

# An Overview of Deep Learning Algorithms and Their Applications in Neuropsychiatry

Gokhan Guney<sup>1,\*</sup>, Busra Ozgode Yigin<sup>1,\*</sup>, Necdet Guven<sup>1</sup>, Yasemin Hosgoren Alici<sup>2</sup>, Burcin Colak<sup>3</sup>, Gamze Erzin<sup>4</sup>, Gorkem Saygili<sup>1</sup>

<sup>1</sup>Department of Biomedical Engineering, Ankara University, <sup>2</sup>Department of Psychiatry, Baskent University, <sup>3</sup>Department of Psychiatry, Ankara University, <sup>4</sup>Department of Psychiatry, Ankara Dışkapı Training and Research Hospital, Ankara, Turkey

Deep learning (DL) algorithms have achieved important successes in data analysis tasks, thanks to their capability of revealing complex patterns in data. With the advance of new sensors, data storage, and processing hardware, DL algorithms start dominating various fields including neuropsychiatry. There are many types of DL algorithms for different data types from survey data to functional magnetic resonance imaging scans. Because of limitations in diagnosing, estimating prognosis and treatment response of neuropsychiatric disorders; DL algorithms are becoming promising approaches. In this review, we aim to summarize the most common DL algorithms and their applications in neuropsychiatry and also provide an overview to guide the researchers in choosing the proper DL architecture for their research.

**KEY WORDS:** Deep learning; Neuropsychiatry; Artificial neural networks; Convolutional neural networks; Recurrent neural networks; Generative adversarial networks.

## INTRODUCTION

During the past few decades diagnosing neuropsychiatric disorders, finding etiology and predicting prognosis has been dragging considerable attention. Various psychiatric disorders such as bipolar, schizophrenia spectrum, anxiety, addiction, etc. exhibit common pathological features in terms of genetic, behavioral, and neuroimaging patterns which causes a challenge in diagnosis and prognosis [1]. At the point where the classical research methods are inefficient, machine learning algorithms in particular deep learning (DL), provide promising solutions with their complex, nonlinear nature. Their success in revealing complex patterns under data has a remarkable impact on prediction, classification and data analysis. Moreover,

their flexibility, adaptation, and learning capability sourcing from millions of parameters make DL algorithms essential in big data analysis [2].

Conventional machine learning techniques (shallow learners) require distinguishing features for successful classification. This brings a nontrivial challenge especially when there is a common pattern between different groups in data. In contrast, DL algorithms do not require explicitly extracted features since they can extract their own from raw data. While their feature extraction capability brings a demand for more training data, it increases their flexibility to learn complex and distinguishing patterns between different classes in data [3].

Considering all of the above-mentioned advantages, DL algorithms are essential tools in classification of neuropsychiatric disorders and foreseeing their prognosis. Therefore, DL algorithms start to lead the research progress in neuropsychiatry.

In this review, we summarized different types of DL algorithms and their applications in the diagnosis of neuropsychiatric disorders. Firstly, we defined the basics and modified types of networks. Then we discussed the appli-

**Received:** June 30, 2020 / **Revised:** August 31, 2020

**Accepted:** September 5, 2020

**Address for correspondence:** Gorkem Saygili

Department of Biomedical Engineering, Ankara University,  
Golbasi 50. yil Yerleskesi Bahcelievler Mh, K Blok, Ankara 06830,  
Turkey

E-mail: gorkemsaygili@ankara.edu.tr

ORCID: <https://orcid.org/0000-0002-9049-2138>

\*These authors contributed equally to this study.

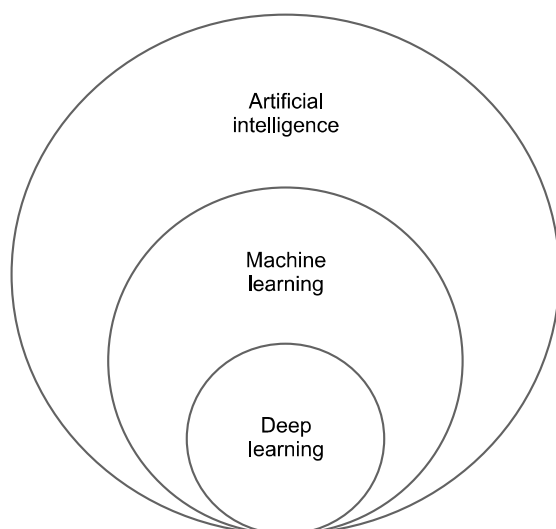
© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

cations of DL algorithms on the most common neuropsychiatric disorders. Finally, we elaborate on their limitations and discuss possible future directions. Although there are many types of DL algorithms, in this review, we limit our attention to the types that are commonly used in the diagnosis of the most common neuropsychiatric disorders.

## DEEP LEARNING CONCEPTS AND ARCHITECTURES

“Artificial intelligence” (AI) was exposed in the 1950s and refers to the ability of machines to perform some operations as skillfully as humans. Data mining is the application of algorithms used to reveal similar motifs with similar sequences in the data. “Machine learning” emerged in the 1980s and has become more popular with the use of data mining. “Deep learning” started to be used in the 2010s and it is known as a custom type of machine learning. It is a learning model that performs calculations used in machine learning over many layers at once, discovers the features that are needed to be hand-crafted in conventional machine learning, and reveals highly complex patterns inside the data. Figure 1 shows the relationship between these different AI disciplines [3].

To understand the underlying concept of DL, the basics of machine learning algorithms must be understood. Machine learning algorithms are described as algorithms



**Fig. 1.** The diagram shows deep learning is a subfield of machine learning, which is a subfield of artificial intelligence.

that learn the intrinsic pattern of data and generally categorized into three classes namely, supervised, unsupervised and semi-supervised learning. In supervised learning, there is a corresponding label to specify the target for every input. For example, an image classification algorithm is fed using images with labels from different categories (a structural magnetic resonance imaging [sMRI] of a subject with schizophrenia spectrum and healthy control, etc.). The algorithm learns the underlying structure of data according to given labels and produces an output score that corresponds to the target label [3]. In unsupervised learning, the algorithm learns the intrinsic structure of the data without providing any labels. These algorithms can be used to cluster similar examples in the data and these algorithms are helpful to identify the anomalies in the dataset [4]. Semi-supervised learning can be considered as halfway between supervised and unsupervised learning. In this learning, the algorithm is fed with some labeled examples in addition to unlabeled data [5]. Usually, this learning is used to improve the accuracy of a classifier with additional unlabeled data, commonly used in fields such as natural language processing, computer vision, etc. The process of exploring the intrinsic pattern of data is called training. In this process, using training data (split from the whole data), the algorithm is trained and some error metrics (training error, the difference between real and calculated output for train data) are calculated through the parameter optimization step. After training, the algorithm produces an output for unseen data, so-called test data (rest of the data after training data split), and its performance is validated. In the testing part, some error metrics can be also calculated which are named as testing or generalization error (the difference between real and calculated output for test data). This error is described as the difference between expected and produced output. Briefly, the performance of an algorithm can be determined by two important measures: training and test errors. These important factors address the two major problems in machine learning: under fitting and over fitting. Under fitting happens when the complexity of the algorithm is not adequate to represent the intrinsic structure of the data. There could be several reasons for this: If the model is not strong enough, over-regulated, or has not been trained long enough. This means that the network does not learn relevant patterns in training data. In contrast, over fitting happens when the complexity of the

learning algorithm is high and the data that is provided for training is small. While it is often possible to achieve high accuracy in the training set, what is really desired is to develop models that generalize well to a test set (or data they haven't seen before). But overfitting prevents this. Over fitting can be tackled by incorporating more data whereas under fitting can be solved by increasing the complexity of the algorithm such as adding more hidden layers [6]. The user-defined design parameters such as the number of hidden layers, number of neurons in each layer, and types of activation functions are called hyper-parameters. These parameters should be tuned carefully on a validation set that is different from the training and test data to reach optimum performance.

A human brain has billions of neurons. Neurons are interconnected nerve cells that are involved in the processing and transmitting of chemical and electrical signals. An artificial neuron is the most fundamental unit of a DL algorithm that mimics certain neuron parts, such as dendrite, cell body, and an axon, using simplified mathematical models. A perceptron (the predecessors of artificial neurons) is a neuron unit and shown in Figure 2A. The concept of perceptron was first introduced in 1958 by psychologist Frank Rosenblatt and further refined and analyzed by Minsky and Papert in 1969 [7]. The task of each neuron is taking its input, multiplying it by its weight, and passing the sum from activation function to other neurons. The neural network learns how to classify an input by adjusting its weight according to previous examples. By

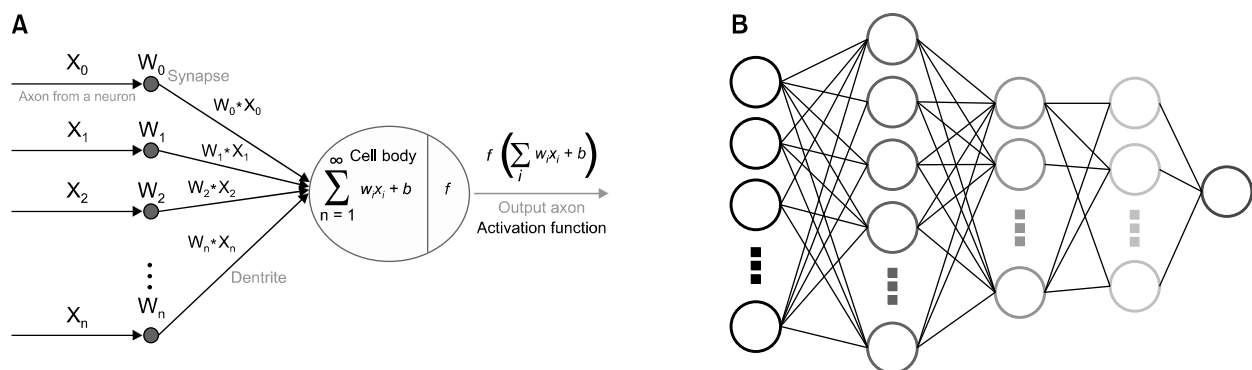
combining these neurons, artificial neural networks (ANNs) are constructed. The aim of training in ANNs is to find the weight values that eventually predict correct labels of the samples provided to the network. Reaching the optimum weight values of the network means that the samples can make generalizations about the event represented by the samples.

DL and neuroscience have recently been intertwined. Neuroscience can help to validate already existing DL techniques and provide rich inspiration for new types of algorithms and architectures. Also, some new techniques developed with DL algorithms can be used in neuroscience to help to understand neuropsychiatric disorders [8].

In the first part of the review, we outline the underlying concepts of DL and we provide brief description of the most common DL architectures used in the field of neuropsychiatry, including ANNs, convolutional neural networks (CNNs), recurrent neural networks (RNNs) and generative adversarial networks (GANs).

### Artificial Neural Networks

One important drawback of perceptrons is their limited number of layers which confines the complexity of the classifier. Furthermore, although linear classifiers provide sufficient performance for many tasks, not all data is linearly separable which makes nonlinear classification a necessity. To alleviate these problems, ANNs with high-level representation with error propagation has been proposed



**Fig. 2.** (A) Perceptron, the smallest part of the artificial neural network (ANN) model, is defined by the linear function  $y = W \cdot x + b$ . In biological neural networks, information from the axon is collected by the dendrites and processed by the cell body to generate electrical pulses and chemical signals. Communication between two different neurons is achieved by means of neurotransmitters in the synapses between the axons and dendrites of two adjacent neurons when the neuron meets the threshold level. Similarly, in ANNs, each input,  $x_i$ , is weighted by  $w_i$  according to its contribution to obtain the final output,  $f(y)$ . The output unit is obtained by passing the weighted sum of the inputs through an activation function. (B) ANN architecture with multiple layers. It has 1 input layer (the first layer), 3 hidden layers (in between layers), and finally 1 output layer (the last layer) with 1 output units.

by Rumelhart *et al.* [9]. ANN consists of several neurons in input, output, and multiple hidden layers. Having multiple hidden layers increases the complexity of the classifier and enables revealing complex patterns inside data. Furthermore, each neuron applies an activation function such as a sigmoid function to the weighted combination of its inputs to establish nonlinearity. Figure 2B shows an example of an ANN architecture with three hidden layers. The number of hidden layers and neurons in each layer are important hyper-parameters that are set by the programmer and affect the overall complexity of the classifier.

The learning stage consists of two parts: a forward pass and back propagation. In the forward pass, an error is calculated with the recent weights. In literature, there are several methods to find optimal solutions to minimize the error term. One of them and arguably the most common method is known as Stochastic Gradient Descent (SGD). In SGD, the gradient of the error term is calculated and the parameters are updated to minimize the error using the training data. This process occurs in the back propagation step which constitutes the main learning process [10]. ANNs are good for general purpose, classification tasks in neuropsychiatry using data from electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), genetic and psychiatric survey data, etc., due to fast trainability, easy implementation, and smaller data set require-

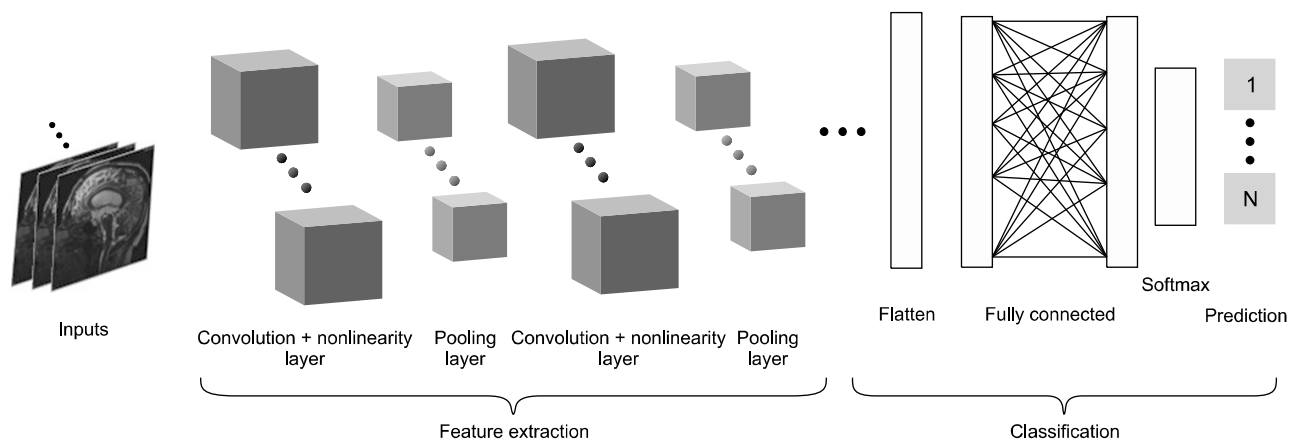
ment compare to other methods [11].

### Convolutional Neural Networks

CNNs are specialized ANNs that apply convolutional kernels in at least one of its layers. CNNs were originally inspired by the primate visual system and explore spatial invariances in the data [2]. A basic CNN architecture is shown in Figure 3.

Various methods have been proposed to improve the performance of CNNs, such as the use of different activation and loss functions, parameter optimization, regularization and restructuring of processing units. In the design of new CNN architectures, these components are increasingly combined in more complex and interconnected ways and even replaced by other more convenient processes. Numerous CNNs have been implemented since the late 1980s to the present day. The first CNN architecture was introduced by Lecun *et al.* [12].

CNN networks can automatically extract patterns from images using filters (kernels) and they need relatively less pre-processing time in comparison to other handcrafted features [13]. In general, medical imaging systems produce three-dimensional (3D) images (MRI, computed tomography [CT]). 3D CNN architectures have been proposed to process these images. In two-dimensional (2D) CNNs, features are computed only from a 2D space (such



**Fig. 3.** Convolutional neural network (CNN). A CNN contains two basic parts: feature extraction and classification. The feature extraction part consists of successive convolutional and pooling layers. A convolutional layer applies convolutional filters called a kernel to the image for exploring low and high-level structures. These structures are obtained by shifting these kernels, so called convolution, in the image with a set of weights. After multiplying the elements of these kernels with the corresponding receiving field elements, a feature map is obtained. These maps are passed through nonlinear activation function (e.g., a rectified linear unit). The task of pooling layer is to reduce the feature map size and the total number of parameters to be optimized in the network. It works by gathering similar information in the neighborhood of the receptive field and find a representative value (e.g., maximum or average) within this local region. Flatten layer converts matrices from the convolution layers into a one-dimensional array for the next layer. Fully connected layer computes the final outputs using back propagation and gradient descent as for standard artificial neural networks.

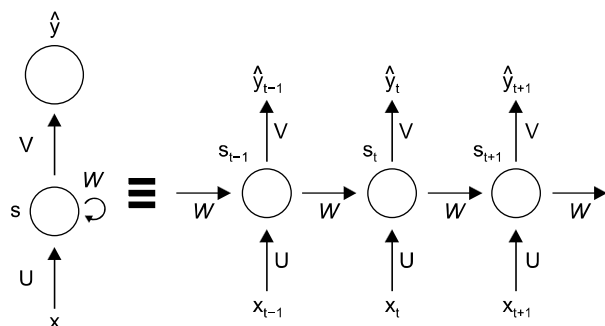
as X-Ray or 2D ultrasound images), whereas in 3D CNNs, they are computed from a 3D volume. New 3D CNN architectures have recently been proposed and implemented in neuroimaging tasks which are indispensable methods for studying neuropsychiatric disorders, and have yielded very promising results compared to other methods. The applications of the above-mentioned algorithms of neuropsychiatry will be discussed in the next section.

### Recurrent Neural Networks

Since the 1990's RNNs have become an important research area. This network is designed to learn sequential or time-dependent patterns in data. In comparison to a standard feed-forward neural network, an RNN architecture also uses connections to form directed cycles. In Figure 4, a simple RNN structure is presented.

In an ordinary feed-forward network, previous predictions are not used for predicting the output. In RNN, however, decisions are made using the previous results due to recurrent connections. Until now, many variants of RNNs have been proposed by researchers. Elman [13] and Jordan [14] networks, known as simple recurrent networks, can be considered as the first publications in the historical development of RNNs.

In RNNs, recurrent connections make minor changes in architecture that raise a dynamic system with many new behaviors. Training of these architectures is difficult than previously-discussed. However, once trained, RNNs can be run forward in time to produce predictions of future outcomes or states. RNNs are widely used in longitudinal studies where observing temporal variations of a signal



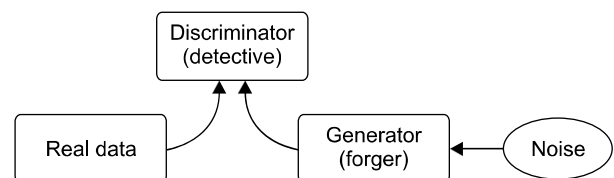
**Fig. 4.** Recurrent neural network: Given architecture has an input layer  $X$ , hidden layer  $S$  and output layer  $\hat{y}$ . In the network,  $X_t$ ,  $\hat{y}_t$ , and  $S_t$  define the current input, output and states respectively.  $U$  and  $W$  are the weights of the relevant layer and  $V$  is the output function.  $S_t$  is calculated using the information from previous state as:  $S_t = f(UX_t + WS_{t-1})$  and,  $\hat{y}_t$  is calculated as:  $\hat{y}_t = V(S_t)$ .

are crucial such as fNIRS and EEG-related tasks.

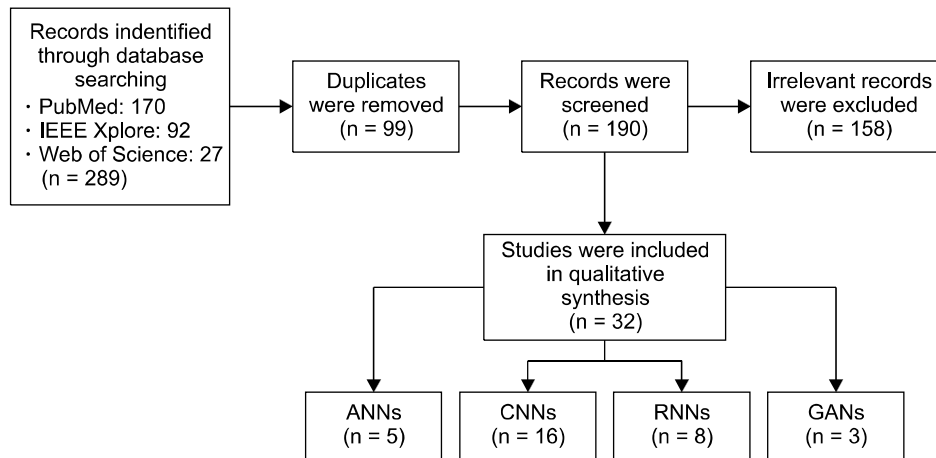
### Generative Adversarial Networks

Goodfellow *et al.* [15] proposed GANs in 2014. GANs consists of two opposing neural networks (generator and discriminator) that learn to create a new data set with the same statistics of the input data and discriminate between the real and the generated data in the training stage. Generator stands for producer network, while discriminator stands for distinctive network. The purpose of the generator is to fit a suitable curve for the distribution of real data and generate new samples. The discriminator is fed with fake and real images, and it produces binary output as fake (0) or real (1) [15]. Goodfellow explains the GANs briefly that while the generator is a 'counterfeiter' team that tries to make tables similar to real ones, the discriminator is like a detective team that tries to understand fake or real ones. The metaphor used by Ian Goodfellow to explain the GANs model is shown in Figure 5.

In adversarial models, the generator's parameters are not directly updated using components from the training data but the gradient components flowing from the discriminator, which provides considerable statistical advantage [15]. GANs can transfer the raw inputs to outputs or serve as a post-processing step to filter images, adversarial training can be used to supply structure consistency and the generator and discriminator parts can be used as a feature extractor or the discriminator part can be used directly as a classifier [16]. Due to mentioned properties, GANs are specifically used for segmentation, reconstruction and classification tasks for different imaging modalities.



**Fig. 5.** The metaphor used by Ian Goodfellow to explain the generative adversarial networks (GANs) model. GANs consists of two different network structures; generator and discriminator networks. While the discriminator network creates new data from a sample database, the discriminator network tries to distinguish between real and fake samples by looking at the data produced by the generator with some noise.



**Fig. 6.** Flow diagram for study selection (modified from Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement). ANNs, artificial neural networks; CNNs, convolutional neural networks; RNNs, recurrent neural networks; GANs, generative adversarial networks.

**Table 1.** Studies using ANN in neuropsychiatry

Reference	Method	Year	Modality	Application	Result
Vyškovský <i>et al.</i> 2016 [17]	ANN	2016	MRI	Schizophrenia classification	Overall accuracy = 68%
Jafri and Calhoun 2006 [16]	ANN	2006	fMRI	Schizophrenia classification	Accuracy = 76%
Fonseca <i>et al.</i> 2018 [18]	ANN	2018	Array collection data	Classification of bipolar and schizophrenia disorders	Accuracy = 90%
Lins <i>et al.</i> 2017 [19]	ANN	2017	Array collection data	Classification of mild cognitive impairment and dementia	Sensitivity = 98% and Specificity = 96%
Narzisi <i>et al.</i> 2015 [20]	ANN	2015	Array collection data	Classification of children with a positive response to TAU	Accuracy = 89.24%

ANN, artificial neural network; MRI, magnetic resonance imaging, fMRI, functional MRI; TAU, treatment as usual.

## APPLICATION OF DEEP LEARNING ALGORITHMS TO NEUROPSYCHIATRIC DISORDERS

To identify previous practices of DL in neuropsychiatric studies of psychiatric or neurological disorders, our search was carried out on 31st January 2020 in various search databases (PubMed, IEEE Xplore, and Web of Science) using the following search terms: (“deep learning” OR “deep architecture” OR “artificial neural network” OR “convolutional neural network” OR “recurrent neural network” OR “generative adversarial network”) AND (neuropsychiatry OR psychiatry OR “psychiatric disease”). A total of 289 articles were reached in the first search and duplicates were excluded. As a next step, we selected studies that focus on DL models for neuropsychiatric research and provided cross-references; this identified a total of 32 articles that were relevant to our review. We organized these studies according to the types

of DL architectures such as ANNs, CNNs, RNNs, and GANs. The strategy that we follow for choosing related articles in this survey is represented in the flow-chart given in Figure 6. These studies are summarized in Tables 1–4 which provides the following information: general type of architecture, type of data used as input (modality), diagnostic groups being investigated, and results as various performance metrics.

### ANNs for Classification of Neuropsychiatric Disorders

Studies using ANN in neuropsychiatry are shown in Table 1. Vyškovský *et al.* [17] used ANN for Schizophrenia spectrum disorder (SZ) classification from MRI scans of 104 subjects. They compared its performance with support vector machine (SVM) classifier and achieved accuracies up to 68% using it together with SVM. Functional magnetic resonance imaging (fMRI) scans were also used for the same task by Jafri and Calhoun [16] achieving around 76% classification accuracy. In a different study

Table 2. Studies using CNN in neuropsychiatry

Reference	Method	Year	Modality	Application	Result
Gupta <i>et al.</i> 2013 [22]	Simple CNN + Sparse Automatic Encoder	2013	sMRI	Classification of MCI and AD	Accuracy = 94.7% for NC vs. AD Accuracy = 86.4% for NC vs. MCI Accuracy = 88.1% for AD vs. MCI
Payan and Montana 2015 [23]	3D CNN	2015	3D MRI	Classification of MCI and AD	Accuracy = 95.4% for NC vs. AD Accuracy = 92.1% for NC vs. MCI Accuracy = 86.8% for AD vs. MCI
Hosseini-asl <i>et al.</i> 2018 [24]	3D CNN with pre-trained 3D convolution automatic encoder	2018	3D sMRI	Classification of MCI and AD	Accuracy = 99.3% for NC vs. AD Accuracy = 94.2% for NC vs. MCI Accuracy = 100% for AD vs. MCI
Wang <i>et al.</i> 2018 [25]	CNN	2018	sMRI	Classification of AD	Accuracy = 97.65% Sensitivity = 97.96% Specificity = 97.35% for NC vs. AD
Duc <i>et al.</i> 2020 [26]	3D CNN	2020	fMRI	Classification of AD	Accuracy = 85.27% for NC vs. AD
Sarraf and Ghassem 2016 [27]	LeNet and GoogleNet	2016	sMRI and fMRI	Classification of AD	Accuracy = 94.32% for NC vs. AD with the fMRI Accuracy = 97.88% for NC vs. AD with the sMRI
Spasov <i>et al.</i> 2018 [28]	3D CNN	2018	sMRI, genetic measures (APOe4) and clinical assessment	Classification of AD	Accuracy = 99% for NC vs. AD Sensitivity = 98% for NC vs. AD Specificity = 100% for NC vs. AD
Liu <i>et al.</i> 2020 [29]	ResNet and 3D DenseNet	2020	sMRI	Classification of MCI and AD	Accuracy = 88.9% AUC = 92.5% for AD vs. NC Accuracy = 76.2% AUC = 77.5% for MCI vs NC
Farooq <i>et al.</i> 2017 [30]	GoogleNet and ResNet	2017	sMRI	Classification of AD, EMCI, and LMCI	Accuracy = 98.88% for GoogleNet Accuracy = 98.01% for ResNet-18 Accuracy = 98.14% for ResNet-152
Korolev <i>et al.</i> 2017 [31]	Plain 3D CNN (VoxCNN) and ResNet with six VoxRes blocks	2017	3D sMRI	Classification of AD, EMCI, and LMCI	Accuracy (VoxCNN) = 79% for NC vs. AD Accuracy (VoxCNN) = 63% for NC vs. LMCI Accuracy (ResNet) = 54% for AD vs. EMCI Accuracy (ResNet) = 80% for NC vs. AD Accuracy (ResNet) = 61% for NC vs. LMCI Accuracy (ResNet) = 56% for AD vs. EMCI
Senanayake <i>et al.</i> 2018 [32]	ResNet, DenseNet, and GoogleNet	2018	3D MR volumes and neuropsychological measure based feature vectors	Classification of MCI and AD	Accuracy = 79% for NC vs. AD Accuracy = 74% for NC vs. MCI Accuracy = 77% for AD vs. MCI
Zou <i>et al.</i> 2017 [33]	3D CNN	2017	Resting-state fMRI signals	Classification of ADHD	Accuracy = 65.67%
Zou <i>et al.</i> 2017 [34]	Multi-modality 3D CNN	2017	fMRI and sMRI	Classification of ADHD	Accuracy = 69.15%
Chen <i>et al.</i> 2019 [35]	3D CNN and 2D CNN	2019	A new form of representation of multi channel EEG data	Detection of personalized spatial-frequency abnormality in EEGs from children with ADHD	Accuracy = 90.29% ± 0.58% AUC = 0.96 ± 0.01
Campese <i>et al.</i> 2019 [36]	SVM, 2D CNN, and three different 3D architectures (VNet, UNet, and LeNet)	2019	2D and 3D sMRI	Classification of SZ and BP	AUC score: 86.30 ± 9.35 using VNet + SVM for Dataset A AUC score: 71.63 ± 12.87 using VNet for Dataset B for binary classification of SZ vs. NC AUC score: 66.43 ± 12.15 using UNet for Dataset A AUC score: 75.52 ± 13.71 using UNet for Dataset B for binary classification of BP vs. NC
Choi <i>et al.</i> 2017 [37]	3D CNN (PD Net)	2017	FP-CIT SPECT	Classification of PD	Accuracy = 96% for the PPMI dataset Accuracy = 98.8% for the SNUH dataset

CNN, convolutional neural network; 3D, three-dimensional; 2D, two-dimensional; ResNet, residual networks; DenseNet, densely connected networks; MRI, magnetic resonance imaging; fMRI, functional MRI; sMRI, structural MRI; EEG, electroencephalography; FP-CIT, dopamine-transporterscintigrafie; SPECT, single photon emission computerized tomography; MCI, mild cognitive impairment; AD, Alzheimer's disease; EMCI/LMCI, early/late mild cognitive impairment; ADHD, attention deficit and hyperactivity disorder; SZ, spectrum disorder; BP, bipolar disorder; PD, Parkinson's disease; NC, normal cognitive; AUC, area under the curve; SVM, support vector machine; PPMI, Parkinsons progression markers initiative; SNUH, Seoul National University Hospital.

**Table 3.** Studies using RNN in neuropsychiatry

Reference	Method	Year	Modality	Application	Result
Petrosian <i>et al.</i> 2000 [38]	RNN	2000	EEG	Prediction of epileptic seizures	Existence of preictal stage in some minutes reported as feasible to predict seizure
Petrosian <i>et al.</i> 2001 [39]	RNN	2001	EEG	Early prediction of AD	Sensitivity = 80% Specificity = 100%
Wang <i>et al.</i> 2018 [40]	LSTM	2018	Array collection data	AD progression prediction	Accuracy = 99% ± 0.0043
Dakka <i>et al.</i> 2017 [41]	LSTM	2017	4D fMRI	Learning invariant markers of schizophrenia disorder	Average accuracy using LSTM = 66.4% Average accuracy using R-CNN = 64.9% Average accuracy using SVM = 57.9%
Kumar <i>et al.</i> 2019 [42]	RNN	2019	CT, MRI, and PET	Classification of dementia, AD, and autism disorders	Dementia Accuracy = 82.8% AD Accuracy = 72.2% Autism Accuracy = 78.2%
	BRNN				Dementia Accuracy = 95.3% AD Accuracy = 89.6% Autism Accuracy = 91.9%
Talathi 2017 [43]	GRU	2017	EEG	Early epileptic seizure detection	Accuracy = 99.6%
Che <i>et al.</i> 2017 [44]	GRU	2017	Parkinson's progression markers initiative (PPMI) challenge dataset	Personalized predictions of Parkinson's disease	Personalized LR RMSE = 0.658 Personalized SVM RMSE = 0.695 Multiclass LR RMSE = 0.719 Multiclass SVM RMSE = 0.742 LSTM RMSE = 0.785 KNN RMSE = 0.957
Yao <i>et al.</i> 2019 [45]	IndRNN	2019	EEG	Classification of epileptic seizure	IndRNN Average accuracy = 87% ± 0.03 LSTM Average accuracy = 84.4% ± 0.02 CNN Average accuracy = 82.9% ± 0.02

RNN, recurrent neural network; LSTM, long-short term memory; BRNN, bidirectional RNN; GRU, gated recurrent unit; IndRNN, independent RNN; EEG, electroencephalography; 4D, four-dimensional; MRI, magnetic resonance imaging; fMRI, functional MRI; CT, computed tomography; PET, positron emission tomography; AD, Alzheimer's disease; CNN, convolutional neural network; RCNN, recurrent CNN; SVM, support vector machine; LR, logistic regression; KNN, K nearest neighbors; RMSE, root mean square error.

**Table 4.** Studies using RNN in neuropsychiatry

Reference	Method	Year	Modality	Application	Result
Truong <i>et al.</i> 2018 [46]	DCGAN	2018	EEG	Seizure prediction	AUC = 80%
Wei <i>et al.</i> 2018 [47]	cGAN	2018	Multimodal MRI	Predicting myelin content	Dice index between ground truth and prediction = 0.83
Palazzo <i>et al.</i> 2017 [48]	LSTM and cGAN	2017	EEG	Reading the mind	Maximum test accuracy = 83.9% for the LSTM-based EEG feature encoder Inception scores is 5.07 and inception classification accuracy is 0.43 for overall

RNN, recurrent neural network; DCGAN, deep convolutional generative adversarial network; cGAN, conditional generative adversarial network; EEG, electroencephalography; MRI, magnetic resonance imaging; AUC, area under the curve; LSTM, long-short term memory.

[18], bipolar disorder (BP) and SZ were classified from normal controls using array collection data of Stanley Neuropathology Consortium databank. They excluded patients over 65 years old and achieved an accuracy of around 90%. ANN's have also been used for mild cognitive impairment (MCI) and Alzheimer's disease (AD) classification [19]. They used a dataset of cognitive tests

(Mini-Mental State Examination, Semantic Verbal Fluency Test, Clinical Dementia Rating and Ascertainig Dementia) from 151 individuals of which 126 having a diagnosis of either dementia of Alzheimer type (n: 56) or MCI (n: 70) and achieved an accuracy higher than 90% with a sensitivity and specificity of 98% and 96%, respectively. Narzisi *et al.* [20] used ANN to explore the variables in-



volved in the positive response to treatment as usual (TAU) in autism, and classified children with a positive response to TAU (reduction in Autism Diagnostic Observation Schedule; Child Behavior Checklist and Parenting Stress Index scores) with 85–90% of global accuracy.

In neuropsychiatry, methods called network-based statistics (NBS) are also used as an alternative to ANN to identify functional or structural connection differences in the data. NBS was first presented by Zalesky *et al.* [21] along with a case-control study that identifies disconnected subnets in chronic SZ patients with resting-state functional MRI data. It was mentioned in this study that the neuroimaging data of NBS can play an important role in network analysis.

### CNNs for Classification of Neuropsychiatric Disorders

The most frequent use of CNNs in these areas draw attention for AD and MCI however, CNNs have been also used in the classification of diseases such as attention-deficit/hyperactivity disorder (ADHD), SZ, BP, and Parkinson's disease (PD). Studies using CNNs to classify these diseases from healthy individuals have used a range of neuroimaging modalities including sMRI, fMRI, resting-state fMRI (rs-fMRI), single-photon emission computed tomography (SPECT) and a combination of different modalities or clinical assessments.

Studies using CNN in neuropsychiatry are shown in Table 2. In one of the early studies, Gupta *et al.* [22] trained a sparse automatic encoder to learn features from natural images, then applied it to sMRI data via a CNN. This method outperforms all previous methods in which learned features were extracted from the Alzheimer's disease neuroimaging initiative (ADNI) dataset. A few years later, Payan and Montana [23] found comparable classification accuracies using features that were learned from 3D sMRI images instead of 2D. This could potentially be explained by the fact that 3D brain images contain more useful patterns for classification. By further expanding 3D CNN, Hosseini *et al.* [24] proposed to predict AD with a deep 3D-CNN that learns the general characteristics of AD biomarkers and could adapt to different domain data sets. There are other researches [25-29] that use subtypes of CNN for the classification of AD or MCI vs. normal cognitive (NC) profile. Recent studies have shown that the functional organization of the brain is dynamic. It can be deduced from the study by Sarraf and Ghassem [27] that

sMRI can be useful to identify patients with MCI and AD.

Some studies classify the ADNI dataset into four different classes as AD, early (E-MCI) and late (L-MCI) stages of MCI, and NC. One of these studies is the deep CNN for multi-class classification developed by Farooq *et al.* [30] using structural MRI images. This framework was constructed using both of the two state-of-the-art CNN models; namely GoogleNet and ResNet. With the presented framework, GoogleNet achieved the highest accuracy. Korolev *et al.* [31] compared two separate network architectures that classify images from the ADNI data set into the above-mentioned four different classes. One of them is the plain 3D CNN model, and the other is a ResNet architecture. They reported that the networks learned to accurately classify AD subjects from the NC, but had difficulty distinguishing them from E-MCI and L-MCI. Senanayake *et al.* [32] inspired by the concepts underlying the ResNet, DenseNet, and GoogleNet architectures, developed a model to classify MRI scans of subjects diagnosed with AD, MCI, and NC.

Zou *et al.* [33] introduced a new 3D CNN architecture to automatically diagnose ADHD using rs-fMRI signals for assisting psychiatrists to diagnose ADHD. They reported that their architecture provided better performance (65.67%) in the ADHD dataset than other studies in the literature. In their later work, Zou *et al.* [34] have suggested combining low-level imaging attributes from both fMRI and sMRI data. With this new architecture, they increase the accuracy of the ADHD dataset to 69.15%. Although CNNs have been generally studied on images obtained from different modalities, they have also been used on non-image data by converting the data into images. For example, Chen *et al.* [35] have converted EEG data into 2D topographic maps by applying an azimuthal equation projection in their studies to identify EEG abnormalities of ADHD children with precise spatial frequency resolution. Later, they applied 3D CNN algorithm on their data and obtained reasonably high performance (accuracy  $90.29 \pm 0.58\%$  and area under curve value  $0.96 \pm 0.01$ ).

Campese *et al.* [36] considered and compared shallow machine learning models, 2D CNN, and three different 3D CNN architectures (VNet, UNet, and LeNet) for the classification of psychiatric disorders, such as SZ and BP. According to their experimental results, 3D CNN models were the most successful. It was concluded that working on the whole 3D structure of the brain improves overall

performances, and spatial information about the position of each voxel is important and could be used to further improve the performance.

Choi *et al.* [37] aimed to develop an automated SPECT interpretation system based on DL for objective diagnosis. Their primary goal was to create a more accurate interpretation system to refine the imaging diagnosis of PD with SPECT. They trained the model using a 3D CNN architecture, namely PD Net and tested it on Parkinson's progression markers initiative (PPMI) and Seoul National University Hospital (SNUH) datasets. They trained their system to classify PD patients with normal controls. In the PPMI dataset, the accuracy values for rater 1 and rater 2 were 90.7% and 84%, respectively, while the accuracy for PD Net was 96% and 98.8% for the SNUH dataset.

### RNNs for Classification of Neuropsychiatric Disorders

RNNs are known as one of the most powerful types of DL algorithms designed to learn underlying patterns of time series data. This power of RNNs, made them favorable tools to use them for diagnosis, prediction and decision support purposes.

Until today, RNNs have been widely used in many biomedical applications. An important part of these studies was presented in the fields of neuropsychiatric disorders and these studies are briefly overviewed in the following sections and Table 3.

In 2000 and 2001, two different studies were presented by Petrosian *et al.* [38,39]. Rather than using extracted features, they used raw EEG signals with RNNs for the first time. In the first study, they predict epileptic seizures from intracranial and extracranial EEG recordings using a simple RNN network. They reported that the presence of the preictal stage in EEG signals could indicate upcoming epileptic seizures. In the latter study, their aim was early recognition of AD with a simple RNN network. Under the eyes-closed condition, they reported 80% sensitivity and 100% specificity. For AD, another study was conducted based on long-short term memory (LSTM) to predict the progression of AD [40]. Using the ability to learn long-term dependencies of LSTMs, Dakka *et al.* [41] presented a comparative study on SZ. In the study, the classification performance of SVM, Region-based Convolutional Neural Networks (R-CNN), and LSTM networks were compared utilizing fMRI data. Results showed that the LSTM network outperformed SVM and produced slightly

better performance than R-CNN (~1%).

Bidirectional long short-term memory (BI-LSTM) networks have also attracted the interest of many researchers in the fields of psychiatry and neuroscience. A comprehensive study about the classification of dementia other than AD, and autism disorders published in 2019 [42]. This study covered a comparative analysis of simple RNN and BI-LSTM network using MRI, CT, and positron emission tomography images for these three disorders. Due to the ability to learn from both past and future inputs, the outcome of the study showed that BI-LSTM achieved around 13.6% higher accuracy than simple RNN for all disorders.

In literature, gated recurrent unit (GRU)-based networks were also implemented to predict epileptic seizures [43] and PD [44] accurately. Researchers proposed a GRU network for seizure detection using publicly available data. One of these studies was published in 2019 [45] in predicting epileptic seizures. In this study, independently RNN (IndRNN) has applied the first time the seizure/non-seizure classification. In the results, compared to the other two common algorithms (LSTM and CNN) IndRNN provided the best accuracy.

### GANs in Neuroscience

Nowadays, GANs are used in many areas, such as image conversion (low-resolution to multi-resolution), image segmentation, reconstruction, de-noising, registration, classification, and completing the missing parts of an image. Additionally, GANs are used with medical images such as MRI to classify neuropsychiatric disorders such as multiple sclerosis (MS).

Since GANs are relatively new in the field, studies employing GANs have a few examples in the neuroscience literature. Studies using GAN in neuropsychiatry are shown in Table 4. One of these studies was published in 2018, by Truong *et al.* [46] on predicting seizure. In this study, a deep convolutional GAN (DCGAN) was used to reveal the underlying relevant structures from EEG signals, and results were investigated in three different scenarios on two datasets to observe the system's overall performance. Results showed that, compared to the fully supervised CNN, DCGAN achieved approximately 6% and 12% lower performance for two datasets.

In 2018, researchers studied on MS to learn myelin content using adversarial training [47]. They proposed a Sketcher-Refinery GANs which consists of two condi-

tional GANs (cGANs) to predict the myelin content from multimodal MRI. Using this method, the ability to predict myelin content at the voxel level was evaluated. The results of the evaluation concluded that demyelination at the lesion sites and the myelin content in normal-appearing white matter could be predicted with high accuracy.

In another study, Palazzo *et al.* [48] proposed a deep network model using L-STM and cGAN on reading the mind. They aimed to generate the picture shown in the subject with cGAN after removing the distinctive features of the picture using L-STM from the subject's EEG signals. The resulting images are not the same but can most probably match the images that the user is looking at.

## DISCUSSION

Detecting and making differential diagnosing of neuropsychiatric disorders at their early stages has been a challenging problem. DL algorithms provide highly accurate, generalizable solutions for such problems compared to traditional approaches. Different from conventional statistical methods, DL algorithms do not require explicit assumption about each parameter and its distribution and use optimization techniques, in particular gradient descent [10], to find the relevant parameters together with their appropriate value. Considering their advantage of finding and optimizing hundreds or even millions of relevant parameters rather than using prior, explicit assumptions, DL provides considerable advantages compared to statistical approaches in terms of revealing intrinsic patterns for performing diagnosis and prognosis research in neuropsychiatry.

Different DL algorithms are used depending on input data and the task. ANNs are among the first DL architecture and used for general classification purposes in neuropsychiatry. In contrast, CNNs are preferred for neuroimaging studies since their convolutional layers extract their own image-related features with the convolutional kernels. In addition to images, CNNs can also be used with one-dimensional (1D) sequential data with 1D convolutional kernels. However, CNNs explore spatial features and might lose temporal patterns in the data. RNNs can exploit long-term temporal information due to their architectural state behavior (memory). Hence, RNNs are generally used for analyzing temporal and sequential data in neuropsychiatric research. GANs are relatively new

compared to other architectures and have just begun to take place in the neuropsychiatric research.

Besides many advantages, DL algorithms also have a number of important limitations. DL-based classification algorithms generally require very large data sizes compared to typical sample sizes collected in neuropsychiatric studies [2]. Areas where DL typically outperforms other ML methods, and shallow networks such as image recognition or speech analysis may have larger databases [49]. Training models with many parameters on small sample sizes pose a serious challenge to find solutions that will generalize well to the population [50], so researchers continue to turn to traditional ML applications due to their limited sample sizes. DL algorithms have also some limitations such as the black-box problem, requiring large training set, selecting an appropriate network, highly complex, and intractable calculations between layers so called black-box problem, and need for high computational power.

**Black-box problem:** Since feature extraction is performed automatically in DL algorithms, why network performed well or why the modified network failed cannot be fully explained. This problem can prevent the researchers from understanding causal relations in neuropsychiatric disorders [51].

**Data requirement:** When the number of training samples is insufficient, the network cannot learn underlying hidden patterns causing over fitting. Hence, DL algorithms require large amounts of data that are hard to collect in many neuropsychiatric experiments [52].

**Architecture selection:** There are no networks that provide the best results for any problem, so different algorithms have to be tried, and this complicates the network selection.

**Computational power:** DL algorithms optimize the large amount of parameters demanding huge processing load. Recently, cloud computing has been used to train large networks on platforms such as Google Colab and Amazon Web Services.

## FUTURE ASPECTS

It is very likely that the successes of DL algorithms in neuropsychiatric disorders will likely keep its rapid growth in the near future. New architectures such as capsule networks [53] and new hardwares were designed specifically

for DL architectures to provide promising solutions to the drawbacks of DL algorithms.

## CONCLUSION

In this paper, we provide a broad overview of DL algorithms that are used in the field of neuropsychiatry. Considering a wide range of different architectures, we focus particularly on the four different types of DL algorithms that have been used recently for analyzing neuropsychiatric disorders. In addition to providing an overview, our aim is also to guide researchers in choosing the proper DL architecture for solving their problems in neuropsychiatry and provide a perspective for future research.

### ■ Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

### ■ Author Contributions

Conceptualization: Gamze Erzin, Gorkem Saygili, Burcin Colak, Yasemin Hosgoren Alici. Data acquisition: Gokhan Guney, Busra Ozgode Yigin, Gamze Erzin, Necdet Guven. Formal analysis: Busra Ozgode Yigin, Gokhan Guney, Gamze Erzin, Gorkem Saygili. Supervision: Gorkem Saygili, Gamze Erzin, Burcin Colak, Yasemin Hosgoren Alici. Writing – original draft: Busra Ozgode Yigin, Gokhan Guney, Gamze Erzin, Necdet Guven, Yasemin Hosgoren Alici, Burcin Colak, Gorkem Saygili. Writing – review & editing: Busra Ozgode Yigin, Gokhan Guney, Gamze Erzin, Necdet Guven, Yasemin Hosgoren Alici, Burcin Colak, Gorkem Saygili.

### ■ ORCID

Gokhan Guney <https://orcid.org/0000-0003-0522-5877>  
 Busra Ozgode Yigin <https://orcid.org/0000-0003-4803-5504>  
 Necdet Guven <https://orcid.org/0000-0001-7181-7876>  
 Yasemin Hosgoren Alici <https://orcid.org/0000-0003-3384-8131>  
 Burcin Colak <https://orcid.org/0000-0002-1691-2886>  
 Gamze Erzin <https://orcid.org/0000-0001-8002-5053>  
 Gorkem Saygili <https://orcid.org/0000-0002-9049-2138>

## REFERENCES

1. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A,

Jones-Hagata LB, et al. *Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry* 2015;72:305-315.

2. Durstewitz D, Koppe G, Meyer-Lindenberg A. *Deep neural networks in psychiatry. Mol Psychiatry* 2019;24:1583-1598.
3. LeCun Y, Bengio Y, Hinton G. *Deep learning. Nature* 2015; 521:436-444.
4. Goodfellow I, Bengio Y, Courville A. *Deep learning. Cambridge:MIT Press;2016.*
5. Thomas P. *Semi-supervised learning by Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien (review). IEEE Trans Neural Netw* 2009;20:542.
6. Karsoliya S. *Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. Int J Eng Trends Technol* 2012;3:714-717.
7. Minsky M, Papert SA. *Perceptrons: an introduction to computational geometry. Cambridge:MIT Press;1969.*
8. Gonzalez RT, Riascos JA, Barone DAC. *How artificial intelligence is supporting neuroscience research: a discussion about foundations, methods and applications. In: Barone D, Teles E, Brackmann C, editors. LAVCN 2017: Computational neuroscience. Cham:Springer;2017. p.63-77.*
9. Rumelhart DE, Hinton GE, Williams RJ. *Learning internal representations by error propagation. California:University of California;1985 Sep. Report No.: 49.*
10. Amari S. *Backpropagation and stochastic gradient descent method. Neurocomputing* 1993;5:185-196.
11. Kocyigit Y, Alkan A, Erol H. *Classification of EEG recordings by using fast independent component analysis and artificial neural network. J Med Syst* 2008;32:17-20.
12. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. *Backpropagation applied to handwritten zip code recognition. Neural Comput* 1989;1:541-551.
13. Elman JL. *Finding structure in time. Cognit Sci* 1990;14:179-211.
14. Jordan MI. *Serial order: a parallel distributed processing approach. Adv Psychol* 1997;121:471-495.
15. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. *Generative adversarial nets. In: Neural Information Processing Systems 2014; Dec 8-13, 2014; Montreal, QC, Canada. Poster.*
16. Jafri MJ, Calhoun VD. *Functional classification of schizophrenia using feed forward neural networks. Conf Proc IEEE Eng Med Biol Soc* 2006;Suppl:6631-6634.
17. Vyškovský R, Schwarz D, Janoušová E, Kašpárek T. *Random subspace ensemble artificial neural networks for first-episode schizophrenia classification. In: 2016 Federated Conference on Computer Science and Information Systems (FedCSIS); Sep 11-14, 2016; Gdansk, Poland.*
18. Fonseca MB, de Andrades RS, de Lima Bach S, Wiener CD, Oses JP. *Bipolar and schizophrenia disorders diagnosis using artificial neural network. Neurosci Med* 2018;9:209-220.
19. Lins AJCC, Muniz MTC, Garcia ANM, Gomes AV, Cabral RM, Bastos-Filho C. *Using artificial neural networks to select the*

- parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. *Comput Methods Progr Biomed* 2017;52:93-104.
20. Narzisi A, Muratori F, Buscema M, Calderoni S, Grossi E. Outcome predictors in autism spectrum disorders preschoolers undergoing treatment as usual: insights from an observational study using artificial neural networks. *Neuropsychiatr Dis Treat* 2015;11:1587-1599.
  21. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. *Neuroimage* 2010;53:1197-1207.
  22. Gupta A, Ayhan MS, Maida AS. Natural image bases to represent neuroimaging data. *Proc 30th International Conference on Machine Learning* 2013;28:987-994.
  23. Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *arXiv. 1502.02506 [Preprint]. 2015 [cited 2020 Nov 2]. Available from: <https://arxiv.org/abs/1502.02506>.*
  24. Hosseini-Asl E, Ghazal M, Mahmoud A, Aslantas A, Shalaby AM, Casanova MF, et al. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci (Landmark Ed)* 2018;23:584-596.
  25. Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J Med Syst* 2018;42:85.
  26. Duc NT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B. 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. *Neuroinformatics* 2020;18:71-86.
  27. Sarraf S, Ghassem T. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv. 070441 [Preprint]. 2016 [cited 2020 Nov 2]. Available from: <https://doi.org/10.1101/070441>.*
  28. Spasov SE, Passamonti L, Duggento A, Lio P, Toschi N. A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. *Annu Int Conf IEEE Eng Med Biol Soc* 2018;2018:1271-1274.
  29. Liu M, Li F, Yan H, Wang K, Ma Y, Alzheimer's Disease Neuroimaging Initiative, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *Neuroimage* 2020;208:116459.
  30. Farooq A, Anwar S, Awais M, Rehman S. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *2017 IEEE International Conference on Imaging Systems and Techniques (IST); Oct 18-20, 2017; Beijing, China.*
  31. Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); Apr 18-21, 2017; Melbourne, VIC, Australia.*
  32. Senanayake U, Sowmya A, Dawes L. Deep fusion pipeline for mild cognitive impairment diagnosis. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); Apr 4-7, 2018; Washington, DC, USA.*
  33. Zou L, Zheng J, McKeown MJ. Deep learning based automatic diagnoses of attention deficit hyperactive disorder. In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP); Nov 14-16, 2017; Montreal, QC, Canada.*
  34. Zou L, Zheng J, Miao C, McKeown MJ, Wang Z. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access* 2017;5:23626.
  35. Chen H, Song Y, Li X. Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. *J Neural Eng* 2019;16:066046.
  36. Campese S, Lauriola I, Scarpazza C, Sartori G, Aiolfi F. Psychiatric disorders classification with 3D convolutional neural networks. In: *Oneto L, Navarin N, Sperduti A, Anguita D, editors. Recent advances in big data and deep learning. Cham:Springer;2019.*
  37. Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *Neuroimage Clin* 2017;16:586-594.
  38. Petrosian A, Prokhorov D, Homan R, Dasheiff R, Wunsch DC. Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Neurocomputing* 2000;30:201-218.
  39. Petrosian AA, Prokhorov DV, Lajara-Nanson W, Schiffer RB. Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG. *Clin Neurophysiol* 2001;112:1378-1387.
  40. Wang T, Qiu RG, Yu M. Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks. *Sci Rep* 2018;8:9161.
  41. Dakka J, Bashivan P, Gheiratmand M, Rish I, Jha S, Greiner R. Learning neural markers of schizophrenia disorder using recurrent neural networks. *arXiv. 1712.00512 [Preprint]. 2017 [cited 2020 Nov 2]. Available from: <https://arxiv.org/abs/1712.00512>.*
  42. Kumar PSJ, Yuan Y, Yung Y, Hu W, Pan M, Li X. Bi-directional recurrent neural networks in classifying dementia, Alzheimer's disease and autism spectrum disorder. In: *Kumar PSJ, editor. The art of fixing Alzheimer's disease. Pittsburgh:Dorrance. Publishing Co.;2019. p.4-51.*
  43. Talathi SS. Deep Recurrent Neural Networks for seizure detection and early seizure detection systems. *arXiv. 1706.03283 [Preprint]. 2017 [cited 2020 Nov 2]. Available from: <https://arxiv.org/abs/1706.03283>.*
  44. Che C, Xiao C, Liang J, Jin B, Zho J, Wang F. An RNN architecture with dynamic temporal matching for personalized predictions of Parkinson's disease. In: *Proceedings of the 2017 SIAM International Conference on Data Mining; Apr 27-29, 2017; Houston, TX, USA.*

45. Yao X, Cheng Q, Zhang GQ. *A novel independent RNN approach to classification of seizures against non-seizures*. arXiv. 1903.09326 [Preprint]. 2019 [cited 2020 Nov 2]. Available from: <https://arxiv.org/abs/1903.09326>.
46. Truong N, Kuhlmann L, Bonyadi M, Kavehei O. *Semi-supervised seizure prediction with generative adversarial networks*. arXiv. 1806.08235 [Preprint]. 2018 [cited 2020 Nov 2]. Available from: <https://arxiv.org/abs/1806.08235>.
47. Wei W, Poirion É, Bodini B, Durrleman S, Ayache N, Stankoff B, et al. *Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training*. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical image computing and computer assisted intervention- MICCAI 2018*. Cham:Springer;2018. p.514-522.
48. Palazzo S, Spampinato C, Kavasidis I, Giordano D, Shah M. *Generative adversarial networks conditioned by brain signals*. In: *2017 IEEE International Conference on Computer Vision (ICCV); Oct 22-29, 2017; Venice, Italy*.
49. He K, Zhang X, Ren S, Sun J. *Delving deep into rectifiers: surpassing human-level performance on ImageNet classification*. In: *2015 IEEE International Conference on Computer Vision (ICCV); Dec 7-13, 2015; Santiago, Chile*.
50. Whelan R, Garavan H. *When optimism hurts: inflated predictions in psychiatric neuroimaging*. *Biol Psychiatry* 2014; 75:746-748.
51. Zednik C. *Solving the black box problem: a normative framework for explainable artificial intelligence*. *Philos Technol* 2019. doi: <https://link.springer.com/article/10.1007/s13347-019-00382-7>.
52. Camilleri D, Prescott T. *Analysing the limitations of deep learning for developmental robotics*. In: *Biomimetic and Biohybrid Systems. 6th International Conference, Living Machines 2017; July 26-28, 2017, Stanford, CA, USA*.
53. Sabour S, Frosst N, Hinton GE. *Dynamic routing between capsules*. In: *31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017; Long Beach, CA, USA*.