Depression diagnosis by deep learning using EEG signals: A Systematic Review

Atefeh Safayari[1], Hamidreza Bolhasani[2+]

Abstract— **Depression is considered by WHO as the main contributor to global disability and it poses dangerous threats to approximately all aspects of human life, in particular public and private health. This mental disorder is usually characterized by considerable changes in feelings, routines, or thoughts. With respect to the fact that early diagnosis of this illness would be of the critical importance in effective treatment, some developments have occurred in the purpose of depression detection. EEG signals reflect the working status of the human brain which are considered the most proper tools for a depression diagnosis. Deep learning algorithms have the capacity of pattern discovery and extracting features from the raw data which is fed into them. Owing to this significant characteristic of deep learning, recently, these methods have intensely utilized in the diverse research fields, specifically medicine and healthcare. Thereby, in this article, we aimed to review all papers concentrated on using deep learning to detect or predict depressive subjects with the help of EEG signals as input data. Regarding the adopted search method, we have finally evaluated 22 articles between 2016 and 2021. This article which is organized according to the systematic literature review (SLR) method, provides complete summaries of all exploited studies and compares the noticeable aspects of them. Moreover, some statistical analyses have been performed to gain a depth perception of the general ideas of the latest pieces of research in this area. A pattern of a five-step procedure has also been established by which almost all reviewed articles have fulfilled the goal of depression detection. Finally, open issues and challenges in this way of depression diagnosis or prediction and suggested works as the future directions have been discussed.**

Keywords: **Deep Learning; Depression; Electroencephalogram; EEG; Depressive Disorders; Systematic Literature Review**

---

[1] Department of Computer Science, Kharazmi University of Tehran, Tehran, Iran.

[2] 2 Department of Computer Engineering, Islamic Azad University Science and Research Branch, Tehran, Iran.

+ Corresponding Author: hamidreza.bolhasani@srbiau.ac.ir

1. Introduction

It has been announced by the World Health Organization (WHO) that according to the estimation, in excess of 322 million people all around the world suffer from depression by which this mental disorder has become the leading contributor to the causing of disability. Depressive patients are typically recognized by symptoms such as a feeling of sadness, hopelessness, and guilty, loss of interest, concentration, and energy, changes in appetite, sleep, and other routines [1].

Poverty, unemployment, tragic events of life, physical disorders, and problems with alcohol or drug consumption are considered as determining factors leading to depression [1]. Recently, the Covid19 pandemic has advanced the cause of depression, and its consequent conditions such as imposing lockdown, going into quarantine, and practicing social distancing are considered other main reasons for experiencing depression [2-6].

With regard to the fact that depression poses an unprecedented threat to public health and have some adverse effects on depressive people such as committing suicide [7], and also considering this matter that early diagnosis can result in providing timely and more effective treatment, devising and developing an efficient and reliable method of depression detection or even prediction would be of the high importance.

Electroencephalogram (EEG) signals which naturally have nonstationary, highly complex, non-invasive, and nonlinear structure, involves human brain activities and working status. Due to this complexity, available abnormality would be difficult to detect with the naked eyes. These properties have made physiological signals are deemed to be valuable tools for depression detection [8-9].

Deep learning is defined as a hierarchy structure with a series of algorithms which have some hidden neurons [10]. These models provide computers with the ability of building complex concepts from simple statements [11]. These learnt concepts utilize to build next layers. Furthermore, in these methods, multiple processing layers are responsible for pattern and data structure recognition [12].

This multi-layer method has been recently utilized in different applications of various fields, ranging from agriculture [13], automotive Industry [14-15], and IoT [16] to diverse usage in medicine [17-20]. Because deep learning architecture has the capability of automated learning and extracting features from input raw data, and regarding the limitation of machine learning techniques in this point [21], and also due to difficulties in manual EEG signals analyzing [22], adopting deep learning solutions has been more popular in the related occasions by which the implicit nonlinear features in EEG signals can extract with the least effort [23-25].

Applications of deep learning in depression diagnosis with the assistance of EEG signals have increased in recent times. The main purpose of this article is to conduct an SLR-based review of papers concentrated on deep learning usage in depression detection or prediction using EEG signals. To the best of our knowledge, there was no systematic literature review focused on the mentioned topic and our SLR survey is the pioneer in this field of research.

The important contributions of this review are as follow:

- Designing a complete taxonomy chart of all exploited deep learning algorithms
- Providing concise summaries of all studied researches and drawing a comparison between main aspects of them
- Propose a general pattern of five steps based on all reviewed papers and providing some information related to each stage to analyze them
- Highlighting open issues and future works for this purpose

In the organization of this review, section 2 is dedicated to discuss related works. In section 3 the adopted research method is presented. The provided summaries together with the designed taxonomy are located in section 4. In section 5, a discussion and comparison between studied papers are provided. Section 6 at the end, is devoted to the conclusion.

2. Review of Related Studies

This section is aimed to review the conducted research works which have addressed depressive subjects' diagnosis and prediction by deep learning methods and EEG signals as part of them.

Yasin et al. [26] conducted a review of studies that had adopted both neural network and deep learning approaches to detect two kinds of depression which were Major Depressive Disorder (MDD) and Bipolar Disorder (BD) using EEG signals. It utilized various source engines and a combination of diverse keywords to search among papers that have been published over the last ten years and then extracted some useful information from those. One of the good points about this review was providing different classifications about exploited datasets, methods of analyzing or extracting features, and algorithms in articles. It also draws various tables to present extracted data and provide comparisons between them in different ways. However, the major drawback of this research was that it did not use a sufficient number of articles to review, so approximately five articles as it stated, particularly for MMD diagnosis, was considered an obvious flaw. Besides, those employed articles did not explain as it was required to do to get the general idea and procedure of them.

The presented review by [27] concentrated on articles within which deep learning methods had been employed to study mental disorders that depression was one of the subjects. Detection using clinical data, the use of genetic data in disease diagnosis, analyzing different datasets, and utilizing social media data to estimate the probability of mental illness were the four main categories of this study. Selected papers which were published until April 2019 utilized various kinds of datasets among which the Electroencephalogram dataset type were used in only three articles focused on depression detection or foreseen. This research included a complete representation of evaluated datasets. It also discussed thoroughly the opportunities and challenges which are likely to face by exploiting each dataset. However, due to the fact that it was a comprehensive review concentrated on encompassing diverse cases of mental illness, it covered a few papers related to adopting deep learning for depression diagnosis and prediction using EEG signals with a short explanation.

Khosla et al. [28] carried out a review of articles focused on using EEG signals and different types of classifiers to diagnose some neurological disorders such as depression and monitor other problems such as emotion recognition. It utilized various sources like journals, conferences, books, and thesis to get papers that were mainly published from 1999 to 2019, and only four papers belonged to an earlier time. With respect to the depression diagnosis, just around ten articles were considered. It provided a comprehensive comparison with statistical analyses of exploited techniques and knowledge in articles which were categorized into separate sections, namely, artifacts removal, feature extraction, types of extracted features, feature selection, dimensionality reduction, and classifiers methods. Different adopted datasets were also summarized and exhibited as public and local acquisition classifications based on the manner within which they were collected. Moreover, it contained information about techniques of functional neuroimaging. Covering diverse areas of research in mental health problems, however, prevented it from concentrating on perfectly addressing each area.

Table 1 provides some main information about the reviewed articles contained some analyses of depressive discrimination by adopting deep learning using EEG signals.

Table. 1. Major features of studied researches included depression detection using DL and EEG

Reference	Publication year	Main context
Yasin et al. [26]	2021	BD and MDD detection by neural networks using EEG
Su C et al. [27]	2020	Mental health outcome study using DL
Khosla et al. [28]	2020	Using DL and EEG signals for different applications

3. Research Method

In this article, we have concentrated on utilizing deep learning algorithms to detect or predict depression using EEG signals. These techniques are classified based on the SLR method as an assessment method of the research study [29-32]. This section intends to expound on the adopted SLR-based research method.

The Analytical Questions which are defined in this SLR paper would be as follow. These questions are going to be answered completely in regard to the purpose of this study.

AQ1: What deep learning algorithms have been used to detect or predict depression?

AQ2: Which deep learning methods are mostly preferred to fulfill our goal?

AQ3: What approaches are mainly used to extract features from EEG signals?

AQ4: What are the future and open works related to depression detection or prediction?

The selection process of ultimate papers is carried out by applying some principles and inclusion/exclusion criteria.

Publisher type and index, publish year, citation numbers of articles were four main aspects affected by applied rules. Thereby, only peer-viewed, ISI-indexed papers that were published over the past five years in journals or conferences with more than 10 citations, except those published over two last years, were analyzed. Owning to the fact that published articles in the last year and also in 2020 have been cited fewer than those in other years, researches with zero-citation in 2021, less than 5 quotations for the second half of 2020, and at least 5 for the first half of 2020 were chosen to include in the analysis.

Fig 1 presents detailed information about the evaluation procedure of selected studies. Finally, amongst all related papers to our topic, 22 have selected to consider.

Inclusion principles are formulated as containing studies which have:

- Published between 2016 and 2021
- Focused on adopting deep learning to detect or predict depression using EEG signals
- Adhered to the rules of citation number

Exclusion principles are contained studies which have:

- Not indexed in ISI
- Not written in English
- Not meet the minimum requirements of the citation.
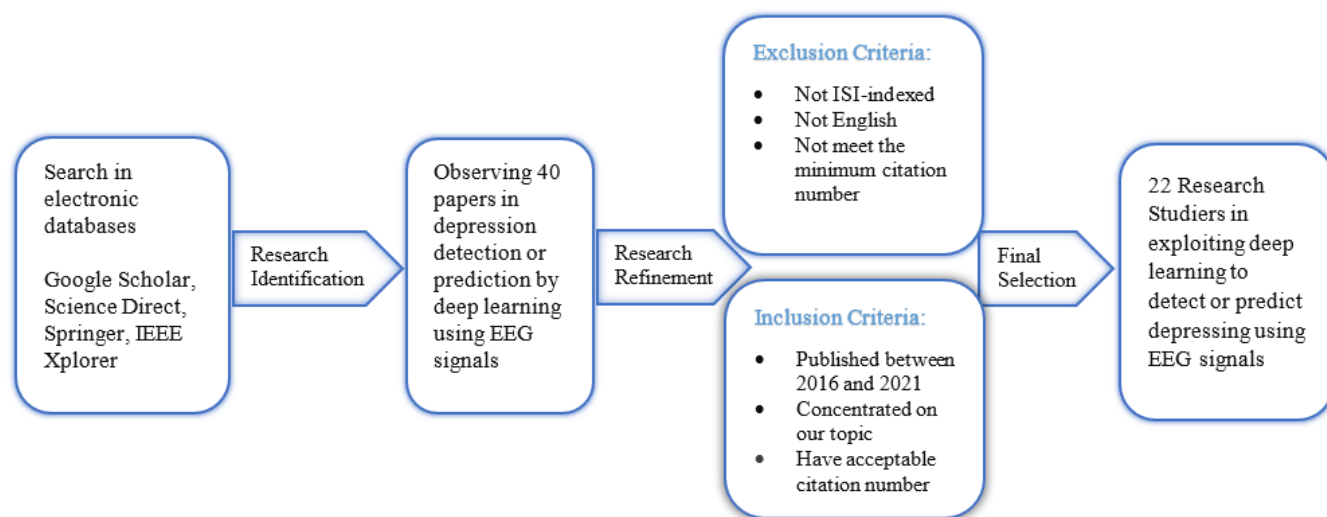
Fig. 1. The papers selection and evaluation flow chart

Fig 2 exhibits the distribution of selected papers by their publishers over the specified time interval. Elsevier, Springer, and IEEE were three most leading publisher. In this SLR review, Google Scholar, Science Direct, and IEEE Xplorer used as scientific electronic databases.
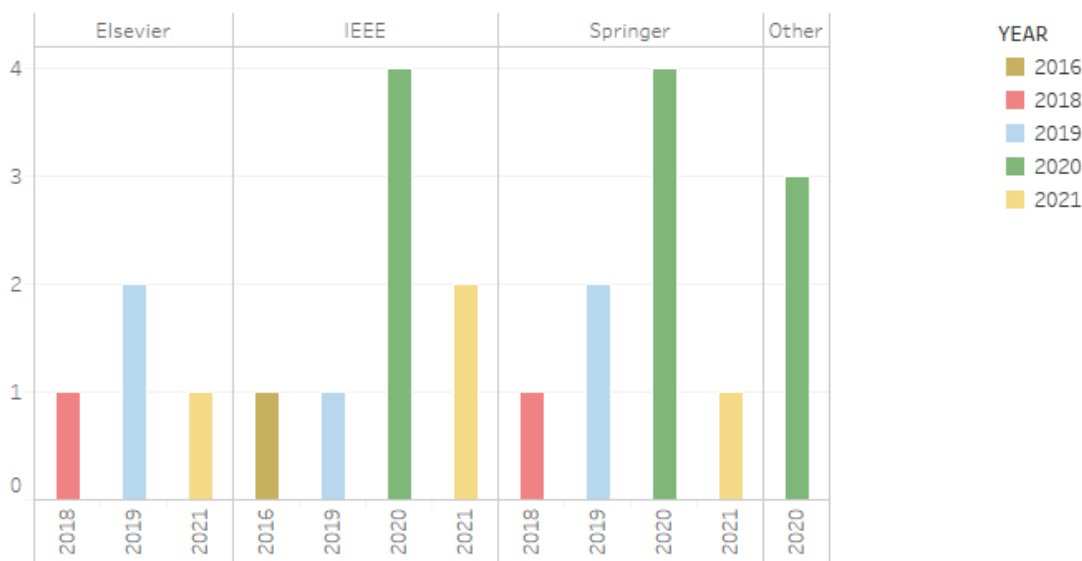
Fig. 2. Distribution of analyzed papers by publisher and year

4. Organization of deep learning algorithms for depression diagnosis using EEG signals

   This part, according to the applied systematic literature review process, presents a review and summary of papers intended to use deep learning algorithms to detect or predict depression with the assistance of EEG signals, in a way that they have adhered to the set of mentioned rules in section 3. Fig.3 exhibits a comprehensive taxonomy of used deep neural networks models to serve the purpose of depressive disorders diagnosis and foreseen. According to the presented taxonomy, all adopted models can be categorized into four main groups, namely CCN-Based models, LSTM models, Combination (combined models of both convolutional networks and LSTM blocks), and Other Algorithms. Due to the fact that all other algorithms of deep learning except CNN and LSTM were less prevalent and less common among all selected papers, with about one application for each of them, these kinds of methods have been classified as Other Algorithms cluster. With respect to the construction of the reviewed articles, the vast majority of them have applied various deep learning algorithms on the same prepared dataset to obtain the best accuracy. So the offered taxonomy chart is classified and drawn based on utilized algorithms from all articles. Thereby, the summary of inspected studies have been arranged in descending chronological order into subsections 4.1 and 4.2, respectively. A detailed analyses of these studies and also the comparison between them has been carried out in subsection 4.3. Furthermore, some major concepts such as main context, advantage, drawback, and new contribution of summarized articles presents in table 2 and table 3. As we have noticed, all articles follow some general steps to discriminate between depressed and normal cases. Subsection 4.3 also attempt to analyze this five-step pattern and fig.4 shows these five key stages. In subsections 4.3.1 to 4.3.5, some extracted information related to each of these five has been presented in tables numbered 4 to 9 by which an opportunity for comparison will provide between examined papers in the point of different techniques and methods which have been employed in each step.
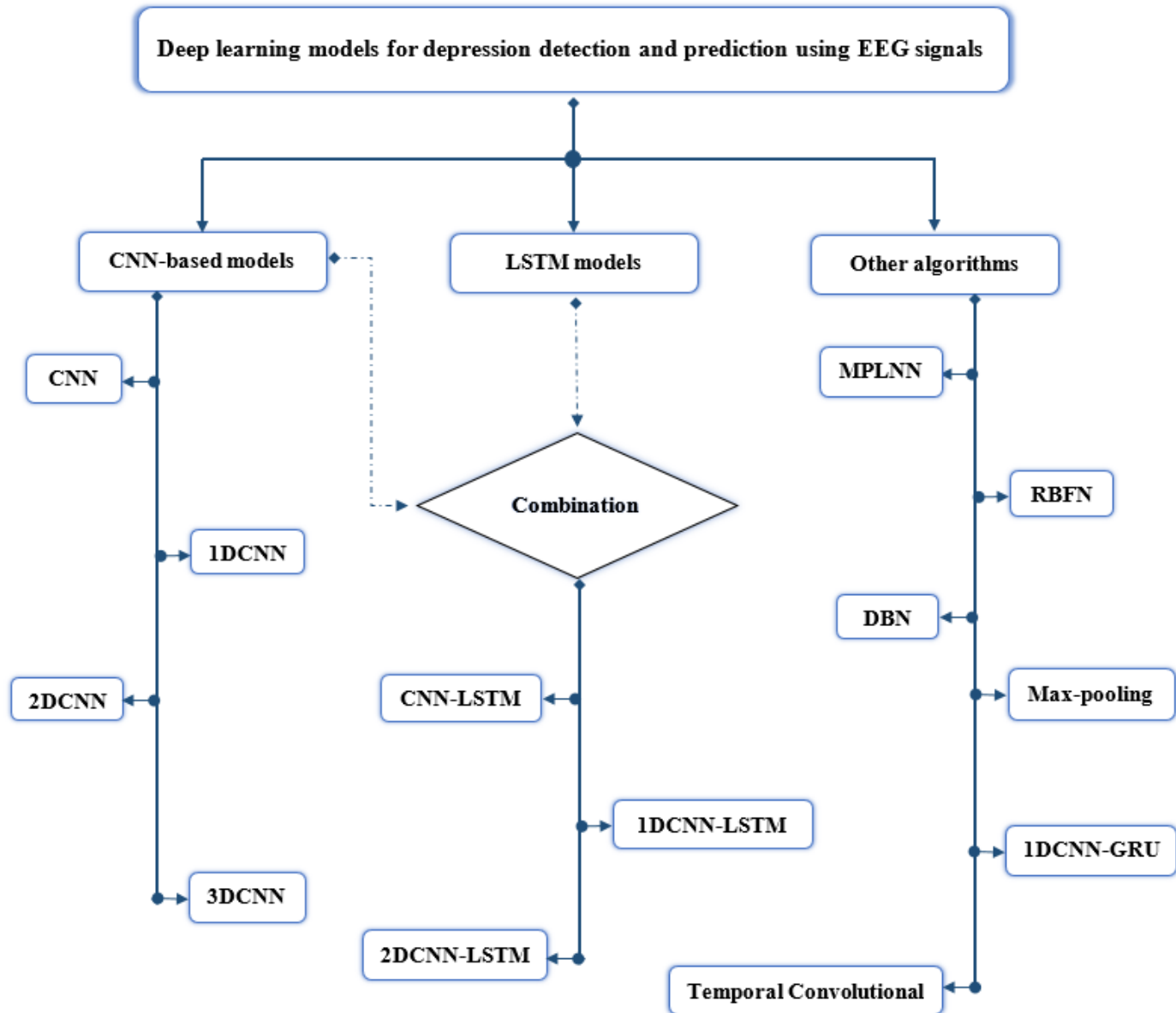
Fig.3. the taxonomy of deep learning models for depression detection using EEG signals

4.1 Deep learning methods for depression detection using EEG signals

Following the SLR structure and organization [29-32], this part is aimed to provide a brief and yet complete introduction to all examined articles to acquaint future researchers with previous works done in the regard of this review's topic. It has been strived for covering all main points and features of papers in the form of summaries, including some facts about used dataset, structure of proposed deep learning model, ways of data preparation, advantages and drawbacks of their adopted method, and their achievements and contribution.

Sharma et al. [33] proposed a new, computer-aided (CAD) method for depression diagnosis called DepHNN which stands for Depression Hybrid Neural Network. Convolutional Neural Network (CNN) and Long-short Term Memory (LSTM) are two deep learning algorithms used to capture the temporal

dependencies in the time-series EEG input signals and to process the sequence learning, respectively. Artifacts in this study have been eliminated by Independent Components Analysis. Time-frequency information also extracted by the Fast Fourier Transform method from EEG signals. To construct a model with high performance, some models with different numbers of LSTM and fully connected layers has been built. Thereby, this result achieved that decreasing in the number of LSTM layers and increasing in fully connected layers numbers, in a way that the former figure is less than the latter one, will lead to reduction of both execution time and the loss value. In this adopted approach, firstly, CNN is in charge of EEG input data transformation into cross-sectional data using a technique named windowing. Then, the transformed output is fed into the LSTM block which has the ability to memorize data from previous steps due to hidden memory cells in its architecture. Finally, fully connected layers are employed to assist in the automatic detection of depressive cases. This Hybrid 6-layer CNN-LSTM model has enjoyed advantages of less time and computation complexity with considerable high accuracy due to using fewer hidden layers and the windowing technique. Overfitting possibility and small input data, however, have been two main problems this research has been confronted with.

A new, 18-layer, CNN-based framework named DeprNet has been proposed by [34] in which both spatial and temporal dimensions of the EEG data have been utilized during training. Produced artifacts by eye blink have been removed from EEG signals with the help of independent component analysis (ICA) Method. Three kinds of interferences which were the low-frequency noise, irrelevant signals, and the baseline noise has been eliminated with the assistance of three filter methods, namely, 0.1-Hz high-pass filter, a 100-Hz low-pass filter, and a 50-Hz notch filter. A construction of 2D matrix has been fed as input data. Five convolution layers were in charge of feature extraction, five batch normalization layers aimed to keep the network steady through output normalization, five max-pooling layers that decreased the sample size of EEG signals, two fully connected layers, and finally, a soft-max Layer employed to classify subjects. This paper has taken advantage of using short 4-s EEG signals from 19 channels which will lead to adopt this method in pragmatic approaches. Besides, obtaining significant-high accuracy over other base models was considered another merit of it. Moreover, it has shown that activities related to the left and right hemispheres of the brain are differently influenced by depression consequences. However, the required technologies to fulfill the goal of depression diagnosis using this method might impose difficulties on clinical settings in using up this method. Complexity might also be a result of using a large number of layers.

Saeedi A et al. [35] practiced using generalized partial directed coherence (GPDC) and direct directed transfer function (DDTF) along with deep learning algorithms to achieve the goal of detection. GPDC and DDTF intended to extract the available connection between EEG signals channels as an effective analysis of brain connectivity, combined with eight frequency bands in pairs. These sixteen connectivity methods aimed to transform 1D EEG signals into 2D images by which the input of the deep learning classifiers is formed. With regards to classification, among five adopted deep learning algorithms, which were 1DCNN, 2DCNN, LSTM, 1DCNN-LSTM, 2DCNN-LSTM, one-dimensional CNN-LSTM had better performance rather than others in terms of accuracy and sensitivity. Although 2DCNN-LSTM was faster and employed more parameters, temporal information might filter out by two-dimensional filtering of this method and lead it to fail in obtaining perfect accuracy. Providing comparison between diverse deep learning algorithms and conversion of EEG signals into connectivity-2D matrix images by the employment of effective connectivity methods are two imperative merits of this research. It also derived the advantage of temporal and spatial information in detection. Small data size, however, is considered a major deficiency of this model. Moreover, there might be more techniques to utilize these features and parameters to get a more precise accuracy.

The proposed model by [36] was depression detection by adopting a three-dimensional convolutional neural network (3D-CNN) using effective connectivity inside the brain default mode network (DMN) region which was estimated from 19-channel EEG signals. The process of removing artifacts by eye blinks, muscles, etc. has been done by EEGLAB software in which the artifact subspace reconstruction (ASR) method was responsible for elimination. To calculate effective connectivity, some continuous segments of 2-sec length extracted from the output of pre-processing to feed into partial directed coherence (PDC) algorithm. The result of PDC was connectivity matrices that covered 19 channels. These output matrices reduced to overlay six channels of DMN area by adopting the subsequent DMN connectivity extraction. These matrices were then utilized as input of the CNN model. Three convolutional layers, three batch normalization (BN), and three rectified linear unit (ReLU) activation layers, a global average pooling layer, a dropout layer, and one fully connected layer were components of this deep network. The fully connected layer aimed to classify using the binary soft-max regression. This approach that according to the assertion of its authors has been taken for the first time, expressed a novel viewpoint in EEG signals study. Overfitting, however, might be one of the most likely issues that this method should address. For evaluation of this new manner, more data is required to clarify and examine this procedure

Qayyum et al. [37] suggested using two deep neural network based models which were called IDCNN-GRU and IDCNN-LSTM on two eye open (EO) and eye close (EC) datasets. Input EEG signals segmented into a time window of one second which has included 19 channels. IDCNN, Gated Recurrent Units (GRU)/long short term memory (LSTM), and classifiers were considered three main parts of both methods. 1DCNN as the first component, which included three CNN, two max-pooling, and two dropout layers, adopted to extract features. Then LSTM/GRU was in charge of sequential learning. Finally, KNN, RF, and SVM machine learning models accomplished the purpose of prediction. Extracted features from only the last dense layer fed into the classification layer (sigmoid function for binary classification) to improve performance. Learning rate, optimizers, and loss functions were training hyper parameters that were exploited in the study. As a result, GRU outperformed LSTM since it is appropriate for short sequences and fewer training parameters and memory is required. Time window segmentation, data size, and extracted features, however, needed more inspections.

Depression detection in [38] carried out using a 12-layer CNN-LSTM model. Recorded EEG signals prepared through two steps. By adopting the FASTER algorithm at first, artifacts eliminated. The FASTER algorithm method was then employed to remove the offset effects and the amplitude scaling issues from EEG signals. The proposed model generally was comprised of three main components which were convolutional neural network, LSTM memory, and classification parts. Three CNN layers and three MaxPooling1D layers formed the first section of the model which exploited to extract features. The second unit included two LSTM layers whose duty was to generate the feature maps by discovering different patterns in EEG signals and then retain the sequence of these learning. Two fully connected layers, one dropout layer, and one Flatten layer designed to carry out the classification task within which the flatten layer aimed to provide fully connected layers with a feature vector by converting the output of LSTM layers. This simple architecture with the flatten layer exhibited that the right hemisphere signal affected the result of diagnosis more than signals from the left hemisphere. In addition to the fact that more data was required to evaluate the accuracy and assertion of the model and also the risk of overfitting, it was expected to utilize some advanced manner of feature extraction and analysis. Moreover, it had inadequate explanation about the used dataset.

Kang M et al. [39] offered a two-dimensional convolution layer-based model using recorded signals of both eye closed (EC) and eye opened (EO) states. After data collection, firstly, power line interference was suppressed by a 50 Hz notch filter. The process of removal artifacts was applied only on the EC dataset.

Thereby, noises by eye blink or muscle movement were eliminated with the help of a low-pass filter by which frequencies above 32Hz are removed. The output signals from preprocessing step were decomposed to Delta (0.5~4), Theta (4~8), Alpha (8~13) Beta (13 ~32) frequency bands. The 11-layer proposed model was comprised of a convolutional neural network included two Conv2D layers with ReLU activation function, two Max-pooling layers, and one dropout layer, fully connected layer contained two dense layers with ReLU activation function and one dropout, and the output layer as the last section composed of one dense layer with ReLU activation function. It also utilized ten-fold cross-validation to evaluate and verify the result of the model. As a result, it obtained a high accuracy compared to existing models on the same dataset. However, the act of generalizing the result of this model was hampered by a lack of adequate data.

Kang M et al. [40] introduced a novel methodology of feature extraction to detect depression which was exploiting asymmetry feature of EEG signals and converting it into 2D images to feed to the convolutional neural network. In the preprocessing level, each channel of raw EEG signals normalized by the min-max normalization method, and then artifacts elimination was done by the independent component analysis (ICA). The output signals were afterward segmented into a time window of four seconds. To generate the asymmetry matrix image, firstly, four power frequency bands which were Delta, Theta, Alpha, and Beta extracted, and based on them the matrix image created. The produced image used as input of a CNN model which had constructed from three two-dimensional convolution layers with the ReLU as activation function, three two-dimensional max-pooling layers, one flatten and one dropout layer, and two fully connected layers. It concluded that the alpha-band asymmetry matrix image obtained the best accuracy rather than other power bands. Although it strived to take advantage of utilizing the dropout layer, data segmentation, and batch normalization to avoid overfitting, lack of sufficient data was still a potential cause of it.

Duan L et al. [41] endeavored to exploit the effects of depression on the MDD patient's signals. To begin, to do pre-processing, finite impulse response (FIR) filter and independent component analysis (ICA) algorithm adopted to remove unnecessary signals and visual artifacts from EEG data, respectively. Then, the fast Fourier transform (FFT) algorithm aimed to extract Theta, Beta, and Alpha frequency bands from pre-processed EEG signals by which two features named the interhemispheric asymmetry value and cross-correlation value computed. The structural and connective changes analysis afterwards conducted by utilizing these extracted features. Combination of all the structural features and connectivity features used to construct three feature matrices; two single-feature matrices which were comprised of three layers of Theta, Beta, and Alpha interhemispheric asymmetry (or cross-correlation), and one 6-layer mixed feature matrix that had composed of two Alpha, two Beta, and two Theta band mixed features layers. In consequence, to determine the most appropriate feature to obtain the best accuracy, this study drew a conclusion that among the structural features, connectivity features, and the mixed features which were fed into CNN, the feature combining method outperforms others. Anyway, other than overfitting which might be one of the weaknesses of this proposed model, by proving more data, use features and formulas might be reconsidered.

Saeedi M et al. [42] aimed to the prediction by a three-step procedure; feature extraction, feature selection, classification. For pre-processing operation, Discrete Wavelet Transform (DWT) employed to decompose EEG signals into detailed and approximate coefficients by which thresholds that matched artifacts omitted. Five prevalent frequency bands called linear features extracted from EEG signals were Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz). In the following, non-linear features are obtained by applying sample entropy and approximate entropy on the wavelet packet coefficients. Then, a genetic algorithm (GA) adopted to choose significant features and reduce feature dimensionality. Two machine learning algorithms, namely, KNN and SVM, and multilayer

perceptron (MLP) algorithm as deep learning methods were exploited to identify depressive cases. In respect of deep learning model utilization, MLP combined with non-linear features achieved a more high accuracy compared to linear features. While using linear features, the Gamma power band provided a high value of accuracy rather than others. Although this research strived for enhance the result of this prediction by deriving advantages of various features from EGG signals, these parameters should be optimized.

Y. Xie et al. [43] designed a CNN model in a combination of functional connectivity of brain networks. To begin, to do pre-processing tasks, after artifacts elimination from EEG signals, the PLI method employed to form a 31-in-31 adjacency matrix with functional connectivity of brain networks which was calculated on the Delta, Alpha, Beta, Theta, and Gamma frequency bands. In this research, a brain network defined as a graph with edges and nodes in which nodes corresponded to particular areas of the brain to record EEG signals, and edges indicated the function of each node. PLI method was also adopted to measure the strength of the established connection between nodes to draw the brain network as input data. They utilized a simple deep learning model named CNN-2 in which there were one convolutional layer and one pooling layer. This study provided the opportunity of making a comparison between six different models in which brain networks and Prefrontal Lateralization methods combined with CNN-2, DBN, and LDA algorithms in pairs. They showed that among all six mentioned models, brain networks with CNN-2 performed the best. Although this proposed method took a new approach in pre-processing step, the simple-adopted convolutional neural network model posed a risk of overfitting to classification output. Moreover, the employment of more data or other techniques such as LSTM might improve the accuracy.

X. Zhang et al. [44] experimented with two models of MDD prediction which were the employment of one-dimensional Convolutional Neural Network (1-D CNN) with and without demographic attention mechanism to indicate the effects of EEG signals and demographic information integration on accuracy. In order to prepare data, discrete wavelet transform and Kalman filtering methods were adopted to omit noisy signals. In addition, time windows of 4-non-overlapping seconds are made from 40-second-high-quality parts of EEG signals. One Hot Encoding and normalization method converted demographics information into 1D. Spatial-temporal features then extracted using 1DCNN. After that, demographics information is indirectly integrated into the mentioned model by the attention mechanism. Lastly, a fully connected layer utilized to produce the final classification through the soft-max function. As a result, CNN with demographic attention outperformed CNN without that information. To more closely investigate the impacts of incorporated demographic information, the DEEP-DREAM method exploited to create two artificial EEG signals from mentioned CNN models and their power spectral density (PSD) measured. It exhibited that mainly Beta frequency bands distinguished two created signals. This study took advantage of demographics information as a new technique to obtain acceptable diagnostic accuracy. However, alongside the lack of data and overfitting, it is required to take a closer look at the integration and influence of this information.

C. Uyulan et al. [45] suggested three convolutional neural network-based models which integrated with advanced computational neuroscience methodology named ResNet-50, MobileNet, and Inception-v3 to apply to the EEG recorded signals from both left and right hemispheres with eye closed states. The produced artifacts as a result of eye and muscle activities were removed by the wavelet transform method. In addition to a notch filter of 50 Hz which was employed to compensate power line interferences, three band-stop with 3 Hz, 35 Hz, and 50 Hz cutoff frequencies and the FastICA method also applied on input signals to carry out a complete process of noise removal. 12th-order Butterworth band pass filter utilized to extract four Delta, Theta, Beta, and Alpha frequency bands. Probable error in both passbands and stopbands minimized by using Parks-McClellan optimal FIR (finite impulse response) filter algorithm. The output signals from

noise removal and pre-processing steps were then converted to a 2D EEG image matrix to input CNN structures. 22-layer ResNet-50 Convolutional model which was considered a residual learning network had the ability to deal with problems about the gradients vanish and led the deep network to better training. MobileNet Convolutional structure used depth-wise, separable convolutions and some parameters to enhance performance. 48-layer Inception-v3 Convolutional model also had some advantages like avoiding loss of information during training. As a consequence, MobileNet architecture performed better compared to two other proposed models and obtained the highest accuracy for hemispheric classification. Furthermore, for frequency band-based classification the best accuracy belonged to ResNet-50. Although this proposed model had some novelty in the point of proposed CNN-based models, clinical environments will be faced with some barriers to implementation due to complexity. Moreover, the gained accuracy should be evaluated with more data.

Z. Wan et al. [46] set the objective of MDD prediction using HybridEEGNet which was a sort of Convolutional Neural Network with the assistance of feature analysis as the second main aspect of this proposed model. Two different sorts of convolutional filters adopted to learn the synchronous and regional characteristics from EEG signals in order to analyze them. Analysis process conducted by using the deep-dreaming algorithm to create feature matrices and the FFT method to examine those generated matrices. A 21-layer deep model employed in this research which constituted of eight convolution layer to learn features, eight max-pooling layers to do down-sampling, one concatenation layer to transform the output of the last-previous layer into one vector and feed it into the next-first fully connected layer, three fully connected layers, and one soft-max layer to categorize. Studying the FFT resulted of analyzing two learned features, particularly the combination of them, provided the fact that they carry some significant information about the depressive cases, namely, they claimed that amplitude ranges and spatial distributions differences in Alpha frequency band might be a distinctive characteristic. It also emphasized the importance of the 4-10 Hz band as the low frequency in the diagnosis goal. Moreover, they stated that Theta and Alpha rhythms played an important part in depression discrimination with considerable performance. In addition to the lack of data that considered a remarkable drawback since led to doubt the accuracy, the manner of extracting features and related achievements to extraction need to be review.

One Dimensional Convolutional Neural Network (1DCNN) and 1DCNN with long short-term memory (LSTM) are two deep learning algorithm employed in [47]. 1DCNN which is aimed to learn and transform temporal information has consisted of three 1D Convolutional layers, two pooling layers, and one dropout layer, and three other layers. The second model with the same purpose as the first architecture but better performance, has comprised of three blocks of one 1D Convolutional layer, one pooling layer, and one dropout layer which were ended up with two LSTM layers. As the start point, the multiple source eye correction (MSEC) method employed to rectify raw input data from which artefacts by eye blinks, muscular or heart activities corrected. Then, those signals segmented into 1 second time window. This proposed model has the merit of rectifying noises generated by eye blinks and muscular activities while EEG recording with applying the multiple source eye correction (MSEC) method session. It also showed that EO signals had a better performance in classification compared to EC signals. Nonetheless, disadvantages of this would be small data size and small time window segmentation. Although it claimed that 1-second window size resulted in best performance, this assertion needed to be proved

Mahato et al. [48] fulfilled the goal of depression detection by employing four deep learning classifiers called multi-layered perceptron neural network (MLPNN), radial basis function network (RBFN), linear discriminant analysis (LDA), and quadratic discriminant analysis. Eight linear features such as Alpha, Beta, Delta, Gamma frequency band power and also Alpha, Beta, Delta, Gamma hemispheric asymmetry, relative wavelet energy (RWE) and wavelet entropy (WE) as non-linear feature extracted from input EEG signals. Dimension reduction of non-linear features occurred with the assistance of principal component analysis

(PCA). Independent component analysis (ICA) aimed to delete artifacts. It was achieved by this paper that the highest accuracy can be obtained by adopting the combination of linear and non-linear features with both MLPNN and RBFN classifiers. Using two non-linear features like WE and RWE together with the LDA classifier produced the second-highest accuracy which is higher than the employment of any mentioned classifiers with any linear features in combination. Although this model disclosed Theta power as a possible way of getting significant-high accuracy to diagnose depressive cases, it required more data to clarify and generalized.

X. Li et al. [49] expounded on taking ensemble learning and deep learning approaches to predict. It also paid close attention to the influence of all three temporal, spectral, and spatial information aspects of EEG signals in analyzing them. EEG signals generated from 128-channel HydroCel Geodesic Sensor Net (HCGSN). Auto-regress model and Hjorth algorithms executed with various time windows to extract power spectral density which was frequency-domain feature and activity that considered the time-domain feature. In this research, EEG signals were transformed into images by adding spatial information through applying the auditory evoked potentials (AEP) method. Two various VGG style architecture exploited in this study in which both ended in a dense layer and a soft-max layer. This study exhibited that the Alpha frequency band outperformed Beta and Theta frequency scales. It also showed that using features played a prominent role in obtaining accuracy. Adopting the AEP method, however, put an obstacle to the performance since it is a distant-based projection method and as a consequence, non-distant aspects were ignored during image conversion. It might be required to review the steps of noisy signals from EEG input.

Ay B et al. [50] studied an 11-layer CNN- LSTM model to discriminate normal and depressed cases using EEG signals from both right and left brain hemispheres. The process of artifact elimination took place manually by specialists. It was comprised of four Conv1D, two Max-Pooling 1D, two Dropout, two dense, and one LSTM layers. The task of feature extraction was carried out by adopting CNN. The employment of LSTM was to perform sequence learning using feature maps which were generated from the output of the CNN. In the end, the detection was accomplished by fully connected layers with the assist of the LSTM output. Besides, it utilized the maximum pooling technique in the max-pooling for input size, memory usage, and the number of parameters reduction. This research reached a noticeable performance due to utilizing some EEG signal properties such as long-term dependencies and local characteristics. In addition to the small data size, manual removal of the artifact was another drawback of this method. Even though this study affirmed that there is no possibility of overfitting, this matter needs to be investigated.

Acharya et al. [51] claimed to pioneer in employment of deep neural network idea and Convolutional Neural Network (CNN) to detect depression. This 13-layer CNN model was comprised of 5 convolutional layers, five pooling layers, and three fully-connected layers in which convolutional layers were responsible for providing important feature obtained from the input EEG signals from the left and right hemisphere to train the algorithm, pooling layers were accountable for reducing of the feature map's size, and fully-connected layers established connection between neurons of a layer with the next one. Finally, results produced by the last fully-connected layer was applied to a soft-max function to detect depressive cases. One of their remarkable achievement was to show that the right hemisphere signals outperformed the left ones to the point of sensitivity and accuracy. However, this method might suffer from false prediction due to lack of adequate data. Moreover, complex architecture because of a lot layers might pose limitations upon using this model.

W. Mao et al. [52] placed a strong emphasis on pre-processing step in which raw EEG signals were transformed and encoded to construct a sample input data to feed into deep learning algorithms. Some primary preparation tasks were comprised of segmentation of 270 seconds, eliminating noise, removing interfering effects of EOG, and extracting Theta, Alpha, and Beta bands, cropping into 0.5-s segments, and

at last, applying the Auto-Regress model (AR model) to calculate the power of electrodes. Power spectrum density of signals exploited to preserve the temporal property of EEG signals. To preserving spatial information of EEG two projection models called distance-based and non-distance methods were employed. The output data fed into deep neural networks to extract both inherent spatial and temporal patterns. After data processing, four deep learning models are CNN, Temporal Convolution, MAX, and LSTM used for classification. The first three models perform better on non-distance mapping frames whilst the LSTM had a lower performance. Although the distance-based method generates frames with more brain activity information, it badly affected the prediction accuracy. Low accuracy and small input data are two downsides of this procedure. Moreover, the performance can get better by optimizing parameter configuration.

Cai et al. [53] made the depression diagnosis by using both deep learning and machine learning algorithm and drew a comparison between them. This research collected EEG signals from Fpl, Fp2, and Fpz channels on the frontal scalp by locating an EEG collector with three electrodes in the mentioned areas. Data pre-processing in this study involved firstly noisy signals removal by wavelet transform method, secondly alpha, theta, beta, and gamma linear features extraction, then normalization, and finally feature matrix construction to feed into deep models. After pre-processing and artifacts elimination, some features were extracted which included both linear and non-linear types. ANN (Artificial Neural Network) and DBN (Deep Belief Network) were deep learning algorithms used to the way of detection purpose. As a consequence, among adopted methods, DBN in combination with extracted features outperformed other ones. This study showed that we can take advantage of the Beta wave power to enhance classification accuracy. However, providing more data would be imperative to reach a more authentic and trustable accuracy. Besides, some problems are likely to arise due to the complexity of big DBN architecture.

4.1.1 Analysis of reviewed articles focused on depression detection

Table 2 presents some major and useful information about all articles in section 4.1 to provide the opportunity for highlighting the contrasts between them and delineating their main context.

Table.2. classification of articles concentrated on depression detection using deep learning

Reference	Publish year	Best method(among all adopted methods)	Advantages	Drawbacks	Contribution
Sharma et al. [33]	2021	Combined model with automatic feature extraction	• Time and computation complexity of model was low • Simple architecture	Not using any dropout layer and consequently having high probability of overfitting	Computer aided deep models
Seal et al. [34]	2021	CNN structure with automated feature extracting	Easy to practical implement since using short time window from input signal	Difficult to implement because of its dependencies on 19-channel EEG signals	CNN framework
Saeedi A et al. [35]	2021	1DCNN combined with LSTM	Converted EEG signals into	Small input data	EEG-based framework

				connecitvity-2D matrices
Khan DM [36]	2021	3D convolutional network using effective connectivity within MDD	Employed new features like PDC and DMN connectivity	The achieved accuracy is doubtful and needed more assessment	Deep architecture
Qayyum et al. [37]	Nov-20	Convolutional neural network combined with GRU	Exhibited that GRU performed better than LSTM is short sequence learning	Time window segmentation length	Proposed both GRU and LSTM to sequence learning
Thoduparambil [38]	Oct-20	Combined model of CNN and LSTM with signal of right hemisphere	Utilized flatter layers	There was not enough information about input data	Hybrid model
Kang M et al. [39]	Oct-20	2DCNN	Performed better than previous-proposed models on the same dataset	Lack of data and manually noise removal	Deep model structure
Kang M et al. [40]	Sep-20	2DCNN with alpha-band asymmetry matrix image	Took comprehensive pre-processing and feature extracting actions	Lack of data	Way of using asymmetry biomarkers
Duan L et al. [41]	Sep-20	CNN with mixed features and using 2 seconds time window	Using different length of time window	Lack of data which may lead to overfitting	Mixed feature construction
Saeedi M et al. [42]	Jul-20	MLPNN using non-linear features and gamma as linear feature	Compared deep learning and machine learning approaches	Extracted features and result should evaluate	Adopted genetic algorithms(GA) to identify appropriate features
Y. Xie et al. [43]	Jul-20	CNN with brain networks	Exploited Adjacency matrix	Did not use dropout layers	Combination of deep models with brain networks and prefrontal Lateralization
X. Zhang et al. [44]	Jul-20	CNN with demographic attention	Effects of using demographic information and not using them compared	Lack of data	Utilized demographic attention mechanism
C. Uyulan et al. [45]	Jun-20	• MobileNet with right hemisphere signals •ResNet-50 with delta power band	Made comparison between three different CNN-based models	Caused difficulty in using in clinical situation due to its complexity	Deep learning approach coupled with advanced

					computational neuroscience methodology
Z. Wan et al. [46]	2020	HybridEEGNet model with alpha band	Outperformed all five baselines models	Required more data	Extracted synchronous and regional EEG features
Mumtaz [47]	2019	One dimensional CNN with end-to-end feature extraction method	Comparison between two different models 1DCCNN and LSTM-1DCNN	Exploited methods in pre-processing steps such as time window segmentation should be evaluate	Customized deep models structure
Mahato et al. [48]	2019	Alpha frequency band with both MLPNN and RBFN	Compared the accuracy of two proposed deep models with diverse number of extracted features	Inadequate data	Deep model
X. Li et al. [49]	2019	7-layer convolutional neural network with alpha power band	Distinction between 7-layer and 8-layer CNN	Employment of AEP method which led to ignore non-distant aspects during signal to image conversion	Comparison between deep and ensemble learning
Ay B et al. [50]	2019	CNN combined with LSTM using right hemisphere signals	Obtained high accuracy	Manual elimination of artifacts	Deep hybrid model
Acharya et al. [51]	2018	Convolutional neural network with feature self-learning of signals from right hemisphere	Significant performance	More data required to evaluate the accuracy	First application of deep learning in depression diagnosis
W. Mao et al. [52]	2018	CNN using the non-distance projection method	Providing comparison between four CNN, Max-pooling, Temporal convolution, and LSTM models	The low accuracy should inspect with more data and other changes in structure	Reconstructing input data
Cai et al. [53]	2016	DBN with beta power band	Adopting three electrodes to collect input signals	Small input data	Utilizing considerable numbers of features

4.2 Deep learning methods for depression prediction using EEG signals

S. D. Kumar et al. [54] researched depression prediction by the LSTM model and the help of feature extraction. To remove artifacts of power line interference from recorded EEG signals from the left and right hemispheres of the brain, a notch filter of 50 Hz was applied. To generate input data, time-domain analysis with moving window segmentation was employed to extract the statistical mean feature. LSTM block comprised of one LSTM layer with ten hidden neurons, a dropout layer of 0.1, and a dense layer. This study compared the introduced method with two other models that were CNN-LSTM and ConvLSTM. Among them, LSTM had the best performance over those structures since it owned the smallest RMSE values (Root mean square error) as model evaluator rather than them. Although simple construction of the suggested model might bring some advantages, especially in the point of implementation, EEG signals could be investigated more to extract more efficient features or parameters to assist in attaining the goal of MDD prediction. Besides, in addition to providing more data, some enhancements in the architecture of the model could applied to improve the accuracy of prediction.

4.2.1 Analysis of reviewed articles focused on depression prediction

Table 3 displays the key concepts of the studies article in section 4.2 in order to assist to comprehend its main idea.

Table.3. detail information of article focused on depression prediction by deep learning

Reference	Publish year	Best method(among all adopted methods)	Advantages	Drawbacks	Contribution
S. D. Kumar et al. [54]	2019	LSTM with statistical mean time-domain feature	Easy to use in clinical environment	Taking inadequate actions on artifacts removal	Deep learning structure

4.3 Detailed analysis of reviewed articles

Subsections 4.1 and 4.2 are devoted to giving brief and effective summaries of all examined papers on depression detection and prediction, respectively. In this subsection, the main goal is to provide a comprehensive analysis of studies while introducing the five-step process and make a thorough comparison between works in the point of all five stages.

With respect to our notice, approximately all studied articles share the same idea of having five general steps to reach the target of depressive disorders recognition or prediction by deep learning using EEG signals. According to fig.4, five parts of the above-mentioned process include collecting EEG signals, eliminating artifacts and noise signals, performing required pre-processing operations, extracting suitable features, and finally classifying the output. Subsections 4.3.1 to 4.3.5 intends to delineate the detailed aspects of utilized methods in each phase per research. Because all of these information has been presented with the related reference, it is also possible to compare the output of studies in the point of each step.

Fig.4. Five general steps to achieve the goal of depression diagnosis by deep learning using EEG signals

4.3.1 Analysis of utilized EEG datasets to diagnose depressive cases by deep learning

The main purpose of table 4 is to impart detailed facts about main properties of datasets used in reviewed articles. Using this table assists us with making a comprehensive comparison between datasets and also can directly affect researchers' choices of data in the future.

As it can be observed, most datasets consisting of high-channels signals have been recorded from a few patients. As a consequence, considering the fact that deep learning models require lots of data to work effectively, this shortage of data might throw doubt on the accuracy of results.

Table.4. Information of used EEG datasets

Dataset collecting location	Used article	Used patient	MDD No.	Normal No.	Patient age	Hemisphere	State	Signal Duration	Sampling frequency	Channels No.
Psychiatry department of Medical College Calicut, Kerala, India.	[54],[51],[50]	30	15	15	20-50	Both	EC[1]/ EO[2]	5 Min	256 Hz	6
Lanzhou University Second Hospital	[52]	34	17	17	22-53	Both	EC	5 Min	256 Hz	16
	[49]	28	14	14	-	-	-	-	-	-
Hospital University Sains Malaysia (HUSM)	[48],[35],[42],[40],[39]	64	34	34	27-53	-	EC	5 Min	256 Hz	19
Beijing Anding Hospital, Capital Medical University	[53]	178	86	92	18-55	Both	-		-	3
	[41]	32	16	16	-	-	-	3 min	-	64
	[46]	35	11	23	21-55	-	EC	8 min	500 HZ	6
The psychology department, University of Arizona, USA[56]	[33]	45	21	24	18-25	-	-	-	500 HZ	64

Changhai Hospital in Shanghai	**[43]**	-	-	-	-	-	-	-	500 HZ	32
Psychiatric hospital	**[44]**	170	81	89	-	Both	EC	90 sec	250 HZ	3
Patient Repository for EEG data[3]	**[38]**	-	-	-	-	Both	-	-	500 H	
Hospital University Sains Malaysia (HUSM).	**[47]**	63	33	30	-	Both	EC/EO	5 Min	256 Hz	19
Neuropsychiatry Istanbul Hospital	**[45]**	92	46	46	20-45	Both	EC	4 Min	125 H	19
MDD Patients and Healthy Controls EEG Data (New)[4]	**[9], [13], [14], [22]**	74	34	30	22-53	Both	EC	5 Min	256 Hz	19

1: Eyes Closed

2: Eyes Open

3: http://predi ct.cs.unm.edu/downloads.php. Accessed 15 Jan 2020

4: https://figshare.com/articles/dataset/EEG_Data_New/42441714

4.3.2 Analysis of interference and artifacts removal methods

In order to perform pre-processing actions, EEG signals are expected to be free of any kind of noise. Tables 5 and 6 represent methods adopted by each article to address interferences and artifacts, respectively. These structures also help us to find out more about the fact that which works have taken advantages of the same method. As an example, most studies have employed A Notch Filter of 50 Hz.

Table.5. Detail information of adopted interferences elimination methods

Interference Type	Interference Elimination Method	Research
Power line	A notch filter of 50 Hz	34,35,36,39,40,42,43,45,47,48,50,53,54
Low-frequency noise	0.1-Hz high-pass filter	34
Irrelevant signals	A 100-Hz low-pass filter	34
High frequency noise, low-frequency drift and my electricity	Filtering	52
EOG	discrete wavelet transform and Kalman filtering	44
EOG	FastICA	38,49,52

Frequencies above 40 Hz	40 Hz cut off frequency low pass filter	44
Frequencies under 1 Hz	1 Hz cut off frequency high pass filter	44
Frequencies above 32/muscle artifacts	32 Hz cut off frequency low pass filter	39,40,48
Frequencies under 0.5	0.5 Hz cut off frequency high pass filter	40,48
Electromyography (EMG).	0.5–40 Hz band-pass filter	45,49,53
0.1 Hz and 70 Hz frequencies	Band-pass filter with cutoff frequencies	35,36,42
Unnecessary signals,	Finite impulse response (FIR) filter	41
Frequencies out of 0.5~100Hz	Butterworth 6th order filter	43
-	3 Hz, 35 Hz, and 50 Hz cutoff frequencies band-stop filters	45

Table.6. Detail information of adopted artifacts removal methods

Artifacts type	Artifact removal method	Research
Eye blinks	Independent component analysis (ICA)	33,34,39,40,48
Eye blinks	Manually(an experienced neurophysiologist)	40,51
Eye blinks	Multiple source eye correction (MSEC) method	47
Eye blinks	Wavelet transform method	42,45,53
Eye blinks	EEGLAB software	36
Eye blinks	Low-pass filter(above 32Hz)	39
Muscles movement	Manually(an experienced neurophysiologist)	50,51
Muscles movement	Multiple source eye correction (MSEC) method	47
Muscles movement	Independent component analysis (ICA)	40,48
Muscles movement	EEGLAB software	36
Muscles movement	Low-pass filter(above 32Hz)	39
Muscles movement	Wavelet transform method	42,45
Heart beats	Multiple source eye correction (MSEC) method	47
Sensor motions	EEGLAB software	36

4.3.3 Analysis of pre-processing operations

To utilize EEG signals to fulfill the goal of depression detection or prediction, they are supposed to undergo some alterations by which they get prepared to use in the next step to extract features. Table 7 shows some information related to this matter for each article.

Table.7. Detail information of taken pre-processing actions

Research	Preprocess Methods
Sharma et al. [33]	• Windowing method of fixed size • Transforms the time-series EEG data into cross-sectional data

Research	Method
Seal et al. **[34]**	• Segmentation of 4-second window time • Convert signals to 2D matrix
Saeedi A et al. **[35]**	Transformation of 1-D EEG signals into 2-D images using the combination of eight extracted frequency bands and extracted effective connectivity images in pair
Khan DM **[36]**	Segmentation of 2-second window time
Qayyum et al. **[37]**	Segmentation of 1-second window time
Thoduparambil **[38]**	Offset effect removal and controlling amplitude scaling problem using Z-score normalization
Kang M et al. **[39]**	-
Kang M et al. **[40]**	• Using min-max normalization • Segmentation of 4 seconds window time • Asymmetry matrix image creation using delta, theta, alpha, and beta frequency bands
Duan L et al. **[41]**	Construction of feature matrix using extracted features
Saeedi M et al. **[42]**	Using Genetic algorithm (GA) as main feature selection method
Y. Xie et al. **[43]**	Adjacency matrix of functional connectivity of brain networks calculation in five extracted bands using Phase Lag Index (PLI)
X. Zhang et al. **[44]**	Partition of 40s high-quality EEG signals into 4 seconds non-overlapping sliding windows
C. Uyulan et al. **[45]**	Converting 1D EEG signal into 2D EED image matrix
Z. Wan et al. **[46]**	Offset effect removal and controlling amplitude scaling problem using Z-score normalization
Mumtaz **[47]**	Segmentation of 1-second window time
Mahato et al. **[48]**	Reduction the dimension of non-linear features by Principal component analysis (PCA)
X. Li et al. **[49]**	Segmentation of 0.2 s, 0.5 s, 1 s ,2 s window time
Ay B et al. **[50]**	-
Acharya et al. **[51]**	Normalization using the Z-score
W. Mao et al**. [52]**	• Interception of 270 seconds segmentation in length from 15 s to 285 s • Employing power spectrum density and two projection models, namely, distance-based and non-distance methods to preserving the spatial and temporal patter in EEG signals
Cai et al. **[53]**	Normalization

4.3.4 Analysis of utilized methods in feature extraction steps

Table 8 explains how each research work has extracted required features from prepared input signals. Most research works have enjoyed convolutional layers as an end-to-end manner to extract features, among which frequency waves features like Alpha, Beta, and so forth have been preferred more.

Table.8. Detail information of feature extraction stage

Research	Feature extraction method	Extracted features

Sharma et al. [33]	• Fast Fourier Transform method • Convolutional layer (end-to-end)	Time-frequency features and local features
Seal et al. [34]	Convolutional layer (end-to-end)	Temporal and spatial information
Saeedi A et al. [35]	• MATLAB (The Math works, Inc., Natick, MA, USA) via the open-source SIFT toolbox) • GPDC and DDTF	• Delta (1–4 Hz), Theta (4–7 Hz), Alpha or Mu (8–12 Hz), Beta1 (12–15 Hz), Beta2 (15–22 Hz), Beta3 (22–30 Hz), Beta (12–30 Hz), Gamma (30–70 Hz) • Effective connectivity image between EEG signals
Khan DM [36]	1. Connectivity matrices 2. DMN connectivity matrices	1. PDC (Partial Directed Coherence) connectivity calculation 2. DMN connectivity extraction
Qayyum et al. [37]	End-to-end	Local characteristics
Thoduparambil [38]	Convolutional layer (end-to-end)	Local features and signals sequence
Kang M et al. [39]	Band-pass filter	Delta (0.5~4), Theta (4~8), Alpha (8~13) Beta (13 ~32)
Kang M et al. [40]	Convolutional layer (end-to-end)	• Spatial features • Asymmetry feature • Image feature
Duan L et al. [41]	1. Fast Fourier transform (FFT) 2. Calculation based on extracted frequency bands	1. Theta (4–8Hz), Alpha (8–13Hz), and Beta (13–40Hz) 2.1 Construct Interhemispheric asymmetry 2.2 Cross-correlation using frequency bands 2.3 structural changes 2.4 connectivity changes
Saeedi M et al. [42]	1. Wavelet packet decomposition 2. Decomposing data	1.1 Approximate entropy 1.2 Sample entropy 2. Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), Gamma (30–50 Hz)
Y. Xie et al. [43]	-	-
X. Zhang et al. [44]	Convolutional layer (end-to-end)	Spatial-temporal features
C. Uyulan et al. [45]	1. 12th-order Butterworth band pass filter 2. Convolutional layers	1. Delta (2-4), Theta (4.5-7.5), Alpha (8-12.5), Beta (13, 30) 2. Feature maps of local features
Z. Wan et al. [46]	Convolutional filters	Synchronous and regional features
Mumtaz [47]	1. 1D Convolutional layer (end-to-end) 2. 1D Convolutional layer +LSTM (end-to-end)	1. Temporal 2. Temporal and spatial
Mahato et al. [48]	1. Welch period gram 2. Wavelet transform (WT)	1. Linear features: • Alpha, beta, delta, gamma frequency band power • Alpha, beta, delta, gamma Interhemispheric asymmetry 2. Non-linear features: • Relative wavelet energy (RWE) • Wavelet entropy (WE) • Approximate entropy (ApEn)
X. Li et al. [49]	1. Fast Fourier Transform (FFT) 2. Auto-regress (AR)	1. Three frequency bands (i.e. theta, alpha, beta) 2.1 Frequency domain feature (i.e. power spectral density) 2.2 Time domain feature (i.e. activity)

Ay B et al. [50]	Convolutional layer (end-to-end)	Deep local features
Acharya et al. [51]	Convolutional layer (end-to-end)	
W. Mao et al. [52]	1. Hanning filter 2. Deep neural networks	1. Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz) bands 2. Iinherent spatial and temporal patterns
Cai et al. [53]	End-to-end	85 linear or non-linear features for four different waveform of EEG signals: Theta wave (4-8Hz), Alpha wave (8-15Hz), Beta wave (13-30Hz), and Gamma wave (30-50Hz)
S. D. Kumar et al. [54]	Time domain analysis	Statistical mean feature

4.3.5 Analysis of exploited deep learning classifiers to diagnose or predict depression

Table 9 provides the summary of used classifier(s), obtained results, and some related facts in articles. In fact, this is aimed to convey the gist of the detection or prediction stage at a glance.

Table.9. Detail information of deep classifiers

Research	Publish year	Deep Classifier(s)	Model performance evaluator	Accuracy percentage	Best accuracy	Activation function of Classifier	DP No./ DP rate [56]
Sharma et al. [33]	2021	LSTM-CNN	-	99.10%	-	Sigmoid	not used
Seal et al. [34]	2021	CNN	-	99.37	-	Softmax	2 / 0.2
Saeedi A et al. [35]	2021	1. 1DCNN 2. 2DCNN 3. LSTM 4. 1DCNN-LSTM 5. 2DCNN-LSTM	-	1. 95.283% ± 2.109 2. 96.226% ± 1.208 3. 89.057% ± 1.849 4. 99.245% ± 1.152 5. 96.415% ± 3.422	1DCNN+LSTM	1. Sigmoid 2. Sigmoid 3. Tanh 4. Sigmoid 5. Sigmoid	• 3 / 0.5 • 1 / 0.5 • 3 / 0.5 • 2 / 0.25 + 1 / 0.5
Khan DM [36]	2021	3DCNN	10/15-fold cross-validation	100%	-	Softmax	1 / 0.2
Qayyum et al. [37]	Nov-20	IDCNN-LSTM	-	99.66	-	Sigmoid	2
Thoduparambil [38]	Oct-20	CNN-LSTM	Random splitting and tenfold cross-validation	• Right hemisphere: 99.07% • Left hemisphere: 98.84%	Right hemisphere	Softmax	1 / 0.2
Kang M et al. [39]	Oct-20	2DCNN	10-fold cross-validation	96.97	-	-	-
Kang M et al. 40]	Sep-20	2DCNN	5-fold cross-validation	98.85%	Asymmetry matrix image in the alpha band	Sigmoid	-
Duan L et al. [41]	Sep-20	CNN	-	94.13%	With the mixed features for the 2s time window	-	3 / 0.5

Saeedi M et al. **[42]**	Jul-20	MLPNN	Tenfold cross validation (10-CV)	88.33%(± 7.2)	-	Sigmoid	-
Y. Xie et al. **[43]**	Jul-20	1. CNN 2. DBN	-	1. Brain Networks + CNN-2 67.67 Prefrontal Lateralization + CNN-2 55.88 2. Brain Networks + DBN 55.56 Prefrontal Lateralization + DBN 62.5	CNN with brain networks	-	-
X. Zhang et al. **[44]**	Jul-20	1DCNN	-	75.29%	-	Softmax	1 / 0.5
C. Uyulan et al. **[45]**	Jun-20	1. ResNet-50 2. MobileNet 3. Inception-v3	5-fold cross-validation	• 92.66% • 90.22%	• MobileNet with right hemisphere signals • ResNet-50 with delta power band	Sigmoid	-
Z. Wan et al**. [46]**	2020	CNN	Tenfold cross-validation method	79.08%	-	Softmax	1 /0.9
Mumtaz **[47]**	2019	1. 1DCNN 2. 1DCNN-LSTM	10-fold cross validation using 19 features	1. 98.32% 2. 95.97%,	1DCNN	Sigmoid	1. 1 / 0.2 2. 2 / 0.2
Mahato et al. **[48]**	2019	• MLPNN • RBFN	-	93.33 ± 1.67%	Combination alpha power and RWE in both MLPNN and RBFN using PCA for reduction	-	-
X. Li et al. **[49]**	2019	1. CNN 2. CNN	-	0.8475	7-Conv using the activity in alpha frequency band	Softmax	0.5
Ay B et al. **[50]**	2019	CNN-LSTM	-	• Avg: 97.82% • Right: 98% • Left: 98.94%	-	Softmax	2 / 0.2
Acharya et al. **[51]**	2018	CNN	-	Left: 93.54 Right: 95.9	Right hemisphere	Softmax	2 / 0.9

W. Mao et al. [52]	2018	1. CNN 2. Max-pooling 3. Temporal convolution 4. LSTM	8-fold cross-validation	• CNN Dis:76.14% ± 6.86% Non-dis :77.20% ± 9.43% • Temporal conv Dis:67.78% ± 7.68% Non-dis :70.95% ± 12.38% • MAX Dis:67.58% ± 5.83% Non-dis :73.53% ± 10.45% • LSTM Dis: 70.94% ± 8.08% Non-dis :67.33% ± 12.59%	CNN using the non-distance projection method	Softmax	1 / 0.5
Cai et al. [53]	2016	DBN (Deep Belief Network)		0.7824	-	-	-
S. D. Kumar et al. [54]	2019	LSTM	RMSE (Root mean square error)	RMSE= 0.000064	-	-	1 / 0.1

5. Discussion and comparison

With regard to the adopted SLR method, previous sections were aimed to provide a comprehensive explanation about the procedure of papers selection, together with summaries of ultimate studies to get the gist of them and detailed analyses of all research works according to the defined five-step pattern. This part offers some statistical analysis and reports of studied articles in this review. Furthermore, the suggested analytical questions in section 3 are provided with complete answers.

AQ1: What deep learning algorithms have been used to detect or predict depression?

A variety of deep learning algorithms utilized to pursue the goal of diagnosis or prediction of depressive subjects with EEG signals has been illustrated with fig.5 according to the designed taxonomy in section 4. Among the displayed methods, LSTM has used for prediction purposes. Majority of papers exploited at least two deep learning structures and applied them on the same prepared dataset to provide more opportunity of getting the best accuracy. Combined models of convolution neural network structure and LSTM blocks have obtained better results than other ones because these representations enjoyed the advantages of memorization capability of LSTM architecture and built-in feature extracting property of CNN constructions. Referring to reviewed studies, it can conclude that using some external methods and combining them with a deep learning structure will also give a significant accuracy, as this technique has been adopted by [42, 51].

Fig. 5. Diversity of adopted deep learning algorithms in evaluated researches

AQ2: Which deep learning methods are mostly preferred to fulfill our goal?

Fig 6 shows the distribution of mentioned deep learning architectures throughout all examined researches. It can see that CNN-based methods include CNN, 1DCNN, 2DCNN, and 3DCNN, which together accounted for approximately half of the utilized deep learning classifiers, which were preferred to other algorithms. Among this category, CNN with 32.4% of the whole percentage was considered the most popular deep learning structure to accomplish the objective of depression discrimination or foreseen.
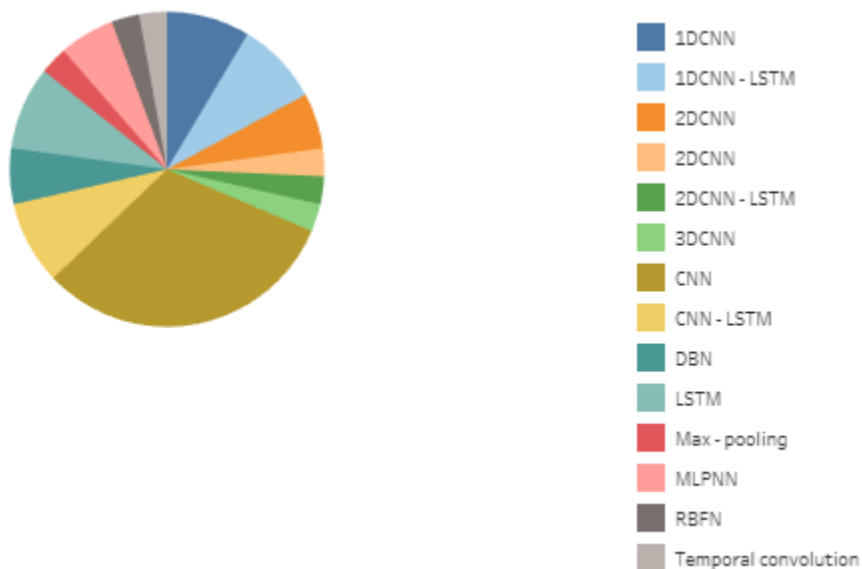
Fig. 6. Distribution of adopted deep learning models across all studied papers

AQ3: What approaches are mainly used to extract features from EEG signals?
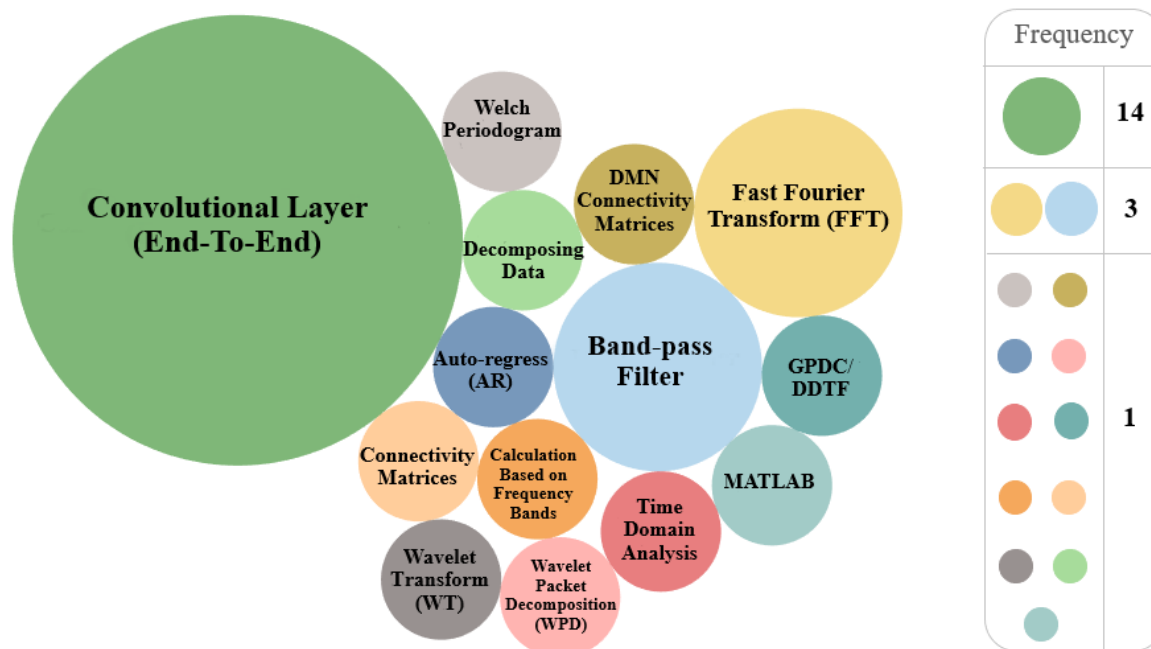
Fig. 7. Frequencies of adopted feature extraction methods

Fig 7 portrays the numbers of usage of techniques employed to extract required features from recorded EEG signals. As it can perceive from the depiction, most studies have selected convolutional layer(s) to derive features as an end-to-end technique. It is true that there are possibilities of extracting some defined features, this manner sounds safe because all aspects of extraction process are handled by the deep model. The second place of frequency has simultaneously occupied by band-pass filter and FFT methods with 3 times usage. Other ones were exploited one time. We observed that some researchers practiced two or more mentioned feature extraction approaches to produces different or combined sets of features.

5.2 Open issues and future works

This article has been organized based on the SLR method. Considering the evaluated researches and their chief characteristics such as main context, advantage, drawback, and contribution, some challenges still need to be addressed. These matters are presented as AQ4:

AQ4: What are the future and open works related to depression detection or prediction?

In this part, issues that studies had confronted with, are discussed according to the introduced five-step procedure, including input data source, data preparation (noise removal and pre-processing) , and implementation of deep learning model. In addition, we will look into a general problem about depression prediction and the importance of resolving it.

Inadequate input data: The lack of input data was the first and major problem that posed serious difficulties for all studied researchers. Overfitting and unreliable accuracy were two consequences from which the detection or prediction process suffered. As we have noticed, some studies strived to overcome these issues. Exploiting one or more dropout layers with various probability rates was one technique that was applied to deal with overfitting. Tackling overfitting was also carried out by utilizing the windowing method and sample generation to produce substantial data records. However, although these approaches might alleviate these difficulties, the problem of insufficient data remains an open issue and would be of primary importance to cope with.

Lack of specific and effective technique for data pre-processing: one problem that almost all works had in common was their challenge for preparing data to use for prediction or detection. In the point of noise removal, some of them have accomplished this task manually, others have utilized various experience-based methods to fulfill this purpose, which have been listed in tables 5 and 6. This issue can be generalized to pre-processing steps too, in which diverse manners such as using time window segmentation of different lengths, AEP method, or transforming signals to images can negatively affect the accuracy of the proposed model. The main reason is rooted in the elimination of some important features of input signals by applying these experimental techniques, loss of non-distance properties of signals by conversion to images for example. So, a specific and reliable way of data preparation is required.

Implementing and usability: Another key problem with which papers have encountered was the implementation of their proposed method(s). Due to the fact that the main objective of the conducted researches was to present a practical method to assist clinical environments and related medical doctors, the construction and architecture of some introduced models have not met this important aspect. To achieve the high accuracy, they usually have adopted deep learning models with many layers that this complexity can erect some barriers in the way of using these proposed methods in depression detection or prediction in real situations. Moreover, as another issue in this regard, in some studies, EEG signals were recorded with a lot of electrodes, such as signals of 19 channels. Clinical situations might face considerable difficulties in the process of collecting these kinds of high-channel signals. Thereby, designing a model compatible with the clinical setting is an important aspect that is more expected to focus on.

Giving less attention to depression prediction: As it has been sought in different search engines for papers on predicting depression, unfortunately, very few works have been done in this respect, in a way that we only could review one article in accordance with our paper selection criteria. Depression, other than seriously affecting people's health or costing their life [57], produces damaging effects on patients' financial condition [58] and imposes some difficulties on the healthcare system and economy [59]. Above these factors, on many occasions, depression is diagnosed late for a variety of reasons, by which sufferers' treatment will face difficulties, and sometimes, it becomes incurable [60]. Considering all these mentioned elements, put more emphasis on the importance of devoting more time and attention to research into the prediction of this dangerous disease.

6. Conclusion

This paper has concentrated on depression detection and prediction by deep learning algorithms using EEG signals. According to the SLR method which has been applied to this study, a comprehensive review was conducted in which some researches focused on our specified topic were evaluated and their main aspects were inspected. Moreover, open issues and futures research directions were discussed. Regarding our purpose and also the fact that most articles drew a comparison between the results of two or more deep learning algorithms on the same prepared dataset, the taxonomy was designed based on a collection of all

used deep learning methods from all researches. By analyzing 22 articles as an output of a complete, elaborated, SLR-based refinement, it was concluded that CNN-based deep learning methods, namely, CNN, 1DCNN, 2DCNN, and 3DCNN are by far the more preferable group among the diverse adopted algorithms with almost 50% of the total in sum. In this classification, CNN with approximately one-third of the whole took first place. Combined models of both mentioned CNN-based algorithms and LSTM blocks were second only to the CNN-based category in popularity. Regarding AQ3, it was also inferred that different researchers have taken advantage of adopting various feature extraction methods to build more appropriate models. Amongst these techniques, the majority of papers aimed to utilize convolutional layers as an end-to-end technique to extract local features. As it has been determined, all analyzed studies follow almost the same procedure of collecting EEG signals, artifacts and noise removal of them, extracting required features from pre-processed signals, and finally, classify depressive and normal subjects using one or more deep learning methods. To sum up, with respect to the SLR method and according to our goal, we aimed to present a comprehensive review to assist next pieces of research with a strong background of this area.

Abbreviations

1DCNN	**One Dimensional Convolution Neural Network**
2DCNN	Two Dimensional Convolution Neural Network
3DCNN	Three Dimensional Convolution Neural Network
AEP	Auditory Evoked Potential
AF	Activation Function
ANN	Artificial Neural Network
AQ	Analytical Question
AR	Auto-regress
ASR	Artifact Subspace Reconstruction
BD	Bipolar Disorder
BN	Batch Normalization
CAD	Computer-aided
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DDTF	Direct Directed Transfer Function
DL	deep learning
DMN	Default Mode Network
DP	dropout layer
DWT	Discrete Wavelet Transform
EC	eye closed
EEG	Electroencephalogram
EO	eye open
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GA	Genetic Algorithm
GPDC	Generalized Partial Directed Coherence
GRU	Gated Recurrent Units

ICA	Independent Component Analysis
IOT	Internet of Things
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LSTM	Long Short Term Memory
MDD	Major Depressive Disorder
MLPNN	multi-layered perceptron neural network
MSEC	Multiple Source Eye Correction
NO	number
PDC	Partial Directed Coherence
PLI	Phase Lag Index
PSD	Power Spectral Density
RBFN	radial basis function network
ReLU	Rectified Linear Unit
RF	Random Forest
RWE	Relative Wavelet Energy
SLR	systematic literature review
SVM	Support Vector Machine
VGG	Visual Geometry Group
WE	Wavelet Entropy
WHO	World Health Organization

Compliance with Ethical Standards:

Atefeh Safayari declares that she has no conflict of interest.

Hamidreza Bolhasani declares that he has no conflict of interest.

Ethical approval:

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent:

Not applicable.

References

[1] W.H. Organization, et al., Depression and Other Common Mental Disorders: Global Health Estimates, Technical Report, World Health Organization, 2017.

[2] Mazza, Mario Gennaro, et al. "Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors." *Brain, behavior, and immunity* 89 (2020): 594-600.

[3] Tang F, Liang J, Zhang H, Kelifa MM, He Q, Wang P. COVID-19 related depression and anxiety among quarantined respondents. Psychology & health. 2021 Feb 1; 36(2):164-78.

[4] Salari, Nader, et al. "Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis." *Globalization and health* 16.1 (2020): 1-11.

[5] Chen F, Zheng D, Liu J, Gong Y, Guan Z, Lou D. Depression and anxiety among adolescents during COVID-19: A cross-sectional study. Brain, behavior, and immunity. 2020 Aug 1.

[6] Bueno-Notivol, Juan, et al. "Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies." *International journal of clinical and health psychology* 21.1 (2021): 100196.

[7] A. Swetaa, R. Gayathri, and V. V. Priya, "Awareness of mental health among teenagers," *Drug Invention Today*, vol. 11, no. 8, pp. 1979–1982, 2019.

[8] J. Shen *et al.*, "Depression Detection from Electroencephalogram Signals Induced by Affective Auditory Stimuli," *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 76-82, doi: 10.1109/ACII.2019.8925528.

[9] Acharya UR, Sudarshan VK, Adeli H, Santhosh J, Koh JE, Puthankatti SD, Adeli A. A novel depression diagnosis index using nonlinear features in EEG signals. European neurology. 2015; 74(1-2):79-83.

[10] Deng, Li, and Dong Yu. "Deep learning: methods and applications." *Foundations and trends in signal processing* 7.3–4 (2014): 197-387.

[11] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016 Nov 18.

[12] Andreas Kamilaris, Francesc X. Prenafeta-Boldú, Deep learning in agriculture: A survey, Computers and Electronics in Agriculture, Volume 147, 2018, Pages 70-90.

[13] Rusk N. Deep learning. Nature Methods. 2016 Jan; 13(1):35-.

[14] Luckow A, Cook M, Ashcraft N, Weill E, Djerekarov E, Vorster B. Deep learning in the automotive industry: Applications and tools. In2016 IEEE International Conference on Big Data (Big Data) 2016 Dec 5 (pp. 3759-3768). IEEE.

[15] Grigorescu S, Trasnea B, Cocias T, Macesanu G. A survey of deep learning techniques for autonomous driving. Journal of Field Robotics. 2020 Apr; 37(3):362-86.

[16] Bolhasani H, Mohseni M, Rahmani AM. Deep learning applications for IoT in health care: A systematic review. Informatics in Medicine Unlocked. 2021 Mar 17:100550.

[17] Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. JAMA internal medicine. 2019 Mar 1; 179(3):293-4.

[18] J. Ker, L. Wang, J. Rao and T. Lim, "Deep Learning Applications in Medical Image Analysis," in *IEEE Access*, vol. 6, pp. 9375-9389, 2018, doi: 10.1109/ACCESS.2017.2788044.

[19] A. Akay and H. Hess, "Deep Learning: Current and Emerging Applications in Medicine and Technology," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 3, pp. 906-920, May 2019, doi: 10.1109/JBHI.2019.2894713.

[20] Moazemi S., Essler M., Schultz T., Bundschuh R.A. (2021) Predicting Treatment Response in Prostate Cancer Patients Based on Multimodal PET/CT for Clinical Decision Support. In: Syeda-Mahmood T. et al. (eds) Multimodal Learning for Clinical Decision Support. ML-CDS 2021. Lecture Notes in Computer Science, vol 13050. Springer, Cham. https://doi.org/10.1007/978-3-030-89847-2_3

[21] Bengio Y, LeCun Y. Scaling learning algorithms towards AI. Large-scale kernel machines. 2007 Aug; 34(5):1-41.

[22] Y. Ding, X. Chen, Q. Fu, S. Zhong, A depression recognition method for college students using deep integrated support vector algorithm, IEEE Access (2020).

[23] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," J. Neural Eng., vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[24] S. In Cho, "Vision-based people counter using CNN-based event classification," IEEE Trans. Instrum. Meas., vol. 69, no. 8, pp. 5308–5315, Aug. 2020.

[25] H. Wang, S. Li, L. Song, L. Cui, and P. Wang, "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," IEEE Trans. Instrum. Meas., vol. 69, no. 6, pp. 2648–2657, Jun. 2020.

[26] Yasin, S., Hussain, S. A., Aslan, S., Raza, I., Muzammel, M., & Othmani, A. (2021). EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks: A review. *Computer Methods and Programs in Biomedicine*, 106007.

[27] Su C, Xu Z, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. Translational Psychiatry. 2020 Apr 22; 10(1):1-26.

[28] Khosla A, Khandnor P, Chand T. A comparative analysis of signal processing and classification methods for different applications based on EEG signals. Biocybernetics and Biomedical Engineering. 2020 Apr 1; 40(2):649-90.

[29] C. Jatoth , G.R. Gangadharan , R. Buyya , Computational intelligence based QoS-aware web service composition: a systematic literature review, IEEE Trans. Serv. Comput. 10 (3) (2017) 475–492.

[30] E. Jafarnejad Ghomi , A. Masoud Rahmani , N. Nasih Qader , Load-balancing al- gorithms in cloud computing: a survey, J. Netw. Comput. Appl. 88 (Supple- ment C) (2017) 50–71.

[31] M. Effatparvar , M. Dehghan , A.M. Rahmani , A comprehensive survey of en- ergy-aware routing protocols in wireless body area sensor networks, J. Med. Syst. 40 (9) (2016) 201.

[32] Asghari, Parvaneh, Amir Masoud Rahmani, and Hamid Haj Seyyed Javadi. "Internet of Things applications: A systematic review." *Computer Networks* 148 (2019): 241-261.

[33] Sharma G, Parashar A, Joshi AM (2021) DepHNN: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. Biomedical Signal Processing and Control 66:102393.

[34] Seal, Ayan, et al. "DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG." *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1-13.

[35] Saeedi A, Saeedi M, Maghsoudi A, Shalbaf A (2020) Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach. Cognit Neurodyn, 1.

[36] Khan DM, Yahya N, Kamel N, Faye I. Automated Diagnosis of Major Depressive Disorder Using Brain Effective Connectivity and 3D Convolutional Neural Network. IEEE Access. 2021 Jan 5; 9:8835-46.

[37] Qayyum, Abdul, Imran Razzak, and Wajid Mumtaz. "Hybrid Deep Shallow Network for Assessment of Depression Using Electroencephalogram Signals." In International Conference on Neural Information Processing, pp. 245-257. Springer, Cham, 2020.

[38] P. P. Thoduparambil, A. Dominic and S. M. Varghese, "EEG-based deep learning model for the automatic detection of clinical depression", Phys. Eng. Sci. Med., vol. 43, pp. 1349-1360, Oct. 2020.

[39] Kang M, Park J, Kang S, Lee Y. Low channel electroencephalogram based deep learning method to pre-screening depression. In2020 International Conference on Information and Communication Technology Convergence (ICTC) 2020 Oct 21 (pp. 449-451). IEEE.

[40] Kang M, Kwon H, Park JH, Kang S, Lee Y. Deep-Asymmetry: Asymmetry Matrix Image for Deep Learning Method in Pre-Screening Depression. Sensors. 2020 Jan; 20(22):6526.

[41] Duan L, Duan H, Qiao Y, Sha S, Qi S, Zhang X, Huang J, Huang X, Wang C (2020) Machine learning approaches for MDD detection and emotion decoding using EEG signals. Front Hum Neurosci 14.

[42] Saeedi, M.; Saeedi, A.; Maghsoudi, A. Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. Phys. Eng. Sci. Med. 2020, 43, 1007–1018.

[43] Y. Xie, B. Yang, X. Lu, M. Zheng, C. Fan, X. Bi, et al., "Anxiety and depression diagnosis method based on brain networks and convolutional neural networks", Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), pp. 1503-1506, Jul. 2020.

[44] X. Zhang, J. Li, K. Hou, B. Hu, J. Shen, J. Pan, et al., "EEG-based depression detection using convolutional neural network with demographic attention mechanism", Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), pp. 128-133, Jul. 2020.

[45] C. Uyulan, T. T. Ergüzel, H. Unubol, M. Cebi, G. H. Sayar, M. N. Asad, et al., "Major depressive disorder classification based on different convolutional neural network models: Deep learning approach", Clin. EEG Neurosci., vol. 52, no. 1, pp. 38-51, 2020.

[46] Z. Wan, J. Huang, H. Zhang, H. Zhou, J. Yang and N. Zhong, "HybridEEGNet: A Convolutional Neural Network for EEG Feature Learning and Depression Discrimination", IEEE Access, vol. 8, pp. 30332-30342, 2020.

[47] W. Mumtaz, and A. Qayyum, A deep learning framework for automatic diagnosis of unipolar depression. International Journal of Medical Informatics, 132 (2019) 103983.

[48] Mahato, S., Paul, S., Detection of major depressive disorder using linear and non-linear features from EEG signals. In: Microsystem Technologies, Springer, 2018, pp 1–12, (2018).

[49] X. Li, X. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, C. Xia, B. Hu, Depression recognition using machine learning methods with different feature generation strategies, Artificial intelligence in medicine 99 (2019) 101696.

[50] Ay B, Yildirim O, Talo M, Baloglu UB, Aydin G, Puthankattil SD, Acharya UR (2019) Automated depression detection using deep representation and sequence learning with EEG signals. J Med Syst 43(7):205.

[51] Acharya U R, Lih S, Hagiwara Y, Hong J, Adeli H and Subha D P 2018 Automated EEG-based screening of depression using deep convolutional neural network Comput. Methods Progr. Biomed. 161 103–13.

[52] W. Mao, J. Zhu, X. Li, X. Zhang, S. Sun, Resting state eeg based depression recognition research using deep learning method, in: International Conference on Brain Informatics, Springer, 2018, pp. 329–338.

[53] Cai, H., Sha, X., Han, X., Wei, S., and Hu, B., Pervasive EEG diagnosis of depression using Deep Belief Network with three-electrodes EEG collector. In 2016 IEEE International Conference on Bioinformatics and Biomedicine, 1239–1246, 2016.

[54] S. D. Kumar and D. Subha, "Prediction of depression from EEG signal using long short term memory (LSTM)", *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, pp. 1248-1253, Apr. 2019.

[55] Frontiers | Larger Error Signals in Major Depression are Associated with Better Avoidance Learning | Psychology (frontiersin.org)

[56] Alekh Agarwal. Sahand Negahban. Martin J. Wainwright. "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions." Ann. Statist. 40 (2) 1171 - 1197, April

[57] Saint Onge JM, Krueger PM, Rogers RG. The relationship between major depression and nonsuicide mortality for US adults: the importance of health behaviors. Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2014 Jul 1; 69(4):622-32.

[58] Kessler RC. The costs of depression. Psychiatric Clinics. 2012 Mar 1;35(1):1-4.

[59] Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of Lost Productive Work Time Among US Workers with Depression. JAMA 2003; 289: 3135–44.

[60] Mehltretter J, Fratila R, Benrimoh D, Kapelner A, Perlman K, Snook E, Israel S, Armstrong C, Miresco M, Turecki G. Differential treatment benefit prediction for treatment selection in depression: a deep learning analysis of STAR* D and CO-MED data. Computational Psychiatry. 2020 Dec;4:61-75.