Rate	Yes	–	2018	USA
Wei et al. (2019)	Yes	Adam RMSprop	Sensitivity Specificity Accuracy	Yes	83.27	2019	China
Wen and Zhang (2018)	No	–	Accuracy	No	–	2018	China
Hussein et al. (2019)	No	Adam	Accuracy Sensitivity Specificity	No	–	2019	Canada USA

Table 10 (continued)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Khan et al. (2018)	Yes	SGD	Roc-Auc Sensitivity	No	–	2018	USA
Yuan et al. (2019)	Yes	Adadelta	F1-Score Roc-Auc Prc Accuracy	Yes	94.37	2019	China USA
Biswal et al. (2018)	No	SGD	Accuracy	Yes	88.2	2018	USA
Chen et al. (2019)	Yes	Nadan	Accuracy Roc-Auc	No	–	2019	China
Türk and Özerdem (2019)	No	adadelta	Accuracy Sensitivity Specificity F1-Score	No	–	2019	Turkey
Jonas et al. (2019)	No	–	Accuracy Roc-Auc Sensitivity Specificity	Yes	87.04	2019	Switzerland
Fürbass et al. (2020)	No	–	Sensitivity Specificity Accuracy	Yes	80.0	2020	Austria Denmark
Phang et al. (2020)	Yes	–	Accuracy Precision Sensitivity Specificity	No	–	2020	Saudi Arabia Malaysia

Table 10 (continued)

Reference	Dropout	Optimizator	Metrics of performance	Validation intra-subject	Accuracy	Year	Countries
Calhas et al. (2020)	Yes	Adam	Accuracy Sensitivity Specificity	Yes	95.0	2020	Portugal Spain
Dubreuil-Vall et al. (2020)	No	Adam	Accuracy Roc-Auc	No	–	2020	USA Spain
Li et al. (2020a)	No	Adam	Accuracy	Yes	80.74	2020	China
Li et al. (2020b)	No	Adam	Accuracy Sensitivity Specificity	No	–	2020	China USA
Sahu et al. (2020)	No	–	Accuracy Roc-Auc	No	–	2020	Switzerland Malaysia India
Naira et al. (2019)	Yes	–	Sensitivity Specificity Accuracy	No	–	2019	Peru
Warrick et al. (2019)	No	–	Prc	Yes	–	2019	Canada USA Venezuela

Acknowledgements This work has been funded by the ECLIPSE project (RTI2018-094283-B-C32) from the Spanish Ministry of Science, Innovation and Universities.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdelhameed AM, Bayoumi M (2019) Semi-supervised EEG signals classification system for epileptic seizure detection. IEEE Signal Process Lett 26(12):1922–1926. https://doi.org/10.1109/LSP.2019.2953870

Acharya UR et al (2012) Automated diagnosis of normal and alcoholic EEG signals. Int J Neural Syst. https://doi.org/10.1142/S0129065712500116

Acharya UR et al (2018a) Automated EEG-based screening of depression using deep convolutional neural network. Comput Methods Program Biomed 161:103–113. https://doi.org/10.1016/j.cmpb.2018.04.012

Acharya UR et al (2018b) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med 100:270–278. https://doi.org/10.1016/j.compbiomed.2017.09.017

Ali NA et al (2020) Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm. IAES Int J Artific Intell 9(1):91–99

Andreasen NC (1982) Negative symptoms in schizophrenia: definition and reliability. Arch Gener Psychiat 39(7):784–788. https://doi.org/10.1001/archpsyc.1982.04290070020005

Andrzejak RG et al (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Phys Rev E. https://doi.org/10.1103/PhysRevE.64.061907

Arjovsky M, Soumith C, Léon B (2017) Wasserstein GAN. In: arXiv: 1701.07875

Association American Psychiatric et al (2013) Diagnostic and statistical manual of mental disorders (DSM-5 R). American Psychiatric Pub

Ay B et al (2019) Automated depression detection using deep representation and sequence learning with EEG signals. J Med Syst. https://doi.org/10.1007/s10916-019-1345-y

Baloglu UB, Yildirim Ö (2019) Convolutional long-short term memory networks model for long duration EEG signal classification. J Mech Med Biol 19(1):1–21. https://doi.org/10.1142/S0219519419400050

Bernardo D et al (2018) Visual and semi-automatic non-invasive detection of interictal fast ripples: a potential biomarker of epilepsy in children with tuberous sclerosis complex. Clin Neurophys 129(7):1458–1466. https://doi.org/10.1016/j.clinph.2018.03.010

Bi XH (2019) Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. Neural Netw 114:119–135. https://doi.org/10.1016/j.neunet.2019.02.005

Biswal S et al (2018) Expert-level sleep scoring with deep neural networks. J Am Med Inf Assoc 25(12):1643–1650. https://doi.org/10.1093/jamia/ocy131

Bourlard H, Kamp Y (1988) Auto-association by multilayer perceptrons and singular value decomposition. Biol Cybern 59(4–5):291–294

Bromley J et al (1994) Signature verification using a "siamese" time delay neural network. Adv Neural Inf Process Syst 7(4):669–688

Budgen D et al (2008) Using mapping studies in software engineering. PPIG. 8:195–204

Calhas D, Enrique R, Rui H (2020) On the use of pairwise distance learning for brain signal classification with limited observations. Artific Intell Med. https://doi.org/10.1016/j.artmed.2020.101852.

Chen H, Yan S, Xiaoli L (2019) A deep learning framework for identifying children with ADHD using an EEG-based brain network. Neurocomputer 356:83–96. https://doi.org/10.1016/j.neucom.2019.04.058

Cho K, et al (2014) On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259

Clarke S, et al (2019) Computer-assisted EEG diagnostic review for idiopathic generalized epilepsy. Epilepsy Behav. https://doi.org/10.1016/j.yebeh.2019.106556. http://www.sciencedirect.com/science/article/pii/S1525505019306018

Craik A, Yongtian H, Contreras-Vidal JL (2019) Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng. https://doi.org/10.1088/1741-2552/ab0ab5

Croskerry P (2003) The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 78(8):775780. https://doi.org/10.1097/00001888-200308000-00003

Dash D et al (2012) Ambulatory EEG: a cost-effective alternative to inpatient video-EEG in adult patients. Epileptic Disorders 14(3):290–297. https://doi.org/10.1684/epd.2012.0529

Dean DA et al (2016) Scaling up scientific discovery in sleep medicine: the national sleep research resource. Sleep 39(5):1151–1164. https://doi.org/10.5665/sleep.5774

Domingues I, et al (2019) Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. Artific Intell Rev, pp 1–68

Dominic A, Aswathy KJ, Surekha MV (2019) Deep learning in computer aided diagnosis of MDD. Int J In-novat Technol Explor Eng 8(6):464–468

Dubreuil-Vall L, Giulio R, Camprodon JA (2020) Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. Front Neurosci. https://doi.org/10.3389/fnins.2020.00251

Dupuy FE et al (2014) EEG and electrodermal activity in girls with attentionde deficit/hyperactivity disorder. Clin Neurophysiol 125(3):491–499. https://doi.org/10.1016/j.clinph.2013.09.007

Emami A, et al (2019) Seizure detection by convolutional neural networkbased analysis of scalp electroencephalography plot images. NeuroImage: Clin 22:101684. https://doi.org/10.1016/j.nicl.2019.101684. http://www.sciencedirect.com/science/article/pii/S2213158219300348

Faulkner HJ, Hisatomi A, Armin M (2012) The utility of prolonged outpatient ambulatory EEG. Seizure 21(7):491–495. https://doi.org/10.1016/j.seizure.2012.04.015

Ferri CP et al (2005) Global prevalence of dementia: a Delphi consensus study. Lancet 366(9503):2112–2117. https://doi.org/10.1016/S0140-6736(05)67889-0

Fisher RS, et al (2005) Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). https://doi.org/10.1111/j.0013-9580.2005.66104.x.

Fürbass F et al (2020) An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: validation against the diagnostic gold standard. Clin Neurophysiol 131(6):1174–1179. https://doi.org/10.1016/j.clinph.2020.02.032

Gallagher A et al (2008) Non-invasive pre-surgical investigation of a 10 yearold epileptic boy using simultaneous EEG-NIRS. Seizure 17(6):576–582. https://doi.org/10.1016/j.seizure.2008.01.009

Geschwind MD, et al (2008) Rapidly progressive dementia. https://doi.org/10.1002/ana.21430

Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448

Gorbachevskaya NN, Borisov S (2002) EEG data of healthy adolescents and adolescents with symptoms of schizophrenia. http://brain.bio.msu.ru/eeg_schizophrenia.htm (visited on06/10/2020)

Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw 18(5–6):602–610

Halász P et al (2004) The nature of arousal in sleep. J Sleep Res 13(1):1–23

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hosseini MP, Hosseini A, Ahi K (2020) A review on machine learning for EEG signal processing in bioengineering. IEEE Rev Biomed Eng. https://doi.org/10.1109/RBME.2020.2969915

Hu L, Zhang Z (2019) EEG signal processing and feature extraction. Springer Singapore, pp. 1–437. https://doi.org/10.1007/978-981-13-9113-2

Hussein R et al (2019) Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. Clin Neurophys 130(1):25–37. https://doi.org/10.1016/j.clinph.2018.10.010

Ieracitano C et al (2020) A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. Neural Netw 123:176–190. https://doi.org/10.1016/j.neunet.2019.12.006

Jankovic J (2008). Parkinson's disease: Clinical features and diagnosis. https://doi.org/10.1136/jnnp.2007.131045

Jeste SS, Frohlich J, Loo SK (2015) Electrophysiological biomarkers of diagnosis and outcome in neurodevelopmental disorders. https://doi.org/10.1097/WCO.0000000000000181. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00019052-201504000-00005

Johnson C, et al (2007) Identification and evaluation of children with autism spectrum disorders. https://doi.org/10.1542/peds.2007-2361.

Jonas S, et al (2019) EEG-based outcome prediction after cardiac arrest with convolutional neural networks: performance and visualization of discriminative features. Human Brain Mapping 40(16):4606–4617. https://doi.org/10.1002/hbm.24724.

Jurcak V, Daisuke T, Ippeita D (2007) 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. NeuroImage 34(4):1600–1611. https://doi.org/10.1016/j.neuroimage.2006.09.024

Khan H et al (2018) Focal onset seizure prediction using convolutional networks. IEEE Trans Biomed Eng 65(9):2109–2118. https://doi.org/10.1109/TBME.2017.2785401

Kieling R, Rohde LA (2012) ADHD in children and adults: diagnosis and prognosis. Curr Top Behav Neurosci 9:1–16. https://doi.org/10.1007/7854_2010_115

Kingma DP, Max W (2014) Auto-encoding variational bayes. In: 2nd international conference on learning representations, ICLR 2014. arXiv: 1312.6114

Kitchenham B, Stuart C (2007) Guidelines for performing systematic literature reviews in software engineering

Kitchenham Barbara A, Budgen D, Pearl Brereton O (2011) Using mapping studies as the basis for further research-a participant-observer case study. Inf Softw Technol 53(6):638–651

LeCun Y et al (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

Li X, et al (2019a) Depression recognition using machine learning methods with different feature generation strategies. Artific Intell Med https://doi.org/10.1016/j.artmed.2019.07.004. http://www.sciencedirect.com/science/article/pii/S0933365719300296

Li Y et al (2019b) Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach. Knowl-Based Syst 164:96–106

Li X et al (2020a) A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography. Front Neurosci. https://doi.org/10.3389/fnins.2020.00192

Li Y et al (2020b) Automatic seizure detection using fully convolutional nested LSTM. Int J Neural Syst. https://doi.org/10.1142/S0129065720500197

Liang W et al (2019) Scalp EEG epileptogenic zone recognition and localization based on long-term recurrent convolutional network. Neurocomputing. https://doi.org/10.1016/j.neucom.2018.10.108

Liu G, Weidong Z, Minxing G (2019) Automatic seizure detection based on s-transform and deep convolutional neural network. Int J Neural Syst. https://doi.org/10.1142/S0129065719500242

Liu Y et al (2020) Deep C-LSTM neural network for epileptic seizure and tumor detection using high-dimension EEG signals. IEEE Access 8:37495–37504. https://doi.org/10.1109/ACCESS.2020.2976156

Lundberg SM, Lee SI (2016) SHapley Additive exPlanations. https://github.com/slundberg/shap

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Guyon I, et al (eds.), Advances in neural information processing systems vol 30. Curran Associates, Inc., pp 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Masci J, et al (2011) Stacked convolutional auto-encoders for hierarchical feature extraction. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in Bioinformatics), vol 6791 LNCS. PART 1, pp 52–59. https://doi.org/10.1007/978-3-642-21735-7_7., https://www.researchgate.net/publication/221078713

Mcloughlin G, Makeig S, Tsuang MT (2014) In search of biomarkers in psychiatry: EEG-based measures of brain function. Am J Med Genet B Neuropsych Genet 165(2):111–121. https://doi.org/10.1002/ajmg.b.32208

Mecarelli O (2019) Electrode placement systems and montages. In: Mecarelli O (eds) Clinical electroencephalography. Springer International Publishing, Cham, pp 35–52. https://doi.org/10.1007/978-3-030-04573-9_4.

Merlin P, Angelin Sarah DD, George ST (2020) Deep learning techniques for EEG signal application review. IETE J EEG Deep Learn Syst Map Stud. https://doi.org/10.1080/03772063.2020.1749143

Miotto R et al (2017) Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 19(6):1236–1246. https://doi.org/10.1093/bib/bbx044

Morabito FC et al (2017) Deep learning representation from electroencephalography of early-stage creutzfeldt-jakob disease and features for diffierentiation from rapidly progressive dementia. Int J Neural Syst 27(2):1–15. https://doi.org/10.1142/S0129065716500398

Mumtaz W, Qayyum A (2019) A deep learning framework for automatic diagnosis of unipolar depression. Int J Med Inform. https://doi.org/10.1016/j.ijmedinf.2019.103983

Murtaza G et al (2019) Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artif Intell Rev 53:1655–1720

Muhammad ZB, Aslam N, Shum Hubert PH (2020) Filtering techniques for channel selection in motor imagery EEG applications: a survey. Artif Intell Rev 53(2):1207–1232

Naira CAT, Cristian JDA (2019) Classi- fication of people who suffer schizophrenia and healthy people by EEG signals using deep learning. Int J Adv Comput Sci Appl 10(10):511–516. https://doi.org/10.14569/ijacsa.2019.0101067.

Noor NSEM, Haidi I (2020) Machine learning algorithms and quantitative electroencephalography predictors for outcome prediction in traumatic brain injury: a systematic review. IEEE Access 8:102075–102092. https://doi.org/10.1109/ACCESS.2020.2998934

Obeid I, Picone J (2016) The temple university hospital EEG data corpus. Front Neurosci. https://doi.org/10.3389/fnins.2016.00196

Oh SL et al (2019) Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals. Appl Sci 9(14):2870. https://doi.org/10.3390/app9142870

Olbrich S, Martijn A (2013) EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response. Int Rev Psychiatry 25(5):604–618. https://doi.org/10.3109/09540261.2013.816269

Olbrich S, Van Dinteren R, Arns M (2016) Personalized medicine: review and perspectives of promising baseline EEG biomarkers in major depressive disorder and attention deficit hyperactivity disorder. https://doi.org/10.1159/000437435.

Organization, World Health (2001) The World Health Report 2001: Mental health: new understanding, new hope. World Health Organization

Petersen K, et al (2008) Systematic mapping studies in software engineering. In: 12th international conference on evaluation and assessment in software engineering (EASE) 12, pp 1–10

Phang CR, et al (2020) A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns. IEEE J Biomed Health Inf 24(5):1333–1343. https://doi.org/10.1109/JBHI.2019.2941222

Picchietti D, Winkelman John W (2005) Restless legs syndrome, periodic limb movements in sleep, and depression. Sleep 28(7):891–898

Podgorelec V (2012) Analyzing EEG signals with machine learning for diagnosing alzheimer's disease. Elektronika ir Elektrotechnika 18(8):61–64. https://doi.org/10.5755/j01.eee.18.8.2627

Poirson P, Idrees H (2013) Multimodal stacked denoising autoencoders. In: Technical Report, pp 1–4

Roth T (2007) Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 3(5 suppl):S7–S10

Roy Y, et al (2019) Deep learning-based electroencephalography analysis: a systematic review. J Neural Eng. https://doi.org/10.1088/1741-2552/ab260c

Ruffini G et al (2019) Deep learning with EEG spectrograms in rapid eye movement behavior disorder. Front Neurol 10:1–9. https://doi.org/10.3389/fneur.2019.00806

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536

Rishi RS, Ram BP (2017) Time-frequency representation using IEVDHM-HT with application to classification of epileptic EEG signals. IET Sci Meas Technol 12(1):72–82

Sahu R, et al (2020) Epileptic seizure detection: a comparative study between deep and traditional machine learning techniques. J Integrat Neurosci 19(1):1–9

Shahin M et al (2017) Deep learning and insomnia: assisting clinicians with their diagnosis. IEEE J Biomed Health Inform 21(6):1546–1553. https://doi.org/10.1109/JBHI.2017.2650199

Sharma RR, et al (2018) Automated system for epileptic EEG detection using iterative filtering. IEEE Sensors Lett 2(4):1–4

Shoeb A (2009) Application of machine learning to epileptic seizure onset detection and treatment. In: Diss. Massachusetts Institute of Technology, pp 157–162. https://dspace.mit.edu/handle/1721.1/54669

Shrestha A, Mahmood A (2019) Review of deep learning algorithms and architectures. IEEE Access 7:53040–53065

Smolensky P (1986) Information processing in dynamical systems: Foundations of harmony theory BT–parallel distributed processing: volume 1: foundations. In: Parallel distributed processing: volume 1: foundations. MIT Press, Cambridge, pp 194–281

Sun M et al (2018) Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm. IEEE Access 6:77596–77605. https://doi.org/10.1109/ACCESS.2018.2883562

Sun S, Zhou J (2014) A review of adaptive feature extraction and classification methods for EEG-based brain-computer interfaces. In: Proceedings of the international joint conference on neural networks, pp 1746–1753. https://doi.org/10.1109/IJCNN.2014.6889525.

Thara TDK, PremaSudha BGP, Xiong F (2019) Epileptic seizure detection and prediction using stacked bidirectional long short term memory. Pattern Recogn Lett 128:529–535. https://doi.org/10.1016/j.patrec.2019.10.034
Tjepkema-Cloostermans MC et al (2019) Outcome prediction in postanoxic coma with deep learning. Crit Care Med 47(10):1424–1432. https://doi.org/10.1097/CCM.0000000000003854
Tsiouris O et al (2018) A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. Comput Biol Med 99:24–37. https://doi.org/10.1016/j.compbiomed.2018.05.019
Türk Ö, Siraç Özerdem M (2019) Epilepsy detection by using scalogram based convolutional neural network from EEG signals. Brain Sci. https://doi.org/10.3390/brainsci9050115
University of Freiburg (2003) EEG database. In: Seizure prediction project freiburg. https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database. Accessed 23 Mar 2021
Vahid A et al (2019) Deep learning based on event-related EEG differentiates children with ADHD from healthy controls. J Clin Med. https://doi.org/10.3390/jcm8071055
Vincent P, et al (2010) Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11(12)
Warrick PA, Lostanlen V, Nabhan Homsi M (2019) Hybrid scattering-LSTM networks for automated detection of sleep arousals. Physiol Meas. https://doi.org/10.1088/1361-6579/ab2664
Wei X, et al (2018) Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. In: BMC medical informatics and decision making. https://doi.org/10.1186/s12911-018-0693-8
Wei Z et al (2019) Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. Biomed Signal Process Control 53:101551. https://doi.org/10.1016/j.bspc.2019.04.028
Weisman D, McKeith I (2007) Dementia with Lewy bodies. Semin Neurol 27(1):42–47. https://doi.org/10.1055/s-2006-956754
Wen T, Zhang Z (2018) Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access 6:25399–25410. https://doi.org/10.1109/ACCESS.2018.2833746
Wittchen HU et al (2011) The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharm 21(9):655–679. https://doi.org/10.1016/j.euroneuro.2011.07.018
Xie N, et al (2020) Explainable deep learning: a field guide for the uninitiated explainable deep learning. Tech. rep. arXiv: 2004.14545v1
Yosinski J et al (2014) How transferable are features in deep neural networks? In: NIPS'14: Proceedings of the 27th international conference on neural information processing systems. MIT Press, Montreal, Canada, pp 3320–3328
Young T, Peppard Paul E, Gottlieb Daniel J (2002) Epidemiology of obstructive sleep apnea: a population health perspective. Am J Respir Crit Care Med 165(9):1217–1239
Yuan Y et al (2019) A multi-view deep learning framework for EEG seizure detection. IEEE J Biomed Health Inform 23(1):83–94. https://doi.org/10.1109/JBHI.2018.2871678
Zerafa R et al (2018) A comparison of a broad range of EEG acquisition devicesis there any difference for SSVEP BCIs? Brain-Comput Interfaces 5(4):121–131. https://doi.org/10.1080/2326263X.2018.1550710
Zerr I et al (2009) Updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease. Brain 132(10):2659–2668. https://doi.org/10.1093/brain/awp191
Zhang T, Chen W (2016) LMD based features for the automatic seizure detection of EEG signals using SVM. IEEE Trans Neural Syst Rehabil Eng 25(8):1100–1108
Zhou C, et al. (2015)A C-LSTM neural network for text classification. In: arXiv preprint arXiv:1511.08630
Zhu J et al (2019) Multimodal mild depression recognition based on EEGEM synchronization acquisition network. IEEE Access 7:28196–28210. https://doi.org/10.1109/ACCESS.2019.2901950